

## **ФОРМИРОВАНИЕ ПРИЗНАКОВ МАШИННОГО ОБУЧЕНИЯ НА ОСНОВЕ МЕТОДОВ ВЫЧИСЛИТЕЛЬНОЙ ТОПОЛОГИИ**

**С.Н. Чуканов**<sup>1</sup>

д-р техн. наук, профессор, ведущий научный сотрудник, e-mail: ch\_sn@mail.ru

**И.С. Чуканов**<sup>2</sup>

студент, e-mail: chukanov022@gmail.com

**С.В. Лейхтер**<sup>3</sup>

старший преподаватель, e-mail: leykhter@mail.ru

<sup>1</sup>Институт математики им. С.Л. Соболева СО РАН, Омский филиал, Омск, Россия

<sup>2</sup>Уральский федеральный университет имени первого Президента России  
Б.Н. Ельцина, Екатеринбург, Россия

<sup>3</sup>Омский государственный университет им. Ф.М. Достоевского, Омск, Россия

**Аннотация.** Использование традиционных методов алгебраической топологии для получения информации о форме объекта связано с проблемой формирования малого количества информации: чисел Бетти и характеристик Эйлера. Центральным инструментом топологического анализа данных является метод персистентной гомологии, который суммирует геометрическую и топологическую информацию в данных с использованием персистентных диаграмм и баркодов. На основе методов персистентной гомологии может быть выполнен анализ топологических данных для получения информации о форме объекта. Построение персистентных баркодов и персистентных диаграмм в вычислительной топологии не позволяет построить гильбертово пространство со скалярным произведением. Возможность применения методов топологического анализа данных основана на отображении персистентных диаграмм в гильбертово пространство; одним из способов такого отображения является метод построения персистентного ландшафта. Его преимущества заключаются в том что он обратим, поэтому он не теряет никакой информации, и имеет свойства персистентности. В работе рассматриваются математические модели и функции представления объектов персистентного ландшафта на основе метода персистентной гомологии. Рассмотрены методы преобразования персистентных баркодов и персистентных диаграмм в функции персистентного ландшафта.

**Ключевые слова:** распознавание образов, многопараметрический персистентный ландшафт, гильбертово пространство, топологический анализ данных.

## Введение

В последние годы возрос интерес к использованию методов алгебраической топологии для топологического анализа данных [1] и применению в различных областях знаний. Целью топологического анализа данных является определение информативных топологических свойств и использование их в качестве дескрипторов.

Ключевым математическим инструментом в топологическом анализе данных является метод персистентных гомологий, который используется для извлечения топологической информации из данных. Рассмотрим способ формирования персистентной гомологии из точек данных в евклидовом пространстве. Целью является получение топологии из конечных данных. Рассмотрим  $r$ -шары (радиуса  $r$ ) для реконструкции топологии. Ожидается, что модель  $r$ -шаров может представлять основные топологические структуры. Если  $r$  мал, то объединение всех  $r$ -шаров состоит из непересекающихся  $r$ -шаров. Если радиусы  $r$ -шаров слишком большие, то объединение становится одним пространственным компонентом. Персистентная гомология [2] рассматривает все значения одновременно и обеспечивает выражение для топологических свойств.

Персистентная гомология может быть визуализирована персистентной диаграммой  $D = \{(b_i, d_i) \in \mathbb{R}^2 \mid i \in I, b_i \leq d_i\}$ . Каждая точка  $(b_i, d_i) \in D$ , которая называется генератором персистентной гомологии, представляет топологическое свойство, появляющееся при  $X_{b_i}$  и исчезающее при  $X_{d_i}$  в модели шаров с изменяющимся радиусом  $r$ ; здесь  $b_i$  – диаметр шара при появлении (birth)  $i$ -й персистентной гомологии;  $d_i$  – диаметр шара при исчезновении (death)  $i$ -й персистентной гомологии. Топологическое свойство с большим значением  $d_i - b_i$  может рассматриваться как структура с высокой персистентностью, в то время как топологическое свойство с низкой персистентностью может рассматриваться как шум. Персистентные диаграммы кодируют топологическую и геометрическую информацию о точках данных.

