

ПРОБЛЕМЫ УСТОЙЧИВОСТИ АЛГОРИТМА КЛАССИФИКАЦИИ КАФЕДР УНИВЕРСИТЕТА

В.И. Ляликова

Гродненский государственный университет им. Я.Купалы, г. Гродно, Беларусь
vialikova@tut.by

При анализе эффективности деятельности кафедр университета руководством часто решается задача их ранжирования. На сегодняшний день нет единого подхода к построению рейтинга изучаемых объектов. Как правило, в качестве рейтинга используется некоторый интегрированный показатель. Для формирования интегрированного показателя используются экспертные методы на различных уровнях анализа:

- отбор показателей для решения поставленной задачи,
- выбор экспертов,
- определение экспертами весов для изучаемых показателей и т.д.

В результате строится рейтинг изучаемых структур, который тут же подвергается критике, в первую очередь за выбранные веса показателей.

В работе для решения поставленной задачи предлагается использовать методы прикладной статистики (факторный, кластерный и дисперсионный анализ), которые позволяют формулировать выводы в терминах вероятностей ошибок. Основная идея этого подхода состоит в том, чтобы не использовать экспертные оценки. Имеется опыт применения факторного анализа для ранжирования экономических объектов. Такого рода исследования проводились в Центральном экономико-математическом институте Российской академии наук при построении рейтинга качества жизни населения в различных регионах [1]. Важность показателей оценивается методами математической статистики, исходя из природы случайности. Применение факторного анализа позволяет отобрать наиболее варьирующие показатели. Если первый главный фактор сохраняет большой процент общей дисперсии исходных показателей, то можно считать, что он является удовлетворительной аппроксимацией всех анализируемых частных критериев.

Такой подход был использован при анализе развития малого предпринимательства в районах Гродненской области [2]. На основании значений первого главного фактора был построен рейтинг районов, отражающий уровень развития малого предпринимательства.

Изучение проблемы ранжирования кафедр университета показало, что для решения поставленной задачи не достаточно использование только первого главного фактора [3, 4].

Предлагается следующий алгоритм ранжирования и классификации кафедр университета:

- 1) отбор показателей для решения поставленной задачи;
- 2) нормировка показателей, то есть приведение их к единой шкале измерения, например, в расчете на одного работающего на кафедре;
- 3) преобразование исходных показателей с помощью метода главных компонент и определение числа главных факторов для построения интегрированного показателя;
- 4) вычисление рейтинга изучаемых объектов на основании значений первых главных факторов;
- 5) разбиение объектов исследования (кафедр) на группы с помощью метода k-средних кластерного анализа по значению интегрированного показателя;
- 6) контроль различия средних значений интегрированного показателя в построенных кластерах с помощью дисперсионного анализа.

Таким образом, ранжирование кафедр Гродненского университета по результатам 2007 года проведен с помощью факторного анализа.

Для вычисления рейтинга была использована следующая формула:

$$R = a_1 F_1 + a_2 F_2 + \dots + a_q F_q. \quad (1)$$

Здесь F_i – значения первых главных факторов, веса главных факторов $a_i = V(Y_i) / V$ представляют собой часть дисперсии исходных показателей, которая обеспечивается первой главной компонентой, $V = \sum_{i=1}^p \sigma_{ii}$ – суммарная дисперсия исходных показателей, $V(Y_i)$ – дисперсия i-ой главной компоненты.

Особенностью построения главных факторов является то, что они центрированы относительно нуля (их средние значения равны нулю). Тогда, в результате применения формулы (1) мы получим значение среднего рейтинга равно нулю. Положительное значение рейтинга свидетельствует о том, что кафедра заняла место выше среднего, отрицательное значение – ниже среднего.

Представляет интерес не только построение рейтинга кафедр, но и классификация их по однородному значению интегрированного показателя. Ставится задача построения групп кафедр, внутри которых различия в рейтинге являются недостоверными (случайными), а различие групп достоверно. Эта задача решается с помощью метода k-средних кластерного анализа. Число кластеров выбирается по возможности больше, при условии, что средние значения рейтинга для различных кластеров различаются достоверно.

Рассмотрим пример классификации кафедр Гродненского университета по учебной работе. Было построено 7 кластеров. В таблице 1 приведены средние значения интегрального показателя, число кафедр и среднее число преподавателей кафедры в соответствующем кластере.

