DOI 10.54596/2958-0048-2024-3-165-171
UDK 63.001
IRSTI 12.21.35

APPLICATION OF MARKOV CHAINS TO FORECASTING TASKS IN SOCIOCENOSES

Kurmasheva L.B.[1*], Neradovskaya Y.W.[2], Kurmashev I.G.[1]
[1]Manash Kozybayev North Kazakhstan University NPLC, Petropavlovsk, Kazakhstan*
[2]Federal State Budget Educational Institution of Higher Education «Saint-Petersburg State University of Economics», Saint Petersburg, Russia
[*]Corresponding author: lbkurmasheva@ku.edu.kz

Abstract

Professional development is an important process that affects people's way of life. Supporting students at the moment of choosing a university, during the learning process, can help them make important career decisions and increase their employability. The paper proposes an approach to modeling the behavior of an applicant using Markov chains, and provides some interpretations. The Markov chain is widely used for modeling and analyzing stochastic systems in various fields of science and technology. The results of the study can be useful for the university administration, career consultants when planning career guidance activities.

Keywords: career guidance, modeling, system, Markov chains.

МАРКОВ ТІЗБЕКТЕРІН СОЦИОЦЕНОЗДАРДАҒЫ БОЛЖАУ МІНДЕТТЕРІНЕ ҚОЛДАНУ

Курмашева Л.Б.[1*], Нерадовская Ю.В.[2], Курмашев И.Г.[1]
[1] «М. Қозыбаев атындағы Солтүстік Қазақстан университеті» КеАҚ Петропавл, Қазақстан*
[2] ФГБОУ «Санкт-Петербург мемлекеттік экономикалық университеті», Санкт-Петербург, Ресей
[*]Хат-хабар үшін автор: lbkurmasheva@ku.edu.kz

Аңдатпа

Кәсіби даму - бұл адамдардың өмір жолына әсер ететін маңызды процесс. Студенттерді университетті таңдау кезінде, оқу процесінде қолдау оларға мансапқа қатысты маңызды шешімдер қабылдауға және олардың жұмысқа орналасуын арттыруға көмектеседі. Жұмыста Марков тізбектерін қолдана отырып, талапкердің мінез-құлқын модельдеу тәсілі ұсынылған, кейбір түсіндірулер келтірілген. Марков тізбегі ғылым мен техниканың әртүрлі салаларында стохастикалық жүйелерді модельдеу және талдау үшін кеңінен қолданылады. Зерттеу нәтижелері кәсіптік бағдарлау іс-шараларын жоспарлау кезінде университет әкімшілігі, мансап бойынша кеңесшілер үшін пайдалы болуы мүмкін.

Кілт сөздер: Кәсіби бағдар, модельдеу, жүйе, Марков тізбектері.

ПРИЛОЖЕНИЕ ЦЕПЕЙ МАРКОВА К ЗАДАЧАМ ПРОГНОЗИРОВАНИЯ В СОЦИОЦЕНОЗАХ

Курмашева Л.Б.[1*], Нерадовская Ю.В.[2], Курмашев И.Г.[1]

[1] НАО «Северо-Казахстанский университет имени М. Козыбаева» Петропавловск, Казахстан*

[2] ФГБОУ ВО «Санкт-Петербургский государственный экономический университет», Санкт-Петербург, Россия

[*]Автор для корреспонденции: lbkurmasheva@ku.edu.kz

Аннотация

Профессиональное развитие является важным процессом, который влияет на жизненный путь людей. Поддержка студентов в моменты выбора вуза, в процессе обучения может помочь им принимать важные решения в отношении карьеры и повысить их трудоустройство. В работе предложен подход к моделированию поведения абитуриента с использованием цепей Маркова, приводятся некоторые интерпретации. Цепь Маркова широко используется для моделирования и анализа стохастических систем в разных областях науки и техники. Результаты исследования могут быть полезными для администрации вуза, консультантов по карьере при планировании профориентационных мероприятий.

Ключевые слова: профориентация, моделирование, система, цепи Маркова.

Introduction

The time of study at the university is a very important stage in the life of every person, because students try themselves as a future specialist, explore potential career paths [1]. According to the theory of self-determination [2], all people have innate psychological needs, and they must be satisfied in order to ensure optimal functioning and development, as well as promote well-being and self-actualization. After receiving a bachelor's degree, some of the students decide to try themselves in science - they enter the master's degree, doctoral studies, some get a job and start building a career [3].

For universities, this process is also one of the most important, since such an indicator as the employment of graduates directly depends on the establishment of the career identity of students. Thus, research and modeling of the processes of the educational path, as well as the choice of a university of study, is relevant both for universities and for the education system as a whole.

