MODELING AND ANALYSIS OF INFORMATION SYSTEMS, VOL. 29, NO. 3, 2022
journal homepage: www.mais-journal.ru

DISCRETE MATHEMATICS IN RELATION TO COMPUTER SCIENCE

Formation of Machine Learning Features Based on the Construction of Tropical Functions

S. N. Chukanov[1], I. S. Chukanov[2]

DOI: 10.18255/1818-1015-2022-3-200-209

[1]Sobolev Institute of Mathematics, SB RAS, Omsk branch, 13 Pevtsova str., Omsk, 644043, Russia.
[2]Ural Federal University named after the First President of Russia B. N. Yeltsin, st. Mira, 19, Yekaterinburg, Sverdlovsk Oblast, Russia.

MSC2020: 68T10 14T25
Research article
Full text in English

Received June 5, 2022
After revision August 24, 2022
Accepted August 26, 2022

One of the main methods of computational topology and topological data analysis is persistent homology, which combines geometric and topological information about an object using persistent diagrams and barcodes. The persistent homology method from computational topology provides a balance between reducing the data dimension and characterizing the internal structure of an object. Combining machine learning and persistent homology is hampered by topological representations of data, distance metrics, and representation of data objects. The paper considers mathematical models and functions for representing persistent landscape objects based on the persistent homology method. The persistent landscape functions allow you to map persistent diagrams to Hilbert space. The representations of topological functions in various machine learning models are considered. An example of finding the distance between images based on the construction of persistent landscape functions is given. Based on the algebra of polynomials in the barcode space, which are used as coordinates, the distances in the barcode space are determined by comparing intervals from one barcode to another and calculating penalties. For these purposes, tropical functions are used that take into account the basic structure of the barcode space. Methods for constructing rational tropical functions are considered. An example of finding the distance between images based on the construction of tropical functions is given. To increase the variety of parameters (machine learning features), filtering of object scanning by rows from left to right and scanning by columns from bottom to top are built. This adds spatial information to topological information. The method of constructing persistent landscapes is compatible with the approach of constructing tropical rational functions when obtaining persistent homologies.

Keywords: persistent homology; persistent landscape; machine learning; RKHS; Hilbert space; tropical functions

INFORMATION ABOUT THE AUTHORS

Sergey N. Chukanov correspondence author	orcid.org/0000-0002-8106-9813. E-mail: ch_sn@mail.ru Leading Researcher, Doctor of Technical Sciences, Professor.
Ilya S. Chukanov	orcid.org/0000-0001-9946-7484. E-mail: chukanov022@gmail.com Student.

Funding: This work was supported by the Program for Fundamental Research of the Siberian Branch of the Russian Academy of Sciences No. I.5.1, project No. 0314-2019-0020 and the Russian Science Foundation, grant No. 22-21-00035.

For citation: S. N. Chukanov and I. S. Chukanov, "Formation of Machine Learning Features Based on the Construction of Tropical Functions", *Modeling and analysis of information systems*, vol. 29, no. 3, pp. 200-209, 2022.

МОДЕЛИРОВАНИЕ И АНАЛИЗ ИНФОРМАЦИОННЫХ СИСТЕМ, ТОМ 29, № 3, 2022
сайт журнала: www.mais-journal.ru

DISCRETE MATHEMATICS IN RELATION TO COMPUTER SCIENCE

Формирование признаков машинного обучения на основе построения тропических функций

С. Н. Чуканов[1], И. С. Чуканов[2]

DOI: 10.18255/1818-1015-2022-3-200-209

[1]Институт математики им. С. Л. Соболева СО РАН, Омский филиал, ул. Певцова, 13, г. Омск, 644043 Россия.
[2]Уральский федеральный университет имени первого Президента России Б. Н. Ельцина, ул. Мира, 19, Екатеринбург, Свердловская обл., Россия.

УДК 004.93'1
Научная статья
Полный текст на английском языке

Получена 5 июня 2022 г.
После доработки 24 августа 2022 г.
Принята к публикации 26 августа 2022 г.

Одним из основных методов вычислительной топологии и топологического анализа данных является персистентная гомология, объединяющая геометрическую и топологическую информацию об объекте с использованием персистентных диаграмм и баркодов. Метод персистентной гомологии из вычислительной топологии обеспечивает баланс между уменьшением размерности данных и характеристикой внутренней структуры объекта. Объединению машинного обучения и персистентной гомологии препятствуют топологические представления данных, метрики расстояния и представление объектов данных. В работе рассматриваются математические модели и функции представления объектов персистентного ландшафта на основе метода персистентной гомологии. Функции персистентного ландшафта позволяют отображать персистентные диаграммы в гильбертово пространство. Рассмотрены представления топологических функций в различных моделях машинного обучения. Приведен пример нахождения расстояния между изображениями на основе построения функций персистентного ландшафта.
На основе алгебры полиномов в пространстве баркодов, которые используются в качестве координат, определяются расстояния в пространстве баркода сопоставлением интервалов от одного баркода к другому и расчета штрафов. Для этих целей используются тропические функции, которые учитывают базовую структуру пространства баркода. Рассмотрены методы построения рациональных тропических функций. Приведен пример нахождения расстояния между изображениями на основе построения тропических функций. Для повышения разнообразия параметров (признаков машинного обучения) построены фильтрации сканирования объекта по строкам слева направо и сканирования по столбцам снизу вверх. Это добавляет пространственную информацию к топологической информации. Метод построения персистентных ландшафтов совместим с подходом построения тропических рациональных функций при получении персистентных гомологий.

