Н.Н. Двоерядкина, Н.А. Чалкина, Т.А. Макарчук

ИСПОЛЬЗОВАНИЕ ТЕОРИИ НЕЧЕТКИХ МНОЖЕСТВ В СОЦИОЛОГИИ

There are considered application of non-numeric objects for estimation socio-economic showings and theirs analysis by methods of statistical interval data.

Современное исследование общественных и социально-экономических процессов в значительной степени опирается на результаты социологических исследований. Социолог – это специалист по информации, которую нужно не только собрать, но и грамотно обработать.

Используя первичную информацию, собранную в результате социологических опросов, необходимо сформировать данные, которые допускают возможность математической обработки.

Формируя данные, исследователь ставит в соответствие значениям переменной, имеющей содержательный смысл, числовые значения. Такое соответствие называется шкалой измерения переменной. В зависимости от свойств переменной выделяют шкалы: номинальную, ранговую, интервальную и шкалу отношений.

Номинальная шкала является самым «низким» уровнем измерения. В этом случае используется только равенство или неравенство значений. Примерами таких переменных являются «пол», «профессия».

Часто значения признака выражают степень проявления какого-либо свойства и могут быть упорядочены. Такая шкала называется ранговой.

Интервальная шкала предполагает, что можно определить не только порядок значений, но и расстояние между значениями. Шкала отношений в дополнение к свойствам интервальной шкалы позволяет измерять пропорции значений.

Для анализа количественных (числовых) данных, заданных шкалой отношений либо интервальной, у социолога имеется богатый арсенал статистических методов (корреляционный, регрессионный анализ, законы больших чисел, центральная предельная теорема и др.). Однако часто ему приходится работать с признаками нечисловой природы, измеренными номинальной или порядковой шкалой. К ним нельзя применить многие классические методы математической статистики, что существенно затрудняет исследования.

В общем случае под нечисловыми данными понимают элементы пространств, не являющихся линейными (векторными), в которых нет операций сложения элементов и их умножения на действительное число.

Основы нечеткой логики были заложены в конце 60-х гг. в трудах американского ученного Л.А. Заде. В настоящее время активно ведутся работы по статистическому анализу нечисловых данных.

В основе алгебры нечеткой логики лежат два основных понятия: нечеткого множества и нечетких операций над ними. Элементами нечеткого множества являются лингвистические переменные.

Нечеткое подмножество A универсального множества U характеризуется функцией принадлежности f(u;A), которая ставит в соответствие каждому элементу u число f(u; A) из отрезка от [0;1].

Лингвистической переменной называют переменную, принимающую в качестве своих значений нечеткие множества.

В нечеткой логике операции: дизъюнкции (or), конъюнкции (and), отрицания (not) импликации (=>) обозначаются и определяются следующим образом:

v(p or q) = max(v(p), v(q));
v (p and q) = min(v(p), v(q));
v (not p) = 1 - v(p);
v(p => q) = min(l, 1- v(p) + v(q)).

Математический аппарат статистики объектов нечисловой природы основан не на свойстве линейности пространства, а на применении симметрии и метрик в нем, поэтому существенно отличается от классического [2]. Для анализа нечисловых данных существует апробированный аппарат вне рамок классического подхода, в частности анализ соответствий, факторный анализ и др. Один из результатов статистических методов анализа нечисловой информации – возможность перевода нечисловых данных в интервальную шкалу (например, с помощью факторного анализа).

В качестве условного примера рассмотрим определение качества знаний по предмету у студентов. Для анализа выберем три переменные: успеваемость, наличие академических способностей, наличие или отсутствие интереса к изучению предмета [1].

Таблица 1

№	Способности	Интерес к предмету	Успеваемость
1	Способный	Заинтересованный	4
2	Способный	Заинтересованный	5
3	Неспособный	Незаинтересованный	2
4	Способный	Незаинтересованный	4
5	Неспособный	Заинтересованный	3
6	Неспособный	Заинтересованный	2
7	Неспособный	Незаинтересованный	2
8	Неспособный	Незаинтересованный	4

Табл. 1 не позволяет сделать какой-то объективный вывод относительно качества знаний. Методами факторного анализа, идея которого состоит в сжатии матрицы признаков в матрицу с меньшим числом переменных, сохраняющую почти ту же самую информацию, что и исходная матрица, можно сконцентрировать исходную информацию, содержащуюся в нескольких переменных в одной латентной характеристике, отвечающей за качество.

