

Нечеткое представление результатов кластеризации студентов

*Харьковская национальная академия городского хозяйства,
Харьковский автомобильно-дорожный университет*

Предложен метод формализации результатов распределения студентов по типологическим группам. Процедура классификации студентов представлена в виде нечеткой функции. Графическим способом определены значения функции принадлежности перехода студентов в другую типологическую группу. Разработаны продукционные модели предпочтения студентов по оценке их компетенций. Получено множество недоминирующих альтернатив для каждой модели поведения студента. Формализованы предпосылки перераспределения студентов по типологическим группам с учетом предпочтений студентов.

Ключевые слова: нечеткое множество, кластеризация, типологические группы, нечеткая функция, функция принадлежности, продукционная модель, язык предикатов.

1. Постановка проблемы

На современном этапе развития образования Украины особое внимание уделяется созданию единой информационно-компьютерной системы управления образованием. Попытки создания такой системы наталкиваются на значительные трудности, связанные с исследованием многообразных слабоструктурированных когнитивных, учебных и образовательных процессов. Преодоление указанных трудностей возможно с помощью разработки методов формализации слабоструктурированных процессов управления в сфере образования на основе теории принятия решений, эвристических методов представления знаний теории искусственного интеллекта, теории нечетких множеств и других методов.

2. Анализ последних исследований и публикаций

В настоящее время имеется значительное количество работ, в которых обсуждаются проблемы математического моделирования и управления процессом обучения студентов [1, 2] с помощью различного математического аппарата, описываются принципы формализации процессов формирования компетентностей студентов [3], а также рассматриваются математические способы прогнозирования успеваемости студентов [4]. Прогнозирование должно быть направлено на обоснование целей, содержания, методов, средств и организационных форм обучения, призванных практически решить проблему индивидуализации образования [5].

3. Постановка задачи

Целью настоящей статьи является представление метода формализации процесса распределения студентов по типологическим группам с учетом мотивационных предпочтений студентов на основе теории нечетких множеств.

В целях прогнозирования успеваемости поток студентов дорожно-строительного факультета ХНАДУ, изучающих дисциплину «Информатика» на первом курсе в первом семестре, в начале семестра был распределен на четыре

типологические группы методом кластерного анализа. При кластеризации были использованы три параметра:

- 1) начальные знания студентов, измеренные специальным тестированием в начале семестра;
- 2) компетенции, приобретенные студентом по первой теме дисциплины, измеренные с помощью тестирования на втором занятии;
- 3) количество пропущенных занятий на момент проведения кластеризации.

Для повышения точности кластеризации необходимо внедрить в процедуру кластеризации в качестве исходных данных факторы, оказывающие влияние на успеваемость студентов, но не учтенные ранее в связи с трудностями в их количественном измерении. К таким факторам относятся: мотивация, индивидуальные способности студентов к обучению, факторы внешней среды и др. На настоящий момент не существует четкого и однозначного руководства по измерению уровня значимости названных факторов. Поэтому будем считать качественные факторы статистически неопределенными параметрами, а процедуру кластеризации студентов на типологические группы исследуем как задачу принятия решений в условиях нечеткой информации.

4. Формализация процедуры кластеризации

Представим процедуру классификации студентов в виде нечеткой функции:

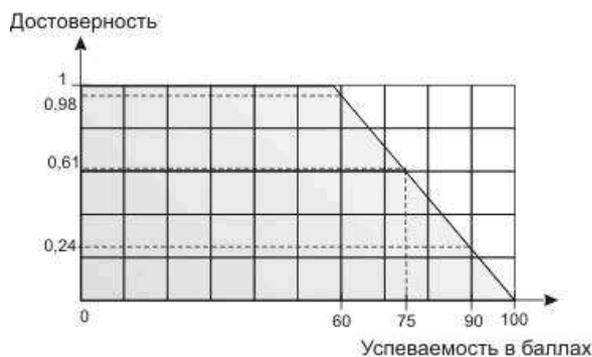
$$\tilde{K} = (X, Y, \tilde{F}), \quad (1)$$

где $X = \{x_i\}, i = \overline{1,4}$ – четкое множество типологических классов, по которым проведено распределение студентов методом кластерного анализа;

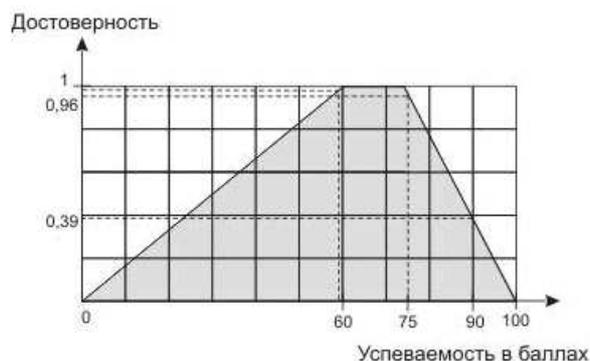
$Y = \{y_j\}, j = \overline{1,4}$ – четкое множество типологических классов, в которые возможен переход;

$\tilde{F} = \{\mu_{\tilde{F}}\langle x_i, y_j \rangle\}$ – нечеткое множество $\tilde{F} \subseteq X \times Y$, где $\mu_{\tilde{F}}\langle x_i, y_j \rangle$ – функция принадлежности, $x_i \in X, i = \overline{1,4}, y_j \in Y, j = \overline{1,4}$.

