

Формирование признаков на основе методов вычислительной топологии

С.Н. Чуканов¹

¹ Институт математики им. С.Л. Соболева СО РАН, Омский филиал,
644043, Россия, г. Омск, ул. Певцова, д. 13

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. С функциями персистентного ландшафта ассоциируется ядро персистентного ландшафта, которое формирует отображение в гильбертово пространство со скалярным произведением. Предложена формула для определения расстояния между персистентными ландшафтами, которая позволяет находить расстояния между изображениями объектов.

Функции персистентного ландшафта отображают персистентные диаграммы в гильбертово пространство. Приведены примеры определения расстояния между изображениями на основании построения функций персистентного ландшафта этих изображений. Рассмотрены представления топологических характеристик в различных моделях вычислительной топологии. Расширены результаты для модулей персистентности с одним параметром на многопараметрические модули персистентности.

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

Цитирование: Чуканов, С.Н. Формирование признаков на основе методов вычислительной топологии / С.Н. Чуканов // Компьютерная оптика. – 2023. – Т. 47, № 3. – С. 482-490. – DOI: 10.18287/2412-6179-CO-1190.

Citation: Chukanov SN. Formation of features based on computational topology methods. Computer Optics 2023; 47(3): 482-490. DOI: 10.18287/2412-6179-CO-1190.

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 r одновременно и обеспечивает выражение для топологических свойств.

Персистентная гомология может быть визуализирована персистентной диаграммой $D = \{(b_i, d_i) \in \mathbb{R}^2 | i \in I, b_i \leq d_i\}$. Каждая точка $(b_i, d_i) \in D$, которая называется генератором персистентной го-

мологии, 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 высокой персистентностью $d_i - b_i$ может рассматриваться как надежная структура, 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 [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.

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:

$$L_k = \left\{ l_j^k = \{b_j^k, d_j^k\}_{j \in \{1, 2, \dots, N_k\}} \right\}.$$

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 гомотопия $Hq(S)$, $q=0,1,\dots$, определяется как векторное пространство. Так как $X_r \subset X_s$ для $r \leq s$ множество $X = \{X_r | r \geq 0\}$ 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 $\underline{D}_q(X) = \{(b_i, d_i) \in \mathbb{R}^2 | i \in I\}$ в виде мультимножества. Персистентная диаграмма $D_q(X)$ определяется несвязным объединением $\underline{D}_q(X)$ и диагонального множества $\Delta = \{(a, a) | a \in \mathbb{R}\}$, учитываемого с бесконечной кратностью. Точку $x = (b, d) \in D_q(X)$ называют генератором персистентной диаграммы. Персистентность точки x 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 γ – 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 Hausdorff расстояние, 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]:

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

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:

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

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 $\lambda(k, t) = k \max \{f_{(b_i, d_i)}(t)\}_{i \in I}$, где $k \max$ 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: $K(x, y) = \langle F(x), F(y) \rangle_H$.

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:

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

Для персистентного ландшафта сформируем p -норму:

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

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

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:

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

Расстояния между персистентными ландшафтами можно определить с помощью нормы L^∞ :

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

или нормы [12, 13]:

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

Пример 1.

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];
plot(qq_x, qq_y); plot(qq_x, qq_y).
```

Используя пакет JavaPlex [12, 13], определим баркоды размерности 0: 2 [0 1,4142), 2 [0 2), [0 ∞); и размерности 1: [2 2,82825); см. табл. 1.

Табл. 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:

$$\begin{aligned} \lambda^{House}(1,t) &= t \cdot st(t, (0 \dots 1]) + (2-t) \cdot st(t, (1 \dots 2]), \\ \lambda^{House}(2,t) &= \\ &= t \cdot st(t, (0 \dots 0,707]) + (1,414-t) \cdot st(t, (0,707 \dots 1,414]), \end{aligned}$$

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];
plot(qq_x, qq_y); plot(qq_x, qq_y).
```

Используя пакет JavaPlex [12, 13], определим баркоды размерности 0: 0: 3 [0, 2,0), [0, 2,233), [0, ∞); и размерности 1: [2,0, 2,828); см. табл. 2.