As a result of the analysis of literary sources, the following groups of methods for modeling the processes of the educational path have been identified:

- Methods using regression models;
- Methods based on the use of neural network technologies;
- Methods based on the application of discriminant analysis;
- Methods of cluster analysis and forecasting;
- Methods based on the use of fuzzy modeling.

The use of certain methods depends on many factors, for example, the purpose of the study, the number of system states, the available data, the modeling period, and many others.

In [4] we examined the influence of various factors on the vocational guidance of applicants using the Matlab & Simulink system. As a result of the simulation, six variants of the system behavior (career guidance of the applicant) were considered using the Kalman filter under the influence of various factors. We have determined the best option for the applicant's professional orientation – the case when the applicant enrolls in preparation courses for entrance exams for the selected educational program at the university. The training program is adjusted depending on the level of training of students.

In [5], models of successful learning (i.e., the probability of not being expelled before graduation) were built using methods of discrete selection, regression analysis of time series and neural networks.

The present study was conducted in order to simulate the choice of a university by an applicant and the educational path of a student using Markov chains. This model can be useful for understanding and predicting the behavior of students when choosing a place of study, for analyzing transitions between different educational conditions, such as admission to a university, change of educational program or level of study and graduation (or expulsion).

The relevance of the study is the construction of forecasts based on mathematical approaches using Markov chains, which allows predicting the dynamics of the development of the indicator in question for objective assessment, correction and management.

Markov chains are successfully used in medical research [6], in psychology [7-8], in the field of UAV applications [9] and many others.

In work [10] the authors evaluate hydrometeorological data from the basins of various rivers using the Markov chain approach combined with artificial intelligence. Such studies are especially important now because they help predict natural disasters and contribute to the development of early warning systems in regions prone to floods and heavy downpours.

Article [11] describes the application of the mathematical apparatus of Markov chains in the process of forming a family trajectory, provides decision-making mechanisms based on this method. The application of the method in psychology shows excellent results in practice.

Medical research has successfully used Markov chains in modeling various processes. Michael A. Kouritzin [12] uses hidden Markov models in tracking disease progression. The paper also proves a new formula for changing the velocity of the Markov chain in continuous time.

In [13] the authors apply Markov chain technology in predicting the concentration of pollutants in indoor air. They prove that the use of mathematical apparatus is computationally efficient in this field.

In practice, Markov chains are used in computational linguistics, natural language Processing (NLP, Natural Language Processing), financial modeling and other processes where it is necessary to analyze a series of sequential events over time [14].

In economic research, it is also often possible to find the use of Markov chains, for example, in [15] an innovative method for assessing income inequality between African countries using a dynamic approach based on a spatial Markov chain was presented. The authors use real statistics for 42 years in 54 African countries, taking into account the peculiarities of the economic development of countries and studying various factors that influence them.

Research methods

As mentioned above, the paper proposes an approach to modeling the behavior of an applicant using Markov chains. As it is known, Markov processes are divided into discrete and continuous time processes. The first variety was considered in the work.

A Markov chain is a Markov random process with discrete time, in which its possible states $s_1, s_2, \dots, s_n$ can be enumerated in advance, and the transition from state to state occurs instantly (by a leap), but only at certain points in time.

After analyzing the literature on the topic of the study, we came to the conclusion that Markov chains are widely used in scientific works and in practice when building models. Markov chains are great for probabilistic modeling and Data Science. The relevance of such modeling remains for systems in which processes occur without aftereffect. Processes without aftereffect occur both in a variety of technical systems and sociotechnical ones.

Visually, the Markov chain is a directed weighted graph, the vertices of which are the states of a given system, and the edges are transitions between them, which can happen with a certain probability [16].

The results of the study

We considered the discrete system "Student Behavior" (S), which is in one of the states $S = \{S_1, S_2, S_3, S_4, S_5\}$:

where S_1 - is admission to a university for a bachelor's degree program (decided on the choice of university, submitted documents for enrollment);

S_2 - continuing education (the state in which the student successfully completes the course and moves on to the next one);

S_3 - changing the direction of study (a situation when a student changes his specialty);

S_4 - student's expulsion from the university (due to academic or financial debt, for health reasons);

S_5 - graduation (the state when a student successfully completes his studies) and a job placement.

We introduced assumptions:

- The system has 5 end states;
- The probability of a student's transition from state to state depends only on the current level of training and does not depend on his academic performance at school, in previous courses, thus, the system has the property of markovity, i.e. we defined this process of system change as a Markov chain. In this case we are dealing with a discrete Markov chain.

Obviously, the S_4 and S_5 states are absorbing, because after the student is expelled or receives a diploma, the system cannot switch to another state.

The transition matrix for the system under consideration has the following form:

$$D_{[4]} = \begin{bmatrix} & S_1 & S_2 & S_3 & S_4 & S_5 \\ S_1 & 0 & 0 & p_{13} & p_{14} & 0 \\ S_2 & 0 & 0 & p_{23} & p_{24} & p_{25} \\ S_3 & 0 & p_{32} & 0 & p_{34} & 0 \\ S_4 & 0 & 0 & 0 & 1 & 0 \\ S_5 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \tag{1}$$

where $\sum p_{ij} = 1$.

