

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ
ОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ КОМПЬЮТЕРНЫХ НАУК

МАТЕМАТИЧЕСКИЕ
СТРУКТУРЫ
И
МОДЕЛИРОВАНИЕ

Выпуск 17

Омск 2007

Журнал «Математические структуры и моделирование». – Омск:
Омск. гос. техн. ун-т, 2007. – Вып. 17. – 125 с.

Редакционная коллегия

Главный редактор

А.К. Гуц

д-р физ.-мат. наук, профессор

Н.Ф. Богаченко

к-т физ.-мат. наук

Е.В. Палешева

к-т физ.-мат. наук

Художественное оформление

В.В. Коробицын

Адрес научной редакции

Россия, 644053, Омск - 53, ул. Грозненская, 11

Омский государственный университет

факультет компьютерных наук

E-mail: guts@univer.omsk.su

msm@cmm.univer.omsk.su

© Омский госуниверситет, 2007

МАТЕМАТИЧЕСКИЕ СТРУКТУРЫ И МОДЕЛИРОВАНИЕ

В журнале публикуются статьи, в которых излагаются результаты исследований по фундаментальной и прикладной математике, теоретической физике и размышления, касающиеся окружающей нас природы и общества.

Публикуются также статьи по информационным технологиям, компьютерным наукам, защите информации, философии и истории математики.

Объекты исследования должны быть представлены в форме некоторых математических структур и моделей.

Журнал является реферируемым. Рефераты статей публикуются в Реферативном журнале и в журналах «Zentralblatt für Mathematik» (Германия) и «Mathematical Reviews» (США).

Электронная версия журнала представлена в сети «Интернет» по адресу:

<http://cmm.univer.omsk.su>

Журнал издается на коммерческие средства факультета компьютерных наук Омского государственного университета.

Наш E-mail:

msm@cmm.univer.omsk.su

Подробную информацию можно найти на Web-сервере:

<http://cmm.univer.omsk.su>

СОДЕРЖАНИЕ

Некто А.Г. Гринь. *Про декана Гуца, удалого молодца*..... 5

Фундаментальная математика

А.Г. Гринь. *О минимальном условии слабой зависимости в центральной предельной теореме для зависимых случайных величин* 13

В.В. Коробицын, В.Б. Маренич, Ю.В. Фролова. *Исследование поведения явных методов Рунге-Кутты при решении систем обыкновенных дифференциальных уравнений с разрывной правой частью* 19

Е.А. Мещеряков. *Классификация плоских полных строго причинных лоренцевых многообразий малых размерностей*..... 26

Прикладная математика и моделирование

Е.В. Ренжина, А.К. Гуц. *Прогнозирование поведения работающих на рынке труда омской области* 40

Теоретическая физика

А.К. Гуц. *Квантовое рождение физической реальности и математическое описание осознания* 47

Ю.А. Лебедев. *Нелинейные семантические аспекты квантовомеханической концепции «соотнесенных состояний» Х. Эверетта и перспективы развития эвереттики*..... 53

Информационные технологии

С.С. Ефимов. *Обзор методов распараллеливания алгоритмов решения некоторых задач вычислительной дискретной математики* 72

Н.С. Иванова. *Стратегия развития электронного документооборота промышленного предприятия*..... 94

О.В. Паппинин. *Оптимизация запросов к базам данных* 100

А.Н. Романов. *Комбинирование методов при построении интеллектуальных систем* 108

продолжение на след.странице



Защита информации

Н.Ф. Богаченко. *Применение теоретико-автоматных моделей в криптографии* 112

Д.Н. Лавров, Е.А. Первушин. *Проверка паролей на надежность с помощью модифицированного фильтра Блума* 121

ФАКУЛЬТЕТ

КОМПЬЮТЕРНЫХ НАУК

Омского государственного университета

Факультет создан 1 мая 2003 г. на базе отделения компьютерных наук, открытого 1 апреля 2001 года при физическом факультете.

Первый набор студентов был осуществлен в 2001 году.

Кафедры и лаборатории. На факультете открыты:

- кафедра информационной безопасности;
- кафедра кибернетики;
- кафедра вычислительных систем;
- научно-исследовательская лаборатория программного обеспечения и компьютерных сетей;
- научно-исследовательская лаборатория социо-кибернетики.
- учебно-научный центр переподготовки и повышения квалификации в области защиты информации и информационной безопасности;
- локальная сетевая академия CISCO;
- академия SUN;
- академия Oracle.

Обучение по специальностям:

1) 075200 – Компьютерная безопасность / анализ безопасности компьютерных систем. Квалификация «математик». Срок обучения 5 лет 6 месяцев. Выпуск в феврале.

2) 220100 – Вычислительные машины, комплексы, системы и сети. Квалификация «инженер». Срок обучения 5 лет. Имеются очно-заочное отделение, а также сокращенная форма обучения (очно – 3 года, заочно – 3,5 года).

Телефон (факс): (+7 3812) 64-83-11

ПРО ДЕКАНА ГУЦА, УДАЛОГО МОЛОДЦА

Некто А.Г. Гринь

Подражание Леониду Филатову

Действующие лица: Идейный Вдохновитель, Студенческий Пастух, Завхоз, Профессор, Лицо Непальской Национальности, Надежда Факультета.

У деканских у ворот
Собрался честной народ.
Токмо не на панихиду,
А совсем наоборот!
Ента антеллекта рать
Вознамылилась решать
Как своёйного декана
С юбилеем поздравлять!

Идейный Вдохновитель:
Энто как же ж, вашу мать,
Нам декана поздравлять?
Чай, не Хранция какая,
Штоб открыточки писать!
Хто не понял - мы in Russia!
Што предложите, мать вашу?
Я полезных нциатив
Никогда не супротив!
Ить за все потом ответит
Предложивший коллефтив!

Студенческий Пастух:
Нету никаких идей!
Говорят, не пьеть, злодей!
Как поздравить, штоб по-русски,
Но без водки и сельдей?

Завхоз:

Дык ить нету ни сельдей,
Ни бельдюг, ни стерлядей!
Даже кильки тошшей нету,
Штоб порадовать людей!
Был тут как-то юбилей,
Хто пришел - тому налей!
А пришли аж аллигархи,
Што помельче и наглей.
Дык вот энти аллигархи,
Погулявши пару дней
Все сожрали на халяву
Окромя токмо гвоздей!
Помнишь был «Педигри Пал»?
Вот тогда же и пропал!
Ить под пиво все едино -
И што Пал, и што не Пал!

Лицо Непальской Национальности:

Енто што за ритуал -
Чуть што - сразу на Непал!
Али чукчей тебе мало,
Штоб непальцев тут трепал!

Идейный Вдохновитель:

Чукчу ты не трожь, студент!
Он фольклорный алимент!
Чуть чаво - он сразу в харю
Фицияльный документ!
Даже мент его не тронет!
Ты ж, студент, ишшо не мент!

А уж чукча Абрамович
Кабы токмо захотел-
Енту вашу ... Джомолунгму
Приобрел б и не вspotел!
Глядя на Вершину Мира
Из окна своей хватиры
В стольном граде Анадыре
Он сказал бы: «как-никак, а
Удалася жизнь, однако!»

Што касаето непальцев.

Я, прожимши стоко лет,
Убедился, што непальцев
Вообще в природе нет!
Обошел я белый свет:
Бар, буфет и туалет.

И не видел я непальцев
Ни вчера и ни в обед!

Лицо Непальской Национальности:

Я б смолчал, глядишь, но пиш!
Разве ж с вами промолчишь,
Коль в лице моем поганют
Государственный престиж!

Идейному Вдохновителю:

О тебе, о подлеце,
Слава аж в Череповце!
Ты всему народу в душу
Наплевал в моем лице!

Дык вот, штоб ты, дядя, знал!

Мудрый лама указал:
Нас, непальцев, в мире стоко,
Што никто и не считал!
Што гласит наш основной
Государственный устой?
Ить его мудрее нет:

Хто произведен на свет

а) *непальцем*

б) *непалкой*

Тот - непалец! Потому
Што б там в пашпорте кому
Не вписали - хошь «хранцуз»,
Хошь «индус», хошь «бялорус» -
Наш он! И иди в Тибет,
Иль в буфет, иль в туалет -
Всюду наши! Ваших нет!

Идейный Вдохновитель:

Енто где же ты, злодей,
Набрался таких идей,
Штоб записывать в непальцы
Всех порядошных людей?
Я вот думаю, зараза,
Ты не засланный ли к нам?
Мы таких засланцев сразу
В ФСБ иль по мордам!
Но поскольку я спокон
Чту порядок и закон -
Гони мне сто евров штрафа -
И пошел отседа вон!

Таперь слушайте сюда
Дамы али господа!
Чем производить непальцев
Мы узнали. Это да!
Токмо мы ж хотели знать
Чем декана поздравлять?
Понимаю, что не тем же,
Растудыть едрена мать!

Профессор:

За бугром, слышал я вот,
Каждый год, аль через год
Аккурат перед банкетом
Вумный содится народ
И гундит о том и энтот,
Не беря ни капли в рот!
Конхеренцией зовется
У буржуев энтот сброд.
Вот бы нам такую здесь
Конхеренцию провести!
Пусть все скажут, что декан наш
В мировом масштабе есть!
И годится ли в подметки
Ему Билл, простите, Гейтс?

(Возможные ответы:

- а) да, вот в подметки и годится;
б) нет, и в подметки не годится)*

Пусть расскажут без воды
Про евойные труды.
Хорошо ли ему плотят
Забугорные фонды?
И в народное хозяйство
Што вложил он и куды?
В обчем, пусть, как на духу
Ему скажут хто есть who?

Студенческий Пастух:

Ты даешь, едрена вошь!
Щас! Все скажут, што ты хошь!
Где ж таких вот камикадзей
Хоть полштуки ты найдешь?

Хто хотить сказать *ему*
Who is *он* и почему -
Вишь вон двери дьяконата -
Заходи по одному!
Там у вас в момент наступит
Просветление уму!

Но вот на худой конец
У меня есть молодец!
Он надежда факультета -
Шибко грамотный, стервец!

Вот нехай он соберет
Ентот ... как его ... народ!
Все расскажет про декана...
Все не надо? ... Тогда вот:
Мол декан с утра до ночи ...
Познавал ... и ... созидал...
Ай, пусть брешет што захочет,
Лишь бы правду не сказал!

Входит **Надежда Факультета**
Студенческий Пастух:

Ты чаво такой смурной?
Што причиной, хто виной?
Аль в навучной библиотеке
Повредился головой?
Аль в студенческой едальне
Суши подали не то?
Али в нашем гандеропе
Хуже выдали пальто?
Аль случился аморал?
Али СПИДом захворал?
Али взял, погорячимшись,
Неберущийся нтеграл?
Докладай без всяких врак
Отчего на сердце мрак!
Мы желаем знать подробно:
Хто, куды, чаво и как!

Надежда Факультета:

Ох, робяты, просто шок!
Я тут выдумал стишок!!!
Ентото на факультете
Не случалося ишшо!
Што? У Гриня был грешок?
Понимаю ваш смешок!
К классикам не приبلудился
Я, как он, пока ишшо.
А потом какой скандал!
Как вчера я низко пал!
Ведь не сразу отличил я
От фрактала мартингал!

Студенческий Пастух:

Ты, скажи, у нас на кой
Шибко грамотный такой?
А декан, вон, не поздравлен,
Дико борется с тоской!
Токмо тщетно тщится он
Обрести покой и сон,
Ить в заботах о студенте
День и ночь проводит он!
Утром пробует абсент -
Сразу мысль: «а как студент?»
И икра не лезет в глотку,
От омаров кайфу нет!
Ночью встанет в туалет -
Снова: «а вот как студент?»
Так и мается, сердешный,
Без продыху много лет!
Ты, надежда, соберешь
Завтра всех, кого захошь,
И расскажешь, чем декан наш
Так хорош, . . . и сяк хорош!

Надежда Факультета:

Постараюсь. Докажу.
Оправдаю. Отслужу.
Как в гестапе на допросе
Все в подробностях скажу!
Но у нас же рынок, вроде,
А, лучшей сказать, базар!
Без "подмазки" не проходит
Положительный пиар!
Разрешенью надо - дай!
Помещенью надо - дай!
Демонстранты - енти тоже -
Чуть чаво - подымут хай!
Им оранжевые польты
Как с майдана подавай!

Завхоз:

Тю ты, господи еси!
Вот всегда так на Руси!
Польт на всех не напасешься,
Лучше даже не проси!

Надежда Факультета:

А уж прессу и TV
И не пробуй - не зови!
Штоб снять крышку с объектива -
Штука баксов! Се ля ви!

Вот тут репу и чеши -
Где такие взять шиши!
Шибче, чем прокуратура
Спонсера тогда ишши!
А коль спонсер не проплатит,
Растуды его в качель,
Ну тогда - кому на паперть,
А кому - и на панель!

Разве што из-за кордона
Вдруг подмогут? Вот бы каб
Нам подбросил из Лондона
Пару миллионов БАБ!

Студенческий Пастух:
Скоко баб??? Штоб мы так жили!
Ох, студент на енто слаб!
Где бы что не говорили -
Все одно сведет на баб!

Идейный Вдохновитель:
А ить прав же он, собака!
Хто щас даром будет рад
Ублажать электорат?
Время рыцарев прошло,
Люди гибнут за бабло!
Материализм, однако!

Тут все знают - я всю жисть
Злостный материалист!
Нет ни часу и ни дня
Штоб не вспомнил мать я!

Но што толку вспоминать!
Вон, уже давно за полночь,
Но кто сможет, вашу мать,
Хошь одну, простите, сволочь,
Окромя меня, назвать,
Хто бы знал бы как декана
С хеппибёздьем поздравлять!

Я бы тоже бы не знал бы,
Но я вумный - нету сил!
Я пошел, и я спросил!
И виновник торжества
Раскололся, как дрова!
Долго тихим теплым словом
Поминал свой юбилей,
Кучерявым жжа глаголом
Уши и сердца людей!

Я ж пристойными словами
Оглашаю суть ответа.
Хошь его и так мы знали.
Лучший, мол, подарок - это
Штобы члены факультета
Все, как лошади, пахали,
Планы перевыполняли,
День и ночь статьи клепали.
Так студентов обаяли,
Штоб и через двадцать лет
Всех рвало неудержимо
На родимый факультет!

Штоб «Неписающий мальчик»,
Што у входа возведен,
Как сокровище культуры
Был в ЮНЕСКО утвержден!

Ну што, члены факультета,
Мы могём сказать на это?
Поздно! - мógем мы сказать!
Поздно, мы уже поэму
Вам успели накропать!
Ту, в которой спросили:
«Енто как же ж, вапу мать,
Нам декана поздравлять?»
И, штоб глыбже то понять,
Надо снова все читать!

О МИНИМАЛЬНОМ УСЛОВИИ СЛАБОЙ ЗАВИСИМОСТИ В ЦЕНТРАЛЬНОЙ ПРЕДЕЛЬНОЙ ТЕОРЕМЕ ДЛЯ ЗАВИСИМЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

А.Г. Гринь

In this article is obtained a minimal in a certain sense conditions of weakly dependence, which make it possible to prove a central limit theorem for sums of dependent random variables.

Пусть $\{\xi_n\}$ – последовательность случайных величин. Будем писать $\xi \stackrel{d}{=} \eta$, $\xi_n \xrightarrow{d} \eta$ и $\xi_n \stackrel{d}{\sim} \eta_n$ в случаях, когда, соответственно, распределения ξ и η совпадают, $\{\xi_n\}$ сходится к η по распределению и когда последовательности $\{\xi_n\}$ и $\{\eta_n\}$ слабо эквивалентны (см., например, [1, § 28.1]). Через $\hat{\xi}_1, \dots, \hat{\xi}_n$ будем обозначать независимые случайные величины, такие, что $\hat{\xi}_k \stackrel{d}{=} \xi_k, k = 1, 2, \dots, n$. Слабая эквивалентность равносильна поточечной сходимости разности характеристических функций величин $\{\xi_n\}$ и $\{\eta_n\}$ к нулю при $n \rightarrow \infty$ [1, с. 393]. Введем еще некоторые обозначения. Пусть $S_{m,n} = \sum_{j=m}^n \xi_j$, $S_{m,n}^{+k} = \sum_{j=m}^n \xi_{j+k}$, $S_n = S_{1,n}$, $S_n^{+k} = S_{1,n}^{+k}$, а если $\mathbf{E}\xi_n^2 < \infty$, $n = 1, 2, \dots$, то обозначим $\sigma_{k,n}^2 = \mathbf{D}S_{k,n}$, $(\sigma_{m,n}^{+k})^2 = \mathbf{D}S_{m,n}^{+k}$, $\sigma_n^2 = \mathbf{D}S_n$, $(\sigma_n^{+k})^2 = \mathbf{D}S_n^{+k}$. Через $\mathcal{N}(0, 1)$ обозначим случайную величину, имеющую стандартное нормальное распределение.

Если

$$\frac{S_n - \mathbf{E}S_n}{\sigma_n} \xrightarrow{d} \mathcal{N}(0, 1), \quad n \rightarrow \infty,$$

то будем говорить, что к последовательности $\{\xi_n\}$ применима центральная предельная теорема, а если при любой последовательности натуральных чисел $k = k(n)$

$$\frac{S_n^{+k} - \mathbf{E}S_n^{+k}}{\sigma_n^{+k}} \xrightarrow{d} \mathcal{N}(0, 1), \quad n \rightarrow \infty,$$

то к последовательности $\{\xi_n\}$ применима однородная центральная предельная теорема.

Предельные теоремы для сумм зависимых случайных величин, как правило, доказываются с определенными заранее условиями слабой зависимости (регулярности), такими, как условия сильного перемешивания, равномерно сильного

перемешивания, полной регулярности (ρ -перемешивания) и т. п. (см., например, [2]). В связи с этим возникает вопрос – насколько выполнение той или иной теоремы обусловлено данным условием регулярности, возможно ли (и в каких границах) ослабление этого условия?

В работе [3] доказан следующий результат.

Теорема 1. Пусть $\{\xi_n, n = 1, 2, \dots\}$ – стационарная последовательность и пусть $\mathbf{E}\xi_n^2 < \infty$, $\mathbf{E}\xi_n = 0$, $n = 1, 2, \dots$. Для того чтобы к последовательности $\{\xi_n\}$ была применима центральная предельная теорема и σ_n^2 являлась правильно меняющейся последовательностью порядка 1, необходимо и достаточно, чтобы при любой последовательности натуральных чисел $m = m(n)$ и при любом действительном t выполнялось условие

$$\frac{S_{n+m}}{\sigma_{n+m}} \underset{d}{\sim} \frac{\hat{S}_n}{\sigma_{n+m}} + \frac{\hat{S}_m}{\sigma_{n+m}}, \quad n \rightarrow \infty \quad (\text{R})$$

и последовательность $\{\sigma_n^{-2} S_n^2\}$ была равномерно интегрируема.

Эту теорему можно интерпретировать так: условие (R) является *минимальным* условием слабой зависимости, при котором справедлива центральная предельная теорема для стационарных последовательностей с правильно меняющейся порядка 1 дисперсией.

В настоящей работе этот результат обобщается на произвольные последовательности случайных величин с конечными вторыми моментами.

Так же, как в предельных теоремах для независимых случайных величин, мы ограничимся только последовательностями, удовлетворяющими условию равномерной предельной малости: при любом $\varepsilon > 0$ и при любой последовательности натуральных чисел $k = k(n)$

$$\lim_{n \rightarrow \infty} \sup_{k \leq l \leq n+k} \mathbf{P}\{|\xi_l| > \varepsilon \sigma_n^{+k}\} = 0. \quad (\text{UN})$$

Следующее условие является обобщением условия (R) на нестационарные последовательности: при любом действительном t и при любых последовательностях натуральных чисел $m = m(n)$ и $k = k(n)$

$$\frac{S_{n+m}^{+k}}{\sigma_{n+m}^{+k}} \underset{d}{\sim} \frac{\hat{S}_n^{+k}}{\sigma_{n+m}^{+k}} + \frac{\hat{S}_m^{+(k+n)}}{\sigma_{n+m}^{+k}}, \quad n \rightarrow \infty. \quad (\text{RL})$$

В соответствии с определением слабой эквивалентности условие (RL) можно записать так:

$$\Delta(n) = \left| \mathbf{E} \exp \left\{ it \frac{S_{n+m}^{+k}}{\sigma_{n+m}^{+k}} \right\} - \mathbf{E} \exp \left\{ it \frac{S_n^{+k}}{\sigma_{n+m}^{+k}} \right\} \mathbf{E} \exp \left\{ it \frac{S_m^{+(k+n)}}{\sigma_{n+m}^{+k}} \right\} \right| \rightarrow 0,$$

при $n \rightarrow \infty$ и любом $t \in \mathbb{R}$. Введем еще аналог правильного изменения дисперсии сумм (см. [3, лемма 1]): при любых последовательностях натуральных чисел $m = m(n)$ и $k = k(n)$

$$(\sigma_{n+m}^{+k})^2 \sim (\sigma_n^{+k})^2 + (\sigma_m^{+(k+n)})^2, \quad n \rightarrow \infty. \quad (\text{RD})$$

Теорема 2. Пусть $\{\xi_n, n = 1, 2, \dots\}$ – последовательность случайных величин, удовлетворяющая условию (UN), и пусть $\mathbf{E}\xi_n = 0$, $\mathbf{E}\xi_n^2 < \infty$, $n = 1, 2, \dots$. Для того чтобы к последовательности $\{\xi_n\}$ была применима однородная центральная предельная теорема и σ_n удовлетворяла условию (RD), необходимо и достаточно, чтобы выполнялось условие (RL) и при любой последовательности натуральных чисел $k = k(n)$ последовательность $\left\{ \left(\frac{S_n^{+k}}{\sigma_n^{+k}} \right)^2 \right\}$ была равномерно интегрируема.

Доказательство.

Необходимость. Пусть к последовательности $\{\xi_n\}$ применима однородная центральная предельная теорема, то есть при любом $t \in \mathbf{R}$ и любой последовательности натуральных чисел $k = k(n)$

$$\mathbf{E} \exp \left\{ it \frac{S_n^{+k}}{\sigma_n^{+k}} \right\} \rightarrow \exp \left\{ -\frac{t^2}{2} \right\}, \quad n \rightarrow \infty \quad (1)$$

и σ_n удовлетворяет условию (RD). Тогда

$$\left(\frac{S_n^{+k}}{\sigma_n^{+k}} \right)^2 \xrightarrow{d} \mathcal{N}^2(0, 1), \quad \left(\frac{S_n^{+k}}{\sigma_n^{+k}} \right)^2 \geq 0, \quad \mathbf{E} \left(\frac{S_n^{+k}}{\sigma_n^{+k}} \right)^2 = \mathbf{E} \mathcal{N}^2(0, 1) = 1,$$

откуда следует равномерная интегрируемость последовательности $\left\{ \left(\frac{S_n^{+k}}{\sigma_n^{+k}} \right)^2, n = 1, 2, \dots \right\}$ (см., например, [4, теорема 5.4]).

Пусть $t \in \mathbf{R}$ и $m = m(n)$, $k = k(n)$. Так как σ_n^2 удовлетворяет условию (RD), то для любой последовательности натуральных чисел $\{n_1\}$ существуют $0 \leq c \leq 1$ и подпоследовательность $\{n_2\} \subseteq \{n_1\}$, такая, что

$$\left(\frac{\sigma_{n_2}^{+k_2}}{\sigma_{n_2+m_2}^{+k_2}} \right)^2 \rightarrow c, \quad \left(\frac{\sigma_{m_2}^{+(k_2+n_2)}}{\sigma_{n_2+m_2}^{+k_2}} \right)^2 \rightarrow 1 - c, \quad n \rightarrow \infty,$$

где $k_2 = k(n_2)$, $m_2 = m(n_2)$. Если $c = 0$ ($c = 1$), то при $n \rightarrow \infty$

$$(\sigma_{n_2+m_2}^{+k_2})^{-1} S_{n_2}^{+k_2} \rightarrow 0 \quad \left((\sigma_{n_2+m_2}^{+k_2})^{-1} S_{m_2}^{+(k_2+n_2)} \rightarrow 0 \right)$$

по вероятности, следовательно, $\Delta(n_2) \rightarrow 0$, $n \rightarrow \infty$. Если же $0 < c < 1$, то в силу (1)

$$\begin{aligned} \mathbf{E} \exp \left\{ it \frac{S_{n_2+m_2}^{+k_2}}{\sigma_{n_2+m_2}^{+k_2}} \right\} &\sim \exp \left\{ -\frac{t^2}{2} \right\} = \exp \left\{ -\frac{ct^2}{2} \right\} \exp \left\{ -\frac{(1-c)t^2}{2} \right\} \sim \\ &\sim \mathbf{E} \exp \left\{ it \frac{S_{n_2}^{+k_2}}{\sigma_{n_2+m_2}^{+k_2}} \right\} \mathbf{E} \exp \left\{ it \frac{S_{m_2}^{+(k_2+n_2)}}{\sigma_{n_2+m_2}^{+k_2}} \right\} \end{aligned}$$

$n \rightarrow \infty$, то есть снова $\Delta(n_2) \rightarrow 0$, $n \rightarrow \infty$.

Мы доказали, что из любой последовательности $\{\Delta(n_1)\}$ можно выделить сходящуюся к нулю подпоследовательность. Это означает, что $\Delta(n) \rightarrow 0$, $n \rightarrow \infty$, то есть выполнено условие (RL).

Достаточность. Пусть выполнено условие (RL) и при любой последовательности натуральных чисел $k = k(n)$ последовательность $\left\{ \left(\frac{S_n^{+k}}{\sigma_n^{+k}} \right)^2 \right\}$ равномерно интегрируема. В силу известной теоремы Прохорова (см., например, [4]) последовательность $\left\{ \frac{S_n^{+k}}{\sigma_n^{+k}} \right\}$ является относительно компактной, так что из любой последовательности натуральных чисел можно выбрать подпоследовательность $\{n_1\}$, $n_1 = n_1(n)$, такую, что

$$\frac{S_{n_1}^{+k_1}}{\sigma_{n_1}^{+k_1}} \xrightarrow{d} \xi, \quad \frac{S_{m_1}^{+(k_1+n_1)}}{\sigma_{m_1}^{+(k_1+n_1)}} \xrightarrow{d} \eta, \quad \frac{S_{n_1+m_1}^{+k_1}}{\sigma_{n_1+m_1}^{+k_1}} \xrightarrow{d} \zeta, \quad n \rightarrow \infty, \quad (2)$$

где $k_1 = k(n_1)$, $m_1 = m(n_1)$, а ξ, η и ζ – случайные величины. При этом поскольку последовательность $\left\{ \left(\frac{S_n^{+k}}{\sigma_n^{+k}} \right)^2 \right\}$ равномерно интегрируема, то

$$\mathbf{E}\xi^2 = \lim_{n \rightarrow \infty} \mathbf{E} \left(\frac{S_{n_1}^{+k_1}}{\sigma_{n_1}^{+k_1}} \right)^2, \quad \mathbf{E}\eta^2 = 1, \quad \mathbf{E}\zeta^2 = 1 \quad (3)$$

[4, теорема 5.4]. Далее, из ограниченных последовательностей

$$\alpha_{n_1} = \frac{\sigma_{n_1}^{+k_1}}{\sqrt{(\sigma_{n_1}^{+k_1})^2 + (\sigma_{m_1}^{+(k_1+n_1)})^2}}, \quad \beta_{n_1} = \frac{\sigma_{m_1}^{+k_1}}{\sqrt{(\sigma_{n_1}^{+k_1})^2 + (\sigma_{m_1}^{+(k_1+n_1)})^2}}$$

выберем подпоследовательности $\{\alpha_{n_2}\}$ и $\{\beta_{n_2}\}$, такие, что

$$\alpha_{n_2} \rightarrow \alpha, \quad \beta_{n_2} \rightarrow \beta, \quad n \rightarrow \infty, \quad \alpha^2 + \beta^2 = 1. \quad (4)$$

Тогда при $n \rightarrow \infty$

$$\frac{\widehat{S}_{n_2}^{+k_2} + \widehat{S}_{m_2}^{+(k_2+n_2)}}{\sqrt{(\sigma_{n_2}^{+k_2})^2 + (\sigma_{m_2}^{+(k_2+n_2)})^2}} = \alpha_{n_2} \frac{\widehat{S}_{n_2}^{+k_2}}{\sigma_{n_2}} + \beta_{n_2} \frac{\widehat{S}_{m_2}^{+(k_2+n_2)}}{\sigma_{m_2}^{+(k_2+n_2)}} \xrightarrow{d} \alpha \widehat{\xi} + \beta \widehat{\eta}. \quad (5)$$

Понятно, что $\alpha \widehat{\xi} + \beta \widehat{\eta}$ имеет невырожденное распределение.

Далее, в силу соотношений (RL) и (2)

$$\frac{\widehat{S}_{n_2}^{+k_2} + \widehat{S}_{m_2}^{+(k_2+n_2)}}{\sigma_{n_2+m_2}^{+k_2}} \xrightarrow{d} \zeta, \quad n \rightarrow \infty, \quad (6)$$

где ζ имеет невырожденное распределение. По теореме о сходимости типов [1, с.216] из (5) и (6) вытекает

$$\frac{\sigma_{n_2+m_2}^{+k_2}}{\sqrt{(\sigma_{n_2}^{+k_2})^2 + (\sigma_{m_2}^{+(k_2+n_2)})^2}} \rightarrow C, \quad 0 < C < \infty.$$

Отсюда следует, что вместе с последовательностями $\left\{ \left(\frac{S_{n_2}^{+k_2}}{\sigma_{n_2}^{+k_2}} \right)^2 \right\}$ и $\left\{ \left(\frac{\widehat{S}_{m_2}^{+(k_2+n_2)}}{\sigma_{m_2}^{+(k_2+n_2)}} \right)^2 \right\}$ равномерно интегрируемой является последовательность $\left\{ \left(\frac{\widehat{S}_{n_2}^{+k_2} + \widehat{S}_{m_2}^{+(k_2+n_2)}}{\sigma_{n_2+m_2}^{+k_2}} \right)^2 \right\}$ и из (6) получаем теперь

$$\delta_{n_2} = \frac{(\sigma_{n_2}^{+k_2})^2 + (\sigma_{m_2}^{+(k_2+n_2)})^2}{(\sigma_{n_2+m_2}^{+k_2})^2} \rightarrow \mathbf{E}\zeta^2 = 1, \quad n \rightarrow \infty.$$

Таким образом, мы показали, что для всякой последовательности натуральных чисел найдется подпоследовательность $\{n_2\}$, такая, что $\delta_{n_2} \rightarrow 1$, $n \rightarrow \infty$. Это означает, что $\delta_n \rightarrow 1$, $n \rightarrow \infty$, то есть выполнено (RD).

Пусть $\sigma_{k,n}$ удовлетворяет условию (RD) и выполнено условие (RL). Если $k \leq l \leq n+k$, то в силу условия (UN) $(\sigma_n^{+k})^{-1} \xi_l \rightarrow 0$ по вероятности, а из условия (RD) следует, что

$$\frac{\mathbf{E}\xi_l^2}{(\sigma_n^{+k})^2} \sim \frac{\mathbf{E}\xi_l^2}{(\sigma_{l-1}^{+k})^2 + \mathbf{E}\xi_l^2 + (\sigma_{l+1 < n}^{+k})^2} \geq 1,$$

так что последовательность $\left\{ (\sigma_n^{+k})^{-2} \xi_l^2 \right\}$ равномерно интегрируема, следовательно,

$$(\sigma_n^{+k})^{-2} \mathbf{E}\xi_l^2 \rightarrow 0, \quad n \rightarrow \infty. \quad (7)$$

Пусть $j_0 = k+1$, $j_r = \min\{l > j_{r-1} : \sigma_{j_{r-1},l}^2 \geq m^{-1} (\sigma_n^{+k})^2\}$, $r = 1, 2, \dots, m$. Тогда если $m = m(n) \rightarrow \infty$ растет достаточно медленно, то в силу (7)

$$\sigma_{j_{r-1},j_r}^2 \sim m^{-1} (\sigma_n^{+k})^2, \quad r = 1, 2, \dots, m. \quad (8)$$

Имеем

$$S_n^{+k} = \sum_{r=1}^m X_r, \quad X_r = \sum_{j=j_{r-1}+1}^{j_r} \xi_j, \quad r = 1, \dots, m.$$

В силу условия (RL) если $m = m(n)$ растет достаточно медленно, то

$$(\sigma_n^{+k})^{-1} S_n^{+k} \stackrel{d}{\sim} (\sigma_n^{+k})^{-1} \sum_{r=1}^m \widehat{X}_r, \quad n \rightarrow \infty. \quad (9)$$

Так как

$$\sum_{r=1}^m \mathbf{D}\widehat{X}_r = \sum_{r=1}^m \sigma_{j_{r-1},j_r}^2 \sim \sigma_n^{+k},$$

то для того, чтобы

$$(\sigma_n^{+k})^{-1} \sum_{r=1}^m \widehat{X}_r \xrightarrow{d} \mathcal{N}(0, 1) \quad (10)$$

(то есть чтобы к последовательности серий независимых случайных величин $\left\{ (\sigma_n^{+k})^{-1} \widehat{X}_r, r = 1, \dots, m, n = 1, 2, \dots \right\}$ была применима центральная предельная теорема), достаточно, чтобы выполнялось условие Линдеберга: при любом $\varepsilon > 0$

$$L_n(\varepsilon) = \frac{1}{(\sigma_n^{+k})^2} \sum_{r=1}^m \mathbf{E}\{\widehat{X}_r^2, |\widehat{X}_r| > \varepsilon \sigma_n^{+k}\} \rightarrow 0, \quad n \rightarrow \infty.$$

По условию при каждом $r = 1, \dots, m$ последовательность величин $\left\{ (\sigma_{j_{r-1}, j_r}^2)^{-2} X_r^2 \right\}$ равномерно интегрируема, так что в силу (8) при любом $1 \leq r \leq m$ и любом $\varepsilon > 0$ имеем

$$\frac{1}{\sigma_{j_{r-1}, j_r}^2} \mathbf{E}\left\{ \widehat{X}_r^2, |\widehat{X}_r| > \varepsilon \sigma_n^{+k} \right\} \sim \frac{m}{(\sigma_n^{+k})^2} \mathbf{E}\left\{ \widehat{X}_r^2, |\widehat{X}_r| > \varepsilon \sigma_n^{+k} \right\} \rightarrow 0,$$

так что при достаточно медленно растущих $m = m(n)$

$$L_n(\varepsilon) \leq \max_{1 \leq r \leq m} \frac{m}{(\sigma_n^{+k})^2} \mathbf{E}\left\{ \widehat{X}_r^2, |\widehat{X}_r| > \varepsilon \sigma_n^{+k} \right\} \rightarrow 0, \quad n \rightarrow \infty,$$

и из (9) и (10) следует теперь $(\sigma_n^{+k})^{-1} S_n^{+k} \xrightarrow{d} \mathcal{N}(0, 1), \quad n \rightarrow \infty.$ Теорема доказана. ■

ЛИТЕРАТУРА

1. Лозв М. Теория вероятностей. М.: ИЛ, 1962. 719 с.
2. Ибрагимов И.А., Линник Ю.В. Независимые и стационарно связанные величины. М.: Наука, 1965. 524 с.
3. Гринь А.Г. О минимальном условии слабой зависимости в центральной предельной теореме для стационарных последовательностей // Теория вероятн. и ее примен. 2002. Т.47, N.3. С.554-558.
4. Биллингсли П. Сходимость вероятностных мер. М: Наука, 1977. 351 с.

ИССЛЕДОВАНИЕ ПОВЕДЕНИЯ ЯВНЫХ МЕТОДОВ РУНГЕ-КУТТЫ ПРИ РЕШЕНИИ СИСТЕМ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ С РАЗРЫВНОЙ ПРАВОЙ ЧАСТЬЮ

В.В. Коробицын, В.Б. Маренич, Ю.В. Фролова

Определены понятия системы обыкновенных дифференциальных уравнений с клетчатой структурой и его решения. Приведен пример линейной системы с клетчатой структурой, имеющей циклическое решение. Представлено исследование поведения численного решения для традиционных явных схем Рунге-Кутты разных порядков. Показано поведение погрешности численного решения. Приведены результаты измерений вычислительных затрат, необходимых для достижения заданной точности.

1. Введение

Модели, описываемые системами обыкновенных дифференциальных уравнений, являются удобным инструментом для описания реальных процессов, но, как правило, эти модели ведут себя адекватно лишь в некоторой ограниченной области фазового пространства. В природе же процессы имеют более сложную форму, и поэтому нам приходится прибегать к различным приемам, обеспечивающим необходимыми инструментами исследования систем. Опираясь на известные алгоритмы и методы, мы пытаемся свести задачу к ряду связанных задач, каждая из которых может быть решена известными методами. Но не так легко бывает связать полученные локальные результаты со всей задачей целиком. Так, например, для описания многих физических, биологических и социальных процессов целесообразно воспользоваться моделями, описываемыми системами обыкновенных дифференциальных уравнений, где в правой части стоит функция, имеющая разрыв на некотором подмножестве.

В общем случае необходимо пересмотреть понятие самого решения такого уравнения, поскольку в традиционном понимании справа должна стоять функция, удовлетворяющая условию Липшица. Кроме того, необходимо модернизировать численные методы решения обыкновенных дифференциальных уравнений на случай разрывной правой части. Исследовать аппроксимацию, устойчивость и сходимость численного метода. Эти и многие другие задачи возникают

при исследовании систем обыкновенных дифференциальных уравнений с разрывной правой частью.

В данной работе представлено исследование поведения численного решения на задаче с разрывной правой частью для традиционных явных схем типа Рунге-Кутты разных порядков. Показано поведение погрешности численного решения на конкретном примере. Приведены результаты измерений вычислительных затрат, необходимых для достижения заданной точности.

2. Определение понятия решения разрывной системы

Рассмотрим поведение автономной системы, описываемой набором n параметров $y = (y_1, y_2, \dots, y_n)$ в некоторой односвязной области $D \subset \mathbb{R}^n$. Эта область разделена на две части D_1, D_2 некоторой поверхностью G , заданной алгебраическим уравнением $g(y) = 0$. Таким образом, заданы области $D_1 = \{y : g(y) < 0\}$ и $D_2 = \{y : g(y) > 0\}$. В этих областях поведение системы описывается решением системы обыкновенных дифференциальных уравнений:

$$\frac{dy}{dt} = f_1(y), \quad y \in D_1, \quad (1)$$

$$\frac{dy}{dt} = f_2(y), \quad y \in D_2. \quad (2)$$

Функции f_1, f_2 удовлетворяют условию Липшица.

Решение внутри каждой области D_1 и D_2 находится обычным способом и является непрерывно дифференцируемой кривой. Если рассматривать решение задачи в обеих областях одновременно, то возникает вопрос — что считать решением системы?