Ключевые слова: персистентные гомологии; персистентный ландшафт; машинное обучение; RKHS; гильбертово пространство; тропические функции

ИНФОРМАЦИЯ ОБ АВТОРАХ

Сергей Николаевич Чуканов автор для корреспонденции	orcid.org/0000-0002-8106-9813. E-mail: ch_sn@mail.ru ведущий научный сотрудник, доктор технических наук, профессор.
Илья Станиславович Чуканов	orcid.org/0000-0001-9946-7484. E-mail: chukanov022@gmail.com студент.

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

Для цитирования: S. N. Chukanov and I. S. Chukanov, “Formation of Machine Learning Features Based on the Construction of Tropical Functions”, *Modeling and analysis of information systems*, vol. 29, no. 3, pp. 200-209, 2022.

Introduction

A central tool in topological data analysis is persistent homology, which summarizes geometric and topological information in data using persistent diagrams and barcodes [1–6]. The use of persistent homology in relation to the traditional methods of algebraic topology [7] provides additional information about the shape of the object.

Machine learning can then be performed to analyze the topological data [8]. The application of machine learning methods for complex systems of large dimensions is difficult due to the methods of adequate representation of functions [9, 10]. The use of standard metrics for persistence charts makes it difficult to perform computational operations. Simplifying the application of machine learning methods is to map persistent diagrams to Hilbert space; one way is the persistent landscape method [9, 10]. Its advantages are that it is reversible so it does not lose any information, has persistence properties, has no parameters, and is non-linear.

Persistent landscapes are compatible with the tropical rational functions approach for obtaining persistent homologies [11], since persistent landscapes can be represented by tropical rational functions.

1. Persistence modules, persistence diagrams and barcodes

The persistent module [9] M is composed of vector space $M(a)$ for every real number a and $a \le b$ a linear mapping $M(a \le b) : M(a) \longrightarrow M(b)$ such that for $a \le b \le c : M(b \le c) \circ M(a \le b) = M(a \le c)$. Persistence modules arise in topological data analysis from the homology of a filtered simplicial complex.

In many cases, a persistent module can be represented entirely by a set of intervals called a barcode. Another representation of a barcode is a persistence diagram, consisting of pairs $(b_i, d_i), -\infty < b_i < d_i < \infty, i = 1, \ldots, n$, which are which are the start and end points of the intervals in the barcode. points of the intervals in the barcode. To determine the distance between modules, we will use the Wasserstein distance or the bottleneck distance. These distances induce a topology in the space of persistent diagrams.

If we will consider the sequences of persistence diagrams $D_1, ..., D_n$, then we can consider this sequence as a point $(D_1, ..., D_n)$ in the space of the product of n persistence diagrams with the metric:

$$d((D_1, ..., D_n), (D'_1, ..., D'_n)) = \max\{d_B(D_1, D'_1), \ldots, d_B(D_n, D'_n)\}.$$

2. Persistent landscapes

For a given persistence module M, we can define the persistence landscape as a function $\lambda : \mathbb{N} \times \mathbb{R} \longrightarrow \mathbb{R}$, given by $\lambda(k, t) = \sup(h \ge 0 | \operatorname{rank} M(t - h \le t + h) \ge k)$. For a barcode $B = \left\{ I_j \right\}$, we can define a persistent landscape as $\lambda(k, t) = \sup \left(h \ge 0 | [t - h, t + h] \subset I_j ;\ \text{for at least } k \text{ different } j \right)$

For a persistent diagram $D = \{(a_i, b_i)\}$, we can define a persistent landscape as follows. For $a < b$ we define $f_{(a,b)}(t) = \max(0, \min(a + t, b - t))$. Then $\lambda(k, t) = k_{\max} \{f(a_i, b_i)(t)\}_{i \in I}$, where $k_{\max}$ denotes the k-th largest element.

