

Обчислювальна технологія виявлення пріоритетних дисциплін навчання студентів

КУЗЬМА К.Т.

Дніпропетровський національний університет ім. Олеся Гончара

Розглянуто один з підходів обробки експертних оцінок з використанням множинного порівняльного аналізу.

Рассмотрен один из подходов обработки экспертных оценок с использованием множественного сравнительного анализа.

One of approaches of processing of expert estimations on the basis of plural comparative analysis is considered.

Постановка проблеми. Моніторинг і дослідження процесу навчання є однією з задач управління якістю освітнього процесу, а, отже, одним з ключових елементів Болонського процесу [1].

В цьому контексті актуальною є задача оцінки практичної цінності знань і навиків, які дають вищі навчальні заклади своїм студентам. В зв'язку з великою кількістю дисциплін, які вивчають студенти протягом всього періоду навчання, процес дослідження навчання ускладнюється. Тому постає завдання вибору дисциплін для будь-якого напрямку підготовки, які відповідають освітнім вимогам ринку праці і є пріоритетними для спеціалістів даної кваліфікації.

Аналіз останніх досліджень і публікацій, в яких започатковано розв'язання даної проблеми та постановка задачі. Враховуючи, що оцінку важливості дисциплін навчання студентів відносно їх значення в практичній діяльності майбутнього фахівця будуть здійснювати експерти-роботодавці, задача виявлення пріоритетних дисциплін навчання зводиться до задачі обробки експертних оцінок.

Розв'язок задач управління, обробки експертних оцінок, обробки підсумків експерименту та інші на початковій стадії відбувається шляхом вивчення зв'язків між явищами та процесами. Методи та алгоритми оцінки і аналізу зв'язків між явищами та процесами були започатковані в роботах С.А. Айвазяна, Н. Дрейпера, Е. Ферсера [2-4]. Подальшим розвитком методів множинного аналізу зв'язків між змінними по результатам спостережень займалися О.П. Приставка, О.Г. Байбуз, П.О. Приставка Т.Г. Ємель'яненко, в роботах яких запропоновані обчислювальні схеми кореляційного аналізу зв'язків [5-6] в системах моніторингу. Різні методи аналізу зв'язків дають відмінні результати.

В зв'язку з цим постає задача аналізу існуючих алгоритмів аналізу зв'язків між змінними по результатам спостережень, вибір та застосування обчислювальної технології для виявлення пріоритетних дисциплін навчання студентів.

Виклад основного матеріалу дослідження. Для виявлення та оцінки стохастичного зв'язку між явищами та процесами по результатам спостережень використовують методи кореляційного і регресійного аналізу. Кореляційний аналіз робить оцінку сили стохастичного зв'язку, регресійний – форми зв'язку.

Методи кореляційного і регресійного, які використовують в наукових працях вітчизняних та зарубіжних авторів, наведені на рис.1.

Вибір одного з наведених методів для рішення задачі залежить від об'єкта дослідження та його показників.