Применение методов для получения информации о форме объекта для сложных систем большой размерности затруднено из-за методов адекватного представления функций, так как формирование баркодов не обеспечивает функциональную зависимость [3, 4]. Геометрический анализ характеризует локальную структуру, но приводит к сложности представления данных. Элементы, полученные из топологических моделей определяют глобальную внутреннюю структурную информацию, но редуцируют много локальной структурной информации [5].

Метод персистентных гомологий разработан для многомасштабного представления топологических признаков [1, 2, 6]; метод персистентных гомологий обеспечивает связь между топологическими и геометрическими методами. Основная идея персистентных гомологий – применение фильтрации для присвоения каждому топологическому признаку геометрической размерности. Процесс фильтрации генерирует серии симплициальных комплексов, кодируемых со структурной информацией различных масштабов. Персистентная гомология может быть представлена персистентным баркодом или персистентной диаграммой

мой.

Персистентные гомологии могут быть использованы для анализа топологических данных. Существуют подходы к формированию персистентных гомологий, основанные на построении персистентных диаграмм. Однако, стандартные меры для персистентных диаграмм (например, расстояние Вассерштейна) не подходят для топологического анализа данных из-за большого количества вычислительных операций. Одним из подходов к топологическому анализу данных является отображение персистентных диаграмм в гильбертово пространство на основе формирования функций персистентного ландшафта (landscape functions) [7, 8].

В работе представлено расширение результатов для модулей персистентности с одним параметром на многопараметрические модули персистентности. Многопараметрический персистентный ландшафт сводится к семейству ландшафтов персистентности с одним параметром. В пространстве Лебега измеримых функций  $L^p$  многопараметрических ландшафтов можно построить функции расстояния; естественная структура внутреннего произведения ландшафтных функций порождает положительно определенное ядро. Использование многопараметрических функций персистентного ландшафта повышает разнообразие параметров, характеризующих форму объекта, точность определения расстояния между изображениями объектов.

## 1. Персистентные гомологии

Предположим, что  $k + 1$  точек  $u_0, \dots, u_k \in \mathbb{R}^k$  аффинно независимы. Тогда симплекс это множество точек:

$$C = \left\{ \mu_0 u_0 + \dots + \mu_k u_k \mid \sum_{i=0}^k \mu_i = 1; \mu_i \geq 0, i = 0, \dots, k \right\}.$$

Симплициальный комплекс  $K$  — это множество симплексов, удовлетворяющее условиям: 1) каждая грань  $K$  принадлежит  $K$ ; 2)  $[(\sigma_1 \cap \sigma_2) \in \sigma_1] \wedge [(\sigma_1 \cap \sigma_2) \in \sigma_2]$ .

Функцию на симплициальном комплексе  $K$   $f : K \rightarrow \mathbb{R}; f(\sigma) \leq f(\tau); \tau \in K$ , для любой грани  $\sigma \in \tau$ . Множество подуровней  $K(a) = f^{-1}(-\infty, a); \forall a \in \mathbb{R}$  является подкомплексом в  $K$ , и упорядочение значений  $f$  на симплексах  $K$  индуцирует фильтрацию  $\emptyset = K_0 \subseteq K_1 \subseteq \dots \subseteq K_n = K$ . Включение  $K_i \rightarrow K_j; 0 \leq i \leq j \leq n$  индуцирует гомоморфизм  $f_p^{i,j} : H_p(K_i) \rightarrow H_p(K_j)$  на симплициальных группах гомологий для каждой размерности  $p$ .  $p$ -е персистентные группы гомологий являются образами этих гомоморфизмов;  $p$ -е числа Бетти являются рангами этих групп [2].

