

А.С. КУЦЕНКО, д-р. техн. наук, НТУ „ХПИ” (г. Харьков),
Э.Е. РУБИН, директор ООО „Телесенс КСЦЛ Украина” (г. Харьков)

ВЕРОЯТНОСТНО-СТАТИСТИЧЕСКИЙ ПОДХОД К ОЦЕНКЕ СОДЕРЖАНИЯ И КАЧЕСТВА ВЫСШЕГО ОБРАЗОВАНИЯ

Запропоновано новий підхід до оцінювання вмісту та якості освітньої діяльності, заснований на статистичній інформації про результати навчання студентів за даним фахом. На основі ентропійного методу побудована матриця зв'язків між навчальними дисциплінами. Запропоновано методику оцінки напруженості навчальних планів та класифікації студентів.

If offered the new approach to the substance and quality estimating of the educate activities, founded on the statistical information about students training results in given speciality.

Введение. Вероятностно-статистические методы являются одним из основных инструментов для описания слабоформализованных сложных систем, к разряду которых относятся социальные, в том числе, системы образования [1]. Привлечение математических методов для планирования и управления образовательной деятельностью в основном происходило двумя путями. Одно направление тесно связано с разработкой математических моделей динамики образовательного процесса, позволяющих прогнозировать численности выпускников вузов по различным специальностям, а также их соответствие потребностям общества [2, 3]. Второе направление связано с комплексной информатизацией и автоматизацией образовательной деятельности [4]. Основной акцент в этом направлении сделан на создание методов и алгоритмов автоматизированного контроля знаний и обучения [5], а также на разработку систем дистанционного образования [6]. Вопросы планирования учебного процесса на основе экспертных оценок детально рассмотрены в [7]. Несмотря на широкое привлечение математических методов к задачам планирования образования существенным пробелом следует считать отсутствие количественных оценок содержания образовательных программ и аргументированной методики классификации выпускников высших учебных заведений. Существующая методика классификации выпускников на классы: с отличием, без удовлетворительных оценок или по принципу «среднего балла», не выдерживает никакой критики.

Целью настоящей работы является попытка дать количественную оценку взаимосвязи между учебными дисциплинами и на ее основании предложить процедуру классификации выпускников высших учебных заведений.

Постановка задачи исследования. Для получения количественных оценок взаимосвязи учебных дисциплин и организации в последующем процедуры классификации выпускников предполагается воспользоваться

статистическими данными успеваемости студентов по некоторой специальности. Множество экзаменационных оценок, полученных каждым студентом, можно рассматривать как некоторое сообщение, которое, с одной стороны, несет в себе информацию о взаимосвязях между учебными дисциплинами и об уровне подготовки студентов с другой. Статистические данные успеваемости позволяют построить матрицу взаимосвязей между дисциплинами, на основании которой можно провести группирование дисциплин на блоки по степени связи, выбрать наиболее значимые в каждом из блоков и классифицировать студентов по наиболее значимым дисциплинам.

Информационная оценка содержания высшего образования.

Рассмотрим задачу количественной оценки содержания образования. Интуитивно ясно, что между изучаемыми N дисциплинами существует определенная связь, заключающаяся в том, что знания и умения, полученные в результате изучения дисциплины i , используются при изучении дисциплины j . Таким образом, полученное количество знаний и умений по i -й дисциплине непосредственно влияет на результат обучения по дисциплине j . Введем количественную меру – коэффициент степени связи r_{ij} , отражающий эту взаимосвязь. Коэффициент связи r_{ij} естественно наделить следующими свойствами:

1. $r_{ii} = 1, i = \overline{1, N}$;
2. $0 \leq r_{ij} \leq 1, i, j = \overline{1, N}$;
3. $r_{ij} = r_{ji}$.

Множество коэффициентов связи $r_{ij}, i, j = \overline{1, N}$ образует симметрическую матрицу сходства R :

$$R = \begin{pmatrix} 1 & r_{12} & \dots & r_{1N} \\ r_{21} & 1 & \dots & r_{2N} \\ \dots & \dots & \dots & \dots \\ r_{N1} & r_{N2} & \dots & 1 \end{pmatrix}.$$

Введем понятие блочности множества дисциплин. Предварительно отождествим подмножество C множества дисциплин A с подмножеством номеров $I_C \subset I$, где $I = \{1, 2, 3, \dots, N\}$, а I_C – некоторый набор неравных между собой целых чисел, не превышающих N . Блоком будем называть некоторое подмножество B множества дисциплин A , обладающее свойством

$$r_{ij} \gg r_{ik},$$

где $i, j \in I_B, k \notin I_B, I_B$ – множество индексов, соответствующих блоку B .