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

$$\begin{aligned} \lambda^{House1}(1,t) &= \\ &= t \cdot st(t, (0 \dots 1,116]) + (2,233-t) \cdot st(t, (1,116 \dots 2,233]), \\ \lambda^{House1}(2,t) &= t \cdot st(t, (0 \dots 1]) + (2-t) \cdot st(t, (1 \dots 2]). \end{aligned}$$

Табл. 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) |

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. □

Пример 2. Рассмотрим пример, аналогичный примеру 1, в котором находится расстояние между изображениями стеклянных бутылок.

Аппроксимируем шестнадцатью точками контур 2D-изображения бутылки молока (в нотации Matlab):

```
qq_x = [0 -1 -1,75 -1,75 -1,75 -0,75 -1 -1 1 1 0,75 1,75 1,75 1,75 1 0];
qq_y = [0 0 1 3,75 6,5 9,25 9,25 10 10 9,25 9,25 6,5 3,75 1 0 0];
plot(qq_x, qq_y); plot(qq_x, qq_y);
```

и бутылки шампанского (в нотации Matlab):

```
qq_x = [0 -1,25 -1,75 -1,75 -1,075 -0,4 -0,5 -0,5 0,5 0,5 0,4 1,075 1,75 1,75 1,25 0];
qq_y = [0 0,5 4,6,75 9,5 9,75 10 10,9,75 9,5 6,75 4,0,5 0 0];
plot(qq_x, qq_y);
```

По полученным баркодам сформируем функции персистентных ландшафтов $\lambda(k,t)$ изображения бутылки молока для размерности 0:

$$\begin{aligned} \lambda^{milk}(1,t) &= \\ &= t \cdot st(t, (0 \dots 2,75]) + (5,5-t) \cdot st(t, (2,75 \dots 5,5]); \\ \lambda^{milk}(2,t) &= \\ &= t \cdot st(t, (0 \dots 1,45]) + (2,9-t) \cdot st(t, (1,45 \dots 2,9]); \\ \lambda^{milk}(3,t) &= \\ &= t \cdot st(t, (0 \dots 0,75]) + (1,5-t) \cdot st(t, (0,75 \dots 1,5]); \\ \lambda^{milk}(4,t) &= \\ &= t \cdot st(t, (0 \dots 0,615]) + (1,23-t) \cdot st(t, (0,615 \dots 1,23]); \\ \lambda^{milk}(5,t) &= \\ &= t \cdot st(t, (0 \dots 0,5]) + (1,0-t) \cdot st(t, (0,5 \dots 1,0)); \end{aligned}$$

и изображения для бутылки шампанского:

$$\begin{aligned} \lambda^{champ}(1,t) &= \\ &= t \cdot st(t, (0 \dots 1, 75]) + (3.5 - t) \cdot st(t, (1, 75 \dots 3, 5)); \\ \lambda^{champ}(2,t) &= \\ &= t \cdot st(t, (0 \dots 1, 4]) + (2, 8 - t) \cdot st(t, (1, 4 \dots 2, 8)); \\ \lambda^{champ}(3,t) &= \\ &= t \cdot st(t, (0 \dots 1, 065]) + (2, 13 - t) \cdot st(t, (1, 065 \dots 2, 13)); \\ \lambda^{champ}(4,t) &= \\ &= t \cdot st(t, (0 \dots 0, 615]) + (1, 23 - t) \cdot st(t, (0, 615 \dots 1, 23)); \\ \lambda^{champ}(5,t) &= \\ &= t \cdot st(t, (0 \dots 0, 4]) + (0, 8 - t) \cdot st(t, (0, 4 \dots 0, 8)); \\ \lambda^{champ}(6,t) &= \\ &= t \cdot st(t, (0 \dots 0, 35]) + (0, 7 - t) \cdot st(t, (0, 35 \dots 0, 7)); \\ \lambda^{champ}(7,t) &= \\ &= t \cdot st(t, (0 \dots 0, 13]) + (0, 26 - t) \cdot st(t, (0, 13 \dots 0, 26)); \\ \lambda^{champ}(8,t) &= \\ &= t \cdot st(t, (0 \dots 0, 115]) + (0, 23 - t) \cdot st(t, (0, 115 \dots 0, 23)). \end{aligned}$$

