Автоматика и телемеханика, № 3, 2021

Оптимизация, системный анализ и исследование операций

© 2021 г. Ю.С. ПОПКОВ, д-р техн. наук (popkov@isa.ru)
(Федеральный исследовательский центр
"Информатика и управление" РАН, Москва;
Институт проблем управления им. В.А. Трапезникова РАН, Москва),
Ю.А. ДУБНОВ (yury.dubnov@phystech.edu)
(Федеральный исследовательский центр
"Информатика и управление" РАН, Москва;
Национальный исследовательский университет
"Высшая школа экономики", Москва),
А.Ю. ПОПКОВ, канд. техн. наук (apopkov@isa.ru)
(Федеральный исследовательский центр
"Информатика и управление" РАН, Москва)

ЭНТРОПИЙНО-РАНДОМИЗИРОВАННОЕ ПРОЕКТИРОВАНИЕ[1]

Предложен новый метод рандомизированного проектирования, основанный на энтропийной оптимизации случайных матриц-проекторов (ERP-метод). Введено понятие индикатора компактности матрицы данных, который сохраняется в матрицах-проекциях. Сформулирован алгоритм ERP в виде задачи условной максимизации энтропийного функционала, заданного на функциях плотности распределения вероятностей матриц-проекторов. Рассмотрен дискретный вариант этой задачи, получены условия существования и единственности ее положительного решения. Развиты процедуры реализации энтропийно-оптимальных матриц-проекторов путем сэмплирования функций плотности распределения вероятностей (ПРВ).

Ключевые слова: случайное проектирование, сжатие и растяжение матрицы данных, матрицы-проекторы, индикатор компактности, сэмплирование функции плотности, дисперсионные множества, интерквартильные множества.

DOI: 10.31857/S0005231021030090

1. Введение

Процедуры машинного обучения, ориентированные на задачи классификации, кластеризации и прогнозирования, оперируют данными, представленными в виде матриц "объекты–признаки", из которых формируются обучающие и тестовые коллекции данных. В процессе формирования обучающей

[1] Работа выполнена при поддержке Российского фонда фундаментальных исследований (проект № 20-07-00470).

коллекции и в особенности в результате ее применения для решения упомянутых выше задач часто оказывается, что признаков недостаточно или слишком много, так же как и объектов. Другими словами, возникает необходимость в преобразованиях имеющейся матрицы данных в обучающую и тестовую коллекции.

Эти преобразования касаются как "сжатия", так и "растяжения" матрицы данных. В терминах "объекты–признаки" возникает семейство задач "сжатия": редукция количества признаков, редукция количества объектов и параллельная редукция того и другого; и семейство задач "растяжения": расширение пространства признаков, увеличение количества данных (условных "объектов") и параллельное увеличение того и другого. Если преобразования "сжатия" очень популярны в теории и практике машинного обучения, то преобразованиям "растяжения" уделяется не столь широкое внимание. Хотя исследователи, занимающиеся проблемами классификации и кластеризации, часто сталкивались с тем, что в указанных проблемах в маломерных пространствах признаков они неразрешимы или дают низкую точность, тогда как при увеличении размеров признакового пространства удается решить задачу с хорошей точностью. Аналогичные эффекты наблюдаются и при недостаточном количестве данных.

Проблемам "сжатия" матрицы данных посвящена обширная литература, некоторое представление о которой можно составить по обзорам [1–3] и монографиям [4–7].

Данная проблема вложена в более общую: приближение заданного набора многомерных точек маломерным аффинным многообразием [8]. Здесь следует отметить метод главных компонент (МГК) (Principal Component Analysis (PCA)) [9–11] и его робастные версии [12]. Следует отметить, что МГК применяется для матриц данных ограниченной размерности, так как вычисление собственных чисел является весьма трудоемкой задачей. Для многомерных задач применяют латентный семантический анализ ЛСА (Latent Semantic Analysis (LSA)), основанный на сингулярном разложении матрицы [13].

Редуцированные коллекции данных используются для решения различных задач машинного обучения. Поэтому было бы полезным в процедурах редукции учитывать проблемную ориентацию их использования. Этот эффект достигается в линейном дискриминационном анализе (ЛДА) (Linear Discriminant Analysis (LDA)) [14, 15].

Учет в процедурах редукции последующих применений "сжатых" матриц – тренд, который просматривается в задачах обработки больших объемов данных: распознавание по видеопотокам, классификация коллекций документов и др. В этих случаях применяется метод случайных проекций (RP), базирующийся на асимптотическом эффекте проецирования на пространство достаточно большой размерности, при котором в среднем сохраняется расстояние между точками (лемма Джонсона–Линденштраусса [16]). Экспериментально показано, что для генерации случайных матриц-проекторов может быть использовано стандартное нормальное распределение, а в некоторых случаях равномерное распределение [17–19].

Совсем иная ситуация, когда возникает необходимость в “растяжении” матрицы данных. Наиболее часто такая необходимость имеет место в задачах анализа таких данных, механизмы генерации которых неизвестны. Одной из форм представления данных являются многомерные временные ряды. Примером таких объектов являются процессы, происходящие в мозгу живых существ и, прежде всего, человека, макро-проявления которых фиксируются в виде электрических сигналов ЭЭГ [20]. Аналогичная ситуация имеет место при отображении биржевой деятельности в виде временных рядов котировок акций, валют, товаров и финансовых инструментов [21]. В последнее время весьма популярными становятся исследования электорального поведения населения во время выборов, предсказание которого осуществляется путем анализа ретроспективной динамики процессов выборов, отображенной в соответствующих временных рядах [22].

В этих работах применялись так называемые модельные конструкции временных рядов, а формирование признакового пространства осуществлялось последовательным отбором признаков с привлечением методов фильтрации [23], условной энтропии [24], взаимной информации “признак–класс” [25], кросс-энтропии Кульбака–Ляйблера [26].

Однако в некоторых задачах, частично упомянутых выше, использовать модель временного ряда трудно или невозможно из-за отсутствия знания о механизмах его происхождения. Поэтому важными являются попытки отображения компонент временного ряда в конечный набор чисел — признаков многомерного временного ряда. Полагая, что многомерный временной ряд кодирует свойства объекта, последнее означает, что формируется признаковое пространство объекта.

В [27] была предложена процедура безмодельного отображения временного ряда в конечную совокупность характеризующих его чисел, основанную на теории ε-сложности. В дальнейшем [28, 29] эта процедура была тестирована на задачах бинарной классификации одного класса ЭЭГ-сигналов в области психиатрии. Результаты этих исследований показали, с одной стороны, что технология ε-сложности может служить основой для безмодельной классификации временных рядов, а с другой — что для достижения удовлетворительной точности классификации возникает необходимость в увеличении размерности признакового пространства.

