

УДК 316.334

**Возможности кластерного анализа в обработке результатов  
социологических исследований (на примере опроса безработных граждан  
в Городском центре занятости населения г. Барнаула)**

**Мальцева А.В.**

Кандидат социологических наук,  
доцент кафедры математических методов в социальных науках,  
заведующая лабораторией математического обеспечения  
социальных и психологических исследований факультета социологии  
Алтайского государственного университета

*Современные методы и технологии обработки и анализа данных предоставляют социологу-исследователю огромные возможности для поиска ответов на интересующие его вопросы. Статья посвящена рассмотрению метода кластерного анализа и его возможностей для трактовки социальной информации.*

*Ключевые слова: варианты кластерного анализа, самоорганизующиеся карты признаков Кохонена (СКП), технология Data Mining, рынок труда, предметные ситуации.*

**Метод кластерного анализа и его современные варианты.** Кластерный анализ – один из распространенных методов решения задач, возникающих при анализе социальной и экономической информации [1, с. 89]. Этот вид анализа может применяться для описания структуры социально-территориальных групп, разбиения клиентов рынка на различные по уровню мотивации группы и т.д., т.е. для нахождения групп схожих объектов в выборке данных. «Кластеризация – это группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов. Объекты внутри кластера должны быть похожими друг на друга и отличаться от других, которые вошли в кластеры» [2, с. 43]. В результате решения задачи кластеризации выявляется скрытая структура данных и обнаруживаются неочевидные закономерности в данных. «Кластерный анализ – это способ группировки многомерных объектов, основанный на представлении результатов отдельных наблюдений точками подходящего геометрического пространства с последующим выделением групп как “сгустков” этих точек (кластеров, таксонов)» [3, с. 103; 2, с. 82-89].

Выделяется несколько подходов к решению задач кластерного анализа, которые основаны на различных представлениях о задаче, использовании специфичной для каждой предметной области

дополнительной информации и т.д. Среди них: вероятностный подход [4, с. 146, с. 186-188]; подход, использующий аналогию с центром тяжести [4, с. 220]; иерархический подход [4, с. 251]; подход, основанный на понятии ближайшего соседа [5, с. 251; 6, с. 257]; нечеткие алгоритмы кластерного анализа [4, с. 242]; эволюционный (генетический) подход [6, с. 421-430]; подход, использующий искусственные нейронные сети [7, с. 45, с. 77-79].

Для обработки данных проведенного исследования был применен метод, позволяющий проводить разделение на непересекающиеся классы с произвольным их числом – кластеризация методом Кохонена. В данном случае это имеет ряд преимуществ: возможно решение задач большой размерности, число кластеров может быть не задано; формируют наглядное двумерное отображение множества объектов и т.д. Главной причиной использования самоорганизующихся карт признаков (СКП) для исследовательского анализа данных и извлечения знаний является то, что данный метод не требует никаких априорных предположений о распределении данных, позволяет обнаруживать в наборах данных неизвестные ранее структуры или образы при помощи обучения без учителя. Он был предложен Тьюво Кохоненом в начале 80-х гг. и нашел широкое применение. Технология СКП представляет собой набор аналитических

процедур и алгоритмов, позволяющих преобразовать традиционное описание множества объектов, заданных в многомерном ( $n > 3$ ) пространстве признаков плоской базы данных, в двумерную карту, устроенную таким образом, что близким объектам в многомерном пространстве отвечают рядом стоящие точки (их образы) на карте [8, с. 23; 9, с. 143-144]. В результате трудно анализируемые в совокупности многомерные объекты получают простой и наглядный вид на двумерной карте, которая сохраняет их основные свойства – топологию и распределение в многомерном пространстве.

Применение технологии СКП дает ряд преимуществ: обнаружение групп объектов с одинаковыми характеристиками (далее – кластеров) по их локализованному расположению на специально создаваемой карте кластеров; проверка содержательного описания обнаруженных групп по специфическим особенностям, обнаруженным на карте признаков, а также на проекциях карты кластеров на каждый признак в отдельности; выявление неявных связей и закономерностей между признаками; проведение оценки объектов в динамике, оценка изменений как в целом по структуре кластеров, так и по отдельности; позиционирование на карту новых объектов для придания им статуса (рейтинга); прогнозирование значений одних признаков объектов через другие; фильтрация объектов за счет поисковых уникальных критериев, формируемых в терминах СКП.