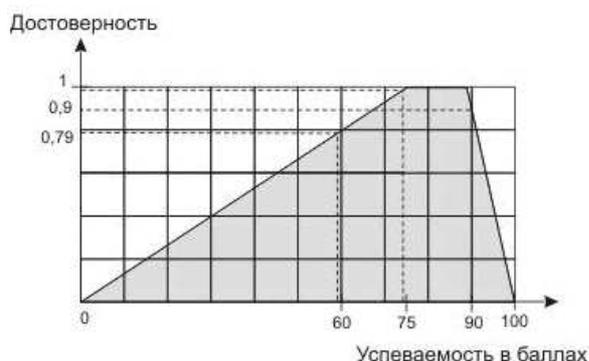
Определим возможность перехода студента в процессе обучения из одного типологического класса в другой. Для этого представим графически результат проведенной кластеризации студентов как функции нечеткого множества (рисунок) с учетом того, что класс «Плохо» находится в пределах успеваемости студентов от 0 до 59 баллов, класс «Удовлетворительно» – в пределах от 60 до 74 баллов, класс «Хорошо» – в пределах от 75 до 89 баллов, класс «Отлично» – в пределах от 90 до 100 баллов.



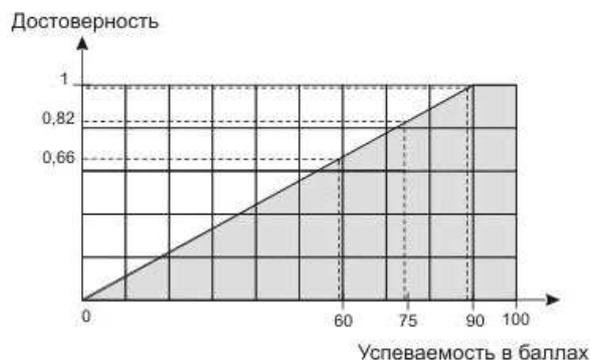
1



2



3



4

Графическое представление функции принадлежности нечеткого множества: 1 – класса «Плохо»; 2 – класса «Удовлетворительно»; 3 – класса «Хорошо»; 4 – класса «Отлично»

Так как граничные значения успеваемости по каждому классу известны, определим достоверность возможного перехода студентов из одной типологической группы в другую (табл. 1).

Таблица 1

Значения функции принадлежности перехода студентов в другую типологическую группу

Из класса	В класс			
	«Плохо»	«Удовлетворительно»	«Хорошо»	«Отлично»
«Плохо»	1	0,98	0,61	0,24
«Удовлетворительно»	0,98	1	0,96	0,39
«Хорошо»	0,79	0,99	1	0,9
«Отлично»	0,66	0,82	0,99	1

Влияние качественных факторов на процедуру кластеризации студентов рассмотрим на примере факторов цели и мотивации обучения, которые определим как предпочтения студентов по оценке их компетентности на экзамене (зачете). Запишем модели поведения студентов на языке предикатов [6]. Обозначим: $МП1, МП2, \dots, МП5$ – соответствующие модели поведения студента; $\forall x$ – любой студент, изучающий дисциплину «Информатика»; P – предикат, отражающий

стремление студента; $V \uparrow$ – оператор возможности; $\lambda(\chi)$ – оценка студента по дисциплине «Информатика»; K_y – класс «Удовлетворительно»; K_x – класс «Хорошо»; K_o – класс «Отлично»:

$$\begin{aligned}
 \text{МП1. } & \forall \chi [P, V \uparrow (\lambda(\chi)) \rightarrow K_y = 60 \leq \lambda(\chi) < 75]; \\
 \text{МП2. } & \forall \chi [P, V \uparrow (\lambda(\chi)) \rightarrow K_x \subseteq K_y = 60 \leq \lambda(\chi) < 90]; \\
 \text{МП3. } & \forall \chi [P, V \uparrow (\lambda(\chi)) \rightarrow K_x = 75 \leq \lambda(\chi) < 90]; \\
 \text{МП4. } & \forall \chi [P, V \uparrow (\lambda(\chi)) \rightarrow K_o \subseteq K_x = 75 \leq \lambda(\chi) < 100]; \\
 \text{МП5. } & \forall \chi [P, V \uparrow (\lambda(\chi)) \rightarrow K_o = 90 \leq \lambda(\chi) < 100].
 \end{aligned} \tag{2}$$