Персистентные гомологии можно представить в виде топологических генераторов (баркодов) — пар появления (BT — birth time) и исчезновения (DT — death time) баркодов, которые можно обозначить как  $l_j^k = \{b_j^k, d_j^k\}, j \in \{1, 2, \dots, N_k\}$ , где  $N_k$  — общее число  $k$ -мерных топологических генераторов [9]. Определим множество баркодов  $k$ -го измерения:  $L_k =$

$\{l_j^k = \{b_j^k, d_j^k\}_{j \in \{1, 2, \dots, N_k\}}\}$ . Топологическая персистентность может быть представлена персистентным баркодом (каждый  $l_j^k$  рассматривается как баркод) или персистентной диаграммой (каждый  $l_j^k$  рассматривается как двумерная точка с координатой  $l_j^k = (b_j^k, d_j^k)$ ) [10].

Пусть  $X = \{x_1, \dots, x_n\}$  – множество точек в метрическом пространстве  $(M, d_M)$ . Чтобы проанализировать топологические свойства  $X$ , рассмотрим модель  $X_r = \bigcup_{i=1}^n B(x_i; r)$ , состоящую из шаров  $B(x_i; r) = \{x \in M \mid d_M(x_i, x) \leq r\}$  с радиусом  $r$ , где  $d_M(x_i, x)$  – метрическое расстояние от точки  $x_i$  до точки  $x$ , и используем гомологии  $H_q(X_r)$  для описания топологии  $X_r$ . Здесь для топологического пространства  $S$  его  $q$ -я гомология  $H_q(S)$ ,  $q = 0, 1, \dots$  определяется как векторное пространство. Так как  $X_r \subset X_s$  для  $r \leq s$  множество  $X = \{X_r \mid r \geq 0\}$  становится фильтрацией. При изменении радиуса новый генератор  $a_i \in H_q(X_r)$  появляется на каком-то радиусе  $r = b_i$  (birth) и исчезает на радиусе  $r = d_i$  (death) большем чем  $b_i$ . Собирая все  $a_i$  ( $i \in I$ ) в фильтрации  $X$ , получаем множество пар  $\underline{D}_q(X) = \{(b_i, d_i) \in \mathbb{R}^2 \mid i \in I\}$  в виде мультимножества. Персистентная диаграмма  $D_q(X)$  определяется несвязным объединением  $\underline{D}_q(X)$  и диагонального множества  $\Delta = \{(a, a) \mid a \in \mathbb{R}\}$ , учитываемого с бесконечной кратностью. Точку  $x = (b, d) \in D_q(X)$  называют генератором персистентной диаграммы. Персистентность точки  $s$  равна:  $\text{pers}(x) = d - b$ .

Желательно, чтобы персистентные диаграммы были устойчивыми при возмущении данных. Мерой для изучения сходства между двумя персистентными диаграммами  $D$  и  $E$  является расстояние bottleneck  $d_B(D, E) = \inf_{\gamma} \sup_{x \in D} \|x - \gamma(x)\|_{\infty}$ , где  $\gamma$  это различные биекции от  $D$  до  $E$ :  $(x \in D) \rightarrow (\gamma(x) \in E)$ . В качестве расстояния между конечными множествами  $X, Y$  в метрическом пространстве  $M$  можно использовать расстояние Хаусдорфа, определяемое формулой:

$$d_H(X, Y) = \max \left\{ \sup_{x \in X} \inf_{y \in Y} d_M(x, y), \sup_{y \in Y} \inf_{x \in X} d_M(x, y) \right\}. \quad (1)$$

Для представления топологической информации были предложены различные функции, основанные на результатах обработки баркодов. Непрерывная персистентная функция Бетти определяется как [11]:

$$f(x; L_k) = \sum_j \exp \left( - \left( x - \frac{b_j^k + d_j^k}{2} \right) (w_j (d_j^k - b_j^k))^{-1} \right), \quad (2)$$

где  $w_j$  – значения весов.

