

Дана робота присвячена прогнозуванню вступу учнів до університету на основі середнього балу за останні 3 роки навчання. Матеріали статті можуть бути використані в школах для аналізу та прогнозування

Ключові слова: дискримінантний аналіз, учень, класифікація

Данная работа посвящена прогнозированию поступления учащихся в университет на основе среднего балла за последние 3 года обучения. Материалы статьи могут быть использованы в школах для анализа и прогнозирования

Ключевые слова: дискриминантный анализ, студент, классификация

This work is devoted to prediction of entered students in the university on the basis of the average score for the last 3 years. Materials of articles can be used in schools for the analysis and prediction

Keywords: discriminant analysis, student classification

ПОБУДОВА МОДЕЛІ ПРОГНОЗУ ВСТУПУ УЧНІВ ДО ВНЗ ЗА ДОПОМОГОЮ ДИСКРИМІНАНТНОГО АНАЛІЗУ

О. М. Халецька*

Контактний тел.: 099-534-44-51

E-mail: olyakhaletskaya@gmail.com

Б. В. Шамша

Кандидат технічних наук, професор*

Контактний тел.: 095-587-33-32

*Кафедра інформаційних управляючих систем
Харківський національний університет радіоелектроніки
пр. Леніна, 14, м. Харків, 61166

1. Вступ

Моніторинг і аналіз обумовлені необхідністю постійного відстеження стану навчально-виховного процесу, окремих його ланок з метою діагностики, аналізу, корекції, прогнозування управлінських дій по досягненню планованого результату.

Прогнозування вступу учнів до ВНЗ, в залежності від вхідних даних, може відбуватися не тільки на рівні учня, класу, а й на рівні шкіл і країни загалом. На основі чого можна планувати розвиток, науковий потенціал України, вносити корективи в освітній процес. Саме тому задача дослідження, аналізу і прогнозування вступу учня до ВНЗ є важливою і актуальною проблемою в Україні.

2. Постановка проблеми

Об'єкт дослідження: спеціалізована школа №3 м. Кролевець Сумської обл. і процес аналізу та прогнозування вступу учнів до ВНЗ.

Предмет дослідження: методи і моделі прогнозування якості навчання студентів на основі аналізу даних взятих із школи за 2008-2011 роки.

Наукова задача: дослідження, аналіз і вибір методів прогнозування з ціллю побудови моделі прогнозу вступу учнів до ВНЗ.

Оскільки прогнозування ніколи не зможе цілком знищити ризик при прийнятті рішень, необхідно явно визначати неточність прогнозу. Звичайно, прийняте рішення визначається результатами прогнозу (при цьому передбачається, що прогноз правильний) з урахуванням можливої помилки прогнозування.

Сказане вище припускає, що прогноуюча система повинна забезпечувати визначення помилки прогнозування, також як і саме прогнозування. Такий підхід значно знижує ризик об'єктивно зв'язаний із процесом прийняття рішень.

Ціль дослідження - провести аналіз методів прогнозування, вибір оптимальних методів для побудови моделі прогнозу вступу учнів до ВНЗ.

Задачі, що допоможе вирішити підсистема прогнозування вступу учнів до ВНЗ: відібрати об'єктивні методи прогнозування і критеріїв оцінки результатів; розробити документацію для методичного супроводу проведення прогнозування; розробити математичне забезпечення системи прогнозування вступу учнів до ВНЗ; випробувати в дії механізм інформування всіх учасників освітньої системи; одержати об'єктивні дані, що свідчать про визначений рівень якості освіти; допомогти класним керівникам в роботі з успішністю учнів; намітити стратегію освітнього процесу.

3. Аналіз досліджень і публікацій

Нові освітні умови системи керування характеризуються децентралізацією, розпадом ієрархічної управлінської системи. В умовах ринку кожен навчальний заклад змушений звернутися до пошуку "свого обличчя" і відповідної структури керування, що перестала бути універсальною. Традиційні компоненти керування - планування, організація, керівництво і контроль, доповнюються новим функціональним складом – аналіз і прогнозування, що забезпечує організованість спільної діяльності студентів і викладачів

і спрямованим на досягнення освітніх цілей розвитку університету.