Определение 1. Решением объединенной системы (1)–(2) будем считать кривую $y(t)$, которая совпадает с решением (1) в области D_1 и решением (2) в D_2 , а также склеивается на границе $G = \{y : g(y) = 0\}$. Склеивание понимается как существование точки $y^* \in G$ и величины $t^* \in \mathbb{R}$, таких, что $\lim_{t \rightarrow t^*-0} y(t) = y^*$ и $\lim_{t \rightarrow t^*+0} y(t) = y^*$.

Определение 2. Множество D_i будем называть клеткой системы обыкновенных дифференциальных уравнений $\frac{dy}{dt} = F(y)$, если D_i является линейно связным подмножеством множества D_C непрерывности функции $F(y)$.

Для всех кусочно-гладких систем множество D_C может быть представлено как объединение конечного числа клеток D_i , $i = 1, \dots, n$

$$D_C = \bigcup_{i=1}^n D_i.$$

Определение 3. Считаем, что система обыкновенных дифференциальных уравнений $\frac{dy}{dt} = F(y)$ имеет клетчатую структуру $\mathcal{D}_s$, если объединение замыканий клеток $\overline{D_i}$, $i = 1, \dots, n$ совпадает с замыканием области определения $\overline{D}$ функции $F(y)$:

$$\mathcal{D}_s : \bigcup_{i=1}^n \overline{D_i} = \overline{D}.$$

Определение 4. Решением системы обыкновенных дифференциальных уравнений $\frac{dy}{dt} = F(y)$ с клетчатой структурой $\mathcal{D}_s$ будем считать непрерывную кривую $y(t)$, удовлетворяющую дифференциальному уравнению в каждой клетке D_i и склеенную на границе между ними. Склеивание понимается как существование точки $y^* \in G$ на поверхности разрыва функции F и величины $t^* \in \mathbb{R}$, таких, что $\lim_{t \rightarrow t^* - 0} y(t) = y^*$ и $\lim_{t \rightarrow t^* + 0} y(t) = y^*$.

Приведем пример линейной системы ОДУ с клетчатой структурой. Пример интересен наличием циклического решения, которое отсутствует в каждой отдельной клетке.

3. Рождение цикла в линейной системе с клетчатой структурой

Рассмотрим систему линейных обыкновенных дифференциальных уравнений с клетчатой структурой, составленной из двух клеток $D_1 : y_1 < \alpha$ и $D_2 : y_1 > \alpha$ на плоскости (y_1, y_2) :

$$\begin{aligned} \frac{dy_1}{dt} &= a_{11}(y_1 - c_1) + a_{12}(y_2 - d_1), \\ \frac{dy_2}{dt} &= a_{21}(y_1 - c_1) + a_{22}(y_2 - d_1), \quad (y_1, y_2) \in D_1, \end{aligned} \quad (3)$$

$$\begin{aligned} \frac{dy_1}{dt} &= a_{22}(y_1 - c_2) + a_{12}(y_2 - d_2), \\ \frac{dy_2}{dt} &= a_{21}(y_1 - c_2) + a_{11}(y_2 - d_2), \quad (y_1, y_2) \in D_2. \end{aligned} \quad (4)$$

В случае, когда $a_{11} = a_{22} = 0$, $a_{12} = a_{21} = 1$, решение каждой из систем

$$\begin{aligned} \frac{dy_1}{dt} &= y_2 - d_1, \\ \frac{dy_2}{dt} &= y_1 - c_1, \quad y_1 < \alpha, \end{aligned} \quad (5)$$

$$\begin{aligned} \frac{dy_1}{dt} &= y_2 - d_2, \\ \frac{dy_2}{dt} &= y_1 - c_2, \quad y_1 > \alpha \end{aligned} \quad (6)$$

имеет одну особую точку (седло). Когда $c_1 < \alpha$ и $c_2 > \alpha$, то особые точки лежат в соответствующих клетках D_1 и D_2 . Кроме того, когда $d_1 = d_2$, — особые точки лежат на прямой ортогональной границе между клетками, а их сепаратрисы пересекаются в некоторой точке на этой границе. В этом случае образуется замкнутая область D_0 , заключенная между сепаратрисами. Решение внутри этой области образует цикл и не выходит за ее пределы. Цикл составлен из

двух гладких кривых, склеенных на границе между клетками (рис.1). Кривые решения описываются функциями

$$\begin{aligned} y_1(t) &= A_1 e^{t-t^*} + A_2 e^{-(t-t^*)} + a_{y1}, \\ y_2(t) &= A_1 e^{t-t^*} + A_2 e^{-(t-t^*)} + a_{y2}, \end{aligned} \quad (7)$$

где $A_1 = ((y_1^* - a_{y1}) + (y_2^* - a_{y2}))/2$, $A_2 = ((y_1^* - a_{y1}) - (y_2^* - a_{y2}))/2$. Значения t^* , y_1^* , y_2^* задают начальные значения, а величины $a_{y1} = c_1$, $a_{y2} = d_1$ при $y_1^* < \alpha$ и $a_{y1} = c_2$, $a_{y2} = d_2$ при $y_1^* > \alpha$.

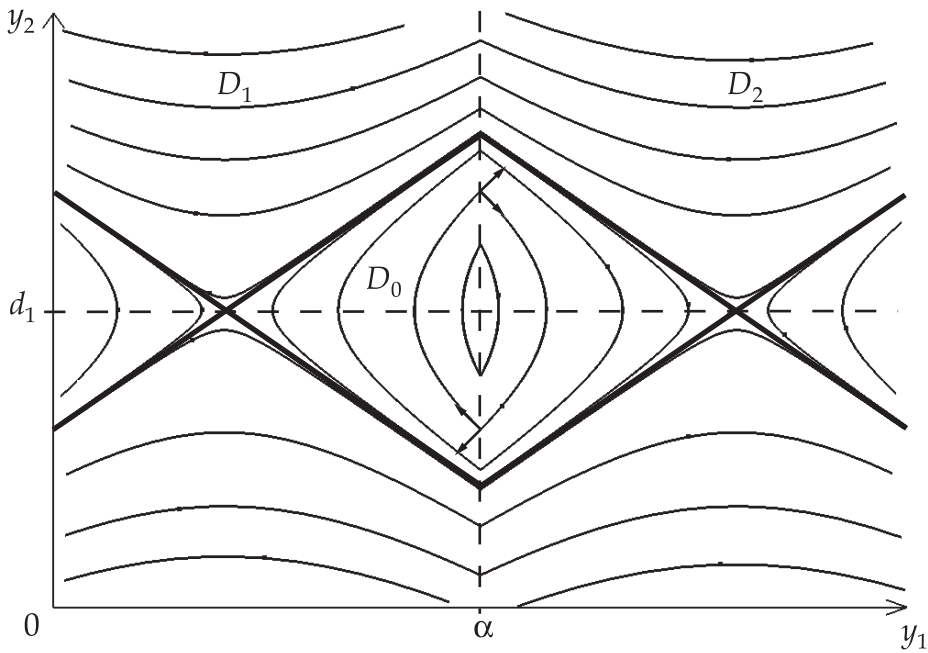


Рис. 1. Фазовый портрет системы (5)-(6)

Длина полупериода решения внутри D_0 задается формулой

$$\frac{T}{2} = \ln \left(\frac{|a_{y1} - \alpha| + \sqrt{(a_{y1} - \alpha)^2 - 4A_1 A_2}}{2A_1} \right).$$

4. Вычислительные эксперименты для системы с клетчатой структурой

Исследуем поведение численного решения на задаче с клетчатой структурой (5)-(6). Очевидно, что в областях D_1 и D_2 известные методы Рунге-Кутты без труда справятся с этой задачей, но что получается при переходе решения из D_1 в D_2 и обратно? Именно на границе происходит скачок погрешности решения из-за резкого изменения правой части уравнения. И дело даже не в том, что сильно меняется правая часть, а в том, что малейшие отклонения от траектории приводят к большим отклонениям при переходе через границу G . Рассмотрим это на нескольких тестовых экспериментах.

Исследуем погрешность численного решения для системы (5)–(6) с заданными параметрами $d_1 = 0.5$, $c_1 = 0.2$, $d_2 = 0.5$, $c_2 = 0.8$, $\alpha = 0.5$ и начальными данными $y_1(0) = 0.499999$, $y_2(0) = 0.3$. Приведем результаты численных экспериментов для пяти явных методов решения систем обыкновенных дифференциальных уравнений:

- 1) явный метод Эйлера (Explicit Euler). Одностадийная явная схема с оценкой погрешности методом Рундсона;
- 2) метод Рунге-Кутты 4-го порядка (Runge-Kutta 4). Классическая четырехстадийная схема с оценкой погрешности по Рундсону и автоматическим управлением длиной шага интегрирования;
- 3) метод Дорманда-Принса 5-го порядка (Dormand-Prince 5(4)). Восьмистадийная схема 5-го порядка с оценкой погрешности по вложенной схеме 4-го порядка;
- 4) метод Рунге-Кутты 2-го порядка (Runge-Kutta 2). Явная двухстадийная схема, предложенная Рунге, с оценкой погрешности по Рундсону;
- 5) метод Рунге-Кутты-Фельберга 4-го порядка (Felberg 4(5)). Наиболее распространенная вложенная схема четвертого порядка с 7-ю стадиями, использующая встроенную схему 5-го порядка для оценки локальной погрешности.

Первый эксперимент предназначен для исследования поведения схем на разрывах. Оценивалась относительная погрешность численного решения на точном решении после преодоления одного цикла длительностью T . В этом случае точное решения попадает в начальную точку, образуя цикл. Значения численного решения показывают насколько оно отклонилось от точного за один цикл. Все рассматриваемые методы решения имеют алгоритм оценки локальной погрешности и выбора длины шага интегрирования на основе заданной точности. Предполагается, что при уменьшении заданной точности погрешность численного решения также должна уменьшаться, обеспечивая тем самым сходимость численного метода. Однако для разрывных систем это не выполняется. Кроме того, заметно значительное уменьшение порядка точности метода. Результаты эксперимента приведены на рис. 2. График зависимости изображен в логарифмическом масштабе: по горизонтальной оси откладывается порядок заданной точности ($-\log_{10}(\text{loc_tolerance})$), по вертикальной — порядок относительной погрешности решения ($-\log_{10}(\text{rel_error})$), вычисляемой по формуле

$$\text{rel_error} = \frac{\|y_n - y^*(T)\|}{\|y_n\|},$$

где $y^*(T)$ — точное решение в точке $t = T$, y_n — численное решение, полученное после n шагов, соответствующих той же точке, $\|\cdot\|$ — евклидова норма.

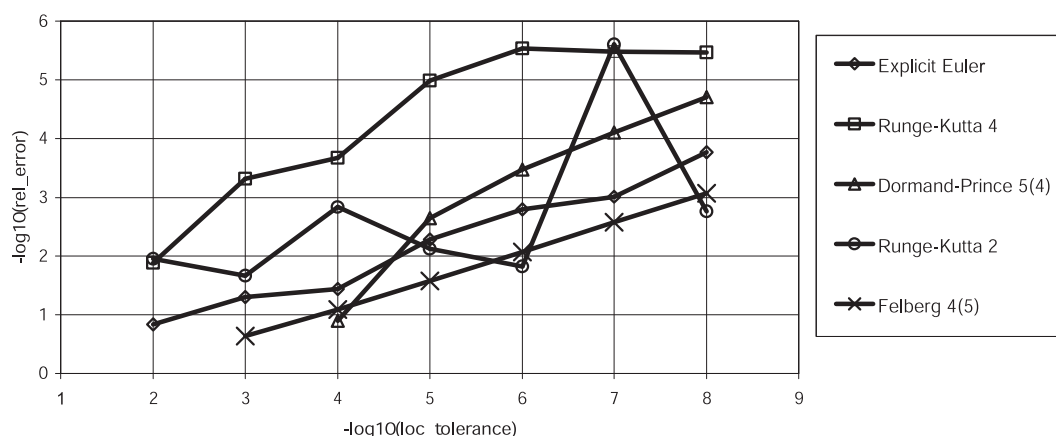


Рис. 2. Зависимость относительной погрешности решения от заданной локальной точности

Как можно видеть из диаграммы, метод Рунге-Кутты-Фельберга имеет большую погрешность, чем метод Эйлера. Метод Рунге-Кутты 2-го порядка ведет себя нестабильно, что отражается в изменении тенденции «роста-падения» погрешности решения. При достаточно высокой заданной точности метод Дорманда-Принса ведет себя лучше метода Эйлера. Наилучшим образом показал себя классический метод Рунге-Кутты 4-го порядка, который обеспечил заданную точность до 5-го порядка включительно. И это значительно лучше других методов.

Безусловно, для нахождения численного решения требуется разный объем вычислительных затрат для различных методов. Оценим вычислительные затраты каждого метода для получения решения с относительной погрешностью, не превышающей заданную. Результаты приведены в таблице 1. Вычислительные затраты оцениваются количеством вычислений правых частей уравнения, которых было достаточно для обеспечения заданной точности.

Таблица 1. Оценка вычислительных затрат для нахождения численного решения

rel_error	Explicit Euler	RK 4	Dormand-Prince	RK 2	Felberg
$5 \cdot 10^{-2}$	711	363	1627	69	9729
$5 \cdot 10^{-3}$	7188	411	1627	2001	89979
$5 \cdot 10^{-4}$	71736	819	15595	—	—

Особенность численного решения задачи с разрывами проявляется в изменении длины периода циклического решения. Так, для наиболее удачной схемы RK4 численное решение постепенно смещается вдоль числовой оси, уменьшая тем самым длину периода решения. Результат смещения продемонстрирован на рис. 3. Начиная с одной точки (слева) решение постепенно смещается (центр) и входит в противофазу (справа).

В такой ситуации оценивать погрешность численного решения относительно точного традиционными методами (как разность между практикой и теорией) будет неразумно. Поэтому имеет смысл разделить погрешность численного решения на две составляющие: смещение по оси времени и отклонение от траек-

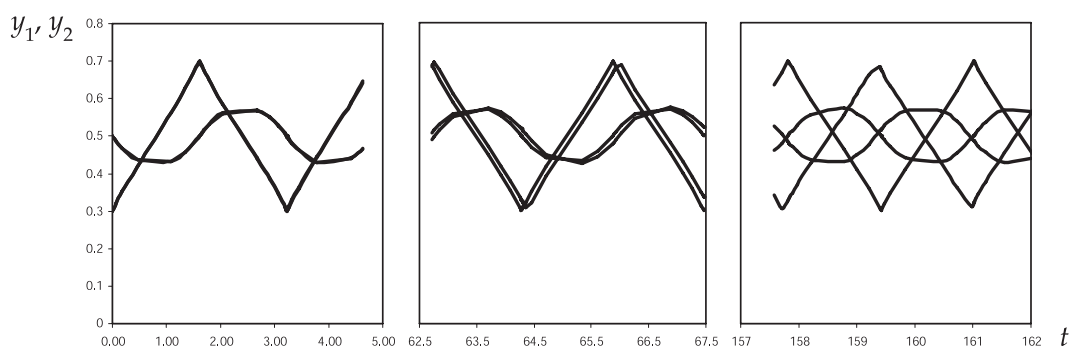


Рис. 3. Смещение численного решения RK4 относительно точного по временной оси

тории точного циклического решения. Результаты оценки смещений численного решения относительно точного после прохождения одного периода ($t = T$) приведены в таблице 2. Обозначения, принятые в таблице: Tol — заданная точность, Δt — смещение по оси времени, Δr — отклонение от траектории точного решения. Как можно видеть из таблицы, сходимость численного решения к точному наблюдаются, причем отклонения от траектории значительно меньше, чем смещение по времени.

Таблица 2. Оценка смещения решения по оси времени и отклонения от траектории

Tol	Δt	Δr
10^{-2}	-0,02149378	0,0009338849
10^{-3}	-0,00018583	0,0002725838
10^{-4}	-0,00034308	0,0000122797
10^{-5}	-0,00001584	0,0000016340
10^{-6}	0,00000491	0,0000004262
10^{-7}	0,00000491	0,0000000504
10^{-8}	0,00000502	0,0000000062

Результаты экспериментов показали, что для преодоления разрывов в правой части системы обыкновенных дифференциальных уравнений лучше подходят методы с оценкой погрешности, основанные на экстраполяции Рундсона, чем методы с вложенной схемой оценки погрешности. А наиболее удачным является классический метод Рунге-Кутты 4-го порядка.

ЛИТЕРАТУРА

1. Деккер К., Вервер Я. Устойчивость методов Рунге-Кутты для жестких нелинейных дифференциальных уравнений. М.: Мир, 1988. 334 с.
2. Филиппов А.Ф. Дифференциальные уравнения с разрывной правой частью. М.: Наука, 1985. 224 с.
3. Хайрер Э., Нёрсетт С., Ваннер Г. Решение обыкновенных дифференциальных уравнений. Нежесткие задачи. М.: Мир, 1990. 512 с.
4. Хайрер Э., Ваннер Г. Решение обыкновенных дифференциальных уравнений. Жесткие и дифференциально-алгебраические задачи. М.: Мир, 1999. 685 с.

КЛАССИФИКАЦИЯ ПЛОСКИХ ПОЛНЫХ СТРОГО ПРИЧИННЫХ ЛОРЕНЦЕВЫХ МНОГООБРАЗИЙ МАЛЫХ РАЗМЕРНОСТЕЙ

Е.А. Мещеряков

Классифицируются полные плоские строго причинные лоренцевы многообразия малых размерностей, не более 8.

Любое полное плоское лоренцево многообразие M может быть реализовано как фактор-пространство $\mathbb{M}_n/\Gamma$, где $\mathbb{M}_n$ – n -мерное пространство-время Минковского, а Γ – дискретная подгруппа группы Пуанкаре, действующая свободно и собственноразрывно на $\mathbb{M}_n$. Заметим, что $\Gamma \cong \pi_1(M)$. Лоренцева метрика определяет в каждом касательном пространстве пару замкнутых выпуклых круглых конусов. Выбор одного из них задает локальное поле конусов. Оно продолжается до глобального поля на M или на двулистной накрывающей M ; условимся считать, что поле определено на M . Это равносильно тому, что линейные части отображений из Γ не переставляют конусов прошлого и будущего в $\mathbb{M}_n$. Если многообразие M не допускает замкнутых времениподобных кривых, то оно называется *причинным*. Для Γ это означает, что орбита любой точки не пересекает конуса с вершиной в этой точке из поля параллельных конусов на $\mathbb{M}_n$, заданного метрикой. Выбор начала координат $o \in \mathbb{M}_n$ позволяет отождествить $\mathbb{M}_n$ с вещественным векторным пространством V , в котором задана лоренцева форма ℓ сигнатуры $(+, - \dots, -)$; причинная структура задается конусом C (один из двух конусов, определяемых соотношением $\ell(v, v) \geq 0$). Будем называть M *строго причинным*, если M причинно, а прошлое и будущее любой точки $p \in M$ замкнуты в некоторой окрестности p (это не означает глобальной их замкнутости). В работе [8] такие многообразия были найдены с точностью до конечных накрытий.

В дальнейшем, если не оговорено противное, лоренцево многообразие M считается плоским, полным и строго причинным. В работе [8] было получено параметрическое описание таких многообразий (с точностью до конечных накрытий). В статье [7] каждому такому многообразию с унипотентной группой голономии сопоставлялась парабола в конусе P_m положительно определенных квадратичных форм на некотором евклидовом пространстве T (здесь $m = \dim T$); там же было показано, что многообразие может быть восстановлено

по этой параболе, рассматриваемой с точностью до линейных автоморфизмов конуса и аффинных замен параметра.

В данной работе уточняется алгоритм восстановления многообразия по характеристической параболе для малых размерностей (не более 8). Любой квадратный трехчлен с матричными коэффициентами, задающий характеристическую параболу, будем называть *характеристическим многочленом* данного многообразия. В статье вводится каноническая запись характеристического многочлена (которая, в случае парабол общего положения, единственна), классифицированы многообразия указанного типа в размерностях до 8, а также найдены размерности пространств модулей таких многообразий.

1. Предварительные сведения

Обозначим через V вещественное векторное пространство, в котором задана лоренцева метрика ℓ сигнатуры $(+, -, \dots, -)$. Будем рассматривать пространство Минковского $\mathbb{M}_n$, где $n = \dim V$, как пару (V, ℓ) . В работе [7] приведен способ построения многообразия M , состоящий из трех этапов:

- (А) фиксируем изотропные векторы v_0, v_1 , такие, что $\ell(v_0, v_1) = 1$, определяем подпространства L, W, N соотношениями

$$L = \mathbb{R}v_0, \quad W = L^\perp, \quad N = W \cap v_1^\perp$$

и выбираем подпространство $T \subseteq N$;

- (В) для произвольного ℓ -симметрического линейного оператора $a' : T \rightarrow T$ и каждого его собственного подпространства Λ_j задаем невырожденный линейный оператор $a''_j : \Lambda_j \rightarrow R = T^\perp \cap N$, после чего полагаем

$$a'' = \sum_j a''_j, \quad a = a' + a'';$$

- (С) выбираем линейный базис в T и определяем Γ как порожденную им подгруппу векторной группы T .

Вектор v называется изотропным, если $\ell(v, v) = 0$. Отображение a определяет аффинное действие T (и поэтому Γ) на $\mathbb{M}_n$:

$$\begin{aligned} \gamma_x(v) &= \lambda_x v + \tau_x, \quad x \in T; \\ \lambda_x v &= v + \ell(v, v_0)ax - \left(\ell(v, ax) + \frac{1}{2} \ell(v, v_0) \ell(ax, ax) \right) v_0; \\ \tau_x &= x - \frac{1}{2} \ell(ax, x) v_0. \end{aligned}$$

Положим

$$n = \dim M, \quad m = \dim T, \quad r = \dim R, \quad k = \dim \ker a$$

и, следуя [7], будем называть набор $\sigma = (n, m, r, k)$ *сигнатурой* M . Очевидно, эти числа удовлетворяют неравенствам

$$m + r + 2 \leq n, \quad r + k \leq m. \quad (1)$$

Согласно [7, Предложение 1], группа T может быть определена инвариантным образом, а именно как алгебраическое замыкание Γ в группе Пуанкаре. Плоские полные строго причинные лоренцевы многообразия $M = \mathbb{M}/\Gamma$ и $\tilde{M} = \mathbb{M}/\tilde{\Gamma}$ будем называть *почти причинно изометричными*, если факторпространства $\mathbb{M}/T$ и $\mathbb{M}/\tilde{T}$, где T и $\tilde{T}$ — алгебраические замыкания Γ и $\tilde{\Gamma}$, соответственно, изометричны. Это равносильно тому, что их построение по алгоритму (A)–(C) различается только последним пунктом.

В работе будут использоваться утверждения из [7], которые сформулированы ниже в удобном для нас виде.

Теорема 1. (i) Пусть M и $\tilde{M}$ — плоские полные строго причинные лоренцевы многообразия. Они причинно изометричны тогда и только тогда, когда их сигнатуры совпадают и

$$Q_M(s) = X^\top Q_{\tilde{M}}(\alpha s + \beta) X \quad (2)$$

для некоторых $X \in \mathrm{GL}(m, \mathbb{Z})$, $\alpha > 0$, $\beta \in \mathbb{R}$. Замена включения $X \in \mathrm{GL}(m, \mathbb{Z})$ на $X \in \mathrm{GL}(m, \mathbb{R})$ дает критерий почти причинной изометричности M и $\tilde{M}$.

(ii) Полином $Q(s) = A + 2sB + s^2C$, где A, B и C — m -матрицы, определяет характеристическую кривую для некоторого строго причинного многообразия M сигнатуры $(n, m, r, 0)$ тогда и только тогда, когда $n \geq m + r + 2$,

$$Q(s) > 0 \text{ для любого } s \in \mathbb{R}, \quad (3)$$

$$C - BA^{-1}B \geq 0, \quad (4)$$

$$r = \mathrm{rank}(C - BA^{-1}B). \quad (5)$$

(iii) Любое многообразие сигнатуры (n, m, r, k) почти причинно изометрично произведению многообразия сигнатуры $(n - k, m - k, r, 0)$ и плоского k -тора. Равенство $k = 0$ равносильно невырожденности C .

(iv) Пусть полиномы $Q_M(s) = A + 2sB + s^2C$, $Q_{\tilde{M}}(s) = \tilde{A} + 2s\tilde{B} + s^2\tilde{C}$ являются характеристическими для почти причинно изометричных многообразий M и $\tilde{M}$. Допустим, что $C > 0$, $\tilde{C} > 0$. Тогда

$$\mathrm{sp}(BC^{-1}) = \phi(\mathrm{sp}(\tilde{B}\tilde{C}^{-1})) \quad (6)$$

для некоторого $\phi \in \mathrm{Aff}(\mathbb{R})$.

Здесь через $\mathrm{Aff}(\mathbb{R})$ обозначена группа аффинных преобразований прямой, сохраняющих её ориентацию.

Замечание 1. Пусть многочлен $Q(s)$ задает характеристическую кривую некоторого строго причинного многообразия M . Тогда при любых $X \in \text{GL}(m, \mathbb{R})$, $\alpha > 0$, $\beta \in \mathbb{R}$ то же самое верно для многочлена $XQ(\alpha s + \beta)X^\top$. Для доказательства достаточно заметить, что эти преобразования сохраняют условия (3), (4) и (5); сигнатура тоже сохраняется.

Согласно (iii), с точностью до почти причинной изометрии можно считать, что многообразие имеет сигнатуру $(n, m, r, 0)$ и что $C > 0$ (тем самым исключается эллиптический случай). В дальнейшем это предполагается по умолчанию.

Будем обозначать через $\text{diag}(d_1, \dots, d_p)$ диагональную p -матрицу с элементами $d_1, \dots, d_p$ по диагонали; $\text{sp}(A)$ обозначает спектр матрицы A ; I – единичная матрица соответствующего размера.

Предложение 1. Для любого плоского полного строго причинного лоренцева многообразия M существует матрица $X \in \text{GL}(m, \mathbb{R})$ и преобразование $\alpha \in \text{Aff}(\mathbb{R})$ такие, что

$$Q'_M = XQ_M(\alpha(s))X^\top = A' + 2sD + Is^2, \quad (7)$$

где $D = \text{diag}(0, \dots, 0, 1, d_2, \dots, d_{m-k})$, k – количество нулей и $1 \leq d_2 \leq \dots \leq d_{m-k}$. Если спектр BC^{-1} содержит более одной точки, то преобразование α определяется этим условием однозначно.

Доказательство. Пусть $C^{1/2}$ – симметрический неотрицательно определенный квадратный корень из C . Поскольку матрица $C^{-1/2}BC^{-1/2}$ симметрическая, существует ортогональная матрица H , такая, что $D' = H^\top C^{-1/2}BC^{-1/2}H$ диагональна. Тогда

$$A + 2sB + Cs^2 = (C^{1/2}H)(A'' + 2sD' + s^2I)(C^{1/2}H)^\top.$$

Согласно замечанию 1, многочлен $A'' + 2sD' + s^2I$ соответствует некоторому полному плоскому строго причинному многообразию $\tilde{M}$, которое почти причинно изометрично M в силу теоремы 1. Поэтому можно считать, что $Q_M = A'' + 2sD' + s^2I$, где D' диагональна.

Заметим, что спектр D совпадает со спектром BC^{-1} , так как эти матрицы подобны согласно (7):

$$D = (XC^{-1/2})(BC^{-1})(XC^{-1/2})^{-1}.$$

Пусть $\text{sp}(BC^{-1}) = \{s_1, \dots, s_m\}$, причем $s_1 = s_2 = \dots = s_k < s_{k+1} \leq \dots \leq s_m$. Если $k < m$, то существует единственное аффинное преобразование прямой, переводящее $\text{sp}(BC^{-1})$ в набор $\{0, \dots, 0, 1, d_2, \dots, d_{m-1-k}\}$ (именно, $\alpha(s) = \frac{s-s_1}{s_{k+1}-s_1}$). ■

Если спектр BC^{-1} состоит из одного собственного числа, то его можно перевести в ноль сдвигом: $\alpha(s) = s - s_1$.

Замечание 2. Сопрягая характеристический многочлен ортогональной матрицей, сохраняющей структуру D , можно добиться того, чтобы диагональные блоки A' , соответствующие одинаковым числам на диагонали D , стали диагональными.

Характеристический многочлен многообразия M будем называть *каноническим*, если он имеет вид:

$$Q_M(s) = A + 2sD + Is^2, \quad (8)$$

где $D = \text{diag}(0, \dots, 0, 1, d_2, \dots, d_{m-k})$, k – количество нулей и $1 \leq d_2 \leq \dots \leq d_{m-k}$. Диагональные блоки матрицы A , отвечающие одинаковым диагональным элементам D , в соответствии с замечанием 2, будем считать диагональными матрицами. Кроме того, будем считать, что диагональные элементы в каждом диагональном блоке матрицы A не убывают и элементы $(k+1)$ -ой строки, кроме элементов на диагонали, неположительны. Например, при $m = 3$ и $D = \text{diag}(d_1, d_1, d_2)$, где $d_1 \neq d_2$, A должна иметь вид $\begin{pmatrix} * & 0 & * \\ 0 & * & * \\ * & * & * \end{pmatrix}$.

Замечание 3. В качестве канонического можно было бы принять вид $I + 2sD + Cs^2$. Выбор (8) в качестве канонического вида был обусловлен тем, что при $C = I$ один из инвариантов многообразия (а именно, $\text{sp}(BC^{-1})$) присутствует в каноническом многочлене явным образом (как множество диагональных элементов матрицы D).

Пусть характеристический многочлен имеет вид $Q_M(s) = Cs^2 + 2Bs + A$. Будем говорить, что плоское полное строго причинное лоренцево многообразие имеет *простой спектр*, если спектр BC^{-1} прост (т. е. у матрицы BC^{-1} нет кратных собственных чисел). Предложение 1 и замечание 2 приводят к следующему утверждению.

Теорема 2. Полные плоские строго причинные многообразия с простым спектром почти причинно изометричны тогда и только тогда, когда совпадают их сигнатуры, а канонические характеристические многочлены сопряжены, как в (2), матрицей вида $X = \text{diag}(\pm 1, \dots, \pm 1)$.

Доказательство. Если сигнатуры многообразий совпадают и многочлены сопряжены указанной матрицей X , то многообразия причинно изометричны согласно теореме 1, (i).

Так как мы условились не различать почти причинно изометричные многообразия, то задача сводится к доказательству единственности канонического характеристического многочлена с точностью до сопряжения матрицами X указанного в теореме вида. Пусть Q_1 и Q_2 – два канонических многочлена многообразия M . Тогда, согласно теореме 1, существует матрица X , такая, что $Q_2 = XQ_1X^\top$. Так как коэффициент при s^2 сохраняется, то X ортогональна. Следовательно, $Q_2 = XQ_1X^{-1}$. Поскольку $\text{sp}(BC^{-1})$ является инвариантом многообразия согласно теореме 1, (iv), матрицы D_1 и D_2 (коэффициенты при s) имеют одинаковый спектр. Так как эти матрицы диагональны, то они отличаются лишь порядком следования диагональных элементов. Поскольку их диагональные элементы упорядочены, $D_1 = D_2$. Поэтому X коммутирует с D_1 . Следовательно, X – блочно-диагональная матрица, причем нетривиальные блоки отвечают одинаковым собственным числам на диагонали. Так как собствен-

ные числа различны, то X диагональна. Из диагональности и ортогональности X и следует утверждение теоремы. ■

Замечание 4. В условиях теоремы 2 можно добиться единственности канонического многочлена, если потребовать, чтобы элементы некоторой фиксированной строки (столбца) матрицы A , кроме элементов на диагонали, были отрицательными (или положительными). Если у A нет строк без нулей, то имеются разные канонические многочлены, отвечающие одному многообразию.

При не простом спектре многообразия канонический характеристический многочлен неединственен. Легко показать, что все многочлены одного многообразия сопряжены блочно-ортогональной матрицей X (блоки соответствуют одинаковым собственным числам D). Согласно определению канонического вида, диагональные блоки A , отвечающие одинаковым числам на диагонали D , можно считать диагональными матрицами. Условие упорядоченности элементов этих матриц (например, неубывание) определяет X с точностью до сопряжения матрицей вида $\text{diag}(\pm 1, \dots, \pm 1)$. Однако если какой-нибудь диагональный блок матрицы A имеет одинаковые элементы, то возникает дополнительная группа блочно-ортогональных матриц, сопряжение которыми не изменяет I , D и диагональные элементы A .

Запись канонического многочлена с точностью до $\text{diag}(\pm 1, \dots, \pm 1)$ позволяет найти размерность пространства модулей многообразий: согласно теореме 2, она равна размерности пространства канонических многочленов. Далее предполагается, что характеристический многочлен записан в каноническом виде.

Лемма 1. Пусть $D = \text{diag}(0, \dots, 0, 1, d_2, \dots, d_{m-k})$ (где k – количество нулей), $A > 0$ и выполнено любое из условий:

- (a) $I - DA^{-1}D > 0$;
- (b) $I - DA^{-1}D \geq 0$ и $\det(Q(s)) > 0$ для всех $s \in \mathbb{R}$.

Тогда

- (c) $Q(s) > 0$ для любого $s \in \mathbb{R}$.

Более того, условия (b) и (c) равносильны.

Доказательство. Так как $A > 0$, то из A извлекается единственный положительный квадратный корень $A^{1/2}$. Преобразуем многочлен $Q(s)$:

$$\begin{aligned} Q(s) &= s^2 I + 2sD + A = A^{1/2}(s^2 A^{-1} + 2sA^{-1/2}DA^{-1/2} + I)A^{1/2} = \\ &= s^2(I - DA^{-1}D) + A^{1/2}(I + sA^{-1/2}DA^{-1/2})^2 A^{1/2}. \end{aligned} \quad (9)$$

Пусть верно (a). Если $s \neq 0$, то $Q(s) > 0$ согласно (9); при $s = 0$ утверждение леммы выполняется, так как $Q(0) = A > 0$.

Покажем, что (b) эквивалентно (c). Если $I - DA^{-1}D \geq 0$, то, согласно (9), $Q(s) \geq 0$; кроме того, из (b) следует, что $Q(s)$ невырождена. Следовательно, $Q(s) > 0$. Обратно, если $Q(s) > 0$ для любого s , то $\det(Q(s)) > 0$ для любого $s \in \mathbb{R}$, и, очевидно, коэффициент при s^2 в (9) неотрицателен, то есть выполняется (b). ■

Для классификации с точностью до причинной почти изометрии (т.е. с точностью до выбора решетки Γ в T) плоских полных строго причинных лоренцевых многообразий достаточно описать все многочлены вида (8), удовлетворяющие условиям пункта (ii) Теоремы 1, и указать, когда полученные многообразия эквивалентны.

2. Классификация

Пусть $Q_M(s)$ — канонический многочлен вида (8). Положим $k_D = \dim(\ker D)$.

Лемма 2. Пусть M — многообразие сигнатуры $(m, n, r, 0)$. Тогда $r \geq k_D$.

Доказательство. Обозначим $F = I - DA^{-1}D$. Поскольку k_D первых диагональных элементов матрицы D равны нулю, $F = \begin{pmatrix} I_{k_D} & \mathbf{0} \\ \mathbf{0} & f \end{pmatrix}$. Утверждение леммы следует из равенства (5), которое в данном случае может быть записано в виде $r = \text{rank}(F)$. ■

Введем обозначения для серий многообразий, выделяемых условиями на сигнатуры и k_D .

$$\begin{aligned} S_1 : & \quad (n, m, m, 0), \quad k_D = m; \\ S_2 : & \quad (n, m, k_D, 0); \\ S_3 : & \quad (n, m, k_D + 1, 0). \end{aligned}$$

Предложение 2. Плоское полное строго причинное лоренцево многообразие сигнатуры $(n, m, r, 0)$ и размерности не более 8 либо принадлежит одной из серий S_1, S_2, S_3 , либо имеет сигнатуру $(8, 3, 3, 0)$ при $k_D = 1$.

Доказательство. Перебирая все возможные варианты для n, m, r и k_D , с соблюдением неравенств $n \leq 8$, $m + r = n - 2$, $r \leq m$ и $k_D \leq r$, получаем таблицу 1, из которой следует предложение. ■

Замечание 5. Предложение 2 не исчерпывает всех многообразий в размерностях до 8, поскольку не указаны сигнатуры (n, m, r, k) при $k \neq 0$. Любое многообразие такой сигнатуры почти причинно изометрично произведению многообразия сигнатуры $(n - k, m - k, r, 0)$ и k -мерного тора $\mathbb{T}^k$, согласно теореме 1, (iii).

На множестве парабол в конусе P_m положительно определенных матриц, рассматриваемых с точностью до преобразований (2) с $X \in \text{GL}(m, \mathbb{R})$, имеется естественная топология, которую можно определить, например, так. Стандартная евклидова метрика ρ в пространстве матриц однозначно определяет вершину параболы — это точка, в которой касательная перпендикулярна оси. Это свойство не зависит от параметризации. Переместим параболу так, чтобы вершина попала в единичную матрицу I . Пусть $\mathcal{B}$ — замкнутый шар достаточно

Таблица 1.

$n = \dim M$	$m = \dim T$	$r = \dim R$	$k_D = \dim(\ker(D))$	Серия
4	1	1	1	S_1
5	2	1	1	S_2
6	2	2	1	S_3
			2	S_1
	3	1	1	S_2
7	3	2	1	S_3
			2	S_2
	4	1	1	S_2
8	3	3	1	отдельный
			2	S_3
			3	S_1
	4	2	1	S_3
			2	S_2
	5	1	1	S_2

малого радиуса относительно ρ с центром в I . Расстояние по Хаусдорфу на множестве $\mathcal{P}$ пересечений парабол с вершинами в I с шаром $\mathcal{B}$ задает метрику на $\mathcal{P}$. Очевидно, она инвариантна относительно стационарной группы $K = \text{SO}(m)$ точки I , которая естественным образом действует на $\mathcal{P}$. Поскольку K компактна, метрика Хаусдорфа на семействе K -орбит в $\mathcal{P}$ определяет метрику на пространстве модулей рассматриваемого класса многообразий фиксированной сигнатуры σ . Обозначим последнее через $\mathcal{M}_\sigma$ и будем считать в дальнейшем, что оно снабжено отвечающей метрике топологией. Очевидно, замыкание множества невырожденных парабол содержит семейство лучей (но не точек, которые соответствуют эллиптическим многообразиям), а семейство многообразий с простым спектром плотно в $\mathcal{M}_\sigma$.

Обозначим через $\mathcal{M}_\sigma(S_k)$ множество многообразий из $\mathcal{M}_\sigma$, принадлежащих серии S_k , $k = 1, 2, 3$.