A persistent landscape can also be thought of as a sequence of functions $\lambda_1, \lambda_2, \ldots : \mathbb{R} \longrightarrow \mathbb{R}$, where λ_k is called the k-th function of the persistent landscape. The function λ_k is piecewise linear with a slope of 0, 1, or -1. Critical points of λ_k are those values of t, at which the slope changes. The set of critical points of the persistent landscape λ is the union of the sets of critical points of functions λ_k. The average persistent landscape of landscapes $\lambda^{(1)}, \ldots, \lambda^{(N)}$ is given by the formula: $\bar{\lambda}(k, t) = N^{-1} \sum_{j=1}^{N} \lambda^{(j)}(k, t)$.

Let M is a persistent module. For $a \le b$ the corresponding Betti number for M is determined by the dimension of the image of the corresponding linear mapping: $\beta^{a,b} = \dim(\operatorname{Im}(M(a \le b)))$.

The rank function: $\lambda : \mathbb{R}^2 \longrightarrow \mathbb{R}$, is the function given by the expression:

$$\lambda(b, d) = \begin{cases} \beta^{b,d} & \text{if} \quad b \le d, \\ 0, & \text{if} \quad b > d. \end{cases}$$

Let's change the coordinates so that the resulting function lean on the upper half-plane. Let $m = 0,5(b+d), h = 0.5(d-b)$. Then the rescaled rank function $\lambda : \mathbb{R}^2 \longrightarrow \mathbb{R}$, has the form:

$$\lambda(m,h) = \begin{cases} \beta^{m-h,m+h} & \text{if} \quad h \geq 0, \\ 0, & \text{if} \quad h < 0. \end{cases}$$

The persistent landscape is a function $\lambda : \mathbb{N} \times \mathbb{R} \longrightarrow \bar{\mathbb{R}}$, where $\mathbb{R}$ denotes the extended real numbers in the interval $[-\infty, \infty]$. Alternatively, it can be considered a sequence of functions $\lambda_k : \mathbb{R} \longrightarrow \bar{\mathbb{R}}$, where $\lambda_k(t) = \lambda(k,t)$. Let's define $\lambda_k(t) = \sup(m \geq 0 | \beta^{t-m,t+m} \geq k)$.

Let a set be given S. The function $F : S \longrightarrow H$, where is the Hilbert space, is called the feature mapping. The kernel on S is such symmetric mapping: $K : S \times S \longrightarrow \mathbb{R}$, that for any n and all $x_1, ..., x_n \in S, a_1, ..., a_n \in \mathbb{R}$: $\sum_{i,j=1}^{n} a_i a_j K\left(x_i, x_j\right) \geq 0$. A Hilbert space with a reproducing kernel (RKHS [12]) on a set S is a Hilbert space of real-valued functions on S such that the point value functional is continuous. For a given mapping of characteristics, there is an associated kernel defined by the formula: $K(x,y) = \langle F(x), F(y) \rangle_H$.

The kernel K has an associated space RKHS H_k, which is the completion of the range of functions $K_x : S \longrightarrow \mathbb{R}$, given by the formula: $K_x(y) = K(x,y), \forall x \in S$, with respect to the scalar product given by the formula $\left\langle K_x, K_y \right\rangle = K(x,y)$.

A persistent landscape is stable in the following sense. Let D, D' be persistent diagrams and λ, λ' be their persistent landscapes. Then for all k: $|\lambda_k(t) - \lambda'_k(t)| \leq d_b(D, D')$, where d_b is the bottleneck distance. For two persistence diagrams $D = \{(a_1,b_1), \ldots, (a_n,b_n)\}$, $D' = \left\{(a'_1,b'_1), \ldots, (a'_n,b'_n)\right\}$, let λ, λ' be the associated persistence landscapes. Then $\|\lambda - \lambda'\|_\infty \leq \|(a_1,b_1,\ldots,a_n,b_n) - (a'_1,b'_1,\ldots,a'_n,b'_n)\|_\infty$.

Since the persistent landscape is a mapping of characteristics from the set of persistent diagrams to , then there is an associated kernel of the persistent landscape, which is determined by the scalar product:

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

One of the advantages of a persistent landscape is that its definition does not include parameters. A set S of persistent diagrams in a vector space is called linear if for two persistent diagrams D_1, D_2: $S(D_1 \cup D_2) = S(D_1) + S(D_2)$; the persistent landscape is non-linear. There are fast algorithms and software for calculating the persistent landscape.

For real-valued functions on $\mathbb{N} \times \mathbb{R}$ we define the p-norm $(1 \leq p \leq \infty)$; for persistent landscapes at $1 \leq p < \infty$: $\|\lambda\|_p = \sum_{k=1}^{\infty} [\int_{-\infty}^{\infty} (\lambda_k(t))^p dt]^{\frac{1}{p}}$; at $p = \infty$: $\|\lambda\|_\infty = \sup_{k.t} \lambda_k(t)$.

