ДОКЛАДЫ АКАДЕМИИ НАУК, 2019, том 488, № 1, с. 21–23

МАТЕМАТИКА

УДК 51-7

КРОСС-ЭНТРОПИЙНАЯ ОПТИМАЛЬНАЯ РЕДУКЦИЯ РАЗМЕРНОСТИ МАТРИЦЫ ДАННЫХ С ОГРАНИЧЕНИЕМ ИНФОРМАЦИОННОЙ ЁМКОСТИ

Академик РАН Ю. С. Попков[1,2,*], А. Ю. Попков[1]

Поступило 29.04.2019 г.

Развивается метод редукции матрицы с неотрицательными элементами, основанный на кросс-энтропийной оптимизации матриц-проекторов при условии сохранения информационной ёмкости исходной матрицы.

Ключевые слова: снижение размерности, прямые и обратные проекции, информационная ёмкость, кросс-энтропия.

DOI: https://doi.org/10.31857/S0869-5652488121-23

1. ВВЕДЕНИЕ

Одна из проблем использования данных связана с их достаточным количеством для решения той или иной задачи. Один из путей определения достаточного количества данных может быть реализован с помощью перебора задач, контроля точности их решения, выбора приемлимой точности и соответствующего ей объёма данных. Базовым инструментом в этой группе является процедура случайных проекций (RP) [1], когда для фиксированной размерности редуцированной матрицы генерируются случайные матрицы проекций на основе нормального распределения, а в некоторых случаях — равномерного распределения [2]. В работе [3] был предложен метод “прямого и обратного проектирования” (DIP-метод — direct-inverse-projection), в котором для каждого фиксированного количества данных определялись энтропийно-оптимальные матрицы-проекторы.

Другой путь — в формировании процедуры редукции с количественно измеряемым функционалом её качества. Наиболее распространённой является процедура PCA (principal component analysis), основанная на методе опорных векторов [4, 5]. Функционал качества в этой процедуре — остаточная дисперсия, которая монотонно зависит от количества опорных векторов и, косвенно, от объёма данных [6]. На основе процедуры PCA развито много модификаций, ориентированных на особенности прикладных задач. Среди них линейный дискриминантный анализ (LDA) [7], в котором используется матрица данных и метки классов объектов, процедуры ICE [8] для анализа цифровых сигналов и NMF [9] для обработки аудиосигналов. Упомянутые методы используют в той или иной степени метод опорных векторов, который весьма чувствителен к ошибкам в данных. Некоторое уменьшение этой зависимости даёт его робастная модификация [10].

Данное сообщение посвящено модификации DIP-метода, построенной на понятии информационной ёмкости матрицы. DIP-метод дополняется условием сохранения информационной ёмкости исходной матрицы данных в заданных пределах. Модифицированный метод редукции матрицы данных (mDIP) моделируется задачей максимизации кросс-энтропийной функции на множестве, описываемом линейным неравенством. Для поиска минимума предлагается использовать мультипликативный алгоритм первого порядка [11, 12]. Получены условия его локальной сходимости.

2. КРОСС-ЭНТРОПИЙНАЯ РЕДУКЦИЯ С ОГРАНИЧЕНИЕМ ИНФОРМАЦИОННОЙ ЁМКОСТИ (mDIP-МЕТОД)

Рассмотрим невырожденную матрицу $U_{(m\times s)} \geq 0$ данных: m — количество объектов, s — количество признаков. Для сокращения размерности признакового пространства с s до r применим метод “прямого” с матрицей-проектором $Q_{(s\times r)} \geq 0$ и “обратного” с матрицей-проектором $T_{(r\times s)} \geq 0$ проектирования [3]. Будем иметь

[1] *Федеральный исследовательский центр “Информатика и управление” Российской Академии наук, Москва*

[2] *Department of Software Engineering, ORT Braude College, Karmiel, Israel*

[*] *E-mail: popkov@isa.ru*

$$Y_{(m\times r)} = U_{(m\times s)} Q_{(s\times r)}, \;\; X_{(m\times s)} = U_{(m\times s)} Q_{(s\times r)} T_{(r\times s)} \leq 0. \quad (1)$$

