УДК 004.942,519.177, 519.217
https://doi.org/10.17721/1812-5409.2021/2.7

Кириченко О.Л.[1], аспірант,
Малик І.В.[1], д.ф.-м.н., доц.,
Остапов С.Е.[1], д.ф.-м.н., проф.

Стохастичні моделі в задачах штучного інтелекту

[1] Чернівецький національних університет імені Юрія Федьковича, 58012, м. Чернівці, вул. Коцюбинського, 2, 58012
e-mail: o.kyrychenko@chnu.edu.ua, i.malyk@chnu.edu.ua, s.ostapov@chnu.edu.ua

O.L. Kyrychenko[1], postgraduate student,
I.V Malyk[1], Dr.Sc., Associate Professor,
S.E. Ostapov[1], Dr.Sc., Professor

Stochastic models in artificial intelligence development

[1] Yuriy Fedkovych Chernivtsi National University, 58012, Chernivtsi, Kotsiubynsky str, 2,

e-mail: o.kyrychenko@chnu.edu.ua, i.malyk@chnu.edu.ua, s.ostapov@chnu.edu.ua

У роботі розглянуто деякі властивості стохастичних матриць великих розмірностей за умов незалежності елементів матриці або за умов незалежності рядків (стовпців). Проаналізовано основні властивості спектру стохастичної матриці. Також розглянуто застосування даних результатів до задач кластеризації та вибору оптимального числа кластерів.

Результати дослідження доповідались на Міжнародній науковій конференції ``Modern Stochastics: Theory and Applications. V'' (MSTA-V).

Ключові слова: стохастична випадкова матриця, спектр матриці, оптимальне число кластерів.

In this paper, we consider some properties of stochastic random matrices of large dimensions under conditions of independence of matrix elements or under conditions of independence of rows (columns). The main properties of stochastic random matrices spectrum are analyzed and the result of convergence to 0 is proved of almost all eigenvalues. Also, the application of these results to clustering problems and selection of the optimal number of clusters is considered. Note that the results obtained in this work are consistent with the Marchenko - Pastur theorem on the asymptotic distribution of eigenvalues of random matrices with independent elements. The results proved in this paper can be interpreted as a law of large numbers and will be used in the study of the asymptotic behavior of the maximum.

Key Words: stochastic random matrix, spectrum of a matrix, optimal number of clusters.

Спектральний аналіз матриць [1, 2, 3] є одним із важливих підходів оптимізації кількості кластерів при кластеризації в задачах штучного інтелекту. Теорія випадкових матриць є однією із математичних моделей розумних мереж (Smart Grid) [4], що описує функціонування великих систем (переважно енергосистем). Одним із основних завдань, пов'язаних із побудовою великої системи є розподіл функцій між основними "центрами" системи. Крім того, вплив випадкових факторів в системі може бути описано саме за допомогою стохастичних випадкових матриць. Надалі, під стохастичною матрицею $A = A_{N\times N}$ будемо розуміти матрицю, що володіє наступними властивостями:

1. $A_{ij} \geq 0$;
2. $\sum_{j=1}^{N} A_{ij} = 1$.

Під випадковою матрицею будемо розуміти матрицю, елементами якої є випадкові величини, визначені на ймовірнісному просторі (Ω, F, P).

Основна ідея роботи полягає у локалізації власних значень стохастичної матриці великої розмірності. Для цього будемо користуватися тим фактом, що одним із власних значень стохастичної матриці, що відповідає незвідному ланцюгу Маркова, є 1 та всі власні значення стохастичної матриці не перевищують за абсолютним значенням одиницю. Таким чином, на основі даної властивості, розроблено алгоритм для визначення оптимальної кількості кластерів k_{opt}.

Розглянемо основні позначення, які будуть використовуватися у даній роботі. Основними математичними об'єктами, які будуть використані нижче є граф та ланцюг Маркова. Граф будемо позначати через $G = (V, E)$, де $V = \{1,2,\dots,N\}$ – множина вершин графа, $E = \{e_1,\dots,e_m\}$ – множина ребер графа, що з'єднують вершини із V. Припустимо, що граф задається матрицею суміжності A:

$$A = A_{N\times N},$$

де елемент A_{ij} рівний ваговому коефіцієнту між вершинами i та j. Стохастична матриця

$$P_{ij} = \frac{A_{ij}}{\sum_{j=1}^{N} A_{ij}}, \tag{1}$$

що відповідає графу G та задається матрицею суміжності A, задається співвідношенням (1). Матриця P й буде основним предметом дослідження в даній роботі. Надалі будемо впорядковувати власні значення матриці A наступним чином

$$|\lambda_1(A)| \geq |\lambda_2(A)| \geq \dots \geq |\lambda_N(A)|.$$

Згідно зауваження зробленого вище, вірні наступні співвідношення

$$\begin{cases} \lambda_1(P) = 1, \\ |\lambda_i(A)| \leq 1, i \geq 1. \end{cases}$$