Карты Кохонена позволяют также представить полученную информацию в простой и наглядной форме путем нанесения раскраски [2, с. 330-331], что дает возможность визуализировать результаты кластеризации, в том числе и многомерные. Карта Кохонена состоит из сегментов прямоугольной или шестиугольной формы, называемых ячейками. Каждая ячейка связана с определенным выходным нейроном и представляет собой «сферу влияния» данного нейрона. Распределение векторов весов нейронов карты происходит на основе конкурирующего обучения. Объекты, векторы признаков которых оказываются ближе к вектору весов данного нейрона, попадают в ячейку, связанную с ним. Тогда распределение объектов на карте в целом соответствует распределению векторов весов нейронов. Следовательно, если объекты на карте расположены близко друг к другу, то и векторы признаков этих объектов близки и наоборот, если ячейки с объектами находятся далеко друг от друга, то и векторы их признаков различаются значительно. Хотя расстояние между объектами уже позволяет сделать выводы о степени их сходства или различия, но важна информация о том, в чем проявляется это сходство и различие, по каким признакам они различаются в наибольшей степени, а по каким – в наименьшей и т.д.

Таким образом, есть два важных фактора: положение объекта на карте (расстояние до других объ-

ектов) и цвет ячейки (см. рис. 1). Очевидно, что при таком способе визуализации на одной карте можно использовать расцветку только по одному признаку, т.е. для значений нескольких признаков, следует строить отдельные карты по каждому из признаков (см. рис. 2).

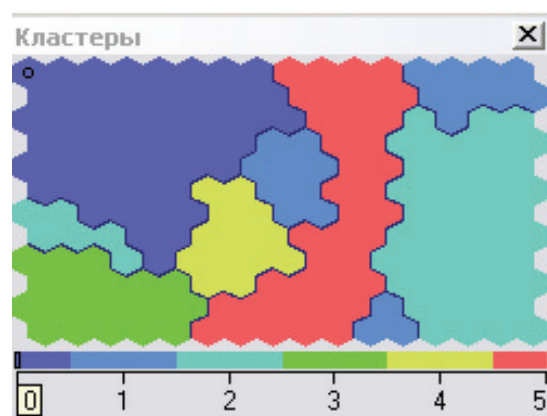


Рис. 1. «Количество кластеров»

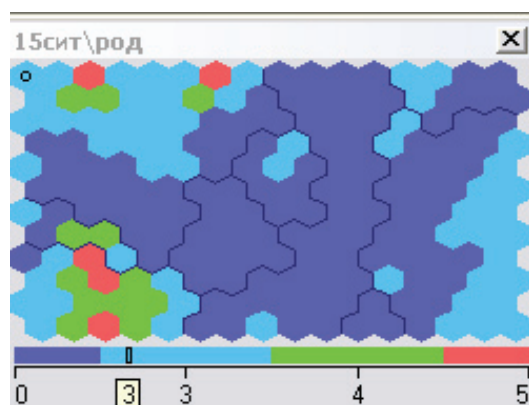


Рис. 2. Карта Кохонена, детализирующая характеристики кластеров группы безработных в ситуации взаимодействия с родственниками по вопросу получения или оказания помощи при ремонте машины

Для обработки результатов исследования была выбрана аналитическая платформа Deductor, позволяющая проводить анализ данных в парадигме «Data Mining». Термин «Data Mining» обозначает не столько конкретную технологию, сколько сам процесс поиска корреляций, тенденций, взаимосвязей и закономерностей посредством различных математических и статистических алгоритмов: кластеризации, создания субвыборок, регрессионного и корреляционного анализа. Цель этого поиска — представить данные в максимально четком виде, что является особенно актуальным для визуализации результатов кластерного анализа [6, с. 7; 10, с. 101; 9].

Большое удобство заключается еще и в возможности анализа полученных статистических выводов не только в виде СКП, но также и в виде профилей кластеров, содержащих все необходимые сведения о распределении признаков.

