

## РОЗРОБКА МЕТОДУ КЛАСИФІКАЦІЇ СТУДЕНТІВ НА ОСНОВІ КЛАСТЕРНОГО АНАЛІЗУ ПРЕЦЕДЕНТІВ В АДАПТИВНІЙ НАВЧАЛЬНІЙ СИСТЕМІ

*Розглянуто стан, проблеми і перспективи розробки адаптивної системи дистанційної освіти. Запропоновано метод класифікації студентів на основі кластерного аналізу прецедентів, побудований шляхом застосування матриці сценарних прикладів, який дозволяє проводити успішне адаптивне навчання студентів в залежності від їх поточних навчальних успіхів та структури навчального матеріалу. Розроблена математична модель адаптивного управління навчанням студента та описана практична реалізація запропонованого методу.*

*В роботі запропоновано метод класифікації студентів на основі кластерного аналізу прецедентів, що застосовується в дистанційній навчальній системі для адаптивного навчання студентів. Побудована матриця сценарних прикладів та виконано групування студентів на основі 5-ти категорій, що дозволяє управляти найбільш типовими ймовірнісними переходами студента по навчальних режимах. Розроблена математична модель адаптивного управління навчанням студента та описана практична реалізація запропонованого методу.*

*Ключові слова: адаптивне навчання, метод класифікації, кластерний аналіз, прецедентне правило.*

M.V. PIKULYAK

Vasyl Stefanyk Precarpathian National University

### DEVELOPMENT OF THE METHOD OF CLASSIFICATION OF STUDENTS ON THE BASIS OF CLUSTER ANALYSIS OF PRECEDENTS IN THE ADAPTIVE EDUCATIONAL SYSTEM

*Abstract – In the article the author describes the method of classification of students on the basis of cluster analysis of precedents, that is used in distance educational system for the adaptive learning of students.*

*The matrix of scenario examples is created, that allows conducting successful adaptive learning of students according to their academic success and current structure of educational material.*

*A cluster analysis is carried out performed and grouping of students is performed based on 5 categories, which allows to manage the most typical probabilistic transitions of a student through educational regimes.*

*A three digit mathematical apparatus of logic for automatic number identification of educational regime is offered that gives an opportunity to generate educational quanta based on mathematical representation of the knowledge structure and automatically create an individualized adaptive content. A mathematical model of adaptive learning management is developed and practical implementation of the offered method is described.*

*Keywords: adaptive learning, model of a student, method of classification, cluster analysis, precedent rule.*

#### Вступ

Питанню розробки адаптивних навчальних систем в даний час приділяється все більше уваги, оскільки впровадження подібних навчальних продуктів в освітній процес, а особливо з поширенням дистанційної форми навчання, набуває все більш вагомого значення в навчальному процесі.

Оскільки різні студенти підходять до початку процесу вивчення нової порції інформації (кванта знань) з різним початковим рівнем знань, то використання адаптивних технологій під час автоматизованого навчання дозволяє індивідуалізувати навчальний процес як за рахунок врахування особистих характеристик студента, відношення його до сприйняття різного роду нових знань так і врахуванням поточних навчальних успіхів.

В роботі вирішується задача класифікації студентів, тобто запропоновано розбиття всіх студентів на групи за результатами поточної тестової перевірки, що дозволяє індивідуалізувати процес дистанційної освіти.

#### Постановка задачі

Задача кластеризації відноситься до статистичної обробки та дозволяє виконати збір даних, що містять інформацію про вибірку об'єктів, і потім впорядковувати об'єкти в порівняно однорідні групи [1].

Для вирішення практичних задач кластеризації застосовують різні класифікаційні методи [2]:

- бейєсівські методи класифікації;
- метричні методи класифікації;
- лінійні методи класифікації (метод стохастичного градієнту, метод логістичної регресії, метод опорних векторів);
- штучні нейронні мережі.

Запропонований в роботі метод кластеризації із застосуванням прецедентного підходу дозволяє не тільки сформувати класи студентів на основі аналізу міри близькості, а й в процесі побудови індивідуальної навчальної траєкторії використати знання про попередні навчальні ситуації та випадки, а також надати допомогу в прийнятті навчальних рішень у процесі засвоєння нових знань.