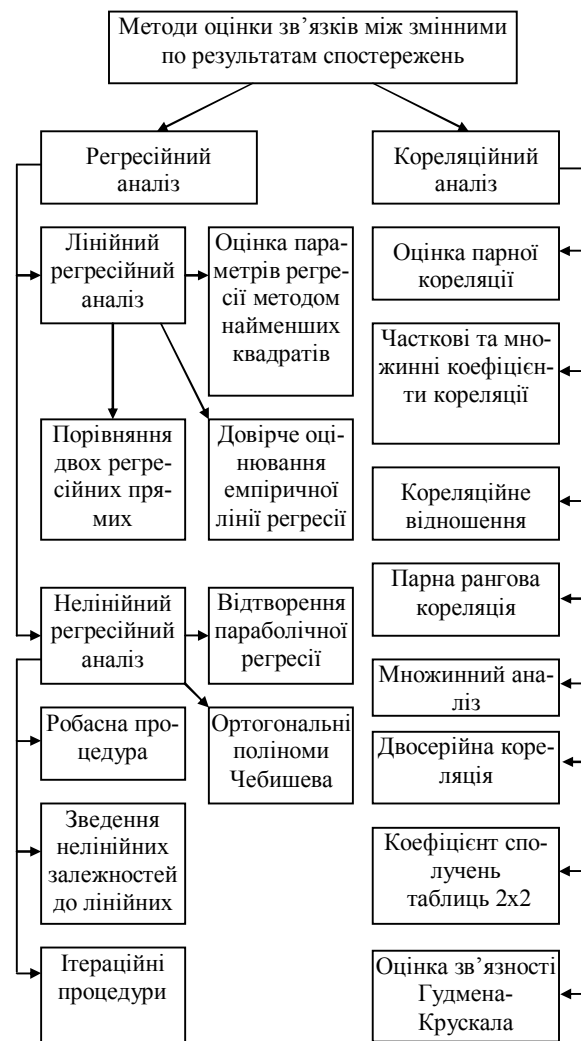


Рис. 1. Методи кореляційного та регресійного аналізу

Враховуючи що початковими даними для рішення поставленої задачі є експертні оцінки, для виявлення пріоритетних дисциплін навчання студентів відповідного напрямку підготовки обрано обчислювальну технологію множинного порівняльного аналізу.

На основі аналізу методик та результатів проведення рейтингів ВУЗів [7], вирішено оцінку дисциплін виконувати по десятибальній шкалі, при цьому експертами будуть працевлаштувачі.

В рамках експерименту розглядалися дисципліни циклу професійної та практичної підготовки і циклу вибіркових навчальних дисциплін напрямку підготовки 6.050102 «Комп'ютерна інженерія» в Миколаївському політехнічному інституті. Експертами виступали директори і головні інженери провідних компаній м. Миколаєва, діяльність яких пов'язана з комп'ютерними та інформаційними технологіями (експерт (1, 2), а також керівний склад вищого навчального закладу (ректор - експерт (3), проректор з наукової роботи – експерт (4), проректор з навчальної роботи – експерт (5)).

Для отримання експертних оцінок використовувався комунікативний метод отримання знань – метод експертного опитування. Так як в рамках задачі передбачається подальша автоматизація процесу опитування, то в даному випадку підходить метод анкетування, який є стандартизованим та більш жорстким.

Сформований набір оцінок наведений в таблиці $m \times n$, де m - кількість спостережень (експертів), n - кількість показників (ознак) (табл. 1).

Таблиця 1. Початкові дані спостережень

Експерти Дисципліни	1.	2.	3.	4.	5.
1. Архітектура комп'ютерів	1	1	2	1	2
2. Комп'ютерні мережі	3	2	1	2	1
3. Програмування та алг. мови	2	1	1	1	1
4. Об'єктне програмування	1	1	2	1	2
5. Мережні інформаційні технології	2	3	1	3	1
6. Моделювання комп. систем	0,5	1	2	1	2
7. Дослідження операцій	0,5	1	1	1	1

Здійснено кореляційний аналіз експертних оцінок на основі обчислювальної схеми множинного порівняльного аналізу.

Обчислювальна схема обробки початкового масиву містить такі процедури [8]:

- оцінка адекватності вимірювань (компетентність експертів);
- узгодженість вимірювань;
- переважання висновків.

1. Оцінка адекватності або компетентності експертів. Найбільш часто компетентність експерта оцінюють по тому, наскільки узгоджені його оцінки з оцінками більшості.

Слідуючи роботі [6], оцінка компетентності експертів проводиться за рекурентною процедурою:

$$x_i^t = \sum_{j=1}^m x_{ij} k_j^{t-1}, \quad i = \overline{1, n}, \quad (1)$$

$$\lambda^t = \sum_{i=1}^n \sum_{j=1}^m x_{ij} x_i^t, \quad t = 1, 2, \dots, \quad (2)$$

$$k_j^t = \frac{1}{\lambda^t} \sum_{i=1}^n x_{ij} x_i^t, \quad \sum_{j=1}^m k_j^t = 1, \quad j = \overline{1, m}, \quad (3)$$

де k_j - оцінка адекватності експерта.