Для нахождения расстояния между контурами 2D-изображений бутылки молока и бутылки шампанского используем соотношение (6):

$$\|\lambda^{milk} - \lambda^{champ}\|_2 = \sqrt{\sum_k \int_{-\infty}^{\infty} |\lambda^{milk}(k,t) - \lambda^{champ}(k,t)|^2 dt}.$$

В результате получим расстояние между аппроксимированными контурами 2D-изображений бутылки молока и бутылки шампанского: $\|\lambda^{milk} - \lambda^{champ}\|_2 = 2,68$, что указывает на возможность сравнения аппроксимированных контуров 2D-изображений и распознавание различий между этими изображениями. □

Выводы по параграфу 2. В параграфе 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. Методы ядра для персистентных диаграмм

Рассмотрим ядро для персистентных диаграмм, называемое персистентным взвешенным ядром Гаусса (PWGK – persistent weighted Gaussian kernel) [3]. Пусть $k^w(x,y) = w(x)w(y)k(x,y)$ – взвешенное ядро весовой функцией $w(\cdot)$; рассмотрим отображение:

$$E_{k^w} : \mu_D \rightarrow \sum_{x \in L} w(x)w(\cdot)k(\cdot, x) \in \mathcal{H}_{k^w}.$$

Для практических целей выбираем ядро Гаусса

$$k_G(x,y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right); \sigma > 0$$

для k и $w_{\text{arc}}(x) = \arctan(C(b_x - a_x)^p)$; $C > 0, p > 0$ для весовой функции.

Персистентное взвешенное ядро Гаусса (PWGK) определяется следующим образом:

$$\begin{aligned} \kappa_{PWGK}(L_k, L'_k, \sigma) &= \\ &= \sum_{l_j^k \in L_k, l_j'^k \in L'_k} w_{\text{arc}}(l_j^k)w_{\text{arc}}(l_j'^k) \exp\left(-\frac{\|l_j^k - l_j'^k\|^2}{2\sigma^2}\right), \quad (7) \\ w_{\text{arc}}(l_j^k) &= \arctan\left(C(d_j^k - b_j^k)^p\right); C, p > 0. \end{aligned}$$

Коэффициент $w_{\text{arc}}(x)$ является возрастающей функцией по отношению к персистентности x . Следовательно, генератор x дает малое значение $w_{\text{arc}}(x)$ при малых x . Изменяя параметры C, p , можно контролировать эффект персистентности.

По расстоянию между множествами точек персистентных диаграмм L_k и L'_k можно оценить расстояние между соответствующими изображениями. Если персистентные диаграммы представлены векторами в RKHS, можно применять к этим векторам методы ядра для определения расстояния между L_k и L'_k . Самый простой выбор – рассмотреть линейное ядро на RKHS:

$$k_L(D, E) = \sum_{x \in L_k} \sum_{y \in L'_k} w_{\text{arc}}(x)w_{\text{arc}}(y)k_G(x, y). \quad (8)$$

Также можно рассмотреть нелинейное ядро на RKHS, такое как ядро Гаусса:

$$k_G(L_k, L'_k) = \exp\left(-\frac{d_{k_G}^{w_{\text{arc}}}(L_k, L'_k)^2}{2\tau^2}\right), \tau > 0, \quad (9)$$

где

$$\begin{aligned} d_{k_G}^{w_{\text{arc}}}(L_k, L'_k)^2 &= \\ &= \sum_{x \in L_k} \sum_{x' \in L_k} w_{\text{arc}}(x)w_{\text{arc}}(x')k_G(x, x') + \\ &+ \sum_{y \in L'_k} \sum_{y' \in L'_k} w_{\text{arc}}(y)w_{\text{arc}}(y')k_G(y, y') - \\ &- 2 \sum_{x \in L_k} \sum_{y \in L'_k} w_{\text{arc}}(x)w_{\text{arc}}(y)k_G(x, y). \end{aligned} \quad (10)$$