Для нахождения латентного фактора необходимо определить коэффициенты корреляции исходных данных по формуле:

$$r_{xy} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sqrt{\overline{x^2} - \overline{x}^2} \cdot \sqrt{\overline{y^2} - \overline{y}^2}}. \qquad (1)$$

Коэффициенты корреляции показателей качества представлены в табл. 2.

Таблица 2

Переменные	Коэффициенты корреляции		
	способность	интерес	успеваемость
Способности	1	0,26	0,77
Интерес	0,26	1	0,23
Успеваемость	0,77	0,23	1

Основные результаты факторного анализа выражаются в наборах факторных нагрузок и факторных весов. Факторные нагрузки являются значениями коэффициентов корреляции каждого из исходных признаков с каждым из выявленных факторов. Для построения матрицы факторных нагрузок необходимо найти собственные числа корреляционной матрицы переменных, решив уравнение:

$$\begin{vmatrix} 1-\lambda & 0,26 & 0,77 \\ 0,26 & 1-\lambda & 0,23 \\ 0,77 & 0,23 & 1-\lambda \end{vmatrix} = 0$$

Для полученной корреляционной матрицы, представленной в табл. 2, собственные числа $\lambda_1 = 1,902$, $\lambda_2 = 0,228$, $\lambda_3 = 0,866$.

Согласно критерию Кайзера значимыми являются только факторы, с собственными значениями, большими 1. Нормированные координаты собственного вектора, соответствующие собственному числу $\lambda_1 = 1,902$, находятся путем решения системы уравнений:

$$\begin{cases} -0,902x_1 + \ 0,26x_2 + \ 0,77x_3 = 0 \\ 0,26x_1 - \ 0,902x_2 + \ 0,23x_3 = 0 \\ 0,77x_1 + \ 0,23x_2 - \ 0,902x_3 = 0 \end{cases}$$

и последующей нормировки по формуле

$$v_{i\,норм} = \frac{v_i}{\sqrt{\sum_i v_i^2}}. \qquad (2)$$

Координаты собственного вектора, соответствующего наибольшему собственному числу, составляют:

$$v_{1\,норм} = (0,967;\ 0,526;\ 0,957)$$

и являются элементами матрицы факторных нагрузок

$$A = \begin{pmatrix} 0,967 \\ 0,526 \\ 0,957 \end{pmatrix}.$$

Элементы матрицы факторных нагрузок являются коэффициентами корреляции между исходными переменными и латентным фактором, отвечающим за качество. Их абсолютные значения показывают наличие достаточно значимой линейной связи между исходными переменными и найденным фактором.

Количественные значения выделенного фактора для каждого из имеющихся объектов содержатся в матрице факторных весов. Значения факторных весов можно рассматривать как значения индекса, характеризующего уровень развития объектов в рассматриваемом аспекте.

Элементы матрицы факторных весов находятся по формуле:

$$P = \left(A^T \cdot A\right)^{-1} \cdot A^T \cdot Z^T, \qquad (3)$$

где A – матрица факторных нагрузок; Z – матрица исходных данных.

Матрица факторных весов P представляет собой вектор-строку, содержащую 8 координат (по количеству наблюдений):

P= (2,5; 2,95; 0,899; 2,254; 1,597; 1,147; 0,899; 1,799).

Числовые значения фактора оценки качества позволяют проранжировать все наблюдения.

Предложенный способ анализа нечисловой информации позволяет решать ряд социологических задач, связанных с использованием порядковых шкал. Однако отметим, что факторный анализ лучше всего использовать для предварительного изучения данных, формирования рабочих гипотез. Особенно удачным может быть его применение при пилотажных исследованиях.

Мы рассмотрели применение факторного анализа к обработке объектов нечисловой информации. Данный метод анализа нечисловых данных особенно хорошо приспособлен для использования в экономике, социологии, педагогических исследований. Разумеется, он не исчерпывает все многообразие фронта научных исследований в этой области. Однако в настоящее время является наиболее доступным и широко реализуемым в различных компьютерных статистических программах.

Исследование поддержано грантом Министерства образования и науки РФ «Развитие научного потенциала высшей школы», регистрационный номер 3.1.1/2265.

1. Крамер Д. Математическая обработка данных в социальных науках: современные методы. – М.: Академия, 2007.
2. Орлов А.И. Эконометрика. – М.: Экзамен, 2003.