КВАНТИТАТИВНАЯ ИСТОРИЯ
QUANTITATIVE HISTORY

ДИСКРИМИНАНТНЫЙ АНАЛИЗ С ИСПОЛЬЗОВАНИЕМ КОНЦЕПЦИИ НЕЧЕТКИХ МНОЖЕСТВ В ЗАДАЧЕ ОЦЕНКИ ДВУХ ПУТЕЙ АГРАРНОЙ ЭВОЛЮЦИИ РОССИИ НА РУБЕЖЕ XIX–XX вв.

DISCRIMINANT ANALYSIS OF THE TWO WAYS OF RUSSIA'S AGRARIAN EVOLUTION AT THE TURN OF 19th — 20th CENTURIES: USING FUZZY SETS CONCEPT

Бородкин Леонид Иосифович
Доктор исторических наук, профессор, заведующий кафедрой исторической информатики исторического факультета МГУ имени М.В. Ломоносова
E-mail: borodkin@hist.msu.ru

Leonid I. Borodkin

Концепции теории нечетких (размытых) множеств получили применение в гуманитарных науках, включая историю и археологию. Эти концепции учитывают специфику задач типологии в социально-гуманитарных исследованиях. Предлагается алгоритм распознавания, в котором значения степени принадлежности объекта обучающей выборки к классам определяются не с помощью эксперта, а алгоритмически; при этом решающее правило является «нечетким» и в режиме экзамена (в развитие подхода, предложенного автором ранее). Дается формальная постановка задачи, излагается алгоритм нечеткой классификации и основанный на его использовании алгоритм нечеткого распознавания (дискриминантного анализа). Во второй части статьи обсуждаются результаты использования предложенного подхода для оценки пространственного соотношения двух путей аграрного развития Европейской России конца XIX — начала XX в.

Ключевые слова: теория нечетких множеств, типология, классификация, алгоритм, распознавание образов, дискриминантный анализ, Европейская Россия конца XIX — начала XX в., аграрное развитие, два пути аграрной эволюции.

The concept of fuzzy set theory have been used in the humanities including researches of historians and archaeologists. This concept allow to take into account the specific problems of classification and typology in the humanities, where, as a rule, there are no rigid boundaries between classes (types) of objects. In this paper we propose an algorithm of pattern recognition in which the degrees of object belonginess to the clusters in the training sample are not determined by an expert but are estimated algorithmically, while the decision rule is fuzzy in the process of recognition. Statement of the problem is given in the second section of the paper, the third one contains a brief description of the algorithm of fuzzy clusterization, and the fourth is based on its use in constructing of the fuzzy pattern recognition algorithm. Finally, the fifth section discusses the results of using the proposed approach in solving the problem of multi-dimensional typology of agricultural development in European Russia at the turn of 19th — 20th centuries, focusing on the space distribution of the two paths of agrarian evolution.

Key words: fuzzy set theory, typology, classification algorithm, pattern recognition, discriminant analysis, European Russia at the turn of 19th — 20th centuries, agricultural development, two paths of agrarian evolution.

ВВЕДЕНИЕ

Один из путей реализации типологического анализа при изучении сложных совокупностей объектов социальной природы связан с применением методов многомерной классификации (кластер-анализ). Большинство этих методов основано на однозначном отнесении каждого объекта к тому или иному классу (типу). Выделяемые с помощью этих методов группы объектов, как правило, позволяют выявить специфику каждого типа. Однако при этом в тени остается внутренняя структура классов, состав ядра типа и его «окружения». Сложным может быть и вопрос об однозначной принадлежности к какому-либо классу «пограничных» объектов (переходного типа).

Адекватный инструмент для решения типологических задач с учетом указанной специфики объектов социальной природы дает теория нечетких множеств (ТНМ). Как указывает основатель этой теории американский профессор Л. Заде, «ее развитие в 60-х годах обязано большей частью своих идей задачам, относящимся к распознаванию образов (классификации). Однако по существу глубинная связь между теорией нечетких множеств и распознаванием образов основана на том обстоятельстве, что большинство реальных классов размыты по своей природе в том смысле, что переход от принадлежности к непринадлежности для этих классов скорее постепенен, чем скачкообразен»[1]. В данной работе мы используем как равноправные оба укоренившихся термина — «нечеткие» и «размытые» множества (перевод с англ. *fuzzy set*).

Концепции ТНМ нашли применение в гуманитарных исследованиях, в том числе в работах историков, принадлежащих к российской школе квантификации[2], в рамках которой историко-типологические исследования активно развивались в 1970–1990-х гг. Так, опыт решения задач многомерной типологии аграрного развития губерний Европейской России конца XIX — начала XX в.[3] выявил необходимость учета указанной специфики. Исходя из этого нами были предложены алгоритмы кластер-анализа (автоматической классификации и классификации с использованием обучающей выборки), основанные на использовании теории нечетких множеств[4]. Использование этих алгоритмов в задачах аграрной типологии губерний позволило расширить возможности интерпретации результатов многомерного анализа, введя в рассмотрение количественные оценки степени типичности объектов каждого класса.