Предложение 3. *Многочлен вида (8) соответствует многообразию серии S_1 тогда и только тогда, когда $D = \mathbf{0}$ и $A > 0$. При этом A диагональна. Семейство многообразий серии S_1 сигнатуры $(n, m, m, 0)$ параметризуется $(m - 1)$ вещественным параметром.*

Доказательство. Условие $k_D = m$ эквивалентно равенству $D = 0$. Поэтому условие (3) принимает вид

$$Is^2 + A > 0 \quad \text{для любого } s \in \mathbb{R},$$

что равносильно неравенству $A > 0$. Так как $C = I$ и $B = D = \mathbf{0}$, то выполняются условия (4) и (5). Первое утверждение следует из теоремы 1, (ii).

Матрица A имеет блочно-диагональную структуру (один блок), что влечет ее диагональность согласно определению канонического многочлена. Таким образом, $A = \text{diag}(a_1, \dots, a_m)$ и $0 < a_1 \leq \dots \leq a_m$. Проводя замену $s \rightarrow \alpha s$ и сопрягая

некоторой скалярной матрицей, можно добиться выполнения равенства $a_1 = 1$. Канонический многочлен однозначно определяется диагональными элементами матрицы A , что и доказывает последнее утверждение. ■

Пусть $Q(s)$ — многочлен вида (8). Введем обозначения:

$$F = I - DA^{-1}D, \quad A^{-1} = \begin{pmatrix} \tilde{A} & b \\ b^\top & G \end{pmatrix} = (a^{ij})_{ij=1}^m,$$

$$D = \text{diag}(0, \dots, 0, d_1, d_2, \dots, d_{m-k_D}),$$

причем $d_1 = 1$; $F = (f_{ij})_{ij=1}^m$. В силу определения канонического вида многочлена, $\tilde{A}$ -диагональная матрица размера k_D . Непосредственные вычисления дают

$$\begin{aligned} f_{ij} &= \delta^{ij} \quad \text{при} \quad i \leq k_D \quad \text{или} \quad j \leq k_D, \\ f_{ij} &= \delta^{ij} - d_{i-k_D} d_{j-k_D} a^{ij} \quad \text{для} \quad i > k_D, j > k_D, \end{aligned} \quad (10)$$

где δ^{ij} — символ Кронекера. Таким образом, F имеет вид

$$F = \begin{pmatrix} I_{k_D} & \mathbf{0} \\ \mathbf{0} & \tilde{F} \end{pmatrix}. \quad (11)$$

Лемма 3. Для канонического характеристического многочлена многообразия серии S_2 условия (4) и (5) эквивалентны тому, что

$$G = \text{diag}(1, d_2^{-2}, \dots, d_{m-k_D}^{-2}). \quad (12)$$

Доказательство. Используя (11), из условия (5) получаем, что $\tilde{F} = 0$. Согласно (10), $\delta^{ij} - d_{i-k_D} d_{j-k_D} a^{ij} = 0$ для всех i, j , таких, что $i > k_D$ и $j > k_D$. Таким образом, $a^{ij} = \frac{\delta^{ij}}{d_{i-k_D} d_{j-k_D}}$. Но элементы с такими индексами образуют матрицу G , откуда следует (12). Таким образом, из (4) и (5) следует утверждение леммы. Обратное получается непосредственной подстановкой A^{-1} (определяемой G) в (4) и (5). ■

Предложение 4. Многочлен $Q(s)$ вида (8) определяет некоторое многообразие M серии S_2 сигнатуры $(n, t, k_D, 0)$ тогда и только тогда, когда матрица A такова, что

(a) $\det Q(s) > 0$ для любого s ;

(b) $A^{-1} > 0$ и выполняется (12).

Множество многообразий таких, что спектр $\tilde{A}$ прост, все d_i различны и $d_i \neq 1$, образует $k_D(m - k_D) + m - 1$ — параметрическое семейство, плотное в $\mathcal{M}_\sigma(S_2)$.

Доказательство. Очевидна эквивалентность (а), (b) и (3)–(5), в силу лемм 1 и 3.

При фиксированных матрицах $\tilde{A}$ и G на элементы матрицы b ограничения могут быть, например, следующим способом: выделим полные квадраты в квадратичной форме, соответствующей матрице A^{-1} . Положительность коэффициентов при квадратах и даст условия на b .

Матрицы D и $\tilde{A}$ задаются своими диагональными элементами, это $m - 1$ параметр; при фиксированных $\tilde{A} > 0$ и $G > 0$ условие $A^{-1} > 0$ выполняется при любой достаточно малой матрице b , что дает еще $k_D(m - k_D)$ параметр. Последнее утверждение очевидно. ■

Лемма 4. Для канонического характеристического многочлена многообразия серии S_3 условия (4) и (5) эквивалентны следующим ограничениям на элементы A^{-1} :

$$a_{i+k_D i+k_D} \leq d_i^{-2} \quad \text{для } i = 1, \dots, m - k_D, \quad (13)$$

$$a_{i+k_D j+k_D} = -\frac{\sqrt{(1-d_i^2 a_{i+k_D i+k_D})(1-d_j^2 a_{j+k_D j+k_D})}}{d_i d_j} \quad \text{при } d_i \neq d_j, \quad (14)$$

$$a_{i+k_D j+k_D} = 0 \quad \text{и } (1 - g_{i+k_D i+k_D} d_i^2)(1 - g_{j+k_D j+k_D} d_j^2) = 0 \quad \text{при } d_i = d_j. \quad (15)$$

При этом в (13) хотя бы одно неравенство строгое.

Доказательство. Покажем, что (4) и (5) влекут (13)–(15). Из (5) следует $\text{rank}(F) = k_D + 1$, то есть $\text{rank}(f) = 1$. Это равносильно тому, что любой минор порядка два равен нулю и $f \neq 0$. Из условия равенства нулю диагональных миноров получаем

$$a_{i+k_D j+k_D} = \epsilon_{ij} \frac{\sqrt{(1-d_i^2 a_{i+k_D i+k_D})(1-d_j^2 a_{j+k_D j+k_D})}}{d_i d_j} \quad \text{для всех } i \neq j \in \overline{1, m - k_D}, \quad (16)$$

где $\epsilon_{ij} = \pm 1$ отвечает за выбор знака. При выборе других миноров получаем соотношения на ϵ , а именно

$$\begin{aligned} \epsilon_{i_1 j_1} \epsilon_{i_2 j_2} &= \epsilon_{i_1 j_2} \epsilon_{i_2 j_1} \quad \text{для всех попарно не равных } i_1, j_1, i_2, j_2, \\ -\epsilon_{i_1 j_1} &= \epsilon_{i_2 j_1} \epsilon_{i_1 i_2}. \end{aligned} \quad (17)$$

Очевидно, что выбор знаков некоторой строки однозначно определяет знаки всех остальных элементов (пусть выбраны знаки k -той строки, в силу симметричности матрицы, и k -того столбца, тогда $\epsilon_{ij} = -\epsilon_{ik} \epsilon_{kj}$, согласно (17)). Согласно определению канонического многочлена элементы $(k_D + 1)$ -ой строки A неположительны. Если эта строка не содержит нулей, то знаки всех элементов определены однозначно (так как заданы знаки одной строки). Если же в соответствующей строке присутствуют нули, то $\epsilon_{i,j}$ определены неоднозначно. Это соответствует сопряжению канонического многочлена матрицей $X = \text{diag}(\pm 1, \dots, \pm 1)$, в которой все элементы кроме одного однозначно определены. Таким образом, условие отрицательности элементов $(k_D + 1)$ -ой строки A (по определению канонического многочлена) и (16) влекут (14). Если $d_i = d_j$, то,

согласно определению канонического вида (A – блочно-диагональная), данная часть матрицы A^{-1} диагональна, откуда следует (15). Кроме того, необходимо, чтобы хотя бы один элемент матрицы f не был равен нулю. Заметим, что матрица f равна нулю, если все ее диагональные элементы равны нулю. Это невозможно, если $\text{rank}(f) = 1$. Таким образом, мы получаем в качестве необходимого следующее условие:

$$a_{i+k_D i+k_D} \neq d_i^{-2} \quad \text{для некоторого } i \in \{1, \dots, m - k_D\}. \quad (18)$$

Из (4) следует $\langle F e_{i+k_D}, e_{i+k_D} \rangle = 1 - d_i^2 a_{i+k_D i+k_D} \geq 0$ для любого $i \in \{1, \dots, m - k_D\}$. В сочетании с (18) это приводит к (13).

Обратно, из (14) и (15) следует (5) (проверяется непосредственной подстановкой). Кроме того, (14) влечет положительность следа матрицы f (так как все диагональные элементы f неотрицательны), что, в сочетании с условием $\text{rank}(f) = 1$ (оно же (5)), дает $f \geq 0$. Поэтому $F \geq 0$, а это и есть (4). ■

Предложение 5. *Многочлен $Q(s)$ вида (8) определяет некоторое унитарное многообразие M серии S_3 , тогда и только тогда, когда*

- (a) $\det Q(s) > 0$ для любого s ;
- (b) матрица $A^{-1} > 0$ удовлетворяет условиям (13)–(15) леммы 4.

Множество многообразий, таких, что спектр $\tilde{A}$ прост, все d_i различны и $d_i \neq 1$, образует $(k_D(m - k_D) + 2m - k_D - 1)$ -параметрическое семейство, плотное в $\mathcal{M}_\sigma(S_3)$.

Доказательство. Покажем, что (a) и (b) эквивалентны (3)–(5). Условия (13)–(15) эквивалентны (4) и (5) (см. лемму 4), тогда как (a) и $A^{-1} > 0$ (или, что тоже самое, $A > 0$) эквивалентны условию (3) (см. лемму 1). Если $A^{-1} > 0$, то небольшое возмущение b не приводит к нарушению условия положительности; это $k_D(m - k_D)$ параметр. Матрицы D , $\tilde{A}$ и G определяются своими диагональными элементами, что дает еще $2m - k_D - 1$ параметр. Последнее утверждение очевидно. ■

Предложение 6. *Многочлен $Q(s)$ вида (8) определяет некоторое унитарное многообразие M сигнатуры $(8, 3, 3, 0)$, при $k_D = 1$, тогда и только тогда, когда*

- (a) $A > 0$;
- (b) элементы матрицы $A^{-1} = (\tilde{a}^{ij})_{ij=1}^3$ удовлетворяют соотношениям

$$a^{2,2} < 1, \quad a^{3,3} < \frac{1}{d_2}, \quad (a^{2,3})^2 < \frac{(1 - a^{2,2})(1 - d_2^2 a^{3,3})}{d_2^2}.$$

Канонические характеристические многочлены многообразий общего положения ($d \neq 1$), имеющих данную сигнатуру, образуют 7-ми параметрическое семейство.

Доказательство. Пусть $D = \text{diag}(0, 1, d)$. Тогда $F = I_3 - DA^{-1}D$ имеет вид:

$$F = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 - a^{2,2} & da^{2,3} \\ 0 & da^{2,3} & 1 - d^2a^{3,3} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & f \end{pmatrix},$$

где f — 2×2 -матрица. Покажем, что условия (3)–(5) влекут (а) и (b). Очевидно, что из (3) следует (а). Так как выполнены условия пункта (а) леммы 1, то (а) и (b) влекут (3).

В условиях предложения условия (4) и (5) эквивалентны условию $F > 0$. Последнее равносильно тому, что $f > 0$. Так как f должна быть положительно определена, то условие $\langle fe_i, e_i \rangle > 0$, где e_i — базисные векторы, необходимо. Таким образом, получены необходимые условия: $1 - a^{2,2} > 0$ и $1 - d^2a^{3,3} > 0$. Таким образом, след f положителен. Так как f имеет порядок 2, то для положительности f необходимо и достаточно, чтобы $\det f > 0$. Это условие и задает нужный вид матрицы A . Поскольку симметрическая матрица A^{-1} определяется своими элементами на диагонали и выше, это дает 6 параметров. Осталось заметить, что матрица D зависит от одного параметра. ■

В качестве следствия из предложений 3 – 6 получаем таблицу 2.

Таблица 2.

Серия	Размерность $M_\sigma(S_i)$
S_1	$m - 1$
S_2	$r(m - r) + m - 1$
S_3	$r(m - r) + r + m - 1$
отдельный	7

Полученные результаты позволяют конкретизировать алгоритм (A)–(C) на странице 27 для размерностей не выше 8.

- (α) Фиксируем изотропные векторы v_0, v_1 , такие, что $\ell(v_0, v_1) = 1$, определяем подпространства L, W, N соотношениями

$$L = \mathbb{R}v_0, \quad W = L^\perp, \quad N = W \cup v_1^\perp.$$

Выбор подпространств N и $T \subseteq N$ в пространстве V определяет сигнатуру

$$(n, m, r, 0) = (\dim V, \dim T, \dim(T_N^\perp), 0).$$

После этого выбирается диагональная неотрицательная матрица D так, чтобы

$$k_D = \dim(\ker D) \leq r,$$

и определяется, к какой серии принадлежит многообразие, согласно таблице 1. Матрица A определяется из предложений 3 – 6 (в зависимости от серии). Это задает характеристический многочлен $Is^2 + 2Ds + A$.

(β) Отображение a определяется из соотношений

$$\begin{aligned}\langle Az, z \rangle &= \langle x, x \rangle, \\ \langle Dz, z \rangle &= \langle a'x, x \rangle, \\ \langle Iz, z \rangle &= \langle a'x, a'x \rangle + \langle a''x, a''x \rangle,\end{aligned}$$

где $x = \iota z \in T$, $z \in \mathbb{R}^m$; кроме того, $\Gamma = \iota \mathbb{Z}^m$ (см. [7] (Теорема 3)).

Согласно [7], алгоритм (A)–(C) позволяет построить любое плоское полное лоренцево строго причинное многообразие.

В следующей теореме сформулированы основные результаты работы.

Теорема 3. (1) Любое плоское полное лоренцево строго причинное многообразие размерности не выше 8 может быть построено по уточненному алгоритму (α)–(β).

(2) Размерности пространств модулей $\mathcal{M}_\sigma$ задаются таблицей 3:

Таблица 3.

Сигнатура $\sigma = (n, m, r, 0)$	Размерность $\mathcal{M}_\sigma$
(4, 1, 1, 0)	0
(5, 2, 1, 0)	2
(6, 2, 2, 0)	3
(6, 3, 1, 0)	4
(7, 3, 2, 0)	6
(7, 4, 1, 0)	6
(8, 3, 3, 0)	7
(8, 4, 2, 0)	9
(8, 5, 1, 0)	8

Доказательство. Пункт (1) следует из возможности построения многообразия по алгоритму (A)–(C) и существования для данного многообразия канонического многочлена. Пункт (2) следует из предложения 2 и таблицы 2. ■

ЛИТЕРАТУРА

1. Abels H. Properly Discontinuous Groups of Affine Transformations: A Survey // Geometriae Dedicata. 2001. N.87. P.309-333.
2. Barbot T. Globally hyperbolic flat space-times // Journal of Geometry and Physics. 2005. N.53. P.123-165.
3. Beem J.K., Ehrlich P.E. Global Lorentzian geometry // Monographs and Textbooks in Pure and Applied Mathematics, 2nd ed., vol. 202, Marcel Dekker, New York. 1996.
4. Charette V., Drumm T., Goldman W., Morill M. Complete Flat Affine and Lorentzian Manifolds // Geometriae Dedicata. 2003. N.97. P.187-198.
5. Ellis G., Hawking S. The large scale structure of space-time // Cambridge Monographs on Mathematical Physics. N.1. Cambridge University Press, London, New York, 1973.

-
6. Fried D. Flat Spacetimes // J. Diff. Geom. 1987. N.26. P.385-396.
 7. Гичев В.М., Мещеряков Е.А. О геометрии плоских полных лоренцевых строго причинных многообразий // Принята к печати в Сибирский Математический Журнал.
 8. Gichev V.M., Morozov O.S. On flat complete causal Lorentzian manifolds // Geometriae Dedicata. 2005. N.116. P.37-59.
 9. Milnor J. On fundamental groups of complete affinely flat manifolds // Adv. in Math. 1977. N.25. P.178-187.
 10. Wolf J.A. Spaces of constant curvature. University of Californnia, Berkley, California, 1972.

ПРОГНОЗИРОВАНИЕ ПОВЕДЕНИЯ РАБОТАЮЩИХ НА РЫНКЕ ТРУДА ОМСКОЙ ОБЛАСТИ

Е.В. Ренжина, А.К. Гуц

В данной работе приводится математическая модель стратегий экономического поведения индивидов на рынке труда, позволяющая осуществлять прогнозирование состояния рынка труда Омской области.

С формированием рыночных отношений, ранее не существовавших в Советской России, в жизни каждого человека большое значение приобретает знание механизмов экономического поведения людей.

Одной из стадий формирования рыночных отношений является появление такого понятия, как рынок труда, который, по мнению И.С. Масловой [1], определяется следующим образом: «рынок труда» – органическая сфера рыночной экономики, выполняющая функцию опосредования через куплю-продажу рабочей силы, соединения вещественных и человеческих факторов производства, поддерживая их сбалансированность в условиях многообразия форм собственности на средства производства и преимущественного права граждан на распоряжение своей рабочей силой и способностями».

Каждый индивид имеет свою стратегию SW экономического поведения на рынке труда, складывающуюся из определенного количества затраченных трудовых ресурсов для получения определенного дохода, то есть разнообразие стратегий экономического поведения людей складывается из трех факторов:

1. Величина трудовых ресурсов.
2. Величина труда.
3. Область труда, в которой соединяются труд и доход в необходимой пропорции [2].

1. Распределение стратегии экономического поведения людей на рынке труда

Экономическое поведение людей формируется под влиянием широкой сферы экономической деятельности и, соответственно, должно охватывать все возмож-

ные варианты соединения труда и дохода – типы стратегий, которые оптимизируют поведение человека на рынке труда. Как показали исследования О.С. Ёлкиной и В.С. Половинко [1], эти варианты формируют матрицу стратегий – некую область Ω в пространстве с осями T, D (труд и доход), которую наглядно можно представить в следующем виде (Рис. 1).

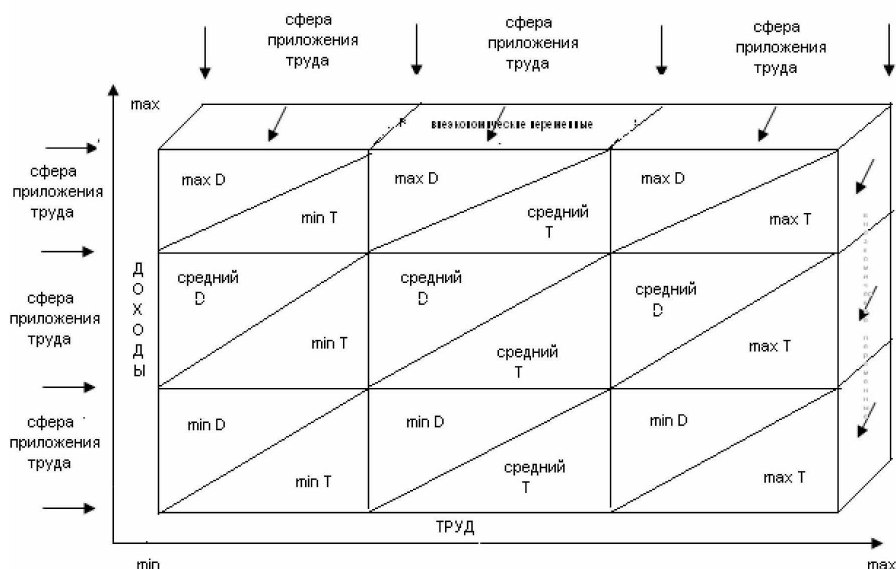


Рис. 1. Область, представляющая стратегии экономического поведения индивидов на рынке труда

В [1] были приведены следующие результаты статистических исследований рынка труда Омской области в 1999 г. (Таблица 1).

2. Численное описание стратегий экономического поведения

В статье [2] рынок труда Омской области моделируется с помощью следующей краевой задачи:

$$\frac{\partial S}{\partial t} = a\Delta S, \quad (1)$$

$$S|_{t=t_0} = S_0, \quad S|_{[0, \infty) \times \partial\Omega} = 0, \quad (2)$$

где Ω – это область приложения труда. Таким образом, мы считаем рынок труда Омской области изолированным пространством.

Решение задачи (1)-(2) имеет вид:

$$S(D, T, t) = \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} A_{k,j} e^{-a^2 \pi^2 (\frac{k^2}{l^2} + \frac{j^2}{m^2}) t} \sin \frac{k\pi D}{l} \sin \frac{j\pi T}{m}. \quad (3)$$

Таблица 1. Распределение стратегий работающих на рынке труда города Омска
1999 г. и 2005 г.

Типы стратегий	Max D/ Max T	Max D/ Cp T	Max D/ Min T	Cp D/ Max T	Cp D/ Cp T	Cp D/ Min T	Min D/ Max T	Min D/ Cp T	Min D/ Min T
Количество респондентов, занятых на рынке труда в процентном соотношении в 1999 г.	14,8	33,2	5,5	4,6	12,5	5,2	3	13,5	7,7
Количество респондентов, занятых на рынке труда в процентном соотношении в 2005 г.	9,64	35,14	5,94	4,34	12,24	5,94	3,74	13,34	9,94

Полагая, что $\Omega = [0, \pi] \times [0, \pi]$, получаем

$$S(D, T, t) = \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} A_{k,j} e^{-a^2(k^2+j^2)t} \sin kD \sin jT. \quad (4)$$

Для проверки адекватности предлагаемой модели рынка труда воспользуемся результатами исследований экономиста О.С. Ёлкиной [2]. Первое статистическое исследование ею было проведено в 1999 году, второе – в 2005 году (см. табл.1).

Рассмотрим приближенное решение задачи, состоящее из первых девяти членов ряда:

$$S(D, T, t) = \sum_{k=1}^3 \sum_{j=1}^3 A_{k,j} e^{-a^2(k^2+j^2)t} \sin kD \sin jT. \quad (5)$$

Формула (5) описывает поведение работающих индивидов в какой-то одной точке (D, T) . Для нахождения девяти коэффициентов $A_{k,j}$ проинтегрируем (5) по каждому из девяти секторов области Ω :

$$\int_{\alpha_1^m}^{\alpha_2^m} \int_{\beta_1^n}^{\beta_2^n} S(D, T, t) dDdT = \sum_{k=1}^3 \sum_{j=1}^3 A_{k,j} e^{-a^2(k^2+j^2)t} \int_{\alpha_1^m}^{\alpha_2^m} \int_{\beta_1^n}^{\beta_2^n} \sin kD \sin jT dDdT, \quad (6)$$

где $[\alpha_1^m, \alpha_2^m]$ и $[\beta_1^n, \beta_2^n]$ ($m, n = 1, 2, 3$) – произвольный сектор области Ω .

Величину

$$\widehat{SW}_{m,n} = \int_{\alpha_1^m}^{\alpha_2^m} \int_{\beta_1^n}^{\beta_2^n} S(D, T, t) dDdT$$

трактуем как число работников, отвечающих стратегии поведения соответствующего сектора области Ω .

Далее для описания модели распределения стратегии поведения работающих на рынке труда для $t=1999$ -й год составляем систему из девяти уравнений:

$$\sum_{k=1}^3 \sum_{j=1}^3 A_{k,j}^{(1999)} e^{-a^2(k^2+j^2)1999} \int_0^{\frac{\pi}{3}} \int_0^{\frac{\pi}{3}} \sin kD \sin jT dDdT = SW_{1,1}(1999);$$

$$\sum_{k=1}^3 \sum_{j=1}^3 A_{k,j}^{(1999)} e^{-a^2(k^2+j^2)1999} \int_0^{\frac{\pi}{3}} \int_{\frac{\pi}{3}}^{\frac{2\pi}{3}} \sin kD \sin jT dDdT = SW_{1,2}(1999);$$

$$\sum_{k=1}^3 \sum_{j=1}^3 A_{k,j}^{(1999)} e^{-a^2(k^2+j^2)1999} \int_0^{\frac{\pi}{3}} \int_{\frac{2\pi}{3}}^{\pi} \sin kD \sin jT dDdT = SW_{1,3}(1999);$$

$$\sum_{k=1}^3 \sum_{j=1}^3 A_{k,j}^{(1999)} e^{-a^2(k^2+j^2)1999} \int_{\frac{\pi}{3}}^{\frac{2\pi}{3}} \int_0^{\frac{\pi}{3}} \sin kD \sin jT dDdT = SW_{2,1}(1999);$$

$$\sum_{k=1}^3 \sum_{j=1}^3 A_{k,j}^{(1999)} e^{-a^2(k^2+j^2)1999} \int_{\frac{\pi}{3}}^{\frac{2\pi}{3}} \int_{\frac{\pi}{3}}^{\frac{2\pi}{3}} \sin kD \sin jT dDdT = SW_{2,2}(1999);$$

$$\sum_{k=1}^3 \sum_{j=1}^3 A_{k,j}^{(1999)} e^{-a^2(k^2+j^2)1999} \int_{\frac{\pi}{3}}^{\frac{2\pi}{3}} \int_{\frac{2\pi}{3}}^{\pi} \sin kD \sin jT dDdT = SW_{2,3}(1999);$$

$$\sum_{k=1}^3 \sum_{j=1}^3 A_{k,j}^{(1999)} e^{-a^2(k^2+j^2)1999} \int_{\frac{2\pi}{3}}^{\pi} \int_0^{\frac{\pi}{3}} \sin kD \sin jT dDdT = SW_{3,1}(1999);$$

$$\sum_{k=1}^3 \sum_{j=1}^3 A_{k,j}^{(1999)} e^{-a^2(k^2+j^2)1999} \int_{\frac{2\pi}{3}}^{\pi} \int_{\frac{\pi}{3}}^{\frac{2\pi}{3}} \sin kD \sin jT dDdT = SW_{3,2}(1999);$$

$$\sum_{k=1}^3 \sum_{j=1}^3 A_{k,j}^{(1999)} e^{-a^2(k^2+j^2)1999} \int_{\frac{2\pi}{3}}^{\pi} \int_{\frac{2\pi}{3}}^{\pi} \sin kD \sin jT dDdT = SW_{3,3}(1999),$$

где $SW_{m,n}(1999)$ берутся из базы данных, полученной при обследовании рынка труда Омской области в 1999 году [2, с.136]. Вычисляя определенные интегралы, мы упрощаем эту систему и получаем ее в следующем виде:

$$\sum_{k=1}^3 \sum_{j=1}^3 A_{k,j}^{(1999)} e^{-a^2(k^2+j^2)1999} \frac{1}{k * j} (\cos 0 - \cos \frac{\pi}{3}k) (\cos 0 - \cos \frac{\pi}{3}j) = SW_{1,1}(1999);$$

$$\begin{aligned}
& \sum_{k=1}^3 \sum_{j=1}^3 A_{k,j}^{(1999)} e^{-a^2(k^2+j^2)1999} \frac{1}{k * j} (\cos 0 - \cos \frac{\pi}{3}k) (\cos \frac{\pi}{3}j - \cos \frac{2\pi}{3}j) = SW_{1,2}(1999); \\
& \sum_{k=1}^3 \sum_{j=1}^3 A_{k,j}^{(1999)} e^{-a^2(k^2+j^2)1999} \frac{1}{k * j} (\cos 0 - \cos \frac{\pi}{3}k) (\cos \frac{2\pi}{3}j - \cos \pi j) = SW_{1,3}(1999); \\
& \sum_{k=1}^3 \sum_{j=1}^3 A_{k,j}^{(1999)} e^{-a^2(k^2+j^2)1999} \frac{1}{k * j} (\cos \frac{\pi}{3}k - \cos \frac{2\pi}{3}k) (\cos 0 - \cos \frac{\pi}{3}j) = SW_{2,1}(1999); \\
& \sum_{k=1}^3 \sum_{j=1}^3 A_{k,j}^{(1999)} e^{-a^2(k^2+j^2)1999} \frac{1}{k * j} (\cos \frac{\pi}{3}k - \cos \frac{2\pi}{3}k) (\cos \frac{\pi}{3}j - \cos \frac{2\pi}{3}j) = SW_{2,2}(1999); \\
& \sum_{k=1}^3 \sum_{j=1}^3 A_{k,j}^{(1999)} e^{-a^2(k^2+j^2)1999} \frac{1}{k * j} (\cos \frac{\pi}{3}k - \cos \frac{2\pi}{3}k) (\cos \frac{2\pi}{3}j - \cos \pi j) = SW_{2,3}(1999); \\
& \sum_{k=1}^3 \sum_{j=1}^3 A_{k,j}^{(1999)} e^{-a^2(k^2+j^2)1999} \frac{1}{k * j} (\cos \frac{2\pi}{3}k - \cos \pi k) (\cos 0 - \cos \frac{\pi}{3}j) = SW_{3,1}(1999); \\
& \sum_{k=1}^3 \sum_{j=1}^3 A_{k,j}^{(1999)} e^{-a^2(k^2+j^2)1999} \frac{1}{k * j} (\cos \frac{2\pi}{3}k - \cos \pi k) (\cos \frac{\pi}{3}j - \cos \frac{2\pi}{3}j) = SW_{3,2}(1999); \\
& \sum_{k=1}^3 \sum_{j=1}^3 A_{k,j}^{(1999)} e^{-a^2(k^2+j^2)1999} \frac{1}{k * j} (\cos \frac{2\pi}{3}k - \cos \pi k) (\cos \frac{2\pi}{3}j - \cos \pi j) = SW_{3,3}(1999).
\end{aligned}$$

Относительно переменных $A_{k,j}^{(1999)}$ данная система является линейной, и для нахождения ее решения используем программный пакет MathCad 11. Результат приведен в виде диаграммы на рис. 2. Слева на рис. 2 даны величины $SW_{m,n}(1999)$ [2, с.136], а справа прогнозируемые на 2005 год стратегии работающих на рынке труда города Омска, найденные с помощью формулы (6). Точнее, по формуле

$$\begin{aligned}
\widehat{SW}_{m,n}(2005) &= \int_{\alpha_1^m}^{\alpha_2^m} \int_{\beta_1^n}^{\beta_2^n} S(D, T, 2005) dD dT = \\
&= \sum_{k=1}^3 \sum_{j=1}^3 A_{k,j}^{(1999)} e^{-a^2(k^2+j^2)2005} \int_{\alpha_1^m}^{\alpha_2^m} \int_{\beta_1^n}^{\beta_2^n} \sin kD \sin jT dD dT. \quad (7)
\end{aligned}$$

Спрогнозированные величины $\widehat{SW}_{m,n}(2005)$ были соотнесены с базой данных $SW_{m,n}(2005)$, полученной при обследовании рынка труда Омской области в 2005 году [2, с.136]. Отклонение не превысило 5%.

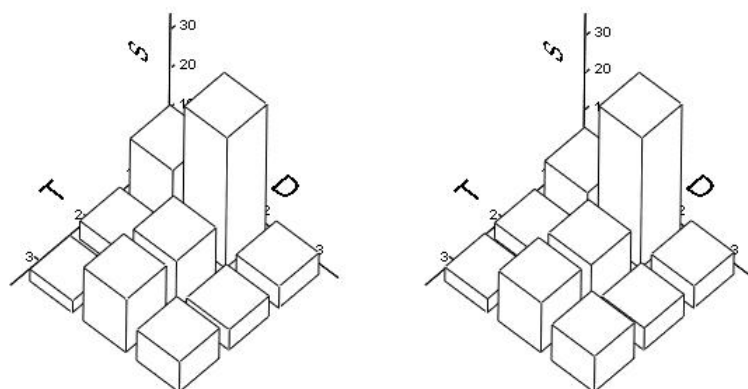


Рис. 2. Базы данных величин $SW_{m,n}(1999)$ [2, с.136] (слева). Прогноз распределения стратегий работающих на рынке труда города Омска в 2005 г. (справа)

Таким образом, использование формулы (2) позволяет с точностью до 5% описывать поведение работающих индивидов на рынке труда г. Омска.

Аналогично, используя коэффициенты $A_{k,j}^{(1999)}$, можно сделать прогноз по рынку труда на 2011-й год (рис. 3).

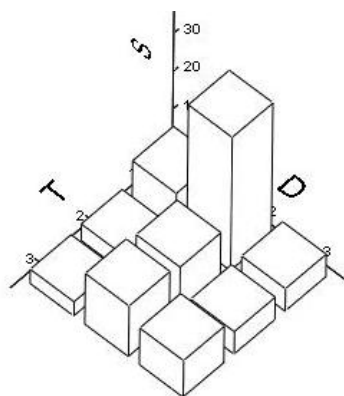


Рис. 3. Расчетное распределение стратегий работающих на рынке труда города Омска в 2011 г.

Из рис. 3 видно, что больших изменений в стратегиях поведения на рынке труда до 2011 года ожидать не приходится. Работники обладают устойчивыми стратегическими направленностями. «Первое исследование было проведено в 1999 году - после экономического кризиса. Условия 2005 года приближаются к условиям, которые мы имели в 1998 году до кризиса» [1]. Таким образом, два исследования были проведены в различных экономических условиях. То, что мы получили практически одинаковые результаты, подтверждает теоретические положения, разработанные О.С. Елкиной и В.С. Половинко, о стабильности и устойчивости стратегических жизненных ориентиров людей на рынке труда [1]. Более того, это свидетельствует о том, что внешние экономические

условия не в значительной степени влияют на распределение стратегий экономического поведения работника на рынке труда, а влияют в большей мере внутренние факторы, их формирующие.

В заключение следует сказать, что наша математическая модель, с одной стороны, позволяет отразить тонкое наблюдение специалистов – экономистов [1] по исследованию рынка труда, а с другой – позволяет создавать модель, прогнозирующую состояние рынка труда.

ЛИТЕРАТУРА

1. Ёлкина О.С., Половинко В.С. Экономическое поведение работников на рынке труда. Омск: Изд-во ОмГПУ, 2001. 278 с.
2. Ёлкина О.С., Гуревич Е.В., Гуц А.К. Математическое моделирование стратегий экономического поведения людей на рынке труда // Математические структуры и моделирование. 2005. Вып.15. С.107-111.
3. Пу Т. Нелинейная экономическая динамика. Ижевск: Издательский дом «Удмуртский университет», 2000. 200 с.

КВАНТОВОЕ РОЖДЕНИЕ ФИЗИЧЕСКОЙ РЕАЛЬНОСТИ И МАТЕМАТИЧЕСКОЕ ОПИСАНИЕ ОСОЗНАНИЯ

А.К. Гуц

Предлагается квантовая модель рождения физического реального мира с помощью индивидуальных сознаний. Реальность – это результат скоррелированных индивидуальных «фантазий». Соотнесение в мозгу «фантазии» с отраженной реальностью есть осознание (момент присутствия).

1. Что такое физическая реальность?

Для исследователя, работающего в естествознании, очевидно существование внешнего мира. Стены, деревья *всем* видны и препятствуют попыткам пройти сквозь них и, следовательно, *реальны*, а сны, фантазии доступны только их владельцу и поэтому *нереальны*. Но виртуальная компьютерная реальность доступна уже всем, но считается, и в этом уверены все, нереальной, хотя в ней о стены также можно разбить лоб. Где грань между реальным и нереальным?

2. Физическая реальность и сознание

Эйнштейн: «Вера в существование внешнего мира, независимого от воспринимающего субъекта, лежит в основе всего естествознания. Но так как чувственное восприятие дает информацию об этом внешнем мире, или о «физической реальности», опосредствованно, мы можем охватить последнюю только путем рассуждений. Из этого следует, что наши представления о физической реальности не могут быть окончательными. Мы всегда должны быть готовы изменить эти представления, т.е. изменить аксиоматическую базу физики, чтобы обосновать факты восприятия логически¹ наиболее совершенным образом» [1, с.136].

Тагор: «Наш мир относителен, его реальность зависит от нашего сознания» [1, с. 130]. «Наука доказала, что стол как твердое тело – это одна лишь

Copyright © 2007 А.К. Гуц.
Омский государственный университет.
E-mail: guts@univer.omsk.su

¹Следует еще задуматься о логике, используемой для обоснования фактов восприятия. Модальные логики, например, как известно, ведут к многовариантным интерпретациям, к множественным (двузначным) мирам (реальностям) в семантике Крипке. – А.Г.

видимость, и, следовательно, то, что человеческий разум воспринимает как стол, не существовало, если бы не было человеческого разума. В то же время следует признать и то, что элементарная физическая реальность стола представляет собой не что иное, как множество отдельных вращающихся центров электрических сил, и, следовательно, также принадлежит человеческому разуму» [1, с.132].

Эйнштейн: «Если есть реальность, не зависящая от человека, то должна быть истина, отвечающая этой реальности, и отрицание первой влечет за собой отрицание последней» [1, с.131].

Тагор: «Существует реальность бумаги, бесконечно далекая от реальности литературы. Для моли, поедающей бумагу, литература абсолютно не существует, но для разума человека литература как истина имеет большую ценность, чем сама бумага. Точно так же, если существует какая-нибудь истина, не находящаяся в рациональном или чувственном отношении к человеческому разуму, она будет оставаться ничем до тех пор, пока мы будем существами с разумом человека» [1, с.133]

3. Физическая реальность создается сознанием через деятельность

Бондаренко: «Мира, в котором была одна истина, уже нет ... Мир традиционной и интуиционистской математики – это разные математики... Мир всегда есть множество миров (или реальностей). И вместе с тем мир един [2, с.16]...

Мир, который мы понимаем, – это мир, который можем задать в нашем пространстве и времени опыта, через структуры и схемы деятельности познающего существа, задающие этой деятельностью свои пространства и время. Этот явленный нам мир, воссозданный на своих основаниях, одновременно допускается самим миром, является фактом мира и вписывается в связи Вселенной. Поэтому мы не можем вернуться в прежнее положение мира ..., в мир до того, как он был воспринят, наблюдался, познавался [2, с.16].

Научный взгляд на мир мы «покупаем» ценой расплаты непонимания сознания, которое научным же образом (объективно, доказательно, контролируемо и рационально) исследовать невозможно (оно уже было использовано в качестве основания теоретического анализа) [2, с.134].

Научные утверждения о вещах мира возможны потому, что мы можем их создать... Если я совершаю какую-то процедуру (она не всегда проговаривается, но имплицитно содержится в любом состоявшемся научном диспуте), связанную с приданием каким-то состояниям моего сознания и опыта свойств, присущих миру вне этих состояний. Это вынесение в действительность не любых состояний сознания и опыта, но лишь тех, о которых я могу говорить доказательно, опытно их проверять [2, с.135].

... Мы имеем дело с миром, который нам явлен через модели и схемы нашей деятельности, а затем, случившись как-то, чем извлекается опыт бытия, сами эти модели и схемы являются элементами мира. Модели фактов сами становятся фактом, *существуют*. Иначе говоря, действия мира в качестве явленных

предметов и вещей (а ведь вещи сами по себе не есть явления) существуют как реальность, а не наше представление [2, с.136-137]» .