В данной работе развивается новый метод “сжатия” и “растяжения” матрицы с не вполне достоверными данными, основанный на построении рандомизированных матриц-проекторов (ERP-метод). Развит алгоритм оптимизации функции плотности распределения вероятностей (ПРВ) матриц-проекторов с учетом сохранения *индикатора компактности* матрицы данных. Алгоритм сформулирован в терминах условной максимизации энтропийного функционала, определенного на функциях ПРВ матриц-проекторов. При этом оптимальные оценки функций ПРВ соответствуют условиям максимальной неопределенности. Реализация рандомизированных матриц-проекторов осуществляется путем сэмплирования оптимальных ПРВ с использованием bootstrap-процедуры и метода Монте-Карло [30, 31]. Рассмотрены задачи рандомизированного проектирования с фиксированными элементами множества матриц-проекторов, на котором определена дискретная функция распределения вероятностей. Сформулирован алгоритм ее оптимизации и выбора полезных числовых характеристик множества случайных матриц-проекторов.

2. Формулировки задач энтропийно-рандомизированного проектирования

1. Точечно-множественное представление матрицы данных: индикатор компактности

Рассмотрим матрицу данных "объекты–признаки":

$$U_{(m\times s)} = \begin{pmatrix} \mathbf{u}^{(1)} \\ \cdots \\ \mathbf{u}^{(m)} \end{pmatrix}, \qquad \mathbf{u}^{(i)} = \left\{u_1^{(i)}, \dots, u_s^{(i)}\right\} \in R^s, \quad i = \overline{1,m}. \tag{2.1}$$

Здесь m — количество объектов, s — количество признаков, R^s — признаковое пространство. Будем полагать, что элементы матрицы данных стандартизованы, т.е. $u_{ij} \in [0,1]$, $i = \overline{1,m}$, $j = \overline{1,s}$.

Точечно-множественным отображением матрицы $U_{(m\times s)}$ *в признаковом пространстве* $R^s_{(m)}$ *будем называть множество* $\mathfrak{U}^s_m$ *точек* $\left\{\mathbf{u}^{(1)}, \dots, \mathbf{u}^{(m)}\right\}$.

В качестве характеристики разброса точек в множестве $\mathfrak{U}^s_m$ определим *индикатор компактности* ρ_U как среднее арифметическое расстояний между точками[2]:

$$\rho_U = \frac{2}{m(m-1)} \sum_{(\alpha,\beta)=1,\, \alpha<\beta}^{m} \varrho\left(\mathbf{u}^{(\alpha)}, \mathbf{u}^{(\beta)}\right), \tag{2.2}$$

где $\varrho\left(\mathbf{u}^{(\alpha)}, \mathbf{u}^{(\beta)}\right)$ — расстояние между точками $\mathbf{u}^{(\alpha)}, \mathbf{u}^{(\beta)}$. Если элементы матрицы данных стандартизованы и выбрано евклидово расстояние, то

$$0 \leqslant \varrho\left(\mathbf{u}^{(\alpha)}, \mathbf{u}^{(\beta)}\right) \leqslant \sqrt{2} \qquad \text{для всех } (\alpha,\beta) = \overline{1,m},$$

$$0 \leqslant \rho_U \leqslant 1. \tag{2.3}$$

Определим следующие классы трансформаций матрицы данных:

- (mr)-*класс*: $U_{(m\times s)} \to Y_{(m\times r)}$, множество $\mathfrak{Y}^r_m$ в пространстве $R^r_{(m)}$;
- (ns)-*класс*: $U_{(m\times s)} \to B_{(n\times s)}$, множество $\mathfrak{B}^s_n$ в пространстве $R^s_{(n)}$;
- (nr)-*класс*: $U_{(m\times s)} \to Z_{(n\times r)}$, множество $\mathfrak{Z}^r_n$ в пространстве $R^r_{(n)}$.

[2] Могут быть выбраны и другие индикаторы, например максимальное расстояние между точками и др.

2. Случайные матрицы-проекторы и их вероятностные характеристики

1. Проектирование на пространство $R^r_{(m)}$ (генерация множества $\mathfrak{Y}^r_m$).

Эту операцию будем осуществлять с помощью матрицы-проектора $Q_{(s\times r)}$:

$$
Y_{(m\times r)} = \begin{pmatrix} \mathbf{y}^{(1)} \\ \cdots \\ \mathbf{y}^{(m)} \end{pmatrix} = U_{(m\times s)} Q_{(s\times r)};
$$
$$
y_j^{(i)} = \sum_{k=1}^{s} u_{ik} q_{kj}, \quad i = \overline{1,m}; \quad j = \overline{1,r}. \tag{2.4}
$$

Матрица-проектор $Q_{(s\times r)}$ случайная интервального типа, т.е.

$$
Q_{(s\times r)} \in \mathcal{Q} = [Q^-, Q^+]. \tag{2.5}
$$

Вероятностные свойства матрицы-проектора характеризуются непрерывно-дифференцируемой функцией плотности распределений вероятностей (ПРВ) $P(Q)$, определенной на $\mathcal{Q}$.

Векторы $\mathbf{y}^{(i)}$ образуют множество $\mathcal{Y}^r_m$ из m r-мерных точек. Определим расстояние между ними $\varrho(\mathbf{y}^{(\alpha)}, \mathbf{y}^{(\beta)} \,|\, Q_{(s\times r)})$ и индикатор компактности

$$
\rho_Y(Q) = \frac{2}{m(m-1)} \sum_{(\alpha,\beta)=1,\, \alpha<\beta}^{m} \varrho\left(\mathbf{y}^{(\alpha)}, \mathbf{y}^{(\beta)} \,|\, Q_{(s\times r)}\right). \tag{2.6}
$$

Поскольку значения элементов матрицы-проектора — случайные числа, то среднее расстояние (2.6) принимает случайные значения.

Определим математическое ожидание индикатора компактности:

$$
\bar{\rho}_Y = \int_{\mathcal{Q}} P(Q) \rho_Y(Q) dQ. \tag{2.7}
$$

2. Проектирование на пространство $R^{(n\times s)}$ (генерация множества $\mathfrak{B}^s_n$).

Эту операцию осуществим с помощью матрицы-проектора $T_{(n\times m)}$ со случайными элементами:

$$
B_{(n\times s)} = \begin{pmatrix} \mathbf{b}^{(1)} \\ \cdots \\ \mathbf{b}^{(n)} \end{pmatrix} = T_{(n\times m)} U_{(m\times s)};
$$
$$
b_j^{(i)} = \sum_{k=1}^{m} t_{ik}\, u_{kj}, \quad i = \overline{1,n}; \quad j = \overline{1,s}. \tag{2.8}
$$

Матрица-проектор $T_{(n\times m)}$ случайная интервального типа, т.е.

$$
T_{(n\times m)} \in \mathcal{T} = [T^-, T^+]. \tag{2.9}
$$

Ее вероятностные свойства характеризуются функцией ПРВ $W(T)$, определенной на $\mathcal{T}$.