#### Кластерний аналіз прецедентів

Запропонований в роботі метод полягає в тому, що за відомою множиною сценарних прикладів (прецедентних рішень)  $X = \{X_1, X_2, \dots, X_n\}$  та з використанням функції відстані між прецедентами

$P(X, X_i)$  вирішується задача кластеризації студентів за результатами перевірки поточного рівня їх знань.

Кожний сценарний приклад представлений предикатним правилом:

$$\text{ЯКЩО } P_1 \text{ і } P_2 \text{ і } \dots \text{ і } P_j, \text{ ТО } R_k,$$

де  $P_i$  – студентські параметри,  $R_k$  – номер відповідного правила [3].

Математично даний метод дозволяє проаналізувати функціональні залежності між вхідними та вихідними параметрами навчальної системи, в результаті чого відбувається адаптація системи до поточного рівня навченості студента.

Для формалізації методу необхідна послідовна реалізація наступних інформаційних процесів [4]:

1) вхідні параметри представлені масивом прецедентів, що накопичені як за рахунок змодельованих випадків, так і випадками з практики навчальної поведінки студентів:

$$X_i \rightarrow F_1 [X_i, X_{i-1}, \dots, X_{i-j}, \dots, X_{i-n}], \quad (1)$$

де  $X_i$  – окремий прецедент. Дана сукупність утворює матрицю прецедентів.

2) вихідні параметри формуються на основі інформації з багатопараметричної моделі студента, що описується:

а)  $G_i \rightarrow F_2 [G_1, G_2, \dots, G_l]$  – характеристики, що відображають початковий рівень розуміння студентом курсу, що вивчається;

б)  $P_i \rightarrow F_3 [P_1, P_2, \dots, P_k]$  – параметри, що характеризують поточний рівень набутих нових знань.

3) у випадку появи невідомої навчальної ситуації поведінки студента у системі відбувається пошук подібного прецедента, який використовується в ролі аналога з метою його адаптації до поточного випадку:

$$X_i \equiv X_{i-j} = F_4 \begin{cases} 1, & j = j+0 \\ 0, & j = j+1 \end{cases},$$

де 1 – наявність прецедента  $X_i$  в масиві (1), 0 – відповідно відсутність прецедента  $X_i$  в масиві (1);  $j = j+1$  – відповідно нарощення бази прецедентів.

Після того, як нова ситуація буде опрацьована, вона вноситься в базу прецедентів разом зі своїм вирішенням. Завдяки цьому база прецедентів інкрементно розширюється і у випадку ефективного застосування нового елемента для моделювання навчальної траєкторії студента, прецедент, який відповідає опису створеної ситуації, використовується в аналогічних випадках для інших студентів.

Таким чином, загальна формалізація представляється інформаційною технологією конвеєрного виконання функціоналів [4]:

$$F_{AH} = F_1(t) \Rightarrow F_2(t) \Rightarrow F_3(t) \Rightarrow F_4(t) \begin{cases} [j = j+0, & X_j \neq \{X_i, X_{i-1}, \dots, X_{i-j}, \dots, X_{i-n}\}] \\ [j = j+1, & X_j \equiv \{X_i, X_{i-1}, \dots, X_{i-j}, \dots, X_{i-n}\}] \end{cases}$$

Система, побудована за таким принципом є самонавчальною: чим більше прецедентів міститься в базі, тим ширший спектр їхніх можливих значень, тим вища ймовірність знайти найбільш адекватний прецедент по функціональних характеристиках та оптимізаційних параметрах, отже, вища якість прийнятого рішення [5].

### Побудова бази даних прецедентів

Стан окремого студента в момент часу  $t$  представимо у вигляді

$$Y(t) = (M, K_n(t)),$$

де  $M$  – модель студента, побудована на множині  $P_i$ ;

$K_n(t)$  – кількість квантів інформації, вивчених за час  $t$ .