Weighted versions of these norms and internal works can be used; for any non-negative function $w : \mathbb{N} \times \mathbb{R} \longrightarrow \mathbb{R}$: $K_w(D^{(1)}, D^{(2)}) = \langle \sqrt{w} \cdot \lambda^{(1)}, \sqrt{w} \cdot \lambda^{(2)} \rangle$.

Let M, M' be the persistent modules and λ, λ' their corresponding persistent landscapes be. For $1 \leq p \leq \infty$, we define p-landscape distance between M, M' :

$$\Lambda_p(M, M') = \|\lambda - \lambda'\|_p. \quad (2)$$

Table 1. Barcodes of Image 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)

Example 1. Let's consider the image House from five points $[\ -1,0;\quad 1,0;\quad 1,2;\quad -1,2;\quad 0,3\ ]$. Let's find barcodes of dimension 0: $2[\ 0\quad 1.4142\ )$, $2[\ 0\quad 2\ )$, $[\ 0\quad \infty\ )$; dimensions 1: $[\ 2\quad 2.82825\ )$; see table 1.

For dimension 0: $\lambda^{House}(1,t) = t \cdot st(t,(0\dots 1]) + (2-t)\cdot st(t,(1\dots 2])$,
$\lambda^{House}(2,t) = t\cdot st(t,(0\dots 0.707]) + (1.414-t)\cdot st(t,(0.707\dots 1.414])$,
where $st(t,(a\dots b])$ is the step function:

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

Consider now the image House1 of five points $[\ -1,0;\quad 1,0;\quad 1,2;\quad -1,2;\quad 0,4\ ]$.
Let's define barcodes of dimension 0: $3[0,2.0)$, $[0,2.233)$, $[0,\infty)$; dimensions 1: $[2.0,2.828)$; see table 2.

Table 2. Barcodes of Image 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)

For dimension 0:
$\lambda^{House1}(1,t) = t\cdot st(t,(0\dots 1.116]) + (2.233-t)\cdot st(t,(1.116\dots 2.233])$,
$\lambda^{House1}(2,t) = t\cdot st(t,(0\dots 1]) + (2-t)\cdot st(t,(1\dots 2])$.
Let us define the norm $L^2 : \|\lambda^{House} - \lambda^{House1}\|_2 = 0.5451$.

3. Tropical features

The term "Tropical mathematics" was introduced by Academician Viktor Pavlovich Maslov [13]. In Mathematical Subject Classification (MSC-2020) there is a rubric "14Txx Tropical geometry". V.P. Maslov began to use such expressions as "tropical picture", "tropical relations" in the early 1980s in connection with the economic situation in Russia in the early years of perestroika and the liberalization of political and economic restrictions in the USSR, by analogy with the slave trade in Tropical Africa.

Tropical algebra is based on the study of the tropical semiring $(\mathbb{R} \cup \{\infty\}, \oplus, \odot)$. In this semiring, addition and multiplication are defined as follows: $a \oplus b := \max(a,b)$ and $a \odot b := a + b$.

Both are commutative and associative. An operator $\odot$ takes a priority, when $\oplus$ and $\odot$ occur in the same expression. There is a distributive law: $a \odot (b \oplus c) = a \odot b \oplus a \odot c$. The following identity holds in tropical arithmetic: $(a \oplus b)^n = a^n \oplus b^n, \forall n$.

Let $x_1,\dots,x_n$ be variables representing elements in the Max-plus semiring. Max-plus monomial expression is any product of these variables where repetition is allowed. Max-plus polynomial expression is a finite linear combination of Max-plus expressions of monomials:

$$p(x_1,\dots,x_n) = a_1 \odot x_1^{a_1^1} \dots x_n^{a_n^1} \oplus a_2 \odot x_1^{a_1^2} \dots x_n^{a_n^2} \oplus \dots \oplus a_m \odot x_1^{a_1^m} \dots x_n^{a_n^m},$$

here the coefficients $a_1, \dots, a_m$ are real numbers and $a_j^i, 1 \le j \le n, 1 \le i \le m$ are non-negative integers. The total power of the Max-plus expression $p(x_1, \dots, x_n)$ is $\deg(p) = \max\limits_{1 \le i \le m}(a_1^i + \dots + a_n^i)$.

Considered as a function $p : \mathbb{R}^n \longrightarrow \mathbb{R}$ has the following properties: p is continuous; p is piecewise linear; p is convex.

Max-plus polynomials is a semiring of equivalence classes for expressions of Max-plus polynomials. In the case of n variables, we denote the semiring by MaxPlus$[x_1, \dots, x_n]$.

Tropical rational expression is quotient:

$$r(x_1, \dots, x_n) = p(x_1, \dots, x_n) \odot q(x_1, \dots, x_n)^{-1} = p(x_1, \dots, x_n) - q(x_1, \dots, x_n), \tag{3}$$

where p, q are Max-plus polynomial expressions.