Векторы $\mathbf{b}^{(i)}$ образуют множество $\mathfrak{B}_n^s$ из s-мерных точек в количестве $n < m$. Определим индикатор компактности в виде

$$\rho_B(T) = \frac{2}{n(n-1)} \sum_{(\alpha,\beta),\, \alpha<\beta}^{n} \varrho(\mathbf{b}^{(\alpha)}, \mathbf{b}^{(\beta)} \,|\, T) \tag{2.10}$$

и его математическое ожидание:

$$\bar{\rho}_B = \int_{\mathcal{T}} W(T)\rho_B(T)dT. \tag{2.11}$$

3. Проектирование на пространство $R_{(n)}^r$ *(генерация множества* $\mathfrak{Z}_n^r$*).*

Эту операцию будем осуществлять с помощью случайных матриц-проекторов $T_{n\times m}$ и $Q_{(s\times r)}$:

$$Z_{(n\times r)} = \begin{pmatrix} \mathbf{z}^{(1)} \\ \cdots \\ \mathbf{z}^{(n)} \end{pmatrix} = T_{(n\times m)} U_{(m\times s)} Q_{(s\times r)}; \tag{2.12}$$

$$z_j^{(i)} = \sum_{p=1}^{m} t_{ip} \sum_{k=1}^{s} u_{pk}\, q_{kj}, \qquad i = \overline{1,n}; \quad j = \overline{1,r}.$$

Указанные матрицы-проекторы — интервального типа:

$$\begin{aligned} &T_{(n\times m)} \in \mathcal{T} = [T^-, T^+], \\ &Q_{(s\times r)} \in \mathcal{Q} = [Q^-, Q^+]. \end{aligned} \tag{2.13}$$

Вероятностные свойства матриц-проекторов характеризуются функцией совместного ПРВ $F(T,Q)$, определенной на множестве

$$\mathcal{F} = \mathcal{T} \bigotimes \mathcal{Q}. \tag{2.14}$$

Векторы $\mathbf{z}^{(i)}$ образуют множество $\mathfrak{Z}_n^r$ из r-мерных точек. Индикатор компактности этого множества:

$$\rho_Z(T,Q) = \frac{2}{n(n-1)} \sum_{(\alpha,\beta),\, \alpha<\beta}^{n} \varrho(\mathbf{z}^{\alpha}, \mathbf{z}^{\beta} \,|\, T, Q). \tag{2.15}$$

Математическое ожидание индикатора компактности имеет вид

$$\bar{\rho}_Z = \int_{\mathcal{F}} F(T,Q)\rho_Z(T,Q)dTdQ. \tag{2.16}$$

Соотношения между размерами (m,s) матрицы данных и размерами $(m,r),(n,s),(n,r)$ матриц-проекций характеризуют ее “сжатие” или “растяжение”: если $r < s$, $n < m$, то “сжатие”; если $r > s$, $n > m$, то “растяжение”.

3. Алгоритмы энтропийно-рандомизированного проектирования

Рассмотрим проектирование на пространство $R^r_{(n)}$, которое осуществляет трансформацию матрицы данных одновременно по двум ее параметрам.

Алгоритм энтропийно-рандомизированного проектирования (*ERP-алгоритм*) представляет собой MEE-оценку функции ПРВ $F(T,Q)$ с включенным в нее условием балансирования (равенства) индексов компактности матриц данных $U_{(m\times s)}$ и проекции $Z_{(n\times r)}$.

MEE-оценка [32] функции $F(B,Q)$ определяется следующим алгоритмом:

$$\mathcal{H}[F(T,Q)] = -\int\limits_{\mathcal{F}} F(T,Q)\,\ln\,F(T,Q)\,dTdQ \Rightarrow \max, \tag{3.1}$$
$$\int\limits_{\mathcal{F}} F(T,Q)dTdQ = 1,$$
$$\int\limits_{\mathcal{F}} F(T,Q)\rho_Z(T,Q)dTdQ = \rho_U.$$

Задача (3.1) является функциональной задачей на условный экстремум ляпуновского типа: все их компоненты являются интегральными функционалами [33].

Интегральная форма этой задачи позволяет использовать для формулировки условий оптимальности функционал Лагранжа со скалярными множителями Лагранжа:

$$\mathcal{L}[F(T,Q)] = \mathcal{H}[F(T,Q)] + \mu\left(1 - \int\limits_{\mathcal{F}} F(T,Q)dTdQ\right) +$$
$$+\,\lambda\left(\rho_U - \int\limits_{\mathcal{F}} F(T,Q)\rho_Z(T,Q)dTdQ\right). \tag{3.2}$$

В терминах производных Гато [33] ($\mathfrak{G}_F$) условия стационарности функционала Лагранжа приобретают следующий вид:

$$\mathfrak{G}_F\mathcal{L}[F(T,Q)] = 0. \tag{3.3}$$

Энтропийно-оптимальная функция ПРВ матриц-проекторов имеет вид [32]

$$F^*(T,Q\,|\,\lambda) = \frac{\exp\left(-\lambda\ \rho_Z(T,Q)\right)}{\mathbb{F}(\lambda)}, \tag{3.4}$$
$$\mathbb{F}(\lambda) = \int\limits_{\mathcal{F}} \exp\left(-\lambda\ \rho_Z(T,Q)\right)dTdQ.$$

Из выражений (3.4) видно, что структура функции ПРВ зависит от выбранной функции расстояний (2.15) и не зависит от параметра λ, определяемого из уравнения баланса индикаторов компактности:

$$\Phi(\lambda) = \int\limits_{\mathcal{F}} \exp\left(-\lambda\, \rho_Z(T,Q)\right)\left(\rho_Z(T,Q) - \rho_U\right) dT dQ = 0. \tag{3.5}$$

Поскольку параметр λ является масштабным, можно использовать его приближенное значение, получаемое из линеаризованного уравнения (3.5):

$$\tilde{\lambda} = \frac{\int_{\mathcal{F}} \left(\rho_Z(T,Q) - \rho_U\right) dT dQ}{\int_{\mathcal{F}} \rho_Z(T,Q)\left(\rho_Z(T,Q) - \rho_U\right) dT dQ}. \tag{3.6}$$

Пример 1. Пусть матрица данных и индикаторы компактности, построенные на евклидовом и манхэттенском расстояниях, имеет вид

$$U_{(3\times 2)} = \begin{pmatrix} 0{,}1 & 0{,}8 \\ 0{,}5 & 0{,}4 \\ 0{,}9 & 0{,}2 \end{pmatrix}; \qquad \rho_U^E = 0{,}506, \quad \rho_U^M = 0{,}933.$$

Рассмотрим проектирование на пространство $R^2_{(1)}$:

$$Z_{(2\times 1)} = \mathbf{z}^{(2)} = T_{(2\times 3)} U_{(3\times 2)} \mathbf{q}(2),$$