Для каждого отдельного баркода можно определить кусочно-линейную функцию  $f(x, l_j^k)$  [7]:

$$f(x, l_j^k) = \begin{cases} 0 & \text{if } x \notin (b_j^k, d_j^k), \\ x - b_j^k & \text{if } x \in \left( b_j^k, \frac{b_j^k + d_j^k}{2} \right], \\ -x + d_j^k & \text{if } x \in \left( \frac{b_j^k + d_j^k}{2}, d_j^k \right). \end{cases} \quad (3)$$

## 2. Персистентные ландшафты

Персистентный ландшафт  $k$ -мерного баркода  $L_k$  – это последовательность функций:  $\lambda_m^{PL} : \mathbb{R} \rightarrow [0, \infty)$ ,  $m = 1, 2, 3, \dots$ , где  $\lambda_m(x)$  –  $m$ -е наибольшее значение  $\{f(x, l_j^k)\}_{j=1}^{N_k}$ . Для баркодов  $B = \{I_j\}$  можно определить функцию персистентного ландшафта как:

$$\lambda(k, t) = \sup (h \geq 0 \mid [t - h, t + h] \subset I_j, \text{ for } \geq k \text{ different } j).$$

Определим функцию для персистентных диаграмм:  $D = \{(b_i, d_i)\}$ ,  $b_i < d_i$ :  $f_{(b,d)}(t) = \max(0, \min(b + t, d - t))$ ; тогда  $\lambda(k, t) = k\text{max} \{f(b_i, d_i)(t)\}_{i \in I}$ , где  $k\text{max}$  обозначает  $k$ -й наибольший элемент.

Пусть задано множество  $S$ . Функция  $F : S \rightarrow \mathcal{H}$ , где  $\mathcal{H}$  – гильбертово пространство, называется функцией отображения признаков. Ядро на  $S$  является таким симметричным отображением  $K : S \times S \rightarrow \mathbb{R}$ , что для любого  $n$  и всех  $x_1, \dots, x_n \in S$ ,  $a_1, \dots, a_n \in \mathbb{R} : \sum_{i,j=1}^n a_i a_j K(x_i, x_j) \geq 0$ .

RKHS (Reproducing kernel Hilbert space) на множестве  $S$  – это гильбертово пространство функций на  $S$ , где точечная оценка – непрерывный линейный функционал. Для заданного отображения характеристик существует ассоциированное ядро, определяемое формулой:

$$K(x, y) = \langle F(x), F(y) \rangle_{\mathcal{H}}.$$

С ядром  $K$  связано гильбертово пространство RKHS  $\mathcal{H}_k$ , которое является пополнением множества функций  $K_x : S \rightarrow \mathbb{R}$ , заданных формулой:  $K_x(y) = K(x, y)$ ,  $\forall x \in S$ , относительно скалярного произведения:  $\langle K_x, K_y \rangle = K(x, y)$ .

Поскольку функция персистентного ландшафта является отображением характеристик из множества персистентных диаграмм в  $L^2(\mathbb{N} \times \mathbb{R})$ , то с ней ассоциируется ядро персистентного ландшафта:

$$K(D^{(1)}, D^{(2)}) = \langle \lambda^{(1)}, \lambda^{(2)} \rangle = \sum_{k=1}^{\infty} \int_{-\infty}^{\infty} \lambda_k^{(1)}(t) \lambda_k^{(2)}(t) dt, \tag{4}$$

Для персистентного ландшафта сформируем  $p$ -норму:

$$\|\lambda\|_p = \sum_{k=1}^{\infty} \left[ \int_{-\infty}^{\infty} (\lambda_k(t))^p dt \right]^{\frac{1}{p}}, \quad 1 \leq p < \infty,$$

и:  $\|\lambda\|_{\infty} = \sup_{k,t} \lambda_k(t)$ ,  $p = \infty$ .

Ядро можно рассматривать как ассоциированное отображение признаков:  $D \rightarrow \sum_{k=1}^{\infty} \lambda_k(D)$ , которое формирует отображение в гильбертово пространство со скалярным произведением:

$$\langle f, g \rangle = \sum_{k=1}^{\infty} \int_{-\infty}^{\infty} f_k(t) g_k(t) dt. \tag{5}$$

Расстояния между персистентными ландшафтами можно определить с помощью нормы  $L^\infty$ :

$$\|\lambda^{PL} - \lambda'^{PL}\|_\infty = \sup_{k,t} |\lambda_k^{PL}(t) - \lambda'_k{}^{PL}(t)|,$$