В настоящее время опубликованы уже десятки работ, в которых излагаются различные подходы к построению алгоритмов автоматической классификации, имеющих целью определение степеней принадлежности объектов к классам — размытым множествам[5]. В последние годы подобные методы предлагаются и в работах археологов[6]. Алгоритмы размытой классификации можно рассматривать как естественное развитие соответствующих «традиционных» алгоритмов.

Значительно менее разработанными являются алгоритмы распознавания (классификации с обучением) нечетких паттернов (образов). При разработке таких алгоритмов предполагается, что «учитель» обладает априорным знанием о принадлежности определенных объектов (обучающей выборки) к тем или иным классам, т. е. речь идет о наличии типичных, характерных для каждого класса объектов («маркеров»), что позволяет сформировать «ядро» каждого класса.

В данной статье предлагается алгоритм распознавания, в котором значения степени принадлежности объекта обучающей выборки к классам определяются не с помощью эксперта, а алгоритмически; при этом решающее правило является «нечетким» и в режиме экзамена (в развитие подхода, разработанного нами ранее)[7]. Формальная постановка задачи дается во втором разделе статьи, в третьем представлена краткая характеристика алгоритма нечеткой класссификации, а в четвертом — основанного на его использовании алгоритма нечеткого распознавания. Наконец, в пятом разделе обсуждаются результаты применения предложенного подхода при решении задачи построения многомерной типологии аграрного развития губерний Европейской России на рубеже XIX–XX вв. и оценки пространственного соотношения двух путей ее аграрного развития[8].

ПОСТАНОВКА ЗАДАЧИ

Пусть каждый исследуемый объект характеризуется набором из $\boldsymbol{m}$ параметров, так что каждому объекту соответствует точка $x \in R^m$. Пусть, далее, имеется обучающая выборка объектов $X = \{x_1,\dots, x_n\}$, о каждом из которых известно, к какому из k классов $P_1,\dots, P_k$ в R^m он принадлежит. Таким образом, для объектов из обучающей выборки можно ввести характеристические функции

$$\omega_l = \begin{cases} 1 \text{ при } & x \in P \\ 0 \text{ при } & x \notin P_l \end{cases} \quad l = 1 \div \mathrm{k}, x \in X.$$

Введем в рассмотрение линейную разделяющую функцию (ЛРФ)

$$F(\alpha, x) = (\alpha, x),\ x \in X, \qquad (1)$$

где $\alpha = (\alpha_1,\dots,\alpha_m)$ — вектор весовых коэффициентов признаков.

Специфика данного подхода заключается в том, что при каждом фиксированном положении разделяющей оси производится поиск «компактных» групп объектов множества ***X*** на этой оси; с этой целью используется алгоритм размытой классификации. Соответствие же сформированных групп исходным классам $P_1,\ldots, P_k$ оценивается с помощью суммарной ошибки распознавания, которая определяется с учетом значений функции принадлежности $\mu_1(x),\ldots,\mu_k(x)$ объектов $x\in X$ к размытым подмножествам $\tilde{P}_1,\ldots,\tilde{P}_k$ множества ***X***. Размытые классы $\tilde{P}_i$, $i = 1 \div k$, образуют размытое *K*-разбиение множества ***X***, т. е.

$$\sum_{i=1}^{k}\mu_i(x)=1,\ \mu_i(x)\geq 0,\ i = 1 \div k,\ x \in X. \qquad (2)$$

«Настроив» коэффициенты α_i, $i = 1 \div m$, по обучающей выборке можно распознавать новые объекты $y \notin \boldsymbol{X}$ из пространства R^m, вычисляя по формуле (1) соответствующие значения ЛРФ и оценивая по полученному значению степени принадлежности этих объектов к каждому из классов.

АЛГОРИТМ ПОСТРОЕНИЯ РАЗМЫТОЙ КЛАССИФИКАЦИИ (ОДНОМЕРНЫЙ СЛУЧАЙ)

Существует широкий класс задач распознавания, в которых классы объектов определенным образом упорядочены на некоторой оси (такой является, например, распространенная задача построения рейтинга по целому ряду показателей, с выделением групп передовых, средних и «слабых» объектов). Использование известного алгоритма Fuzzy ISODATA[9] в таких задачах затрудняется, в частности, тем обстоятельством, что достаточно «удаленные» объекты получают в соответствии с алгоритмом FI веса принадлежности ко всем классам, близкие к $1/k$. Предлагаемый алгоритм размытой классификации для одномерного случая (АРКОС) лишен этого недостатка: поведение «хвостов» функции принадлежности при удалении от центра классов является более естественным, если исходить из потребностей практических задач. Действительно, точки, расположенные на оси левее центра класса $\tilde{P}_1$ или правее центра класса $\tilde{P}_k$, получают в соответствии с представленным ниже алгоритмом АРКОС веса принадлежности, равные 1 для ближайшего класса и нулю — для остальных. Значения $\mu_i(x)$ для «внутренних» точек $x \in X$ (расположенных на оси между центрами классов $\tilde{P}_1$ и $\tilde{P}_k$) тем выше, чем ближе данная точка $x \in X$ к центру соответствующего *l-го* класса ($l = 1 \div k$).

