

СКІНЧЕННІ ПРЕДИКАТНО-КЕРОВАНІ ЛОГІЧНІ МЕРЕЖІ ДЛЯ АНАЛІЗУ ОСВІТНІХ ДАНИХ

Козирев А.Д., Шубін І.Ю.

Кафедра програмної інженерії,
Харківський національний університет радіоелектроніки,

Україна

E-mail: andrii.kozyriev@nure.ua,

igor.shubin@nure.ua

Abstract

The digitalization of higher education has caused an explosive growth of data sets about students: academic performance, attendance, participation in scientific activities and extracurricular activities. Traditional statistical or neural network methods are not always able to simultaneously consider probabilistic and logical dependencies between such parameters. The paper proposes a method of finite predicate-driven logic networks (PDLN), which integrates the apparatus of predicate logic with logic networks that have the Markov property. This approach allows building dynamic models of student states and specifying the probabilities of transitions between them based on the execution of certain predicates. Based on a generalized experiment that models the behavior of several hundred students over many semesters, it is shown that PDLN is better at identifying patterns and determining the importance of success factors than traditional methods. The results obtained indicate that the greatest positive impact on success is the average score and participation in competitions, while simple attendance at classes or active publication activity do not guarantee a high result. Predicate-driven logic networks open prospects for personalized learning and predicting the risk of failure.

Цифровізація освіти породжує величезні масиви даних про діяльність студентів: оцінки, відвідування, участь у наукових і позанавчальних заходах. Традиційні статистичні моделі або нейронні мережі добре працюють з великими даними, проте погано враховують логічні залежності між різними характеристиками чи умови, коли можливі переходи залежать від множини факторів. Автори статті сформулювали завдання створити інтегровану математичну модель, що поєднує ймовірнісні логічні мережі та логіку предикатів. Згідно з описом на ResearchGate, модель передбачає одночасне врахування стохастичних переходів і логічних залежностей та тестування на реальних даних, що підкреслює наукову новизну підходу [1].

Сильні сторони предикатно-керованих логічних мереж (ПКЛМ) полягають у тому, що він поєднує логічну інтерпретованість з можливістю враховувати стохастичні процеси. Використання алгебри скінченних предикатів дає змогу формалізувати залежності між різними показниками та управляти переходами через логічні умови. Подібні підходи вже застосовувалися в різних галузях: наприклад, у робототехніці досліджують розкладання складних предикатів на кон'юнктивні компоненти для побудови логічних мереж; у праці [2] показано, що кон'юнктивна декомпозиція формул скінченних предикатів полегшує побудову моделей, зменшує кількість допоміжних змінних та створює підґрунтя для паралельних обчислень. Поєднання предикатної алгебри з логічними мережами дає можливість моделювати складні функціональні залежності у навчальному процесі. Подальше вдосконалення ПКЛМ можливе через покращення способів оброблення предикатів. У науковій роботі, присвяченій кон'юнктивній декомпозиції формул скінченних предикатів, показано, що двійкова декомпозиція дозволяє мінімізувати кількість допоміжних змінних та полегшує побудову логічних мереж. Це важливо для ПКЛМ, оскільки складні предикати (наприклад, взаємодія між успішністю та участю в олімпіадах) можна розкласти на простіші компоненти, що знизить обчислювальну складність та підвищить ясність моделі. Крім того, у тій же роботі зазначається, що алгебра предикатів придатна для

формалізації знань та легко паралелізується, що може прискорити обчислення на великих наборах даних [2].

Маркові логічні мережі (MLN) – це універсальне поєднання логіки першого порядку та стохастичних графових моделей, яке дає змогу легко формулювати знання у вигляді логічних правил із вагою. Нещодавні дослідження демонструють, що MLN можуть бути використані для покращення виявлення об'єктів поза розподілом (OOD) завдяки вбудуванню людських концептів та апріорних знань. Дослідники запропонували новий алгоритм для навчання інтерпретованих обмежень і показали, що інтеграція MLN підвищує ефективність систем OOD на різних наборах даних [3]. Інші роботи аналізують масштабування вагових параметрів у MLN і реляційній логістичній регресії, доводячи, що правильний вибір масштабу дозволяє розповсюджувати висновки на великі домени, зберігаючи точність моделі [4]. Для ПКЛМ використання MLN може стати потужним інструментом: можна описати логічні правила (наприклад, «студент із високою відвідуваністю та участю в олімпіадах має високу ймовірність успішного завершення») і надати їм вагу, а масштабування дозволить застосовувати модель на великих вибірках.