Обчислювальний процес починається з $t = 1, k_j = \frac{1}{m}$, далі збільшуємо $t: t = t + 1$, повторюємо обчислювання поки процес не зійдеться з деякою заданою точністю ε .

$$X^1 = \{1,40; 1,80; 1,20; 1,40; 2,0; 1,30; 0,90\}, \quad \lambda^1 = 75,5,$$

Підсумком другого та третього наближення є:

$$X^2 = \{1,369; 1,829; 1,204; 1,369; 2,039; 1,266; 0,893\},$$

$$\lambda^2 = 75,72, \quad K^2 = \{0,209; 0,210; 0,185; 0,210; 0,185\},$$

$$X^3 = \{1,370; 1,839; 1,209; 1,370; 2,050; 1,266; 0,895\},$$

$$\lambda^3 = 75,749, \quad K^3 = \{0,209; 0,210; 0,185; 0,210; 0,185\},$$

Аналізуючи результати другого та третього наближення, робимо висновок, що вектор K стабілізувався.

Далі знаходимо межове значення векторів групової оцінки X та адекватності K на підставі виразів $B = XX^T$; $C = X^T X$; $X = \|x_{ij}\|$:

$$B = \begin{pmatrix} 11 & 11 & 8 & 11 & 12 & 10,5 & 6,5 \\ 11 & 19 & 12 & 11 & 20 & 9,5 & 7,5 \\ 8 & 12 & 8 & 8 & 12 & 7 & 5 \\ 11 & 11 & 8 & 11 & 12 & 10,5 & 6,5 \\ 12 & 20 & 12 & 12 & 24 & 11 & 9 \\ 10,5 & 9,5 & 7 & 10,5 & 11 & 10,25 & 6,25 \\ 6,5 & 7,5 & 5 & 6,5 & 9 & 6,25 & 4,25 \end{pmatrix}$$

$$C = \begin{pmatrix} 19,5 & 17 & 12,5 & 17 & 12,5 \\ 17 & 18 & 13 & 18 & 13 \\ 12,5 & 13 & 16 & 13 & 16 \\ 17 & 18 & 13 & 18 & 13 \\ 12,5 & 13 & 16 & 13 & 16 \end{pmatrix}$$

Максимальні власні числа матриць B і C приймають в якості групової оцінки компетентності спостережень та обчислюються як корені рівнянь:

$$|B - \lambda E| = 0, \quad |C - \lambda E| = 0,$$

$$B = \begin{pmatrix} 11 - \lambda & 11 & 8 & 11 & 12 & 10,5 & 6,5 \\ 11 & 19 - \lambda & 12 & 11 & 20 & 9,5 & 7,5 \\ 8 & 12 & 8 - \lambda & 8 & 12 & 7 & 5 \\ 11 & 11 & 8 & 11 - \lambda & 12 & 10,5 & 6,5 \\ 12 & 20 & 12 & 12 & 24 - \lambda & 11 & 9 \\ 10,5 & 9,5 & 7 & 10,5 & 11 & 10,25 - \lambda & 6,25 \\ 6,5 & 7,5 & 5 & 6,5 & 9 & 6,25 & 4,25 - \lambda \end{pmatrix}$$

$$C = \begin{pmatrix} 19,5 - \lambda & 17 & 12,5 & 17 & 12,5 \\ 17 & 18 - \lambda & 13 & 18 & 13 \\ 12,5 & 13 & 16 - \lambda & 13 & 16 \\ 17 & 18 & 13 & 18 - \lambda & 13 \\ 12,5 & 13 & 16 & 13 & 16 - \lambda \end{pmatrix}$$

Максимальні власні числа матриці B обчислюються як корені рівняння:

$$87, \lambda^6 - \lambda^7 - 911,5 \lambda^5 + 1636,0 \lambda^4 = 0$$

Максимальні власні числа матриці C обчислюються як корені рівняння:

$$-\lambda^3 + 87,5\lambda^2 - 911,5\lambda + 1636 = 0$$

Для матриці В та С максимальний корінь рівняння $\lambda_0 = 75,752$.