The semiring of equivalence classes of tropical rational expressions with respect to a functional equivalence relation RTrop$[x_1, \dots, x_n]$ is called the semiring of rational tropical functions.

Any function can be represented by an expression of the form $p \odot q^{-1}$, where p, q are tropical polynomial expressions. The calculation algorithm for p, q is the usual algorithm for adding fractions by finding a common denominator, but performed in tropical arithmetic [7].

The functions contained in MaxPlus$[x_1, \dots, x_n]$ or RTrop$[x_1, \dots, x_n]$ are called tropical functions [Kalisnik].

A tropical function f is symmetric if $f(x_1, \dots, x_n) = f(x_{\pi(1)}, \dots, x_{\pi(n)})$ for each permutation $\pi \in S_n$.

For given variables $x_1, \dots, x_n$, we define elementary symmetric Max-plus polynomials $\sigma_1, \dots, \sigma_n \in$ MaxPlus$[x_1, \dots, x_n]$ by the formulas:

$$\begin{aligned} \sigma_1 &= x_1 \oplus \dots \oplus x_n, \\ &\vdots \\ \sigma_k &= \bigoplus_{\pi \in S_n} x_{\pi(1)} \odot \dots \odot x_{\pi(k)}, \\ &\vdots \\ \sigma_n &= x_1 \odot \dots \odot x_n. \end{aligned} \tag{4}$$

Let us show that the persistent landscape is a tropical rational function.

Let $D = \{(a_i, b_i)\}_{i=1}^n$, $-\infty < a_i < b_i < \infty$, is a persistence diagram. k-th landscape persistence function is given by $\lambda_k(t) = k\max\limits_{1 \le i \le n} f_{(a_i, b_i)}(t)$, where $f_{(a,b)}(t) = \max(0, \min(a + t, b - t))$.

Let's rewrite $f_{(a,b)}$ as a tropical rational expression with one variable t :

$$\begin{aligned} &f(a, b)(t) = \max(0, -\max(-(a + t), t - b)) = \\ &\max(0, -\max((a \odot t)^{-1}, t \odot b^{-1})) = 0 \oplus [(a \odot t)^{-1} \oplus (t \odot b^{-1})]^{-1}. \end{aligned}$$

The right term can be simplified using the usual rules for adding fractions:

$$f_{(a,b)}(t) = 0 \oplus (a + b) \odot t \odot \left(b \oplus a \odot t^2\right)^{-1}.$$

Consider polynomials Max-plus in n variables, $x_1, \dots, x_n$. The elementary symmetric polynomials Max-plus, $\sigma_1, \dots, \sigma_n$, are given by the formula: $\sigma_k(x_1, \dots, x_n) = \bigoplus_{\pi \in S_n} x_{\pi(1)} \odot \cdots \odot x_{\pi(k)}$, where the sum is taken over the elements π of the symmetric group S_n. σ_k is the sum of the k-th largest monomials $x_1, \dots, x_n$. Consequently, $k\max\limits_{1 \le i \le n} x_i = \sigma_k(x_1, \dots, x_n) - \sigma_{k-1}(x_1, \dots, x_n)$.

Thus $\lambda_k(t) = \sigma_k(f_i(t)) \cdot \sigma_{k-1}(f_i(t))^{-1}$, where $\sigma_k(x_i)$ is written for $\sigma_k(x_1, \dots, x_n)$ and $f_i(t)$ for $f_{(a_i, b_i)}(t)$. Therefore, for a fixed persistent diagram D, we have as a tropical rational function λ_k of one variable t.

However, t should be considered as a fixed value and the persistence chart as a variable. Consider $f_t(a, b) = 0 \oplus t \odot a \odot b \odot (b \oplus 2t \odot a)^{-1}$ a tropical rational function of variables a, b. Further,

$$\sigma_k(f_t(a_1, b_1), \dots, f_t(a_n, b_n)) = \oplus_{\pi \in S_n} f_t(a_{\pi(1)}, b_{\pi(1)}), \cdots, f_t(a_{\pi(n)}, b_{\pi(n)}),$$

is a symmetric Max-plus tropical rational function of the variables $a_1, b_1, \dots, a_n, b_n$, and

$$\lambda_{k,t}(a_1, b_1, \dots, a_n, b_n) = \sigma_k(f_t(a_1, b_1), \dots, f_t(a_n, b_n)) \odot \sigma_{k-1}(f_t(a_1, b_1), \dots, f_t(a_n, b_n))^{-1},$$

is also a symmetric tropical rational function of the variables $a_1, b_1, \dots, a_n, b_n$.

Since the mapping of persistence diagrams to persistent landscapes is reversible [9], the persistent landscape gives us a set of tropical rational functions $\lambda_{k,t}$, from which persistent diagrams can be reconstructed.