Дослідження в галузі логіко-керованого машинного навчання показують, що поєднання символічних правил із нейронними моделями може зменшити суперечності між прогнозами та заданими знаннями. У роботі, присвяченій забезпеченню узгодженості нейронних моделей, запропоновано структуру, яка регуляризує модель шляхом інкорпорації логічних правил у функцію втрат. Автори у роботах [5–7] визначають міру неузгодженості на базі правил та формують диференційовану функцію, що дозволяє навчатись на мічених та немічених даних. Такий підхід може бути застосований до ПКЛМ, щоб навчити нейронну мережу прогнозувати фінальний стан студента з урахуванням логічних обмежень, наприклад, що відсутність відвідування знижує ймовірність успішного завершення. Таким чином, предикативні правила виконуватимуть роль регуляризатора, зменшуючи надмірне пристосування та підвищуючи узгодженість прогнозів.

Нейросимвольні системи прагнуть поєднати переваги статистичного навчання та символічного міркування. Огляд нейросимвольного обчислення підкреслює, що інтеграція символічних правил із глибинним навчанням забезпечує інтерпретованість і здатність до узагальнення, а також відкриває нові можливості для вирішення складних завдань штучного інтелекту [8]. Дослідження про синергію між машинним навчанням та міркуванням наводять приклади, коли обмеження та фонові знання допомагають регуляризувати навчання, а машинне навчання, зі свого боку, підтримує логічне виведення, наприклад в індуктивній логічній програмувальності чи нейросимвольному ШІ [9]. Для ПКЛМ використання нейросимвольних моделей означає можливість навчати нейронні мережі, що опановують закономірності у даних, водночас дотримуючись логічних правил. Такий підхід потенційно дозволяє відмовитися від жорсткого усереднення параметрів і враховувати складні нелінійні взаємодії.

Цілями дослідження є: побудова математичної моделі ПКЛМ для аналізу академічних даних; розробка програмного засобу для реалізації моделі; оцінка ефективності підходу на узагальненому наборі даних студентів. Завдання передбачають формування теоретичної бази, розробку моделі, її застосування на практичних даних та аналіз результатів.

Предикатна логіка розширює булеву алгебру, дозволяючи працювати з змінними, які мають скінченні, але не двійкові множини значень. У системах прийняття рішень предикати описують властивості об'єктів, а предикатні рівняння задають відношення між характеристиками. Наприклад, якщо змінна y може набувати значень $\{a_1, a_2, a_3\}$, а змінна x – значень $\{b_1, b_2, b_3, b_4, b_5, b_6\}$, то предикатна залежність може обмежувати, що для $y = a_1$ дозволені тільки $x = b_1$ або $x = b_2$. Відсутність значень b_1 та b_2 автоматично виключає можливість $y = a_1$. Оператори підстановки використовуються для зменшення кількості змінних у моделі шляхом фіксації певних параметрів. Вони можуть бути обмежувальними (зберігають логічні зв'язки) або розподільними (послаблюють зв'язки), що дозволяє спростити предикатні вирази або залишити ключові залежності. У моделі ПКЛМ предикатні рівняння формують базу правил для визначення можливих переходів між станами студентів.

Логічні мережі (або логічні ланцюжки) моделюють послідовності станів, у яких перехід до наступного стану залежить лише від поточного (Марковська властивість). У традиційному вигляді вони описуються матрицею ймовірностей переходів $P(X_{t+1} | X_t)$. Поєднавши ці мережі з предикатною логікою, автори пропонують визначати ймовірності переходів не тільки на основі минулого стану, а й

на основі виконання певних логічних умов (предикатів). Стан системи описується набором істинних предикатів $S_t = \{P_1(x), P_2(y), \dots\}$. Ймовірність переходу до S_{t+1} обчислюється як добуток ймовірності переходу в логічній мережі та ймовірності виконання умовних предикатів, з урахуванням вагових коефіцієнтів різних параметрів. У контексті освітньої аналітики кожен стан може відображати комбіновані показники студента: рівень успішності, відвідуваність, наукову й позанавчальну активність. Предикати можуть відповідати умовам «успішність висока», «відвідуваність низька» тощо. Такий підхід дозволяє моделювати динаміку студентів у часі та оцінювати, як виконання певних умов змінює ймовірність переходу до нового стану. Наприклад, регулярне відвідування занять збільшує ймовірність успішного складання іспиту, що в моделі відображається зміною відповідних переходів.