Позначимо через  $Y(t), Y(t+1)$  – стан студента до і після взаємодії із системою. Адаптивну взаємодію системи із студентом визначимо як  $U(t) = (H_k(t))$ , де  $H_k(t)$  – послідовність навчальних кроків.

Для оцінки результатів навчання побудуємо множину класів станів студента відповідно до критеріїв  $Q = \{Q_l, Q_{pr}\}$ , де  $Q_l$  – клас, до якого відноситься стан студента після сеансу навчання;  $Q_{pr}$  – клас, якого можна було досягти за прогнозом навчальної системи.

Результатом навчання будемо називати функцію  $I(t) = IP\{Q_l, Q_{pr}\}$ , де  $IP\{Q_l, Q_{pr}\}$  – функція оцінки якості навчання.

Таким чином, прецедент представляє собою наступний набір даних:

$$PR(t) = (Y(t), U(t), Y(t+1), I(t)).$$

Оскільки в основу дослідження моделі студента покладено аналіз 5-ти параметрів  $P_i$  ( $P_i$  –

загальний рівень засвоєння навчального матеріалу;  $P_2$  – глибина знань;  $P_3$  – ступінь засвоєння матеріалу;  $P_4$  – якість засвоєння матеріалу;  $P_5$  – час, витрачений на навчання), кожен з яких приймає значення з множини  $\{H, C, B\}$ , що визначена на проміжку  $[0..1]$  («низький» –  $[0..0,4)$ , «середній» –  $[0,4..0,8)$  та «високий» –  $[0,8..1]$  відповідно), то загальна кількість можливих прецедентів  $N$  дорівнює:  $N = 3^5 = 243$ . Це значення отримано на основі комбінаторного принципу добутку про кількість заповнення 5-ти місць трьома елементами. З метою скорочення кількості прецедентів використана теорія розбиття натуральних чисел на цілі невід’ємні доданки, яка дозволяє побудувати 21 базовий прецедент та виконати групування по 6 категоріях (табл. 1).

Таблиця 1

**Групова класифікація студентів з прив’язкою до режиму навчання**

№ за/п	Група	$H + C + B = 5$	Ймовірність вибору режиму		
			$R_1$	$R_2$	$R_3$
1	1	5 0 0	0,9	0,05	0,05
2		4 1 0	0,8	0,1	0,1
3		3 2 0	0,6	0,3	0,1
4		2 3 0	0,3	0,6	0,1
5		4 0 1	0,8	0,1	0,1
6	2	3 1 1	0,6	0,2	0,2
7		2 2 1	0,4	0,4	0,2
8		2 1 2	0,4	0,2	0,4
9		1 4 0	0,1	0,8	0,1
10		1 3 1	0,2	0,6	0,2
11	3	0 5 0	0,05	0,9	0,05
12		0 4 1	0,1	0,8	0,1
13		0 3 2	0,1	0,6	0,3
14	4	1 1 3	0,2	0,2	0,6
15		1 2 2	0,2	0,4	0,4
16	5	0 0 5	0,05	0,05	0,9
17		0 1 4	0,1	0,1	0,8
18		1 0 4	0,1	0,1	0,8
19		0 2 3	0,1	0,3	0,6
20	6*	3 0 2	0,05	0,05	0,05
21		2 0 3	0,05	0,05	0,05

\* – клас неіснуючих (аномальних) прецедентів.

Побудоване групування на основі 5-ти категорій дає можливість управляти найбільш типовими (класичними) ймовірнісними переходами студента по навчальних режимах.

Для підрахунку матриці відстаней між елементами  $P_i$  використана метрика Евкліда

$$d_{ij} = \sqrt{\sum_{k=1}^n (X_{ik} - X_{jk})^2},$$

застосування якої дозволяє підрахувати середню різницю відстаней між різними параметрами та побудувати 6 класів прецедентів.

У результаті проведених експериментальних досліджень різних груп та студентів різного рівня знань кожен параметр  $P_i$  проаналізований з точки зору впливу на загальний рівень знань студента та степінь зв’язку з іншими параметрами. Це дозволило визначити структуру відношень кожного класу (побудована матриця близькості між окремими класами) та ймовірності прийняття рішення про вибір навчального режиму перенавчання  $R_1$ , донавчання  $R_2$  чи навчання  $R_3$ .