где

$$\mathcal{T} = [T^-, T^+], \quad T^- = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \quad T^+ = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix};$$

$$\mathcal{Q} = [\mathbf{q}^-, \mathbf{q}^+], \quad \mathbf{q}^- = \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \quad \mathbf{q}^+ = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Вектор $\mathbf{z}^{(2)}$ имеет следующие компоненты:

$$z_1 = q_1 \sum_{j=1}^{3} t_{1j} u_{j1} + q_2 \sum_{j=1}^{3} t_{1j} u_{j2}, \quad z_2 = q_1 \sum_{j=1}^{3} t_{2j} u_{j1} + q_2 \sum_{j=1}^{3} t_{2j} u_{j2}.$$

Рассмотрим два вида расстояний между элементами вектора $\mathbf{z}$:

— евклидово

$$\rho_Z^E(T,\mathbf{q}) = (z_1 - z_2)^2 = \left[q_1 \sum_{j=1}^{3} u_{j1}(t_{1j} - t_{2j}) - q_2 \sum_{j=1}^{3} u_{j2}(t_{1j} - t_{2j}) \right]^2;$$

— манхэттенское

$$\rho_Z^M(T,\mathbf{q}) = |z_1 - z_2| = \left| q_1 \sum_{j=1}^{3} u_{j1}(t_{1j} - t_{2j}) - q_2 \sum_{j=1}^{3} u_{j2}(t_{1j} - t_{2j}) \right|.$$

Энтропийно-оптимальные ПРВ случайных матриц-проекторов (3.4) имеют вид:

— для евклидового расстояния

$$F_E^*(T,\mathbf{q}\,|\,\lambda) = \frac{\exp\left(-\lambda\ \rho_Z^E(T,\mathbf{q})\right)}{\mathbb{F}(\lambda)}, \tag{3.7}$$

$$\mathbb{F}_E(\lambda) = \int\limits_{\mathcal{F}} \exp\left(-\lambda\,\rho_Z^E(T,\mathbf{q})\right)\, dT\, d\mathbf{q};$$

— для манхэттенского расстояния

$$F_M^*(T,\mathbf{q}\,|\,\lambda) = \frac{\exp\left(-\lambda\ \rho_Z^M(T,\mathbf{q})\right)}{\mathbb{F}(\lambda)}, \tag{3.8}$$

$$\mathbb{F}_M(\lambda) = \int\limits_{\mathcal{F}} \exp\left(-\lambda\,\rho_Z^M(T,\mathbf{q})\right)\, dB\, d\mathbf{q}.$$

Параметр λ определяется из следующих уравнений для евклидова и манхэттенского расстояний соответственно:

$$\int\limits_{\mathcal{F}} \exp\left(-\lambda\,\rho_Z^E(T,\mathbf{q})\right)\left(\rho_Z^E(T,\mathbf{q}) - \delta\,\rho_U^E\right)\, dT\, d\mathbf{q} = 0, \tag{3.9}$$

$$\int\limits_{\mathcal{F}} \exp\left(-\lambda\,\rho_Z^M(T,\mathbf{q})\right)\left(\rho_Z^M(T,\mathbf{q}) - \delta\,\rho_U^M\right)\, dT\, d\mathbf{q} = 0.$$

В равенствах (3.7)–(3.9)

$$\mathcal{F} = \mathcal{T} \bigotimes \mathcal{Q}.$$

Качественное представление о морфологии многомерных функций ПРВ (3.7), (3.8) дают некоторые их сечения.

а) *матрица-проектор T задана*:

$$T^* = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Соответствующие сечения ненормированных функций (3.7), (3.8) имеют вид:

$$F_E^*(T^*,\mathbf{q}) \sim \exp\left[-\lambda\,(0,41q_1 - 0,28q_2)^2\right), \quad (q_1,q_2) \in [-1,1],$$
$$F_M^*(T^*,\mathbf{q}) \sim \exp\left[-\lambda\,|0,41q_1 - 0,28q_2|\right), \quad (q_1,q_2) \in [-1,1].$$

б) *матрица-проектор T и вектор-проектор $\mathbf{q}^{(2)}$ имеют вид*:

$$T = \begin{pmatrix} 0 & 1 & 0 \\ t_{21} & 0 & 1 \end{pmatrix}, \qquad \mathbf{q}^{(2)} = \begin{pmatrix} q_1 \\ -1 \end{pmatrix}.$$

Соответствующие сечения ненормированных функций (3.7), (3.8) имеют вид:

$$F_E^*(T^*, \mathbf{q}) \sim \exp\left[-\lambda\left(-0{,}1q_1t_{21} - 0{,}4q_1 + 0{,}8t_{21} + 0{,}2\right)^2\right), \qquad q_1 \in [-1, 1], \quad t_{21} \in [0, 1],$$

$$F_M^*(B^*, \mathbf{q}) \sim \exp\left[-\lambda\left| -0{,}1q_1t_{21} - 0{,}4q_1 + 0{,}8t_{21} + 0{,}2\right|\right), \qquad q_1 \in [-1, 1], \quad t_{21} \in [0, 1].$$

Заметим, что полученные ПРВ случайных проекторов энтропийно-оптимальны для соответствующей матрицы данных и существенно отличаются от нормального распределения, рекомендуемого для применения при случайном проектировании (см., например [19]).

4. Реализация случайных проекторов и их числовых характеристик

Стандартный способ реализации случайных проекторов состоит в сэмплировании энтропийно-оптимальных ПРВ $P(Q), W(T), F(T, Q)$, при которых сохраняется межвекторное расстояние матрицы данных в ее матрицах-проекциях. В результате сэмплирования ПРВ трансформируются в последовательность случайных матриц-проекторов:

$$F^*(T, Q) \Rightarrow \left(T^{(1)}, Q^{(1)}\right), \dots, \left(T^{(N)}, Q^{(N)}\right), \dots \tag{4.1}$$

Общий метод генерации случайных последовательностей с заданной функцией ПРВ изложен в [34]. Свойство сохранения указанного выше баланса может не выполняться для каждого в отдельности элемента этой последовательности, но “в среднем” оно выполняется. Поэтому в качестве реализуемого эмпирического проектора можно принять матрицы

$$\bar{T} \approx \frac{1}{N} \sum_{i=1}^{N} T^{(i)}, \qquad \bar{Q} \approx \frac{1}{N} \sum_{i=1}^{N} Q^{(i)}. \tag{4.2}$$

Точные значения “средних” матриц-проекторов определяются следующими равенствами:

$$\bar{T} = \int\limits_{\mathcal{F}} T\, F(T, Q) dT dQ, \qquad \bar{Q} = \int\limits_{\mathcal{F}} Q\, F(T, Q) dT dQ. \tag{4.3}$$

Хотя интегралы в этих равенствах многомерные, но подынтегральные функции представляют собой экспоненты от полиномов. Это часто позволяет получить аналитические формулы для вычисления многомерных интегралов.