4. Construction of a feature vector in the space of barcodes

Aadcock A. et al. [14] defined an algebra of polynomials in barcode space that can be used as coordinates. Distances in barcode space are determined by matching intervals from one barcode to another and calculating penalties that involve taking maximums. For this reason, the tropical functions and are suitable, given the basic structure of the barcode space. We will represent the barcode with n intervals as a vector $(x_1, d_1, x_2, d_2, \dots, x_n, d_n)$, where $x_i \geq 0, \forall i$ denotes the left end of the i-th interval and d_i is its length.

Let's build two filtrations: scanning by rows from left to right and by columns from bottom to top. This adds spatial information to topological information. Information from Betti 0 and Betti 1 is used.

We will use six parameters – tropical features of the Max-plus type [11]:

$$\begin{aligned} p_1 &: \bigoplus_i d_i; \\ p_2 &: \bigoplus_{i<j} \left(d_i \odot d_j\right); \\ p_3 &: \bigoplus_{i<j<k} \left(d_i \odot d_j \odot d_k\right); \\ p_4 &: \bigodot_i d_i; \\ p_5 &: \bigodot_i x_i; \\ p_6 &: \bigodot_i \left(\left(\bigoplus_i y_i\right) - y_i\right). \end{aligned} \tag{5}$$

When applied to two sweeps, each with a 0-dimensional and a 1-dimensional barcode, yields a feature vector with a common size $2 \times 2 \times 6 = 24$.

Example 2 Since the image of the number 6 can be obtained from the image of the number 9 with the Euclidean transformation (rotation relative to the center of the image by π rad), the topological characteristics of these images are indistinguishable.

Let us find the topological characteristics of these images using tropical functions. Let's determine the image coordinates of the number 9:

along the x-axis: $q9_x = \left[\begin{array}{cccccccccccccccccccc} 1 & 1 & 2 & 3 & 4 & 4 & 4 & 4 & 4 & 4 & 4 & 3 & 2 & 1 & 1 & 1 & 1 & 2 & 3 & 4 \end{array}\right]$;

along the y-axis: $q9_y = \left[\begin{array}{cccccccccccccccccccc} 2 & 1 & 1 & 1 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 7 & 7 & 7 & 6 & 5 & 4 & 4 & 4 & 4 \end{array}\right]$.

Values of the left ends of the barcodes when scanning from left to right with $\beta_0 = 1, \beta_1 = 0 : x_1^{up} = 1; x_2^{up} = 0$.

Values of the left ends of barcodes when scanning from bottom to top with $\beta_0 = 1, \beta_1 = 0 : x_1^{up} = 1; x_2^{up} = 0$.

Values of the left ends of the barcodes when scanning from left to right with $\beta_0 = 0, \beta_1 = 1 : x_3^{right} = 1$.

Values of the left ends of barcodes when scanning from bottom to top with $\beta_0 = 0, \beta_1 = 1 : x_3^{up} = 4$.

Values of barcode lengths when scanning from left to right with $\beta_0 = 1, \beta_1 = 0 : d_1^{right} = 3; d_2^{right} = 4$.

Values of barcode lengths when scanning from bottom to top with $\beta_0 = 1, \beta_1 = 0 : d_1^{up} = 7; d_2^{up} = 0$.

Values of barcode lengths when scanning from left to right at with $\beta_0 = 0, \beta_1 = 1 : d_3^{right} = 4$.

Values of barcode lengths when scanning from bottom to top at with $\beta_0 = 0, \beta_1 = 1 : d_3^{up} = 4$.

Let's find the values of tropical parameters p_1, p_2, p_4, p_5 for the barcodes of the image of the number 9, when scanning from left to right with $\beta_0 = 1, \beta_1 = 0$: $p_1^{right,\beta_0} = \bigoplus_i d_i = 4$; $p_2^{right,\beta_0} = \bigoplus_{i<j} \left(d_i \odot d_j\right) = 7$; $p_4^{right,\beta_0} = \bigodot_i d_i = 7$; $p_5^{right,\beta_0} = \bigodot_i x_i = 2$.

Parameter values p_1, p_2, p_4, p_5, when scanning from bottom to top at $\beta_0 = 1, \beta_1 = 0$: $p_1^{up,\beta_0} = \underset{i}{\oplus} d_i = 7$; $p_2^{up,\beta_0} = \underset{i<j}{\oplus} \left(d_i \odot d_j\right) = 7$; $p_4^{up,\beta_0} = \underset{i}{\odot} d_i = 7$; $p_5^{up,\beta_0} = \underset{i}{\odot} x_i = 1$.