Проведений кластерний аналіз дозволяє:

- розбити множину  $X_i$  прецедентів на класи подібних об’єктів, що спрощує процес реалізації системи;
- будувати навчальну траєкторію студента по групах, що дозволяє прискорити процес обробки навчальної інформації;
- зменшити об’єм даних, тобто скоротити кількість прецедентів, завдяки використанню тільки базових сценаріїв для кожного класу;
- виділити нетипові (неіснуючі) сценарії, що не використовуються в навчальному процесі та відносяться до класу аномальних прецедентів;

- виконати ієрархію на множині сценаріїв, що дозволяє реалізувати побудову індивідуальної адаптивної траєкторії навчання шляхом переведення студентів з нижчих класів ієрархії у вищі.

### Обробка матриці прецедентів із застосуванням математичної тризначної логіки

Оскільки у розробленій системі рівень оволодіння студентом навчальним матеріалом студентський модуль визначає на основі аналізу 5-ти параметрів  $P_i$ , то вводиться означення тризначного предиката  $P$  на множині  $M$  у вигляді 5-місної функції, визначеної на  $M$  із значеннями з множини  $\{B, C, H\}$ . В такому випадку окреме прецедентне рішення представляється у вигляді предикату  $P$  як:

$$P_1(H_1, H_2, C_3, C_4, H_5), P_2(C_1, C_2, B_3, B_4, C_5), P_3(B_1, C_2, C_3, B_4, B_5) \text{ і т.д.}$$

З метою математичного обрахунку номера навчального режиму для матриці прецедентів будують аналог досконалої диз'юнктивної нормальної форми (ДДНФ) [6]:

$$H_1 H_2 C_3 C_4 H_5 | R_1 \vee \dots \vee C_1 C_2 B_3 B_4 C_5 | R_2 \vee \dots \vee B_1 C_2 C_3 B_4 B_5 | R_3, \quad (2)$$

де  $R_i$  – номер режиму навчання, що відповідає окремому прецеденту.

Припустимо, що за результатами поточного тестування рівень оволодіння навчальним матеріалом для деякого студента  $M_1$  представлений, наприклад, вектором  $P(C_1, C_2, B_3, H_4, B_5)$ . Використовуючи ДДНФ, визначимо номер режиму, по якому необхідно продовжити навчання для студента  $M_1$ .

Беручи до уваги те, що операція кон'юнкції у двозначній логіці визначається як мінімум [7]:

$$(a \wedge b) = \min(a, b), \quad a, b \in [0, 1],$$

емпірично визначені кон'юнктивні правила для означеної модифікованої тризначної логіки:

$$| a_{ik} \wedge a_{jk} | = \begin{cases} 1, & |i - j| = 0 \\ 2/3, & |i - j| = 1 \\ 1/3, & |i - j| = 2 \end{cases}, \quad i, j \in \{1, 2, 3\}; k \in \{1, 2, 3, 4, 5\}$$

де  $H \rightarrow a_1, C \rightarrow a_2, B \rightarrow a_3$  та побудована таблиця відповідності виконання кон'юнктивної операції стосовно різних комбінацій значень  $B, C, H$  (табл. 2).

Таблиця 2

Кон'юнктивні операції модифікованої тризначної логіки			
	$H - [0..0,4)$	$C - [0,4..0,8)$	$B - [0,8..1]$
$H - [0..0,4)$	1	2/3	1/3
$C - [0,4..0,8)$	2/3	1	2/3
$B - [0,8..1]$	1/3	2/3	1