Управління, таким чином, повинне вчасно реагувати на зміни в управляючих об'єктах, повинне мати про них інформацію. Повинний бути зворотний зв'язок, реалізована у формі контролю. Більш того, керування повинне адаптуватися до змін освітньої ситуації, до її нових вимог. Моніторинг і аналіз обумовлені необхідністю постійного відстеження стану навчально-виховного процесу, окремих його ланок з метою діагностики, аналізу, корекції, прогнозування управлінських дій по досягненню планованого результату.

В українських школах поставлена проблема практично не досліджена.

4. Задача дискримінації в загальному вигляді

Для рішення задачі прогнозування якості навчання - чи вступить учень до ВНЗ 1-2 рівня акредитації, чи до ВНЗ 3-4 рівня акредитації по результатам середнього балу за три останні роки навчання, використовуємо дискримінантний аналіз та байєсовську процедуру класифікації.

Дискримінантний аналіз (ДА) є розділом багатомірного статистичного аналізу, який дозволяє вивчати відмінності між двома і більшою кількістю груп об'єктів за декількома змінними одночасно. Дискримінантний аналіз - це загальний термін, що відноситься до декількох щільно зв'язаних між собою статистичних процедур. Ці процедури можна розділити на методи інтерпретації міжгрупових розбіжностей - дискримінації і методи класифікації спостережень за групами. У загальному випадку задача дискримінації формулюється в такий спосіб.

Мається набір об'єктів, розбитий на декілька класів $W_i, i = 1, 2, \dots, k$ (тобто для кожного об'єкта можна сказати, до якого класу він належить). Для кожного об'єкта існують виміри $X = (x_1, x_2, \dots, x_p)$ декількох кількісних характеристик. Необхідно знайти спосіб, що дозволяє на основі наявних характеристик визначити клас, до якого належить об'єкт. Це дозволить для нових об'єктів, що відносяться до того ж набору, визначити групи, до яких вони будуть відноситись. Характеристики, що приміняються для того, що б відрізнити один клас від іншого, називаються дискримінантними змінними.

Для рішення цієї задачі необхідно побудувати функції вимірюємих характеристик, значення яких і пояснюють розбиття об'єктів на групи.

5. Рішення задачі

Вхідні дані:

Таблиця 1

Вибірка даних про середній бал учнів, які вступили до ВНЗ 3-4 рівнів акредитації

| № | 9 клас | 10 клас | 11 клас |
|---|--------|---------|---------|
| 1 | 2 | 3 | 4 |
| 1 | 11,1 | 10,5 | 10,3 |
| 2 | 10,6 | 9,9 | 11,2 |

Продовження таблиці 1

| 1 | 2 | 3 | 4 |
|----------------------------------|------|------|------|
| 3 | 9,6 | 9,2 | 9,5 |
| 4 | 10,2 | 9,1 | 9,4 |
| 5 | 8,3 | 7,8 | 7,8 |
| 6 | 8,5 | 7,5 | 8,3 |
| 7 | 10,3 | 9,9 | 10,5 |
| 8 | 8,2 | 9,8 | 9,6 |
| 9 | 7,8 | 7,4 | 7,2 |
| 10 | 7,9 | 7,4 | 8,8 |
| 11 | 7,6 | 7,8 | 7,2 |
| 12 | 10,6 | 11,0 | 11,2 |
| 13 | 8,0 | 8,1 | 8,5 |
| 14 | 7,8 | 8,5 | 8,3 |
| 15 | 10,6 | 10,3 | 9,2 |
| 16 | 8,3 | 8,9 | 8,9 |
| 17 | 8,3 | 8,5 | 10,1 |
| 18 | 7,9 | 7,6 | 7,8 |
| 19 | 11,2 | 10,1 | 10,8 |
| 20 | 10,2 | 9,1 | 9,1 |
| 21 | 10,2 | 10,3 | 10,6 |
| 22 | 8,5 | 8,9 | 10,3 |
| 23 | 10,9 | 11,3 | 11,2 |
| 24 | 8,3 | 8,6 | 8,4 |
| 25 | 10,3 | 10,6 | 10,2 |
| 26 | 10,1 | 9,3 | 8,9 |
| 27 | 8,5 | 9,1 | 9,3 |
| 28 | 6,9 | 7,5 | 8,1 |
| 29 | 9,9 | 9,1 | 10,5 |
| 30 | 8,3 | 8,4 | 8,9 |
| 31 | 10,3 | 10,9 | 10,1 |
| 32 | 8,9 | 7,7 | 8,1 |
| 33 | 8,9 | 8,1 | 8,2 |
| 34 | 9,1 | 8,2 | 8,6 |
| 35 | 8,1 | 8,2 | 8,5 |
| 36 | 10,6 | 11,3 | 11,1 |
| μ - середнє значення вектора | 9,2 | 9,1 | 9,3 |