Parameter values p_1, p_2, p_4, p_5, when scanning from left to right with $\beta_0 = 0, \beta_1 = 1$: $p_1^{right,\beta_1} = \underset{i}{\oplus} d_i = 4$; $p_2^{right,\beta_1} = \underset{i<j}{\oplus} (d_i \odot d_j) = 4$; $p_4^{right,\beta_1} = \underset{i}{\odot} d_i = 4$; $p_5^{right,\beta_1} = \underset{i}{\odot} x_i = 1$.

Parameter values p_1, p_2, p_4, p_5, when scanning from bottom to top at $\beta_0 = 0, \beta_1 = 1$: $p_1^{up,\beta_0} = \underset{i}{\oplus} d_i = 4$; $p_2^{up,\beta_0} = \underset{i<j}{\oplus} (d_i \odot d_j) = 4$; $p_4^{up,\beta_0} = \underset{i}{\odot} d_i = 4$; $p_5^{up,\beta_0} = \underset{i}{\odot} x_i = 4$.

Let's determine the image coordinates of the number 6:
along the x-axis: $q6_x = \begin{bmatrix} 4 & 4 & 3 & 2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 3 & 4 & 4 & 4 & 4 & 3 & 2 & 1 \end{bmatrix}$;
along the y-axis: $q6_y = \begin{bmatrix} 6 & 7 & 7 & 7 & 7 & 6 & 5 & 4 & 3 & 2 & 1 & 1 & 1 & 1 & 2 & 3 & 4 & 4 & 4 & 4 \end{bmatrix}$.

Values of the left ends of the barcodes when scanning from left to right with with $\beta_0 = 1, \beta_1 = 0 \,:\, x_1^{right} = 1$; $x_2^{right} = 0$.

Values of the left ends of barcodes when scanning from bottom to top at with $\beta_0 = 1$, $\beta_1 = 0 \,:\, x_1^{up} = 1$; $x_2^{up} = 0$.

Values of the left ends of the barcodes when scanning from left to right with $\beta_0 = 0$, $\beta_1 = 1 \,:\, x_3^{right} = 4$.

Values of the left ends of barcodes when scanning from bottom to top with $\beta_0 = 0$, $\beta_1 = 1 \,:\, x_3^{up} = 4$.

Values of barcode lengths when scanning from left to right at with $\beta_0 = 1$, $\beta_1 = 0 \,:\, d_1^{right} = 4$; $d_2^{right} = 0$.

Values of barcode lengths when scanning from bottom to top at with $\beta_0 = 1$, $\beta_1 = 0 \,:\, d_1^{up} = 7$; $d_2^{up} = 0$.

Values of barcode lengths when scanning from left to right with $\beta_0 = 0$, $\beta_1 = 1 \,:\, d_3^{right} = 5$.

Values of barcode lengths when scanning from bottom to top with $\beta_0 = 0$, $\beta_1 = 1 \,:\, d_3^{up} = 8$.

Let's find the values of tropical parameters p_1, p_2, p_4, p_5 for the barcodes of the image of the number 6 , when scanning from left to right at $\beta_0 = 1, \beta_1 = 0$: $p_1^{right,\beta_0} = \underset{i}{\oplus} d_i = 4$; $p_2^{right,\beta_0} = \underset{i<j}{\oplus} \left(d_i \odot d_j\right) = 4$; $p_4^{right,\beta_0} = \underset{i}{\odot} d_i = 4$; $p_5^{right,\beta_0} = \underset{i}{\odot} x_i = 1$.

Parameter values p_1, p_2, p_4, p_5, when scanning from bottom to top at $\beta_0 = 1, \beta_1 = 0$: $p_1^{up,\beta_0} = \underset{i}{\oplus} d_i = 7$; $p_2^{up,\beta_0} = \underset{i<j}{\oplus} \left(d_i \odot d_j\right) = 7$; $p_4^{up,\beta_0} = \underset{i}{\odot} d_i = 7$; $p_5^{up,\beta_0} = \underset{i}{\odot} x_i = 1$.

Parameter values p_1, p_2, p_4, p_5, when scanning from left to right with $\beta_0 = 0, \beta_1 = 1$: $p_1^{right,\beta_1} = \underset{i}{\oplus} d_i = 5$; $p_2^{right,\beta_1} = \underset{i<j}{\oplus} (d_i \odot d_j) = 5$; $p_4^{right,\beta_1} = \underset{i}{\odot} d_i = 5$; $p_5^{right,\beta_1} = \underset{i}{\odot} x_i = 4$.

Parameter values p_1, p_2, p_4, p_5, when scanning from bottom to top at $\beta_0 = 0, \beta_1 = 1$: $p_1^{up,\beta_0} = \underset{i}{\oplus} d_i = 8$; $p_2^{up,\beta_0} = \underset{i<j}{\oplus} (d_i \odot d_j) = 8$; $p_4^{up,\beta_0} = \underset{i}{\odot} d_i = 8$; $p_5^{up,\beta_0} = \underset{i}{\odot} x_i = 4$.

