

МЕТОДОЛОГИЯ ЭКСПЕРТНО-КЛАССИФИКАЦИОННОГО АНАЛИЗА ДАННЫХ В ЗАДАЧАХ АНАЛИЗА РАЗВИТИЯ РЕГИОНАЛЬНЫХ СИСТЕМ¹

© Дорофеев А.А., Дорофеев Ю.А., Покровская И.В.

ИНСТИТУТ ПРОБЛЕМ УПРАВЛЕНИЯ РАН

Abstract. The expert-ranging methods of regional socio-economic systems functioning analysis are discussed. As an example the utilization of especially developed technique for constituent entity RF social development comparative evaluation is described.

ВВЕДЕНИЕ

Анализ функционирования крупномасштабных региональных систем, каждый из которых характеризуется достаточно большим набором показателей является достаточно сложной задачей даже при использовании современных информационно-компьютерных технологий. Эти сложности связаны в первую очередь с большой размерностью задачи: число объектов колеблется от нескольких десятков до нескольких тысяч, число показателей – в пределах нескольких десятков, число моментов времени (при анализе динамики развития) – может достигать многих десятков (при анализе помесечной динамики).

До сих пор во многих экономических исследованиях для анализа и прогнозирования развития региональных систем используются стандартные статистические методы. В большинстве из них исследуемое множество объектов рассматривается как выборка из некоторой генеральной совокупности. Тогда задача заключается в том, чтобы оценить статистические свойства всей генеральной совокупности по статистическим характеристикам и свойствам этой выборки. Однако, для многих прикладных задач вероятностная интерпретация результатов невозможна, да и сама задача не укладывается в рамки классической математической статистики. По этой причине за последние 20-30 лет появилось достаточно много работ, в которых статистический подход в его классическом виде претерпел существенные изменения (см., например, [1]). Примером такой задачи как раз и является рассматриваемая в работе задача анализа социального развития регионов Российской Федерации. Здесь статистическое оценивание играет вспомогательную роль, а главной задачей становится построение «сжатого описания» исходных данных (например, рейтингов регионов), которое можно было бы использовать для принятия качественных управленческих решений, например, распределения финансовой помощи регионам.

¹Работа выполнена при частичной финансовой поддержке РФФИ, проекты 08-07-00349-а, 08-07-00427-а.

1. МЕТОДЫ ЭКСПЕРТНО-КЛАССИФИКАЦИОННОГО АНАЛИЗА В ЗАДАЧЕ АНАЛИЗА РАЗВИТИЯ РЕГИОНОВ

Для получения полного представления о функционировании крупномасштабной региональной системе (в данном случае – об уровне социального развития регионов) предлагается использовать методы классификационного анализа сложноорганизованных данных [2]. Была предложена концепция применения методов классификационного анализа для решения задач анализа развития региональных систем управления. Основная цель этой концепции состоит в следующем:

- структуризация исходного набора показателей с целью выявления относительно небольшого числа информативных параметров;
- структуризация исходного множества объектов, для чего необходимо выделить в пространстве выбранных информативных параметров области близко расположенных друг к другу объектов;
- анализ динамических свойств системы, в том числе выделение характерных траекторий изменения в пространстве параметров положения объектов во времени (типология траекторий), выявление зависимостей между параметрами с учетом временного сдвига и т.д.

Структуризация исходного набора параметров. Опыт использования алгоритмов классификационного анализа показывает, что классификация по всем исходным параметрам далеко не всегда приводит к желаемым результатам, например, из-за наличия помех и ошибок, негативного влияния малоинформативных и шумящих параметров. Поэтому классификацию обычно проводят в пространстве так называемых информативных параметров, которое имеет значительно меньшую размерность, чем исходное.

Структуризацию параметров предлагается проводить методами экстремальной группировки [1], при этом необходимо выбрать число групп, а также тип группировки – с фоновой группой или без неё, в зависимости от уровня «зашумлённости» параметров. Для этой цели используются специальные экспертные процедуры. Результатом экстремальной группировки являются группы параметров и факторы – интегральные параметры-характеристики групп, каждый из которых является линейной комбинацией параметров соответствующей группы и, в определённом смысле, её центром. На базе результатов экстремальной группировки выбираются информативные показатели для исследуемой системы. В качестве таковых выбираются либо сами факторы (интегральные показатели), либо исходные параметры, ближайшие к этим факторам. Основное условие – они должны быть легко интерпретируемы.