Пусть задано некоторое размытое разбиение $P = \{\tilde{P}_1,\ldots,\tilde{P}_k\}$ множества $X \in R^1$ и определены центры классов $V_1 < V_2 < \ldots < V_k$. Положим

$$X_L = \{x\in X \mid x<V1\},\ X_R = \{x\in X \mid x>V_k\},$$
$$X_M = X \setminus (X_L\cup X_R) = \{x\in X \mid V_1\leq x\leq V_k\}.$$

Введем в рассмотрение функционал

$$I(\mu,V) = \sum_{x\in X_L\cup X_R}\sum_{i=1}^{k}\mu_i(x)\|x-V_i\|^2 + \\ +\sum_{x\in X_M}\sum_{i=1}^{k}\mu_i^2(x)\|x-V_i\|^2. \qquad (3)$$

Запись функционала $I(\mu, V)$ в виде суммы двух выражений (3) показывает, что вклад в функционал $I(\mu, V)$ каждой из «внешних» точек $x\in X_L\cup X_R$ определяется так же, как и в функционале «жесткого» алгоритма ISODATA, а вклад каждой из внутренних точек $x\in \boldsymbol{X}_m$ определяется в соответствии с функционалом размытого алгоритма Fuzzy ISODATA.

Как показано ниже, зависимость $\mu(x)$, представленная на рисунке 1, может быть получена в результате решения задачи минимизации функционала как на множестве векторов $V = (V_1,\ldots V_k)$, $V_i \in R^1$, $i=1 \div k$, так и на множестве функций принадлежности $\mu_i(x)$, удовлетворяющих (2).

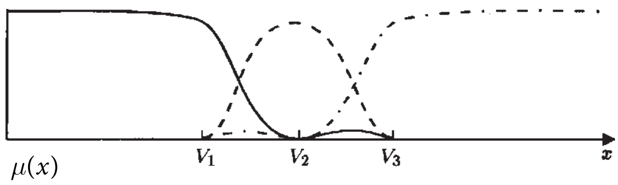

Рис. 1. Графическая интерпретация функций принадлежности, получаемых при использовании алгоритма АРКОС

Теорема[10]. *Пусть* $\mu_i^*(x)$ *и* $V^* = V_1^*,\ldots,V_k^*$, $i = 1 \div k$, $x\in X$, *доставляют минимум функционалу* (3). *Тогда выполняются следующие условия:*

I. *Для любого* $x\in X$ *справедливо:*

1. *Если* $x\in X_L$, *то* $\mu_i^*(x)=\begin{cases}1 \text{ при } i=1,\\ 0 \text{ при } i\neq 1.\end{cases}$ (4)

2. *Если* $x\in X_R$, *то* $\mu_i^*(x)=\begin{cases}1 \text{ при } i=k,\\ 0 \text{ при } i\neq k.\end{cases}$ (5)

3. Если $x\in X_M$, то

$$\mu_i^*(x)=\begin{cases}1 \text{ при } x=V_i,\ i=l,\\ 0 \text{ при } x=V_i,\ i\neq l,\ 1\leq l\leq k.\end{cases} \qquad (6)$$

$$\mu_i^*(x)=\frac{1/\|x-V_i\|^2}{\sum_{j=1}^{k}(1/\|x-V_j\|^2)} \text{ при } x\notin V_i,\ i = 1 \div k. \qquad (7)$$

II. *Для любого* $i = 1 \div k$ *существует такое* $x \in X$, *что* $\mu_i^*(x) \neq 0$.

III. *Компоненты оптимального вектора* V^* *удовлетворяют соотношению*

$$V_i^* = \frac{\sum_{x \in X} (\mu_i^*(x))^2 x}{\sum_{x \in X} (\mu_i^*(x))^2}. \tag{8}$$

Данная теорема устанавливает необходимые условия минимума функционала $I(\mu, x)$ и позволяет сформулировать следующий алгоритм построения размытой классификации.

Алгоритм АРКОС

1. Положим $t = 0$. Зададим $\mu_i^0(x_j)$, удовлетворяющие (2), $x_j \in X$; $i = 1 \div k$, $j = 1 \div n$. Зададим величину $\varepsilon > 0$, определяющую точность вычислений.
2. Вычисляем k взвешенных средних (центров классов) V_i^0, $i = 1 \div k$ по формуле (8).
3. Увеличиваем t на 1.
4. Вычисляем по формулам (4) — (7) значения

$$\mu_i^t(x_j),\ x_j \in X,\ j = 1 \div n,\ i = 1 \div k.$$

5. Вычисляем значения V_i^t, $i = 1 \div k$ по формуле (8).
6. Вычисляем $\delta = \max_{1 \le i \le k, 1 \le j \le n} \left\{ \Delta\mu_i^t(x_j) \right\}$, где $\Delta\mu_i^t(x_j) = \left| \mu_i^t(x_j) - \mu_i^{t-1}(x_j) \right|$.
7. Если $\delta \le \varepsilon$, то работа алгоритма заканчивается, иначе переходим к операции 3.