Euclidean distance between the images of the number ”6” and the number ”9”:

$$\begin{aligned}
&\mathfrak{D}(\ll 6 \gg, \ll 9 \gg)^2 = \sum_i w_i \cdot (p_i^{rigt,\beta_0}(\ll 6 \gg) - p_i^{rigt,\beta_0}(\ll 9 \gg))^2 + \\
&\sum_i w_i \cdot (p_i^{up,\beta_0}(\ll 6 \gg) - p_i^{up,\beta_0}(\ll 9 \gg))^2 + \\
&\sum_i w_i \cdot (p_i^{rigt,\beta_1}(\ll 6 \gg) - p_i^{rigt,\beta_1}(\ll 9 \gg))^2 + \\
&\sum_i w_i \cdot (p_i^{up,\beta_1}(\ll 6 \gg) - p_i^{up,\beta_1}(\ll 9 \gg))^2,
\end{aligned}$$

$\mathfrak{D}\left(\ll 6 \gg, \ll 9 \gg\right) = 8.89$, with weight values: $w_i = 1, \forall i$. □

Conclusion

The paper considers mathematical models and functions for representing persistent landscape objects based on the persistent homology method. The persistent landscape functions allow you to map persistent diagrams to Hilbert space. The representations of topological functions in various machine learning models are considered. An example of finding the distance between images based on the construction of persistent landscape functions is given.

Based on the algebra of polynomials in the barcode space, the distances in the barcode space are determined by comparing the intervals from one barcode to another and calculating penalties. For these purposes, tropical functions are used that take into account the basic structure of the barcode space. Methods for constructing rational tropical functions are considered. An example of finding the distance between images based on the construction of tropical functions is given. To increase the variety of parameters, filtering of object scanning by rows from left to right and scanning by columns from bottom to top are constructed. This adds spatial information to topological information. The method of constructing persistent landscapes is compatible with the approach of constructing tropical rational functions when obtaining persistent homologies.

References

[1] G. Carlsson, “Topology and data”, *Bulletin of the American Mathematical Society*, vol. 46, no. 2, pp. 307–309, 2009. DOI: 10.1090/S0273-0979-09-01249-X.

[2] H. Edelsbrunner and J. Harer, *Computational topology: an introduction*. American Mathematical Soc., 2010.

[3] A. J. Zomorodian, *Topology for computing*. Cambridge UP, 2005.

[4] R. Ghrist, “Barcodes: the persistent topology of data”, *Bulletin of the American Mathematical Society*, vol. 15, no. 1, pp. 61–75, 2008. DOI: 10.1090/S0273-0979-07-01191-3.

[5] S. N. Chukanov, “Comparison of objects’ images based on computational topology methods”, *Informatics and Automation*, vol. 18, no. 3, pp. 1043–1065, 2019. DOI: 10.15622/sp.2019.18.5.1043-1065.

[6] S. N. Chukanov, “The Comparison of Diffeomorphic Images based on the Construction of Persistent Homology”, *Automatic Control and Computer Sciences*, vol. 54, no. 7, pp. 758–771, 2020. DOI: 10.3103/S0146411620070056.

[7] A. Hatcher, *Algebraic Topology*. Cambridge UP, 2005.

[8] C. Hofer, R. Kwitt, M. Niethammer, and A. Uhl, “Deep learning with topological signatures”, in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017, pp. 1633–1643.

[9] P. Bubenik, “The persistence landscape and some of its properties”, *Topological Data Analysis*, pp. 97–117, 2020. DOI: 10.1007/978-3-030-43408-3_4.

[10] P. Bubenik, “Statistical Topological Data Analysis Using Persistence Landscapes”, *Journal of Machine Learning Research*, vol. 16, no. 1, pp. 77–102, 2015.

[11] S. Kalisnik, “Tropical coordinates on the space of persistence barcodes”, *Foundations of Computational Mathematics*, vol. 19, no. 1, pp. 101–129, 2019. DOI: 10.1007/s10208-018-9379-y.

[12] R. Kwitt, S. Huber, M. Niethammer, W. Lin, and U. Bauer, “Statistical topological data analysis: a kernel perspective”, *Advances in Neural Information Processing Systems*, vol. 28, pp. 3052–3060, 2015.

[13] V. P. Maslov, “Motivation and essence of the term ”Tropical mathematics””, *Russian Journal of Mathematical Physics*, vol. 27, no. 4, pp. 478–483, 2020. DOI: 10.1134/S106192082004007X.

[14] A. Aadcock, E. Carlsson, and G. Carlsson, "The Ring of Algebraic Functions on Persistence Bar Codes", *Homology, Homotopy and Applicationse*, vol. 18, pp. 381–402, 2016. DOI: 10.4310/HHA.2016.v18.n1.a21.