4. Квантовый мир и физическая реальность

Внешний мир, который вместо классической механики Ньютона мы должны описывать с помощью квантовой механики, является объективной физической реальностью лишь в рамках эвереттовской интерпретации квантовой механики [3, с. 176], свойства которого могут быть неклассическими.

В рамках квантовой механики внешнему миру ставится в соответствие вектор в гильбертовом пространстве $|\psi\rangle$, который является суперпозицией некоторого набора базисных состояний

$$|\psi\rangle = \sum_{k=1}^{\infty} c_k |\psi_k\rangle. \quad (1)$$

Векторы $|\psi_k\rangle$ – это собственные векторы некоторой наблюдаемой физической величины A . Все они представляют собой классические миры, в которых физическая величина A имеет различные значения a_i .

В копенгагенской интерпретации квантовой механики реальным является только один из этого бесконечного набора потенциальных классических миров. Он дается нам при измерении величины A . Это и есть объективная физическая реальность, которую наблюдает наше сознание.

В эвереттовской интерпретации квантовой механики [4] одинаково реальными являются все классические миры, все они являются физическими реальностями, но каждый из них воспринимается своим наблюдателем, а тот, который наблюдаем мы, – это наша физическая реальность. Многообразие миров по Эверетту рождается в момент, когда Наблюдатель измеряет величину A . При этом происходит расщепление миров, задаваемое в виде формулы (1) с одновременным расщеплением сознания на сознания, наблюдающие свой классический мир.

Ниже мы переинтерпретируем Эверетта, заменяя Наблюдателя на материальную Природу, которая преобразуется при взаимодействии с субъектами. В ходе этих взаимодействий рождаются разнообразные реальные физические реальности. У Эверетта Наблюдатель и процедура измерения, ведущая к расщеплению миров, на наш взгляд, – откупная дань традиционному подходу к квантовой механике, на которую пошел Эверетт, предложивший и без того революционный подход к описанию Мира вне нас и одновременно нас содержащему.

5. Модель квантового рождения физической реальности

Предполагается, что существует нечто, обозначаемое как 0 и называемое *материальной Природой*, которая существует объективно, т.е. независимо от того, есть ли люди или их нет вообще (например, все погибли в мировом катаклизме).

Природе приписываем ψ -функцию состояния $|\psi^0\rangle$. События $A, B, \dots, C$, происходящие в Природе, оставляют свой «след», и это записываем следующим образом:

$$|\psi_{[A,B,\dots,C]}^0\rangle.$$

Субъект S , наделенный индивидуальным сознанием, которое обладает «идеей» (= «фантазией») A с собственной волновой функцией $|\phi_i\rangle$ в системе S , вступает во взаимодействие с Природой, начальное состояние которой $|\psi^0\rangle$. Результатом взаимодействия, проистекающим во *времени* и занимающим промежуток времени $[0, T]$, является попытка реализации идеи A . Иными словами, начальное состояние ($t = 0$)

$$|\psi^{S+0}\rangle = |\phi_i\rangle|\psi_{[...] }^0\rangle \quad (2)$$

переходит в новое состояние

$$|\tilde{\psi}^{S+0}\rangle = |\phi_i\rangle|\psi_{[...a_i]}^0\rangle,$$

где a_i характеризует состояние $|\phi_i\rangle$, т.е. отражает i -ю форму реализации «идеи» (фантазии) A . Под преобразованием мы понимаем решение $|\psi^{S+0}\rangle(t)$ уравнения Шредингера с начальным данным (2) при $t = 0$ и с

$$|\tilde{\psi}^{S+0}\rangle = |\psi^{S+0}\rangle(T).$$

Мы описали идеальный случай, когда индивидуальное сознание остается в собственном состоянии $|\phi_i\rangle$. В общем случае, если начальное состояние индивидуального сознания является несобственным, а общим состоянием $\sum_i a_i |\phi_i\rangle$, конечное состояние системы «субъект-Природа» будет иметь вид

$$|\tilde{\psi}^{S+0}\rangle = \sum_i a_i |\phi_i\rangle |\psi_{[...a_i]}^0\rangle. \quad (3)$$

Мы видим, что в каждом элементе суперпозиции $|\phi_i\rangle|\psi_{[...a_i]}^0\rangle$ состояние Природы есть особенное собственное состояние *потенциальной реальности*, и, более того, состояние Природы описывает Природу как определенно состоящую из набора потенциальных реальностей. Таким образом, Природа ветвится! В каждой потенциальной реальности субъект обнаружит различные наблюдаемые значения a_i идеи A , и его сознание разветвится, оказываясь в состояниях $|\phi_i\rangle$.

Рассмотрим еще более общую ситуацию, когда имеем несколько субъектов $S_1, S_2, \dots, S_n$, находящихся в состояниях $|\psi^{S_2}\rangle, |\psi^{S_1}\rangle, \dots, |\psi^{S_n}\rangle$. Пусть первый субъект S_1 начинает реализовывать идею-фантазию A^1 . Тогда начальное состояние

$$|\psi^{S_1+S_2+\dots+S_n+0}\rangle = |\psi^{S_1}\rangle|\psi^{S_2}\rangle\dots|\psi^{S_n}\rangle|\psi_{[...] }^0\rangle$$

преобразуется в конечное состояние

$$|\psi^{S_1+S_2+\dots+S_n+0}\rangle = \sum_i a_i^1 |\phi_i^{S_1}\rangle |\psi^{S_2}\rangle\dots|\psi^{S_n}\rangle |\psi_{[...a_i^1]}^0\rangle, \quad (4)$$

где $|\phi_i^{S_1}\rangle$ – собственные функции сознания субъекта S_1 .

И здесь мы видим ветвление сознания и Природы.

Если теперь во взаимодействие вступает второй субъект S_2 , реализующий свою идею-фантазию A^2 , то состояние (4) даст состояние

$$|\psi_2^{S_1+S_2+\dots+S_n+0}\rangle = \sum_i \sum_j a_i^1 a_j^2 |\phi_i^{S_1}\rangle |\phi_j^{S_2}\rangle |\psi^{S_3}\rangle \dots |\psi^{S_n}\rangle |\psi_{[\dots a_i^1 a_j^2]}^0\rangle. \quad (5)$$

Действия второго субъекта ведут к новому ветвлению Природы и индивидуального сознания и, естественно, того, что он наблюдает.

Потенциальная реальность $|\psi_{[\dots a_i]}^0\rangle$, навязываемая Природе субъектом S_1 , будет поддержана вторым субъектом в форме a_j^2 , если потенциальная реальность $|\psi_{[\dots a_i^1 a_j^2]}^0\rangle$ коррелирует с потенциальной реальностью $|\psi_{[\dots a_i^1]}^0\rangle$. Под корреляцией можно понимать, например, пропорциональность двух указанных потенциальных реальностей как векторов гильбертова пространства состояний Природы, т.е. если

$$\exists \lambda \in \mathbb{C} (|\psi_{[\dots a_i^1 a_j^2]}^0\rangle = \lambda |\psi_{[\dots a_i^1]}^0\rangle).$$

И вообще после действия r субъектов по реализации своих идей-фантазий ($r \leq n$) получим состояние

$$\begin{aligned} & |\psi_r^{S_1+S_2+\dots+S_n+0}\rangle = \\ & = \sum_{i,j,\dots,m,k} a_i^1 a_j^2 \dots a_m^{r-1} a_k^r |\phi_i^{S_1}\rangle |\phi_j^{S_2}\rangle \dots |\phi_m^{S_{r-1}}\rangle |\phi_k^{S_r}\rangle |\psi^{S_{r+1}}\rangle \dots |\psi^{S_n}\rangle |\psi_{[\dots a_i^1 a_j^2 \dots a_m^{r-1} a_k^r]}^0\rangle. \end{aligned} \quad (6)$$

Если число субъектов достаточно велико, то реализуемые ими идеи в форме потенциальных реальностей приведут к рождению² актуальной **физической реальности**

$$|\psi_{[\dots a_i^1 a_j^2 \dots a_m^{r-1} a_k^r]}^0\rangle, \quad (7)$$

если

$$\exists \lambda \in \mathbb{C} (|\psi_{[\dots a_i^1 a_j^2 \dots a_m^{r-1} a_k^r]}^0\rangle = \lambda |\psi_{[\dots a_i^1 a_j^2 \dots a_m^{r-1}]}^0\rangle = \dots = \lambda |\psi_{[\dots a_i^1 a_j^2]}^0\rangle = \lambda |\psi_{[\dots a_i^1]}^0\rangle).$$

Очевидно, таких физических реальностей может быть много – это всё эвереттовские параллельные вселенные [4], но одновременно существует множество потенциальных реальностей, так и не ставших актуальной физической реальностью. Это происходит потому, что действия индивидуальных сознаний не оказались скоррелированными. Иначе говоря, не все идеи-фантазии реализуются; многие из них остаются снами-миражами.

6. Осознание

Осознание – это отдавание себе отчёта, что мы есть в данном месте и в данное время, т.е. *присутствуем*. Как это описать математически? Традиционно считается, что мир в нас отражается, отражается в нашем мозгу. Но почему мозг знает, что он отражает реальность?

²Нескоррелированные потенциальные реальности остаются с нами в форме грёз субъектов, их породивших.

Физическая реальность (7) есть, в частности нечто, созданное по «матрице» M , состоящей из набора идей-фантазий $A, B, \dots$. Это первый этап на пути к осознанию – рождение физической реальности. На втором этапе рожденная реальность отражается, т.е. воспринимается мозгом. В мозгу появляется отпечаток M' . На третьем этапе матрица M сравнивается с отпечатком M' . При совпадении (почти совпадении) мозг «видит себя в реальности (7)». Это и есть отдавание себе отчёта о присутствии (местонахождении), т.е. акт осознания: «..сущее, которое мы сами всегда суть и которое среди прочего обладает бытийной возможностью спрашивания, мы терминологически схватываем как *присутствие*» [5, с.22].

ЛИТЕРАТУРА

1. Эйнштейн А. Собрание научных трудов. Т.4. М.: Наука, 1967.
2. Бондаренко И.А. Феноменология бытийственного сознания. М., 2000.
3. Менский М.Б. Человек и квантовый мир. Фрязино: «Век2», 2005.
4. Everett H. "Relative state" formulation of quantum mechanics // Reviews of Modern Physics. 1957. V.29. P.454-462.
5. Хайдеггер М. Бытие и время. М.: Фолио, 2003.

НЕЛИНЕЙНЫЕ СЕМАНТИЧЕСКИЕ АСПЕКТЫ КВАНТОМЕХАНИЧЕСКОЙ КОНЦЕПЦИИ «СООТНЕСЕННЫХ СОСТОЯНИЙ» Х. ЭВЕРЕТТА И ПЕРСПЕКТИВЫ РАЗВИТИЯ ЭВЕРЕТТИКИ

Ю.А. Лебедев

Рассмотрено смысловое содержание основных понятий эвереттики – памяти волновой функции и соотнесенного состояния. Показано, что эвереттовская память может трактоваться как модуляция несущей волны Де Бройля волновой функции Ψ . Указана связь эвереттической и фрактальной («голографической») интерпретаций квантовой механики. Установлено, что понятие эвереттических склеек как проявления эффектов информационной нелинейности при взаимодействии квантовых объектов латентно присутствует в тексте основополагающей статьи Х. Эверетта. Обсуждаются физические аспекты смысла информации. Дана оценка количества склеек в зависимости от количества собственных состояний Наблюдателя и Объекта. Предложены схемы экспериментов по доказательству наличия параметров памяти у Ψ -функции и оценке вероятностей конкретных склеек в соотнесенных состояниях. Обозначены пути развития эвереттики при переходе к описанию сильно взаимодействующих систем и показано, что перспективным полем экспериментальных исследований проявления многомирия является химия.

Выдвинутая в 1957 году Х. Эвереттом концепция соотнесенных состояний [1], приведшая к возникновению эвереттики как новой междисциплинарной области духовной деятельности [2], основывалась на анализе процесса измерения в квантовой механике¹.

Во введении Эверетт отмечал, что с целью построения непротиворечивой квантовой механики изолированных систем «необходимо сформулировать абстрактные модели наблюдателей, которые сами по себе, в пределах теории,

Copyright © 2007 **Ю.А. Лебедев**.
МГТУ им. Н.Э. Баумана, г. Москва.
E-mail: petroff@mail.ru

¹Однако к настоящему времени она вышла далеко за пределы этой сравнительно узкой теоретической области. Один из ведущих современных физиков А.Д. Линде отмечает: «Я полагаю, что теория Эверетта – одна из лучших интеллектуально честных попыток интерпретации квантовой механики» [19].

могут трактоваться как физические системы, рассмотреть изолированные системы, содержащие таких модельных наблюдателей во взаимодействии с другими подсистемами, выявить изменения, которые происходят в наблюдателе вследствие взаимодействия с ближайшими подсистемами, и интерпретировать изменения на знакомом языке опыта» [3].

Одним из важнейших, революционизирующих физику понятий, вводимых Эвереттом, является *память как параметр волновой функции*. Вот как определяет это понятие сам Эверетт: «Когда мы имеем дело с системой, в которой наблюдатель представлен квантовомеханически, мы приписываем ему функцию состояния ψ^0 . Когда состояние ψ^0 описывает наблюдателя, память которого содержит представления событий $A, B, \dots, C$, мы обозначаем этот факт, вводя последовательность памяти как дополнение в скобках, и записываем:

$$\psi^0_{[A,B,\dots,C]}. \quad (1)$$

Поэтому символы $A, B, \dots, C$, которые мы временно принимаем, символизируют конфигурацию памяти, находящуюся в соответствии с прошлым опытом наблюдателя. Эти конфигурации могут рассматриваться как отверстия в бумажной ленте, след в магнитной катушке, конфигурации переключающих реле, и даже как конфигурации ячеек мозга. Мы только требуем, чтобы они были способны к интерпретации: «наблюдатель испытал последовательность событий». (Мы иногда пишем точки в последовательности памяти, указывая возможное присутствие предыдущих воспоминаний, которые являются несоответствующими рассматриваемому случаю.)» [3].

Введение параметра памяти в основополагающую характеристику состояния любой физической системы с неизбежностью приводит к тому, что процессы взаимодействия систем становятся принципиально нелинейными и необратимыми [4]. Иными словами, наличие памяти у волновой функции объясняет возникновение «стрелы времени» в мультиверсе, ибо последняя является однозначным следствием возрастания объема памяти при *любых взаимодействиях*.

Более того, появление памяти среди параметров самой фундаментальной физической функции революционным образом меняет мировоззренческую картину – физика выходит из состояния негуманитарной абстрактности, где в «объективной пространственно-временной среде» взаимодействуют безликие массы и заряды, и переходит в категорию конкретно-исторических дисциплин. *У каждого электрона, фотона, нейтрино появляется собственная история*, и потому картина мироздания необозримо усложняется и «одухотворяется». Вполне вероятно, что при этом должен произойти качественный скачок – для сохранения целостности и принципиальной обозримости мировоззренческой картины должны появиться новые фундаментальные понятия на качественно новом уровне.

С учетом этого понятно, почему из объяснений Эверетта неясно, какие физические сущности являются носителями памяти на уровне квантовых объектов и как формально-математически фиксируется этот параметр в уравнении для ψ , поскольку объяснения относятся лишь к описанию *проявлений* этого квантового свойства на макроуровне.

Принципиально важным является и то, что параметры памяти в эвереттовской трактовке явно *не зависят от времени*. Это значит, что параметры относятся к «мгновенному состоянию», а их возможное изменение связано с процессами последующих взаимодействий. Отсюда же следует и то, что память «абсолютна» – ничто и никогда не «забывается». «Рукописи не горят» – этот булгаковский образ можно считать художественным предвидением квантовомеханических идей Х. Эверетта. Другое дело – *вскрытие и использование* содержащейся в памяти информации. Очевидно, что вопрос о физическом смысле существования эвереттовской памяти должен быть отделен от вопроса о механизме ее «считывания» (это явно «последующий процесс») и дальнейшего проявления в процессах последующих взаимодействий.

Дискретность и независимость параметров памяти от «ньютоновского времени» свидетельствуют в пользу реляционной природы физического времени, хотя явно не исключают и его субстанциональную трактовку. Ясно, что данная тема требует отдельной подробной проработки.

Для прояснения физического смысла «временно принятых» Эвереттом символов $\dots A, B, \dots, C$ необходимо вспомнить, что ψ описывается волновым уравнением. Но только для гармонической волны для ее описания нужны всего три параметра – частота, амплитуда и фаза (или некоторый другой набор параметров, однозначно связанный с этими тремя). Более сложные формы колебаний требуют введения дополнительных параметров, индивидуализирующих данный колебательный процесс.

В практике исследований электромагнитных волн, например, эти параметры выявляются формализмом «динамики диссипативных систем (фрактальность, дробные операторы, негауссова статистика, распределения с тяжелыми хвостами, режим детерминированного хаоса, существование странных аттракторов в фазовом пространстве отраженных сигналов, их топология и т.д.)» [5].

Физический смысл порождающих волновую функцию процессов является предметом дискуссии и проясняется в аксиоматических интерпретациях. Последней по времени и наиболее общей по охвату поля физических идей является интерпретация струнных теорий [6]. В этих теориях точечным физическим объектам приписывается «линейная протяженность», частицы представляются проявлениями «струн» и введенные квантовой механикой «волновые» или «колебательные» их свойства получают образно-логическое обоснование. Однако математические свойства этих объектов пока не получили общепринятого описания, и в данной работе в дальнейшем в качестве «нулевого приближения» принимается интерпретация Де Бройля, согласно которой всякий квантовый объект («частица») представляет собой систему, состоящую из «собственно частицы» и неразрывно с ней связанной стационарной «волны Де Бройля» [7]. Такая система является строго изолированной, поскольку стационарность волны Де Бройля порождается именно тем, что она по определению является «вселенской сущностью». Эта интерпретация, как представляется, весьма точно соответствует и духу концепции Эверетта, в которой изолированность системы является основополагающим условием.

Концепция Де Бройля в настоящее время считается несколько архаической

и «перегруженной наглядностью», что затрудняет ее логический анализ из-за возникающих «классических ассоциаций», не имеющих смысла в квантовом мире.

Оставляя в стороне вопрос о плодотворности различных аксиоматических систем и об источниках предпочтения тех или иных моделей квантовой реальности, надеюсь, что искренние и искушенные в современной физике «предпочтители вектора состояний» легко переведут мои высказывания о волнах Де Бройля и волновых функциях на этот, более близкий им язык.

В физических экспериментах до сих пор были обнаружены только явления дифракции и интерференции волн Де Бройля, в которых проявляется их частотная характеристика.

Волна Де Бройля для данной квантовой системы является простейшим выражением ее волновой функции и содержит информацию только об импульсе этой системы. Ни структурной, ни историко-событийной информации о системе волна Де Бройля не имеет.

В то же время простейший расчет показывает, что частоты Де Бройлевских волн для «рядовых» макрообъектов настолько велики, что кажутся явно «избыточными» для фиксации столь ограниченной информации. Так, для тела массой порядка килограмма частота волны Де Бройля составляет около 10^{47} Гц, для типичной «средней» звезды типа Солнца – 10^{77} Гц, а для «типичной» галактики типа туманности Андромеды – 10^{88} Гц. Для нашего универса в целом эта частота может быть оценена как $10^{99} - 10^{100}$ Гц! (Любопытно отметить, что верхний предел – гугол – получается при оценке общей массы «видимой части» материи универса 10^{52} кГ [8] и учете того, что около 90% его массы – это темная энергия и темное вещество).

В связи с этим можно предположить, что сама по себе волна Де Бройля является только «физическим инструментом» для записи (фиксирования) всего богатства информации о данной квантовой системе. Такой инструмент является очень удобным «носителем памяти», а эвереттовские параметры памяти являются параметрами модуляции волновой функции, возникающей на основе волны Де Бройля.

Передача информации всегда сопровождается передачей энергии, что, вследствие однозначной связи массы и энергии, неизбежно должно приводить к формированию новых волн Де Бройля, накладывающихся на основную «волну-носитель»².

²С поразительной точностью и в блестящей художественной форме эту идею ещё в 1975 г. представил С. Лем в своем рассказе «Профессор Доньда». Вот краткая суть идеи Лема в переводе В. Язневича: «Информация имеет массу. Любая. Безразлично какая. Смысл не имеет ни малейшего значения. Атомы также одинаковы, что в камне, что в моей голове. Информация весит, но ее масса неслыханно мала. Знания всей энциклопедии весят около миллиграмма... Вес информации не был бы чем-нибудь необычным, если бы не последствия... Существует критичная масса информации, так же, как существует критичная масса урана. Мы приближаемся к ней. Не конкретно мы здесь, а вся Земля. Приближается к ней каждая цивилизация, создающая компьютеры. Развитие кибернетики – это западня, расставленная Природой для Разума... Учитывается сосредоточенность, а не количество информации. Так же, как и с ураном. Аналогия неслучайна! Рассредоточенный в земле и скалах уран не опасен. Условием

В этом случае индексы эвереттовской памяти $\dots A, B, \dots, C$ есть не что иное, как фиксация наличия у результирующей волны Де Бройля специфических колебательных мод $\dots A, B, \dots, C$.

Специфичность этих мод проявляется в том, что при их сложении результирующая волна сохраняет свою «несущую частоту», обеспечивающую воспроизводимость общей картины явлений интерференции и дифракции, т.е. при этом сохраняется «общий тон» волны, однако форма этой волны перестает быть гармонической, и в течение каждого периода воспроизводится колебание определенной сложной формы, проявления которого можно уподобить проявлению такого качества акустической волны, как тембр.

До последнего времени такая трактовка наталкивалась на сложность сохранения неизменности формы результирующей волны вследствие дисперсии скоростей отдельных мод. Не пытаясь детализировать общую картину распространения волн Де Бройля, можно отметить, что недавнее открытие темной энергии и темного вещества позволяет надеяться, что в их свойствах будет обнаружено некое качество, которое позволяет им играть роль своеобразных «дирижеров», синхронизирующих распространение отдельных мод волн Де Бройля. Эти надежды основаны на том, что дисперсия – это проявление свойств среды, и, пока свойства «темной среды» совершенно неизвестны, дозволительно считать, что они соответствуют сделанным предположениям.

Неизбежно возникающие при такой трактовке музыкальные аналогии, вероятно, не являются случайными и свидетельствуют о том, что античные философские традиции, и пифагорейство в первую очередь, далеко не исчерпали своих креативных потенциалов, а их прозорливость ещё не осознана в должной мере.

Предложенный механизм формирования памяти у квантовых объектов позволяет обосновать и такое качество окружающего нас мира, которое Д. Бом обозначил как «голографичность» [9] – наличие у любого обладающего массой фрагмента универса информации обо всем универсе в целом.

В другой терминологической базе это качество может трактоваться как фрактальность, что позволяет использовать при его анализе уже имеющийся математический аппарат, например [10]. Разработка применения математической теории фракталов применительно к рассматриваемым вопросам является актуальной задачей.

В рамках предложенного механизма возникновения эвереттовской памяти такое свойство легко объясняется тем, что при разделении любого массивного объекта каждый его фрагмент «уносит» практически всю информацию, имевшуюся в исходном объекте (теле). Такое объяснение основано на том, что «Де Бройлевская информация» не локализована в исходном теле и «записана на общей массе» с колоссальным избытком информационной емкости ее носите-

взрыва является отделение от примесей и концентрация. Так и здесь. Информация в книгах или в головах может быть значительной, но остается пассивной. Так же, как и рассеянные частицы урана. Нужно ее объединить! ... Выше критической точки начнется цепная реакция. *Obiit animus, natus est atomus!* (Исчезла душа, родился атом! <лат.>) – В.Я.) Информация исчезает, потому что превращается в материю...» [20].

ля – массы.

Разумеется, в связи с тем, что масса каждого фрагмента все-таки меньше массы исходного объекта, частота Де Бройлевской «волны-носителя» у него будет меньше и какие-то тонкие элементы структуры результирующей волны (а следовательно, и какая-то отраженная в них информация) неизбежно утрачиваются. Однако, учитывая колоссальную частотную избыточность для тел с макромассами, такие утраты не должны серьезно обеднять «базу данных» о породившем их объекте у любого макрофрагмента.

С математической точки зрения это свойство физических тел можно считать отражением фрактальной структуры материального мира, т.е. информационного подобия всякой его части целому.

Наличие «тембровых мод» у волн Де Бройля может быть проверено экспериментально при соответствующей организации экспериментов по дифракции квантовых объектов. Для этого можно предложить, например, такую принципиальную схему эксперимента (рис.1).

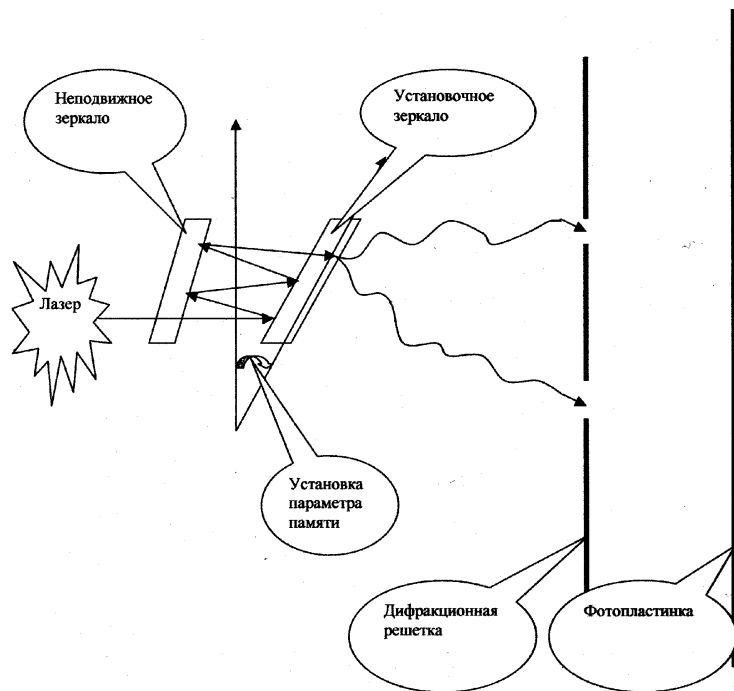


Рис. 1. Схема опыта по регистрации «тембровых мод» волны Де Бройля. Путем поворота установочного зеркала можно задать число внутренних отражений фотона N_E в системе зеркал, которое и будет характеристикой его эвереттовской памяти. Исследуется вид интерференционной картины на фотопластинке в зависимости от значения N_E

Монохроматическое излучение лазера (поток квантов с гармонической формой волны Де Бройля) проходит через отверстие в неподвижном зеркале и падает на установочное зеркало, угол наклона которого можно задавать так, что число внутренних отражений каждого фотона перед тем, как он покинет систему зеркал через отверстие в установочном зеркале, будет являться определенным числом N_E . Это число – по сути экспериментальная «метка» – должно

являться параметром памяти каждого фотона и определяет форму его Де Бройлевской волны, формируя соответствующую тембровую моду. Далее падающий на дифракционную решетку поток фотонов дает соответствующую дифракционную картину, фиксируемую на детекторе (например, фотопластинке).

В соответствии со сделанными предположениями на «стандартную дифракционную картину» для данной длины волны лазера должна накладываться новая, обусловленная той тембровой модой (в зависимости от N_E), которая интерферирует на данной решетке.

Они могут быть разделены при достаточно большой разности частот исходного излучения и тембровой моды.

Сложность постановки такого эксперимента связана с тем, что тембровые моды при достаточно малых N_E могут оказаться волнами из совершенно другого диапазона частот – вплоть до сверхдлинных радиоволн. И для их фокусировки и фиксации необходимо использовать совершенно другие «зеркала» и детекторы. Но это же обстоятельство может быть использовано и для убедительной демонстрации эффекта именно эвереттовской памяти – возникновение когерентного радиоизлучения, синхронного с оптическим.

На этом принципе может быть построен и частотный преобразователь когерентного излучения. Однако нужно учитывать, что чем меньше N_E , тем меньше интенсивность его тембровой моды.

Повышение интенсивности тембровой моды с ростом N_E должно сопровождаться соответствующим уменьшением интенсивности исходного излучения. И в пределе вся энергия исходного излучения может быть трансформирована в энергию тембровой моды, т.е. тембровая мода станет тождественной с исходным излучением. При этом емкость эвереттовской памяти исходного излучения будет исчерпана.

Проведенный анализ предлагаемого эксперимента показывает, что емкость эвереттовской памяти любого объекта действительно ограничена, как и утверждал сам Эверетт: «В любом известном запоминающем устройстве вследствие ограниченной емкости его памяти ветвление не продолжается бесконечно, но должно остановиться в некоторой точке» [3].

Аналогично можно модифицировать и классические опыты У.Г. и У.Л. Брэггов, и Девиссона, и Джермера для рентгеновских лучей и пучков электронов соответственно.

Важно заметить, что Эверетт особо подчеркивает – наблюдатель в его понимании необязательно разумен (качественные особенности разумного наблюдателя будут рассмотрены ниже), наблюдателем может являться и просто *взаимодействующая* с выделенной подсистемой другая физическая система (в рассмотренном случае наблюдателем является фотопластинка), т.е. эффекты эвереттовской памяти могут быть обнаружены *при любом взаимодействии*. Очевидно, что характер взаимодействия подсистем в зависимости от энергии взаимодействия может быть различным. Однако в дальнейшем анализе Эверетт сосредоточился только на таких взаимодействиях, которые происходят при «хороших наблюдениях».

После введения параметра памяти в волновую функцию наблюдателя Эве-

ретт так характеризует рассматриваемые им взаимодействия: «Мы начнем с определения того, что представляет собой «хорошее» наблюдение. Хорошее наблюдение величины A , имеющей собственную функцию ϕ_i в системе S , наблюдателем, начальное состояние которого ψ^0 , состоит во взаимодействии, которое, в указанном промежутке времени, преобразует каждое (полное) состояние

$$\psi^{S+0} = \varphi_i \psi_{[...] }^0 \quad (2)$$

в новое состояние

$$\psi^{S+0'} = \varphi_i \psi_{[...a_i]}^0, \quad (3)$$

где a_i характеризует состояние ϕ_i . (Нужно понимать, что $\psi_{[...a_i]}^0$ – различны для каждого i состояния. Для более точной нотации следовало бы написать $\psi_{i[...a_i]}^0$, но никакой путаницы не может возникнуть, если мы просто позволим ψ_i^0 быть внесенной в указатель только индексом символа конфигурации памяти. Символ a_i отражает, например, регистрацию собственного значения. Штриховой индекс в данном случае относится не к 0, а к символу в целом.).

Таким образом, во-первых, мы требуем, чтобы состояние системы, *если это собственное состояние*, оставалось неизменным, и, во-вторых, чтобы состояние наблюдателя изменилось так, чтобы описать наблюдателя, который «знает», из какой собственной функции оно получено; то есть в памяти наблюдателя зарегистрирована некоторая характеристика типа собственного значения, которая присуща ϕ_i .» [3].

Как видно из этой цитаты, из всех видов взаимодействия Эверетт обратил особое внимание на случай, при котором повторное измерение даст тот же самый результат. Это случай, когда наблюдаемая система находится в стационарном состоянии и динамика собственных значений измеряемой величины либо нулевая, либо циклическая с достаточно большой частотой. При этом «хорошее измерение» возможно только при таком физическом энергообмене наблюдателя с рассматриваемой системой, когда уровень этого энергообмена ниже некоторого «порогового», изменяющего значение измеряемого параметра на величину, превышающую предел чувствительности рецептора наблюдателя³. Энергообмен при таком взаимодействии обеспечивает *только* изменение параметров памяти волновых функций наблюдаемой системы и наблюдателя с созданием «отметки» о произведенном «зондировании».

Такие взаимодействия можно назвать энергоинформационными. Их энергия полностью переходит только в информационные моды Де Бройлевских волн взаимодействующих систем.

Например, таким условиям с достаточной точностью отвечает измерение скорости автомобиля с помощью радара.

Однако выполнение условий «хорошего измерения» какого-то параметра не означает, что другие параметры измеряемой системы останутся неизменными и даже того, что в результате изменения этих других параметров не изменится и измеряемый. В рассмотренном примере может оказаться, что в автомобиле

³Пользуясь уже введенными музыкальными аналогиями, можно сказать, что это взаимодействие изменяет тембр волны Де Бройля в басовой его части за пределами слышимости.

установлен «антирадар» и водитель изменит скорость машины под воздействием осознания смысла полученной от него информации.

Таким образом, к определению Эверетта следует добавить, что рассматриваемый им случай относится к системам, в которых отсутствует семантический анализ информации.

Отсюда следует, что для расширения полноты квантового описания реальности следует принять во внимание разницу понятий «информации» и её «смысла».

Для «расшифровки» информации, «овеществления» ее семантических свойств наблюдатель должен приложить *волевое усилие, основанное на желании* придать смысл информации. Это усилие «объективно фиксируется» тем, что смысл оказывается нелинейной функцией исходной информации. И проявленные таким образом её семантические свойства изменяют состояние только самого Наблюдателя, но не Объекта наблюдения.

В общем случае *состояние системы может зависеть от того, присутствует ли в ней хотя бы один элемент, способный к семантическому анализу информации*⁴.

Присутствие такого элемента – разумного наблюдателя – резко меняет физическую ситуацию. В известной статье М.Б. Менского даже утверждается, «что разделение квантового состояния на компоненты, соответствующие альтернативным результатам измерения, не только связывается с сознанием наблюдателя, но полностью отождествляется с ним» [11, с.413]. К этому можно добавить, что и сам процесс наблюдения в таком случае приобретает новое качество – поиск смысла происходящего.

Смысл *придается* информации ее респондентом и осознается *только им*. Когда 18 июля 1936 года (по свидетельству И.Г. Эренбурга) в радиоэфире прозвучала фраза «Над всей Испанией безоблачное небо» и начался мятеж военных, переросший в одну из самых знаменитых гражданских войн XX века, только *те, кто хотели этого*, поняли, что эти слова *имеют смысл* сигнала к выступлению. Осознание *смысла* информации, т.е. наполнение ее семантическим содержанием – это и есть принципиальное и решающее отличие Сознания

⁴Отсюда следует, что к даваемому Эвереттом определению «Большинство «наблюдателей», обсуждаемых в квантовой механике, относится к фотоэлементам, фотографическим пластинам, и тому подобным устройствам, чья механистическая сущность едва ли может быть оспорена» [3] нужно подходить с пониманием обстоятельств публикации. Очевидно, что Эверетт по «тактическим» диссертационным соображениям и под прямым «давлением» Де Витта, опасавшегося негативной реакции консервативного большинства своих коллег-физиков, в данном случае «осторожничает» и намеренно упрощает ситуацию. То, что это так, следует из его замечания, адресованного вдумчивому читателю: «В последующем каждый читатель, если у него нет желания рассматривать наблюдателей в более привычном смысле, может ограничиться этим классом проблем на том же самом механистическом уровне анализа» [3]. Намек более чем прозрачен: у такого – «вдумчивого» – читателя, конечно, есть желание рассмотреть «нормального» Наблюдателя, а потому он не должен «ограничиться этим классом проблем». Ту же направленность имеет и другое замечание: «В качестве модели для наблюдателей мы, *если пожелаем* (курсив Ю.Л.), можем рассматривать автоматически функционирующие машины, обладающие чувствительным датчиком, связанным с регистрирующим устройством и способные к регистрации прошлых сенсорных данных и конфигураций машины» [3].

от Разума, также обладающего способностью к волевому ветвлению мультиверса. Смысл – это «второй этаж» информации, его нелинейная характеристика, возникающая на основании осознания Наблюдателем своего «Я».

Покажем, что учет смысла информации позволяет выявить принципиально новый класс квантовых состояний.

Это – «семантически противоречивые» состояния, а потому, вероятно, они и не привлекли внимания самого Эверетта. Позже они были введены в эвереттику из соображений симметризации описания систем с эвереттическими ветвлениями и получили название «склеек» [12].

В общем случае склейки – это процессы взаимодействия ветвей мультиверса и потому должны проявляться в виде различных феноменов и ноуменов. Можно полагать, что склейки – это совокупность качественно весьма разнородных эффектов. Их выявление и классификация должны стать одной из важных эвереттических задач.

В связи с этим нужно уточнить, что рассматриваемые в данной работе склейки – это проявление эффектов информационной нелинейности при взаимодействии квантовых объектов. Формальное квантовомеханическое описание этого явления дано в [13]. Развивая полученные в [13] результаты, рассмотрим систему S , состоящую из наблюдателя $\xi_i^{S_1}$ (подсистема S_1) и наблюдаемой $\eta_j^{S_2}$ подсистемы (подсистема S_2). Ограничим состояния и S_1 и S_2 только двумя возможностями: $i = 1, 2; j = 1, 2$. Тогда, согласно принципу суперпозиции (в обозначениях, принятых Эвереттом в [3]):

$$\psi^S = \sum_{i,j} \xi_i^{S_1} \eta_j^{S_2}, \quad (4)$$

ψ^S , выраженная соотнесенными в эвереттовском смысле состояниями⁵ в S_2 для $\xi_1^{S_1}$ и $\xi_2^{S_1}$, будет соответственно равна:

$$\psi_{\xi_1}^S(S_2; rel \xi_1, S_1) = N_1(a_{11}\xi_1\eta_1 + a_{12}\xi_1\eta_2) \quad (5)$$

и

$$\psi_{\xi_2}^S(S_2; rel \xi_2, S_1)^S = N_2(a_{21}\xi_2\eta_1 + a_{22}\xi_2\eta_2), \quad (6)$$

где N_1 и N_2 – нормировочные множители.

Соответственно ψ^S , выраженная соотнесенными в эвереттовском смысле состояниями в S_1 для $\eta_1^{S_2}$ и $\eta_2^{S_2}$, будет равна:

$$\psi_{\eta_1}^S(S_1; rel \eta_1, S_2) = N_3(a_{11}\eta_1\xi_1 + a_{21}\eta_1\xi_2) \quad (7)$$

⁵Вот что сообщил по этому поводу Л.В. Ильичев: «Собственно, сам термин «соотнесённое состояние» отражает свойства зацепленного состояния и есть просто приведение подобных членов по отношению к выбранному состоянию η_i и вынос последнего за скобки».[21] А общие свойства запутанных состояний хорошо охарактеризовал А. Аспек, автор знаменитых экспериментов по проверке неравенств Белла, на примере пары запутанных фотонов: «Запутывание (entanglement) безусловно является свойством, выходящим за рамки любого пространственно-временного описания по Эйнштейну: пары запутанных фотонов должны рассматриваться как единый глобальный объект, который нельзя рассматривать в виде составленного из разделенных в пространстве и времени отдельных объектов с хорошо определенными свойствами» [22].