З розв'язку рівнянь (1-3) отримаємо $x_1=1,369$, $x_2=1,829$, $x_3=1,204$, $x_4=1,369$, $x_5=2,039$, $x_6=1,267$, $x_7=0,893$, $k_1=0,207$, $k_2=0,209$, $k_3=0,186$, $k_4=0,208$, $k_5=0,186$, які приймають як групову оцінку та оцінку адекватності експертів.

Таким чином, після ранжування коефіцієнтів k_i ($k_2=0,209$, $k_4=0,209$; $k_1=0,207$; $k_3=0,186$; $k_5=0,186$), можна зробити висновок, що другий, четвертий та перший експерти найбільш адекватні.

2. Процедура оцінювання узгодження оцінок експертів базується на дисперсійному коефіцієнті конкординації. Початкову матрицю спостережень $\|x_{ij}\|$, $i = \overline{1, n}$, $j = \overline{1, m}$, перетворюють у матрицю рангів $\|R_{ij}\|$, на основі якої обчислюються характеристики:

$$r_i = \sum_{j=1}^m R_{ij}, \quad i = \overline{1, n}; \quad D = \frac{1}{n-1} \sum_{i=1}^n (r_i - \bar{r})^2, \quad \bar{r} = \frac{1}{n} \sum_{i=1}^n r_i;$$

$$W = \frac{D}{D_{max}}, \quad 0 < W \leq 1,$$

де W – дисперсійний коефіцієнт конкординації, Ранги початкових даних наведені в табл. 2.

Таблиця 2. Ранги початкових даних

Експерти Дисципліни	1	2	3	4	5
1. Архітектура комп'ютерів	1	1	1	1	1
2. Комп'ютерні мережі	2	2	2,5	2	2,5
3. Програмування та алг. мови	3	3,5	2,5	3,5	2,5
4. Об'єктне програмування	4	3,5	4	3,5	4
5. Мережні інформаційні технології	5	5	5	5	5
6. Моделювання комп. Систем	6,5	6,5	6	6,5	6
7. Дослідження операцій	6,5	6,5	7	6,5	7

Для знаходження D_{max} вважаємо, що має місце відсутність пов'язаних рангів. Тоді отримаємо:

$$\sum_{i=1}^n R_{ij} = \frac{1}{2} n(n+1), \quad j = \overline{1, m}, \quad \bar{r} = \frac{1}{2} (n+1)m,$$

$$S = \sum_{i=1}^n \left(\sum_{j=1}^m r_{ij} - \bar{r} \right)^2, \quad D = \frac{1}{n-1} \sum_{i=1}^n (r_i - \bar{r})^2 = \frac{1}{n-1} S,$$

$$D_{max} = \frac{m^2(n^3 - n)}{12(n-1)},$$

На основі виразів обчислюються коефіцієнт конкординації:

$$W = \frac{12S}{m^2(n^3 - n)}, \quad (4)$$

Якщо мають місце зв'язані ранги, то

$$W = \frac{12S}{m^2(n^3 - n) - m \sum_{j=1}^m T_j}, \quad (5)$$

де $T_j = \sum_{k=1}^{H_j} (h_k^3 - h_k)$ - показник зв'язаних рангів у j -му ранжуванні; H_j - кількість груп рівнозначних рангів у j -му ранжуванні; h_k - кількість рівнозначних рангів у k -й групі зв'язаних рангів при ранжуванні за j .

Значущість W , а отже, перевірка гіпотези $H_0: W = 0$ при $n > 7$ відбувається на підставі статистичної характеристики для незв'язаних рангів

$$\chi^2 = Wm(n-1), \quad (6)$$

або для зв'язаних рангів

$$\chi^2 = \frac{12S}{mn(n+1) - \frac{1}{n-1} \sum_{j=1}^m T_j}, \quad (7)$$

які мають $\chi^2 - \delta i \zeta i i \bar{a} \bar{e}$ $\zeta v = n - 1$ кількістю степенів вільності.

