

Мальцева Анна Васильевна

Maltseva Anna Vasilievna

ФГБОУ ВПО «Алтайский государственный университет»

Altai state university

Доцент кафедры математических методов в социальных науках,

заведующая лабораторией математического обеспечения

социальных и психологических исследований

Associate professor of chair of mathematical methods in social sciences,

head of laboratory of mathematical assistance of sociological

and psychological researches

E-mail: annamaltseva@rambler.ru

**О частной методике сегментации рынка труда с использованием
современных технологий обработки и анализа данных**

**On particular technique of labour-market segmentation based on application of current
technologies of data processing**

Аннотация: В статье рассматриваются общие теоретические подходы и методика сегментации рынка труда с использованием самоорганизующихся карт Кохонена. Данный метод не требует априорных предположений о распределении данных; позволяет обнаруживать в наборах данных неизвестные ранее структуры или образ, открывает социологам возможность анализа сведений из производственных баз данных. Современные же информационные технологии позволяют сделать этот процесс максимально эвристичным.

Ключевые слова: Рынок труда, авторская методика сегментации, современные методы анализа данных

The Abstract: The article covers procedure of labour-market segmentation with application of Kohonen Map method. This method is not based on any prior guess and let one to find our quite new structures and modes in data sets. Besides current technologies of data analysis (analytical platform) provides a researcher with some new and useful ability.

Keywords: Labor market, original method of segmentation, current methods of data analysis

Рынок труда является неотъемлемой частью жизни современного общества. Возможные экономические потрясения и необходимость оперативного реагирования и выработки управленческих решений как на уровне одной организации, так и управления социальными процессами, требуют применения соответствующих методик анализа производственной информации. Нарастающие объемы информации ставят все новые цели перед учеными-практиками, делая актуальным обоснование новых взглядов на решение традиционных проблем и применения современных технологий.

Научный коллектив лаборатории математического обеспечения социальных и психоло-

гических исследований Алтайского государственного университета в составе асп. О.В. Чудовой., доц. Н.Е. Шилкиной и под руководством автора статьи, по заказу Алтайского Управления по труду и занятости населения провел аналитический проект «Сегментация рынка труда». Эмпирической базой анализа явилась официальная часть открытого рынка труда, сведения о которой собираются территориальными отделениями Управления по труду и занятости в производственной базе данных[15].

Подходы к сегментации рынка труда

Существует большой пласт литературы от работ А. Смита и Д. Рикардо, до современных авторов: А. Маршалла, А. Пигу, М. Фридмена, Ф. Хайека и многих других, где обоснованы подходы к трактовке природы рынка труда. Существуют источники, обосновывающие основания сегментации рынков вообще и рынков труда, в частности, от работ Дж. Данлопа, П. Доринджера, М. Пайора, К. Керра, Р. Буайе и других. В ряде источников уже проводилось обобщенное рассмотрение оснований классификации рынков труда [16,21], ряд имен и работ является классикой в трактовке понятия сегментация [4,10], а некоторые источники содержат критику общепризнанных оснований сегментации за некоторую недоработанность в определении терминов и критериев [22]. Термин «маркетинг рынка» стал неотъемлемой частью не только коммерческих предприятий, но и сферы государственного управления, где уверенно существуют понятия «маркетинг трудовых ресурсов», «сегментация рынка труда» [18]. В целом приводимые в подобных работах взгляды отражают попытку общетеоретического осмысления феномена рынка труда и происходящих на нем процессов, описания присущих ему характеристик и свойств. Характерной чертой является достаточно жесткая укорененность выводов в социально-теоретических воззрениях на природу общественных отношений, экономическое устройство и структуру общества. В особенности последнее соображение часто служит вариантом основания для сегментации и рынка труда, когда под сегментами скорее понимается проекция структуры общественного устройства. Наблюдаемые пересечения смысловых полей основополагающих понятий допустимы для такого уровня, но, в случае актуализации задачи подробного эмпирического описания отдельных сторон рынка труда, становится очевидной потребность в новых подходах[17].