В численных (модельных) компьютерных экспериментах монотонно убывающие последовательности $\left\{ J(\mu^{t+1}, V^{t+1}) \right\}$, t = 0,1,..., сходились к решению задачи минимизации функционала (3).

АЛГОРИТМ ПОСТРОЕНИЯ ЛИНЕЙНОЙ РАЗДЕЛЯЮЩЕЙ ФУНКЦИИ С ИСПОЛЬЗОВАНИЕМ РАЗМЫТЫХ РАЗБИЕНИЙ

Предлагаемый алгоритм построения ЛРФ с использованием размытых разбиений основан на итерационной процедуре. Вначале при фиксированных значениях параметров α_i, $i = 1 \div m$ вычисляются (при помощи алгоритма АРКОС) субоптимальные значения функций принадлежности $\mu_i(x)$, $i = 1 \div k$, $x \in X$, и суммарная ошибка распознавания Ф. Затем производится варьирование α_i, $i = 1 \div m$, и поиск среди них таких значений α_i, $i = 1 \div m$, которые приводят к уменьшению Ф, и т.д. При этом на каждом шаге итерации для каждого фиксированного вектора α также вычисляются (по алгоритму АРКОС) соответствующие этому вектору $\mu_l(x)$, $l = 1 \div k$, $x \in X$, и Ф. Алгоритм осуществляет поиск ЛРФ, минимизирующей Ф. Суммарная ошибка распознавания Ф оценивает степень сходства размытого разбиения $\tilde{P}_1, \ldots, \tilde{P}_k$, построенного с помощью алгоритма АРКОС, с априорным разбиением $P_1, \ldots, P_k$ множества ***X***, которое задается характеристической функцией $\omega_l(x)$, $l = 1 \div k$, $x \in X$:

$$\begin{aligned} \Phi(\mu, \omega) &= \sum_{l=1}^{k} \sum_{x \in P_l} (1 - \mu_l(x)) = \\ &= \sum_{x \in X} \sum_{l=1}^{k} (1 - \mu_l(x)) \varpi_l(x) = \sum_{x \in X} \Delta\mu(x), \end{aligned} \tag{9}$$

где $\Delta\mu(x) = \sum_{l=1}^{k} (1 - \mu_l(x)) \omega_l(x)$ — ошибка распознавания объекта $x \in X$.

Минимизация Ф производится с помощью поискового алгоритма. Если в результате очередного шага варьирования параметров α_i, $i = 1 \div m$, значение Ф уменьшается, соответствующие значения α_i, $i = 1 \div m$, запоминаются. Для варьирования вектора используется метод локальных вариаций (МЛВ)[11].

Перейдем к описанию алгоритма минимизации суммарной ошибки распознавания (АМСОР), каждая итерация которого включает выполнение алгоритма размытой классификации АРКОС.

Алгоритм АМСОР

1. Положим $t = 0$. Зададим произвольные начальные значения параметров α_i^o, $i = 1 \div m$ и величину ε^*, определяющую точность вычислений. Положим $\Phi^0 = n$.
2. Увеличиваем t на 1. Вычисляем по формуле (1) значения ЛРФ: $F^t(x_j)$, $j = 1 \div n$. Положим $z_j^t = F^t(z_j)$, $j = 1 \div n$.
3. Вычисляем в соответствии с алгоритмом АРКОС значения функций принадлежности $\mu_l^t(z_j)$, $l = 1 \div k$, $j = 1 \div n$.
4. Определяем по формуле (9) суммарную ошибку распознавания Φ^t.
5. Вычисляем $\Delta\Phi^t = \Phi^t - \Phi^{t-1}$. Если $\Delta\Phi^t \ge 0$, то переходим к операции 6. Если же $\Delta\Phi^t < 0$, то фиксируем Φ^t, α_i^t, $i = 1 \div m$. Если $\Delta\Phi^t \ge 0$ и $\left| \Delta\Phi^t \right| < \varepsilon^*$, то работа алгоритма заканчивается; иначе переходим к операции 6.
6. Вычисляем новые значения параметров α_i^t, $i = 1 \div m$ с помощью МЛВ. Переходим к операции 2.

Сходимость предложенного алгоритма АМСОР определяется свойствами сходимости алгоритмов МЛВ и АРКОС.

Полученные с помощью алгоритма АМСОР значения параметров α_i^*, $i = 1 \div m$ и центров классов V_l^*, $l = 1 \div k$, могут использоваться далее для распознавания объектов, не входящих в обучающую выборку ***X***. С этой целью для новой точки $y \notin X$ вычисляется по формуле (1) значение $F(\alpha^*, y)$ ЛРФ, а затем с помощью формул (4) — (7) определяются веса принадлежности $\mu_l(y)$, $l = 1 \div k$.