Для нашої задачі маємо:

$$\bar{r} = 20, \quad S = \sum_{i=1}^7 \left(\sum_{j=1}^5 r_{ij} - 20 \right)^2 = 671,5,$$

$$T_1 = 2^3 - 2 = 6, \quad T_2 = 2^3 - 2 + 2^3 - 2 = 12,$$

$$T_3 = 2^3 - 2 = 6, \quad T_4 = 2^3 - 2 + 2^3 - 2 = 12, \quad T_5 = 2^3 - 2 = 6,$$

$$W = \frac{12 \cdot 671,5}{5^2(7^3 - 7) - 5 \cdot 42} = 0,98,$$

$$\chi^2 = \frac{12 \cdot 671,5}{5 \cdot 7 \cdot 8 - \frac{1}{6} \cdot 42} = 29,52 \quad \chi_{a,v}^2 = 12,6$$

Враховуючи, що $\chi_{a,v}^2 < \chi^2$ ($12,6 < 29,52$), гіпотеза про узгодження оцінок експертів є правильною.

3. Вибір переважного висновку або визначення взаємодії ранжувань проводиться на підставі коефіцієнтів рангової кореляції, наприклад Спірмена.

$$\text{Для випадку зв'язаних рангів: } \rho = \frac{K_{12}}{\sqrt{D_1 D_2}},$$

де:

$$K_{12} = \frac{1}{n-1} \sum_{j=1}^n (r_{1j} - \bar{r}_1)(r_{2j} - \bar{r}_2);$$

$$D_1 = \frac{1}{n-1} \sum_{j=1}^n (r_{1j} - \bar{r}_1)^2; \quad D_2 = \frac{1}{n-1} \sum_{j=1}^n (r_{2j} - \bar{r}_2)^2;$$

$$\bar{r}_1 = \frac{1}{n} \sum_{j=1}^n r_{1j}; \quad \bar{r}_2 = \frac{1}{n} \sum_{j=1}^n r_{2j}$$

Якщо відсутні зв'язані ранги, то:

$$\bar{r} = \bar{r}_1 = \bar{r}_2 = \frac{n+1}{2}, \quad D = D_1 = D_2 = \frac{n(n+1)}{12},$$

$$\rho = \frac{12}{n^3 - n} (r_{1j} - \bar{r})(r_{2j} - \bar{r}), \quad (8)$$

$$\hat{\rho} = 1 - \frac{6}{n^3 - 1} \sum_{j=1}^n (r_{1j} - r_{2j})^2 \quad (9)$$

$$\rho = 1 - \frac{6}{7^3 - 1} \cdot 4 = 0.9298$$

Для випадку зв'язаних рангів коефіцієнт кореляції Спірмена одержують за формулою:

$$\tilde{\rho} = \frac{\rho + T_1 + T_2}{\sqrt{(1 - T_1)(1 - T_2)}}, \quad (10)$$

де ρ - значення, обчислене для зв'язаних рангів;

$$T_1 = \frac{3}{n^3 - n K_1} \sum K_1(K_1 - 1); \quad T_2 = \frac{3}{n^3 - n K_2} \sum K_2(K_2 - 1);$$

При $\rho = 1$ має місце однакове ранжування, при $\rho = -1$ - протилежне, при $\rho = 0$ - незалежне ранжування.

Статистичне значення ρ перевіряється на значущість згідно із статистичною характеристикою

$$\varepsilon = \frac{1}{\sqrt{n-1}} \hat{O}^{-1} \left(\frac{1-\alpha}{2} \right), \quad (11)$$

де $\hat{O}^{-1}(\bullet)$ - квантиль нормального розподілу.

Якщо $|\rho| < \varepsilon$, то оцінка коефіцієнта кореляції ρ є значущою.

Значення квантилю нормального розподілу Стюдента для нашого випадку $\hat{O}^{-1}(\bullet) = 2,4469$ (кількість степенів вільності -6, $\alpha=0,05$)

$$\varepsilon = \frac{1}{\sqrt{n-1}} \hat{O}^{-1} \left(\frac{1-\alpha}{2} \right) = \frac{1}{\sqrt{6-1}} 2,4469 = 1,1$$

Вибір переважного висновку здійснюється на основі порівняння сум рангів або сум часток (ваг) за кожною ознакою. Для нашого випадку оберемо характеристику суми ваг. Враховуючи що $|\rho| < \varepsilon$, можна зробити висновок про значущість оцінки коефіцієнта кореляції ρ .