**Методика опроса и оцифровки.** Осенью 2009 г. был проведен опрос безработных, обращающихся в Городской центр занятости населения г. Барнаула. В литературе существует практика называть такую аудиторию участниками официальной части открытого рынка труда. Было опрошено 90 человек, с соблюдением пропорции по полу и в возрасте до пенсионного, приходивших в центр занятости в течение одного дня. Методом сбора информации выступило полуструктуризованное интервью, в ходе которого каждый респондент отвечал на вопросы о своем участии или неучастии в перечисленных предметных ситуациях. Список предметных ситуаций был составлен в ходе анализа услуг, предлагаемых в Барнауле различными организациями и частными лицами, и уточнен на предмет актуальности в ходе предваряющих исследование фокус-групп с безработными гражданами:

1. Отремонтировать сантехнику.
2. Помочь с ремонтом квартиры.
3. Отремонтировать автомобиль.
4. Научить водить машину.
5. Отремонтировать компьютер.
6. Перевезти или перенести груз.
7. Перевести текст на иностранный язык.
8. Помочь выбрать профессию (Вам или Вашему ребенку).
9. Профессионально проконсультировать по вопросам здоровья.
10. Составить налоговую декларацию, проконсультировать по вопросам налогообложения.
11. Проконсультировать по вопросам капитального ремонта дома.
12. Проконсультировать по юридическим вопросам.
13. Предоставить место для недельного проживания.
14. Пригласить Вас отдохнуть на даче.
15. Одолжить крупную сумму денег.
16. Помочь при переезде.
17. Организовать праздник.
18. Устроить ребенка в детский сад.
19. Устроить ребенка в школу или перевести в другую школу.
20. Посоветовать Вам или ребенку спорт, клуб, секцию, спортивную школу.
21. Помочь поступить в учебное заведение (Вам или Вашему ребенку).
22. Проконсультировать по вопросам внешности, одежды.
23. Обсудить взаимоотношения в семье.
24. Посидеть с детьми.
25. Помочь с уходом за больным
26. Обсудить вопросы взаимоотношений в семье.
27. Проконсультировать по обращению с домашними животными.
28. Найти работу.
29. Выбрать политическую позицию.

Было выдвинуто предположение, что степень участия безработных в предлагаемых предметных

ситуациях в отношениях с родственниками и друзьями может служить способом проекции их трудовых способностей на рынке труда. Поскольку далеко не секрет, что записанные в документах человека профессиональные навыки или вербализуемые им самим возможности могут не совпадать с реальными. По каждой ситуации респондент указывал на то, обращался ли он за помощью по данному вопросу (отдельно) к родственникам и друзьям, обращались ли к нему (отдельно) родственники и друзья, было ли совместное выполнение деятельности по данной ситуации.

Были разработаны следующие коды оцифровки результатов: «0» – по ситуации не было обращения или взаимодействия, «3» – сам безработный обращался за помощью, «4» – к безработному обращались за помощью, «5» – было совместное выполнение работ по данной предметной ситуации. В текстовом документе был создан файл данных, отражающих результаты ответов респондентов в контексте взаимодействия с родственниками и друзьями по отдельности. В дальнейшем этот файл был обработан в среде аналитической платформы «Deductor Studio Academic». Построение СКП проводилось с тремя итерациями, всякий раз количество кластеров было равно шести.

#### **Результаты кластерного анализа ответов участников официальной части открытого рынка труда.**

Условное название кластера 1 «Ремонт квартиры». Возраст респондентов от 20 до 59 лет. В кластере 1 значимыми являются следующие ситуации: устроить ребенка в школу или перевести в другую школу; устроить ребенка в детский сад; проконсультировать по обращению с домашними животными; проконсультировать по вопросам внешности, одежды; помочь с ремонтом квартиры; отремонтировать сантехнику; посоветовать Вам или ребенку спортклуб, секцию, спортшколу. Для данного кластера значимыми с точки зрения активности респондентов как участников одностороннего и взаимного действия по решению названной проблемы являются ситуации необходимости ремонта квартиры и сантехники. В этих ситуациях доминирующими участниками взаимодействия выступают родственники. Респонденты обращаются к родственникам за помощью, в то же время востребованы услуги респондентов в решении данного вопроса и встречаются ситуации взаимного обращения. Другие ситуации, вошедшие в кластер, отражают одностороннее обращение респондентов к друзьям, знакомым и родственникам и направлены на получение различных консультаций.