или нормы [12, 13] :

$$\|\lambda^{PL} - \lambda'^{PL}\|_p = \left[ \sum_k \int_{-\infty}^{\infty} |\lambda_k^{PL}(t) - \lambda'_k{}^{PL}(t)|^p dt \right]^{\frac{1}{p}}, 1 \leq p < \infty. \quad (6)$$

**Пример.** Аппроксимируем контур 2D изображения House пятью точками одинаковой яркости и одинакового цвета в нотации Matlab:

$$qq_x = [-1, 1, 1, -1, 0]; qq_y = [0, 0, 2, 2, 3].$$

Используя пакет JavaPlex [12, 13], определим баркоды размерности 0:  $2 \left[ 0 \ 1, 4142 \right)$ ,  $2 \left[ 0 \ 2 \right)$ ,  $\left[ 0 \ \infty \right)$ ; и размерности 1:  $\left[ 2 \ 2, 82825 \right)$ ; см. таблицу 1.

Таблица 1. Баркоды изображения House

Barcode	dim	birth	peak	death
bar1,2	0	(0,0)	(0.707, 0.707)	(1.41,0)
bar3,4	0	(0,0)	(1,1)	(2,0)
bar5	1	(2,0)	(2.414,0.414)	(2.828,0)

Получим функции персистентного ландшафта изображения (см. (3)) для размерности 0:

$$\lambda^{House}(1, t) = t \cdot st(t, (0 \dots 1]) + (2 - t) \cdot st(t, (1 \dots 2]),$$

$$\lambda^{House}(2, t) = t \cdot st(t, (0 \dots 0.707]) + (1.414 - t) \cdot st(t, (0.707 \dots 1.414]),$$

где  $st(t, (a \dots b])$  – ступенчатая функция:

$$st(t, (a \dots b]) = \begin{cases} 1 & \text{if } t \in (a \dots b], \\ 0 & \text{if } t \notin (a \dots b]. \end{cases}$$

Аппроксимируем контур 2D изображения House1 пятью точками одинаковой яркости и одинакового цвета в нотации Matlab:

$$qq_x = [-1, 1, 1, -1, 0]; qq_y = [0, 0, 2, 2, 4].$$

Используя пакет JavaPlex [12, 13], определим баркоды размерности 0:  $3 \left[ 0, 2.0 \right)$ ,  $\left[ 0, 2.233 \right)$ ,  $\left[ 0, \infty \right)$ ; и размерности 1:  $\left[ 2.0, 2.828 \right)$ ; см. таблицу 2.

Таблица 2. Баркоды изображения House1

Barcode	dim	birth	peak	death
bar1,2,3	0	(0,0)	(1.0, 1.0)	(2.0,0)
bar4	0	(0,0)	(1.116, 1.116)	(2.233,0)
bar5	1	(2,0)	(2.414,1.298)	(2.828,0)

Получим функции персистентного ландшафта изображения (см. (3)) для размерности 0:

$$\lambda^{House1}(1, t) = t \cdot st(t, (0 \dots 1.116]) + (2.233 - t) \cdot st(t, (1.116 \dots 2.233]),$$

$$\lambda^{House1}(2, t) = t \cdot st(t, (0 \dots 1]) + (2 - t) \cdot st(t, (1 \dots 2]).$$

Определим расстояние между изображениями на основании нормы  $L^2$ , используя соотношение (6) и методы топологического анализа данных:  $\|\lambda^{House} - \lambda^{House1}\|_2 = 0.5451$ .

Использование методов традиционной алгебраической топологии не позволяет различить изображения House и House1, так как они имеют одинаковые числа Бетти.  $\square$

*Выводы по разделу 2.* В разделе рассмотрен метод отображения персистентных диаграмм в гильбертово пространство на основе построения функций персистентного ландшафта. Его преимущества в том что он обратим, поэтому он не теряет никакой информации, и имеет свойства персистентности. Нахождение расстояния между объектами (изображениями) с использованием функций персистентного ландшафта (по формуле (6)) значительно уменьшает объем вычислительных операций по сравнению методом нахождения расстояния по формуле Л. Вассерштейна [2].