и

$$\psi_{\eta_2}^S(S_1; rel\eta_2, S_2) = N_4(a_{12}\eta_1\xi_1 + a_{22}\eta_2\xi_2), \quad (8)$$

где N_3 и N_4 – нормировочные множители.

Иными словами, наблюдатель, находясь в собственных состояниях ξ_1 и ξ_2 , входит в состав разных суперпозиций состояний с наблюдаемой подсистемой S_2 – (5) и (6) соответственно. Аналогично объект наблюдения – подсистема S_2 – в собственных состояниях η_1 и η_2 входит в состав разных – (7) и (8) – суперпозиций с наблюдателем.

С эвереттической точки зрения все слагаемые этих суперпозиций являются равно реальными и каждое – это одна из ветвей мультиверса или граней кристалла Менского [14].

Сделав такое сильное утверждение, Эверетт сосредоточился на анализе вопроса о его согласованности с предсказаниями «классической» квантовой механики. Здесь же предпринимается попытка рассмотреть семантические аспекты принципа суперпозиции и соотношенного состояния.

Рассмотрим разобранную ситуацию на конкретном примере классического двухщелевого эксперимента. В данном случае совершенно неважно, какой именно агент проявляет свои волновые свойства, – электрон, фотон или что-то иное.

Схема такого эксперимента приведена на рис. 2.

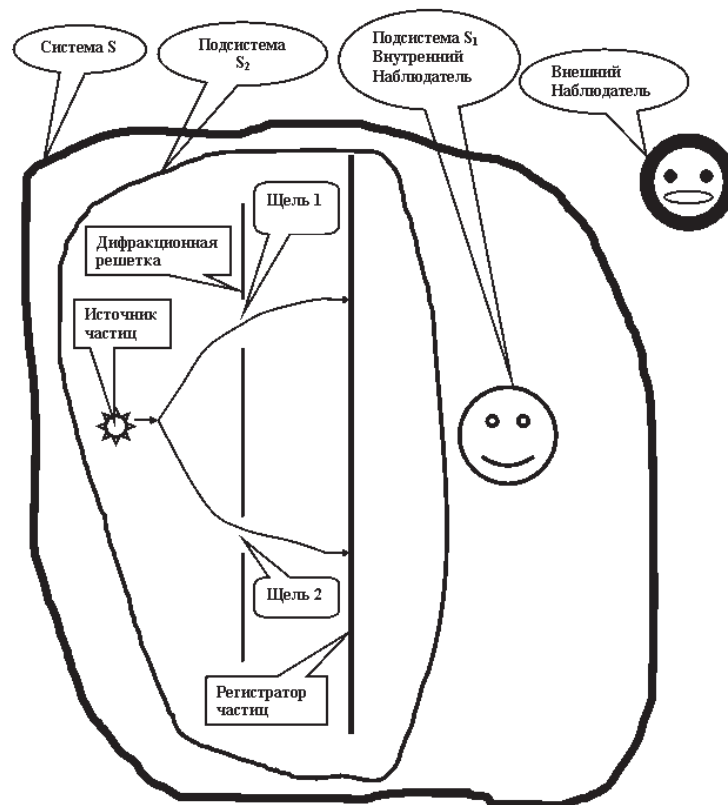


Рис. 2. Анализ двухщелевого эксперимента Внешним Наблюдателем

Система S включает в себя подсистему наблюдателя S_1 и подсистему на-

блюдаемого объекта (эксперимента) S_2 . Подсистема S_2 состоит из источника частиц, дифракционной решетки и детектора (скажем, цифровой матрицы).

Кроме того, в эксперименте должен принимать участие и Внешний Наблюдатель (его можно рассматривать как элемент метасистемы), который и будет фиксировать различные соотнесенные состояния S_1 и S_2 .

Необходимость наличия такого *особого* Внешнего Наблюдателя логически неизбежна и следует из текста статьи самого Эверетта – такими наблюдателями являются и её автор, и читатели, которые рассматривают эвереттовскую «изолированную систему» извне.

Но такая возможность разрушает классическую изолированность системы S и явно противоречит намерению Эверетта построить квантовую механику истинно изолированных систем.

Содержательное и плодотворное разрешение этого парадокса может быть достигнуто в рамках самой эвереттики, о чем будет сказано ниже.

Каким же событиям в ветвях мультиверса соответствуют слагаемые суперпозиций (5) - (8)? Прежде всего отметим, что в данном случае индексы у состояний ξ_1 и ξ_1 могут быть сопоставлены индексам эвереттовской записи памяти наблюдателя [... a_i] и имеют физический смысл фиксации наблюдателем прохода частицы через щели 1 или 2 в дифракционной решетке соответственно.

При такой трактовке члены суперпозиций (5) и (6) имеют следующий смысл:

1. $a_{11}\xi_1\eta_1$ – наблюдатель зафиксировал прохождение частицы через щель 1, и в подсистеме S_2 частица прошла через щель 1. Это событие соответствует одной из возможных «классических альтернатив» копенгагенской трактовки квантовой механики и, в соответствии с ней, «реализуется» с вероятностью, пропорциональной $|a_{11}|^2$.

2. $a_{12}\xi_1\eta_2$ – наблюдатель S_1 зафиксировал прохождение частицы через щель 1, а в подсистеме S_2 частица прошла через щель 2!⁶. Это не соответствует ни одной из возможных «классических альтернатив» копенгагенской трактовки квантовой механики и, в соответствии с этим, «реализуется» с вероятностью равной нулю, и, следовательно, $a_{12} = 0$. В эвереттике же именно такое событие и рассматривается как склейка. Данное слагаемое является явно противоречащим картине «хорошего наблюдения» по Эверетту, принятой в позднейших работах, в частности в широко известных статьях М.Б. Менского, однако его появление однозначно следует из принятого формализма и потому требует физического истолкования.

Физический смысл такого типа склеек тесно связан с наличием Внешнего наблюдателя, который *знает* о реальности обоих состояний – ξ_1 и η_2 – и потому может оценить возникновение состояния $\xi_1\eta_2$ (т.е. данный исход наблюдения) как «невероятность», «алогичность», «семантическую абсурдность».

Для наблюдателя же S_1 любой исход будет являться просто «экспериментальным фактом», более или менее редким. Если частота какого-то исхода настолько мала, что сопровождающее его явление кажется «случайным», он и

⁶Такой исход соответствует, например, ситуации «неожиданной потери»: вы «только что» положили на стол винтик от дужки очков, а «едва моргнули» – он исчез!

относит его к этой категории, не пытаясь установить ускользающую закономерность. Именно поэтому мы столь спокойно относимся к многочисленным фактам «бытовых склеек» – «пропавших очков», «потерянных вещей и документов» и т.п., а в научной практике вводим понятие «экспериментальной ошибки». С эвереттической точки зрения нет «ошибочных измерений», а есть редкие ветвления и склейки, которые относятся к типу «чуда».

При этом, если наблюдатель S_1 обладает способностью к семантическому анализу, он, так же, как и Внешний наблюдатель, может *предположить* «невероятность» произошедшего события и для проверки его «объективности» (в случае, когда это возможно) попытается провести повторный эксперимент, воссоздающий «те же» начальные условия.

Однако такая попытка заведомо обречена на неудачу из-за стрелы времени. Тех же условий достичь не удастся – память наблюдателя уже содержит информацию о произошедшем событии.

Результатом повторного эксперимента будет альтернатива: либо наблюдатель снова получит «чудесный результат» и потому сочтет, что в первом случае его «подвела память» и «на самом деле» все его прошлое соответствует проводимому эксперименту, либо повторный эксперимент подтвердит «ошибочность» полученного «чудесного результата».

С эвереттической точки зрения такая последовательность событий является ветвлением из точки склейки.

Таким образом, сугубо формально *склейка типа «чудо» является элементом суперпозиции состояния системы «Наблюдатель-Объект» (4), описываемым таким произведением Ψ -функций образующих его подсистем, в котором индексы i и j сомножителей не равны друг другу.*

При этом возникает вопрос о природе Внешнего Наблюдателя. Очевидно, что в рамках существующей научной парадигмы множество Внешних Наблюдателей пусто, ибо единственная непосредственно данная нам истинно изолированная система – универс – включает в себя все сущее. В этом смысле можно понять категоричность Джона А. Уилера: «Наблюдения не делаются извне системы неким супернаблюдателем. Для использования теории «наблюдения извне» нет подходящего наблюдателя» [15]. Такой взгляд был оправдан на начальном этапе развития эвереттики, когда ещё не была осознана связь эвереттики с психофизической проблемой и не была столь ясно определена роль сознания в квантовых процессах. Однако в настоящее время, с современной эвереттической точки зрения, Внешним Наблюдателем для данного универса может являться любой наблюдатель из любой другой ветви мультиверса, в том числе и другая ветвь мультивидуума наблюдателя S_1 .

Согласно развиваемым в эвереттике взглядам о мультивидууме (термин предложен Ю.В. Никоновым [16]) у него столько «Я», в скольких ветвях мультиверса он присутствует. Но кроме этого есть ещё и «НАД-Я», объединяющее эти «универсные Я» и являющееся, собственно, *МУЛЬТИвидуумом*.⁷

⁷Разумеется, у работающих в эвереттике исследователей на этот счет существуют различные мнения. Вот что сообщил мне А. Костерин: «По-моему, мультивидуум не имеет надличностной сущности, т.е. сверхсознания. Мультивидуум проявляется через бессознательное в

При этом сущностное единство обоих наблюдателей (то, что выше названо «НАД-Я») снимает вопрос о нарушении изолированности системы S – обмен информацией между наблюдателями физически происходит не через ее границу, а является особым квантовым процессом (можно предположить, что типа «туннелирования информации»), природу и закономерности которого ещё предстоит изучить.

Другим направлением исследований в этой связи представляется весьма интересная область, развиваемая А.К. Гуцем и Е.В. Палешевой, – изучение спинорных духов [17]. Эти квантовые объекты интерпретируются ими как «тенивые частицы» Д. Дойча. И именно спинорные духи могут рассматриваться как претенденты на физическую основу передачи информации между различными воплощениями мультивидуума.

Таким образом, парадокс наличия Внешнего наблюдателя изолированной системы разрешается в эвереттике «офизичиванием метафизичности» – введением в физическую картину новой сущности – мультивидуума. И поставленная в основополагающей статье Эверетта задача – построение квантовой механики для изолированной системы – оказывается гёделевской задачей и разрешается в рамках новой метасистемы «Мультиверс-Мультивидуум».

В заключении анализа слагаемого $a_{12}\xi_1\eta_2$ отметим, что в общем случае, при анализе соотношения (4), копенгагенская интерпретация не требует обязательного равенства нулю a_{12} . Из этого ясно виден конвенционализм копенгагенской интерпретации, порождающей парадокс «коллапса волновой функции».

3. $a_{21}\xi_2\eta_1$ – наблюдатель зафиксировал прохождение частицы через щель 2, а в подсистеме S_2 частица прошла через щель 1! Случай, полностью аналогичный предыдущему, и описывает склейку, «симметричную» рассмотренной ранее ⁸.

4. $a_{22}\xi_2\eta_2$ – наблюдатель зафиксировал прохождение частицы через щель 2 и в подсистеме S_2 частица прошла через щель 2. Это соответствует одной из возможных «классических альтернатив» копенгагенской трактовки квантовой механики и, в соответствии с ней, «реализуется» с вероятностью, пропорциональной $|a_{22}|^2$.

В более сложных случаях, когда система и наблюдатель могут находиться в большем числе состояний ($i, j = 1, 2, 3, \dots, k$), число членов суперпозиции возрастает.

каждой личности, и сознание его растёт вместе с сознанием каждой личности. Осознание каждой составляющей личностью мультивидуума как целого – это и есть проявляющееся общее сознание. Эффект «покровителя», «предзнания» и т.д. появляется в результате взаимодействия различных частей мультивидуума через бессознательное. Полностью самосознающее состояние мультивидуума – это конечно внемирное состояние, но и в мире сем мы можем к нему приближаться». [23] Не вступая в полемику, замечу, что указанные проявления мультивидуума у конкретного «Я» в равной степени характерны для обеих гипотез о сущности сознания мультивидуума.

⁸Интерпретационно случай весьма схож с интерпретацией функции состояния $a_{12}\xi_1\eta_2$. Однако и для того, и для другого случаев, можно представить и «зеркальную» интерпретацию: винтик был положен под правую руку, а найден – под левой! С учетом изложенного, «классический наперсточник», сознание которого очевидно не управляет движениями пальцев (они, в современном представлении психофизиологии, управляются подсознанием), каким-то образом влияет на взаимодействие ветвей мультиверса, вызывая желательные для него склейки.

Если записать матрицу $||a_{ij}||$, то в «классическом» информационном подходе «реальная» суперпозиция будет описываться только диагональной квадратной подматрицей порядка $n = \min(i, j)$.

Однако для внешнего Метанаблюдателя, который, как показано выше, есть элемент физически не пустого множества, нет никаких логических оснований для исключения «нелогичных» элементов матрицы $||a_{ij}||$ – склеек, кроме аргументов «здравого смысла», которые, как свидетельствует история физики, неоднократно опровергались экспериментально.

Поскольку пока не существует теоретических способов расчета элементов a_{ij} , «стихийный эксперимент» свидетельствует только о том, что в большинстве случаев они достаточно малы.

Для более строгой оценки коэффициентов a_{ij} можно предложить провести статистический анализ баз данных по «ошибкам» операторов таких специальностей, как авиадиспетчеры и космонавты, поскольку профессиональная деятельность этих категорий операторов достаточно документирована и имеет богатую статистику. Также целесообразно было бы провести прямые психофизиологические эксперименты по тестированию адекватности реакции наблюдателя на некоторый модельный событийный ряд.

Поскольку уравнения (7)–(8) содержат те же самые слагаемые, которые были проанализированы выше, никаких принципиальных отличий от уже рассмотренных соотнесенных состояний Внешний Наблюдатель не обнаружит. Однако наличие полного спектра соотнесенных состояний системы S позволит ему определить, какого типа склейки возможны для данного состояния наблюдателя S_1 или объекта S_2 .

Проведенный анализ позволяет утверждать, что количество состояний-склеек типа «чудо» в суперпозиционных представлениях систем, имеющих i состояний наблюдателя и j состояний объекта, может быть вычислено по формуле:

$$N_{\text{склеек}} = ij - \min(i, j). \quad (9)$$

Уравнение (9) однозначно свидетельствует о том, что $N_{\text{склеек}}$ – конечное число.

В связи с тем, что «информационный энергообмен» является только частным случаем взаимодействия систем, можно уверенно предположить, что при теоретическом анализе более активных процессов, когда их энергия достаточна для непосредственного изменения наблюдаемых параметров состояния системы, будут обнаружены новые эффекты и проявления *динамики* многомирия, до сих пор математически описывавшегося только как *состояние*⁹.

Разумеется, явления «энергетического взаимодействия» встречаются на практике столь часто, что неоднократно описывались и анализировались с позиций квантовой механики.

При этом следует особо подчеркнуть, что получаемая в таких процессах наблюдателем информация, в отличие от энергоинформационных процессов,

⁹Если опять воспользоваться музыкальными образами, такое взаимодействие радикальным образом меняет тембр Де Бройлевской волны объекта наблюдения, и Наблюдатель «слышит» всё ту же мелодию (система «объект-наблюдатель» остается изолированной), скажем, песню «Валенки», но исполняемую не на балалайке, а симфоническим оркестром.

относится *исключительно к прошлому* наблюдаемой системы. Так, например, в известном опыте Штерна-Герлаха устанавливается информация о величине проекции магнитного момента атома *в момент его прохождения между полюсами магнита*. Вопрос об этой характеристике после взаимодействия с магнитом в копенгагенской интерпретации может быть решен только в результате нового измерения.

Эвереттическая же трактовка позволяет предсказать весь «спектр *настоящих*» этого атома. Правда, какое из этих настоящих станет реальностью того «Я» мультивидуума, которое в следующий момент будет вспоминать об условиях проведенного над этим атомом эксперимента, – это отдельная эвереттическая проблема, сводящаяся к решению вопроса об информационном взаимодействии ветвей «Я».

Особенно перспективной с эвереттической точки зрения областью исследования «энергетических взаимодействий» является химия. Это объективно связано с тем, что при изучении химических процессов исследователь имеет дело с системами, содержащими «термодинамические количества» отдельных объектов (обычно моли веществ, каждый из которых содержит $6,02 \cdot 10^{23}$ частиц). Химические процессы протекают на молекулярном уровне, т.е. в области наиболее яркого проявления квантовых законов¹⁰. А современный уровень развития физико-химических методов анализа дает возможность обнаруживать весьма редкие (маловероятные) исходы молекулярных взаимодействий и подтверждать это статистически достоверно. Да и достигнутое в химии временное разрешение – фемтосекундная спектроскопия – позволяет проявлять весьма тонкие аспекты квантового взаимодействия.

И именно химия дает примеры «материальных капель мультиверса» – объектов, которые зримо демонстрируют физически реальные системы, находящиеся в состоянии квантовой суперпозиции. Это так называемые «переходные состояния», или «активированные комплексы», возникающие в процессах химического превращения реагентов в продукты химических реакций.

Другим примером подобного рода объектов являются квантовые точки и квантовые кубиты – новые объекты, порожденные потребностями нанотехнологий. Электроны, ионы или атомы, входящие в состав квантовых точек и находящиеся в специально вызванном возбужденном состоянии, могут выполнять функции кубитов – рабочих элементов квантовых компьютеров. И отработка способов их создания является одной из приоритетных химических задач в об-

¹⁰С.В.Дёмин в связи с этим спросил: «...для меня не совсем понятен вопрос – какова, с точки зрения эвереттовской трактовки, область «наиболее яркого» проявления квантовых законов?» [24]. Ответ на него важен для понимания места квантовой механики в современном мире. Разумеется, понятие молекулярного уровня уже является образным выражением. Но важность именно этого уровня вызвана тем, что мир «классической квантовой механики» – это именно электронно-атомно-молекулярный мир. Там лежат все экспериментальные основы этой теории, и там продолжается накопление опытных данных. Там выявлялась особая мировоззренческая позиция, с которой сейчас эвереттика пытается посмотреть на «наш макромир». Это естественно: квантовая механика в период своего становления только «оглядывалась по сторонам», а мы «здесь и теперь», с этой, «обустроенной» ею позиции, пытаемся «поднять глаза в небо».

ласти высоких технологий.

При решении задач, обозначенных в последнем примере, возникают и другие квантовые проблемы, связанные с эвереттикой. Так, вследствие того что на наноуровне становится весьма существенной роль флуктуаций (в частности, плотности равновесного с данным объектом теплового излучения, описанные ранее в [18]), придётся описывать термодинамическое состояние нанообъектов как суперпозицию состояний с различной температурой. И здесь важную роль будет играть учет состояний-склеек.

Следует особо подчеркнуть, что нанотехнологии являются тем физическим пределом, который атомная структура вещества поставила перед развитием микроминиатюризации. И тот факт, что именно в ходе реализации потенциальных возможностей нанотехнологий существенную роль начинают играть слабо изученные эффекты многомирия, показывает, что в этой области должен произойти переход к новому технологическому скачку – сознательному использованию этих эффектов в практических целях.

В заключение следует отметить, что рассмотренные нелинейные семантические аспекты основных понятий, введенных Х. Эвереттом в квантовую механику, – параметры памяти волновой функции и соотнесенного состояния – для их более ясного понимания должны быть рассмотрены на более глубоком физико-математическом уровне ¹¹.

Вместе с тем уже представленные здесь результаты дают возможность проведения экспериментальной проверки идей Х. Эверетта и тем самым укрепляют надежду на то, что с помощью развития на этой основе понятийного аппарата эвереттики будут вскрыты новые важные свойства мультиверса.

Благодарности

Автор благодарит д.ф.-м.н. Л.В. Ильичева (Новосибирск, Россия) за весьма плодотворную дискуссию и ценные замечания, к.ф.-м.н. П.Р. Амнуэля (Бейт-Шеан, Израиль) за обсуждение текста и советы, А.М. Костерина (Нюрнберг, Германия), С.В. Дёмина (Омск, Россия) за плодотворные вопросы и обсуждение текста, Д.А. Клеопова (Москва, Россия) за его доклад на Российском междисциплинарном семинаре по темпорологии «Переоткрытие времени: вклад Пригожина в междисциплинарный диалог», катализировавший появление некоторых идей данной работы.

¹¹В равной степени можно утверждать, что углубление и философского и филологического анализа будут способствовать прояснению этих эвереттических понятий. Ведь происходящее сейчас переосмысление роли и места Сознания в физической картине мироздания на философско-филологическом уровне можно уподобить возникновению суперпозиционного состояния « «многомирие» + «многомирие» ». И забытое в связи с орфографической реформой 1918 года слово «*mir*» снова оказывается востребованным для осмысления реальности. Лично мне это стало ясно в результате дискуссии с барнаульским философом Ю.А.Помазным, за что я ему искренно признателен. Вот как обозначил задачу Ю.А.Помазной: «...нужно отличить «Мир» от «Природа» ... Первое – проявление активности души (сознания), результат каковой активности – *Mir* (пусть будет так написано); второе – проявление иной активности» [25].

ЛИТЕРАТУРА

1. Everett H. III. "Relative State" Formulation of Quantum Mechanics // Reviews of Modern Physics. 1957. V.29, N.3. P.454-462.
2. Глобалистика. Международный междисциплинарный энциклопедический словарь / Гл. ред. и сост. И.И. Мазур и А.Н. Чумаков. Москва-Санкт-Петербург-Нью-Йорк, 2006. С.1013.
3. Хью Эверетт III. Формулировка квантовой механики через «соотнесенные состояния» / Пер. Ю.А. Лебедева на сайте МЦЭИ – <http://www.everettica.org/art/Ever2.pdf>.
4. Лебедев Ю.А. Природа необратимости с позиций эвереттики / В сб.: «Необратимые процессы в природе и технике», ч.2, ФИАН, 2007. С.595-598.
5. Потапов А.А., Лактюшкин А.В., Корешков Д.В. Индикатриссы рассеяния электромагнитных волн фрактальной поверхностью, синтезированной на основе модификаций недифференцируемой функции Вейерштрасса / В сб.: «Необратимые процессы в природе и технике», ч.1, ФИАН, 2007. С.40-43.
6. Грин Б. Элегантная вселенная. Суперструны, скрытые размерности и поиски окончательной теории. М., Едиториал УРСС, 2005. 288 с.
7. Невеский Н.Е. О законе фазовой гармонии Луи Де Бройля (опубликовано в «Библиотеке электронных публикаций» сайта Института исследований природы времени,
– http://www.chronos.msu.ru/RREPORTS/nevessky_o_zakone/nevessky_o_zakone.htm.
8. Ефремов М., Каширин С. Скрытая масса. На сайте «Наша Галактика»
– http://galaxias.narod.ru/struktur_n-s_soln.htm#скрытая%20масса.
9. Талбот М. Голографическая вселенная. Новый вид поля. Сайт «Библиотека «Полка букиниста». – http://polbu.ru/talbot_holuniverse/ch13_all.html.
10. Потапов А.А. Фракталы в радиофизике и радиолокации: Топология выборки. М.: Университетская книга, 2005. 848 с.
11. Менский М.Б. Концепция сознания в контексте квантовой механики // УФН. 2005. Т.175, N.4. С.413-435.
12. Лебедев Ю.А. Неоднозначное мироздание. Кострома, 2000. С.106.
13. Лебедев Ю.А. Феноменологический формализм описания взаимодействующих ветвей мультиверса, мультивидуума и явления «склеек ветвей мультиверса» // Математические структуры и моделирование. 2006. Вып.16. С.59-66.
14. Менский М.Б. Квантовая механика: новые эксперименты, новые приложения и новые формулировки старых вопросов // УФН. 2000. Т.170, N.6. С.631-648.
15. Wheeler J.A. Assessment of Everett's "Relative State" Formulation of Quantum Theory // Reviews of Modern Physics. 1957. V.29, N.3.
16. Никонов Ю.В. Многомировая интерпретация диссоциативных расстройств психики // Сознание и физическая реальность. 2006. Т.11, N.2. С.32-35.
17. Палешева Е.В.. Спинорные поля с нулевым тензором энергии-импульса. Автореферат диссертации на соискание ученой степени кандидата физико-математических наук. Омск, 2004. 16 с.
18. Лебедев Ю.А. Радиационно-неравновесные фазы. Деп. ВИНТИ N.6672-B85 от 16.09.85, Ук.N.2 за 1985 г. 17 с.
19. Линде А.Д. Письмо к организаторам Семинара, посвященного 50-летию публикации статьи Х. Эверетта. – <http://www.everettica.org/news.php3>.

-
20. Язневич В. «Закон Доньды» Станислава Лема.
– <http://mindspring.narod.ru/lib/lem/kg/kg02905.html>.
 21. Ильичев Л.В. Частное сообщение по e-mail от 28.11.06 13:58.
 22. Аспек А. Теорема Белла: наивный взгляд экспериментатора / Перевод М.Х. Шульмана доклада А. Аспека на конференции в Вене в декабре 2000 г., посвященной памяти Д. Белла. Опубликован в «Библиотеке электронных публикаций» сайта Института исследований природы времени.
– http://www.chronos.msu.ru/RREPORTS/aspek_teorema_bella.pdf.
 23. Костерин А.М. Частное сообщение по e-mail от 15.12.06 15:05.
 24. Дёмин С.В. Частное сообщение по e-mail от 26.03.07. 06:38.
 25. Помазной Ю.А. Частное сообщение по e-mail от 19.10.06. 15:54.

ОБЗОР МЕТОДОВ РАСПАРАЛЛЕЛИВАНИЯ АЛГОРИТМОВ РЕШЕНИЯ НЕКОТОРЫХ ЗАДАЧ ВЫЧИСЛИТЕЛЬНОЙ ДИСКРЕТНОЙ МАТЕМАТИКИ

С.С. Ефимов

В статье рассматриваются представители параллельных алгоритмов, относящихся к решению задач таких разделов вычислительной дискретной математики, как числовые ряды, комбинаторные задачи, линейная алгебра, вычисления в узлах сеток и решеток, обработка графов.

Дана характеристика параллельных методов сортировки чет-нечетной перестановкой, подсчетом сравнений, двухпутевым слиянием, Шелла, расширения Брюса-Вагара для быстрой сортировки, метода параллельного двоичного поиска, алгоритмов Фокса и Кэннона умножения матриц, методов Гаусса и простой итерации решения систем линейных уравнений, параллельного решения краевой задачи, «жадного» алгоритма Прима построения минимального охватывающего дерева, алгоритм Дейкстры поиска кратчайшего пути и др.

Рост вычислительной мощности компьютерных систем, появление современных суперкомпьютеров, кластеров рабочих станций сделали возможным решение многих задач дискретной математики, требующих выполнения больших объемов вычислений за приемлемое расчетное время.

Существует целый ряд отраслей, в которых возникает необходимость решения подобных задач [1, 2]: квантовая физика (физика элементарных частиц, ядерная физика, квантовая теория поля); статистическая физика; физика молекул, физика плазмы; квантовая химия; науки о Земле (физика атмосферы, метеорология, климатология, геофизика, физика океана); биология, экология; экономика и эконометрия (вычислительная экономика, макроэкономика, теория массового обслуживания и теория оптимального управления, финансовая деятельность); социальные науки; математическая лингвистика (распознавание речи, анализ текста и автоматический перевод); информатика (ведение баз данных, распознавание образов, распределенные вычислительные системы); механика сплошных сред (гидродинамика и газодинамика, теория сопротивления

материалов); баллистика; медицина, фармацевтика; промышленность, в том числе автомобиле- и авиастроение, нефте- и газодобыча, дизайн и другие.

Приведенный список не является исчерпывающим. Постоянное повышение мощности компьютерных систем приводит к тому, что задачи, которые еще в недалеком прошлом не могли быть решены в реальном масштабе времени, успешно решаются благодаря использованию параллельных алгоритмов, реализуемых на многопроцессорных системах, обладающих высоким быстродействием.

Попытки выполнения обычных последовательных алгоритмов решения известных задач дискретной математики на многопроцессорных вычислительных системах во многих случаях не приводят к повышению быстродействия. Для реального уменьшения времени решения требуется применение специальных параллельных вычислительных алгоритмов, учитывающих архитектурные особенности многопроцессорных систем. В связи с тем, что решения в прикладных областях, подобных приведенным в списке, сводятся к ряду известных задач дискретной математики, большой практический интерес представляют параллельные алгоритмы их решения.

Последовательным реализациям методов решения многих задач дискретной математики уделено достаточно много внимания в академической науке, в опубликованных статьях и монографиях. Параллельные аналоги известных последовательных алгоритмов к настоящему времени не были должным образом освещены в публикациях и не привлекали к себе необходимого внимания математиков и исследователей. Это объяснялось сначала отсутствием, а затем недостаточным развитием многопроцессорных вычислительных систем.

На современном этапе аппаратное обеспечение многопроцессорных систем развивается быстрыми шагами: в 2007 г., по сообщению Government Computer News, предполагается появление в США (компания Cray) суперкомпьютера производительностью более 1 петафлопса (10^{15} операций с плавающей запятой в секунду). В настоящее время предел быстродействия суперкомпьютера достигает 280×10^{12} флопсов (количество его процессоров превышает 130 тысяч) [3]. Поэтому любые исследования, связанные с развитием параллельных алгоритмов для многопроцессорных быстродействующих систем, а также популяризация и освещение в печати подобных алгоритмов представляются актуальными.

В статье будут рассмотрены преимущественно алгоритмы, предназначенные для выполнения на распределенных вычислительных системах типа кластеров рабочих станций. Параллельным алгоритмам, ориентированным на выполнение в вычислительных системах с общей памятью, также будет уделено некоторое внимание в тех случаях, когда это необходимо для сравнения с алгоритмами, предназначенными для выполнения в многопроцессорных системах с распределенной памятью.

В связи с тем, что полный список задач дискретной математики является весьма обширным, а объем настоящей публикации ограничен, в ней представлен обзор параллельных алгоритмов задач, относящихся к числу типовых и наиболее часто используемых. Приводимый материал позволит читателю получить представление о приемах, используемых при распараллеливании алгоритмов

решения различных задач дискретной математики, а при необходимости более детального знакомства – обратиться к соответствующим первоисточникам, на которые в работе имеются ссылки.

В статье рассматриваются представители параллельных алгоритмов, относящихся к решению задач таких разделов вычислительной дискретной математики, как:

- числовые ряды;
- комбинаторные задачи;
- линейная алгебра;
- вычисления в узлах сеток и решеток;
- обработка графов.

Рассмотрим представителей каждой из названных групп.

1. Числовые ряды

К этому разделу относятся задачи определения различных констант (числа π , числа e и др.), представляемых в виде числового ряда [4], нахождение определенных интегралов, суммирование большого количества констант и подобные этому задачи, определение всех частичных сумм, выполнение операций параллельного префикса (сканирования). Здесь же рассматриваются задачи, решаемые аналогично перечисленным выше: поиска глобального максимума, минимума, вычисления значений логических выражений над длинными последовательностями величин и т. п.

Алгоритмы, позволяющие выполнять параллельное суммирование с помощью многопроцессорных систем с общей памятью, могут выполняться с помощью каскадной схемы суммирования, модифицированной каскадной схемы [5]. В названном первоисточнике [5] приводится также схема вычисления всех частных сумм.

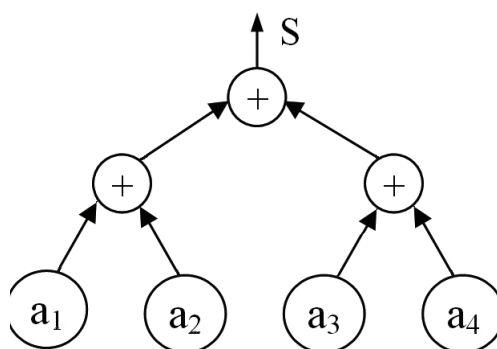


Рис. 1. Каскадная схема

Каскадная схема суммирования имеет вид бинарного дерева, в листьях которого записаны члены ряда, а во всех остальных узлах – операции сложения. Например, чтобы найти сумму $S = a_1 + a_2 + a_3 + a_4$, потребуется схема, представленная на рис. 1.

Эффективность рассмотренной схемы определяется как

$$S_p = T_1/T_p = (n - 1)/\log_2 n ,$$

где T_1 - трудоемкость алгоритма для одного процессора, T_p - для сети из p процессоров.

Эффективность использования процессоров вычисляется по формуле

$$E_p = T_1/pT_p = (n - 1)/((n/2) \log_2 n) .$$

Для данной схемы $E_p \rightarrow 0$ при $n \rightarrow \infty$.

Если допустить, что каскадная схема вычисления выполняется с помощью кластера, каждый узел которого делает одну операцию сложения, то время на пересылку промежуточных результатов будет неизмеримо больше времени вычисления. Применение кластера по такой схеме оказывается нецелесообразным. Однако при использовании многопроцессорного суперкомпьютера с общей памятью каскадная схема является очень эффективной. Ограничением будет являться максимально допустимое на современном этапе количеством процессоров в одном суперкомпьютере – примерно 130 тыс. И то это в предположении, что все процессоры компьютера можно задействовать для вычислений по каскадной схеме. Решение задачи суммирования большого количества слагаемых, измеряемого сотнями тысяч и миллионами чисел, потребует использовать другие вычислительные схемы.

В этом случае для решения задачи с помощью многопроцессорной вычислительной сети с распределенной памятью (кластера) можно предложить, например, использовать каскадные схемы в комбинации с операцией редукции по сумме (см. рис. 2), если многопроцессорные узлы кластера реализуют каскадную схему, а редукцию по сумме выполнит обычный последовательный процессор. Схема обработки для такой реализации получается, если лес, состоящий из бинарных деревьев, соединить связями от корней образующих этот лес деревьев с новым узлом – корневым узлом сформированного дерева. Фрагменты части рисунка, расположенные ниже штрих-пунктирной линии, выполняются на многопроцессорных узлах кластера. Фрагмент, который расположен выше штрих-пунктирной линии, соответствует однопроцессорному корневому узлу.

Аналогом этой схемы является схема, в которой вместо редукции выполняется обычное последовательное суммирование (рис. 3). На этом рисунке фрагменты части рисунка, расположенные ниже штрих-пунктирной линии, также выполняются на многопроцессорных узлах кластера. Фрагмент, который расположен выше штрих-пунктирной линии, соответствует однопроцессорному корневому узлу. Однако вычислительная трудоемкость данной схемы примерно такая же, как и у схемы на рис. 2, так как редукция по сумме выполняется на обычном последовательном компьютере.

Выбор одной из двух схем зависит от программного обеспечения, используемого для последовательного суммирования, от эффективности его реализации в конкретной программной системе.

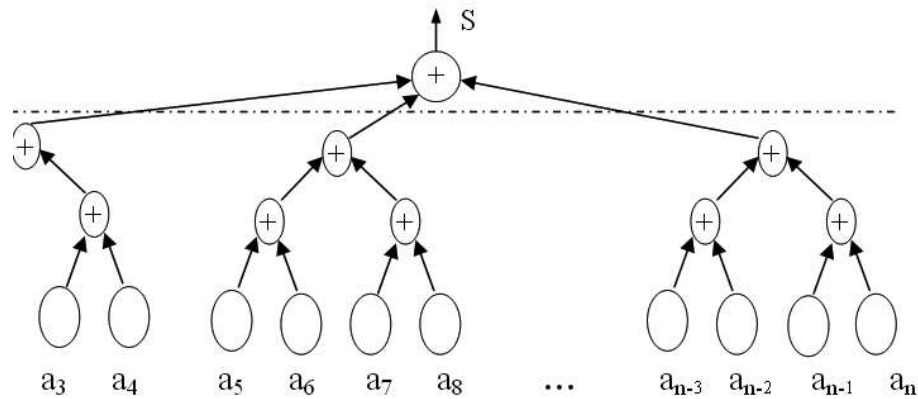


Рис. 2. Каскадная схема с редукцией по сумме

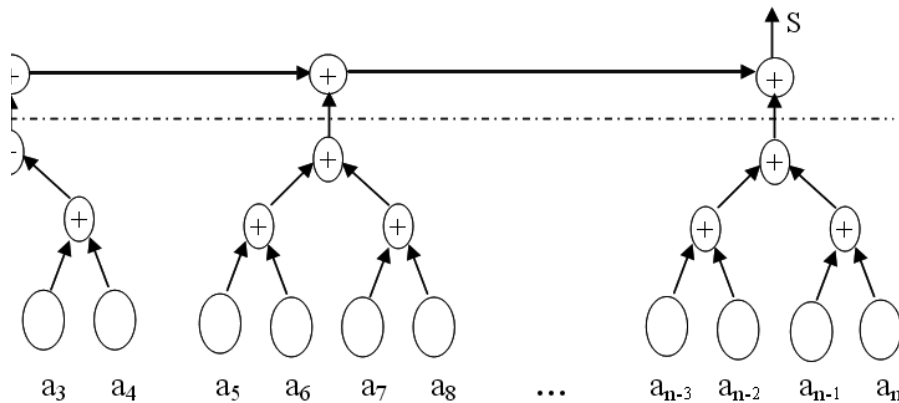


Рис. 3. Каскадно-последовательная схема

Если же в распоряжении имеется кластер, позволяющий лишь на одном узле реализовать каскадную схему, а на других – только последовательное вычисление, то можно использовать другую схему из [5] (см. рис. 4). Фрагменты части рисунка, расположенные ниже штрих-пунктирной линии, выполняются на однопроцессорных узлах кластера. Фрагмент, который расположен выше штрих-пунктирной линии, соответствует многопроцессорному корневому узлу.

Время сортировки рассмотренного метода определяется как

$$T_p = 2 \log_2 n, \quad p = (n / \log_2 n).$$

Ускорение и эффективность определяются формулами

$$S_p = (n - 1) / 2 \log_2 n, \quad E_p = (n - 1) / 2n.$$

По сравнению с обычной каскадной схемой ускорение уменьшается в два раза, однако оно стремится к 0.5 при росте n ($E_p \rightarrow 0.5$ при $n \rightarrow \infty$).