“Расстояние” между указанными матрицами будем оценивать функционалом кросс-энтропии [12, 13]:

$$H(X_{(m\times s)}|U_{(m\times s)}) = H(Q_{(s\times r)}, T_{(r\times s)}|U) =$$

$$= -\sum_{i=1}^{m}\sum_{j=1}^{s} h_{ij}(Q_{(s\times r)}, T_{(r\times s)}|U_{(m\times s)}), \quad (2)$$

где

$$h_{ij}(Q_{(s\times r)}, T_{(r\times s)}|U_{(m\times s)}) =$$

$$= x_{ij}(Q_{(s\times r)}, T_{(r\times s)}|U_{(m\times s)}) \ln \frac{x_{ij}(Q_{(s\times r)}, T_{(r\times s)}|U_{(m\times s)})}{u_{ij}}, \quad (3)$$

$$x_{ij}(Q_{(s\times r)}, T_{(r\times s)}|U_{(m\times s)}) = \sum_{\mu=1}^{r}\sum_{\nu=1}^{s} t_{\mu,j} q_{\nu,\mu} u_{i,\nu}, \quad (4)$$

$$i = 1, 2, \ldots, m; \; j = 1, 2, \ldots, s.$$

Т е о р е м а 1. *Существует множество* $\mathcal{G}_{(Q,T)}$ *неотрицательных матриц* Q, T:

$$(Q, T) \in \mathcal{G}_{(Q,T)} =$$

$$= \left\{ \sum_{i=1}^{m}\sum_{j=1}^{s} x_{ij}(Q, T) \ln \frac{x_{ij}(Q, T)}{u_{ij}} + ms \geq 0, \; Q \geq 0, \; T \geq 0 \right\}, \quad (5)$$

на котором матричный гессиан функции $H(Q, T|U) \leq 0$.

Под матричным гессианом понимается матрица вторых производных скалярной функции $H(Q,T)$ по элементам матриц Q, T [14]. Доказательство следует из вида элементов матричного гессиана.

Введём понятие информационной ёмкости прямоугольной неотрицательной матрицы $U_{(m\times s)}$. Рассмотрим функционал

$$\mathcal{I}(U) = \sum_{(i,j)=1}^{m,s} u_{ij} \ln u_{ij}. \quad (6)$$

Для $0 \leq u_{ij} \leq 1$, $\mathcal{I} \leq 0$ и для $u_{ij} > 1$, $\mathcal{I} > 0$. Минимальное значение $\mathcal{I}_{\min} = -mse^{-1}$. Определим информационную ёмкость матрицы $U_{(m\times s)}$ в виде

$$E_U = \mathcal{I}(U) + e^{-1}ms \geq 0. \quad (7)$$

Оптимальные элементы неотрицательных матриц Q,T (1) будем определять, максимизируя кросс-энтропию (2) на множестве $\mathcal{G}_{Q,T}$ (5) (где существует единственный локальный максимум $H(Q,T)$) при условии, что снижение информационной ёмкости редуцированной матрицы Y (1) не более допустимого:

$$H(Q_{(s\times r)}, T_{(r\times s)}|U) \Rightarrow \max,$$

$$(Q, T) \in \mathcal{D}(Q) = \{(Q) : E_U - E_Y(Q)\} \leq \delta, \quad (8)$$

где

$$E_Y(Q) = \sum_{(i,j)=1}^{m,r} y_{ij}(Q) \ln y_{ij}(Q) + e^{-1}mr,$$

$$y_{ij}(Q) = \sum_{k=1}^{s} u_{i,k} q_{k,j}, \quad i = 1, 2, \ldots, m, \; j = 1, 2, \ldots, r. \quad (9)$$