Кроме того проблему вычисления многомерных интегралов можно решить приближенными методами: методом Монте-Карло и методами полиномиальной аппроксимации подынтегральных функций.

Функция ПРВ $F^*(T, Q)$ определена на ограниченном множестве $\mathcal{F}$ (2.14). Поэтому в качестве реализуемого проектора может быть использована пара матриц

$$(T^*, Q^*) = \arg\max_{\mathcal{F}} F(T, Q). \tag{4.4}$$

5. Случайные матрицы-проекторы с заданными значениями элементов

1. Множество $(0,1)$-матриц-проекторов и его упорядочение

Рассмотрим процедуру проектирования на пространство $R^r_{(m)}$, осуществляемого матрицей-проектором $Q_{(s\times r)}$ со случайными элементами, принимающими значения 0 или 1.

Представим матрицу вектором-строкой длины sr. Количество различных последовательностей из 0 и 1, т.е. количество различных векторов-строк равно $N = 2^{rs}$. В качестве примера для $s = 3$, $r = 1$ таких векторов будет 8:

$$000, 100, 010, 001, 110, 011, 101, 111;$$

для $s = 4$, $r = 1$ их будет 16:

$$\begin{gathered} 0000, 1000, 0100, 0010, 0001, 1100, 0110, 0011, \\ 1001, 0101, 1010, 1111, 1110, 0111, 1011, 1101. \end{gathered}$$

Для матрицы-проектора $Q_{(s\times r)}$ существует конечное множество $\mathbb{Q}$ ее $(0,1)$-реализаций:

$$\mathbb{Q} = \{Q^{(1)}, \ldots, Q^{(N)}\}, \qquad N = 2^{sr}. \tag{5.1}$$

Полагая, что реализации случайные, их вероятностные свойства будем характеризовать функцией дискретного распределения вероятностей (ДРВ) $W(\alpha)$ с дискретным носителем $\alpha \in \mathcal{A} = [1, N]$, где

$$W(\alpha) = \mathbf{w} = \{w_\alpha\, \alpha = \overline{1, N}\}. \tag{5.2}$$

Матрица-проекция

$$\begin{gathered} Y^{(\alpha)}_{(m\times r)} = \begin{pmatrix} \mathbf{y}^{(\alpha)}_{(1)} \\ \vdots \\ \mathbf{y}^{(\alpha)}_{(m)} \end{pmatrix} = U_{(m\times s)} Q^{(\alpha)}_{(s\times r)}; \\ y^{(\alpha,i)}_j = \sum_{k=1}^{s} u_{ik} q^{(\alpha)}_{kj}, \quad i = \overline{1, m}; \quad j = \overline{1, r}. \end{gathered} \tag{5.3}$$

Индикатор компактности множества $\mathfrak{Y}^{(\alpha)}$:

$$\rho^{(\alpha)}_Y(Q) = \frac{2}{m(m-1)} \sum_{(\kappa,\beta)=1,\, \kappa<\beta}^{m} \varrho^{(\alpha)} \left(\mathbf{y}^{(\kappa)}, \mathbf{y}^{(\beta)} \,|\, Q^{(\alpha)}_{(s\times r)}\right). \tag{5.4}$$

Упорядочим последовательность (5.1) в соответствии со следующей последовательностью неравенств:

$$0 < \rho^{(1)}_Y(Q) < \rho^{(2)}_Y(Q) < \cdots < \rho^{(N)}_Y(Q). \tag{5.5}$$

2. МЕЕ-оценка функций ДРВ

МЕЕ-оценка функции ДРВ $W(\alpha)$ определяется следующим алгоритмом:

$$H(\mathbf{w}) = -\sum_{\alpha=1}^{N} w_\alpha \ln w_\alpha \Rightarrow \max \tag{5.6}$$

при ограничениях:

— нормировка

$$\sum_{\alpha=1}^{N} w_\alpha = 1, \tag{5.7}$$

— баланс индикаторов компактности

$$\sum_{\alpha=1}^{N} w_\alpha\, \rho_Y^{(\alpha)}(Q) = \rho_U. \tag{5.8}$$

Задача (5.6)–(5.8) является конечномерной задачей на условный экстремум с вогнутой целевой функцией (информационной энтропией) и линейными ограничениями-равенствами.

Рассмотрим функцию Лагранжа

$$L(\mathbf{w}, \lambda) = H(\mathbf{w}) + \mu\left(1 - \sum_{\alpha=1}^{N} w_\alpha\right) + \lambda\left(\rho_U - \sum_{\alpha=1}^{N} w_\alpha \rho_Y^{(\alpha)}(Q)\right). \tag{5.9}$$

Из условий стационарности этой функции получаем, что энтропийно-оптимальное распределение вероятностей имеет вид

$$w_\alpha^* = \frac{\exp\left(-\lambda \rho_Y^{(\alpha)}(Q)\right)}{\sum\limits_{\alpha=1}^{N} \exp\left(-\lambda \rho_Y^{(\alpha)}(Q)\right)}, \qquad \alpha = \overline{1, N}, \tag{5.10}$$

где параметр λ определяется из следующего уравнения:

$$\sum_{\alpha=1}^{N} \exp\left(-\lambda \rho_Y^{(\alpha)}(Q)\right)\left(\rho_Y^{(\alpha)}(Q) - \rho_U\right) = 0, \qquad -\infty < \lambda < \infty. \tag{5.11}$$

Воспользуемся заменой $z = \exp(-\lambda)$. Тогда уравнение (5.11) примет вид

$$\varphi(z) = \sum_{\alpha=1}^{N} z^{h_\alpha} c_\alpha = 0, \quad z \geqslant 0; \quad c_\alpha = (h_\alpha - \rho_U), \quad h_\alpha = \rho_Y^{(\alpha)}. \tag{5.12}$$

Теорема 1. Пусть существует $\alpha = l$ такое, что

$$c_\alpha < 0 \quad \text{для } \alpha = 1, \dots, l,$$
$$c_\alpha \geqslant 0 \quad \text{для } \alpha = l+1, \dots, N, \tag{5.13}$$

и согласно (5.5)

$$0 < h_1 < h_2 < \dots < h_l < h_{l+1} < \dots < h_N. \tag{5.14}$$

Тогда уравнение (5.12) *имеет единственное положительное решение* $z^* > 0$.

Доказательство приведено в Приложении.

3. Стратегии выбора "подходящей" матрицы-проекции из $\mathbb{Q}$ (5.1)

Энтропийно-оптимальная ДРВ $W^*(\alpha) = w^*_\alpha$, $\alpha = \overline{1, N}$ (5.12) характеризует вероятности реализации элементов (матриц-проекторов) из множества $\mathbb{Q}$ (5.1). Но для формирования конкретной матрицы-проекции $Y_{(m\times r)}$ необходима конкретная матрица-проектор из множества $\mathbb{Q}$. Можно предложить следующие стратегии для осуществления указанного выбора.