Последний вариант (рис. 5) соответствует случаю, когда все узлы кластера, включая корневой, являются многопроцессорными. Он позволяет ускорить вычисление результата на заключительном этапе объединения частичных резуль-

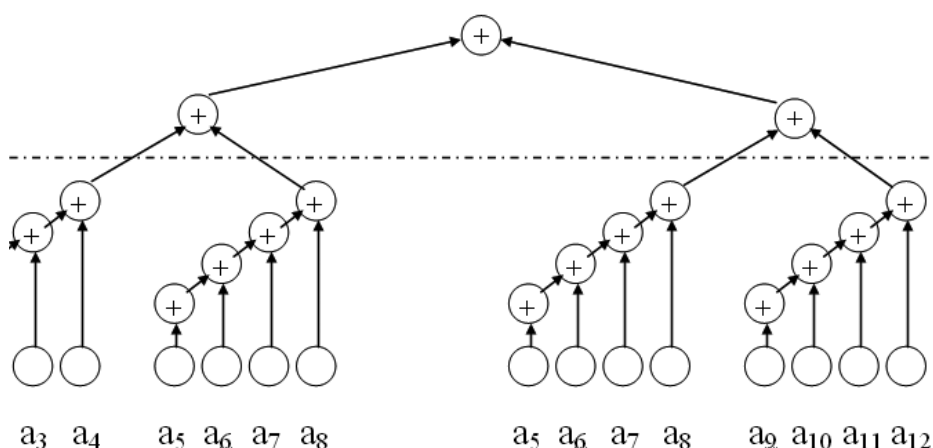


Рис. 4. Последовательно-каскадная схема

татов, а также использовать узлы с ограниченным количеством процессоров в каждом из них.

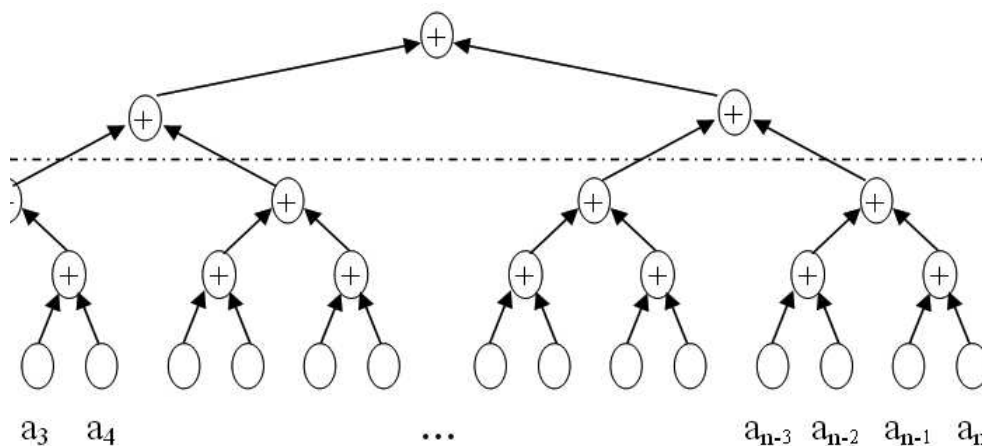


Рис. 5. Каскадно-каскадная схема

Алгоритм вычисления всех частичных сумм [5] рассматривать не будем, так как в представленном виде он не предназначен для распределенной обработки.

Для понимания схемы вычисления определенного интеграла с помощью кластера следует воспользоваться свойством его аддитивности [6]

$$\int_a^b f(x) dx = \sum_{i=0}^{p-1} \int_{a_i}^{b_i} f(x) dx ,$$

где $a_i = i * d$, $b_i = a_i + d$, $d = (b - a)/p$, p - число процессоров (не считая корневого). Тогда с учетом сказанного схема вычисления будет иметь вид, представленный на рис. 6.

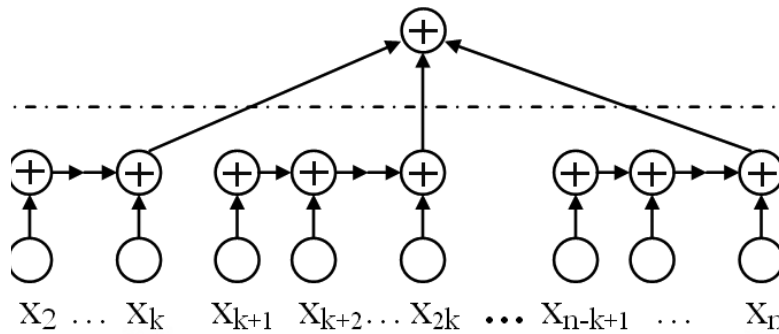


Рис. 6. Схема вычисления определенного интеграла

По данной схеме суммируется n значений с помощью $p + 1$ узла кластера. Каждому процессору, кроме корневого, достается по $k = n/p$ слагаемых. После нахождения всех последовательных сумм (ниже пунктирной линии) корневой процесс выполняет редукцию по сумме этих значений. Если корневой узел кластера является многопроцессорным, то он может вместо редукции выполнить сложение по каскадной схеме.

Для вычисления суммы числового ряда, например числа π [7], можно воспользоваться схемой, аналогичной той, которая представлена на рис. 6, изменив в ней порядок группирования слагаемых следующим образом. Первый узел кластера будет вычислять сумму $S_1 = x_1 + x_{k+1} + \dots + x_{n-k+1}$, второй процессор найдет сумму $S_2 = x_2 + x_{k+2} + \dots + x_{n-k+2}$, ..., p -й процессор найдет $S_p = x_k + x_{2k} + \dots + x_n$. После этого корневой процессор выполнит редукцию и вычислит $\pi = S = S_1 + S_2 + \dots + S_p$. Такой порядок группировки слагаемых легче программируется, позволяет более равномерно загрузить вычислительной работой процессоры в том случае, если сложность вычисления последних членов ряда (если они представлены в виде некоторого выражения) существенно отличается от сложности вычисления начальных.

Таким образом, идея распараллеливания вычислений при подсчете значений числовых рядов заключается в том, что все слагаемые делятся некоторым образом в зависимости от метода на определенное количество промежуточных групп, после чего подсчет сумм в группах выполняется одновременно разными процессорами вычислительной системы. Затем значения частичных сумм складываются для получения конечного результата. Способы вычисления сумм в промежуточных группах и в конечной группе зависят от используемого метода.

Эту же идею можно использовать при поиске максимальных, минимальных значений среди последовательности величин, подсчете значений логических выражений с операциями И, ИЛИ и другими бинарными функциями, как и с функциями от большего числа аргументов.

2. Комбинаторные задачи

Среди задач, относящихся к данной группе, можно обратить внимание, в первую очередь, на параллельные методы сортировки и поиска данных.

Суть многих параллельных алгоритмов сортировки [8] состоит в том, что упорядочиваемые данные разбиваются на какие-либо части в зависимости от метода, каждая из которых сортируется по отдельности разными процессорами. Затем упорядоченные части объединяются некоторым образом, опять же в зависимости от метода. Эти два этапа могут повторяться необходимое число раз.

Из методов параллельной сортировки в качестве примеров рассмотрим наиболее типичных представителей: метод пузырька, сортировку Шелла, быструю сортировку.

Для параллельной реализации метода пузырьковой сортировки используется модификация, называемая методом чет-нечетной перестановки [9]. В простейшем случае сортировка заключается в чередовании двух этапов. На первом этапе массив разбивается на пары $(a_0, a_1), (a_2, a_3), (a_4, a_5), \dots, (a_{n-2}, a_{n-1})$, начинающиеся с четных индексов. В каждой паре слева помещается наименьший элемент из двух, справа – наибольший из них. На втором этапе аналогичная обработка выполняется в парах $(a_1, a_2), (a_3, a_4), (a_5, a_6), \dots, (a_{n-3}, a_{n-2})$, начинающихся с нечетных индексов. Максимум через n повторений массив оказывается упорядоченным. Иллюстрация приведена на рис. 7. Стрелками обозначены пары элементов, которые могут поменяться местами.

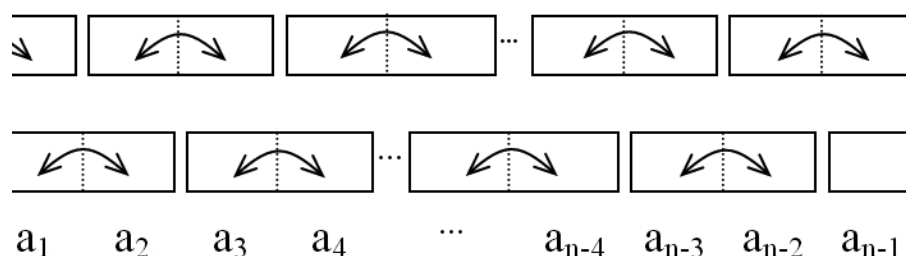


Рис. 7. Чет-нечетная сортировка

В рассмотренном методе подразумевается, что число процессоров равно числу n элементов массива A . Такая реализация может быть использована в многопроцессорном компьютере, но для кластера рабочих станций она не годится, так как затраты на пересылку соседних элементов будут превышать время на выполнение операций по сортировке.

Чтобы приспособить метод к реализации в распределенной вычислительной системе, состоящей из p процессоров, все n сортируемых элементов делятся на части, содержащие по n/p элементов. Каждая из частей рассылается в локальную память соответствующего процессора и предварительно сортируется в нем каким-либо быстрым методом. Затем чередуются два этапа:

1) Обмен фрагментами, состоящими из n/p элементов, между соседними процессорами (рис. 8). Этот обмен выполняется по очереди:

- 1а) внутри четных пар процессоров с номерами (0,1), (2,3), (4,5) ...,
 1б) внутри нечетных пар процессоров с номерами (1,2), (3,4), (5,6) ...

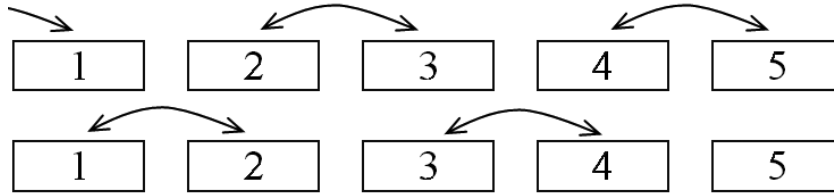


Рис. 8. Порядок обмена фрагментами в чет-нечетной сортировке

2) В каждом процессоре выполняется операция «сравнить и разделить» (compare-split). При ее выполнении процессор объединяет два упорядоченных фрагмента в один упорядоченный удвоенной длины $2n/p$. Для этого можно использовать метод сортировки слиянием («слить» два в один). Затем левый процессор пары оставляет у себя левую половину результата, а правый процессор – правую половину (рис. 9).

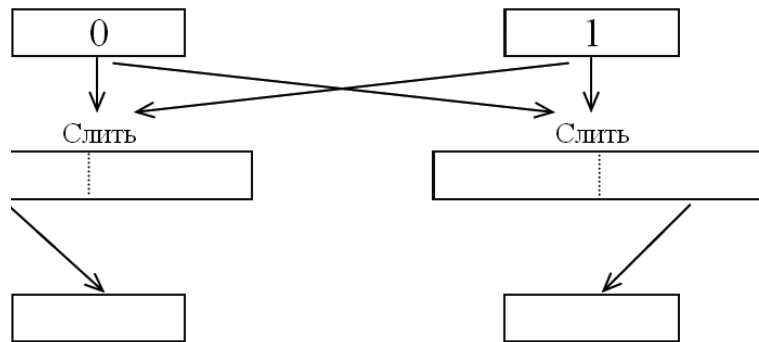


Рис. 9. Операция «Сравнить и разделить»

Для выполнения слияния необязательно выполнять слияние элементов до получения результата длиной $2n/p$. Достаточно в левом процессоре сливать фрагменты, начиная слева, до получения результата длиной n/p , а в правом процессоре сливать, начиная справа, также до получения результата длиной n/p . С более подробной иллюстрацией чет-нечетной сортировки можно ознакомиться в [10].

Вычислительная трудоемкость метода равна

$$T_p = (n/p) \log(n/p) + 2n .$$

Показатели ускорения и эффективности для p процессоров определяются как

$$S_p = n \log n / ((n/p) \log(n/p) + 2n) ,$$

$$E_p = n \log n / (p((n/p) \log(n/p) + 2n)) .$$

Рассмотрим теперь метод сортировки Шелла [11]. При последовательной реализации он имеет то преимущество, что перестановки выполняются между элементами, расположенными друг от друга на большом расстоянии. Для этого в массиве из n элементов на первом этапе сортируются элементы внутри групп, включающих элементы с номерами $[0, n/2]$, $[1, n/2+1]$, $[2, n/2+2]$, $\dots$, $[n/2-1, n-1]$. На втором этапе группы образуются из элементов, расположенных в 2 раза ближе друг к другу. В каждой группе находится в 2 раза больше элементов, то есть: $[0, n/4, n/2, 3n/4]$, $[1, n/4+1, n/2+1, 3n/4+1]$, $\dots$, $[n/4-1, n/2-1, 3n/4-1, n-1]$. На следующем этапе расстояние между элементами уменьшается еще в 2 раза, а число элементов в группе увеличивается в 2 раза и т.д. На последнем этапе сортируется весь массив как одна группа.

На рис. 10 представлена иллюстрация сортировки методом Шелла для массива из 8 элементов. Верхние связи соответствуют четырем группам 1-го этапа, нижние связи – двум группам 2-го этапа.

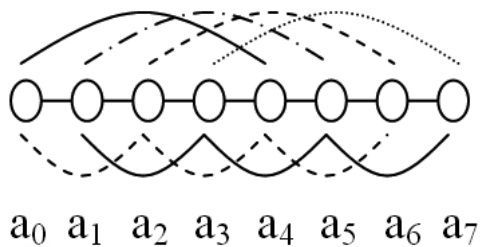


Рис. 10. Метод Шелла

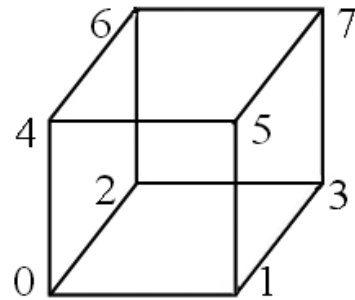


Рис. 11. Топология куба

Чтобы получить параллельную реализацию метода Шелла, необходимо объединить процессоры коммуникационной сети в топологию N -мерного гиперкуба [5]. Для этого в сети должно быть 2^N процессоров. Тогда максимальное расстояние между элементами будет равно N (2^N узлов, связанных последовательно, имели бы наибольшее расстояние $2^N - 1$). На рис. 11 представлен пример 3-мерного гиперкуба (имеет вид обычного куба), у которого последовательность нумерации вершин (узлов сети) позволяет использовать код Грея при обходе в порядке $0 - 1 - 3 - 2 - 6 - 7 - 5 - 4$. Код Грея, который используется для установления соответствия двух систем нумерации, характеризуется тем, что номера соседних процессоров отличаются только одним битом двоичного представления.

На предварительном этапе выполняется деление исходного массива на 2^N частей и рассылка каждой из них соответствующему узлу кластера. Затем все узлы кластера сортируют свои части массива каким-либо быстрым методом.

Перед началом процесса сортировки между узлами, являющимися соседними в структуре гиперкуба, организуются коммуникационные связи. На первом этапе в последовательной реализации метода Шелла обмен должен выполняться между удаленными друг от друга парами узлов с номерами: $0 - 4, 1 - 5, 2 -$

6, 3 – 7. В параллельной реализации операция «сравнить и разделить» также будет выполняться между этими же парами узлов. Однако расстояние между узлами в структуре гиперкуба будет равно не четырем, а единице (см. рис. 10). На следующем этапе, как и в последовательной реализации метода, в группах с номерами 0 – 2 – 4 – 6 и 1 – 3 – 5 – 7 выполняются этапы обычной чет-нечетной сортировки до тех пор, пока не прекратится изменение порядка значений. Операция «сравнить и разделить» всегда проводится таким образом, что меньшая часть блока остается в процессоре, соответствующем вершине гиперкуба с меньшим номером, а большая часть – в процессоре с большим номером.

На заключительном этапе упорядоченные части собираются в единый упорядоченный массив. Обмен фрагментами на всех этапах (предварительном, основном и заключительном) в топологии гиперкуба выполняется существенно быстрее, чем в линейной топологии. Чем больше размерность гиперкуба, тем более эффективно использование метода Шелла.

Трудоемкость рассмотренного метода при числе итераций L (от 2 до p) определяется значением

$$T_p = (n/p) \log(n/p) + (2n/p) \log p + L(2n/p) .$$

Перейдем к рассмотрению метода быстрой сортировки. Последовательная реализация [12] заключается в неоднократном разделении упорядочиваемых значений на два фрагмента, таких, что элементы первого меньше всех элементов второго. К полученным частям снова применяется этот же прием и т.д. Для того чтобы распределить элементы левой и правой части на два фрагмента, в каждой из них выбирается ведущий элемент, после чего элементы, которые меньше его значения, переносятся в один фрагмент, а большие – в другой. При удачном выборе ведущих элементов трудоемкость алгоритма T_1 пропорциональна выражению $n \log n$. Среднее значение [11] определяется формулой $T_1 = 1.4n \log n$.

Для параллельной реализации метода быстрой сортировки (расширение Брюса Вагара [8] – т. н. гипербыстрая сортировка) так же, как и в случае метода Шелла, создается коммуникационная топология в виде N -мерного гиперкуба для кластера, содержащего 2^N процессоров. Для сортировки n значений предварительно каждому узлу рассылается блок из n/p элементов. Затем N раз выполняется следующая последовательность действий:

- выбирается ведущий элемент с рассылкой его значения всем процессорам гиперкуба;
- блок в каждом процессоре разбивается с учетом значения ведущего элемента на две части;
- организуются связи между парами процессоров, двоичные коды номеров которых отличаются только в N -ой позиции;
- выполняется взаимный обмен частями внутри каждой пары процессоров так, чтобы у процессоров, содержащих 0 в двоичном представлении своего номера, осталось по две части с меньшими значениями элементов, а у процессоров с 1 в номере – две части с большими значениями.

Каждый раз после однократного выполнения перечисленных действий гиперкуб порядка N оказывается разбитым на два гиперкуба порядка $N - 1$, таких, что у гиперкуба с нулем в старших разрядах двоичных номеров его процессоров все значения упорядочиваемых элементов меньше всех значений другого гиперкуба, старший разряд двоичных номеров процессоров которого содержит единицу.

Чтобы облегчить удачный выбор ведущих элементов, желательно после рассылки каждому узлу гиперкуба блока значений выполнить их предварительную сортировку, а также обеспечить однородное распределение данных между процессорами.

Вычислительная сложность алгоритма, определяемая временем сортировки в процессорах, временем выбора ведущих элементов и сложностью разделения блоков, вычисляется по формуле:

$$T = (n/p) \log(n/p) + \log p + \log(n/p) \log p .$$

Можно предложить также параллельную реализацию сортировки подсчетом сравнений. Последовательная реализация метода [11] заключается в том, что в массиве A из n элементов каждый элемент A_i ($i = \overline{1, n-1}$) сравнивается с элементами A_j ($j = \overline{i+1, n}$). Если $A_i > A_j$, то соответствующий элемент вспомогательного массива B_i увеличивается на единицу, иначе $B_j = B_j + 1$. Элементам массива B до начала обработки присваиваются единицы. После выполнения всех сравнений он используется для указания местоположений элементов исходного массива в массиве-результате: $C_i = A(B_i)$ ($i = \overline{1, n}$).

Для параллельной реализации массив A , состоящий из n элементов, рассылается всем p узлам кластера. Затем каждый k -й узел ($k = \overline{0, p-1}$) реализует в своей локальной памяти метод подсчета сравнений между элементами k -го фрагмента массива A и всеми элементами массива B , то есть a_i ($i = kn/p, \dots, (k+1)n/p-1$) и a_j ($j = \overline{i+1, n}$). На следующем этапе выполняется суммирование одноименных элементов массива B из всех узлов с накоплением результата в одном узле, то есть операция редукции по сумме MPI_Reduce, если говорить в терминах интерфейса с передачей сообщений MPI.

Сортировка простым двухпутевым слиянием [11] также допускает параллельную реализацию. Предположим, что для сортировки используется сеть из двух процессоров, а числа в каждом из двух сливаемых фрагментов упорядочены по возрастанию. Перед началом работы оба фрагмента рассылаются каждому из процессоров сети. Первый процессор будет выполнять слияние фрагментов, начиная слева, второй – начиная справа. В средней части фрагментов слияние прекращается, а оба слитых фрагмента-результата после пересылки одного из них в какой-либо из двух процессоров объединяются в средней части (с учетом значений граничных элементов). Чтобы процесс объединения не требовал существенных затрат времени, числа в сливаемых фрагментах не должны существенно различаться по значениям, по крайней мере в средней части.

Изложенный в предыдущем абзаце прием можно обобщить на 4 процессора, разбив сливаемые фрагменты на две части и разослав левые части на 1-й и 2-й

процессоры, а правые части – на 3-й и 4-й. Затем каждая пара процессоров реализует рассмотренный выше алгоритм слияния двух фрагментов. Полученные фрагменты аналогичным образом сливаются вместе. Можно реализовать метод на 8-ми, 16-ти и т.д. процессорах.

Рассмотрев методы сортировки данных, остановимся на некоторых методах параллельного поиска. Алгоритм двоичного поиска среди упорядоченных величин в параллельной реализации [13] предполагает разбиение ключей, среди которых выполняется поиск, на последовательные группы по N/p элементов. Каждый из p процессоров сети производит двоичный поиск в своей группе элементов и, в случае обнаружения искомого ключа, передает сведения о нем основному процессору.

Время выполнения такого поиска имеет порядок $O(\log(N/p))$, а стоимость – $O(p \log(N/p))$. Если число процессоров равно $p = \log N$, то время выполнения имеет порядок $O(\log N)$, а стоимость имеет порядок $O(\log^2 N)$.

Параллельная реализация метода последовательного поиска среди неупорядоченных данных в сети из процессоров также может быть выполнена путем разбиения N значений на группы из N/p элементов с последующим поиском искомой величины каждым процессором в своей части и пересылкой результата центральному процессору. Время выполнения метода будет иметь порядок (N/p) , а стоимость – порядок (N) . То есть порядок стоимости остается таким же, как и для последовательного поиска на одном процессоре, однако время поиска удастся уменьшить (по сравнению с (N)) примерно в p раз.

3. Задачи линейной алгебры

Типичным представителем задач, относящихся к линейной алгебре, является умножение матрицы на матрицу. При последовательной реализации используется формула

$$C_{ij} = \sum_{k=1}^m A_{ik} B_{kj} \quad (i = \overline{1, n}; j = \overline{1, m}).$$

Методы решения этой задачи зависят от выбранной топологии вычислительной сети и от способа распределения матриц A , B , C по процессорам. Если сеть содержит p процессоров, то каждую из трех матриц можно распределить между ними одним из 4-х способов: на p горизонтальных полос, на p вертикальных полос, на сетку размера p на p , совсем не разбивать на части. Тогда, в зависимости от способов распределения матриц, получается $4^3 = 64$ варианта решения задачи [14]. Однако большинство из этих способов являются неэффективными.

Рассмотрим в качестве примера так называемый ленточный алгоритм. Если учесть специфику задачи, формулу последовательного алгоритма, то более логично разбить матрицы A и C на p горизонтальных полос, матрицу B – на p вертикальных полос [15]. Схема разбиения представлена на рис. 12.

Сеть в этом случае будет иметь топологию «кольцо». При использовании топологии «кольцо» предполагается, что горизонтальные полосы матрицы A , содержащие по n/p строчек, перед началом вычислений рассылаются в локаль-

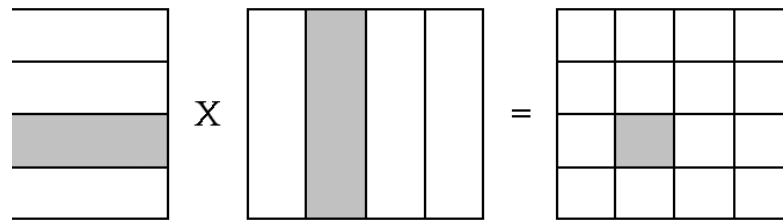


Рис. 12. Схема разбиения матриц при умножении по ленточному алгоритму

ную память соответствующих процессоров и постоянно там находятся. Вертикальные полосы матрицы B , содержащие по m/p столбцов, рассылаются в локальную память соответствующих процессоров. В процессе вычислений каждый i -й процессор ($i = \overline{1, p}$) находит произведение элементов i -й горизонтальной полосы на элементы находящейся у него вертикальной полосы. В первом процессоре на этом же этапе в массиве C получается фрагмент, расположенный на одну полосу ниже и правее и т.д. На втором этапе вертикальные полосы массива B циклически сдвигаются на одну полосу, то есть перемещаются в локальную память соседних процессоров. После этого вычисляются элементы соответствующих фрагментов в массивах C каждого процессора. Затем снова выполняется сдвиг полос массива B и т.д. p раз. На последнем этапе производится сбор горизонтальных полос массива C из всех процессоров в корневой процессор.

Если матрицу B не разрезать на вертикальные полосы, то сдвиг полос не потребуется. Однако в этом случае понадобится в p раз больше локальной памяти для каждого процессора. Для организации вычислений, при которой схема разбиения матриц A и B представляет собой прямоугольную решетку, используются методы Фокса и Кэннона [9, 12]. Такой прием называют геометрическим способом распараллеливания. В данных методах матрицы A , B , C разбиваются, как показано в правой части рис. 12. Каждый из процессоров, которые можно объединить по топологии тора или иным удобным способом, отвечает за вычисление только своего прямоугольного фрагмента матрицы C . В процессе вычислений происходит обмен фрагментами в горизонтальном направлении для матрицы A и в вертикальном направлении для матрицы B .

Вычислительную трудоемкость рассмотренного блочного метода можно оценить как $T_p = 2n^3/p$.

Представление матриц в виде решетки, с одной стороны, увеличивает объем информации, пересылаемой между процессорами, однако, с другой стороны, обмен может быть совмещен с обработкой данных. Ленточный метод умножения позволяет снизить объем пересылаемых данных. Но разница в объемах пересылаемых данных уменьшается при увеличении числа процессоров в соответствии с выражением $(1 + 1/\sqrt{p})$. Кроме того, использование геометрического способа распараллеливания позволяет помещать прямоугольные блоки данных в процессорах с учетом близости их физического расположения. Это также позволяет ускорить пересылку данных. Разбиение матриц по принципу прямоугольной решетки дает возможность строить эффективные алгоритмы для

обработки разреженных матриц.

В качестве примеров решения систем линейных уравнений рассмотрим две модификации метода Гаусса и метод простой итерации [15]. Требуется решить систему линейных уравнений вида $AX = F$, где A – матрица коэффициентов системы, X – вектор неизвестных, F – вектор свободных членов.

Модификации метода Гаусса различаются способом распределения частей матрицы A и вектора F в локальной памяти узлов кластера.

В первой модификации матрица A и вектор F разрезаются горизонтальными полосами, количество которых определяется числом компьютеров в кластере: 0-я полоса помещается в локальную память 0-го компьютера, 1-я полоса – в память 1-го компьютера, ..., $(p-1)$ -я полоса – в память $(p-1)$ -го компьютера.

Прямой ход метода Гаусса, выполняющий преобразование матрицы A к треугольному виду, заключается в следующем. На первом шаге верхняя строка 0-го компьютера рассылается всем компьютерам с 1-го по $(p-1)$ -й. Затем во всех полосах с ее помощью исключается крайний слева столбец во всех строках всех полос, кроме нее самой (в этой строке коэффициент при неизвестном x_0 должен стать равным единице). Очевидно, что обработка полос в каждом компьютере будет выполняться одновременно.

На втором шаге следующая сверху строка из 0-го компьютера рассылается всем остальным узлам. Затем во всех строках всех узлов, кроме 0-го, исключается следующий слева столбец. В 0-м узле это выполняется для строк, расположенных ниже текущей, а в ней самой коэффициент при x_1 должен стать равным единице. На последующих шагах аналогичная рассылка оставшихся в матрице A строк и обработка выполняется для всех строк всех узлов кластера в направлении сверху вниз.

Вид матрицы A размера 16×16 , получившейся после преобразования к треугольному виду, приведен на рис. 13 (для 4-х узлов). Символом v на рисунке обозначены вещественные числа.

Обратный ход метода Гаусса выполняется аналогично, но начинается с нижней строки последнего компьютера.

Недостатком рассмотренной модификации метода Гаусса является то, что по мере увеличения номера рассылваемой строки верхние относительно нее компьютеры при прямом ходе метода и нижние при обратном начинают простаивать, пока остальные не завершат обработку своих строк. То есть вычислительная работа между компьютерами оказывается распределенной неравномерно.

Во второй модификации метода Гаусса матрица A и вектор F распределяются между узлами следующим образом: 0-я строка заносится в 0-й компьютер, 1-я – в 1-й, ..., $(p-1)$ -я – в $(p-1)$ -й. Затем p -я строка снова располагается под 0-ой в 0-ом компьютере, $(p+1)$ -я – в 1-ом и т.д.

Обработка строк при таком распределении выполняется следующим образом. Сначала текущей для рассылки и обработки является 0-я строка в 0-ом узле, затем 0-я в 1-ом узле и т.д. вплоть до 0-ой в $(p-1)$ -м узле. Затем текущими являются первые строки последовательно из 0-го, 1-го, ..., $(p-1)$ -го узлов и т.д. Вид матрицы A после прямого хода метода для данной модификации представлен на рис. 14.

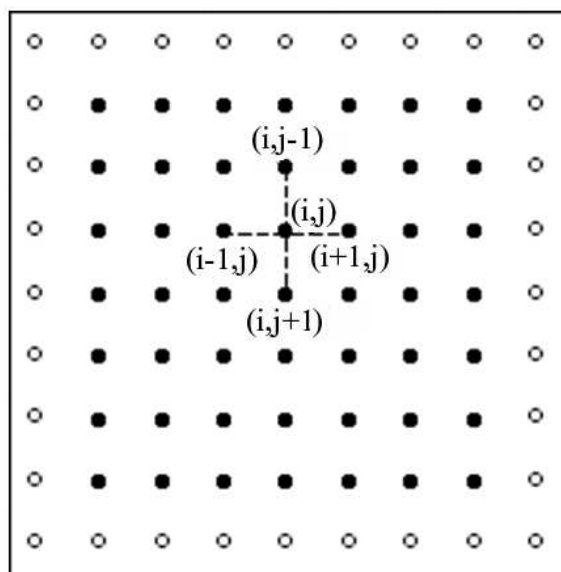


Рис. 15. Иллюстрация двумерной краевой задачи

В отличие от задачи вычисления произведения матриц, в данном случае разбиение области на p подобластей (по числу процессоров) выполняется с перекрытием соседних подобластей (рис. 16). В связи с тем, что при расчете значений точек крайнего (не граничного ряда) требуются точки, хранящиеся в локальной памяти соседнего процессора, добавляются граничные (заштрихованные) ряды, в которые через сеть передаются необходимые значения. Во внутренних точках каждой подобласти расчет выполняется по формуле

$$M_{ij} = 0.25(M_{i-1,j} + M_{i,j-1} + M_{i+1,j} + M_{i,j+1}) \quad (j = \overline{1, 1000}).$$

Номер горизонтального ряда принимает значения в диапазоне $i = 1 \dots 250$ для процессора P_0 , $i = 251 \dots 500$ для процессора P_1 и т.д. После выполнения очередного итерационного шага определяется максимальное значение погрешности во всех областях и, если необходимая точность не достигнута, выполняется обмен граничными строками и повторяется очередной шаг итерационного процесса.

При увеличении числа процессоров и, соответственно, подобластей время решения задачи с помощью рассмотренного метода уменьшается. Ускорение по закону Амдала определяется не только числом процессоров, но и величиной доли последовательных операций. Кроме того, при дальнейшем росте числа процессоров и увеличении времени на обмен между узлами кластера начинает проявляться действие сетевого закона Амдала. В результате этого рост ускорения не только прекращается, но и начинает уменьшаться.

5. Параллельные алгоритмы на графах

Многие реальные задачи прикладного характера, как и теоретические задачи, могут быть решены с использованием математических моделей различных про-

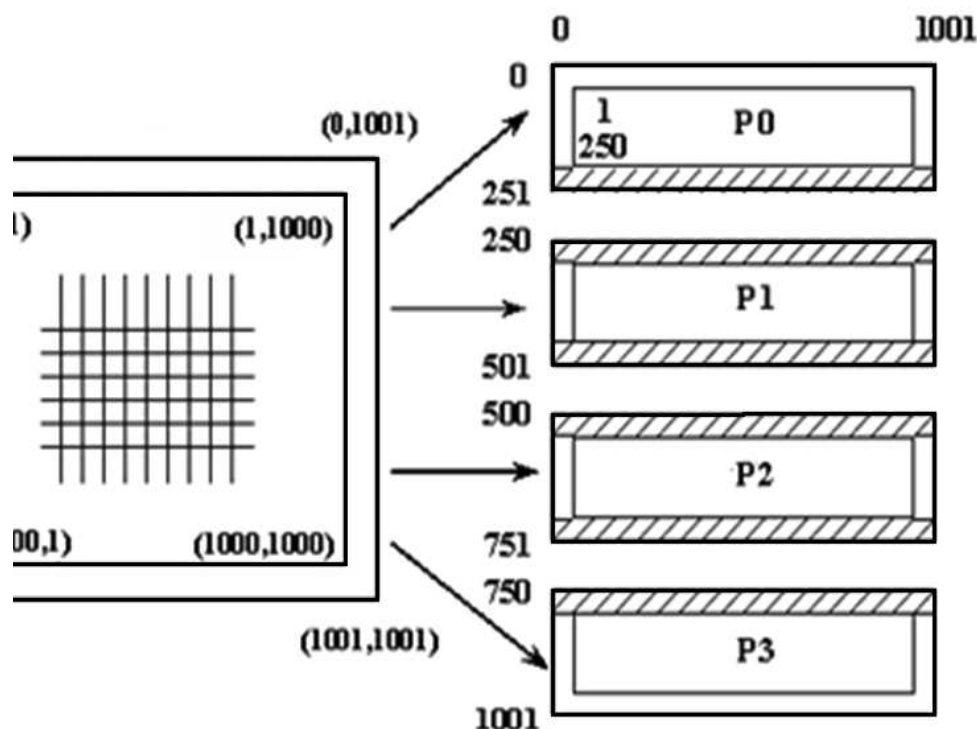


Рис. 16. Разбиение расчетной области на подобласти с перекрытием

цессов, явлений, систем, представленных в виде взвешенных графов. В качестве примеров можно привести линии телефонной связи, электрические сети, задачи составления расписаний, маршрутов самолетов. Для иллюстрации алгоритмов на графах рассмотрим два параллельных алгоритма из данной области: задачи поиска минимального охватывающего дерева и задачи поиска кратчайшего пути [13, 16].

В качестве практического применения алгоритма построения минимального охватывающего дерева можно назвать решение задачи построения сети персональных компьютеров с прокладыванием наименьшего количества соединительных линий связи.

Алгоритм Прима построения минимального охватывающего (остового) дерева относится к так называемым «жадным» алгоритмам. Подобные алгоритмы, базирясь на ограниченном наборе данных, выбирают лучшее решение, которое только возможно при наличии этого неполного набора данных. По данному алгоритму сначала выбирается произвольный начальный узел. Затем формируется начальная «кайма», включающая только узлы, непосредственно связанные с узлами, имеющимися в построенной части охватывающего дерева (то есть на первом шаге – с начальным узлом). На следующем этапе из каймы извлекается узел с наименьшим весом и присоединяется к остовому дереву (остову). После этого создается новая кайма за счет добавления к предыдущей кайме узлов, связанных с только что добавленным к остову. Опять извлекается узел с минимальным весом для добавления к остову и т. д., пока не закончатся все

вершины графа.

Параллельная реализация рассмотренного алгоритма для графа из N узлов в сети из p процессоров предполагает выделение для обработки каждому из процессоров по N/p узлов. Перед началом построения охватывающего дерева центральный процессор сети рассылает всем процессорам сведения о структуре графа и весах узлов, перечень выделенных конкретным процессорам узлов и номер начального узла остова.

На каждом проходе алгоритма процессоры определяют среди узлов своей группы номер узла, имеющего дугу с наименьшим весом, связанную с остовом. Эти номера узлов отправляются от процессоров сети центральному процессору, который выбирает из них узел графа с наименьшим весом, пополняет этим узлом остов и рассылает узлам сети сведения о новом остове. Затем повторяется очередной проход алгоритма до тех пор, пока не будут исчерпаны все узлы графа. Общее число проходов равно $N - 1$.

Сложность данного алгоритма оценивается величиной (N^2/p) при стоимости (N^2) . Это намного меньше, чем сложность (2^N) метода «грубой силы», то есть метода сплошного перебора. Оптимальным для рассмотренного алгоритма считается число процессоров, равное $N/\log N$.

Задача поиска кратчайшего пути имеет важное практическое значение в приложениях, для которых веса дуг означают время, стоимость, расстояние, затраты и т. п.

Алгоритм Дейкстры отличается от рассмотренного алгоритма тем, что минимизируется не расстояние от группы узлов каждого процессора до остова, а расстояние от них до корня остова. При числе процессоров $N/\log N$ и использовании топологии гиперкуба трудоемкость метода имеет порядок $N \log N$.

Еще один вариант решения рассмотренного алгоритма поиска кратчайшего пути [13] заключается в следующем. Строится матрица связности графа размера $N \times N$. На пересечении i -ой строки и j -го столбца записан 0, если $i = j$; символ ∞ , если узлы не связаны; значение веса, если узлы связаны. Данная матрица указывает веса между узлами, находящимися на расстоянии в одну дугу, и обозначается как A^1 . После этого создается матрица A^2 , которая хранит сумму весов дуг между узлами, удаленными друг от друга на одну или две дуги. Затем по этим двум матрицам уже можно получить матрицы A^3 или A^4 и т. д. до A^{N-1} . На каждом шаге получения из матрицы A^1 матриц $A^2, A^4 \dots$ выполняются операции поиска минимума из суммы элементов матриц (сумм весов дуг, соединяющих узлы).

Если в известном алгоритме умножения матриц трактовать сложение как операцию взятия минимума, а умножение как операцию сложения, то алгоритм поиска кратчайшего пути сводится к алгоритму умножения матриц. Следовательно, распараллеливание алгоритма решения задачи поиска кратчайшего пути может быть выполнено так же, как распараллеливание алгоритма умножения матриц рассмотренными выше методами (ленточным, Фокса, Кэннона).

Для многих из рассмотренных параллельных алгоритмов разработаны программы, исходные тексты которых можно найти в сети Интернет. Ряд программ, реализующих данные алгоритмы, входит в состав специализирован-

ных параллельных библиотек. В качестве примера можно назвать библиотеки ATLAS, Aztec, BlockSolve95, Distributed Parallelization at CWP, DOUG, GALOPPS, JOSTLE, NAMD, P-Sparslib, PIM, ParMETIS, PARPACK, PBLAS, PETSc, PGAPack, PLAPACK, ScaLAPACK, SPRNG [17]. Они содержат программный код, реализующий параллельные алгоритмы линейной алгебры, сеточных методов, методов Монте-Карло, генетических алгоритмов, рендеринга изображений, квантовой и молекулярной химии и другие.