Таблиця 2

Вибірка даних про середній бал учнів, які вступили до ВНЗ 1-2 рівнів акредитації

| № | 9 клас | 10 клас | 11 клас |
|---|--------|---------|---------|
| 1 | 2 | 3 | 4 |
| 1 | 5,6 | 5,2 | 5,9 |
| 2 | 7,1 | 7,2 | 6,3 |
| 3 | 6,9 | 7,0 | 6,2 |
| 4 | 5,4 | 5,2 | 5,1 |
| 5 | 6,2 | 6,1 | 5,9 |
| 6 | 6,7 | 6,9 | 7,3 |
| 7 | 4,4 | 4,5 | 4,2 |
| 8 | 5,5 | 5,2 | 5,1 |
| 9 | 5,9 | 6,2 | 6,1 |

Продовження таблиці 2

| | | | |
|--------------------------|-----|-----|-----|
| 1 | 2 | 3 | 4 |
| 10 | 5,2 | 5,8 | 5,4 |
| 11 | 8,1 | 7,9 | 6,3 |
| 12 | 4,9 | 4,8 | 5,2 |
| середнє значення вектора | 6,0 | 6,0 | 5,8 |

Класифікація учнів проводиться по 2-х основних класифікаційних ознаках – середнім балом за 3 останні роки навчання.

$p = 3$ - кількість параметрів класів.

Відповідно до умови задачі студенти розділені на 2 класи:

W1 - учні, які вступили до ВНЗ 3-4 рівнів акредитації;

W2 - учні, які вступили до ВНЗ 1-2 рівнів акредитації.

Відповідно, X - вектор оцінок студентів, має 3 складові:

$$X = (x_1, x_2, x_3).$$

Нехай маємо об'єкт, якому відповідає вектор спостережень $x = (x_1, x_2, \dots, x_p) \in Y$. Потрібно віднести його на основі цих спостережень до популяції W1 з розподілом $N(\mu_1, \Sigma^{p \times p})$ до популяції W2 з розподілом $N(\mu_2, \Sigma^{p \times p})$,

де $\mu_i = (\mu_{i1}, \mu_{i2}, \dots, \mu_{ip})$, $i = 1, 2$; μ_1, μ_2 - вектора середніх; Σ - матриця коваріацій.

Подібні задачі можуть вирішуватися в декількох варіантах: коли відомі апіорні і апостеріорні вірогідності і вартості помилкової класифікації; коли невідомі апіорні і апостеріорні вірогідності і вартості помилкової класифікації; коли відомі середні значення векторів вхідних даних; коли невідомі середні значення векторів вхідних даних; коли закон розподілення – нормальний; коли закон розподілення – ненормальний.

В даній роботі передбачається, що вектори середніх μ_1, μ_2 і матриця коваріацій Σ - невідомі, закон розподілення – нормальний, невідомі апіорні вірогідності і вартості помилкової класифікації. Якщо x_{11}, \dots, x_{1n_1} і x_{21}, \dots, x_{2n_2} - незалежні випадкові вибірки з популяцій W1, W2 відповідно, то можна оцінити μ_i вибірковою вектором середніх.

Апіорні імовірності q_1, q_2 віднесення X_i до того чи іншого класу можна оцінити величинами $q_1 = \frac{n_1}{n_1 + n_2}$, $q_2 = \frac{n_2}{n_1 + n_2}$,

де n_1 - об'єм вибірки №1, n_2 - об'єм вибірки №2.