Теорема 1. Нехай виконуються наступні умови:

1. Всі елементи матриці $A = A_{N\times N}$ є незалежними та мають одинаковий розподіл
$$A_{ij} \sim \mathrm{Distr},$$
де розподіл $Distr$ має носієм підмножину $(0,\infty)$ та володіє скінченним моментом порядку $2+\delta, \delta > 0$;
2. Елементи матриці переходу за 1 крок P_{ij} визначаються із співвідношення (1)
$$P_{ij} = \frac{A_{ij}}{\sum_{k=1}^{N} A_{ik}}.$$

Тоді за ймовірністю має місце збіжність

$$\lambda_2(P) \to 0,$$

при $N \to \infty$.

Доведення. Для доведення даного факту відзначимо, що значення $\lambda_2(N)$, що відповідає значенням матриці

$$P_{ij} = E\left(\frac{A_{ij}}{\sum_{j=1}^{N} A_{ij}}\right)$$

рівне 0. Для доведення цього факту розглянемо власні значення матриці P у наступному вигляді

$$\det(\lambda I - P) = \lambda^n + c_1\lambda^{n-1} + \cdots c_{n-1}\lambda + c_n,$$

де коефіцієнти c_i визначаються за наступною формулою

$$c_i = (-1)^i \sum_k \det\big(M_k(i)\big),$$

де $M_k(i)$ – головні мінори матриці P розмірності $i \times i$. Використовуючи дане представлення, отримаємо

$$c_1 = -\sum_{k=1}^{N} P_{kk} = -\sum_{k=1}^{N} \frac{A_{kk}}{\sum_{j=1}^{N} A_{kj}}.$$

Згідно означення значень c_i, отримаємо на основі припущення 1 теореми

$$E(c_1) = -E\left(\sum_{k=1}^{N} \frac{A_{kk}}{\sum_{j=1}^{N} A_{kj}}\right) = -\sum_{k=1}^{N} \frac{1}{N} = -1.$$

Для всіх інших значень $c_i, i \geq 2$, отримаємо

$$E(c_i) = (-1)^i E\left(\sum_k \det\big(M_k(i)\big)\right) = (-1)^i \sum_k E\Big(\det\big(M_k(i)\big)\Big).$$

Розглянемо значення $E\Big(\det\big(M_k(i)\big)\Big)$. Для цього скористаємося тим фактом, що

$$E\big(\det(Z)\big) = \det\big(E(Z)\big)$$

при умові, що рядки (або стовпці) матриці Z незалежні векторні випадкові величини. У нашому випадку, рядки матриці P є незалежні в сукупності, тому має місце співвідношення:

$$E\Big(\det\big(M_k(i)\big)\Big) = \det\Big(E\big(M_k(i)\big)\Big),$$

причому всі значення матриці $E\big(M_k(i)\big)$ рівні $\frac{1}{N}$. Таким чином

$$E\Big(\det\big(M_k(i)\big)\Big) = 0.$$

Використовуючи дані обчислення, знайдемо, що "усереднений характеристичний поліном" матриці P рівний

$$E(\det(\lambda I - P)) = E(\lambda^n + c_1\lambda^{n-1} + \cdots c_{n-1}\lambda + c_n) =$$
$$= \lambda^n + E(c_1)\lambda^{n-1} + \cdots E(c_{n-1})\lambda + E(c_n) =$$
$$= \lambda^n - \lambda^{n-1}.$$

Отже, розв'язками характеристичного рівняння

$$E(\det(\lambda I - P)) = 0$$

є два значення

$$\lambda_1 = 1, \lambda_i = 0, i = 2,\dots,N.$$

Тобто, 'усереднене значення' власного значення λ_2 рівне 0 за умов теореми.