1. Матрица-проекция с максимальной вероятностью. Пусть в (5.5) разность $\Delta_\rho = \rho_Y^{(N)} - \rho_Y^{(1)}$ достаточно велика. Это означает, что в ДРВ (5.12) максимальное значение вероятности существенно больше минимального. Тогда реализуемая матрица-проектор имеет номер

$$\alpha^* = \arg \max_{1 \leqslant \alpha \leqslant N} w^*_\alpha. \tag{5.15}$$

2. Средняя матрица-проектор. В случае, когда $\Delta_\rho = \rho_Y^{(N)} - \rho_Y^{(1)}$ мала, возрастание последовательности $\{\rho^{(\alpha)}\}$ происходит с небольшой скоростью. Поэтому целесообразнее использовать осреднение для определения реализуемой матрицы-проектора:

$$\tilde{\alpha} = \sum_{\alpha=1}^{N} \alpha w^*_\alpha, \qquad \alpha^* = \lfloor \tilde{\alpha} \rfloor, \tag{5.16}$$

где $\lfloor \bullet \rfloor$ — целая часть числа $\bullet$.

3. Медианная матрица-проектор определяется следующим равенством:

$$\sum_{\alpha=1}^{\alpha^*} w^*_\alpha = \sum_{\alpha=\alpha^*+1}^{N} w^*_\alpha. \tag{5.17}$$

4. Осреднение матриц-проекций. Данная стратегия состоит из следующих этапов. На первом этапе выделяются подмножества матриц-проекторов, обладающие определенными свойствами. На втором — формируется подмножество соответствующих им матриц-проекций. Затем вычисляется средняя матрица-проекция.

Если второй и третий этапы связаны со стандартными для математической статистики вычислениями, то первый требует дополнительных комментариев. Выделение подмножеств матриц-проекторов основано на правилах, которые отображают желаемые свойства этих подмножеств. В зависимости от правила выделения подмножества матриц-проекторов будем различать дисперсионные и интерквартильные подмножества.

Рассмотрим *дисперсионные подмножества*, для чего вычислим среднеквадратичное значение (с.к.з.) переменной α:

$$\sigma^*_\alpha = \lfloor \tilde\sigma \rfloor, \qquad \tilde\sigma = \sqrt{D_\alpha}, \quad D_\alpha = \sum_{\alpha=1}^{N} (\alpha^* - \alpha)^2 w_\alpha. \tag{5.18}$$

Определим дисперсионный интервал

$$\mathcal{I}_\alpha = \Big[\alpha^* - \sigma^*_\alpha, \alpha^* + \sigma^*_\alpha\Big]. \tag{5.19}$$

Дисперсионное подмножество:

$$\tilde{\mathbb{Q}} = \Big\{Q^{(\alpha^*-\sigma^*_\alpha)}, \ldots, Q^{(\alpha^*+\sigma^*_\alpha)}\Big\}. \tag{5.20}$$

Для каждого элемента этого подмножества вычисляются соответствующие матрицы-проекции:

$$\tilde{\mathcal{Y}} = \Big\{Y^{(\alpha^*-\sigma^*_\alpha)}_{(m\times r)}, \ldots, Y^{(\alpha^*+\sigma^*_\alpha)}_{(m\times r)}\Big\}. \tag{5.21}$$

На последнем этапе вычисляется средняя по дисперсионному множеству матрица-проекция:

$$\bar Y^D_{(m\times r)} = \frac{1}{2\sigma^*_\alpha} \sum_{\alpha=\alpha^*-\sigma^*_\alpha}^{\alpha^*+\sigma^*_\alpha} Y^{(\alpha)}_{(m\times r)}. \tag{5.22}$$

Рассмотрим процедуру формирования *интерквартильных подмножеств*. Энтропийно-оптимальная функция распределения вероятностей, согласно (5.10), имеет вид

$$F^*(\beta) = \sum_{\alpha=1}^{\beta} w^*_\alpha, \qquad \beta \in [0, N]. \tag{5.23}$$

Напомним, что κ^*-квартилью случайной переменной α является множество $\mathcal{A}^{(\kappa^*)}$ ее значений $[1, \beta^*]$, где β^* есть целая часть решения уравнения $F^*(\beta) = \kappa^*$.

Рассмотрим два квартильных множества: $\mathcal{A}^{(\kappa^*_1)}$ и $\mathcal{A}^{(\kappa^*_2)}$, причем $\kappa^*_1 > \kappa^*_2$. Каждому из этих множеств соответствуют подмножества матриц-проекторов:

$$\mathbb{Q}(\kappa^*_1) = \Big\{Q^{(1)}, \ldots, Q^{(\alpha(\kappa^*_1))}\Big\},$$
$$\mathbb{Q}(\kappa^*_2) = \Big\{Q^{(1)}, \ldots, Q^{(\alpha(\kappa^*_2))}\Big\}, \tag{5.24}$$

где $\alpha(\kappa)$ – количество элементов в подмножестве $\mathbb{Q}(\kappa)$. Образуем подмножество матриц-проекторов

$$\mathbb{Q}(\kappa_1^* - \kappa_2^*) = \left\{Q^{(\alpha(\kappa_2^*))}, \ldots, Q^{(\alpha(\kappa_1^*))}\right\}, \tag{5.25}$$

которое является $(\kappa_1^* - \kappa_2^*)$-интерквантильным подмножеством $\mathcal{I}^{(\kappa_1^* - \kappa_2^*)}$ энтропийно-оптимальных матриц-проекторов. Количество элементов в этом подмножестве равно $M_{(\kappa_1^* - \kappa_2^*)} = \alpha(\kappa_1^*) - \alpha(\kappa_2^*)$.

Для каждого элемента подмножества $\mathcal{Q}(\kappa_1^* - \kappa_2^*)$ (5.25) вычисляются соответствующие матрицы-проекции

$$\mathcal{Y}(\kappa_1^* - \kappa_2^*) = \left\{Y_{(m\times r)}^{\alpha(\kappa_2^*)}, \ldots, Y_{(m\times r)}^{\alpha(\kappa_1^*)}\right\}. \tag{5.26}$$

Средняя по подмножеству $\mathcal{I}^{(\kappa_1^* - \kappa_2^*)}$ матрица-проекция:

$$\bar{Y}_{(m\times r)}^{K} = \frac{1}{M_{(\kappa_1^* - \kappa_2^*)}} \sum_{\alpha=\alpha(\kappa_2^*)}^{\alpha(\kappa_1^*)} Y_{(m\times r)}^{\alpha}. \tag{5.27}$$

6. Алгоритм энтропийно-рандомизированного проектирования с заданными значениями элементов матриц-проекторов

$dERP$-алгоритм состоит из двух последовательных этапов. На первом вычисляется энтропийно-оптимальная ДРВ на множестве заданных матриц-проекторов, сохраняющая индикатор компактности матрицы данных. Второй этап состоит в реализации “подходящей” матрицы-проектора из заданного множества с учетом энтропийно-оптимальной ДРВ.