For visualization, the Figure 1 shows a graph of the states of the system.

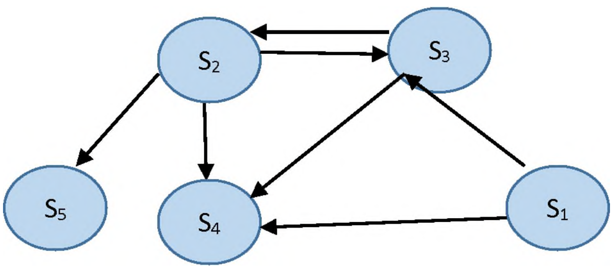

Figure 1. Graph of system S states

As it can be noticed on the graph - no arcs come out of states S_4 and S_5. These states are absorbing because when the process enters these states, it stops.

To bring the matrix D to the canonical form, we reduced it to a block form (2):

$$D_{[4]} = \begin{vmatrix} I & O \\ R & Q \end{vmatrix} \quad (2)$$

where $I_{[3]} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ – the unit matrix;

$O_{[2,3]} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ – the zero matrix;

$Q_{[3]} = \begin{bmatrix} 0 & 0 & p_{13} \\ 0 & 0 & p_{23} \\ 0 & p_{32} & 0 \end{bmatrix}$ - a matrix describing internal transitions in a system in an irrevocable set of states;

$R_{[3,2]} = \begin{bmatrix} p_{14} & 0 \\ p_{24} & p_{25} \\ p_{34} & 0 \end{bmatrix}$ - a matrix describing transitions in the system under consideration from an irrevocable set of states to an absorbing set.

To clarify, we will set specific probabilities. Let $p_{13} = 0.5$ $p_{14} = 0.5$, $p_{23} = 0{,}6$ $p_{24} = 0.4$, $p_{32} = 0.8$, $p_{34} = 0.2$ ($\sum p_{ij} = 1$).

Then we have the fundamental matrix *F:*

$$F = (I_{[2]} - Q_{[2]})^{-1} \quad (3)$$

For our case F:

$$F = \begin{bmatrix} 1 & 0{,}769 & 0{,}962 \\ 0 & 1{,}923 & 1{,}154 \\ 0 & 1{,}538 & 1{,}923 \end{bmatrix}$$

To get the average number of hits of the system in a certain state before absorption, we multiplied the fundamental matrix by a column vector, the elements of which are units, and eventually we got F*:

$$F^* = \begin{bmatrix} 2.731 \\ 3.077 \\ 3.461 \end{bmatrix} \quad (4)$$

Then the absorption probability matrix C is equal to:

$$C = \begin{vmatrix} 0.465 & 0.535 \\ 0.208 & 0.792 \\ 0.293 & 0.707 \end{vmatrix} \quad (5)$$

Judging by (5), the student who has already completed at least one year is the least likely to be expelled and successfully graduate, and the most likely to be expelled are the newly enrolled students.

Discussion

If necessary, we can expand a variety of conditions (add academic leave, summer semester, admission of an applicant to doctoral programs, etc.). We can also build models for different groups of students (for example, in different specialties or courses). And to facilitate calculations, we can use the built-in libraries of programming languages (for example, the *markovify* library in Python is widely used in research). It is necessary to periodically update the model based on new real data and changing conditions in the educational system.

Conclusion

The paper shows that in order to predict the educational path, it is advisable to accumulate information about the dynamics of student transitions between states using discrete Markov chains. The model can be improved by increasing the number of states and refining the transition matrix based on real data. However, it should be borne in mind that in multi-element systems with a large number of states, analytical modeling based on the theory of Markov processes becomes difficult. In this case, it is recommended to use the method of dynamics of averages. Its meaning also consists in using the markovity of the process, for cases of determining the average characteristics of the states of the simulated system, but calculations will be much simpler.

Markov chains can provide valuable forecasts for university administrations, helping to improve strategies for student retention and successful graduation. Career guidance providers can use the model to conduct activities that effectively support students in achieving their career goals. In the future, it is planned to implement it in one of the programming languages.