Таблица 1 – Средние значения интегрального показателя в построенных кластерах

Кластер	Среднее значение рейтинга	Число кафедр в кластере	Среднее число преподавателей кафедры
1	1,037	1	9,0
2	0,457	11	11,55
3	0,197	7	9,85

Продолжение табл.			
4	0,005	11	12,35
5	-0,075	12	13,27
6	-0,210	18	11,94
7	-0,354	8	14,87

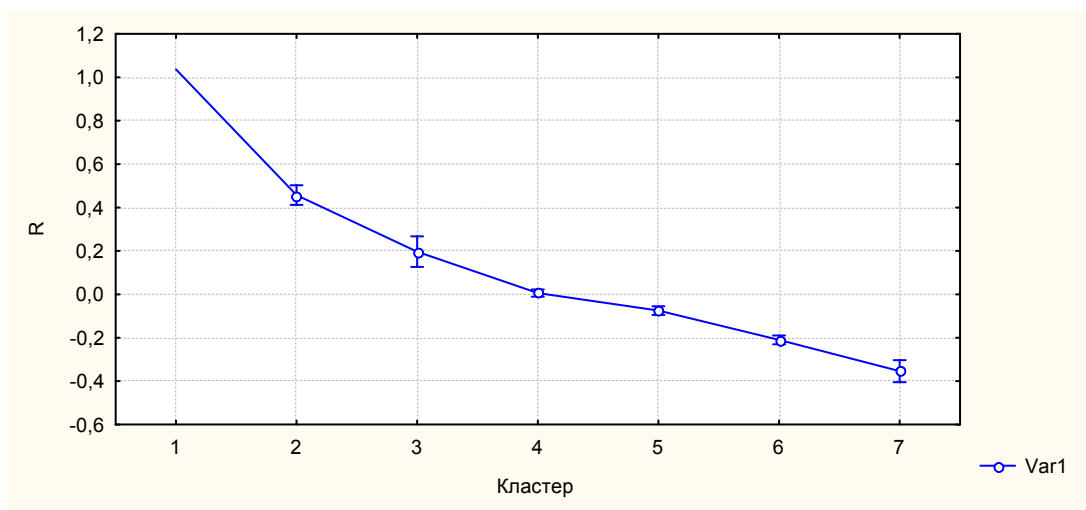


Рисунок 1. Средние значения рейтинга в кластерах

Первый кластер представлен одной кафедрой, во второй входят 11 кафедр, в третий – 7 и так далее. Можно было бы предположить, что чем меньше кафедра, тем выше ее результаты, так исходные показатели взяты в расчете на одного преподавателя. Такой вывод подходит только для последнего кластера, где среднее число преподавателей кафедры составило около 15 человек. На рисунке 1 приведен график средних значений интегрального показателя в кластерах и 95-процентные доверительные интервалы для этих средних. Видно, что наибольший разброс рейтинга имеется в третьем и седьмом кластерах, четвертый, пятый и шестой кластеры содержат очень однородные кафедры.

В данном случае распределения случайной величины R не соответствует нормальному распределению ($\chi^2 = 8,13$, $p = 0,004$).

Был проведен анализ причин несоответствия рейтинга нормальному распределению, а также сильного отличия ведущей кафедры от остальных с точки зрения математической статистики.

Изучив исходную базу данных, можно обнаружить «выбросы» - данные, сильно отличающиеся от остальных. Так, например, одна кафедра зафиксировала 87 профориентационных мероприятий.

С целью изучения устойчивости алгоритма классификации было сделано предположение, что это ошибочное значение. Далее этот выброс был изменен на число 9 по количеству членов кафедры. В результате получим следующую классификацию кафедр (таблица 2).

Таблица 2 – Средние значения интегрального показателя в кластерах после удаления выброса

Кластер	Среднее значение рейтинга	Число кафедр в кластере
1	0,649	6
2	0,350	7
3	0,125	10
4	-0,002	11
5	-0,111	12
6	-0,252	19
7	-0,485	3

После удаления выброса интервал изменения рейтинга уменьшился. Первый кластер содержит не 1, а 6 кафедр, хотя ведущая кафедра сохранила свое место. Отметим, что в этом случае рейтинг соответствовал нормальному распределению ($\chi^2 = 3,42$, $p = 0,064$). Таким образом, точность сбора информации существенным образом влияет на классификацию изучаемых объектов.

Литература

1. Aivazian S.A. Synthetic indicators of quality of life and sustainable development: construction and utilization for social-economic management. Proceedings of the Seventh International Conference "Computer Data Analysis and Modeling". Minsk. 2004. v.2. С.10-18.
2. Ляликова В.И., Ли Чон Ку. Сравнительный анализ состояния и динамики малого бизнеса в районах Гродненской области. - «Экономика и управления». - 2006.- №3.- С. 128-134.
3. Ляликова В.И., Калинина Е.Н. Автоматизированная система ранжирования структур университета. Современный университетский комплекс и регион. Материалы международной научно-практической конференции. – Гродно. - 2007. – С.174-179.
4. Lialikova V.I. A Rating Algorithm in University Management. Computer data analysis and modeling. Complex stochastic data and systems. – Minsk. – 2007. – V2. - С.108-111.