Условное название кластера 2 «Поиск работы». Возраст респондентов от 22 до 59 лет. В кластере 2 значимыми являются следующие ситуации: проконсультировать по вопросам внешности, одежды; организовать праздник; помочь поступить в учебное

заведение; устроить ребенка в школу; устроить ребенка в детский сад; найти работу; помочь выбрать профессию Вам или Вашему ребенку; составить налоговую декларацию; выбрать политическую позицию; перевести текст на иностранный язык; обсудить взаимоотношения в семье; отремонтировать автомобиль; проконсультировать по обращению с домашними животными; посоветовать Вам или Вашему ребенку спортклуб, секцию, спортшколу; проконсультировать по вопросам капитального ремонта дома; проконсультировать по юридическим вопросам. Для данного кластера с точки зрения односторонней активности респондентов значимыми являются ситуации выбора и приобретения профессии, поиска работы, устройства ребенка в детский сад, в школу. Преимущественно отмечается одностороннее обращение респондентов к родственникам. Не исключены случаи обращения к респондентам за консультациями по вопросам, требующим профессиональных знаний.

Условное название кластера 3 «Выбор и приобретение профессии». В данном кластере объединяются ситуации: помочь поступить в учебное заведение Вам или Вашему ребенку; помочь выбрать профессию Вам или Вашему ребенку; проконсультировать по вопросам внешности, одежды; помочь при переезде. Доминирующими участниками взаимодействия в этих ситуациях выступают друзья, знакомые. Отмечается одностороннее обращение как респондентов, так и обращение к респондентам. Возраст респондентов от 19 до 57 лет позволяет считать, что в этой группе присутствуют как те, кто планирует выбор и приобретение профессии для себя, так и те, кто решает эти проблемы в жизни своих детей. Такое сочетание ситуаций позволяет трактовать действия респондентов как направленные на ожидание изменений в жизни будь то выбор профессии, решение вопросов проживания или внешнего вида.

Условное название кластера 4 «Семейные и бытовые проблемы». В данном кластере объединяются ситуации: проконсультировать по вопросам внешности, одежды; обсудить взаимоотношения в семье; организовать праздник; помочь с ремонтом квартиры; помочь с уходом за больными; посоветовать Вам или ребенку спортклуб, секцию, спортшколу; пригласить Вас отдохнуть на даче; помочь выбрать профессию Вам или Вашему ребенку; перевести текст на иностранный язык; одолжить крупную сумму денег; отремонтировать компьютер; профессионально проконсультировать по вопросам здоровья; предоставить место для недельного проживания; найти работу; выбрать политическую позицию; перевезти или перенести груз; проконсультировать по вопросам капитального ремонта дома. Для данного кластера значимыми являются ситуации необходимости оказания помощи, организации отдыха, вопросы взаимоотношений в семье.

Участниками взаимодействия являются друзья, знакомые и родственники. Отмечается одностороннее обращение респондентов, одностороннее обращение к респондентам и взаимное обращение. Данный кластер составляют исключительно ответы женщин в возрасте от 22 лет до 51 года. Интересно, что к женщинам обращаются даже в тех случаях, когда в соответствии со стереотипами следует ожидать помощи от мужчин, а именно когда нужно помочь с ремонтом квартиры, проконсультировать по вопросам капитального ремонта дома, перенести или перевезти груз, отремонтировать компьютер и выбрать политическую позицию.

Условное название кластера 5 «Обучение и воспитание детей». Для данного кластера значимыми являются следующие ситуации: устроить ребенка в школу; устроить ребенка в детсад; посидеть с детьми; посоветовать Вам или ребенку спортклуб, секцию, спортшколу; выбрать политическую позицию; составить налоговую декларацию; профессионально проконсультировать по вопросам здоровья; отремонтировать автомобиль. Участниками взаимодействия выступают друзья, знакомые и родственники. Преимущественно отмечается одностороннее обращение респондентов, иногда взаимное обращение. Респонденты, как мужчины, так и женщины в возрасте от 27 до 48 лет. Наибольшая односторонняя активность проявляется в вопросах обучения и воспитания детей. По этим вопросам респонденты обращаются к друзьям, знакомым и родственникам.