References:

1. Anne-Kathrin Kleine, Antje Schmitt, Barbara Wisse. Students' career exploration: A meta-analysis//Journal of Vocational Behavior. – 2021, Volume 131 https://doi.org/10.1016/j.jvb.2021.103645
2. Ryan R.M., & Deci E.L. Self-determination theory: Basic psychological needs in motivation, development, and wellness//The Guilford Press. – 2017 https://doi.org/10.1521/978.14625/28806
3. Mélanie Bourret, Catherine F. Ratelle, André Plamondon, Geneviève Boisclair Châteauvert. Dynamics of parent-adolescent interactions during a discussion on career choice: The role of parental behaviors and emotions//Journal of Vocational Behavior. – 2023, Volume 141 https://doi.org/10.1016/j.jvb.2022.103837
4. Kurmasheva L.B., Kulikov V.P., Kulikova V.P., Serbin V.V. Modeling of a system for assessing the quality of career guidance of applicants upon admission to a university// vestnik Almatinskogo universiteta energetiki i svyazi, No 4(63) 2023, P.119-130. https://doi.org/10.51775/2790-0886_2023_63_4_119
5. M.V. Bochenina, V. Neradovskaya, L.B. Kurmasheva, V.P. Kulikova. Uspeshnost' obucheniya v vuze: statistiko-ekonometricheskij analiz//Vestnik Altajskoj akademii ekonomiki i prava. – 2022, № 8, P.32-36.
6. Qi Zhen, Anxiao Zhang, Regina Bokel, Yue Zheng, Jing Li, Yiming Du, Qi Zhang. A Markov chain-based approach for assessing respiratory infection risk in a multi-zone office building//Journal of Building Engineering. - 2024, Volume 90 https://doi.org/10.1016/j.jobe.2024.109328
7. Sapphire Yu Han, Aart C. Liefbroer, Cees H. Elzinga. Mechanisms of family formation: an application of Hidden Markov Models to a life course process// Advances in Life Course Research. - 2020, Volume 43 https://doi.org/10.1016/j.alcr.2019.03.001
8. S. Onggo, M. Pidd, D. Soopramanien, D.J. Worthington. Behavioural modelling of career progression in the European Commission//European Journal of Operational Research. - 2012, Volume 222, Issue 3, P.632-641 https://doi.org/10.1016/j.ejor.2012.05.027
9. Qian Ning, Guiping Tao, Bingcai Chen, Yinjie Lei, Hua Yan, Chengping Zhao. Multi-UAVs trajectory and mission cooperative planning based on the Markov model // Physical Communication. – 2019, Volume 35 https://doi.org/10.1016/j.phycom.2019.100717

10. Pankaj Chauhan, Muhammed Ernur Akiner, Rajib Shaw, Kalachand Sain. Forecast future disasters using hydro-meteorological datasets in the Yamuna river basin, Western Himalaya: Using Markov Chain and LSTM approaches // Artificial Intelligence in Geosciences. - 2024, Volume 5 https://doi.org/10.1016/j.aiig.2024.100069
11. Sapphire Yu Han, Aart C. Liefbroer, Cees H. Elzinga. Mechanisms of family formation: an application of Hidden Markov Models to a life course process // Advances in Life Course Research. – 2020, Volume 43 https://doi.org/10.1016/j.alcr.2019.03.001
12. Michael A. Kouritzin. Sampling and filtering with Markov chains // Signal Processing. - 2024, Volume 225 https://doi.org/10.1016/j.sigpro.2024.109613
13. Xiaoxiao Ding, Haotian Zhang, Weirong Zhang, Yingli Xuan. Non-uniform state-based Markov chain model to improve the accuracy of transient contaminant transport prediction//Building and Environment, Volume 245, 2023 https://doi.org/10.1016/j.buildenv.2023.110977
14. Vichugova A. Markovskie cepi dlya analiza dannyh i NLP: teoriya i praktika // Shkola bol'shih dannyh. – 2022 https://bigdataschool.ru/blog/what-is-markov-chain-and-how-to-use-it-in-nlp-and-graph-analysis.html
15. Abderrahim Kerkouch, Aziz Bensbahou, Intissar Seyagh, Jamal Agouram. Dynamic analysis of income disparities in Africa: Spatial Markov chains approach // Scientific African, Volume 24,2024, https://doi.org/10.1016/j.sciaf.2024.e02236.
16. Kel'bert M.Ya., Suhov Yu.M. Veroyatnost' i statistika v primerah i zadachah. - M.: MCNMO. – 2017. T. 2. – P.560

Information about the authors:

Kurmasheva L.B. – Corresponding author, мaster, lecturer of the Department of Information and Communication Technologies, Kozybayev University, Petropavlovsk, Kazakhstan; e-mail: lbkurmasheva@ku.edu.kz;
Neradovskaya Y.W. – PhD in Economics, Associate Professor, Department of Statistics and Econometrics, Federal State Budget Educational Institution of Higher Education "Saint-Petersburg State University of Economics", Saint Petersburg, Russian Federation; e-mail: neradovskaya.yu@unecon.ru;
Kurmashev I.G. – Candidate of Technical Sciences, Head of Chair Information and Communication Technologies, Kozybayev University, Petropavlovsk, Kazakhstan; e-mail: ikurmashev@ku.edu.kz.