Розглядаючи вектор  $P$  як один кон'юнктивний член ДДНФ та використовуючи означені вище кон'юнктивні правила, в результаті множення вектора  $P$  на матрицю прецедентів, представлену у формі (2):

$$\begin{aligned} & C_1 C_2 B_3 C_4 H_5 \wedge (H_1 H_2 C_3 C_4 H_5 | R_1 \vee \dots \vee C_1 C_2 B_3 B_4 C_5 | R_2 \vee \dots \vee B_1 C_2 C_3 B_4 B_5 | R_3) = \\ & = \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{2}{3} \cdot 1 \cdot 1 | R_1 \vee \dots \vee 1 \cdot 1 \cdot 1 \cdot \frac{2}{3} \cdot \frac{2}{3} | R_2 \vee \dots \vee \frac{2}{3} \cdot 1 \cdot \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{1}{3} | R_3 = \\ & = 0,1 | R_1 \vee \dots \vee 0,4 | R_2 \vee \dots \vee 0,07 | R_3 = 0,4 | R_2 \end{aligned} \quad (3)$$

отримано номер режиму  $R_2$ , причому ймовірність вибору системою цього режиму дорівнює 0,4.

При обчисленні номера режиму у формулі (3) береться до уваги, що диз'юнкція декількох значень у тризначній логіці (за аналогією двозначної) дорівнює максимуму серед цих значень:  $(a \vee b \vee c) = \max(a, b, c)$ .

Описаний механізм дозволяє математично визначити номер навчального режиму, продовження навчання по якому дає змогу адаптувати навчальний контент відносно успіхів, які досягнув студент на деякому поточному етапі навчання.

На рисунку 1 показано графічне відображення кластерного аналізу для різних груп студентів, що дозволяє провести аналіз переходу студента по групах під час адаптивного подання навчальної інформації.

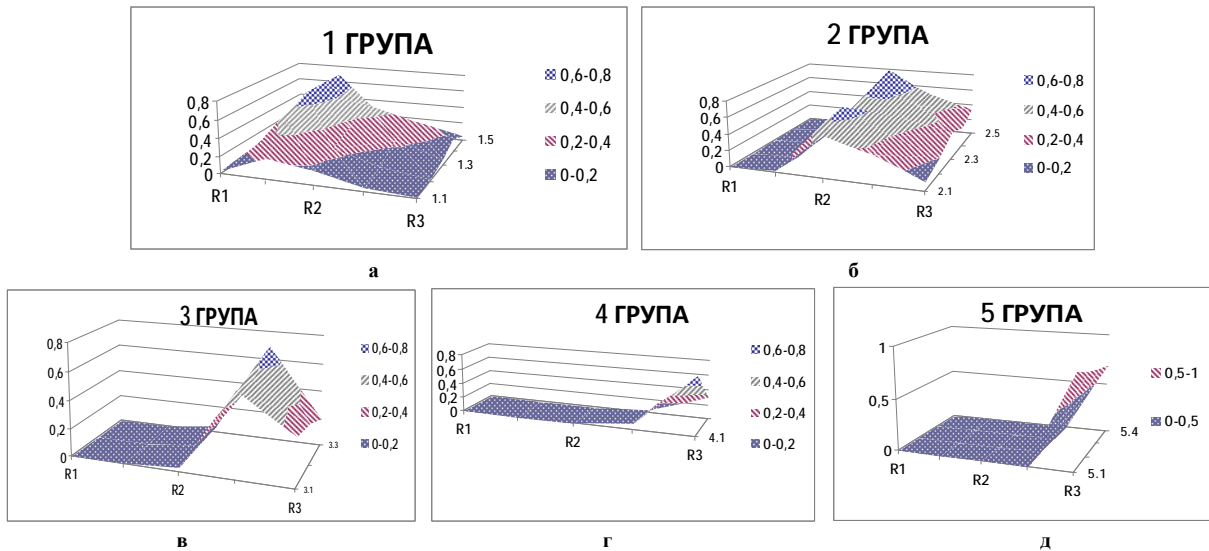


Рис. 1. Групове представлення класифікації студентів

**Приклад практичної реалізації з використанням інструментального засобу ALS**

Під час вивчення дисципліни «Бази даних та інформаційні системи» студент  $M_2$  за результатами поточної тестової перевірки отримав наступні значення студентських параметрів:  $P_1 - H$ ;  $P_2 - H$ ;  $P_3 - C$ ;  $P_4 - C$ ;  $P_5 - C$ .