Пример 2. Рассмотрим изображение House из пяти точек $[-1 \ 0; 1 \ 0; 1 \ 2; -1 \ 2; 0 \ 4]$ с баркодами в размерности $0: 2 [0 \ 1,4142], 2 [0 \ 2], [0 \ \infty]$ и изображение House1 из пяти точек $[-1 \ 0; 1 \ 0; 1 \ 2; -1 \ 2; 0 \ 4]$ с баркодами в размерности $0: 3 [0, 2, 0], [0, 2, 233], [0, \infty]$.

Определим расстояние $d_{k_G}(House, House_1)^2$ на основе соотношения (9) при $w_{arc}^{House} = w_{arc}^{House_1} = 1, 2\tau^2 = 1$:

$$\begin{aligned} \sum_{x \in L_k} \sum_{x' \in L_k} k_G(x, x') &= 4,838; \\ \sum_{y \in L'_k} \sum_{y' \in L'_k} k_G(y, y') &= 5,841; \\ \sum_{x \in L_k} \sum_{y \in L'_k} k_G(x, y) &= 4,098; \\ d_{k_G}^{w_{arc}}(L_k, L'_k)^2 &= 2,482. \end{aligned}$$

Для оценивания расстояния между изображениями по формуле (8): $k_G(L_k, L'_k) = 0,0836$. □

Выводы по параграфу 3. В параграфе рассмотрена структура ядра для топологического анализа данных на основе анализа персистентных диаграмм. Использование персистентного взвешенного ядра позволяет повысить точность определения расстояния между изображениями объектов.

4. 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 $\{\mathbf{a} \leq \mathbf{b} | a_i \leq b_i, \forall i=1, \dots, n\}$, имеем отображение включения $X_{\mathbf{a}} \rightarrow X_{\mathbf{b}}$. Если H – функтор гомологий, то применение этого функтора к набору $\{X_{\mathbf{a}}\}_{\mathbf{a} \in \mathbb{R}^n}$ и соответствующим отображениям включения приведет 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.

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 $rk: \mathbb{R}^{2n} \rightarrow \mathbb{R}$ задается формулой:

$$rk(\mathbf{b}, \mathbf{d}) = \begin{cases} \beta(\mathbf{b}, \mathbf{d}) & \text{if } \mathbf{b} \leq \mathbf{d}, \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

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 $\lambda: \mathbb{N} \times \mathbb{R}^n \rightarrow \mathbb{R}$:

$$\begin{aligned} \lambda(k, \mathbf{x}) &= \\ &= \sup\{\varepsilon \geq 0: \beta(\mathbf{x} - \mathbf{h}, \mathbf{x} + \mathbf{h}) \geq k, \forall \mathbf{h} \geq 0, \|\mathbf{h}\|_{\infty} \leq \varepsilon\}. \end{aligned} \quad (12)$$

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 $\lambda_{\mathbf{w}}: \mathbb{N} \times \mathbb{R}^n \rightarrow \mathbb{R}$:

$$\begin{aligned} \lambda_{\mathbf{w}}(k, \mathbf{x}) &= \\ &= \sup\{\varepsilon > 0: \beta(\mathbf{x} - \mathbf{h}, \mathbf{x} + \mathbf{h}) \geq k, \forall \mathbf{h} \geq 0, \|\mathbf{h}\|_{\infty}^{\mathbf{w}} \leq \varepsilon\}. \end{aligned} \quad (13)$$

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

$$\lambda_p(k, \mathbf{x}) = \left\| \left(\sup\{h_i \geq 0: \beta(\mathbf{x} - h_i \mathbf{e}_i, \mathbf{x} + h_i \mathbf{e}_i) \geq k\} \right)_i \right\|_p. \quad (14)$$

Определим ландшафтное q -расстояние:

$$d_{\lambda_p}^{(q)}(M, M') = \|\lambda_p(M) - \lambda_p(M')\|_q, \quad (15)$$

где M, M' – многопараметрические персистентные модули.