Таким чином обчислювальна схема множинного порівняльного аналізу експертних оцінок може бути описана наступним алгоритмом [6]:

1. Задаються значення змінних $t = 1, k_j = \frac{1}{m}$.
2. Обчислюються значення x_j^t по формулі (1).
3. Обчислюються значення λ^t по формулі (2).
4. Збільшуємо $t: t = t + 1$. Обчислюємо значення k_j^t по формулі (3).
5. Повторюємо пункти 2-4, поки процес не зійдеться з деякою заданою точністю ε .
6. Виконуємо оцінку узгодженості експертів. Якщо мають місце зв'язані ранги (ранги, що співпадають), то коефіцієнт конкордації обчислюють по формулі (5), в іншому випадку по формулі (4).
7. Обчислюють статистичну характеристику χ^2 для зв'язаних рангів по формулі (7), для незв'язаних по формулі (6).

8. Якщо $\chi_{a.v}^2 \leq \chi^2$, то приймається гіпотеза про узгодженість точок зору експертів, інакше робимо висновок, що оцінки експертів не узгоджені.
9. Здійснюють вибір переважного висновку. Обчислюють коефіцієнт кореляції Спірмена по формулі (8) або (9) для незв'язаних рангів та по формулі (10) для зв'язаних рангів в першому та другому ранжуванні.
10. Обчислюють статистичну характеристику ε по формулі (11).
11. Якщо $|\rho| < \varepsilon$, то оцінка коефіцієнта кореляції ρ є значущою.

Висновки

В результаті проведеного аналізу встановлено, що пріоритетними є дисципліни «Мережні інформаційні технології» та «Комп'ютерні мережі». Саме ці дисципліни навчання є пріоритетними для напрямку підготовки 6.050102 «Комп'ютерна інженерія». Тому, для державної атестації студентів, моделювання процесу отримання знань або будь-якої іншої перевірки рівня знань студентів з даного напрямку підготовки, доцільно брати саме ці дисципліни.

ЛІТЕРАТУРА

1. Основы развития высшего образования Украины в контексте Болонского процесса (документы и материалы 2003-2004 гг.) / Под ред. В.Г. Кремня, авт. кол.: Степко М.Ф., Боллобаш Я.Я., Шинкарук В.Д., Грубинко В.В., Бабий И.И. – Киев – Тернополь: Изд-во ТДПУ, 2004. – 147 с.
2. Айвазян С.А. Прикладная статистика: исследование зависимости /С.А. Айвазян, И.С. Енюков, Л.Д. Мешалкин. – М.: Финансы и статистика, 1984. – 475 с.
3. Дрейпер Н. Прикладной регрессионный анализ / Н. Дрейпер, Г. Смит. – М.: Финансы и статистика, 1986. – 322 с.
4. Фёрстер Е. Методы корреляционного и регрессионного анализа / Е.Фёрстер, Б. Рёнц. – М.: Финансы и статистика, 1987. – 222 с.
5. Байбуз О.Г. Сплайны в надежности О.Г.Байбуз, А.Ф. Приставка, - Д.: Изд-во Днепропетр. ун-та, 2003. - 246 с.
6. Принятие решений в системах мониторинга / Т.Г. Емельяненко, А.В. Зборовский, А.Ф. Приставка, Б.Е. Собко. – Днепропетровск: РИК НГУ, 2005. – 224 с.
7. Рейтинговая оценка вузов: международный и украинский опыт. – Режим доступа: < http://www.yourcompass.org/img/Compass%202009_InternationalExperience%20Ru.pdf>.
8. Статистична обробка даних / В.П. Бабак., А.Я. Білецький, О.П. Приставка, П.О. Приставка. – К.: МІВВЦ, 2001. – 388 с.