Для доведення збіжності до 0 випадкової величини λ_2 розглянемо дисперсії діагональних елементів

$$D\left(\sum_{i=1}^{N} P_{ii}\right) = \sum_{i=1}^{N} D(P_{ii}) = ND(P_{11}),$$

оскільки всі значення P_{ii} є незалежними та однаково розподілені. Обчислимо дисперсію випадкової величини

$$D(P_{11}) = E(P_{11}^2) - \frac{1}{N^2}.$$

Обчислимо другий момент для випадкової величини P_{11}^2, використовуючи існування та обмеженість другого моменту:

$$E(P_{11}^2) = E\left(\frac{A_{11}}{\sum_{j=1}^{N} A_{1j}}\right)^2.$$

Для великих N на основі центральної граничної теореми отримаємо

$$X = \sum_{j=1, j\neq i}^{N} A_{ij} = \sqrt{N-1} \sum_{j=1, j\neq i}^{N} \frac{(A_{ij} - EA_{ij} +)}{\sqrt{N-1}} =$$

$$\sigma\sqrt{N-1} \sum_{j=1, j\neq i}^{N} \frac{(A_{ij} - EA_{ij})}{\sigma\sqrt{N-1}} + Nm,$$

де $m = EA_{ij}$, $\sigma^2 = D(A_{ij})$. Використовуючи дане представлення, отримаємо

$$X \approx \sqrt{N-1}Z + Nm.$$

Тут випадкова величина Z має нормальний розподіл із параметрами (0,1). На основі цього

$$E(P_{11}^2) = E\left(\frac{A_{11}}{\sum_{j=1}^{N} A_{1j}}\right)^2 \approx$$

$$E\left(\frac{A_{11}}{A_{11} + \sqrt{N-1}Z + Nm} I_{\{Z>0\}}\right)^2 =$$

$$E\left(\frac{A_{11}}{A_{11} + \sqrt{N-1}Z + Nm} I_{\{Z>0\}}\right)^2$$

$$\approx \frac{1}{N-1} E\left(\frac{\widetilde{A}_{11}}{\widetilde{A}_{11} + Z + \sqrt{N-1}m} I_{\{Z>0\}}\right)^2,$$

де

$$\widetilde{A}_{11} = \frac{A_{11}}{\sqrt{N-1}}.$$

Використовуючи дане представлення, отримаємо

$$E(P_{11}^2) = O\left(\frac{1}{N^2}\right)$$

та

$$D(P_{11}) = O\left(\frac{1}{N^2}\right).$$

Таким чином, використовуючи дане представлення, отримаємо

$$D\left(\sum_{i=1}^{N} P_{ii}\right) = O\left(\frac{1}{N}\right).$$

Для доведення того факту, що $\lambda_2(N)$ прямує до 0 за ймовірністю, використаємо наступне співвідношення

$$\sum_{i=1}^{N} \left(\mu_i(A) - \mu_i(B)\right)^2 \leq Tr(A-B)^2,$$

де $\mu_i(A), \mu_i(B)$ – власні значення матриць A та B, впорядковані за спаданням.

Використовуючи дане співвідношення, отримаємо наступну оцінку для дисперсії $\lambda_2(N)$:

$$E(\lambda_2(N))^2 = D(\lambda_2(N))^2 \leq$$

$$\leq \sum_{i=1}^{N} D(\lambda_i(N)) \leq E\left(\sum_{i=1}^{N}\sum_{k=1}^{N}\left(P_{ik} - \frac{1}{N}\right)\left(P_{ki} - \frac{1}{N}\right)\right)$$

$$=$$

$$E\left(\sum_{i=1}^{N}\left(P_{ii} - \frac{1}{N}\right)^k\right) = \sum_{i=1}^{N} E\left(P_{ii} - \frac{1}{N}\right)^2 =$$

$$= \sum_{i=1}^{N} D(P_{ii}) = D\left(\sum_{i=1}^{N} P_{ii}\right).$$

Отже,

$$D(\lambda_2(N)) = O\left(\frac{1}{N}\right).$$

Тоді, з врахуванням доведених вище представлень для моментів P_{ii}, твердження теореми є наслідком нерівності Чебишова.