Условное название кластера 6 «Профессиональные знания». Он объединяет следующие ситуации: обсудить взаимоотношения в семье; организовать праздник; найти работу»; посоветовать Вам или ребенку спортклуб, секцию, спортшколу; пригласить Вас отдохнуть на даче; перевести текст на иностранный язык; научить водить машину; отремонтировать компьютер; помочь с ремонтом квартиры. Для данного кластера значимыми являются ситуации, решение которых требует специальных знаний и навыков. Участниками взаимодействия выступают друзья, знакомые и родственники. Отмечается одностороннее обращение респондентов, обращение к респондентам и взаимное обращение. Респондентами, ответы которых объединены в данном кластере являются женщины в возрасте от 21 года до 38 лет.

Таким образом, применение метода кластерного анализа в рамках подхода построения самоорганизующихся карт признаков Кохонена позволило выявить особенности поведения в ситуациях взаимодействия сходными с трудовыми в социальной группе участников официальной части открытого рынка труда. Результаты подобных исследований могут предоставить сотрудникам служб занятости, кадровых агентств и бирж труда сведения о реальном потенциале их клиентов, поскольку из желания выглядеть более привлекательно последние часто

могут выдавать желаемые качества за действительные. Кроме того, в теоретическом плане мониторинг изменений или стабильности подобных характеристик поведения безработных в ситуациях трудового типа поможет оценивать состояние трудовых ресурсов, которое П.Б. Струве называл «личной годностью» и считал неотъемлемой характеристикой населения страны, которая стремится быть прогрессивной [11, с. 202-203].

### *Литература:*

1. Чудова О.В. Кластерный анализ социологической информации // Материалы IV всероссийской научной конференции «Сорокинские чтения» «Отечественная социология: обретение будущего через прошлое». – М.: СП Мысль, 2008. – С. 89-92.
2. Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных к знаниям. – СПб.: Питер, 2010. – 704 с.
3. Кузнецов Д.Ю., Трошина Т.Л. Кластерный анализ и его применение // Ярославский педагогический вестник. – 2006. – Вып. 4. – С. 103-107.
4. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Классификация и снижение размерности. – М.: Финансы и статистика, 1989. – 607 с.
5. Дуда Р., Харт П. Распознавание образов и анализ сцен. – М.: Мир, 1976. – 512 с.
6. Berry M.J.A., Linoff G.S. Data Mining Techniques for Marketing, Sales and Customer Relationship Management. – New York: Wiley Publ. Inc. 2004. – 648 p.
7. Круглов В.В., Дли М.И., Голунов Р.Ю. Нечеткая логика и искусственные нейронные сети. – М.: Физматлит, 2000. – 224 с.
8. Комашинский В.И., Смирнов Д.А. Нейронные сети и их применение в системах управления и связи. – М.: Горячая линия–Телеком, 2003. – 93 с.
9. Мальцева А.В. Сегментация рынка труда с использованием интеллектуальных алгоритмов анализа данных // Материалы Международного молодежного научного форума «ЛОМОНОСОВ-2010» / Отв. ред. И.А. Алешковский, П.Н. Костылев, А.И. Андреев, А.В. Андриянов. [Электронный ресурс] – М.: МАКС Пресс, 2010. – 1 электрон. опт. диск (CD-ROM); 12 см.
10. Berthold M., Hand D.J. (Eds.) Intelligent Data Analysis. – Berlin: Springer-Verlag, 2007. – 514 p.
11. Струве П.Б. Patriotica: Политика, культура, религия, социализм. – М.: Изд-во Республика, 1997. – 527 с.

## **Opportunities Provided by Cluster Analysis in Processing of Results of Sociological Survey (as Exemplified by Polling of Unemployed Citizens in Town Employment Center of Barnaul)**

**A. Maltseva**  
**The Altay State University**

*Modern methods and techniques of processing and analysis of data provide the sociologist with enormous opportunities in the process of searching for answers for raised questions. The article refers to the method of cluster analysis and opportunities it provides for interpretation of social information.*

*Key words: variants of cluster analysis, self-organizing maps of Kohonen, technology of Data Mining, labour market, subject situations.*