Отметим, что алгоритм не изменится, если характеристические функции $\varpi_l(x)$ будут заменены задаваемыми априорно функциями $\mu_l^A(x)$ принадлежности объектов к классам. Тогда формула (9) примет вид:

$$\Phi(\mu^A, \mu^B) = \sum_{l=1}^{k} \sum_{x \in P_l} \left| \mu_l^A(x) - \mu_l^B(x) \right|,$$

где $\mu_l^A(x)$ — априорно известные значения функций принадлежности; $\mu_l^B(x)$ — функции принадлежности, вычисляемые по алгоритму АРКОС.

О ДВУХ ТИПАХ СОЦИАЛЬНОЙ АГРАРНОЙ СТРУКТУРЫ ГУБЕРНИЙ ЕВРОПЕЙСКОЙ РОССИИ НА РУБЕЖЕ XIX–XX вв.

Одна из крупных задач при изучении экономической истории России конца XIX — начала XX в. связана с построением типологии аграрного развития ее регионов. Существенный аспект этой типологии определяется пространственным распределением социальных факторов этого развития. Попытки «жесткого» определения понятия «тип» в историко-типологическом исследовании нередко входят в противоречие с принципиальной «размытостью» границ между типами. Это обстоятельство и приводит к использованию концепций теории нечетких множеств в типологических задачах экономической истории.

Построение многомерной социальной аграрной типологии губерний Европейской России в работах И. Д. Ковальченко и Л. И. Бородкина[12] проводилось на основе 8 признаков:

1) доля наемных сельскохозяйственных рабочих по отношению к местным работникам;
2) число сельскохозяйственных рабочих в расчете на десятину посева;
3) доля дворянской земли в удобной земле, %;
4) отношение проданных частновладельческих земель к общей их площади;
5) отношение арендованной крестьянами земли к надельной земле;
6) доля безлошадных и однолошадных в общем числе дворов;
7) доля дворов с 4 и более лошадьми в общем числе дворов;
8) поденная плата сельскохозяйственным рабочим в уборку урожая, коп.

Таблица 1 содержит значения указанных восьми признаков для всех 50 губерний Европейской России на рубеже XIX–XX вв. (номер столбца соответствует номеру признака[13]).

Кластер-анализ данных о 50 губерниях в пространстве 8 указанных признаков выявил 3 макротипа социальной структуры аграрного развития губерний Европейской России на рубеже XIX–XX вв. (названных И. Д. Ковальченко крестьянским, помещичьим и помещичье-крестьянским); каждый из них состоял из двух типов (групп губерний): 1 — помещичий прибалтийский тип; 2 — помещичий западный тип; 3 — крестьянский степной тип; 4 — крестьянский северо-восточный тип; 5 — помещичье-крестьянский промышленный тип; 6 — помещичье-крестьянский земледельческий тип.

Общий характер соотношения выявленных типов социальной аграрной структуры показывает, что существует промежуточное по всем аспектам положение двух типов (промышленного и земледельческого) социальной аграрной структуры. Полюсы представлены, с одной стороны, прибалтийским и степным, а с другой стороны, западным и северо-восточным типами. Однако эти типы представляли полюса далеко не по одним и тем же аспектам.

Анализ, проведенный И. Д. Ковальченко[14], показал, что наиболее существенные различия в социальной аграрной структуре губерний Европейской России связаны с двумя путями аграрной эволюции — *помещичьим («юнкерским») и крестьянским («фермерским»)*. Эти пути (варианты развития) могли выступать как в чистом виде, так и переплетаясь в тех или иных комбинациях. Выявленные типы социальной аграрной структуры Европейской России на рубеже XIX–XX вв. представляют различные вариации помещичьего и крестьянского путей аграрной эволюции. Прибалтийский и западный типы были разновидностями аграрной эволюции, в которой доминировало помещичье хозяйство. Степной и северо-восточный типы выражают аграрную эволюцию, в которой господствовало крестьянское хозяйство.

Таблица 1

Показатели социальной аграрной структуры губерний Европейской России на рубеже XIX–XX вв.

Губерния	Показатели							
	1	2	3	4	5	6	7	8
Архангельская	0,033	0,000	0,334	0,053	0,09	0,770	0,026	59
Астраханская	0,135	0,038	0,334	0,295	0,09	0,675	0,079	47
Бессарабская	0,053	0,224	0,287	0,127	0,01	0,505	0,079	44
Виленская	0,073	0,391	0,319	0,025	0,05	0,718	0,023	30