Таким образом, множество дисциплин A может быть разбито на непересекающиеся блоки

$$A = \bigcup_{k=1}^K B_k.$$

Дисциплины каждого из блоков характеризуются высокой степенью связи $r_{ij} \geq \bar{r}$ между собой, степени связи между дисциплинами различных блоков удовлетворяет неравенству $r_{ik} \leq \underline{r}$. Пара чисел \bar{r}, \underline{r} ($\bar{r} > \underline{r}$) определяет количество блоков K в учебном плане. Вместо матрицы R введем в рассмотрение вектор \bar{R} , имеющий размерность $l = \frac{N(N-1)}{2}$ и представляющий собой множество элементов матрицы R , расположенных над главной диагональю:

$$\bar{R} = (r_1, r_2, \dots, r_l).$$

Для получения количественных оценок степени связи учебных дисциплин, изучаемых по данной специальности, введем функцию распределения $P(r, \Delta)$, представляющую собой количество пар дисциплин, коэффициенты связи между которыми находятся в интервале $(r, r + \Delta)$, где Δ – выбранный шаг дискретизации по r . Таким образом, $P(r, \Delta)$ представляет собой целочисленную функцию дискретного аргумента.

Шаг дискретизации Δ определяется на основании выбранного количества узлов S :

$$\Delta = \frac{1}{S-1},$$

и вместо функции $P(r, \Delta)$ можно рассматривать пронумерованную последовательность целых чисел

$$\bar{P} = \{P_1, P_2, \dots, P_S\},$$

где $P_k = P(k\Delta, \Delta)$ и $\sum_{k=1}^S P_k = l$ – условие нормировки.

Вместо вектора частот \bar{P} будем также рассматривать вектор относительных частот

$$P = (p_1, p_2, \dots, p_S),$$

где $p_k = \frac{P_k}{l}$ и $\sum_{k=1}^S p_k = 1$.

Поскольку распределение P имеет структуру дискретного распределения вероятностей, то для получения числовых характеристик взаимосвязи учебных дисциплин могут быть привлечены все разнообразные подходы теории вероятностей и математической статистики. Рассмотрим некоторые числовые характеристики, принятые в теории вероятностей, и дадим их интерпретацию в терминах данной предметной области.

1. Математическое ожидание:

$$m_r = \sum_{k=1}^S p_k \Delta k$$

представляет собой среднее значение коэффициента связи между дисциплинами. В предельных случаях $m_r = 1$, если все дисциплины «дублируют» друг друга, и $m_r = 0$, если дисциплины абсолютно не связаны между собой.

2. Дисперсия:

$$D_r = \sum_{k=1}^S p_k (\Delta k - m_r)^2$$

характеризует степень «разброса» величины коэффициентов связи между дисциплинами.

3. Асимметрия:

$$S_r = \sum_{k=1}^S p_k (\Delta k - m_r)^3$$

характеризует степень блочности множества дисциплин. С увеличением числа блоков S_r возрастает и убывает с его уменьшением.

Таким образом, при наличии матрицы коэффициентов связи между дисциплинами, можно получить количественные оценки (m_r, D_r, S_r) в той или иной мере характеризующие учебный план и учебные программы по различным специальностям.

В качестве параметров учебного процесса будем рассматривать экзаменационные оценки знаний X_j , полученных студентами по j -й дисциплине. Таким образом, результат обучения одного студента определяется вектором его оценок $X = \{X_1, X_2, \dots, X_N\}$. Если по данной специальности обучается m студентов, то можно рассматривать вектор X

как случайный N -мерный вектор, имеющий t реализаций. При этом шкала оценок предполагается одинаковой по всем дисциплинам, а количество градаций шкалы M может выбираться произвольным образом. В такой интерпретации учебного процесса степень взаимосвязи между дисциплинами может быть оценена статистическими методами. В качестве меры степени связи могут быть взяты любые широко применяемые параметры, рассмотренные в литературе по многомерным статистическим методам. Наибольший интерес представляет информационный критерий связи, который удовлетворяет всем требованиям, предъявляемым к матрице степеней связи R . Кроме того, по определению, такой критерий наиболее естественно отражает количество знаний и умений, содержащихся в одной дисциплине и необходимых для изучения другой.