Дослідженням оптимальної кількості кластерів автори займались в роботах [5,6]. Були розглянуті методи знаходження оптимальної кількості кластерів k-Core decomposition [7] та «метод ліктя» [8]. Цими методами визначались оптимальна кількість кластерів та центри кластерів для різних сегментів веб-простору. Розглянемо тепер задачу визначення оптимальної кількості кластерів на основі стохастичної матриці P, що задає переходи в деякій системі. Також будемо припускати, що стохастична матриця P має блочний вигляд, тобто має місце співвідношення

$$P = \begin{pmatrix} P^{(1)} & Err^{(1,2)} & \cdots & Err^{(1,k_{opt})} \\ Err^{(2,1)} & P^{(2)} & \cdots & Err^{(2,k_{opt})} \\ \vdots & \vdots & \ddots & \vdots \\ Err^{(k_{opt},1)} & Err^{(k_{opt},2)} & \cdots & P^{(k_{opt})} \end{pmatrix}, (2)$$

де $P^{(i)}$ задають ймовірності переходу в i – му кластері та $Err^{(i,j)}$ – ймовірності переходу з i – го кластеру в j – ий кластер. Матриці $Err^{(i,j)}$ будемо характеризувати як "помилки" переходів між сусідніми кластерами. Нехай зв'язки всередині кластера є набагато сильнішими ніж зв'язки між кластерами. Дане припущення будемо виражати наступним чином

$$\min_{i\in 1,\dots,k_{opt}} \min_{m\in 1,\dots,n_i} \sum_{j=1}^{n_i} P_{mj}^{(i)} \geq$$

$$\geq \alpha \max_{i\in 1,\dots,k_{opt}} \max_{m\in 1,\dots,n_i} \left(1-\sum_{j=1}^{n_i} P_{mj}^{(i)}\right),$$

де n_i – розмірність матриці $P^{(i)}$, $\alpha \in (1,\infty)$ - деякий параметр подібності. Зауважимо, що при $\alpha \to \infty$ всі помилки $Err_{k,m}^{(i,j)} \to 0$. В цьому випадку матриця P перетворюється в блочно - діагональну матрицю вигляду

$$P' = \begin{pmatrix} P^{(1)} & 0 & \cdots & 0 \\ 0 & P^{(2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & P^{(k_{opt})} \end{pmatrix}. \quad (3)$$

Тобто, матриця P' в цьому випадку задає чітко визначені k_{opt} кластери даних. Крім того, для матриці P, заданої співвідношенням (3),

$$\lambda_1(P') = \cdots = \lambda_{k_{opt}}(P') = 1$$

та

$$|\lambda_i(P') - 1| > \varepsilon, i > k_{opt},$$

де параметр ε залежить від розмірності системи (кількості об'єктів кластеризації) та параметру α. Для визначення оптимальної кількості кластерів ми будемо користуватися наступним співвідношенням

$$k_{opt} = \#\left\{\lambda_i : |\lambda_i(P) - 1| \leq \frac{1}{\alpha\sqrt{N}}\right\}. \quad (4)$$

Дане співвідношення є наслідком Теореми 1 та асимптотичної поведінки власних значень $\lambda_i(P)$, $i \leq k_{opt}$.

Зауважимо, що результати отримані в роботі узгоджуються із теоремою Марченко – Пастура [9] про асимптотичний розподіл власних значень випадкових матриць із незалежними елементами. Також, слід відмітити, що результати теореми можуть трактуватися як закон великих чисел. Крім того, визначення оптимальної кількості кластерів будується на асимптотиці власних значень матриці P при $N \to \infty$. Одержані результати можуть бути застосовані для визначення оптимальної кількості кластерів Grid system, складних мереж, при дослідженнях структури веб-простору тощо.

Список використаних джерел

1. *Andrew Y. Ng, Michael Jordan, and Yair Weiss.* (2002) On spectral clustering: Analysis and an algorithm, in NIPS, (2002).
2. *Frank Lin and William W. Cohen.* (2010) Power iteration clustering, in ICML(to appear), (2010).
3. *Zhidong Bai, Zhaoben Fang, Ying-Chang Liang* (2014). Spectral Theory of Large Dimensional Random Matrices and Its Applications to Wireless Communications and Finance Statistics : Random Matrix Theory and Its Applications. University of Science and Technology of China Press? World Scientific.
4. *Robert C. Qiu, Paul Antonik* (2017). Smart Grid using Big Data Analytics. A Random Matrix Theory Approach. Wiley Online Library, 2017.
5. *Кириченко О.Л.* Проведення оптимальної кластеризації структури веб-простору [Текст] / О.Л. Кириченко, С.Е. Остапов, І.Я. Кановський // Міжнародна науково-практична конференція «Проблеми інформатики та комп'ютерної техніки» (ПІКТ-2017, 05-08 жовтня). Праці конференції. – Чернівці: Видавничий дім «Родовід», 2017. – С. 67-69.