Окончание таблицы 1

Губерния	Показатели							
	1	2	3	4	5	6	7	8
Витебская	0,056	0,319	0,463	0,047	0,04	0,629	0,031	38
Владимирская	0,024	0,108	0,293	0,111	0,02	0,825	0,013	58
Вологодская	0,017	0,007	0,424	0,021	0,02	0,711	0,015	39
Волынская	0,030	0,358	0,256	0,043	0,02	0,561	0,046	36
Воронежская	0,014	0,177	0,293	0,196	0,01	0,582	0,109	51
Вятская	0,018	0,024	0,111	0,022	0,01	0,554	0,087	34
Гродненская	0,055	0,406	0,119	0,031	0,04	0,747	0,016	32
Донская	0,040	0,071	0,372	0,087	0,01	0,463	0,175	80
Екатеринославская	0,043	0,213	0,313	0,244	0,01	0,408	0,177	78
Казанская	0,010	0,083	0,258	0,052	0,00	0,692	0,033	34
Калужская	0,023	0,159	0,275	0,109	0,02	0,581	0,063	53
Киевская	0,035	0,350	0,318	0,092	0,03	0,680	0,035	40
Ковенская	0,208	0,241	0,370	0,065	0,15	0,540	0,143	37
Костромская	0,029	0,134	0,356	0,092	0,02	0,794	0,010	49
Курляндская	0,298	0,384	0,334	0,128	0,24	0,334	0,392	50
Курская	0,021	0,217	0,296	0,209	0,01	0,482	0,149	52
Лифляндская	0,305	0,407	0,334	0,255	0,27	0,411	0,191	43
Минская	0,050	0,504	0,467	0,062	0,03	0,625	0,039	39
Могилевская	0,033	0,342	0,309	0,104	0,03	0,382	0,148	40
Московская	0,023	0,158	0,310	0,215	0,03	0,791	0,017	61
Новгородская	0,032	0,140	0,394	0,109	0,02	0,632	0,041	42
Нижегородская	0,024	0,151	0,316	0,102	0,02	0,769	0,034	43
Олонецкая	0,029	0,005	0,081	0,002	0,04	0,707	0,030	46
Оренбургская	0,045	0,037	0,595	0,035	0,02	0,359	0,340	51
Орловская	0,025	0,210	0,276	0,178	0,02	0,571	0,095	40
Пензенская	0,019	0,230	0,306	0,169	0,01	0,592	0,079	41
Пермская	0,028	0,235	0,324	0,047	0,02	0,541	0,114	43
Петербургская	0,068	0,267	0,446	0,149	0,06	0,757	0,023	51
Подольская	0,033	0,354	0,469	0,063	0,02	0,426	0,157	31
Полтавская	0,047	0,257	0,404	0,340	0,04	0,777	0,028	47
Псковская	0,027	0,168	0,425	0,163	0,02	0,611	0,041	40
Рязанская	0,020	0,186	0,270	0,196	0,01	0,688	0,055	54
Самарская	0,044	0,071	0,264	0,413	0,02	0,430	0,251	47
Саратовская	0,031	0,193	0,421	0,275	0,01	0,545	0,127	52
Симбирская	0,026	0,153	0,346	0,225	0,01	0,652	0,053	39
Смоленская	0,033	0,202	0,486	0,324	0,03	0,430	0,119	44
Таврическая	0,097	0,165	0,339	0,273	0,02	0,311	0,316	69
Тамбовская	0,020	0,195	0,292	0,196	0,01	0,579	0,114	46
Тверская	0,029	0,116	0,295	0,224	0,03	0,667	0,026	48
Тульская	0,024	0,295	0,250	0,196	0,01	0,552	0,080	55
Уфимская	0,027	0,130	0,401	0,067	0,01	0,541	0,171	40
Харьковская	0,021	0,180	0,443	0,083	0,01	0,704	0,037	52
Херсонская	0,052	0,202	0,338	0,576	0,01	0,414	0,183	64
Черниговская	0,041	0,207	0,337	0,155	0,03	0,495	0,134	40
Эстляндская	0,177	0,666	0,334	1,162	0,18	0,543	0,117	47
Ярославская	0,040	0,130	0,281	0,141	0,04	0,834	0,008	55

Результаты выявления социальной аграрной типологии губерний Европейской России на рубеже XIX–XX вв., полученные с помощью кластер-анализа, подводят к вопросу: с каким типом социальной аграрной структуры — помещичьим или крестьянским — имеют больше сходства губернии третьего, помещичье-крестьянского типа социальной аграрной структуры, в котором наиболее заметно переплетались два пути аграрной эволюции.

Один из возможных подходов к решению данного вопроса дают методы распознавания образов. Действительно, если включить в обучающую выборку несколько губерний, явно относящихся к помещичьему типу аграрной эволюции, и несколько губерний с выраженными чертами крестьянского типа, то с помощью методов распознавания образов можно определить, к какому из этих типов ближе каждая из рассматриваемых губерний.

Использование в данной задаче предложенного алгоритма АМСОР позволяет для каждого объекта (губернии) определить степень его принадлежности к классам. Так, если очередной объект «попадает» на середину дискриминантной оси (т. е. между центрами двух классов, определенных по обучающей выборке), то он относится к обоим классам с одинаковыми степенями (весами) принадлежности, равными 0,5. Чем ближе объект к центру класса на дискриминантной оси, тем ближе к 1 значение веса принадлежности его к данному классу.