Пример 3. Поскольку изображение числа 6 можно получить из изображения числа 9 с помощью евклидова преобразования (поворота относительно центра изображения на π рад), то топологические характеристики этих изображений неразличимы.

Определим координаты точек изображения цифры 9 (в нотации Matlab):

$$9_x = [1 1 2 3 4 4 4 4 4 4 3 2 1 1 1 2 3 4];$$

$$9_y = [2 1 1 1 1 2 3 4 5 6 7 7 7 7 6 5 4 4 4 4].$$

Традиционным методом определения баркодов является метод изменения радиусов r шаров [2]. Найдем баркоды при сканировании изображений цифр 9 и 6 слева направо (то есть прямая вертикальная линия сканируется слева направо, при этом становится известна информация об изображении слева от прямой, но неизвестна информация об изображении справа от прямой) и при сканировании изображений снизу вверх, то есть прямая горизонтальная линия сканируется снизу вверх, при этом становится известна информация об изображении снизу от прямой, но неизвестна информация об изображении сверху от прямой) (см. табл. 3 и 4). Баркоды, полученные при сканировании изображений слева направо

во и снизу вверх, повышают разнообразие информации об объекте (изображении) и повышают надежность определения характеристик; например, расстояния между изображениями.

Табл. 3. Баркоды при сканировании изображения цифры 9 слева направо и снизу вверх

| Баркоды при сканировании изображения цифры 9 слева направо (right) [16] | | |
|---|---------------|---------------|
| β | | $\beta_1 = 1$ |
| Баркоды | [1,5]; [1,4] | [1,5] |
| Баркоды при сканировании изображения цифры 9 снизу вверх (up) [16] | | |
| β | $\beta_0 = 1$ | $\beta_1 = 1$ |
| Баркоды | [1,8] | [4,8] |

Функции персистентного ландшафта для $\beta_0 = 1$:

$$\begin{aligned} \lambda^{right}(1, x_{right}) &= t \cdot st(t, (1 \dots 3]) + (5-t) \cdot st(t, (3 \dots 5]); \\ \lambda^{right}(2, x_{right}) &= t \cdot st(t, (1 \dots 2,5]) + (4-t) \cdot st(t, (2,5 \dots 4]); \\ \lambda^{up}(1, x_{up}) &= t \cdot st(t, (1 \dots 4,5]) + (8-t) \cdot st(t, (4,5 \dots 8]). \end{aligned}$$

Определим координаты точек изображения цифры 6 (в нотации Matlab):

$$\begin{aligned} 6_x &= [4 \ 4 \ 3 \ 2 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 2 \ 3 \ 4 \ 4 \ 4 \ 4 \ 3 \ 2 \ 1]; \\ 6_y &= [6 \ 7 \ 7 \ 7 \ 7 \ 6 \ 5 \ 4 \ 3 \ 2 \ 1 \ 1 \ 1 \ 2 \ 3 \ 4 \ 4 \ 4 \ 4]. \end{aligned}$$

Табл. 4. Баркоды при сканировании изображения цифры 6 слева направо и снизу вверх

| Баркоды при сканировании изображения цифры 6 слева направо (right) | | |
|--|---------------|---------------|
| β | $\beta_0 = 1$ | $\beta_1 = 1$ |
| Баркоды | [1,5] | [4,5] |
| Баркоды при сканировании изображения цифры 6 снизу вверх (up) | | |
| β | $\beta_0 = 1$ | $\beta_1 = 1$ |
| Баркоды | [1,8]; [6,7] | [1,8] |

Функции персистентного ландшафта для $\beta_0 = 1$:

$$\begin{aligned} \lambda^{right}(1, x_{right}) &= t \cdot st(t, (1 \dots 3]) + (5-t) \cdot st(t, (3 \dots 5]); \\ \lambda^{up}(1, x_{up}) &= t \cdot st(t, (1 \dots 4,5]) + (8-t) \cdot st(t, (4,5 \dots 8]); \\ \lambda^{up}(2, x_{up}) &= t \cdot st(t, (6 \dots 6,5]) + (7-t) \cdot st(t, (6,5 \dots 7]). \end{aligned}$$