### 3. Многопараметрические персистентные ландшафты

Пусть  $X$  – топологическое пространство и  $f : X \rightarrow \mathbb{R}^n$ , называемая фильтрующей функцией. Можно связать семейство топологических подпространств, индексированных векторами  $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{R}^n$ , индуцированными  $F: X_{\mathbf{a}} = \{x \in X : f(x)_i < a_i \forall i = 1, \dots, n\}$ ; это известно как фильтрация множества подуровней. Для любого  $\mathbf{b} \in \mathbb{R}^n$  такого, что  $\{\mathbf{a} \leq \mathbf{b} \mid a_i \leq b_i, \forall i = 1, \dots, n\}$ , имеем отображение включения  $X_{\mathbf{a}} \mapsto X_{\mathbf{b}}$ . Если  $H$  – функтор гомологий, то применение этого функтора к набору  $\{X_{\mathbf{a}}\}_{\mathbf{a} \in \mathbb{R}^n}$  и соответствующим отображениям включения приведёт к семейству векторных пространств  $\{H(X_{\mathbf{a}})\}_{\mathbf{a} \in \mathbb{R}^n}$  и линейных отображений  $\{H(X_{\mathbf{a}}) \rightarrow H(X_{\mathbf{b}})\}_{\mathbf{a} \leq \mathbf{b}}$ , известному как многопараметрический персистентный модуль с множеством подуровней [14, 15].

Пусть  $M$  – многопараметрический персистентный модуль, тогда при  $\mathbf{a} \leq \mathbf{b}$  функция  $\beta(\cdot, \cdot)$ , задающая соответствующее число Бетти, является ранговым инвариантом  $M$  :

$$\beta(\mathbf{a}, \mathbf{b}) = \dim(\text{Im}(M_{\mathbf{a}} \rightarrow M_{\mathbf{b}})).$$

Многопараметрическая ранговая функция  $\text{rk} : \mathbb{R}^{2n} \rightarrow \mathbb{R}$  задается формулой:

$$\text{rk}(\mathbf{b}, \mathbf{d}) = \begin{cases} \beta(\mathbf{b}, \mathbf{d}) & \text{if } \mathbf{b} \leq \mathbf{d}, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Перемасштабированная ранговая функция  $r : \mathbb{R}^{2n} \rightarrow \mathbb{R}$  :

$$r(\mathbf{m}, \mathbf{h}) = \begin{cases} \beta(\mathbf{m} - \mathbf{h}, \mathbf{m} + \mathbf{h}) & \text{if } \mathbf{h} \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Многопараметрический персистентный ландшафт рассматривает максимальный радиус, в котором  $k$  признаков сохраняются в каждом (положительном) направлении через  $\mathbf{x}$  в пространстве параметров  $\lambda : \mathbb{N} \times \mathbb{R}^n \rightarrow \mathbb{R}$  :

$$\lambda(k, \mathbf{x}) = \sup \{ \varepsilon \geq 0 : \beta(\mathbf{x} - \mathbf{h}, \mathbf{x} + \mathbf{h}) \geq k, \forall \mathbf{h} \geq 0, \|\mathbf{h}\|_\infty \leq \varepsilon \}. \quad (8)$$

Пусть  $\mathbf{w} \in \{\mathbf{u} \in \mathbb{R}^n : u_i > 0, \|\mathbf{u}\|_\infty = 1\}$  – весовой вектор, соответствующий перемасштабированию пространства параметров  $\mathbb{R}^n$ . Определим  $\mathbf{w}$ -взвешенную норму  $\|\mathbf{h}\|_\infty^{\mathbf{w}}$ :  $\|\mathbf{h}\|_\infty^{\mathbf{w}} = \|(\mathbf{w} \odot \mathbf{h})\|_\infty$ .  $\mathbf{w}$ -взвешенный персистентный ландшафт представляет собой функцию  $\lambda_{\mathbf{w}} : \mathbb{N} \times \mathbb{R}^n \rightarrow \mathbb{R}$ :

$$\lambda_{\mathbf{w}}(k, \mathbf{x}) = \sup \{ \varepsilon > 0 : \beta(\mathbf{x} - \mathbf{h}, \mathbf{x} + \mathbf{h}) \geq k, \forall \mathbf{h} \geq 0, \|\mathbf{h}\|_\infty^{\mathbf{w}} \leq \varepsilon \}. \quad (9)$$