У такій ситуації неможливо знайти процедуру класифікації, що була б оптимальною в змісті вартості помилкової класифікації. Однак можна показати, що якщо параметри в узагальненої байєсовської процедури замінити їхніми обґрунтованими оцінками, то в результаті очікувана вартість помилкової класифікації буде убувати при $n_1, n_2 \rightarrow \infty$. Оскільки приведені вище оцінки обґрунтовані, узагальнена процедура байєсовської класифікації на основі оцінок параметрів полягає в наступному: спочатку зважується система

рівнянь із заміною μ_{ij} на \bar{x}_{ij} , де $i = 1, 2, j = 1, \dots, p$, і заміною σ_{jm} на s_{jm} , $t = 1, \dots, p$. Потім отримані оцінки коефіцієнтів $\alpha_1, \dots, \alpha_p$ (позначимо їх $\alpha_1, \dots, \alpha_p$) використовуються для визначення значення дискримінантної функції z_{ij} для кожного вектора спостережень x_{ij} . Далі ξ_i оцінюються величинами

$$\bar{z}_i = \frac{1}{n_j} * \sum_{l=1}^{n_j} z_{il},$$

а σ_z^2 , - величиною

$$\sigma_z^2 = \sum_{j=1}^p \sum_{m=1}^p a_j s_{jm} a_m.$$

Таким чином, узагальнена байєсовська процедура оцінювання складається у віднесенні x до W_1 , якщо

$$z = \sum_{i=1}^p a_i x_i \geq \frac{\bar{z}_1 + \bar{z}_2}{2} + \ln \frac{q_1 C(1/2)}{q_2 C(2/1)},$$

і до W_2 - у протилежному випадку.

Вибіркова відстань Махаланобіса

$$D_2 = (\bar{z}_1 - \bar{z}_2)^2 / s_z^2$$

є оцінкою для $\Delta 2$.

У результаті роботи програм дискримінантного аналізу, як правило, одержуємо наступне: оцінки коефіцієнтів дискримінантної функції a_1, \dots, a_p ; значення дискримінантної функції z_{ij} для кожного вектора спостережень; вибіркові середні \bar{z}_1, \bar{z}_2 ; вибіркова відстань Махаланобіса D^2 . Ця інформація достатня для запису процедури класифікації.

Математичне очікування для даних класів оцінок в 5-тибальній системі відповідно рівні:

$$\mu_1 = (9.2; 9.1; 9.3); \mu_2 = (6; 6; 5.8).$$

Коваріаційна і кореляційна матриці, а також перевірка нормальності закону розподілу вхідних даних були проведені в Statistica.

Розрахуємо апіорні імовірності помилкової класифікації об'єкта.

$$q_1 = \frac{n_1}{n_1 + n_2} = 36 / (36 + 12) = 0.75.$$

$$q_2 = \frac{n_2}{n_1 + n_2} = 12 / (36 + 12) = 0.25.$$

Вартість помилкової класифікації прийемо рівною 3000 грн.

Розрахунок постійних коефіцієнтів $\alpha_1 \dots \alpha_p$, що максимізують відстань Махаланобіса:

Розв'яжемо систему рівнянь типу

$$\begin{aligned} \alpha_1 S_{11} + \alpha_2 S_{12} + \dots + \alpha_p S_{1p} &= \mu_{11} - \mu_{21}, \\ \alpha_1 S_{21} + \alpha_2 S_{22} + \dots + \alpha_p S_{2p} &= \mu_{12} - \mu_{22}, \\ \dots & \dots \dots \dots \dots \dots \dots \\ \alpha_1 S_{p1} + \alpha_2 S_{p2} + \dots + \alpha_p S_{pp} &= \mu_{1p} - \mu_{2p}. \end{aligned}$$

де

S_{ip} - відповідні елементи матриці коваріацій.

$$\begin{aligned} 0.51\alpha_1 + 0.23\alpha_2 + 0.33\alpha_3 &= 9.2 - 6 = 3.2, \\ 0.23\alpha_1 + 0.51\alpha_2 + 0.27\alpha_3 &= 9.1 - 6 = 3.1, \\ 0.33\alpha_1 + 0.27\alpha_2 + 0.58\alpha_3 &= 9.3 - 6.8 = 2.5. \end{aligned}$$