Сложность и трудоемкость разработки распределенных программ с использованием интерфейса передачи сообщений долгое время тормозили широкое развитие многопроцессорных систем с распределенной памятью. Появление в этой области библиотек подпрограмм позволило избавить программистов от рутинной работы. Тем не менее наличие таких библиотек предполагает четкое понимание механизмов работы параллельных алгоритмов, заложенных в программные модули этих библиотек. С другой стороны, далеко не все известные методы используются в программных модулях, входящих в состав библиотек.

Таким образом, в работе выполнен обзор методов распараллеливания алгоритмов решения ряда типовых задач, относящихся к области вычислительной дискретной математики, преимущественно на основе распределенных вычислительных систем, не имеющих общей памяти.

Рассмотрены такие задачи, как определение значений числовых рядов (вычисление определенных интегралов; каскадные схемы суммирования: каскадно-последовательная, последовательно-каскадная, каскадно-каскадная; определение максимальных и минимальных значений среди больших последовательностей величин; определение значений длинных выражений, построенных на основе логических и других операций), комбинаторные задачи (методы сортировки: чет-нечетной перестановки, Шелла, расширение Брюса-Вагара для быстрой сортировки; методы параллельного двоичного поиска среди упорядоченных данных и последовательного поиска среди неупорядоченных величин), задачи линейной алгебры (умножение матриц: ленточный алгоритм, алгоритм Фокса, алгоритм Кэннона; две модификации метода Гаусса решения систем линейных уравнений; метод простой итерации решения СЛУ), краевая задача (метод простой итерации с использованием явного итерационного алгоритма), параллельные алгоритмы на графах («жадный» алгоритм Прима построения минимального охватывающего дерева, алгоритм Дейкстры поиска кратчайшего пути, алгоритм поиска кратчайшего пути сведением задачи к умножению матриц).

ЛИТЕРАТУРА

1. Корнеев В.В. Параллельные вычислительные системы. М.: Нолидж, 1999. 320 с.
2. Проект: ИММ УрО РАН / Моделирование.
– <http://www.parallel.ru/russia/map/data/project29.html>.
3. Список TOP500 наиболее мощных компьютеров мира. – <http://www.parallel.ru/>.
4. Фихтенгольц Г.М. Основы математического анализа (Т.2). СПб.: изд-во «Лань», 2001. 464 с.

-
5. Гергель В.П., Стронгин Р.Г. Основы параллельных вычислений для многопроцессорных вычислительных систем. Учебное пособие: Изд. 2-е, доп-е. Н.Новгород: изд-во ННГУ, 2003.
 6. Богачев К.Ю. Основы параллельного программирования. М.: БИНОМ. Лаборатория знаний, 2003. 342 с.
 7. Дацюк В.Н., Букатов А.А., Жегуло А.И. Методическое пособие по курсу «Многопроцессорные системы и параллельное программирование». Ростов-на-Дону: РГУ, 2000.
 8. Миллер Р., Боксер Л. Последовательные и параллельные алгоритмы: Общий подход. М: Бином. Лаборатория знаний, 2006. 406 с.
 9. Библиотека параллельных алгоритмов ParaLib. Руководство пользователя. Нижний Новгород: НГУ, 2004. 29 с.
 10. Ефимов С.С. Параллельные вычисления. Методические указания к лабораторным работам. Омск: Изд-во: Наследие. Диалог-Сибирь, 2006. 32 с.
 11. Кнут Д.Э. Искусство программирования. Т.3. Сортировка и поиск: Уч. пособие. М.: Изд. дом «Вильямс», 2000. 832 с.
 12. Kumar V., Grama A., Gupta A., Karypis G. Introduction to Parallel Computing. Second Edition. Addison Wesley, 2003. 856 с.
 13. Макконелл Д. Основы современных алгоритмов. М: Техносфера, 2004. 368 с.
 14. Букатов А. А., Дацюк В. Н., Жегуло А. И. Программирование многопроцессорных вычислительных систем. Ростов-на-Дону: Изд-во ООО «ЦВВР», 2003. 208 с.
 15. Корнеев В.Д. Параллельное программирование в MPI. Москва-Ижевск: Институт компьютерных исследований, 2003. 304 с.
 16. Кормен Т.Х., Лейзерсон С.Е., Ривест Р.Л. Алгоритмы: построение и анализ. М: БИНОМ: МЦНМО, 2004.
 17. Специализированные параллельные библиотеки.
– http://www.parallel.ru/tech/tech_dev/par_libs.html.

СТРАТЕГИЯ РАЗВИТИЯ ЭЛЕКТРОННОГО ДОКУМЕНТООБОРОТА ПРОМЫШЛЕННОГО ПРЕДПРИЯТИЯ

Н.С. Иванова

Рассматриваются основные принципы и решения в разработках систем компьютерного ведения документов

Результатом работы конструктора являются чертежи с большим числом изменений и согласований и, возможно, результаты некоторых расчетов. Насколько конструкция удачна, надежна и технологична, часто определяется лишь после ее изготовления. При традиционном способе разработки КД, основанному на двухмерном рисовании, возможны пропуски или ошибки в простановке размеров, неоднозначность прочтения чертежей, задание «невозможной» геометрии, наличие проникновения деталей в сборке, которые крайне сложно выявить. И дело не в конструкторских ошибках, основная проблема – в качестве проработки.

А также такие документы, как договоры, прайс-листы, приказы, деловая переписка, финансовая документация и другие, хранятся как в бумажном, так и в электронном виде (часто неупорядоченно). При этом практически неизбежно возникает дублирование информации, а также не обеспечивается поддержка разных версий документов и безопасность их хранения. То есть и при преобладании электронной формы хранения документов они не всегда используются эффективно. Потребность в эффективном управлении электронными документами и привела к созданию систем электронного документооборота.

Существуют оценки, что до 90 процентов времени сотрудников тратится на так называемую обеспечивающую функцию, а именно на поиск необходимых для работы документов. Эта проблема усугубляется при коллективном использовании документов, когда надо найти документы, созданные другим сотрудником. Соответственно, существует возможность практически на порядок повысить эффективность сотрудников; сократить расходы на копирование, канцелярские принадлежности и т. п.; сократить время на передачу документов между исполнителями. Кроме того, немаловажно отметить еще и фактор повышения безопасности при работе с документами – организация глубокой системы

защиты документов, в зависимости от операций и пользователей, позволяет защитить документы от несанкционированного доступа.

Кроме того, запись всех операций с документами позволяет восстановить всю историю действий с ними.

Настоящая стратегия развития направлена на реализацию концепции безбумажной технологии производства.

Элементами единого информационного поля предприятия должны быть:

- система автоматизированного проектирования различного уровня;
- автоматизированная система технологической подготовки производства;
- автоматизированная система документоучета;
- система электронного архива.

А также должно быть:

- лицензионное программное обеспечение;
- подготовка кадров;
- пересмотр стандартов предприятия.

1. Общие требования к системе документооборота

Первоначально рассмотрим общие требования к системе электронного документооборота.

1. Масштабируемость.

Желательно, чтобы система документооборота могла поддерживать как пять, так и более тысячи пользователей и была способна наращивать свою мощность соответствующего аппаратного обеспечения. Выполнение такого требования может быть обеспечено с помощью поддержки промышленных серверов баз данных производства таких компаний, как Sybase, Oracle, Informix и др., которые существуют практически на всех возможных программно-аппаратных платформах, тем самым обеспечивая самый широкий спектр производительности.

2. Модульность.

Система документооборота должна состоять из отдельных модулей, интегрированных между собой.

3. Открытость.

Система документооборота должна иметь открытые интерфейсы для возможной доработки и интеграции с другими системами.

2. Задачи, решаемые системами документооборота

Спектр задач и соответственно необходимая система автоматизации определяются стадией жизненного цикла документа, которую необходимо поддерживать.

Жизненный цикл документа.

1. Стадия разработки документа, которая может включать:

- собственно разработку содержания документа;
- оформление документа;
- утверждение документа.

В том случае если документ находится на стадии разработки, он считается неопубликованным, и права на документ определяются правами доступа конкретного пользователя.

2. Стадия опубликованного документа, которая может включать:

- активный доступ;
- архивный документ;
- краткосрочное хранение;
- долгосрочное хранение;
- уничтожение документа.

Когда документ переходит на вторую стадию, он считается опубликованным, тогда права на документ остаются только одни – доступ на чтение. В качестве примера опубликованного документа можно привести шаблон стандартного бланка предприятия. Кроме права доступа на чтение могут существовать права на перевод опубликованного документа в стадию разработки.

В зависимости от конкретной стадии жизненного цикла документа, с которым имеет дело архивная система, архивы подразделяются на следующие типы:

статические архивы документов (либо просто архивы) – системы, которые имеют дело только с опубликованными документами.

Динамические архивы документов (либо системы управления документами) – системы, имеющие дело как с опубликованными документами, так и с теми, которые находятся в разработке.

3. Принципы управления хранением

Данные находятся в сохранности только тогда, когда они хранятся более чем на одном носителе и более чем в одном месте.

Существуют гибкие методики, такие, как RAID, включая «зеркалирование» (mirroring), однако хотя они и повышают надежность системы, но не защищают ни от катастрофы, ни тем более от случайного удаления.

4. Организация поиска документов

Наряду с организацией хранения документов необходимо их быстро и эффективно искать. Если со скоростью поиска все понятно, то для пояснения понятия эффективности поиска рассмотрим модели поиска. Существует два подхода к поиску документов. Первый подход состоит в том, что в процессе поиска ищется документ, который точно существует в системе, и задача состоит в том, чтобы процесс поиска свелся к нахождению требуемого документа или документов. Этот подход применяется в 90 процентах всех случаев поиска. Второй подход состоит в том, что ищутся все документы по интересующему вопросу. Для этого подхода присущи такие понятия, как полнота - соответствие между найденными документами по данному запросу и действительному списку документов и шум - соответствие документов, отвечающих запросу, и документов, нерелевантных ему.

Существует два основных типа поиска. Атрибутивный - каждому документу присваивается набор определенных атрибутов, присваиваемых документу во время его размещения в архиве. В дальнейшем документ ищется на совпадение значений этих атрибутов полям запроса. К атрибутам документа можно отнести имя документа, время создания, имя автора, название подразделения, тип документа (факс, письмо, контракт, спецификация и т. д.). Очевидно, что список таких атрибутов может увеличиваться. Совокупность атрибутов документа называется карточкой документа. Эти поля могут заполняться произвольно или из predetermined справочников.

Второй тип поиска, носит название полнотекстовый – автоматически обрабатывается все содержание документа, и затем по любому слову, входящему в данный документ, можно найти сам документ. Соответственно, существует зависимость между типами поиска и подходами к поиску. Для поиска известного документа, очевидно, более применим атрибутивный поиск, для исследовательского поиска - полнотекстовый. Существует комбинация полнотекстового и атрибутивного поиска, когда атрибуты документа обрабатываются так же, как все содержание документа. Полнотекстовый поиск зависит от формата документа и языка, на котором он создан. Электронный документ любого формата необходимо предварительно преобразовывать в плоский текст для обработки системой полнотекстового поиска, следовательно, любая такая система должна содержать в своем составе конвертеры форматов. Зависимость от языка выражается в следующих факторах: поиск документа станет более полным, если в результате запроса будут найдены не только документы, которые точно соответствуют слову в запросе, но и те документы, в которых присутствуют различные словоформы слова в запросе. Данная технология носит название нормализации. Причем эффективность метода зависит от применяемого алгоритма нормализации. Для русского языка наиболее эффективен словарный метод - слово нормализуется на основе словарей, в которых содержатся словоформы. Кроме словарного метода, может применяться эвристический метод нормализации, когда слово может быть приведено к нормальному виду путем выполнения определенных правил, описывающих алгоритмику нормализации.

5. Проблемы внедрения систем электронного документооборота

При внедрении систем электронного документооборота приходится решать не только специфические проблемы, обусловленные многосторонней сложностью такого рода систем, но и иные проблемы, характерные для процессов коренной реорганизации деятельности предприятия. Известно, что переход на электронный документооборот можно с полным правом назвать именно коренным изменением организационного и административного устройства любого предприятия.

Проблемы внедрения системы электронного документооборота (СУД) приведены ниже:

1. Проблема информированности.

Чтобы руководство организации пришло к выводу о необходимости внедрения СУД, оно должно, как минимум, знать о существовании таких систем, для чего они предназначены и как осуществить их внедрение. Есть и другая проблема: с чего начать внедрение СУД?

2. Организационные проблемы.

На предприятии с течением времени уже сложилась определенная организационная структура и сформированы свои методы управления и контроля. Внедряемая СУД, в подавляющем большинстве случаев, на первых порах оказывается как бы «чужеродным телом» для коллектива предприятия. Это происходит потому, что хорошо построенная СУД является своего рода «лакомусовой бумажкой», и многие недостатки в функционально-структурном построении предприятия проявляются уже на первых этапах процесса внедрения СУД. Также если предприятие в значительной степени организационно не готово к внедрению СУД, то такое внедрение либо весьма затруднено, либо вовсе невозможно. Хорошей аналогией может послужить попытка установить более мощный двигатель на автомобиль со слабой ходовой частью. Результат известен заранее.

3. Психологические проблемы.

Как видится СУД большинству сотрудников предприятия? Если они обладают поверхностной информацией, то руководству предприятия СУД представляется как защита от всех неурядиц, то и дело возникающих из-за небрежного отношения к документам. А исполнители считают, что СУД – это что-то среднее между электронной почтой и привычным редактором. И только потом, по мере более детального ознакомления с системой, руководство вдруг с удивлением обнаруживает, что оно тоже должно будет работать на компьютере. Для немалого числа руководителей старой закалки это оказывается психологическим барьером. Им куда привычнее работать непосредственно с людьми: «вызвал на ковер», «дал накачку», увидел страх в глазах подчиненного – приходишь к мысли, что не зря занимаешь кресло. У исполнителей же часто возникает ощущение, что с внедрением СУД появился еще один начальник, который постоянно стоит за спиной. Действительно, ведь теперь совершенно точно можно узнать при желании, кто, что, когда и сколько делает. Раньше можно было сколько угодно вешать начальству «лапшу на уши», мол, искал такой-то и такой-то документ по всем этажам (хотя на самом деле играл в преферанс на компьютере). Вот и возникает у плохого начальства и у нерадивых работников психологический дискомфорт и полное неприятие СУД. Хорошо, если это выражается только в заявлении на увольнение. Чаще мы получаем стойких скрытых врагов, всячески сопротивляющихся такому нововведению, как СУД, которая воочию покажет их несостоятельность и бесполезность для предприятия.

4. Проблема кадров.

Внедрение СУД подразумевает, что все основные участники документооборота на предприятии должны уметь работать на компьютере. Но что делать, если основной костяк руководства предприятия получил образование 15 – 20 лет назад? По своему опыту и профессиональным навыкам они могут полностью соответствовать занимаемым должностям, но они никогда не обучались

и не работали на ПК. Отправлять их на учебу? Но, как правило, если пришли к мысли о необходимости внедрения СУД, то интенсивность труда на таком предприятии весьма высока. Это значит, что обучение сотрудников с отрывом от производства практически невозможно. А факультативное обучение может оказаться неэффективным и будет приводить лишь к повышенной усталости работников предприятия.

Хочется лишь подчеркнуть, что все эти проблемы должны приниматься во внимание.

6. Что нужно сделать для оптимального внедрения системы электронного документооборота

1. Обязательная аттестация всего задействованного персонала. Аттестацию провести в виде теста на 35 вопросов.
2. Обучение персонала, не прошедшего аттестацию.
3. Повторная аттестация после обучения.
4. Выбор и утверждение программного обеспечения.
5. Паспортизация автоматизированных рабочих мест.
6. Создание ЭЦП.
7. Автоматизация регистрации документов и заданий.
8. Обеспечение эффективного взаимодействия сотрудников в рамках работ по документам.
9. Разработка единой электронной библиотеки.
10. Осуществление мгновенного поиска информации.
11. Контроль выполнения работ, инициируемых документами и заданиями.
12. Создание динамического архива.
13. Создание статического архива.
14. Обеспечение разграничения прав доступа сотрудников к информации.
15. Создание форума предприятия для быстрого решения возникающих по проектированию вопросов.

ЛИТЕРАТУРА

1. Антопольский А.Б. Информационные ресурсы России // Научные и технические библиотеки. 2000. N.1.
2. ГОСТ 2.004-88. ЕСКД Общие требования к выполнению конструкторских и технологических документов на печатающих и графических устройствах вывода ЭВМ.
3. ГОСТ Р 51141-98. Делопроизводство и архивное дело. Термины и определения.
4. РД 1-000-99 CALS технологии. Терминологический словарь.

ОПТИМИЗАЦИЯ ЗАПРОСОВ К БАЗАМ ДАННЫХ

О.В. Пашинин

Рассматривается возможность вывода текущих запросов к базе данных из множества уже обработанных запросов

1. Введение

В настоящее время бурно развивается Интернет и Интранет, а также технологии с ними связанные. Корпоративные сети охватывают даже самые отдаленные производственные участки предприятия. Пропускная способность каналов связи до некоторых структурных подразделений недостаточна для корректной и адекватной работы части программных продуктов, построенных на технологии «клиент-сервер». В большинстве своем эти программные продукты так или иначе касаются баз данных. В данной ситуации чаще всего проблема решается либо за счет установки дополнительного сервера в удаленном структурном подразделении и пользователь работает именно с этим сервером, либо за счет увеличения пропускной способности канала передачи данных, что не всегда экономически и технологически оправдано.

Проблему можно решить и другим способом. Во-первых, вместо дополнительного сервера в удаленном структурном подразделении ставится рабочая станция (будем называть ее – буфер), которая хранит результаты последних запросов. Во-вторых, оптимизируется запрос к базе данных.

Вопросу оптимизации запросов посвящено множество статей и обзоров [1, 2, 6]. Можно выделить два основных метода оптимизации – статистический и алгебраический. Статистический метод базируется на системе оценок, статистике базы данных и допущения модели. Применение различных эвристик сужает пространство поиска, и выбирается оптимальный план выполнения запроса [7]. Алгебраический метод основан на применении к запросу операций реляционной алгебры и математической логики, благодаря чему на выходе получается эквивалентный канонический запрос [4, 8].

Используя алгоритм, предложенный в этой статье, сначала проверяется возможность вывода текущего запроса из предыдущих запросов, хранящихся в буфере и обработанных за данное время, и только в случае невозможности вывода запрос полностью или частично передается на сервер, находящийся на головном предприятии.

2. Определения и обозначения

Рассматриваемый в данной статье подход предполагает использование стандартного запроса $S = \pi_X(\sigma_F(R_1 \bowtie R_2 \bowtie \dots \bowtie R_n))$. Заполнение памяти буфера происходит следующим образом. Пользователь посылает запрос к головному серверу. Этот запрос идет к серверу не напрямую, а через буфер. Будем считать, что на буфере установлен программный модуль (ПМ), который делит поступающий запрос на две части: $S_1 = \sigma_F(R_1 \bowtie R_2 \bowtie \dots \bowtie R_n)$ (редуцированный запрос) и $S_2 = \pi_X(S_1)$. Запрос S_1 отправляется на головной сервер, а запрос S_2 выполняется средствами ПМ. В результате для каждого запроса S в буфере помимо результата выполнения самого запроса S хранится таблица, соответствующая запросу S_1 .

Также положим, что логическое выражение в запросе приведено к ДНФ. Для каждого запроса S алгоритм применяется последовательно к каждому дизъюнкту в логическом выражении F .

Например, $F = \{A_1 = x \vee (A_2 = y \wedge A_3 < z)\}$. Тогда алгоритм применяется сначала к $S_1^1 = \sigma_{A_1=x}(R)$, а затем к $S_1^2 = \sigma_{A_2=y \wedge A_3 < z}(R)$.

Если результатом работы алгоритма является возможность вывода текущего запроса из предыдущих полностью или частично, то к множеству таблиц на выходе алгоритма необходимо применить аппарат математической логики и реляционной алгебры.

Введем следующие обозначения:

S – запрос на момент времени t к базе данных (DB).

$F(S)$ – логическое выражение в запросе S .

$r = S(DB)$ – таблица (результат выполнения запроса).

$S' = \{S_1, S_2, \dots, S_k\}$ – множество запросов, находящихся в системном буфере на момент времени t .

$r' = \{r_1, r_2, \dots, r_k\}$ – множество таблиц, соответствующих запросам из множества S' .

Сначала рассмотрим случай, когда база данных содержит только одну таблицу $R(A_1, A_2, \dots, A_n)$. Далее обобщим рассуждения на случай нескольких таблиц.

Рассмотрим множество запросов S' и ограничим его редуцированными запросами. Т. е. вместо множества $\{\pi_X(\sigma_{F_1}(R)), \pi_Y(\sigma_{F_2}(R)), \dots, \pi_Z(\sigma_{F_k}(R))\}$, где X, Y, Z – подмножества множества атрибутов $\{A_1, A_2, \dots, A_n\}$, к которым применяется операция проекции, имеем дело с множеством $s' = \{\sigma_{F_1}(R), \sigma_{F_2}(R), \dots, \sigma_{F_k}(R)\}$.

Введем определения:

Определение 1. Пусть g – булев вектор, соответствующий запросу. Он формируется следующим образом: $g(j) = 1$, если в запросе S при выполнении операции селекции накладывались любые ограничения на j -ый атрибут (отрицание, сравнение, равенство с элементом из множества значений данного атрибута). Вектор g будем называть **характеристическим вектором** для запроса S .

Например, $R = (\text{№ чит. билета}, \text{назв. книги}, \text{назв. изд-ва}, \text{дата заказа}, \text{№ заказа})$. Запросу $\pi_{A_1}(\sigma_{A_2='Незайка' \wedge A_4='01.01.2000'}(R))$ будет соответствовать харак-

теристический вектор $g = (01010)$; запросу $\pi_{A1}(\sigma_{A1 < '100' \& A3 = 'Наука'}(R))$ будет соответствовать вектор $g = (10100)$.

Определение 2. Прямоугольную матрицу, составленную из коэффициентов характеристических векторов, назовем **характеристической матрицей**.

Например, $g_1 = (1010)$, $g_2 = (1100)$, $g_3 = (1010)$, $g_4 = (1000)$, $g_5 = (1000)$, тогда

$$G = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

3. Алгоритм проверки выводимости запроса S из множества S' в случае одной таблицы

Входные данные:

S – текущий запрос;

$F = F(S)$ – логическое выражение в запросе S ;

$g = g(S)$ – характеристический вектор запроса S ;

$S' = \{S_1, \dots, S_k\}$ – множество запросов в системном буфере;

$F' = \{F_1, \dots, F_k\}$ – множество логических выражений, соответствующих запросам в системном буфере, приведенные к ДНФ;

$g' = \{g_1, \dots, g_k\}$ – множество характеристических векторов, соответствующих запросам в системном буфере;

$r' = \{r_1, r_2, \dots, r_n\}$ – множество таблиц, находящихся в системном буфере на момент времени t ;

G' – характеристическая матрица множества характеристических векторов g' ;

G – вспомогательная матрица такой же размерности, как и G' , на начальном этапе она составлена из одних нулей. Область значений элементов $(0, 1, *)$.

Выходные данные:

S'' – множество запросов, из которых возможно вывести S .

Алгоритм:

III.1. Если существует F_i такое, что $F_i \equiv F$, то $S'' = \{S_i\}$. *Конец.*

Иначе, разбиваем F на дизъюнкты F^k и переходим на шаг 2 с очередным дизъюнктом.

III.2. Рассматриваем вектор g . Смотрим на каком месте стоит очередная единица. Пусть на месте p имеет место $g(p) = 1$ (т.е. на атрибут A_p накладывается условие в логическом выражении $F^k(S)$).

Если непройденных единиц нет, то переходим на шаг 5.

Иначе ищем среди векторов множества g' векторы $g_{t1}, \dots, g_{tm}$, такие, что $g_{ti}(p) = g(p) = 1$, $i = 1 \dots m$.

Если хотя бы один такой вектор есть, то переходим на шаг 3, иначе переходим к следующей единице.

III.3. Если строка матрицы G , соответствующая вектору g_{ti} не помечена, то помечаем ее и копируем в неё соответствующую строку из матрицы G' .

III.4. Проверяем справедливость $F_{ti}^k(p) \supseteq F(p)$, $i = 1 \dots m$.

Если условие выполняется, то для данного номера t_i в матрицу G в строку t_i , столбец p ставим звездочку и переходим на шаг 2.

III.5. Вычеркиваем нулевые строчки в матрице G , не помеченные на шаге 3, – тем самым мы ограничиваемся запросами, в которых при выполнении селекции не накладывались ограничения на атрибуты $A_{t1}, A_{t2}, \dots, A_{tm}$.

Вычеркиваем строчки, в которых стоят «1» там, где у вектора g стоят нули, – тем самым мы исключаем те запросы, в которых были ограничения на другие атрибуты, нежели в S .

Переходим на шаг 6.

III.6. Просматриваем полученную матрицу по строкам.

Проверяем для каждой i -ой строки, есть ли строка, состоящая только из $\{0, *\}$: $\forall p \ G[i, p] \neq 1$.

Если такие строки есть, то запрос S выводим и $S'' = \{S_i\}$. *Конец.*

4. Обобщение алгоритма проверки выводимости запроса S из множества S' в случае n таблиц

Пусть в нашей базе данных n таблиц $R_1, R_2, \dots, R_n$.

Тогда запрос S можно представить следующим образом:

$S = \pi_X(\sigma_F(R_{t1} \bowtie R_{t2} \bowtie \dots \bowtie R_{tk}))$, где $1 \leq k \leq n$.

Опять обозначим через s редуцированный запрос, соответствующий запросу S : $s = \sigma_F(R_{t1} \bowtie R_{t2} \bowtie \dots \bowtie R_{tk})$.

Определение 3. Пусть дана реляционная база данных DB ;

$R_1, R_2, \dots, R_n$ – таблицы DB ;

$S = \pi_X(\sigma_F(R_{t1} \bowtie R_{t2} \bowtie \dots \bowtie R_{tk}))$ – некоторый запрос к DB ;

$s = \sigma_F(R_{t1} \bowtie R_{t2} \bowtie \dots \bowtie R_{tk})$ – редуцированный запрос.

Тогда **расширенным характеристическим вектором** запроса S в DB будем называть булев вектор eg (enhanced – расширенный), формирующийся следующим образом: $eg[(n_1 + \dots + n_{ti-1}) + j] = 1$, если в запросе s накладывались ограничения на j -ый атрибут, $2 \leq t_i \leq k$, n_i – размерность i -ой таблицы.

Например, дана DB «библиотека» (R_1, R_2, R_3, R_4).

$R_1 = (\text{№ издательства, наименование изд-ва, адрес изд-ва})$;

$R_2 = (\text{№ книги, название книги, наименование изд-ва, ФИО автора})$;

$R_3 = (\text{ФИО читателя, адрес читателя, № чит. билета})$;

$R_4 = (\text{№ чит. билета, № книги, дата, № заказа})$.

Пронумеровав атрибуты, формально получим:

$R_1 = (A_1, A_2, A_3)$;

$R_2 = (A_4, A_5, A_2, A_6)$;

$R_3 = (A_7, A_8, A_9)$;

$$R_4 = (A_9, A_4, A_{10}, A_{11}).$$

Запросу $s = \sigma_{A_2='Мир' \& A_6='Иванов'}(R_1 \bowtie R_2)$ будет соответствовать вектор $eg = (01000100000)$.

Запросу $s = \sigma_{A_2='Питер' \& A_5='Война и мир' \& A_6='Л.Толстой' \& A_9='27'}(R_1 \bowtie R_2 \bowtie R_4)$ будет соответствовать вектор $eg = (01001100100)$.

Определение 4. Пусть дана реляционная база данных DB ;

$R_1, R_2, \dots, R_n$ – таблицы DB ;

$S = \pi_X(\sigma_F(R_{t1} \bowtie R_{t2} \bowtie \dots \bowtie R_{tk}))$ – некоторый запрос к DB ;

$s = \sigma_F(R_{t1} \bowtie R_{t2} \bowtie \dots \bowtie R_{tk})$ – редуцированный запрос.

Тогда **маской** запроса S в DB будем называть булев вектор mg размерности n , построенный следующим образом: $mg[i] = 1$, если в запросе s таблица $R_i \in (R_1, R_2, R_3, R_4)$, и $mg[i] = 0$ в противном случае.

Например, для DB из предыдущего примера:

запросу $s = \sigma_{A_2='Мир' \& A_6='Иванов'}(R_1 \bowtie R_2)$ будет соответствовать маска $mg = (1100)$.

Запросу $s = \sigma_{A_2='Питер' \& A_5='Война и мир' \& A_6='Л.Толстой' \& A_9='27'}(R_1 \bowtie R_2 \bowtie R_4)$ будет соответствовать маска $mg = (1101)$.

Добавим к алгоритму для одной таблицы 0-шаг:

$$s = \sigma_F(R_{t1} \bowtie R_{t2} \bowtie \dots \bowtie R_{tm}).$$

Ищем среди запросов множества

$$s' = \{s_1, s_2, \dots, s_k\} = \{\sigma_{F1}(R_{11} \bowtie \dots \bowtie R_{1p}), \dots, \sigma_{Fk}(R_{k1} \bowtie \dots \bowtie R_{kq})\}$$

те запросы, для которых выполняется следующее: маска этих запросов должна быть идентична маске запроса s .

Обозначим $R = R_{t1} \bowtie R_{t2} \bowtie \dots \bowtie R_{tm}$ и в алгоритм входим с этой таблицей R . Вместо вектора g в алгоритме используем расширенный вектор eg .

5. Пример работы алгоритма

Рассмотрим пример работы алгоритма с одной таблицей.

На входе имеем:

- Таблицу $R(A_1, A_2, A_3, A_4, A_5)$.
- $R = (\text{№ чит. билета, назв. книги, назв. изд-ва, дата заказа, № заказа})$.
- Текущий запрос S :

$$S = \pi_{A1}(\sigma_{A2='Незнайка' \& A4='01.03.2000'}(R)).$$

- Характеристический вектор g запроса S : $g = (01010)$.

- Логическое выражение $F = F(S)$:

$$F = (A_2 = \text{'Незнайка'} \& A_4 = \text{'01.01.2000'}).$$

- Множество запросов в системном буфере $S' = \{$

$$S_1 = \pi_{A1}(\sigma_{A2=\text{'Незнайка'} \& A4 > \text{'02.02.2000'}}(R));$$

$$S_2 = \pi_{A2}(\sigma_{A1=\text{'1710'}}(R));$$

$$S_3 = \pi_{A5}(\sigma_{A2=\text{'Мишкина каша'} \& A4=\text{'01.03.2000'}}(R));$$

$$S_4 = \pi_{A4}(\sigma_{A2=\text{'О, женщины...'} \& A3=\text{'Новый мир'}}(R));$$

$$S_5 = \pi_{A2}(\sigma_{A3=\text{'Просвещение'} \& A4=\text{'01.01.2000'}}(R));$$

$$S_6 = \pi_{A2}(\sigma_{A5=\text{'211'}}(R))\}.$$

- Множество характеристических векторов $g' = \{g_1 = (01010), g_2 = (10000), g_3 = (01010), g_4 = (01100), g_5 = (00110), g_6 = (00001)\}.$

- Матрица G' :

$$G' = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

- Множество логических выражений $F' = \{F_1, \dots, F_6\}$:

$$F' = \{$$

$$(A_2 = \text{'Незнайка'} \& A_4 > \text{'02.02.2000'});$$

$$(A_1 = \text{'1710'});$$

$$(A_2 = \text{'Мишкина каша'} \& A_4 = \text{'01.03.2000'});$$

$$(A_2 = \text{'О, женщины...'} \& A_3 = \text{'Новый мир'});$$

$$(A_3 = \text{'Просвещение'} \& A_4 = \text{'01.01.2000'});$$

$$(A_5 = \text{'211'})\}.$$

III.1. Нет таких F_i , что $F_i \equiv F$, $i = 1 \dots 6$. Переходим на ш. 2.

III.2. Очередная единица стоит на месте № 2. Находим векторы g_1, g_3, g_4 , у которых вторая координата равна «1». Переходим на ш. 3.

III.3. Имеем матрицу G :

$$G = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Переходим на ш. 4.

Ш.4.

$$F_1(2) = \{A_2 = \text{'Незнайка'}\} \supseteq F(2);$$

$$F_3(2) = \{A_2 = \text{'Мишкина каша'}\} \not\supseteq F(2);$$

$$F_4(2) = \{A_2 = \text{'О, женщины'}\} \not\supseteq F(2).$$

Имеем матрицу G :

$$G = \begin{pmatrix} 0 & * & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Переходим на шаг 2.

Ш.2. Очередная единица стоит на месте № 4 Находим векторы g_1, g_3, g_5 , у которых четвертая координата равна «1». Переходим на ш. 3.

Ш.3. Имеем матрицу G :

$$G = \begin{pmatrix} 0 & * & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Переходим на ш. 4.

Ш.4.

$$F_1(4) = \{u[4] > \text{'02.02.2000'}\} \supseteq F(4);$$

$$F_3(4) = \{u[4] = \text{'01.03.2000'}\} \supseteq F(4);$$

$$F_5(4) = \{u[4] = \text{'01.01.2000'}\} \not\supseteq F(4).$$

Имеем матрицу G :

$$G = \begin{pmatrix} 0 & * & 0 & * & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & * & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Переходим на ш. 2.

Ш.2. Больше непройденных единиц нет. Переходим на ш. 5.

Ш.5. Вычеркиваем нулевые строчки и строчки, в которых стоят единицы в столбцах, отличных от № 2 и № 4.

Имеем матрицу G :

$$G = \begin{pmatrix} 0 & * & 0 & * & 0 \\ 0 & 1 & 0 & * & 0 \end{pmatrix}$$

Переходим на ш. 6.

III.6. Получаем набор строк $\{1, 3\}$. Первая строка состоит только из $\{0, *\}$. Значит, из запроса № 1, находящегося в буфере, можно вывести текущий запрос.
Конец.

ЛИТЕРАТУРА

1. Кузнецов С.Д. Методы оптимизации выполнения запросов в реляционных СУБД / Тем. изд. «Итоги науки и техники. Вычислительные науки». Т.1. Москва, ВИНТИ, 1989, С.76-153
2. Чаудхари С. Методы оптимизации запросов в реляционных системах // СУБД. 1998. N.3. С.22-36.
3. Чери С., Готлоб Г. Логическое программирование и базы данных. Пер с англ. М.: Мир, 1992.
4. Latha S.Colby. A recursive algebra and query optimization for nested relations // ACM SIGMOD Record archive. 1989. V.18, N.2. P.273-283.
5. Yannis E. Ioannidis and Raghu Ramakrishnan. Containment of Conjunctive Queries: Beyond Relations as Sets // ACM Trans. Database Syst. 1995. V.20, N.3. P.288-324.
6. Zykin S.V. Relation queries execution under the estimators control // International conference ADBIS-95. Moscow. 1995. V.2 P.52-55.
7. Sun W., Yu C. Semantic Query Optimization for Tree and Chain Queries // IEEE Transactions on Knowledge and Data Engineering. 1994 V.6, N.1 P.136-151.
8. Maher M., Wang J. Optimizing Queries in Extended Relational Databases // Proceedings of the 11th International Conference on Database and Expert Systems Applications. 2000. September 04-08. P.386-396.

КОМБИНИРОВАНИЕ МЕТОДОВ ПРИ ПОСТРОЕНИИ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ

А.Н. Романов

Статья посвящена рассмотрению двух классов методов, используемых при построении модели интеллектуальной системы, а также возможности их совместного использования в рамках этой модели. Речь идёт о комбинировании логического подхода и методов, использующих нейронные сети для работы интеллектуальной системы.

Для построения интеллектуальных систем можно выбрать целый ряд методов, которые, в зависимости от поставленных перед системой целей, могут реализовать эти цели с тем или иным успехом. В нашей статье мы рассмотрим два таких метода. Первым будет метод логического программирования, обычно довольно успешно применяющийся при конструировании «экспертных систем», а вторым – метод, при котором в основу достижения «интеллектуальности» системы положено использование свойств нейронов и нейронных сетей.

Обычно эти два метода редко «конкурируют» друг с другом в рамках какой-либо специализированной задачи, так как случай «экспертных систем» обычно требует наличия некоторой базы знаний, которая и является основой для принятия решений. Конечно, эта база знаний может расширяться и изменяться – при наличии свойства обучаемости у системы, однако и эти процессы, как правило, подчинены определённым законам и в некоторой степени «заранее известны». Тот же фактор «заранее известности» в той или иной степени присутствует и при использовании нейронных сетей, хотя начальными данными здесь является не набор знаний, а набор определённого рода ситуаций, в которых известно «желаемое» поведение системы. По этому набору определяются начальные характеристики нейронов и связей между ними (если таковые имеются) и, соответственно, начальное состояние всей нейронной сети.

И тот и другой подходы имеют свои преимущества в определённых ситуациях. Нашей же целью, в частности, является выбор примера, когда было бы вполне разумно и оправданно применить оба упомянутых метода. Конечно, не в противовес друг другу, а в целях их взаимного дополнения. Зададимся целью выбора задачи, которая должна обладать свойством необходимого наличия начальной базы знаний (как и в экспертной системе), а также свойствами «обучаемости» и наличия ситуаций (условий) с неоднозначным их разрешением.

Присутствие базы знаний и необходимости выбора решения с использованием этой базы знаний оправдывает применение метода логического программирования. Свойство же неоднозначности решений позволяет нам применить методы нейронных сетей, когда подобные неоднозначности разрешаются с использованием накопленного за время работы системы опыта, а правила выбора какого-либо решения из предложенного «экспертной составляющей» набора решений могут меняться со временем. Таким образом, мы логически приходим к разрешению вопроса «обучаемости» подобной интеллектуальной системы – обучаемость также направлена по двум независимым путям. Первое направление затрагивает изменение/дополнение начальной базы знаний, которая используется для принятия решений экспертной составляющей нашей системы, а второе направление затрагивает подстройку свойств нейронов и их связей в нейронной сети.

В качестве примера, реализующего наши цели, попробуем рассмотреть задачу построения медицинской интеллектуально-экспертной системы. Однако, чтобы более ярко проявить именно идею двойной оценки ситуации различными методами, выберем для формирования рекомендаций методы восточной медицины. Дело в том, что именно методы восточной медицины, по нашему мнению, является довольно ярким примером использования идей дуальности, т.е. идей комплексных оценок ситуации с использованием иногда принципиально различных методов. Кроме того, интересна и сама оценка ситуации, которая приводит в выбору метода лечебного воздействия. Дело в том, что реальные врачи могут оценить одну и ту же ситуацию совершенно различными способами и, соответственно, выбрать различные способы воздействия, которые, тем не менее, приведут-таки к желаемому результату.