Сегментация рынка труда в парадигме «Как есть»

Достойной альтернативой видится использование современного методного арсенала математической науки. В этой связи важным представляется ограничение анализируемого пространства рынка труда по основаниям: внешний и внутренний рынок, или – открытый и скрытый, причем здесь возможна их дополнительная конкретизация по признаку официальной (зарегистрированной) части и неофициальной (незарегистрированной). Разработанная методика предлагает решение задачи сегментации рынка труда именно по таким параметрам, что дает новую трактовку рынка труда, в общем, а в частности, позволяет использовать полученные таким образом данные в качестве инструмента оперативного решения проблем занятости, поскольку включаемые в анализ переменные предполагают описание характеристик трудоспособных лиц свободных от занятости; безработных, ищущих другое место работы; граждан впервые вступающих в трудовую деятельность или ищущих рабочее место после перерыва в работе по разным причинам. Именно описанная категория представителей рынка рабочей силы потенциально (и вполне реально) является клиентами государственных и негосударственных служб занятости. Поскольку рынок труда немыслим без второй своей составляющей – рынка вакансий, то описываемая методика учитывает сегментацию не только рынка рабочей силы, но и рынка официально регистрируемых службами занятости вакансий. Кстати, подобная мысль уже была высказана в контексте необходимости теоретического анализа особенностей рынка труда малых городов[2,3]. Методика предполагает получение дополнитель-

ного эмпирического наполнения выводов о специфике и природе процессов, происходящих на рынке труда.

Обоснованным в этой связи является определение открытого рынка труда как всего экономически активного населения ищущего работу и нуждающегося в профориентации, подготовке и переподготовке, а также все вакантные рабочие и ученические места во всех секторах экономики. Также для определения объекта анализа по описываемой методике, важным является пояснение, что возможно более подробное изучение официальной части открытого рынка труда, включающего в себя свободную рабочую силу и вакансии, зарегистрированные в органах службы занятости, а также ученические места в системе профобразования (государственных и негосударственных учебных заведениях). Обозначенные предпосылки являются основополагающими для представляемой методики сегментации рынка труда с использованием методов интеллектуального анализа данных. Дополнительным аргументом в пользу предлагаемой методики является соображение о возможности преодоления директивности определения сегментов до проведения самой процедуры анализа эмпирических данных. Разработанная методика предполагает анализ данных в парадигме «As is» или «Как есть», что дает возможность не только применить современные математические модели анализа данных, но и осуществить оценку ранее незамечаемых аспектов существования рынка труда, что в свою очередь дает неоспоримое преимущество для оперативного принятия управленческих решений в области занятости населения. Дополнить предыдущую мысль можно тем, что в особенности в периоды экономических кризисов или реформ, именно незанятое по различным причинам население (безработные, ищущие другое место работы, впервые вступающие в трудовую деятельность или ищущие занятости после перерыва в работе или в качестве подработки) является наиболее актуальной целевой аудиторией для государственного регулирования.

Метод многомерного анализа данных в сегментации рынка труда

Обоснуем далее математические методы проведения подобной аналитической работы с использованием методов многомерного анализа данных логико-алгебро-геометрического направления[14,20]. Методы многомерного анализа наиболее действенный количественный инструмент исследования социально-экономических процессов, описываемых большим числом характеристик. К ним относится: кластерный анализ, распознавание образов, факторный анализ. В настоящее время все большую популярность в прикладных областях знания набирает парадигма «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 виде, четко отражающем происходящие процессы, а также построить модель, при помощи которой можно прогнозировать процессы, критичные для планирования и принятия управленческих решений. Достаточно ценной является возможность получения доступа к ранее недоступной для социолога области эмпирических массивов – данным из производственных баз данных. Существующие сегодня прикладные пакеты статистического анализа данных, в особенности аналитические платформы, позволяют реализовать возможности «Data Mining» в рамках процесса «Knowledge Discovery in Databases» или «Извлечения знаний из баз данных». Преимущество предлагаемой методики в ее адаптивности к изменениям экономики, нововведениям в области управления коммерческими и государственными структурами, с одной стороны, и возможность автоматизации анализа на основе сценарного метода на базе современных аналитических платформ, с другой[23,24].