Преобразуем задачу (8), (9) к стандартному виду, используя процедуру векторизации матриц:

$$(Q, T) \to \mathbf{q} \in R^{mr}, \;\; \mathbf{t} \in R^{mr}. \quad (10)$$

Тогда задача (8), (9) примет следующий вид:

$$H(\mathbf{q}, \mathbf{t}) \Rightarrow \max,$$

$$(\mathbf{q}, \mathbf{t}) \in \mathcal{D}(q) = \{\mathbf{q} : E_Y(\mathbf{q}) - E_U \geq \delta\}, \quad (11)$$

$$\mathbf{q} \geq 0, \;\; \mathbf{t} \geq 0.$$

3. УСЛОВИЯ ОПТИМАЛЬНОСТИ И МУЛЬТИПЛИКАТИВНЫЙ АЛГОРИТМ

Определим функционал Лагранжа в виде

$$\mathcal{L}(\mathbf{q}, \mathbf{t}, \lambda) = P(\mathbf{q}, \mathbf{t}, \lambda) - \langle \mu, \mathbf{q} \rangle - \langle \nu, \mathbf{t} \rangle, \quad (12)$$

$$P(\mathbf{q}, \mathbf{t}, \lambda) = H(\mathbf{q}, \mathbf{t}) + \lambda\left(\delta - E_U + E_Y(\mathbf{q})\right),$$

$$\mathbf{q} \in R_+^{mr}, \;\; \mathbf{t} \in R_+^{mr}, \;\; \mu \in R^{mr}, \;\; \nu \in R^{mr}.$$

Необходимые условия оптимальности Куна—Таккера в терминах функции $P(\mathbf{q}, \mathbf{t}, \lambda)$ имеют вид [12]

$$P_{\mathbf{q}}(\mathbf{q}^*, \mathbf{t}^*, \lambda^*) \geq 0, \;\; \mathbf{q}^* \otimes P_{\mathbf{q}}(\mathbf{q}^*, \mathbf{t}^*, \lambda^*) = 0, \;\; \mathbf{q}^* \geq 0, \; \mathbf{t} \geq 0,$$

$$P_{\mathbf{t}}(\mathbf{q}^*, \mathbf{t}^*, \lambda^*) \geq 0, \;\; \mathbf{q}^* \otimes P_{\mathbf{t}}(\mathbf{q}^*, \mathbf{t}^*, \lambda^*) = 0, \;\; \mathbf{q}^* \geq 0, \; \mathbf{t}^* \geq 0,$$

$$P_{\lambda}(\mathbf{q}^*, \mathbf{t}^*, \lambda^*) \leq 0, \;\; \lambda^* \otimes P_{\lambda}(\mathbf{q}^*, \mathbf{t}^*, \lambda^*) = 0, \quad (13)$$

$$\mathbf{q}^* \geq 0, \; \mathbf{t}^* \geq 0, \; \lambda^* \geq 0.$$

Алгоритм решения этой системы имеет вид:

а) начальный шаг

$$(\mathbf{q}^0, \mathbf{t}^0) > 0, \quad (\mathbf{q}^0, \mathbf{t}^0) \in \mathcal{G}_{Q,T};$$

б) итерационный шаг

$$\mathbf{q}^{s+1} = \mathbf{q}^s \otimes \left(\mathbf{1} - \gamma P_{\mathbf{q}}(\mathbf{q}^s, \mathbf{t}^s, \lambda^s)\right),$$

$$\mathbf{t}^{s+1} = \mathbf{t}^s \otimes \left(\mathbf{1} - \alpha P_{\mathbf{t}}(\mathbf{q}^s, \mathbf{t}^s, \lambda^s)\right),$$

$$\lambda^{s+1} = \lambda^s \left(\mathbf{1} + \beta P_{\lambda}(\mathbf{q}^s, \mathbf{t}^s, \lambda^s)\right).$$

Доказательство сходимости основано на исследовании поведения функции Ляпунова

$$V(\mathbf{q},\mathbf{t},\lambda)=\sum_{i=1}^{mr}(q_i-q_i^*)-q_i^*(\ln q_i-\ln q_i^*)+$$
$$+\sum_{i=1}^{mr}(t_i-t_i^*)-t_i^*(\ln t_i-\ln t_i^*)+$$
$$+(\lambda-\lambda^*)-\lambda^*(\ln\lambda-\ln\lambda^*) \qquad (14)$$