Для удобства использования интегральных показателей по каждому из них обычно делается одномерная классификация объектов. Благодаря этому интегральный показатель легко преобразовать в параметр, измеряемый в качественной шкале, например, его значения можно характеризовать в таких терминах, как «низкие», «средние» и «высокие». Для этой цели в работе используется алгоритм одномерной m -локальной оптимизации, разработанный в [3].

Структуризация исходного множества объектов. Выявление структуры объектов производится в пространстве информативных параметров X . Для этой

цели используются алгоритмы автоматической классификации [3], применение которых подразумевает выбор:

- вида критерия качества;
- классификации с фоновым классом или без него, т.е. отбрасываются ли «далёкие» (шумящие) объекты;
- типа размытости – четкая, размытая, с размытыми границами, четкая с размытым фоном, размытая с четким фоном и т.д.

Результатом классификации является вектор функций принадлежности объектов к классам $H(x) = (h_1(x), \dots, h_r(x))$, $x \in X$ и описание самих классов (например, эталонов) [2]. Для того чтобы результаты классификации можно было использовать в практических задачах, важно не только насколько экономно она представляет исходную информацию, но и насколько эта классификация удобна для интерпретации в содержательных терминах. В этой связи в приложениях часто используются алгоритмы построения так называемых «хорошо интерпретируемых классификаций» [4].

Выявление динамических свойств исследуемой системы. Для этой цели в работе использовался алгоритм структурного прогнозирования развития сложных динамических объектов, разработанный в [5]. Основная идея предложенного метода решения этой задачи состоит в том, что требуется прогнозировать не точные значения параметров, описывающих состояние объекта, а лишь класс (тип) объекта в рамках некоторой структуры изучаемого множества объектов.

Предполагается, что вектор значений параметров $x_j(t) = (x_j^{(1)}(t), x_j^{(2)}(t), \dots, x_j^{(k)}(t))$ достаточно полно характеризует состояние j -го объекта в момент времени t . А это, в свою очередь означает, что взаиморасположение множества точек $x_1(t), \dots, x_n(t)$ в k -мерном пространстве признаков X отражает реальную структуру (типологию) исследуемого множества объектов. Для выявления этой структуры в работе используется комплексный алгоритм автоматической классификации [3]. С этой целью в момент времени t_1 производится автоматическая классификация (кластеризация) n точек в пространстве X на небольшое (3-5) число классов r , каждый из которых и характеризует тип объекта (в пределах изучаемого множества). Вводится понятие модели (эталона) класса $a_i(t)$, $i = 1, \dots, r$ (чаще всего – это центр класса) [2]. Для каждой точки кроме принадлежности к классу вычисляются расстояния до эталонов всех классов $R_{ij}(t)$, $i = 1, \dots, r$, $j = 1, \dots, n$.

Вопрос содержательной интерпретации полученных классов (типов) решается с помощью специальных экспертных процедур.

В момент времени t_2 и во все последующие моменты времени, для которых известны значения соответствующих параметров, каждая точка относится к тому или иному классу с помощью одного из алгоритмов распознавания образов с учителем. В работе для этой цели использовался алгоритм метода потенциальных функций, который в спрямляющем пространстве эквивалентен алгоритму ближайшего среднего [1]. На каждом шаге, после того, как определена принадлежность всех точек к тому или иному классу, производится пересчет эталонов и для каждой точки с предыдущего шага пересчитываются, а для каждой новой точки вычисляются расстояния до новых эталонов. Такая процедура выполняется для всех m моментов времени. В итоге

для каждого объекта получается последовательность (траектория) из n позиций. В каждой позиции находится $r+1$ число, первое из которых – это номер класса, к которому относился этот объект в соответствующий момент времени, а последующие числа – это значения расстояний до центров классов в тот же момент времени. Требуется спрогнозировать номер класса (тип объекта), к которому будет относиться каждый объект в момент времени t_{m+1} .