Декартово произведение функций  $p$ -ландшафта соответствует использованию функции персистентного ландшафта по каждой координате и последующему применению  $p$ -нормы,  $\lambda_p : \mathbb{N} \times \mathbb{R}^n \rightarrow \mathbb{R}$  :

$$\lambda_p(k, \mathbf{x}) = \|(\sup \{ h_i \geq 0 : \beta(\mathbf{x} - h_i e_i, \mathbf{x} + h_i e_i) \geq k \})_i\|_p.$$

Определим ландшафтное  $q$ -расстояние:

$$d_{\lambda_p}^{(q)}(M, M') = \|\lambda_p(M) - \lambda_p(M')\|_q, \quad (10)$$

где  $M, M'$  – многопараметрические персистентные модули.

*Выводы по разделу 3.* В разделе представлены методы формирования инвариантов для многопараметрических модулей персистентности, которые расширяют результаты для модулей персистентности с одним параметром на многопараметрические модули персистентности. Использование многопараметрических функций персистентного ландшафта повышает разнообразие параметров, характеризующих форму объекта, точность определения расстояния между изображениями объектов.

## Заключение

Для повышения производительности моделей топологического анализа данных необходимо ввести функциональные возможности, способные сохранить внутреннюю информацию данных и уменьшить размерность данных. Использование традиционных методов алгебраической топологии для получения информации о форме объекта связано с проблемой формирования малого количества информации. Основным инструментом топологического анализа данных является метод персистентной гомологии, который суммирует геометрическую и топологическую информацию в данных. Использование метода персистентной



гомологии по отношению к традиционным методам алгебраической топологии даёт дополнительную информацию о форме объекта. Применение методов топологического анализа данных для сложных систем большой размерности затруднено из-за методов адекватного представления функций, так как формирование баркодов не обеспечивает функциональную зависимость. Использование стандартных метрик для персистентных диаграмм затрудняет выполнение вычислительных операций. Построение персистентных баркодов и персистентных диаграмм в вычислительной топологии не позволяет построить гильбертово пространство со скалярным произведением. Возможность применения методов топологического анализа данных основано на отображении персистентных диаграмм в гильбертово пространство; одним из способов такого отображения является метод построения персистентного ландшафта.

В работе рассматриваются математические модели персистентных гомологий и функции персистентных ландшафтов представления признаков в методах топологического анализа данных. Функции персистентных ландшафтов отображают диаграммы персистентности в гильбертово пространство. Рассмотрена структура ядра для анализа персистентных диаграмм. Представлены инварианты для многопараметрических модулей персистентности, которые расширяют результаты для модулей персистентности с одним параметром на многопараметрические модули персистентности. Использование многопараметрических функций персистентного ландшафта повышает разнообразие параметров, характеризующих форму объекта, точность определения расстояния между изображениями объектов.

### Благодарности

Работа выполнена при поддержке Программы фундаментальных научных исследований СО РАН I.5.1., проект № 0314-2019-0020 и Российского научного фонда, грант № 22-21-00035.

### ЛИТЕРАТУРА

1. Carlsson G. Topology and data // *Bulletin of the American Mathematical Society*. 2009. V. 46, No. 2. P. 255–308.
2. Edelsbrunner H., Harer J.L. *Computational topology: an introduction*. American Mathematical Society. 2010.
3. Kusano G., Hiraoka Y., Fukumizu K. Persistence weighted Gaussian kernel for topological data analysis // *International conference on machine learning — PMLR*. 2016. V. 3. IEEE. P. 2004–2013.
4. Hofer C., Kwitt R., Niethammer M., Uhl A. Deep learning with topological signatures // *Advances in neural information processing systems*. 2017. V. 30.
5. Hatcher A. *Algebraic Topology*. Cambridge UP, 2005.
6. Zomorodian A.J. *Topology for computing*. V. 16. Cambridge UP, 2005.
7. Bubenik P. The persistence landscape and some of its properties / *Topological Data Analysis*. Springer, Cham, 2012. P. 97–117.