на решениях системы:

$$\frac{d\mathbf{q}}{dt}=-\mathbf{q}\otimes P_{\mathbf{q}}(\mathbf{q},\mathbf{t},\lambda),$$
$$\frac{d\mathbf{t}}{dt}=-\mathbf{t}\otimes P_{\mathbf{t}}(\mathbf{q},\mathbf{t},\lambda), \qquad (15)$$
$$\frac{d\lambda}{dt}=\lambda P_{\lambda}(\mathbf{q},\mathbf{t},\lambda).$$

Источники финансирования. Работа поддержана Российским фондом фундаментальных исследований (проекты 17–29–03119, 17–29–02115).

СПИСОК ЛИТЕРАТУРЫ

1. *Bingham E.*, *Manilla H.* Random projection in dimensionality reduction: Application to image and next data // Proc. of the Seventh ACM SICKDD Intern. Conf. on Knowledge Discovery and Data Mining. N.Y.: Association for Computing Machinery. P. 245–250.
2. *Achlioptas D.* Database-friendly random projection // Proceeding PODS'01, Proc. of the twentith ACM SIGMOD-SIGACT-SIGART symposium on principles of data base systems. 2001. P. 274–281.
3. *Попков Ю.С.*, *Дубнов Ю.А.*, *Попков А.Ю.* Энтропийная редукция размерности в задачах рандомизированного машинного обучения // АиТ. 2018. № 11. С. 106–123.
4. *Кендал М.Дж.*, *Стюарт А.* Статистические выводы и связи // Пер. с англ. М.: Наука, 1973.
5. *Jollife I.T.* Principle Component Analysis // N.Y.: Springer-Verlag, 2002.
6. *Айвазян С.А.*, *Бухштабер И.М.*, *Енюков И.С.*, *Мешалкин Л.Д.* Прикладная статистика: Классификация и снижение размерности // М.: Финансы и статистика, 1989.
7. *Fisher R.A.* The Use of Multiple Measurements in Taxonomic Problems // Annals of Eugenics. 1936. V. 7. P. 179–188.
8. *Comon P.*, *Jutten C.* Handbook of Blind Source Separation, Independent Component Analysis and Applications. Oxford: Acag. Press, 2010.
9. *Michael W.*, *et all.* Algorithms and Application for Approximate Nonnegative Matrix Factorization // Computational Statistics and Data Analysis. 2007. V. 52. P. 150–173.
10. *Поляк Б.Т.*, *Хлебников М.В.* Метод главных компонент: робастные версии //АиТ. 2017. № 3. С. 130–148.
11. *Попков Ю.С.* Мультипликативные схемы итерационных алгоритмов II: выпуклое программирование // Техническая кибернетика. 1993. № 2. С. 31–45.
12. *Попков Ю.С.* Теория макросистем: равновесные модели. М.: УРСС, 1999. 317 с.
13. *Kullback S.*, *Leibler R.A.* On information and Sufficiency. Ann. of Math. Statistics. 1951. V. 22(1). P. 79–86.
14. *Magnus J.*, *Neudecker H.* Matrix Differential Calculus with Applications in Statistics and Econometrics. John Willey and Sons. 1999.

CROSS-ENTROPY OPTIMAL DIMENSIONALITY REDUCTION WITH CONDITION ON INFORMATION CAPACITY

Academician of the RAS Yu. S. Popkov[1,2,*], A. Yu. Popkov[1]

[1] *Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, Moscow, Russian Federation*

[2] *Department of Software Engineering, ORT Braude College, Karmiel, Israel*

Received April 29, 2019

Using a data leads to a problem of its sufficiency to solve specific task. Proposed paper is devoted to a modification of direct-inverse projection method (DIP-method) based on an idea of information capacity. DIP-method is updated with a condition on maintaining the information capacity in given ranges. Modified dimensionality reduction method (mDIP) based on the problem of minimization cross-entropy function on a set defined by linear inequality. Minimization of the function is suggested to perform by the first-order multiplicative algorithm. There obtained conditions of local convergence.

Keywords: dimensionality reduction, direct and inverse projections, information capacity, cross-entropy.