Будем различать следующие модели поведения студента:

МП1 – лишь бы сдать (не получить «плохо»);

МП2 – сдать желательно на «хорошо» или «отлично» (не получить «плохо», может быть «удовлетворительно»);

МП3 – сдать только на «хорошо» или «отлично» (не получить «плохо» или «удовлетворительно»);

МП4 – сдать желательно на «отлично» (не получить «плохо» или «удовлетворительно», может быть «хорошо»);

МП5 – сдать только на «отлично» (не получить «плохо», «удовлетворительно» или «хорошо»).

Каждую модель поведения студента опишем нечетким множеством $\tilde{M}_s = \{\mu_{\tilde{K}}(y_j)\}$, $s = \overline{1,5}$, $j = \overline{1,4}$, где s – множество моделей поведения студентов; j – множество типологических классов, в которые возможен переход. Получим: $\tilde{M}_1 = \{0; 1; 1; 1\}$; $\tilde{M}_2 = \{0; 0,5; 1; 1\}$; $\tilde{M}_3 = \{0; 0; 1; 1\}$; $\tilde{M}_4 = \{0; 0; 0,5; 1\}$; $\tilde{M}_5 = \{0; 0; 0; 1\}$.

Представим нечеткую функцию (1) в виде матрицы инциденций:

$$\tilde{K} = \begin{vmatrix} 1 & 0,98 & 0,61 & 0,24 \\ 0,98 & 1 & 0,96 & 0,39 \\ 0,79 & 0,99 & 1 & 0,9 \\ 0,66 & 0,82 & 0,99 & 1 \end{vmatrix}. \tag{3}$$

Тогда с учетом моделей поведения студента из матрицы (3) получим множество матриц нечетких отношений предпочтения студента $\tilde{K}_s = \tilde{K} \cap \tilde{P}_s$ с функцией принадлежности $\mu_{\tilde{K}_{s_{nl}}}(x_i, y_j) = \min[\mu_{\tilde{P}_s}(y_j); \mu_{\tilde{K}}(x_i, y_j)]$, где s – множество моделей поведения студентов; i – множество типологических классов, по которым проведено распределение студентов; j – множество типологических классов, в которые возможен переход.

Получим:

$$\begin{aligned}
 \tilde{K}_1 &= \begin{vmatrix} 0 & 0,98 & 0,61 & 0,24 \\ 0 & 1 & 0,96 & 0,39 \\ 0 & 0,99 & 1 & 0,9 \\ 0 & 0,82 & 0,99 & 1 \end{vmatrix}; \\
 \tilde{K}_2 &= \begin{vmatrix} 0 & 0,5 & 0,61 & 0,24 \\ 0 & 0,5 & 0,96 & 0,39 \\ 0 & 0,5 & 1 & 0,9 \\ 0 & 0,5 & 0,99 & 1 \end{vmatrix}; \\
 \tilde{K}_3 &= \begin{vmatrix} 0 & 0 & 0,61 & 0,24 \\ 0 & 0 & 0,96 & 0,39 \\ 0 & 0 & 1 & 0,9 \\ 0 & 0 & 0,99 & 1 \end{vmatrix}; \\
 \tilde{K}_4 &= \begin{vmatrix} 0 & 0 & 0,5 & 0,24 \\ 0 & 0 & 0,5 & 0,39 \\ 0 & 0 & 0,5 & 0,9 \\ 0 & 0 & 0,5 & 1 \end{vmatrix}; \\
 \tilde{K}_5 &= \begin{vmatrix} 0 & 0 & 0 & 0,24 \\ 0 & 0 & 0 & 0,39 \\ 0 & 0 & 0 & 0,9 \\ 0 & 0 & 0 & 1 \end{vmatrix}.
 \end{aligned} \tag{4}$$

Определим отношения строгого предпочтения \tilde{K}_s^C для каждой модели поведения студента с функцией принадлежности, вычисленной по формуле (5):

$$\mu_{\tilde{K}_s^C}(x, y) = \max[\mu_{\tilde{K}_s}(x, y) - \mu_{\tilde{K}_s}(y, x); 0]. \tag{5}$$

Получим:

$$\tilde{K}_1^C = \begin{pmatrix} 0 & 0,98 & 0,61 & 0,24 \\ 0 & 0 & 0 & 0 \\ 0 & 0,03 & 0 & 0 \\ 0 & 0,43 & 0,09 & 0 \end{pmatrix}; \dots; \tilde{K}_5^C = \begin{pmatrix} 0 & 0 & 0 & 0,24 \\ 0 & 0 & 0 & 0,39 \\ 0 & 0 & 0 & 0,9 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad (6)$$