Найдем q – расстояние между изображениями цифры 9 и цифры 6:

$$\begin{aligned} d(M(9), M(6)) &= \\ &= \left(\sum_i \left(\left[\sum_k \int_{-\infty}^{\infty} |\lambda^i(k, M^i(9)) - \lambda^i(k, M^i(6))|^p dt \right]^{\frac{1}{p}} \right)^q \right)^{\frac{1}{q}}; \\ i &= right, up. \end{aligned}$$

При $p = q = 2$ получим:
в случае $i = right$:

$$\begin{aligned} \left[\sum_k \int_{-\infty}^{\infty} |t \cdot st(t, (1 \dots 2,5]) + (4-t) \cdot st(t, (2,5 \dots 4])|^2 dt \right]^{0,5} &= \\ &= 2,449; \end{aligned}$$

в случае $i = up$:

$$\begin{aligned} \left[\sum_k \int_{-\infty}^{\infty} |t \cdot st(t, (6 \dots 6,5]) + (7-t) \cdot st(t, (6,5 \dots 7])|^2 dt \right]^{0,5} &= \\ &= 4,425. \end{aligned}$$

q – расстояние между многопараметрическими персистентными ландшафтами равно:

$$d(M(9), M(6)) = (2,449^2 + 4,425^2)^{0,5} = 5,057;$$

то есть изображение цифры 9 отличается от изображения цифры 6.

Выводы по параграфу 4. В параграфе 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 гомологий, который суммирует геометрическую и топологическую информацию в данных. Использование методов персистентной гомологии по отношению к традиционным методам алгебраической топологии дает 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. Его преимущества в том, что он обратим, поэтому он не теряет никакой информации и имеет свойства персистентности. 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 исследований СО РАН I.5.1, проект № 0314-2019-0020 и Российского научного фонда, грант № 22-21-00035.

References

- [1] Carlsson G. Topology and data. Bulletin of the American Mathematical Society 2009; 46(2): 255-308. DOI: 10.1090/S0273-0979-09-01249-X.
- [2] Edelsbrunner H, Harer JL. Computational topology: an introduction. American Mathematical Society, 2010. ISBN: 978-0-8218-4925-5.
- [3] Kusano G, Hiraoka Y, Fukumizu K. Persistence weighted Gaussian kernel for topological data analysis. Int Conf on Machine Learning (PMLR) 2016: 2004-2013.
- [4] Hofer C, et al. Deep learning with topological signatures. NIPS'17: Proc 31st Int Conf on Neural Information Processing Systems 2017: 1633-1643.
- [5] Hatcher A. Algebraic topology. Cambridge UP; 2005. ISBN: 978-0-521-79160-1.
- [6] Zomorodian AJ. Topology for computing. Cambridge University Press; 2005. ISBN: 978-0-521-83666-1.
- [7] Bubenik P. The persistence landscape and some of its properties. In Book: Topological data analysis. Cham: Springer; 2020: 97-117. DOI: 10.1007/978-3-030-43408-3_4.
- [8] Pun CS, Xia K, Lee SX. Persistent-homology-based machine learning and its applications--A survey. arXiv preprint. 2018. Source: <https://arxiv.org/abs/1811.00252>. DOI: 10.48550/arXiv.1811.00252.
- [9] Ghrist R. Barcodes: the persistent topology of data. Bulletin of the American Mathematical Society 2008; 45(1): 61-75. DOI: 10.1090/S0273-0979-07-01191-3.
- [10] Mischaikow K, Nanda V. Morse theory for filtrations and efficient computation of persistent homology. Discrete & Computational Geometry 2013; 50(2): 330-353. DOI: 10.1007/s00454-013-9529-6.
- [11] Xia K. A quantitative structure comparison with persistent similarity. arXiv preprint. 2017. Source: <https://arxiv.org/abs/1707.03572>. DOI: 10.48550/arXiv.1707.03572.
- [12] Chukanov SN. Comparison of objects' images based on computational topology methods. Informatics and Automation 2019; 18(5): 1043-1065. DOI: 10.15622/sp.2019.18.5.1043-1065.
- [13] Chukanov SN. The comparison of diffeomorphic images based on the construction of persistent homology. Automatic Control and Computer Sciences 2020; 54(7): 758-771. DOI: 10.3103/S0146411620070056.
- [14] Vipond O. Multiparameter persistence landscapes. J Mach Learn Res 2020; 21(61): 1-38.
- [15] Botnan MB, Lesnick M. An introduction to multiparameter persistence. arXiv preprint. 2022. Source: <https://arxiv.org/abs/2203.14289>. DOI: 10.48550/arXiv.2203.14289.
- [16] Adcock A, Carlsson E, Carlsson G. The ring of algebraic functions on persistence bar codes. arXiv preprint. 2013. Source: <https://arxiv.org/abs/1304.0530>.
- [17] Kwitt R, et al. Statistical topological data analysis-a kernel perspective. NIPS'15: Proc 28th Int Conf on Neural Information Processing Systems 2015; 2: 3070-3078.
- [18] Sriperumbudur BK, Fukumizu K, Lanckriet GRG. Universality, characteristic kernels and RKHS embedding of measures. J Mach Learn Res 2011; 12(7): 2389-2410.