Щоб продемонструвати можливості ПКЛМ, було створено узагальнену модель, яка симулює дані приблизно 200 студентів протягом десяти семестрів. Для кожного студента фіксувалися чотири параметри: середній бал (GPA) за семестр, кількість відвіданих аудиторних годин, кількість наукових публікацій та рівень участі в олімпіадах і конкурсах. Кожен параметр кодувався трьома категоріями: низький (1), середній (2), високий (3).

Стан студента в семестрі визначався як кортеж з чотирьох значень (G, A, P, C) . Для спрощення розрахунків застосовували середнє арифметичне цих параметрів: $S_{n,m} = (G_{n,m} + A_{n,m} + P_{n,m} + C_{n,m})/4$. Це зменшувало число унікальних станів з 81 до неперервного діапазону від 1,0 до 3,0, що суттєво оптимізувало матрицю переходів.

Модель працювала таким чином: на основі історичних даних будувалася матриця ймовірностей переходів між станами для кожного студента; ймовірності коригувалися з урахуванням виконання предикатів (наприклад, «відвідуваність >2» чи «кількість публікацій =1»). Для кожного студента прогнозувався ймовірніший фінальний стан після завершення семестру, обчислюючись добуток ймовірностей переходів по всіх семестрах. Розроблений програмний модуль дозволяв візуалізувати мережу станів, їхні зв'язки та зміну ймовірностей.

Результати узагальненого експерименту підтвердили висновки оригінальної статті. Більшість студентів зосереджувалася в діапазоні станів 1,5–2,0, що відповідає середнім показникам за всіма параметрами. Перехід до високих станів (2,5–3,0) був рідкісним та вимагав суттєвих змін одразу у кількох параметрах. Розподіл ймовірностей переходів показав, що стабільні зміни в межах близьких станів мають ймовірність близько 0,1–0,24, а різкі стрибки трапляються рідко.

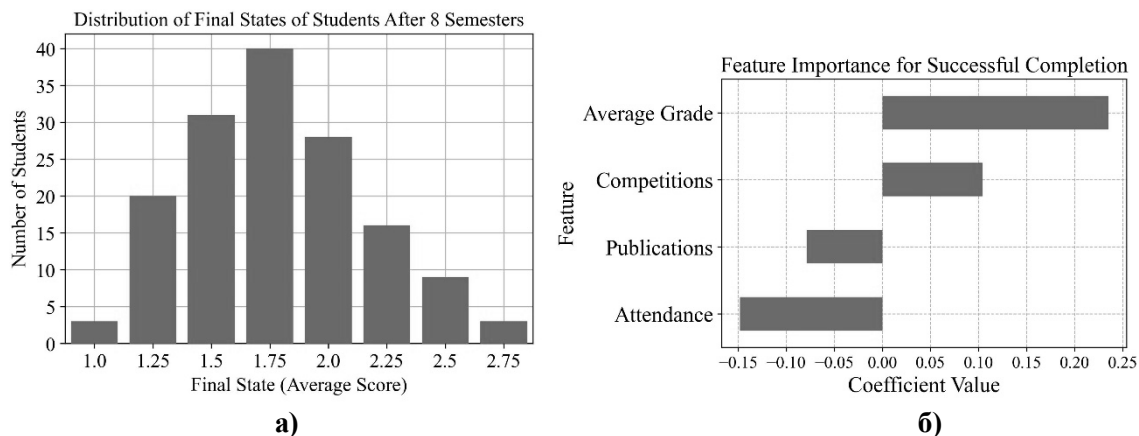


Рис. 1. а) Результати розподілення станів студентів; б) оцінка важливості освітніх ознак

Для оцінки важливості ознак застосовували множинну лінійну регресію, де цільовою змінною був фінальний стан студента. Аналіз показав, що найбільший позитивний внесок у досягнення високого фінального стану роблять середній бал та участь у конкурсах, тоді як відвідуваність та кількість публікацій можуть мати нейтральний або навіть негативний ефект. Це узгоджується з гіпотезою, що активність у науково-творчих заходах стимулює глибше розуміння матеріалу та розвиток м'яких навичок, а просте накопичення годин чи публікацій не гарантує якісного засвоєння.

Проведений узагальнений експеримент засвідчив, що предикатно-керовані логічні мережі забезпечують детальніший аналіз академічних даних порівняно з традиційними підходами. Модель дозволяє не лише прогнозувати ймовірність переходів між станами, але й інтерпретувати, які

предикати (умови) сприяють або заважають переходу. Це забезпечує прозорість і можливість підлаштовувати освітній процес відповідно до реальних потреб студентів.