Побудова математичної моделі адаптивного управління навчанням студента реалізована в середовищі MATLAB. Для цього використовується функція приналежності [8], яка кожному значенню параметра  $P_i$  ставить у відповідність число з інтервалу [0..1].

Для кожного параметра  $P_i$  формується послідовність трапецієподібних функцій, визначених на множині  $\{B, C, H\}$  (рис. 2).

В загальному випадку трапецієподібна функція приналежності представляється наступним аналітичним виразом:

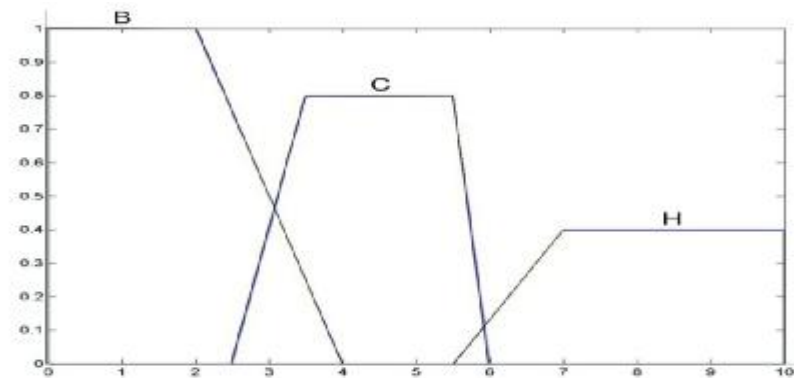


Рис. 2. Вигляд трапецієподібних функцій приналежності для значень B, C, H

$$f_{\text{трап}} = \left. \begin{array}{l} 0, \quad x \leq a \\ \frac{x-a}{b-a}, \quad a \leq x \leq b \\ 1, \quad b \leq x \leq c \\ \frac{d-x}{d-c}, \quad c \leq x \leq d \\ 0, \quad d \leq x \end{array} \right\},$$

де  $a, b, c, d$  – деякі числові параметри, що приймають довільні числові значення і впорядковані відношенням:  $a \leq b \leq c \leq d$ .

Використовуючи матрицю прецедентів, сформовану експертами на основі таблиці 1, побудовано інструментальний засіб ALS, реалізований на мові Delphi. Даний програмний продукт шляхом застосування математичного апарату трізначної математичної логіки автоматично відносить студента до певної групи кластеризації та вибудовує подальшу навчальну траєкторію на основі застосування обчисленого прецедента. Так, на рис. 3 відображено вікно результатів програмної обробки навчання студента  $M_2$ , якому відповідає вектор  $P(H, H, C, C, C)$  поточних параметрів та функція приналежності, що описується вектором числових значень кожної терми та визначає вузлові значення координат трапецієподібної функції.