Приложение. Метод ядра в машинном обучении [17, 18]

В задачах анализа данных нас интересует классификация данных во входном пространстве с помощью разделения гиперплоскостью. Однако использование линейного разделения ограничивает эффективность такого подхода. Можно использовать нелинейное разделение во входном пространстве, и метод RKHS (Reproducing kernel Hilbert space) обеспечивает основу для достижения этого разделения.

Рассмотрим симметричную меру подобия, называемую ядром:

$$\kappa : \Omega \times \Omega \rightarrow \mathbb{R}, (x, x') \rightarrow \kappa(x, x'), x, x' \in \Omega \in \mathbb{R}^d.$$

Так как $\Omega \in \mathbb{R}^d$, то существует возможность рассмотреть евклидово скалярное произведение для вычисления мер подобия: $\kappa(x, y) = x^T y$.

Предположим, что κ – это положительно определенное ядро с действительными значениями и Ω – непустое множество. Если $\mathbb{R}^\Omega = \{f : \Omega \rightarrow \mathbb{R}\}$, то отображение признаков – это такая функция, что $\Phi : \Omega \rightarrow \mathbb{R}^\Omega, x = \kappa(x, \cdot)$

Φ отображает образы в функции на \mathbb{R}^Ω . Это позволяет нам встраивать данные в векторное пространство признаков:

$$\mathcal{F} = \left\{ \sum_{i=1}^n \alpha_i \kappa(x_i, \cdot) \mid n \in \mathbb{N}, x_i \in \Omega, \alpha_i \in \mathbb{R}, i = 1, \dots, n \right\}.$$

Используя эту концепцию, можно построить предгильбертово пространство. Пусть $f, g \in \mathcal{F}$ ассоциируются с образами $x_i, x_j \in \Omega; i = 1, \dots, n; j = 1, \dots, n'$:

$$f = \sum_{i=1}^n \alpha_i \kappa(x_i, \cdot), g = \sum_{j=1}^{n'} \beta_j \kappa(x'_j, \cdot); \alpha_i, \beta_j \in \mathbb{R}.$$

Определим внутреннее (скалярное) произведение:

$$\langle f, g \rangle = \sum_{j=1}^{n'} \sum_{i=1}^n \alpha_i \beta_j \kappa(x_i, x'_j) = \sum_{j=1}^{n'} \beta_j f(x'_j) = \sum_{i=1}^n \alpha_i g(x_i). \tag{16}$$