Розв'язав дану систему рівнянь отримаємо:

$$\begin{aligned} \alpha_1 &= 4.535, \\ \alpha_2 &= 4.136, \\ \alpha_3 &= 0.196. \end{aligned}$$

Дискримінанта функція має вигляд:

$$\begin{aligned} Z &= \alpha_{i1}x_1 + \alpha_{i2}x_2 + \dots + \alpha_{ip}x_p, \\ Z &= 4.535x_1 + 4.136x_2 + 0.196x_3. \end{aligned}$$

Математичне очікування дискримінантної функції для кожного класу (дискримінанта функція розраховується для кожного об'єкту обох класів):

$$\xi_i = \sum_{j=1}^p \alpha_j \mu_{ij},$$

Для першого – 80,77, для другого – 52,7, середнє значення – 66,7.

Визначення границь класів.

$$K = \ln \frac{q_2 C(1|2)}{q_1 C(2|1)} = \ln \frac{0,75 * C(1|2)}{0,25 * C(2|1)} = 1.09.$$

Загальна процедура класифікації Байеса складається в відношенні X до W_1 , якщо

$$\sum_{j=1}^p \alpha_j x_j < \frac{\xi_1 + \xi_2}{2} + \ln \frac{q_2 C(1|2)}{q_1 C(2|1)},$$

і до W_2 , якщо

$$\sum_{j=1}^p \alpha_j x_j \geq \frac{\xi_1 + \xi_2}{2} + \ln \frac{q_2 C(1|2)}{q_1 C(2|1)},$$

$$\frac{\xi_1 + \xi_2}{2} + \ln \frac{q_2 C(1|2)}{q_1 C(2|1)} = 66.7 + 1.09 = 67.79$$

Значить, якщо значення дискримінантної функції більше, ніж 67,79, об'єкт буде віднесений до класу W_1 (учні, які вступили до ВНЗ 3-4 рівнів акредитації) і до класу W_2 в іншому випадку.

Дана процедура мінімізує помилкову вартість неправильної класифікації.

Визначимо дисперсії дискримінантної функції:

$$\sigma_z^2 = \sum_{v=1}^p \sum_{j=1}^p \alpha_v S_{vj} \alpha_j = 2.585,$$

$$\Delta^2 = D^2 = \frac{(\xi_1 - \xi_2)^2}{\sigma_z^2} = 6.71 / 2.585 = 2.595.$$

Отримаємо вірогідність помилкової класифікації:

Найчастіше параметри розподілу не представляють труднощів при розрахунку. Складність - це оцінка ймовірності помилкової класифікації.

У випадку коли параметри розподілів відомі, значення ймовірностей помилкової класифікації $Pr(2|1)$ і $Pr(1|2)$ даються формулами

$$Pr(2|1) = \Phi \left(\frac{K - \frac{1}{2} \Delta^2}{\Delta} \right), \tag{1}$$

$$Pr(1|2) = \Phi \left(\frac{-K - \frac{1}{2} \Delta^2}{\Delta} \right) \tag{2}$$

У випадку, коли параметри оцінюються, існує кілька методів оцінки цих ймовірностей.

Метод 1. Оскільки D^2 є оцінкою Δ^2 , у формулах (1) і (2) можна замінити Δ^2 на D^2 . Треба примітити, що такі оцінки будуть зміщеними, тобто дійсна ймовірність помилкової класифікації буде в середньому більше, ніж така оцінка. Перевагою методу є простота таких оцінок: їх легко одержати за результатами роботи програми.

Метод 2. Цей метод складається в класифікації кожного елемента вибірки об'єму n_1 з популяції W_1 і вибірки об'єму n_2 із W_2 . Якщо m_1 - число спостережень із W_1 віднесених до W_2 , і m_2 - число спостережень з W_2 , класифікованих у W_1 , то $Pr(2|1) = m_1 / n_1$, $Pr(1|2) = m_2 / n_2$. Цей метод дає більше зміщення, чим попередній.

6. Висновки

Використовуючи побудовану модель прогнозування вступу учнів до ВНЗ (на прикладі прогнозу чи вступить учень до ВНЗ 1-2 рівня акредитації, чи до ВНЗ 3-4 рівня акредитації по результатам середнього балу за три останні роки навчання) можна здійснювати корекції в освіті.