В качестве прогнозной модели для каждого объекта в этом алгоритме используется марковская цепь с r состояниями и матрицей переходных вероятностей $P_j = \|p_{ji}\|$. С помощью специального алгоритма на каждом шаге для каждого периода времени (года) производится пересчет соответствующих переходных вероятностей p_{ji} с использованием информации о значениях расстояний от каждого объекта до центров классов [5]. Этот алгоритм позволяет учесть особенности всей прошлой траектории изменения положения каждого объекта в пространстве информативных параметров.

2. СРАВНИТЕЛЬНЫЙ АНАЛИЗ СОЦИАЛЬНОГО РАЗВИТИЯ СУБЪЕКТОВ РФ

Разработанная методика использовалась для сравнительного анализа социального развития субъектов Российской Федерации (47 показателей для 79 регионов за 3 года). Применение алгоритма экстремальной группировки позволило разбить 47 исходных показателей на 6 групп: доходы населения (13 показателей); расходы и сбережения (14 показателей); потребление продуктов питания (7 показателей); демографические характеристики (4 показателя); характеристики социальной напряженности (6 показателей); объем финансовой помощи из межрегиональных фондов (2 показателя). Для последующей классификации и формирования рейтинга регионов было отобрано шесть показателей, ближайших к факторам групп: среднедушевой доход, доля оплаты труда в среднедушевом доходе, превышение доходов над расходами, число пенсионеров на 1000 чел. населения, уровень безработицы, общий объем финансовой помощи (ОФП) на душу населения.

В результате классификационного анализа в шестимерном пространстве этих показателей было получено 7 классов регионов (с использованием процедуры экспертной коррекции). В зависимости от целей исследования классифицируются либо все регионы за все три года (каждый регион в каждом году рассматривается как самостоятельный объект), либо все регионы отдельно для каждого года. Классификация первого типа позволяет анализировать динамику развития регионов, прослеживая на протяжении исследуемого периода переходы регионов из одного класса в другой, или, что то же самое, изменение рейтинга региона. При составлении текущего рейтинга удобнее использовать классификацию второго типа, поскольку в этом случае результаты получаются более обозримыми. На базе этой классификации получен рейтинг социального развития регионов для каждого из 3 лет, проанализирована динамика рейтингов регионов по основным показателям, представленная в табл. 1.

Для укрупненной оценки социально-экономической ситуации в регионе наибольший интерес представляет пара показателей: «Среднедушевой доход» и «Уровень безработицы». По этой паре показателей по данным за последний из трёх анализируемых лет был построен рейтинг регионов.

Таблица 1. Динамика рейтингов регионов по основным показателям

Показатели	Изменение рейтинга за 2 года (количество регионов)			
	+2	+1	0	- 1
Среднедушевой доход	2	48	29	-
Доля оплаты труда в среднедушевом доходе	6	37	33	3
Превышение доходов над расходами	-	6	55	18
Число пенсионеров на 1000 чел. населения	-	9	70	-
Уровень безработицы	8	44	24	3
Общий объем финансовой помощи	-	31	48	2

Для этого с помощью алгоритма хорошо интерпретируемых классификаций [4] было построено 16 классов.

При построении рейтингов по двум и более показателям одновременно возникает проблема многокритериальности: как упорядочить два объекта, один из которых имеет более высокий рейтинг по одному показателю, а второй – по другому показателю. В общем случае в таких ситуациях приходится прибегать к мнению экспертов, оценивающих сравнительную важность приращений значений показателей. В данном случае с помощью экспертной процедуры было сформулировано следующее простое правило упорядочения в конфликтных ситуациях: при низких значениях уровня безработицы упорядочение производится по среднедушевому доходу; при низких значениях среднедушевого дохода упорядочение ведется по уровню безработицы. Тогда классы регионов упорядочиваются так, как показано в табл. 2.

Кроме того, каждый рейтинговый класс получил интегральную характеристику по всем основным показателям.