Функция κ , определенная на $\Omega \times \Omega$, является воспроизводящим ядром если и только если существуют гильбертово пространство \mathcal{H} и отображение $\Phi: \Omega \rightarrow \mathcal{H}$, такие что $\kappa(x, x') = \langle \Phi_x, \Phi_{x'} \rangle_{\mathcal{H}}; \forall x, x' \in \Omega$. С точки зрения внутреннего произведения пространства: $\kappa(x, x') = \langle \kappa(\cdot, x), \kappa(\cdot, x') \rangle_{\mathcal{H}_{\kappa(\Omega)}}$. Дополнительно $\langle f, g \rangle = \langle g, f \rangle$ и:

$$\langle f, f \rangle = \sum_{i,j=1}^n \alpha_i \alpha_j \kappa(x_i, x_j) \geq 0, \left\langle \sum_{i=1}^n c_i f_i, \sum_{j=1}^n c_j f_j \right\rangle \geq 0.$$

Это означает, что $\langle \cdot, \cdot \rangle$ является положительно определенным ядром в пространстве признаков.

Равенство $\langle f, f \rangle = 0$ подразумевает $f = 0$ и $|f(x)|^2 = |\kappa(x, \cdot), f|^2 \leq \kappa(x, x) \cdot \langle f, f \rangle$. Таким образом, $\langle \cdot, \cdot \rangle$ – это хорошо определенное скалярное произведение.

Вспоминая воспроизводящее свойство положительно определенных ядер, получим, что для всех функций из \mathcal{F} имеем $\langle \kappa(x, \cdot), f \rangle = f(x)$, и в частности (kernel trick):

$$\langle \Phi(x), \Phi(x') \rangle = \kappa(x, x'). \tag{17}$$

Сведения об авторе

Чуканов Сергей Николаевич, 1951 года рождения. Д-р техн. наук, профессор; ведущий научный сотрудник ФГБУН Института математики им. С.Л. Соболева СО РАН (Омский филиал). Область научных интересов: системы управления летательными аппаратами, распознавание образов. Число научных публикаций – 100. E-mail: ch_sn@mail.ru.

ГРНТИ: 28.23.15.

Поступила в редакцию 11 июля 2022 г. Окончательный вариант – 8 октября 2022 г.

Formation of features based on computational topology methods

*S.N. Chukanov*¹

¹ *Sobolev Institute of Mathematics, SB RAS, Omsk Branch,
644043, Omsk, Russia, st. Pevtsova 13*

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, namely, 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, topological data can be analyzed 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 mapping persistent diagrams into a Hilbert space; one of the ways of such mapping is a method of constructing a persistence landscape. It has an advantage of being reversible, so it does not lose any information and has persistence properties.

The paper considers mathematical models and functions for representing persistence landscape objects based on the persistent homology method. Methods for converting persistent barcodes and persistent diagrams into persistence landscape functions are considered. Associated with persistence landscape functions is a persistence landscape kernel that forms a mapping into a Hilbert space with a dot product. A formula is proposed for determining a distance between the persistence landscapes, which allows the distance between images of objects to be found.

The persistence landscape functions map persistent diagrams into a Hilbert space. Examples of determining the distance between images based on the construction of persistence landscape functions for these images are given. Representations of topological characteristics in various models of computational topology are considered. Results for one-parameter persistence modules are extended onto multi-parameter persistence modules.

Keywords: pattern recognition, multivariate persistent landscape, Hilbert space, topological data analysis.

Citation: Chukanov SN. Formation of features based on computational topology methods. *Computer Optics* 2023; 47(2): 482-490. DOI: 10.18287/2412-6179-CO-1190.

Acknowledgements: This work was supported by the Basic Research Program of the Siberian Branch of the Russian Academy of Sciences I.5.1, Project no. 0314-2019-0020 and the Russian Science Foundation, Project no. 22-21-00035.

Authors' information

Sergey Nikolayevich Chukanov (b. 1951) – Doctor of Technical Sciences, professor; Leading Researcher, Sobolev Institute of Mathematics of the Siberian Branch of the Russian Academy of Sciences. Research interests: spacecraft control systems, pattern recognition. The number of publications – 100. E-mail: ch_sn@mail.ru.

Received July 11, 2022. The final version – October 8, 2022.