8. Pun C.S., Xia K., Lee S.X. Persistent-Homology-based Machine Learning and its Applications – A Survey. arXiv preprint arXiv:1811.00252. 2018.
9. Ghrist R. Barcodes: the persistent topology of data // Bulletin of the American Mathematical Society. 2008. V. 45, No. 1. P. 61–75.
10. Mischaikow K., Nanda V. Morse theory for filtrations and efficient computation of persistent homology // Discrete & Computational Geometry. 2013. V. 50, No. 2. P. 330–353.
11. Xia K. A quantitative structure comparison with persistent similarity. arXiv preprint arXiv:1707.03572. 2017.
12. Chukanov S. N. Comparison of objects' images based on computational topology methods // Informatics and Automation. 2019. V. 18, No. 5. P. 1043–1065.
13. Chukanov S. N. The Comparison of Diffeomorphic Images based on the Construction of Persistent Homology // Automatic Control and Computer Sciences. 2020. V. 54, No. 7. P. 758–771.
14. Vipond O. Multiparameter Persistence Landscapes // J. Mach. Learn. Res. 2020. V. 21, No. 61, P. 1–38.
15. Botnan M. B., Lesnick M. An introduction to multiparameter persistence. arXiv preprint arXiv:2203.14289. 2022.

## **FORMATION OF MACHINE LEARNING FEATURES BASED ON COMPUTATIONAL TOPOLOGY METHODS**

**S.N. Chukanov**<sup>1</sup>

Professor, Dr.Sc. (Technical), Leading Researcher, e-mail: a@a.ru

**I.S. Chukanov**<sup>2</sup>

Student, e-mail: chukanov022@gmail.com

**S.V. Leykhter**<sup>3</sup>

Senior Lecturer, e-mail: leykhter@mail.ru

<sup>1</sup>Sobolev Institute of Mathematics, Omsk branch, Russia

<sup>2</sup>Ural Federal University named after the first President of Russia B.N. Yeltsin,  
Ekaterinburg, Russia

<sup>3</sup>Dostoevsky Omsk State University, Omsk, Russia

**Abstract.** The use of traditional methods of algebraic topology to obtain information about the shape of an object is associated with the problem of forming a small amount of information: Betti numbers and Euler characteristics. The central tool for topological data analysis is the persistent homology method, which summarizes the geometric and topological information in the data using persistent diagrams and barcodes. Based on persistent homology methods, analysis of topological data can be performed to obtain information about the shape of an object. The construction of persistent barcodes and persistent diagrams in computational topology does not allow one to construct a Hilbert space with a scalar product. The possibility of applying the methods of topological data analysis is based on the mapping of persistent diagrams into the Hilbert space; one of the ways of such mapping is the method of constructing a persistent landscape. Its advantages are that it is reversible, so it does not lose any information, and it has persistence properties. The paper considers mathematical

models and functions for representing persistent landscape objects based on the persistent homology method. Methods for converting persistent barcodes and persistent diagrams into persistent landscape functions are considered. Associated with persistent landscape functions is a persistent landscape kernel that forms a mapping into a Hilbert space with a dot product. A formula is proposed for determining the distance between persistent landscapes, which allows you to find the distance between images of objects. The persistent landscape functions map persistent diagrams to Hilbert space. Examples of determining the distance between images based on the construction of persistent landscape functions for these images are given. Representations of topological characteristics in various models of computational topology are considered. Extended results for single-parameter persistence modules to multi-parameter persistence modules.

**Keywords:** pattern recognition, multi-parameter persistent landscape, Hilbert space, topological data analysis.

*Дата поступления в редакцию: 20.11.2022*