Алгоритм 1 ($dERP$-алгоритм).

Шаг 0. Нормировка матрицы-данных

$$u_{ij} := \frac{u_{ij} - u_{min}}{u_{max} - u_{min}}, \qquad i = \overline{1, m}; \quad j = \overline{1, s}.$$

Шаг 1. Вычисление индикатора компактности матрицы данных

$$\rho_U := \frac{2!(m-2)!}{m!} \sum_{\beta,\gamma=1,\, \gamma<\beta}^{m} \varrho(\mathbf{u}^{(\beta)}, \mathbf{u}^{(\gamma)}).$$

Шаг 2. Генерация множества $\mathfrak{Q}$ матриц-проекторов с элементами 0,1

$$Q^{(\alpha)}, \qquad \alpha = \overline{1, N}, \quad N = 2^{rs}.$$

Шаг 3. Формирование множества $\mathfrak{Y}$ матриц-проекций с элементами

$$y_{i,k}^{(\alpha)} := \sum_{\nu=1}^{s} u_{i,\nu} q_{\nu,k}^{(\alpha)}, \qquad i = \overline{1, m}, \quad k = \overline{1, r},$$

и векторами-строками

$$\mathbf{y}_\nu^\alpha := \left\{ y_{\nu,1}^{(\alpha)}, \ldots, y_{\nu,r}^{(\alpha)} \right\}, \qquad \nu = \overline{1, m}.$$

Шаг 4. Вычисление индикаторов компактности матриц-проекций

$$\rho_Y^{(\alpha)} := \frac{2!(m-2)!}{m!} \sum_{\nu,\mu=1,\, \nu<\mu}^{m} \varrho(\mathbf{y}^{(\nu)}, \mathbf{y}^{(\mu)}), \qquad \alpha = \overline{1, N}.$$

Шаг 5. Определение значения множителя Лагранжа λ^* из уравнения

$$\sum_{\alpha=1}^{N} \exp\left(-\lambda \rho_Y^{(\alpha)}(Q)\right) \left(\rho_Y^{(\alpha)}(Q) - \rho_U\right) = 0, \qquad -\infty < \lambda < \infty,$$

и определение энтропийно-оптимальной ДРВ

$$w_\alpha^* = \frac{\exp\left(-\lambda \rho_Y^{(\alpha)}(Q)\right)}{\sum\limits_{\alpha=1}^{N} \exp\left(-\lambda \rho_Y^{(\alpha)}(Q)\right)}, \qquad \alpha = \overline{1, N}.$$

Шаг 6. Формирование и выбор стратегий реализации энтропийно-оптимальных матриц-проекторов:

— *матрица-проекция с максимальной вероятностью*

$$Q^{(\alpha^*)} : \alpha^* = \arg\max_\alpha w(\alpha \,|\, \lambda^*);$$

— *средняя матрица-проекция*

$$Q^{(\tilde{\alpha})} : \tilde{\alpha} = \sum_{\alpha=1}^{N} \alpha w_\alpha^*;$$

— *медианная матрица-проектор*

$$Q^{(\hat{\alpha})} : \sum_{\alpha=1}^{\hat{\alpha}} = \sum_{\alpha=\hat{\alpha}+1}^{N} w_\alpha^*;$$

— *средняя по дисперсионному подмножеству матрица-проекция*

$$\bar{Y}_{(m\times r)}^{D} = \frac{1}{2\sigma_\alpha^*} \sum_{\alpha=\alpha^*-\sigma_\alpha^*}^{\alpha^*+\sigma_\alpha^*} Y_{(m\times r)}^{(\alpha)},$$

— *средняя по* $(\kappa_1^* - \kappa_2^*)$*-интерквартильному подмножеству матрица-проекция*

$$\bar{Y}_{(m\times r)}^{K} = \frac{1}{M_{(\kappa_1^*-\kappa_2^*)}} \sum_{\alpha=\alpha(\kappa_2^*)}^{\alpha(\kappa_1^*)} Y_{(m\times r)}^{\alpha}.$$

Матрицы-проекции, вычисляемые на 6-м шаге алгоритма, могут использоваться как исходные матрицы данных для решения задач машинного обучения.

7. Заключение

Предлагается новый метод “сжатия” и “растяжения” матрицы данных, основанный на энтропийно-рандомизированном проектировани (ERP). Сформулирован алгоритм ERP в виде задачи условной максимизации функционала энтропии, определенного на функциях ПРВ матриц-проекторов. Получены условия существования и единственности положительного решения в указанной задаче. Рассмотрен случай конечных множеств матриц-проекторов, на которых существуют дискретные функции распределения вероятностей (ДРВ) и сформулирован алгоритм их энтропийной оптимизации. Реализация энтропийно-рандимизированных проекций осуществляется сэмплированием оптимальных ПРВ (ДРВ), генерацией множеств матриц-проекторов и использованием их числовых характеристик: средних, медианных, средних по дисперсионному или интерквартильному множеству.

ПРИЛОЖЕНИЕ

Доказательство теоремы 1. Воспользуемся условиями (5.13), (5.14), и уравнение (5.12) представим в виде

$$\phi(z) = \frac{\sum\limits_{\alpha=l+1}^{N} z^{h_\alpha} c_\alpha}{\sum\limits_{\alpha=1}^{l} z^{h_\alpha} c_\alpha} = 1. \tag{П.1}$$

Согласно (5.5) $h_{l+1} - h_1 > 0$, и

$$\phi(z) = z^{(h_{l+1}-h_1)} \frac{\sum\limits_{\alpha=l+2}^{N} z^{h_\alpha} c_\alpha + c_{l+1}}{\sum\limits_{\alpha=2}^{l} z^{h_\alpha} c_\alpha + c_1}. \tag{П.2}$$

Отсюда следует, что

$$\phi(0) = 0.$$

Представим теперь функцию $\phi(z)$ в виде

$$\phi(z) = z^{(h_N-h_l)} \frac{\sum\limits_{\alpha=l+1}^{N-1} z^{h_\alpha - h_N} c_\alpha + c_N}{\sum\limits_{\alpha=1}^{l-1} z^{h_\alpha - h_l} c_\alpha + c_l}. \tag{П.3}$$

Нетрудно увидеть, что

$$\phi(\infty) = +\infty.$$

Рассмотрим производную

$$(\text{П.4}) \qquad \phi'(z) = \frac{\sum\limits_{\beta=1}^{l} \sum\limits_{\alpha=l+1}^{N} c_\beta \, c_\alpha (h_\alpha - h_\beta) z^{(h_\beta - h_\alpha)}}{\left(\sum\limits_{\alpha=1}^{l} c_\alpha \, z^{h_\alpha}\right)^2} > 0.$$

Следовательно, функция $\phi(z)$ монотонно-возрастающая, и уравнения (П.1), (П.2) имеют единственное положительное решение $z^* > 0$.