На этом же материале использовался алгоритм структурного прогнозирования социального развития субъектов Российской Федерации, разработанный в [5]. Как уже говорилось выше, в качестве прогнозной модели для каждого объекта в этом алгоритме используется марковская цепь с r состояниями и матрицей переходных вероятностей $P_j = \|p_{ji}\|$. Результаты использования полученной прогнозной модели, не смотря на малый период наблюдений (всего три года), был оценён экспертами Минздравсоцразвития как очень хорошие. Эти результаты продемонстрировали высокую эффективность использованных алгоритмов при выборе информативных показателей, «оптимального» числа классов, а также прогнозной модели.

ЗАКЛЮЧЕНИЕ

Описана методика сравнительного анализа развития региональных социально-экономических системы, которые характеризуются достаточно большим набором

Таблица 2. Рейтинг регионов по показателям «Среднедушевой доход» и «Уровень безработицы»

Р	П1	П2	Регионы
1	1	1	Москва
2	2	1	Чукотский АО
3	2	2	Тюменская область
4	3	2	Респ. Коми, Респ. Саха (Якутия), Магаданская обл., Сахалинская обл.
5	3	3	Мурманская область, Камчатская область
6	4	1	Московская обл., Ярославская обл., Вологодская обл., Санкт-Петербург, Респ. Татарстан, Самарская обл.
7	4	2	Челябинская обл., Респ. Карелия, Архангельская обл., Волгоградская обл., Респ. Башкортостан, Нижегородская обл., Пермская обл., Свердловская обл., Респ. Хакасия, Красноярский край, Кемеровская обл., Омская обл., Приморский край, Хабаровский край, Еврейская автономная обл.
8	4	3	Астраханская обл., Ростовская обл., Иркутская обл., Томская обл.
9	5	1	Костромская обл., Липецкая обл., Тверская обл., Тульская обл., Новгородская обл., Ульяновская обл.
10	5	2	Белгородская обл., Брянская обл., Воронежская обл., Калужская обл., Курская обл., Орловская обл., Рязанская обл., Тамбовская обл., Калининградская обл., Ленинградская обл., Псковская обл., Краснодарский край, Ставропольский край, Респ. Мордовия, Удмуртская Респ., Чувашская Респ., Кировская обл., Пензенская обл., Саратовская обл., Алтайский край
11	6	2	Ивановская область
12	5	3	Владимирская обл., Смоленская обл., Респ. Адыгея, Карачаево-Черкесская Респ., Респ. Северная Осетия-Алания, Оренбургская обл., Курганская обл., Респ. Алтай, Новосибирская обл., Читинская обл., Амурская область
13	5	4	Республика Бурятия
14	5	3	Республика Марий-Эл
15	5	5	Кабардино-Балкарская Республика, Республика Тыва
16	6	5	Республика Дагестан, Республика Ингушетия, Республика Калмыкия

показателей. Основу методики составляют алгоритмы структуризации набора исходных параметров и множества исследуемых объектов. Рассмотрены результаты использования методики для решения задачи сравнительного анализа социального развития регионов России за 3 года.

СПИСОК ЛИТЕРАТУРЫ

1. Браверман Э.М., Мучник И.Б. Структурные методы обработки эмпирических данных. – М.: Наука, 1983. – 464 с.
2. Бауман Е.В., Дорофеев А.А. Классификационный анализ данных. // Труды Международной конференции по проблемам управления. Том 1. – М.: СИНТЕГ, 1999. – С. 62-77.
3. Дорофеев Ю.А. Комплексный алгоритм автоматической классификации и его применение для анализа и принятия решений в больших системах управления. / Теория активных систем. Труды международной научно-практической конференции. / –М.: ИПУ РАН. 2007. – С. 39-42.
4. Дорофеев А.А., Чернявский А.Л. Алгоритмы построения хорошо интерпретируемых классификаций. / Проблемы управления. №2, 2007. – С. 83-84.
5. Дорофеев Ю.А., Дорофеев А.А. Методы структурно-классификационного прогнозирования многомерных динамических объектов. Искусственный интеллект, № 2, 2006. С. 138-141

Статья поступила в редакцию 27.04.2008