Распознавание проводилось на основе тех же 8 показателей, которые использовались при многомерной кластеризации губерний Европейской России. В обучающую выборку были включены 4 губернии, в которых доминировал помещичий тип аграрной эволюции (Эстляндская, Курляндская, Лифляндская и Ковенская губернии), а также 6 губерний (Херсонская, Таврическая, Екатеринославская, Донская, Самарская и Оренбургская) с явным преобладанием крестьянского («фермерского») типа. Таким образом, в 8-мерном пространстве указанных показателей была построена дискриминантная ось вида (1), на которой 10 губерний из обучающей выборки четко разделились на два соответствующих класса. Затем программе предъявлялись поочередно все 50 объектов (губерний), заданных значениями 8 показателей. Подстановкой значений этих показателей в формулу (1) определялось значение дискриминантной функции для данного объекта, а затем на основе известного положения центров обоих классов на дискриминантной оси в соответствии с (7) — (8) вычислялись веса принадлежности каждого объекта к обоим классам.

Увеличение значения дискриминантной функции соответствует в данном случае возрастанию степени принадлежности к помещичьему типу, а уменьшение ее значений указывает на доминирование крестьянского типа аграрной эволюции (см. табл. 2). Координаты центров классов на дискриминантной оси, полученные по обучающей выборке, равны 5.80 (для помещичьего типа) и –46,05 (для крестьянского типа).

Для интерпретации результатов распознавания введем пороговое значение $\mu_0 = 0{,}50$ веса принадлежности. Будем считать объект принадлежащим к данному классу, если вес его принадлежности к этому классу больше порогового значения.

Как следует из таблицы 2, в соответствии с этим правилом в 22 губерниях из 50 доминировал помещичий, а в 27 губерниях преобладал крестьянский тип аграрной эволюции (Уфимская губерния характеризуется равными весами принадлежности к обоим типам: $\mu_1 = \mu_2 = 0{,}50$). В число 22 губерний первого типа вошли все 12 губерний прибалтийского и западного подтипов, объединенные по результатам кластер-анализа в один тип, названный помещичьим типом социальной аграрной структуры. Характерно, что из остальных 10 губерний с «доминированием помещичьего пути» почти все были отнесены по результатам кластер-анализа к земледельческому подтипу помещичье-крестьянского типа (исключение составляют лишь Петербургская и Полтавская губернии). Это Орловская, Пензенская, Бессарабская, Черниговская, Симбирская, Смоленская, Псковская и Пермская губернии. Показательно, что веса принадлежности этих 10 губерний к «помещичьему» типу аграрной эволюции оказались намного меньше, чем у 12 губерний двух указанных подтипов помещичьего типа социальной аграрной эволюции. Для 10 указанных губерний помещичье-крестьянского типа значения весов принадлежности к «юнкерскому пути» не превышают 0,65; в то же время для 12 губерний помещичьего типа значения этих весов не опускаются ниже 0,77 (см. табл. 2).

В число 27 губерний с высоким весом принадлежности к крестьянскому типу аграрной эволюции вошли все 12 губерний степного и северо-восточного подтипов. Именно для этих 12 губерний характерны наиболее высокие веса принадлежности к «фермерскому» типу аграрной эволюции. Кроме того, в этот класс включены 9 губерний промышленного подтипа помещичье-крестьянского типа социальной аграрной структуры (т. е. все губернии, кроме двух) и оставшиеся 6 губерний земледельческого типа.

Полученные результаты дают некоторое (разумеется, приближенное) представление о соотношении потенциальных возможностей двух путей аграрной эволюции в Европейской России на рубеже XIX–XX вв. (с доминированием крестьянского типа). Важным представляется вывод о том, что алгоритмы распознавания образов, использующие концепции нечетких множеств, дают адекватный инструмент для построения проблемно-ориентированной классификации в исследованиях социально-экономических систем. Этот вывод подтверждается также отмеченной согласованностью результатов, полученных с помощью традиционных методов многомерной классификации и новых, более гибких алгоритмов распознавания образов.

Таблица 2

Соотношение двух путей аграрной эволюции в России на рубеже XIX–XX вв.: результаты дискриминантного анализа данных по 8 показателям социальной аграрной структуры