Исходя из вышесказанного, можно предположить, что и моделирование поведения врача (работающего именно с такой точки зрения), производимое нашей интеллектуальной системой, может привести к целому набору категорий возможных решений. Каждая из этих категорий будет по-своему верной и вполне жизнеспособной с точки зрения её практической реализации. Как раз в такой ситуации, по нашему мнению, вполне оправдано использовать оценку другим методом, дополнив, возможно, картину какими-либо ещё вспомогательными данными. В данном случае выбор набора категорий может осуществляться логическими методами, а определение наиболее подходящей (по оценке интеллектуальной системы) категории из уже определённого ранее набора, может быть осуществлено методами, использующими нейронные сети. То есть нашей целью будет в том числе и то, чтобы интеллектуальная составляющая экспертной системы, исходя из каких-либо второстепенных факторов, сама выбирала один из нескольких вариантов или, по крайней мере, давала некоторые рекомендации по этому поводу.

Как уже упоминалось ранее, для реализации первой экспертной составляющей можно предложить логический подход. Так, можно завести множество утверждений, из которых будут извлекаться рекомендации по излечиванию «элементарных» составляющих болезни. Наборы же правил будут сводить произвольные формулировки (заявления) о синдромах заболевания к элементар-

ным составляющим, заложенным в утверждениях, и тем самым находить возможные варианты излечения. Здесь, как мы уже упоминали, вариантов разрешения ситуации будет появляться несколько. Так, например, если считать, что уже выбрана категория воздействия и функциональная система, на которую будет оказываться воздействие, то возможно «распараллеливание» вариантов по набору зон воздействия, множества которых могут вообще не пересекаться между собой. Таким образом, в качестве результата работы экспертной составляющей может выдаваться набор видов лечения (в нашем примере это может быть: акупунктура, прогревание, акупрессура, травы и т.д.), а также набор возможных рецептов воздействия в рамках каждого вида. Заметим, что количество возможных вариантов ответов напрямую будет зависеть от количества методов, содержащихся в базе данных нашей экспертной системы. С точки зрения возможной эффективности результата, увеличение количества методов в базе будет увеличивать шанс выбрать более удачное решение, применимое к какому-либо конкретному случаю.

Вторую составляющую можно условно обозначить как интеллектуальную. Как уже было сказано, основной задачей интеллектуальной составляющей нашей экспертной системы будет являться выбор варианта (или нескольких вариантов) из множества решений, предложенных экспертной составляющей. Возможна также и такая реализация системы, когда осуществляется не выбор варианта, а некоторая оценка предложенных экспертной составляющей решений. Здесь, в качестве возможного варианта реализации поставленной задачи, можно предложить использование нейронной сети. Как уже отмечалось, оценка решений может проходить даже не столько по основным данным (в нашем случае – по синдрому), сколько по дополнительному набору информации. Итак, задавая на входе основные и дополнительные данные, на выходе мы хотим получить, во-первых, рекомендации по выбору категории лечебного воздействия и, во-вторых, рекомендации по выбору одного из возможных рецептов (которые уже составлены экспертной составляющей). В качестве инструмента по реализации интеллектуальной составляющей системы в рамках использования нейронных сетей можно рассмотреть вариант применения нейронной сети Хопфилда, где на вход сети будем подавать бинарные коды исходных характеристик (как основных, так и дополнительных), а на выходе сети будем получать коды упоминавшихся уже рекомендаций по выбору возможного подходящего варианта решения.

Напомним также, что не следует забывать и про проблему обучаемости нейронной сети, ведь именно обучаемость должна обеспечить в конце концов корректную работу интеллектуальной составляющей. Следует заметить, что теоретически возможен вариант, когда все необходимые знания будут заложены в интеллектуальную составляющую заранее, и тогда обучаемость не потребует, однако такой вариант представляется нам менее реально осуществимым, нежели вариант с обучаемостью. Тем не менее и его следует рассматривать как один из возможных.

Итак, мы рассмотрели идеи по построению интеллектуальной экспертной системы, которая, по нашему мнению, должна совмещать в себе логический

подход экспертной системы и оценку найденных экспертной составляющей решений с помощью работы нейронной сети. Эта идея дуальности о совмещении двух оценок одной ситуации была рассмотрена на примере применения к подходу экспертной медицинской оценки в том варианте, в котором сама идея комбинирования методов заложена изначально. Поэтому и идея принципиально разных оценок одной ситуации здесь, по нашему мнению, вполне уместна.

ЛИТЕРАТУРА

1. Братко И. Алгоритмы искусственного интеллекта на языке PROLOG. М.: Вильямс, 2004.
2. От моделей поведения к искусственному интеллекту. / Под ред. В.Г. Редько. М.: КомКнига, 2006.

ПРИМЕНЕНИЕ ТЕОРЕТИКО-АВТОМАТНЫХ МОДЕЛЕЙ В КРИПТОГРАФИИ

Н.Ф. Богаченко

В статье обсуждаются вопросы применения классических моделей теории автоматов в криптографии. Проведен обзор методов построения симметричных криптосистем и криптосистем с открытым ключом на основе последовательных автоматов.

В настоящее время «Теория автоматов» в классическом ее представлении (абстрактный и структурный синтез последовательных автоматов [1, 3]) – это достаточно изученная и методически обеспеченная дисциплина. Но вместе с тем представляет интерес применение такого математического объекта, как «автомат», в различных смежных отраслях. Остановимся на вопросах, связанных с возможностью использования автоматного подхода в криптографии.

Изложение материала не будет исчерпывающим. Целью работы является обсуждение нескольких криптосистем на идейном уровне и демонстрация принципиальной возможности использования в них последовательных автоматов (автоматов с памятью). Несмотря на то что предложенные криптосистемы являются скорее иллюстративными, они служат отправной точкой для дальнейших исследований.

1. Классические модели теории автоматов

Последовательным автоматом (или *автоматом с памятью*) называется система из пяти элементов

$$S = (A, X, Y, \delta, \lambda), \quad (1)$$

где $A = \{a_1, \dots, a_p\}$ – множество внутренних состояний; $X = \{x_1, \dots, x_n\}$ – множество входных сигналов (или входной алфавит); $Y = \{y_1, \dots, y_m\}$ – множество выходных сигналов (или выходной алфавит); $\delta : D_\delta \subseteq A \times X \rightarrow A$ – функция переходов; $\lambda : D_\lambda \subseteq A \times X \rightarrow Y$ – функция выходов.

Если необходимо, в систему (1) добавляется шестой элемент $a_1 \in A$ – начальное состояние автомата.

В дальнейшем будем рассматривать только детерминированные последовательные автоматы с конечными множествами A , X и Y [3].

Предполагается, что автомат функционирует в дискретные моменты времени $t = 1, 2, \dots$. В каждый момент времени t автомат, находясь в определенном состоянии a_t множества A , воспринимает на входном канале сигнал $x_t \in X$, выдает на выходном канале сигнал $y_t \in Y : y_t = \lambda(a_t, x_t)$ и переходит в новое состояние $a_{t+1} \in A : a_{t+1} = \delta(a_t, x_t)$.

Другими словами, автомат S , находясь в начальном состоянии a_1 , индуцирует *автоматное отображение* S_{a_1} , которое ставит в соответствие входному слову $\xi = x_1, \dots, x_l$ выходное слово $\psi = y_1, \dots, y_l$. При этом автомат проходит последовательность внутренних состояний $\alpha = a_1, \dots, a_{l+1}$.

В зависимости от области определения функции выходов различают две функциональные модели последовательных автоматов:

1. *Модель Мили* (или *автомат Мили*):

$$\begin{cases} a_{t+1} = \delta(a_t, x_t); \\ y_t = \lambda(a_t, x_t). \end{cases} \quad (2)$$

2. *Модель Мура* (или *автомат Мура*):

$$\begin{cases} a_{t+1} = \delta(a_t, x_t); \\ y_t = \lambda(a_t). \end{cases} \quad (3)$$

Для задания автоматов удобно использовать табличное представление (см. табл. 1, 2, 3).

Таблица 1. Таблица переходов автомата Мили

	a_1	$\dots$	a_p
x_1	$\delta(a_1, x_1)$	$\dots$	$\delta(a_p, x_1)$
$\dots$	$\dots$	$\dots$	$\dots$
x_n	$\delta(a_1, x_n)$	$\dots$	$\delta(a_p, x_n)$

Таблица 2. Таблица выходов автомата Мили

	a_1	$\dots$	a_p
x_1	$\lambda(a_1, x_1)$	$\dots$	$\lambda(a_p, x_1)$
$\dots$	$\dots$	$\dots$	$\dots$
x_n	$\lambda(a_1, x_n)$	$\dots$	$\lambda(a_p, x_n)$

Таблица 3. Таблица переходов автомата Мура

	$\lambda(a_1)$	$\dots$	$\lambda(a_p)$
	a_1	$\dots$	a_p
x_1	$\delta(a_1, x_1)$	$\dots$	$\delta(a_p, x_1)$
$\dots$	$\dots$	$\dots$	$\dots$
x_n	$\delta(a_1, x_n)$	$\dots$	$\delta(a_p, x_n)$

На множестве состояний автомата определим отношение эквивалентности: $a_i \sim a_j$, если $\forall \xi \in X^*$ (X^* – множество всевозможных слов над алфавитом X) выполняется равенство: $S_{a_i}(\xi) = S_{a_j}(\xi)$. Объединяя состояния автомата в классы эквивалентности, можно перейти к автомату, эквивалентному исходному, с минимально возможным числом состояний [3]. Далее будем рассматривать минимизированные последовательные автоматы.

2. Автоматная модель шифратора

Следуя [2], построим модель шифрующего автомата. Для этого необходимо дать несколько определений.

Шифром (или *криптосистемой*) называется система из пяти элементов

$$\Sigma = (K, \tilde{X}^*, \tilde{Y}^*, E, D), \quad (4)$$

где K – множество ключей; $\tilde{X}^* \subseteq X^*$ – множество открытых текстов; $\tilde{Y}^* \subseteq Y^*$ – множество закрытых текстов; $E_k : \xi \rightarrow \psi$ ($\xi \in \tilde{X}^*$, $\psi \in \tilde{Y}^*$) – правило зашифрования для $k \in K$; $D_k : \psi \rightarrow \xi$ – правило расшифрования для $k \in K$. При этом должны выполняться следующие свойства:

1. $\forall \xi \in \tilde{X}^*, \forall k \in K: D_k(E_k(\xi)) = \xi$ (отсюда следует, что отображение E_k инъективно).
2. $\tilde{Y}^* = \bigcup E_k(\xi)$, где объединение берется по всем $k \in K$ и $\xi \in \tilde{X}^*$.

Очевидно, что функции переходов и выходов автомата, моделирующего работу шифра, должны зависеть от ключа k . В качестве такого *шифрующего автомата* можно рассмотреть следующую систему:

$$S = (A \times K, X, Y, \delta, \lambda), \quad (5)$$

функциональная модель которой представляется автоматом Мили:

$$\begin{cases} \delta((a_t, k), x_t) = \delta_k(a_t, x_t) = (a_{t+1}, k); \\ \lambda((a_t, k), x_t) = \lambda_k(a_t, x_t) = y_t. \end{cases} \quad (6)$$

Перейдем теперь к вопросам расшифрования. С точки зрения автоматной модели для этого необходимо построить обратный автомат или, более строго, обратить автоматное отображение.

Автомат $S^{-1} = (\tilde{A}, Y, X, \tilde{\delta}, \tilde{\lambda})$ называется *обратным (слева)* к автомату $S = (A, X, Y, \delta, \lambda)$, если $\forall a \in A \exists \tilde{a} \in \tilde{A}$ такое, что $\forall \xi \in X^*$ выполняется равенство $S_{\tilde{a}}^{-1}(S_a(\xi)) = \xi$. Далее, пока не оговорено противное, будем рассматривать левую обратимость.

Очевидно, что для существования обратного автомата необходимо и достаточно потребовать инъективность автоматного отображения S_a для любого начального состояния a . Инъективность автоматного отображения достигается требованием инъективности функции выходов λ при фиксированном состоянии a , то есть частичной функции $\lambda_a : X \rightarrow Y$. Это, в свою очередь, означает, что $|\lambda_a^{-1}(y)| \leq 1$ [2].

Итак, для однозначности расшифрования на шифрующий автомат (5), (6) необходимо наложить условие инъективности частичной функции выходов λ_a .

Инъективность λ_a приводит к следующему алгоритму построения обратного автомата [2]:

1. Если $|\lambda_a^{-1}(y)| = 1$, то $\tilde{\lambda}(a, y) = \lambda_a^{-1}(y)$ и $\tilde{\delta}(a, y) = \delta(a, \lambda_a^{-1}(y))$.
2. В противном случае, когда $|\lambda_a^{-1}(y)| = 0$, $\tilde{\lambda}(a, y)$ – произвольный $x \in X$, $\tilde{\delta}(a, y)$ – произвольное $a \in A$.

Если мощности алфавитов X и Y совпадают, то обратный автомат строится однозначно. В общем случае должно выполняться $|X| \leq |Y|$. Если неравенство строгое, то обратный автомат является частичным.

Заметим, что для обратимости автомата необходимо и достаточно, чтобы в его табличном представлении в каждом столбце таблицы выходов все выходные сигналы были различны.

Для иллюстрации предложенных алгоритмов и моделей рассмотрим автомат Мили S , заданный таблицами 4, 5. Если произвольное значение сигнала или состояния обозначить прочерком, то автомат S^{-1} представляется таблицами 6, 7.

Таблица 4. Переходы автомата S

	a_1	a_2	a_3
x_1	a_3	a_1	a_1
x_2	a_1	a_3	a_2

Таблица 5. Выходы автомата S

	a_1	a_2	a_3
x_1	y_3	y_1	y_2
x_2	y_1	y_3	y_1

Таблица 6. Переходы автомата S^{-1}

	a_1	a_2	a_3
y_1	a_1	a_1	a_2
y_2	—	—	a_1
y_3	a_3	a_3	—

Таблица 7. Выходы автомата S^{-1}

	a_1	a_2	a_3
y_1	x_2	x_1	x_2
y_2	—	—	x_1
y_3	x_1	x_2	—

В простейшем случае ключом можно считать сам автомат. Пусть a_3 — начальное состояние. В качестве открытого текста рассмотрим последовательность $\xi = x_1x_1x_2x_1x_1$. Тогда закрытый текст $\psi = S_{a_3}(\xi)$ очевидным образом определяется по таблицам 4, 5:

$$\begin{array}{lcl} \xi = & x_1 & \left| \begin{array}{c|c|c|c|c} x_1 & x_2 & x_1 & x_1 & \end{array} \right| \\ \alpha = & a_3 & \left| \begin{array}{c|c|c|c|c} a_1 & a_3 & a_2 & a_1 & \end{array} \right| a_3 \\ \psi = & y_2 & \left| \begin{array}{c|c|c|c|c} y_3 & y_1 & y_1 & y_3 & \end{array} \right| \end{array}$$

Процесс расшифрования проводится по таблицам 6, 7:

$$\begin{array}{lcl} \psi = & y_2 & \left| \begin{array}{c|c|c|c|c} y_3 & y_1 & y_1 & y_3 & \end{array} \right| \\ \alpha = & a_3 & \left| \begin{array}{c|c|c|c|c} a_1 & a_3 & a_2 & a_1 & \end{array} \right| a_3 \\ \xi = & x_1 & \left| \begin{array}{c|c|c|c|c} x_1 & x_2 & x_1 & x_1 & \end{array} \right| \end{array}$$

Далее сформулируем ряд замечаний.

Одним из преимуществ представленного «автоматного» способа шифрования является то, что одинаковые фрагменты входной последовательности в общем случае шифруются различными блоками. Это видно даже на нашем примере: подпоследовательность x_1x_1 встречается во входном слове дважды, и ей соответствуют подпоследовательности y_2y_3 и y_1y_3 закрытого текста.

Если потребовать, чтобы частичная функция λ_x также была инъективной, то таблица выходов автомата (без строки состояний и столбца входных сигналов) будет представлять собой латинский прямоугольник – элементы в линиях (как в столбцах, так и в строках) не повторяются. Существующие оценки числа латинских прямоугольников [9] могут служить основой для анализа криптоустойчивости алгоритмов «автоматного» шифрования.

Следует отметить, что итогом теоретико-автоматной модели шифра является принципиальная возможность построения криптосистемы в виде двух последовательных автоматов (шифратора и дешифратора), блоки памяти которых совпадают [2].

3. Криптосистемы с открытым ключом, основанные на последовательных автоматах

Идея использования автоматой модели в криптографии с открытым ключом не нова. Так, в Китае с 80-х годов прошлого века ведутся работы в области создания асимметричных криптосистем, основанных на конечных автоматах [11]. Здесь следует остановиться на терминологии. *Конечный автомат* – это модель распознавателя, являющаяся эквивалентом автоматной грамматики относительно порождаемых языков [4, 10]. С точки зрения классической теории автоматов, в частности последовательных абстрактных моделей, конечный автомат – это автомат Мура, в общем случае частичный, выходные сигналы которого заданы в алфавите $\{0, 1\}$ [4].

В работе [8] предложена криптосистема с открытым ключом, основным элементом которой является последовательный автомат. Как уже отмечалось, произвольный последовательный автомат S реализует автоматное отображение $S(\xi) = \psi$. При этом и входное, и выходное слова имеют одинаковую длину: $|\xi| = |\psi|$.

Автомат S^{-1} с входным алфавитом Y и выходным алфавитом X , переводящий выходное слово ψ обратно во входное ξ , – это, как и ранее, обратный автомат, но теперь нам понадобится обратимость справа.

Автомат $S^{-1} = (\tilde{A}, Y, X, \tilde{\delta}, \tilde{\lambda})$ называется *обратным (справа)* к автомату $S = (A, X, Y, \delta, \lambda)$, если $\forall a \in A \exists \tilde{a} \in \tilde{A}$ такое, что $\forall \psi \in Y^*$ выполняется равенство $S_a(S_{\tilde{a}}^{-1}(\psi)) = \psi$. Отметим, что для доказательства правой обратимости необходимо и достаточно потребовать, чтобы автоматное отображение S_a было сюръективным для любого начального состояния a , или предъявить эквивалентное требование – сюръективность частичной функции выходов λ_a [2]. В табличном представлении автомата для правой обратимости необходимо и достаточно, чтобы в каждом столбце таблицы выходов каждый выходной сигнал встречался как минимум один раз.

Алгоритм построения правого обратного автомата достаточно прост [2]: $\tilde{\lambda}(a, y) = x$ и $\tilde{\delta}(a, y) = \delta(a, x)$, где x – произвольный элемент из множества $\lambda_a^{-1}(y)$, которое в силу сюръективности не пусто.

Определим теперь *инверсию с задержкой* z , где z – некоторое натуральное число. Пусть $\psi = S(\xi)$. Инверсия с задержкой S_z^{-1} , получая на входе слово $\psi\mu$,

где μ – произвольное слово длины z , порождает на выходе слово $\nu\xi$, где ν – также слово длины z .

Представленная в [8] криптосистема с открытым ключом строится следующим образом. В качестве ключа зашифрования открываются z и S_z^{-1} , а S хранится как секретный ключ расшифрования. Зашифрование исходного текста ψ происходит с помощью выбора произвольного слова μ длины z и применения автомата S_z^{-1} к слову $\psi\mu$. При расшифровании криптотекста легальный получатель игнорирует первые z букв и применяет автомат S к оставшемуся слову. Отметим, что в данном методе используется правая инверсия с задержкой.

Обсуждаемый в [6] асимметричный алгоритм шифрования также основан на паре автоматов S и S_z^{-1} . На шифрующий автомат S накладывается требование отсутствия инъективности частичной функции выходов λ_a (тогда автомат S необратим слева). При дополнительных условиях возможно построение левой инверсии с задержкой S_z^{-1} , что в общем случае является трудной задачей. Знание особенностей структуры автомата S дает легкий алгоритм получения S_z^{-1} (секретный ключ).

4. Шифрование с помощью «раскрашивания» на автоматах Мура

Достаточно много криптосистем используют подход, который может быть сформулирован как *шифрование с помощью «раскрашивания»* [8]. «Краска» связывается с каждой буквой исходного текста. Для простоты будем полагать, что исходные тексты являются двоичными последовательностями. Тогда нам необходимы только две краски – белая (бит «0») и черная (бит «1»).

Ключ зашифрования должен давать метод, который порождает произвольно много элементов, раскрашенных белой краской, и произвольно много элементов, раскрашенных черной краской. Примером таких элементов могут являться слова некоторого фиксированного алфавита. Биты шифруются как слова, раскрашенные соответствующим образом. Для различных появлений бита «0» (соответственно «1») должны выбираться различные слова, раскрашенные белой (соответственно черной) краской.

Для генерации «белых» и «черных» слов представляется возможным использовать автоматную модель. Пусть $S = (A, X, Y, \delta, \lambda, a_{н.с.})$ – автомат Мура, в котором выходной алфавит является двоичным: $Y = \{0, 1\}$ и $a_{н.с.}$ – начальное состояние.

Процесс зашифрования открытого двоичного текста заключается в следующем. Текущий бит i ($i \in \{0, 1\}$) шифруется на автомате S как произвольное входное слово ξ алфавита X с тем ограничением, что последний символ выходного слова $\psi = S(\xi)$ равен « i ». Следует отметить, что в автомате Мура, в отличие от модели Мили, выходное слово строится со сдвигом на один такт, то есть не учитывается выходной сигнал, приписанный начальному состоянию.

При расшифровании текущее слово ξ криптотекста подается на вход автомата S : последний символ выходного слова $\psi = S(\xi)$ и будет являться искомой

«краской». Однозначность расшифрования очевидна. Естественно, что сам автомат S необходимо держать в секрете.

Таблица 8. Переходы автомата Мура S

	1	0	1	0	1
	a_1	a_2	a_3	a_4	a_5
a	a_4	a_4	a_1	a_2	a_2
b	a_1	a_1	a_5	a_3	a_1
c	a_2	a_1	a_3	a_5	a_4

Рассмотрим предложенный способ шифрования на примере. Пусть автомат Мура S задан таблицей 8 и в качестве начального состояния выбрано состояние a_1 . Зашифруем последовательность «00101». Одним из вариантов закрытого текста является последовательность слов: « $a\ bc\ b\ c\ ab$ ». Расшифрование продемонстрируем на примере слова « ab »:

$$\begin{array}{lcl} \xi = & a & b \\ \alpha = & a_1 & a_4 \quad a_3 \\ \psi = & & 0 \quad 1 \end{array}$$

Так как последний символ выходного слова равен «1», то и соответствующий бит открытого текста – это единица.

Очевидно, что все криптосистемы, основанные на шифровании с помощью раскрашивания, имеют тенденцию к недопустимо большому росту длины криптотекста, так как чем больше различных слов, пригодных для шифрования одного бита (а следовательно, чем больше длина каждого слова), тем более стойким является криптоалгоритм.

Этот недостаток не столь существенен, если подобные криптосистемы использовать не для шифрования сообщений, а для шифрования ключей некоторого симметричного алгоритма, который в дальнейшем и будет применяться для передачи сообщений [11].

По этой же причине шифрование с помощью раскрашивания наиболее широко используется в асимметричных криптосистемах. Открытый ключ зашифрования – это метод выбора слов, сопоставимых каждому биту. В идеальной ситуации задача определения «краски» для заданного слова является трудно-решаемой, в то время как знание секретного ключа позволяет легко решать задачу расшифрования [8].

Рассмотренный алгоритм шифрования также может быть преобразован в криптосистему с открытым ключом. Как уже отмечалось, автомат Мура с двоичным выходным алфавитом является конечным автоматом.

Пусть дан автомат Мура $S^0 = (A, X, Y = \{0, 1\}, \delta, \lambda^0, a_{\text{н.с.}})$. Это конечный автомат, определяющий автоматный язык $L(S^0)$ – множество таких слов входного алфавита X , что последний символ соответствующих выходных слов равен «1». Построим автомат $S^1 = (A, X, Y, \delta, \lambda^1, a_{\text{н.с.}})$ по правилу: $\lambda^1(a, x) = \neg \lambda^0(a, x)$. Этот автомат задает автоматный язык $L(S^1)$.

Согласно [8], выберем две грамматики Γ^0 и Γ^1 с одним и тем же терминальным алфавитом X так, чтобы языки $L(\Gamma^0)$ и $L(\Gamma^1)$, порождаемые этими грамматиками, удовлетворяли условию: $L(\Gamma^i) \cap L(S^i) \neq \emptyset$ ($i \in \{0, 1\}$).

Построим две новые грамматики $\tilde{\Gamma}^0$ и $\tilde{\Gamma}^1$ такие, что $L(\tilde{\Gamma}^i) = L(\Gamma^i) \cap L(S^i)$ ($i \in \{0, 1\}$). Построение пересечения языков – стандартное. Отметим лишь, что в общем случае языки не замкнуты относительно операции пересечения, то есть тип результирующего языка в иерархии Хомского может быть произвольным [10].

Пара $(\tilde{\Gamma}^0, \tilde{\Gamma}^1)$ открывается в качестве ключа зашифрования – бит i шифруется как произвольное слово в $L(\tilde{\Gamma}^i)$ ($i \in \{0, 1\}$). Автоматы S^0 и S^1 хранятся в качестве секретного ключа.

Перехватчик для расшифрования должен определить принадлежность каждого слова криптотекста языку $L(\tilde{\Gamma}^i)$ ($i \in \{0, 1\}$), то есть решить задачу распознавания, являющуюся трудной для произвольного контекстного языка, а для языка типа 0 – в общем случае неразрешимой.

Легальный получатель решает легкую задачу принадлежности слова криптотекста одному из автоматных языков $L(S^0)$ или $L(S^1)$.

В силу построения автоматов S^0 и S^1 , языки $L(S^0)$ и $L(S^1)$ не пересекаются (они являются дополнениями друг друга) [8, 10]. Тем самым расшифрование всегда будет однозначным.

Рассмотренные в статье криптосистемы не являются единственно возможной сферой применения «автоматного» подхода в криптографии.

Автоматные модели используются в формальном анализе криптографических протоколов. Существующие алгоритмы основаны на следующем подходе. Если предположить, что злоумышленник может удалять сообщения из канала связи и помещать в канал связи созданные им сообщения, то любое сообщение, посланное легальным пользователем, можно рассматривать как сообщение, отправленное злоумышленнику, и любое сообщение, принятое легальным пользователем, как сообщение, полученное от злоумышленника. Тогда систему можно интерпретировать как автомат, используемый злоумышленником для генерации слов [5].

Представляет интерес идея построения однонаправленных хэш-функций на основе последовательных автоматов. Основное свойство однонаправленной функции – простота вычисления и сложность (практическая невозможность) инвертирования. Тогда в качестве такой функции можно рассмотреть необратимое (или трудно обратимое) автоматное отображение. Требования на хэш-функцию накладывают ограничения и на «однонаправленный» автомат S : трудно подобрать два различных входных слова ξ_1 и ξ_2 таких, что $S(\xi_1) = S(\xi_2)$, а изменение одного символа во входном слове ξ ведет к изменению в среднем половины символов выходного слова $\psi = S(\xi)$.

В работах [6, 7] рассматриваются автоматные однонаправленные отображения, для которых сложность построения обратного автомата приближается к экспоненциальной зависимости.

В заключение отметим, что для представленных в обзоре криптосистем не обсуждались такие вопросы, как оценка сложности, область применения, сопоставимость с другими шифрами. Важно было показать принципиальную возможность приложения классических теоретико-автоматных моделей к задачам криптографии.

ЛИТЕРАТУРА

1. Баранов С.И. Синтез микропрограммных автоматов. Ленинград: Энергия, 1974. 216 с.
2. Белов Е.Б., Зубов А.Ю., Погорелов Б.А., Проскурин Г.В., Черемушкин А.В., Шанкин Г.П., Шурупов А.Н. Криптографические методы защиты информации. Учебно-методическое пособие для курсов повышения квалификации преподавателей вузов учебно-методического объединения по образованию в области информационной безопасности. Часть 2. Москва. 2003.
3. Богаченко Н.Ф., Файзуллин Р.Т. Синтез дискретных автоматов. Учебно-методическое пособие. Омск: Издательство Наследие. Диалог-Сибирь, 2006. 150 с.
4. Богаченко Н.Ф., Файзуллин Р.Т. Автоматы, грамматики, алгоритмы. Учебно-методическое пособие. Омск: Издательство Наследие. Диалог-Сибирь, 2006. 140 с.
5. Давыдов А.Н. Формальный анализ криптографических протоколов: методы, основанные на моделях конечных автоматов // Труды научно-технической конференции «Безопасность информационных технологий». Пенза. 2005. – <http://beda.stup.ac.ru/rv-conf/v06/009/index.html>.
6. Копыленко В.М. Криптографическая конечно-автоматная модель. – http://zhurnal.lib.ru/k/kopylenko_w_m/afr2.shtml.
7. Копыленко В.М. Однонаправленная функция с «секретом» (trap-door function) на базе КАМСИ (Конечно-автоматная модель, сохраняющая информацию). – http://zhurnal.lib.ru/k/kopylenko_w_m/af_7.shtml.
8. Саломеа А. Криптография с открытым ключом. М.: Мир, 1995.
9. Холл М. Комбинаторика. М.: Мир, 1970.
10. Шамашов М.А. Теория формальных языков. Грамматики и автоматы. Учебное пособие. Самара: Университет Наяновой, 1996.
11. Шнайер Б. Прикладная криптография. Протоколы, алгоритмы, исходные тексты на языке Си. М.: ТРИУМФ, 2003.

ПРОВЕРКА ПАРОЛЕЙ НА НАДЕЖНОСТЬ С ПОМОЩЬЮ МОДИФИЦИРОВАННОГО ФИЛЬТРА БЛУМА

Д.Н. Лавров, Е.А. Первушин

В статье представлены алгоритмы проверки паролей на надежность, а также генерации произносимых и легконабираемых на клавиатуре паролей. При реализации алгоритмов использовался модифицированный алгоритм Блума.

1. Введение

Обычно проблема удостоверения подлинности пользователя решается с помощью паролей. Слабость этого метода заключается в том, что ненадежный пароль может быть легко угадан. Процедура угадывания состоит в переборе по словарю наиболее вероятных паролей.

В основном пароль – это компромисс между надежностью и удобством для пользователя. Если есть возможность применить программу, хранящую пароли в зашифрованном виде, то, воспользовавшись генератором случайных паролей, можно устанавливать длинные надежные пароли. Но в противном случае, когда пользователь вручную вводит пароль, на выбор влияет запоминаемость: пользователь ставит пароль, который сможет запомнить и быстро набрать. При установке пароля рекомендуется осуществить проверку на надежность: первая проверка — есть ли такой пароль в словаре.

Пользователь может применить генератор, который по некоторым правилам создает произносимые пароли (произносимые пароли легче запоминаются). Вскрытие пароля может вестись по словарю, каждое слово которого удовлетворяет этим правилам. Можно вести перебор в порядке убывания вероятности их генерации, если генерируемые пароли неравномерно распределены. Здесь встает вопрос о размере пространства генерируемых паролей. Если размер относительно мал, то пароль, полученный генератором, следует считать ненадежным. В этом случае в проверку на надежность следует включить тест на соответствие пароля правилам, по которым генератор его построил.

Прямой поиск в словаре — неэффективная по времени и занимаемому месту операция. Спаффорд предложил использовать фильтр Блума [1] для отбраковки паролей из словаря [2]. Фильтр гарантированно отбрасывает пароли из словаря, но может отбросить и пароли, не включенные в словарь. Фильтр состоит из фиксированного количества хэш-функций и битового массива, изначально заполненного нулями. Добавление пароля состоит в следующем: каждая хэш-функция переводит слово в индекс в массиве, далее биты по этим индексам устанавливаются в единицу. При запросе таким же образом вычисляются индексы. Проверяются биты, находящиеся в таблице по этим индексам. Если среди них хотя бы один ноль, то это слово не было добавлено, иначе пароль отбрасывается, но есть вероятность отбросить ранее не добавленное слово. Эта вероятность зависит от размера массива, количества функций и количества добавленных слов и может быть сколь угодно малой. Скорость запроса не зависит от количества добавленных слов.

Модифицируем фильтр Блума следующим образом: алгоритм добавления:

Введем параметры: $n > 0$, $0 < s \leq n$ — целые числа, например $n = 3$, $s = 1$. При добавлении слова разобьем его на n -граммы — слова длины n . Каждая следующая n -грамма отстоит от предыдущей на s символов. Пусть L — длина слова. Тогда количество n -грамм равно

$$\left\lceil \frac{(L - n)}{s} \right\rceil + 1,$$

где округление происходит до большего целого. Возможно, последняя n -грамма будет содержать меньше, чем n , символов. Далее полученные n -граммы заносятся в фильтр.

2. Алгоритм проверки

Для проверки слова на принадлежность словарю требуется проверить все n -граммы, полученные из проверяемого слова. Как и прежде, возможны коллизии, то есть отбрасывание ранее не добавленных n -грамм. Если все n -граммы отброшены, то слово будет отброшено. Если хотя бы одна из n -грамм принята, то такое слово не добавлялось в фильтр.

Рассматривая n -граммы как правила, можно так же генерировать произносимые пароли. Пусть ячейка в таблице представляет целое число. При добавлении n -граммы, как и раньше, вычисляются значения хэш-функций, которые являются индексами в таблице. Значения чисел, находящихся в таблице по этим индексам, увеличиваются на единицу. Чтобы посчитать, сколько раз была добавлена n -грамма, необходимо вычислить хэш-функции от этой n -граммы и найти минимум среди чисел, находящихся в таблице по этим индексам (при условии, что удаления из таблицы не происходят, то есть хранимые числа не уменьшаются). Результат, возможно, не будет верным из-за коллизий или переполнения ячеек. Далее таким же образом добавляется эта n -грамма, укороченная до $n - s$ символов. Таким образом, после добавления словаря для последо-

вательности $n - s$ символов $a_1 \dots a_{n-s}$ и s символов $b_1 \dots b_s$ можно вычислить вероятность того, что после $a_1 \dots a_{n-s}$ следует $b_1 \dots b_s$.

Для хранения частот можно было использовать другую структуру данных, например n -мерную матрицу, количество элементов которой равно длине алфавита в степени n . Но преимущество предложенного метода в том, что он не зависит от алфавита, а скорость запроса по-прежнему не зависит от количества добавленных слов.

Пусть N – длина массива, K – количество хэш-функций, после добавления D элементов вероятность того, что K случайным образом выбранных ячеек не пусты, равна

$$P = \left(1 - \left(1 - \frac{1}{N} \right)^{KD} \right)^K.$$

При фиксированных D и N значение P минимизируется при $K = \ln 2 \cdot N/D$ [3]. Вероятность, что хотя бы одна ячейка из K случайно выбранных ячеек пуста, равна $1 - P$. Пусть из всего множества n -грамм G добавлено G_1 . Обозначим $G_2 = G/G_1$. При добавлении новой n -граммы ее значение будет верным, если хотя бы одна из K ячеек, соответствующих этой n -грамме, хранит ноль – это произойдет в $(1 - P) * G_1$ случаях из G_1 . При запросе n -грамм из G_2 неверный результат будет возвращен в $G_2 \cdot P$ случаях, верный – $G_2 - P \cdot G_2$. Таким образом, доля неправильных значений при запросе из G равна $1 - \frac{(1-P)G_1 + G_2 - P \cdot G_2}{G_1 + G_2} = P$. В формулу $P = (1 - (1 - 1/N)^{KD})^K \approx (1 - e^{-KD/N})^K$ подставим $K = \ln 2 \cdot N/D$. Получим:

$$(1/2) \ln 2 \cdot N/D = P \ln(1/2) \ln 2 \cdot N/D = \ln PD/N = -\ln 22 / \ln P,$$

при $P = 0,01 = 1\% D/N \approx 0,1043$. То есть хранение 10% от количества всех n -грамм с учетом того, что частоты 99 процентов n -грамм будут сохранены верно, займет не больше памяти, чем количество всех n -грамм.

3. Генерация паролей

Чтобы выбрать случайным образом один элемент (часть слова длины n или $n - s$) из заданного множества $w_1, \dots, w_d$ с учетом частоты их добавления, посчитаем сумму, сколько раз был добавлен каждый из элементов. В пределах полученной суммы случайным образом выберем неотрицательное число. Повторный перебор этих слов с вычислением суммы ведется до достижения выбранного числа. Последнее слово будет выбранным. Таким образом, те элементы, которые не были добавлены, не будут выбраны.

Пусть задан алфавит, из букв которого будут генерироваться пароли (возможно, включающий конечный символ). Начало слова выбирается случайным образом из всех последовательностей длины $n - s$. Пока не будет достигнута максимальная длина пароля или не встретится конечный символ, происходит генерация новых s символов по $n - s$ последним сгенерированным символам $a_1 \dots a_{n-s}$. Для этого из множества $\{a_1 \dots a_{n-s}, b_1, \dots, b_s | b_i - \}$ выбирается один

элемент, последние s символов которого станут продолжением генерируемого слова.

4. Генерация паролей с фильтрацией для удобного набора на клавиатуре

Зададим множество правил - биграмм, которые состоят из букв заданного алфавита, причем буквы каждой отдельной биграммы находятся рядом на клавиатуре либо обе буквы одинаковы. Генерация слова осуществляется по ранее описанному алгоритму, но из меньшего набора n -грамм. Под ограничения не попадают первые два символа. Остальные символы удовлетворяют следующему условию: пусть $a_1a_2b_1 \dots b_m$ - сгенерированный пароль, для каждого символа b_i из множества $b_1, \dots, b_m$ в множестве правил есть биграмма, второй символ которой b_i , а первый находится в слове $a_1a_2b_1 \dots b_{i-1}$ на расстоянии от одного (т.е. эти символы рядом) до заданного числа r , например 4.

Для представления множества правил используем фильтр Блума в классическом варианте. Задавая вероятность коллизии, можно влиять на вероятность допуска правил, не включенных в исходное множество. Таким образом, не все символы сгенерированного слова будут удовлетворять правилам. Это расширяет пространство генерируемых паролей.

ЛИТЕРАТУРА

1. Bloom B.H. Space/time trade-offs in hash coding with allowable errors // Communications of the ACM. N.13(7). P.422-426. July 1970.
2. Spafford E.H. Opus: Preventing weak password choices // Computer and Security. N.11 P.273-278. May 1992.
3. Broder A., Mitzenmacher M. Network applications of Bloom filters: A survey // In Proc. of the 40th Annual Allerton Conference on Communication, Control, and Computing. P.636-646. 2002.

Математические структуры и моделирование

Вып. 17

Журнал

Редактор Е.В. Брусницына

Подписано в печать 20.09.2007.
ОП. Формат 60 × 84 1/8. Печ.л. 12,7. Уч.-изд.л. 21,0.
Тираж 120 экз.

Отпечатано в Полиграфическом центре КАН
644050, г. Омск, пр. Мира, 32, ком. 11, тел. (381-2) 65-47-31
Лицензия ПЛД № 58-47 от 21.04.97 г.