СПИСОК ЛИТЕРАТУРЫ

1. *Carreira-Perpinan M.A.* A review of dimension reduction techniques. Tachnical report CS-96-09, Department of Computer Science, University of Sheffield, 1997.
2. *Imola K.* A survey of dimension reduction techniques. Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, 2002.
3. *Cunningham P.* Dimension Reduction. Technical Report UCD-CSI-2007-7, University College Dublin, 2007.
4. *Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д.* Прикладная статистика. Классификация и снижение размерности. М.: Финансы и статистика, 1989.
5. *Friedman J., Hastie T., Tibshirani R.* Element of Statistical Learning: Prediction, Inference and Data Mining. Springer, 2001.
6. *Bishop C.* Pattern Recognition and Machine Learning, ser. Information Science and Statistics, Springer, 2006.
7. *Comon P., Jutten C.* Handbook of Blind Source Separation. Independent Component Analysis and Applications. Academic Press, Oxford UK, 2010.
8. *Bruckstein A.M., Donoho D.L., Elad M.* From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images. SIAM Rev. 2009. V. 51. No. 1. P. 34–81.
9. *Pirson K.* On lines and planes of closest fit to systems of points in space // Philosophical Magazine. 1901. V. 2. P. 559–572.
10. *Кендал М.Дж., Стюарт А.* Статистические выводы и связи / Пер. с англ. М.: Наука, 1973.
11. *Jolliffe I.T.* Principal Component Analysis. N.Y. Springer-Verlag, 2002.
12. *Поляк Б.Т., Хлебников М.В.* Метод главных компонент: робастные версии // АиТ. 2017. № 3. С. 130–148.

    Polyak B.T., Khlebnikov M.V. Principle component analysis: Robust versions // Autom Remote Control 2017. V. 78. P. 490–506.

    https://doi.org/10.1134/S0005117917030092.
13. *Deerwester S.C., Dumias S.T., Landaurer T.K., Furnas G.W., Harshman R.A.* Indexing by latent semantic analysis // J. American Society of Information Sciences. 1990. V. 41. No. 6. P. 391–407.
14. *Fisher R.A.* The Use of Multiple Measurements in Taxonomic Problems // Annals of Eugenics. 1936. V. 7. P. 179–188.
15. *McLachlan G.J.* Discriminant Analysis and Statistical Pattern Recognition. Wiley Interscience, 2004.

16. *Johnson W.B., Lindenstrauss J.* Extension Of Lipshitz mapping into Hilbert space // In: Conference in modern analysis and probability. American Mathematical Society. 1984. V. 26. P. 189–206.
17. *Achlioptas D.* Database-friendly random projections // Proceedings of the twentieth ACM Symposium on Principles of database systems. P. 274–281.
18. *Bingham E., Mannila H.* Random projection in dimensionality reduction: applications to image and text data // Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining. P. 245–250.
19. *Vempala S.S.* The random projection method. American Mathematical Society, 2005. V. 65.
20. *Ганин И.П., Косиченко Е.А., Каплан А.Я.* Особенности ЭЭГ-реакций на эмоционально значимые стимулы в тезнологии интерфейса мозг-компьютер на волне P300 // Журн. высш. нерв. деятельности им. И.П. Павлова. 2017. Т. 67. № 4. С. 453–463.
21. *Huber F., Zorner T.O.* Threshold cointegrationin international exchange rates: A Bayesian approach // International Journal of Forecasting. 2019. V. 35. P. 458–473.
22. *Kosinskia M., Stillwella D., Groepelb T.* Private traits and attributtes are predictable from digital records of human behaviour // Proceedings of the National Agademy of Sciences of the USA. 2013. 110(15). P. 5802–5805.
23. *Blum A., Langly P.* Selection of relevant feature and examples in machine learning // Artificial Intelligence. 1997. V. 97. Nos. 1–2. P. 245–271.
24. *Cover T.M., Thomas J.A.* Elements of Information Theory. N.Y.: John Wiley and Sons, 1991.
25. *Peng H.C., Long F., Ding C.* Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redudancy // IEEE Transaction on Pattern Analysis and Machine Intelligence. 2005. V. 27. No. 8. P. 1226–1238.
26. *Zhang Y., Li S., Wang T., Zhang Z.* Divergence-based feature selection for separate classes // Neurocomputing. 2013. V. 101. P. 32–42.
27. *Дарховский Б.С., Каплан А.Я., Шишкин С.Л.* О подходе к оценке сложности кривых (на примере электроэнцефалограммы человека) // АиТ. 2002. № 3. С. 134–140.

    Darhovsky B.S., Kaplan A.Ya., Shishkin S.L. On an Approach to the Estimation of the Complexity of Curves (Using as an Example an Electroencephalogram of a Human Being) // Autom. Remote Control. 2002. V. 63. No. 3. P. 468–474.
28. *Дарховский Б.С., Пирятинская А.* Новый подход к проблеме сегментации временных рядов произвольной структуры // Труды МИАН им. В.А. Стеклова. 2014. Т. 287. С. 61–74.
29. *Darhovsky B., Piryatinska A.* Quickest detection of changes in the generating mechanism of a time series via the ε-complexity of continuous functions // Sequantial Analysis. 2014. V. 33. P. 231–250.
30. *Efron B.* Bootstrap Methods: Another Look at the Jackknife // Annals of Statistics. 1979. V. 7. No. 1. P.1–26.
31. *Bach F.R.* Bolasso: model consistent Lasso estimation through the bootstrap // Proceedings of the 25-th International Conference on Machine Learning. 2008. P. 33–40.
32. *Попков Ю.С.* Асимптотическая эффективность оценок максимальной энтропии // Доклады Российской Академии Наук. Математика, Информатика, Процессы Управления. 2020. Т. 493. С. 104–107.

    https://doi.org/10.31857/S2686954320040165.

Popkov Y.S. Asymptotic Efficiency of Maximum Entropy Estimates // Doklady Mathematics. 2020. V. 102. No. 1. P. 350–352.

https://doi.org/10.1134/S106456242004016X.

33. *Иоффе А.Д., Тихомиров В.М.* Теория экстремальных задач. М.: Наука, 1974.
34. *Дарховский Б.С., Попков Ю.С., Попков А.Ю., Алиев А.С.* Метод генерации случайных векторов с заданной функцией плотности распределения вероятностей // АиТ. 2018. № 9. С. 31–45. https://doi.org/10.31857/S000523100001408-2.

    Darkhovsky, B.S., Popkov, Y.S., Popkov, A.Y., Aliev, A.S. A Method of Generating Random Vectors with a Given Probability Density Function // Autom. Remote Control. 2018. V. 79. No. 9. P. 1569–1581.

Статья представлена к публикации членом редколлегии А.И. Кибзуном.

Поступила в редакцию 12.07.2020

После доработки 23.09.2020

Принята к публикации 28.10.2020