Губернии	Вес принадлежности к типу		Значение дискриминантной функции $F(x)$
	крестьянскому	помещичьему	
Архангельская	1,00	0,00	–53,46
Астраханская	0,84	0,16	–37,55
Бессарабская	0,42	0,58	16,21
Виленская	0,00	1,00	15,48
Витебская	0,06	0,94	2,47
Владимирская	0,92	0,08	–42,04
Вологодская	0,71	0,29	–31,24
Волынская	0,01	0,99	5,05
Воронежская	0,65	0,35	–27,94
Вятская	0,68	0,32	–29,69
Гродненская	0,00	1,00	11,71
Донская	1,00	0,00	–66,53
Екатеринославская	1,00	0,00	–50,91
Казанская	0,52	0,48	–21,19
Калужская	0,73	0,27	–32,09
Киевская	0,15	0,85	–0,80
Ковенская	0,23	0,77	–6,09
Костромская	0,68	0,32	–29,32
Курляндская	0,21	0,79	–4,99
Курская	0,59	0,41	–24,76
Лифляндская	0,03	0,97	4,38
Минская	0,00	1,00	20,55
Могилевская	0,11	0,89	0,28
Московская	0,88	0,12	–39,61
Новгородская	0,52	0,48	–21,07
Нижегородская	0,54	0,46	–22,23
Олонецкая	0,96	0,04	–44,14
Оренбургская	0,83	0,17	–37,32
Орловская	0,38	0,62	–13,83
Пензенская	0,35	0,65	–12,28
Пермская	0,37	0,63	–13,47
Петербургская	0,42	0,58	–16,15
Подольская	0,00	1,00	13,17
Полтавская	0,38	0,62	–13,88
Псковская	0,41	0,59	–15,68
Рязанская	0,70	0,30	–30,40
Самарская	0,79	0,21	–35,32
Саратовская	0,60	0,40	–25,17

Окончание таблицы 2

Губернии	Вес принадлежности к типу		Значение дискриминантной функции $F(x)$
	крестьянскому	помещичьему	
Симбирская	0,44	0,56	–17,02
Смоленская	0,40	0,60	–15,17
Таврическая	1,00	0,00	–46,41
Тамбовская	0,52	0,48	–21,11
Тверская	0,71	0,29	–31,18
Тульская	0,51	0,49	–20,53
Уфимская	0,50	0,50	–19,98
Харьковская	0,62	0,38	–26,15
Херсонская	0,84	0,16	–37,63
Черниговская	0,37	0,63	–13,13
Эстляндская	0,00	1,00	27,00
Ярославская	0,82	0,18	–36,97

ПРИМЕЧАНИЯ

1 Заде Л. А. Размытые множества и их применение в распознавании образов и кластер-анализе // Классификация и кластер. М., 1980. С. 208.

2 Бородкин Л. И. Квантитативная история в системе координат модернизма и постмодернизма // Новая и новейшая история. 1998. № 5.

3 Ковальченко И. Д., Бородкин Л. И. Аграрная типология губерний Европейской России на рубеже XIX–XX вв. (Опыт многомерного анализа) // История СССР. 1979. № 1; Borodkin L. I. and Koval'chenko I. D. Two Paths of Agrarian Evolution in European Russia: An Essay in Multivariate Analysis. In: Russian Review. V. 47. 1988, № 4; Borodkin L. Defining agricultural regions in Russia: fuzziness in multivariate classification of historical data. // History and Computing. 2009. № 11 (1–2).

4 Бородкин Л. И. Многомерный статистический анализ в исторических исследованиях. М., 1986; Бородкин Л. И., Гарскова И. М. Программное обеспечение FUZZYCLASS в историко-типологическом исследовании // История и компьютер: новые информационные технологии в исторических исследованиях и образовании. Scripta Mercaturae Verlag, St.Katharinen, 1993.

5 См., напр.: Заде Л. А. Указ. соч.; Bezdek J. C. Pattern recognition with fuzzy objective function algorithms. N. Y; — London, 1981; Dunn J. C. A fuzzy relative of the ISODATA process and its use in detecting compact well separated clusters. In: J. Cybernetics, 1974. V. 3. № 3; Bock H. G. Clusteranalyse mit unscharten Partitionen. — Klassifikation und Erkenutnis, III. Fuchtagung der Gesellschaft für Kiassifikation. Konigstein, 1979.

6 См., напр.: Jasiewicz Ja., Hildebrandt-Radke I. Using multivariate statistics and fuzzy logic system to analyse settlement preferences in lowland areas of the temperate zone: an example from the Polish Lowlands. // Journal of Archaeological Science. 2009. № 36.

7 Бородкин Л. И., Стадник О. Е. Алгоритм построения решающего правила в задаче распознавания образов с использованием размытых множеств // Автоматика и телемеханика. 1985. № 11.

8 Читатель, понявший содержательный смысл применения алгоритмов распознавания с использованием ТНМ и испытывающий затруднения при чтении математизированных разделов 2–4, может перейти непосредственно к разделу 5, в котором дается интерпретация полученных результатов историко-типологического исследования.

9 Dunn J. C. Op. cit.

10 Бородкин Л. И., Стадник О. Е. Указ. соч.

11 Поляк Э. Численные методы оптимизации. М., 1974.

12 Ковальченко И. Д., Бородкин Л. И. Указ. соч.; Borodkin L. I. and Koval'chenko I. D. Op. cit.

13 См. приложение к статье: Ковальченко И. Д., Бородкин Л. И. Указ. соч.

14 Ковальченко И. Д. Методы исторического исследования. М., 1987.