Для каждой модели поведения студента вычислим множество недоминирующих альтернатив \tilde{K}_s^{HD} с функцией принадлежности (7):

$$\mu_{\tilde{K}_s^{HD}}(y) = 1 - \sup_{x \in X} \mu_{\tilde{K}_s^C}(x, y). \quad (7)$$

Множество недоминирующих альтернатив для каждой модели поведения студента представим в виде табл. 2:

Таблица 2

Соответствие успеваемости и предпочтений студентов

Y	Класс «Плохо»	Класс «Удовлетворительно»	Класс «Хорошо»	Класс «Отлично»
$\mu_{\tilde{K}_1^{HD}}(y)$	1	0,02	0,39	0,76
$\mu_{\tilde{K}_2^{HD}}(y)$	1	0,5	0,39	0,76
$\mu_{\tilde{K}_3^{HD}}(y)$	1	1	0,04	0,61
$\mu_{\tilde{K}_4^{HD}}(y)$	1	1	0,5	0,6
$\mu_{\tilde{K}_5^{HD}}(y)$	1	1	1	0,1

Табл. 2 показывает достоверность несоответствия студента типологической группе с определенной моделью поведения. Выделив множество четко недоминирующих альтернатив, для которых $\mu_{\tilde{K}_s^{HD}}(y) = 1$, получаем подгруппы студентов, у которых предпочтения по успеваемости не соответствуют возможностям.

5. Выводы

Таким образом, разработаны продукционные модели предпочтения студентов по оценке их компетенций; формализованы предпосылки перераспределения студентов по типологическим группам с учетом предпочтений студентов.

Список литературы

1. Добрынина, Н. Ф. Математические модели распространения знаний и управления процессом обучения [Текст] / Н. Ф. Добрынина // Фундаментальные исследования. – 2009. – № 7. – С. 7–9.

2. Зеневич, А. М. Математическое моделирование процесса обучения [Текст] / А. М. Зеневич, С. Я. Жукович // Подготовка научных кадров высшей квалификации в целях обеспечения инновационного развития экономики: материалы конференции / под ред. И. В. Войтова и др. – Минск: ГУ БелИСА, 2006. – 146 с.

3. Белова, Л. А. Логико-математические основы управления учебными процессами высших учебных заведений [Текст]: моногр. / Л. А. Белова, К. А. Метешкин, О. В. Уваров. – Харьков, Восточно-региональный центр гуманитарно-образовательных инициатив, 2001. – 272 с.

4. Ясинский, И. Ф. Опыт прогнозирования успеваемости студентов с помощью нейросетевой технологии [Текст] / И. Ф. Ясинский, М. Б. Семенова // Вестник Ивановского государственного энергетического университета. – Иваново: ГОУ ВПО Ивановск. гос. энергет. ун-т им. В. И. Ленина, 2007. – Вып. 4. – С. 29 – 31.

5. Гершунский, Б. С. Образовательно-педагогическая прогностика: теория, методология, практика [Текст]: учеб. пособие / Б. С. Гершунский. – М.: Флинта: Наука, 2003. – 756 с.

6. Метешкин, К. А. Кибернетическая педагогика: теоретические основы управления образованием на базе интегрированного интеллекта [Текст]: моногр. / К. А. Метешкин. – Х.: Междунар. Слав. ун-т, 2004. – 400 с.

Рецензент: д.т.н., проф., зав. каф. А. В. Бажинов, ХНАДУ, Харьков

Поступила в редакцию 06.09.2012

Нечіткє подання результатів кластеризації студентів

Запропоновано метод формалізації результатів розподілу студентів за типологічними групами. Процедура класифікації студентів наведено у вигляді нечіткої функції. Графічним способом визначено значення функції наявності переходу студентів в іншу типологічну групу. Розроблено продукційні моделі переваги студентів за оцінкою їхніх компетенцій. Одержано множину недомінуючих альтернатив для кожної моделі поведінки студента. Формалізовано передумови перерозподілу студентів за типологічними групами з урахуванням переваг студентів.

Ключові слова: нечітка множина, кластеризація, типологічні групи, нечітка функція, функція наявності, продукційна модель, язык предикатів.

Fuzzy Results Representation Concerning Students Clustering

The method of student's distribution result formalization according to typological groups is offered. The procedure of student's classification is represented in the form of the function. The values of the adjective function of typological classes fuzzy set are determined by means of the graphical method. There have been developed production models of student's preferences according to the estimation of their competence. There were also obtained a variety of no dominant alternatives for each student's behavior model/ The prerequisites of student's redistribution according to typological groups taking into account student's preferences are formalized.

Keywords: fuzzy set, clustering, typological groups, fuzzy function, adjective function, production model, predicate language.