Заходу для корекції подальшої успішності учнів містять у собі:

- листа батькам із указівкою не тільки результатів навчання учня на кожному етапі і його поточних оцінок, але і тенденції успішності;
- обговорення на розширених засіданнях комісій із запрошенням батьків;
- формування навчальної мотивації;
- розвиток професійних інтересів;
- проведення додаткових занять по дисципліні з метою більш доступного викладу важкого матеріалу;
- прогнозування кінцевої вірогідності вступу учнів до ВНЗ.

Діагностування якості знань і умінь учнів має потребу в систематичному відстеженні ступеня навченості з метою поетапного рішення навчальних задач, встановлення й усунення пробілів в освоєваному матеріалі з наступною корекцією в ході навчального процесу і прогнозування змісту і технології навчання.

Література

1. Афіфі, А. Статистичний аналіз з використанням ЕОМ [Текст] / А. Афіфі, С. Ейзен. - Москва: «Мир», 1982 - 480 с.
2. Шамша, Б.В. Математичне забезпечення інформаційно-управляючих систем [Текст] / Б.В. Шамша. - Харків: ТОВ «Сміт», 2005 - 448 с.

Обгрунтована необхідність вибору критеріїв ефективності РЧС та його оцінки для впровадження перспективних радіотехнологій на користь перерозподілу смуг частот, який супроводжується вивільненням ресурсу радіочастотного спектра
Ключові слова: РЧС, радіочастотний ресурс, конверсія

Обоснована необходимость выбора критериев эффективности РЧС и его оценки для внедрения перспективных радиотехнологий в интересах перераспределения полос частот, сопровождающееся высвобождением ресурса радиочастотного спектра
Ключевые слова: РЧС, радиочастотный ресурс, конверсия

The necessity of choice RFS performance criteria and assessment for the introduction of advanced radio technologies for reallocation of frequency bands, accompanied by the release of radio spectrum resource is grounded
Keywords: RFS, radio frequency resource, conversion

УДК 621.38.001:006.354

ВОПРОСЫ ДИНАМИЧЕСКОГО УПРАВЛЕНИЯ В ЗАДАЧАХ КОНВЕРСИИ РЕСУРСА РАДИОЧАСТОТНОГО СПЕКТРА

Л. А. Токарь

Кандидат технических наук, старший преподаватель
 Кафедра телекоммуникационных систем
 Харьковский национальный университет
 радиоэлектроники
 пр. Ленина, 14, г. Харьков, Украина, 61166
 Контактный тел.: (057) 702-13-20, 097-456-96-65
 E-mail: tkc2006@ukr.net

Введение

Радиочастотный ресурс (РЧР) является ограниченным ресурсом. Радиочастоты используются как в коммерческих интересах, в частности, для подвижной связи и радиовещания, так и в интересах органов общественного сектора.

Вопрос регулирования радиочастотного спектра (РЧС) является актуальным для развития систем и средств связи различного назначения, так как спрос на РЧР превышает предложение на его использование.

Рост потребностей в свободном радиочастотном спектре для внедрения перспективных радиотехнологий в гражданском потребительском секторе обуславливает важность задачи повышения эффективности управления РЧС, достигаемой, в частности, за счет перераспределения и конверсии РЧР.

Основная часть

Существующие технологии, а также множество частных технологий (технологий не коммерческой деятельности), используют средства связи и включают:

технологии обработки сигналов, сетевые технологии, технологии передачи и обработки данных, технологии доступа в телефонную сеть общего пользования, технологии биллинга, технологии защиты информации, технологии производственных процессов.

Переход на цифровые методы передачи стирает различия между сигналами различных служб для объединения их в единый информационный поток, с последующей его унификацией, образуя единое телекоммуникационное пространство, характеризующееся универсальностью построения сетевой архитектуры.

Развитие прогрессивных информационно-телекоммуникационных технологий в первую очередь влияет на систему распределения и управления РЧР.

Задача повышения эффективности использования РЧС за счет его перераспределения и конверсии является государственной задачей, требующей своевременного решения вопросов управления в этой области, а также методически обоснованных и эффективных критериев и показателей оценки по перераспределению требуемых участков спектра с целью перехода на новую технологическую платформу.