Для определения информационной степени связи r_{ij} между дисциплинами i и j составим матрицу относительных частот появления всех допустимых пар оценок по этим дисциплинам

$j \backslash i$	3	4	5	(2)
3	P_{33}	P_{34}	P_{35}	
4	P_{43}	P_{44}	P_{45}	
5	P_{53}	P_{54}	P_{55}	

Тогда количество информации, содержащейся в дисциплине i по отношению к дисциплине j можно определить в виде $J_{ij} = H_i + H_j - H_{ij}$, где H_i , H_j – энтропии распределения оценок по i и j дисциплинам соответственно, а H_{ij} – энтропия двумерного распределения оценок по дисциплинам i и j .

Указанные энтропии вычисляются по известным соотношениям теории информации на основании матрицы частот (2):

$$H_i = - \sum_j \sum_k P_{jk} \cdot \ln \sum_k P_{jk} ;$$

$$H_j = - \sum_i \sum_k P_{ki} \cdot \ln \sum_k P_{ki} ;$$

$$H_{ij} = - \sum_{k,s} P_{ks} \cdot \ln P_{ks} .$$

В качестве меры степени связи дисциплин r_{ij} примем нормированное

значение информации J_{ij} :

$$r_{ij} = J_{ij} / \sqrt{H_i H_j} ,$$

удовлетворяющее условиям (1).

Вернемся теперь к проблеме классификации выпускников по результатам множества экзаменов. Как было ранее отмечено, различные арифметические манипуляции с оценками ничем не обоснованы, поскольку экзаменационные оценки сами по себе носят качественный характер. Кроме того, классификация выпускников на основе всего массива оценок не дает возможности выделить те или иные перспективные направления дальнейшей деятельности выпускника. В связи с этим предлагается следующий подход к проблеме классификации выпускников высших учебных заведений:

1. На основании статистических данных по результатам экзаменов проводится разбиение учебных дисциплин на блоки.

2. В каждом из блоков на основе информационного критерия связи выбирается наиболее информативная учебная дисциплина i^* :

$$i^* = \arg \max_{i \in I_B} \sum_{j \in I_B} r_{ij} .$$

3. Классификация выпускников осуществляется по наиболее информативным дисциплинам в каждом из блоков.

Заключение. В настоящей работе изложены общие принципиальные подходы к информационной оценке содержания высшего образования и качества подготовки выпускников. Практическая реализация изложенного подхода требует обработки большого числа экспериментальных данных и привлечения экспертов для оценки рациональных величин количественных характеристик учебного процесса, предложенных в работе.

Список литературы: 1. Глас Дж., Стэнли Дж. Статистические методы в педагогике и психологии. – М.: Прогресс, 1976. – 495 с. 2. Бермант М.А., Семенов Л.К., Сулицкий В.Н. Математические модели и планирование образования. – М.: Наука, 1972. – 111 с. 3. Кононенко І.В., Голоскокова Є.О. Прогнозування попиту на освітні послуги та на випускників вищих навчальних закладів // Вестник Национального технического университета „ХПИ”. – Харьков: НТУ „ХПИ”. – 2003. – № 18. – С. 131 – 134. 4. Шаронова Н.В. Проблеми моделювання інформаційних процесів управління освітою як соціально-економічною системою, що розвивається // Теорія і практика управління соціальними системами. – Харків: НТУ „ХПИ”. – 2004. – № 2. – С. 58 – 64. 5. Оксамитна Л.П. Вимір розподільної здатності в адаптивних навчальних системах контролю знань та навчання // Вісник ХДТУ. – Херсон: ХДТУ, 2003. – № 2 (18). – С. 460 – 462. 6. Пономаренко Л.А., Степанова Я.М. Алгоритмічна структура процесу дистанційного навчання // Проблеми підвищення ефективності інфраструктури. – К.: НАУ. – 2002. – Вип. 8. – С. 151 – 156. 7. Петров Э.Г., Радванская Л.Н., Шаронова Н.В. Самовершенствование преподавателя. Учебное пособие. / – Херсон: ОЛДИ-плюс, 2002. – 144 с.

Поступила в редакцию 13.09.2004