References

1. ANDREW Y. NG, MICHAEL JORDAN, AND YAIR WEISS. (2002) *On spectral clustering: Analysis and an algorithm*, in NIPS, (2002).
2. FRANK LIN AND WILLIAM W. COHEN. (2010) *Power iteration clustering*, in ICML(to appear), (2010).
3. ZHIDONG BAI, ZHAOBEN FANG, YING-CHANG LIANG (2014). *Spectral Theory of Large Dimensional Random Matrices and Its Applications to Wireless Communications and Finance Statistics : Random Matrix Theory and Its Applications*. University of Science and Technology of China Press, World Scientific.
4. ROBERT C. QIU, PAUL ANTONIK (2017). *Smart Grid using Big Data Analytics. A Random Matrix Theory Approach*. Wiley Online Library, 2017.
5. KYRYCHENKO O.L. *Provedennia optymalnoi klasteryzatsii struktury veb-prostoru* [Tekst] / O.L. Kyrychenko, S.E. Ostapov, I.Ia. Kanovskyi // Mizhnarodna

naukovo-praktychna konferentsiia «Problemy informatyky ta kompiuternoi tekhniky» (PIKT-2017, 05-08 zhovtnia). Pratsi konferentsii. – Chernivtsi: Vydavnychyi dim «Rodovid», 2017. – Pp. 67-69.

6. *Кириченко О.Л.* Застосування методу k-core decomposition для проведення оптимальної кластеризації / О.Л. Кириченко, С.Е. Остапов Інформаційні технології: наука, техніка, технологія, освіта, здоров'я: тези доповідей XXVII міжнародної науково-практичної конференції MicroCAD-2019, 15-17 травня 2019 р.: у 4 ч. Ч. IV. / за ред. проф. Сокола Є.І. – Харків: НТУ «ХПІ». – С. 155.
7. *Sheng-Tzong Cheng, Yin-Chun Chen, and Meng-Shuan Tsai.* (2017) Using k-Core Decomposition to Find Cluster Centers for k-Means Algorithm in GraphX on Spark, CLOUD COMPUTING 2017: The Eighth International Conference on Cloud Computing, GRIDs, and Virtualization.
8. *Kuraria, Amit & Jharbade, Nitin & Soni, Manish*. (2018). Centroid Selection Process Using WCSS and Elbow Method for K-Mean Clustering Algorithm in Data Mining. International Journal of Scientific Research in Science, Engineering and Technology. Pp. 190-195. 10.32628/IJSRSET21841122.
9. *В. А. Марченко, Л. А. Пастур*, Распределение собственных значений в некоторых ансамблях случайных матриц, Матем. сб., 1967, том 72(114), номер 4, С. 507–536.

6. KYRYCHENKO O.L. *Zastosuvannia metodu k-core decomposition dlia provedennia optymalnoi klasteryzatsii* / O.L. Kyrychenko, S.E. Ostapov Informatsiini tekhnolohii: nauka, tekhnika, tekhnolohiia, osvita, zdorovia: tezy dopovidei KhXVII mizhnarodnoi naukovo-praktychnoi konferentsii MicroCAD-2019, 15-17 travnia 2019 r.: u 4 ch. Ch. IV. / za red. prof. Sokola Ye.I. – Kharkiv: NTU «KhPI». – pp. 155
7. SHENG-TZONG CHENG, YIN-CHUN CHEN, AND MENG-SHUAN TSAI. (2017) *Using k-Core Decomposition to Find Cluster Centers for k-Means Algorithm in GraphX on Spark*, CLOUD COMPUTING 2017: The Eighth International Conference on Cloud Computing, GRIDs, and Virtualization.
8. KURARIA, AMIT & JHARBADE, NITIN & SONI, MANISH. (2018). *Centroid Selection Process Using WCSS and Elbow Method for K-Mean Clustering Algorithm in Data Mining.* International Journal of Scientific Research in Science, Engineering and Technology. Pp. 190-195. 10.32628/IJSRSET21841122.
9. V. A. MARCHENKO, L. A. PASTUR, *Raspredelenye sobstvennыkh znachenyi v nekotorыkh ansambliakh sluchainыkh matryts,* Matem. sb., 1967, tom 72(114), nomer 4, Pp. 507–536.

Надійшла до редакції 28.08.2021