Порівняння з лінійною регресією та нейронними мережами показало, що ПКЛМ краще враховує динаміку змін параметрів та залежності між ними. Зокрема, модель виявила, що участь у конкурсах та високі оцінки мають найбільший ваговий коефіцієнт у прогнозуванні успіху, тоді як високі показники відвідуваності чи наукових публікацій іноді відволікають від навчання, знижуючи загальну успішність. Традиційні методи часто не виділяють цього ефекту, оскільки припускають лінійну залежність між усіма параметрами.

Крім того, модель дозволяє персоналізувати поради студентам. Наприклад, студент зі стабільним середнім балом, але низьким рівнем участі у конкурсах, може отримати рекомендацію активніше брати участь у командних олімпіадах або проєктах, оскільки це істотно підвищує ймовірність переходу до високого стану. Інший студент із високою відвідуваністю, але низьким середнім балом може дізнатися, що просто відвідувати заняття недостатньо — варто зосередити увагу на якісному освоєнні матеріалу. Такий рівень деталізації недоступний у простіших моделях.

Дослідження відкриває перспективи для подальшої роботи. По-перше, модель можна розширити за рахунок врахування психологічних і соціально-економічних факторів — рівня стресу, психічного здоров'я, фінансових умов сім'ї тощо. По-друге, інтеграція машинного навчання (наприклад, методів підкріплення) допоможе адаптивно налаштовувати ваги предикатів та підвищити точність прогнозів. Нарешті, перевірка моделі у різних освітніх контекстах — від онлайн-курсів до університетів різних країн — дозволить оцінити її універсальність та скоригувати структуру предикатів під конкретні умови. ПКЛМ має потенціал стати основою персоналізованих систем підтримки студентів, що сприяють успішному та усвідомленому навчальному процесу.

Література

1. Dudar, Z., Kozyriev, A. (2024), Finite predicate-driven logic networks method for enhanced education data analysis. *Radioelectronic and Computer Systems*, Vol. 2024, No. 3, P. 205–215. DOI: <https://doi.org/10.32620/reks.2024.3.14>.
2. Volkov, V., Shubin, I., Pliekhova, G., Kopytkov, D., Volkova, T. (2025), Decomposition of conjunctive formulas in the algebra of finite predicates for modeling logical structures of transport networks. *Transport, Ecology – Sustainable Development*, Vol. 32, P. 128–144. URL: <https://ecovarna.org/pdf/2025Eko18.pdf>.
3. Kirchheim, K., Ortmeier, F. (2025), Improving out-of-distribution detection with Markov logic networks. arXiv preprint arXiv:2506.04241. DOI: <https://doi.org/10.48550/arXiv.2506.04241>.
4. Weitkämper, F. Q. (2025), Scaling the weight parameters in Markov logic networks and relational logistic regression models. *Machine Learning*, Vol. 114, No. 3. DOI: <https://doi.org/10.1007/s10994-024-06635-7>.
5. Li, T., Wang, H., Chen, B. та ін. (2019), A logic-driven framework for consistency of neural models. arXiv preprint arXiv:1909.00126. DOI: <https://doi.org/10.48550/arXiv.1909.00126>.
6. Zhang, L. та ін. (2025), Data augmentation for sparse multidimensional learning performance data using generative AI. arXiv preprint arXiv:2409.15631. DOI: <https://doi.org/10.48550/arXiv.2409.15631>.
7. Wang, W., Yang, Y., Wu, F. (2022), Towards data- and knowledge-driven artificial intelligence: a survey on neuro-symbolic computing. arXiv preprint arXiv:2210.15889. DOI: <https://doi.org/10.48550/arXiv.2210.15889>.
8. Baaj, I., Amel, K. R. та ін. (2024), Synergies between machine learning and reasoning: an introduction by the Kay R. Amel group. *International Journal of Approximate Reasoning*, Vol. 171, Article 109206. DOI: <https://doi.org/10.1016/j.ijar.2024.109206>.
9. Vechirska, I., Sokolova, A., Valenda, N. (2025), Construction of a formal model of the excursion selection process for students: a case study of the “Push” school in Kharkiv in the form of a logical AFP-network. *Computer Systems and Information Technologies*, No. 3, P. 17–27. DOI: <https://doi.org/10.31891/csit-2025-3-2>.