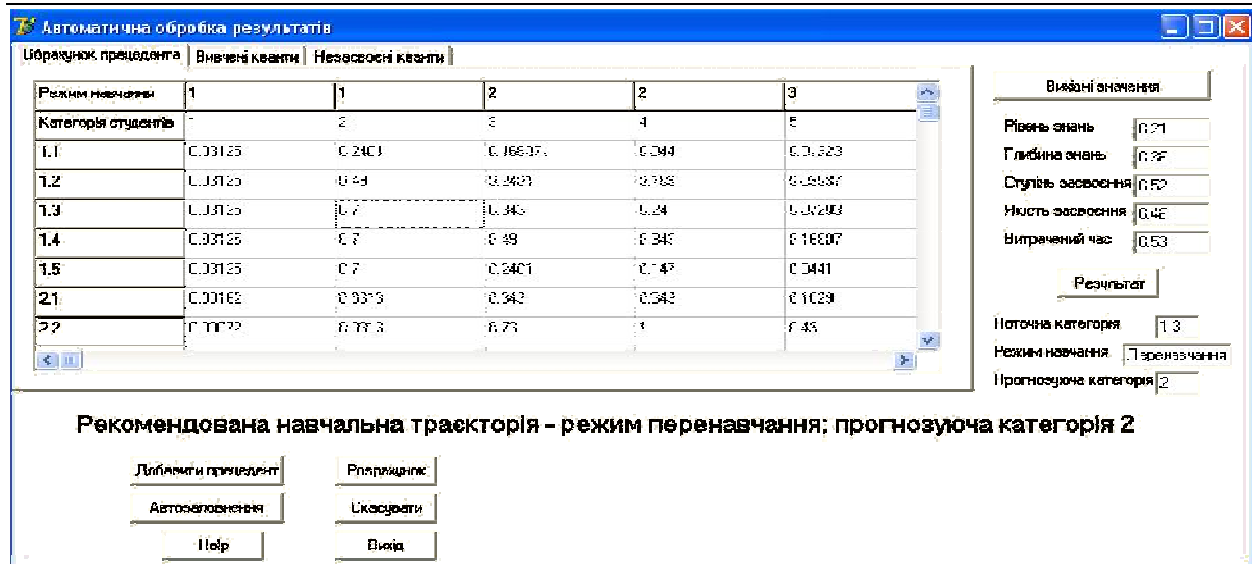


Рис.3 – Вікно програмної реалізації кластерного аналізу

Таким чином, за результатами поточної тестової перевірки студенту М навчальна програма формує набір невивчених навчальних квантів, та направляє його траєкторію навчання по прецеденту з максимальним значенням 0.7, який передбачає використання режиму повторного вивчення навчального матеріалу з прогнозуючим попаданням у 2-гу категорію.

Отриманий програмний продукт дозволяє змоделювати довільну навчальну ситуацію, описуючи її вектором поточних студентських параметрів (для кожного параметра вводиться набір термів та трапецеподібна функція приналежності, що відповідає кожній термі) та матрицею прецедентів.

### Висновки

Запропонований метод класифікації студентів відрізняється від відомих тим, що побудований на основі матриці прецедентів, представлена у вигляді сценарних прикладів і алгебри висловлювань та дозволяє не тільки віднести студента до окремого класу, але й вибрати найбільш адаптований до нього сценарний приклад.

Це дає можливість проводити успішне адаптивне навчання студентів в залежності від їх поточних навчальних успіхів та структури навчального матеріалу.

### Література

1. Арефьев В. П. Кластерный анализ результатов оценивания знаний в системе заочного обучения с использованием дистанционных образовательных технологий / В. П. Арефьев, А. А. Михальчук, Н. М. Филипенко // Современные проблемы науки и образования. – 2013. – № 3 (Электронный журнал). – Режим доступа : <http://www.science-education.ru/109-9506>.
2. Воронцов К. В. Математические методы обучения по прецедентам (теория обучения машин) [Электронный ресурс] / К.В. Воронцов .– Режим доступа: <http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf>
3. Федорук П. И. Использование сценарных примеров знаний при построении индивидуальной учебной траектории / П. И. Федорук, Н. В. Пикуляк // Программные продукты и системы. – Тверь, 2011. – № 2 (94). – С. 89–94.
4. Pikulyak M. The method of formalization of adaptive learning model based on precedents matrix / Mykola Pikulyak // Досвід розробки та застосування приладо-технологічних САПР в мікроелектроніці : CADSM 2015 : матеріали XIII Міжнародної науково-практичної конференції. – Львів : Вид-во Львівської політехніки, 2015. – С. 189–192.
5. Юдин В. Н. Гибридный подход к построению систем поддержки решений / В. Н. Юдин, Л. Е. Карпов // Труды Института системного программирования РАН (ИСП РАН). – 2013. – Т. 24. – С. 447–456.
6. Никольская И. Л. Математическая логика : учебник / И. Л. Никольская. – М. : Высш. школа, 1981. – 127 с.
7. Колмогоров А. Н. Математическая логика / А. Н. Колмогоров, А. Г. Драгалін. – Изд. 3-е, стереот. – М. : КомКнига, 2006. – 240 с.
8. Заде Л. Понятие лингвистической переменной и его применение к принятию приближенных решений / Л. Заде. – М.: Мир, 1976. – 165 с.

Рецензія/Peer review : 21.9.2015 р.

Надрукована/Printed : 2.11.2015 р.

Стаття рецензована редакційною колегією