Кластеризация с использованием самоорганизующихся карт Кохонена

Сегментация рынка – деление (дифференциация) любого рынка на отдельные части (сегменты) с учетом множества критериев и факторов. Выявление структуры и скрытых закономерностей рынка часто может быть выполнено только на основе решения задачи кластеризации. Кластерный анализ – один из распространенных методов решения задач, возникающих при анализе социальной и экономической информации. Примером может служить описание структуры социально-территориальных групп, разбиение клиентов рынка на различные по уровню мотивации группы и других задач, то есть нахождение групп схожих объектов в выборке данных. Кластеризация – это процесс разбиения множества объектов на заданное или неизвестное число кластеров на основании некоторого математического критерия качества кластеризации. 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)[12]. «Кластер» (cluster) в английском языке означает «сгусток», «гроздь винограда», «скопление звезд». Данный метод исследования получил развитие в последние годы в связи с возможностью компьютерной обработки больших баз данных. Многообразие алгоритмов кластерного анализа обусловлено множеством различных критериев, отражающих те или иные аспекты качества автоматической группировки. При этом практически все методы характеризуются высокой вычислительной сложностью и чувствительностью к точности задания априорной информации о количестве кластеров или максимально допустимом размере кластера, о предполагаемой структуре кластеров.

Предполагаемая выборка объектов исследования O_1, \dots, O_n (участники рынка труда: вакансии или клиенты службы занятости) может быть сформирована в результате отбора некоторых представителей генеральной совокупности. Использование методов Data Mining осуществляется для выборки данных, содержащих n записей. Из нее формируется k классов (групп объектов); число классов может быть, как выбрано заранее, так и не задано (в последнем случае оптимальное количество кластеров должно быть определено автоматически). Решение задачи кластеризации предполагает на основании анализа параметров объекта P_1, \dots, P_m (данных об участниках рынка из баз данных) выделение схожих объектов и представление результата в форме, удобной для восприятия. Набор P_1, \dots, P_m может включать переменные разных типов (количественные, качественные, порядковые). Решением задачи кластерного анализа является разбиение, удовлетворяющее некоторому условию оптимальности. Этот критерий может представлять собой некоторый функционал, выражающий уровни желательности различных разбиений и группировок[12]. Этот функционал часто называют целевой функцией. Задачей кластерного анализа является задача оптимизации, то есть нахождение минимума целевой функции при некотором заданном наборе ограничений.

Кластерный анализ включает в себя следующие этапы [5]: 1) получение репрезентативной выборки, формирование системы переменных и выделение наиболее информативных признаков; 2) вычисление некоторой меры близости между объектами (признаками). Это означает переход от таблицы X «объект-признак» к таблице «объект-объект» $S = \|s_{ij}\|_{n \times n}$ (таблица «признак-признак» $R = \|r_{ij}\|_{k \times k}$), где s_{ij} – мера близости между объектами O_i и O_j (r_{ij} – мера близости между объектами P_i и P_j); 3) применение конкретного метода кластерного

анализа; 4) проверка достоверности и интерпретация полученных результатов. Каждый из перечисленных этапов играет важную роль при использовании кластерного анализа для решения прикладных задач. Выделяется несколько подходов к решению задачи кластерного анализа, которые основаны на различных представлениях о задаче, использовании специфичной для каждой предметной области дополнительной информации. Среди них вероятностный подход [1]; подход, использующий аналогию с центром тяжести [8]; подход, основанный на понятии ближайшего соседа [1]; иерархический подход; нечеткие алгоритмы кластерного анализа [9]; подход, использующий искусственные нейронные сети [11,19].

В рамках решаемой задачи целесообразно применить методы позволяющие проводить разделение на не пересекающиеся классы с произвольным их числом. Кластеризация методом Кохонена в данном случае имеет ряд преимуществ: возможно решение задач большой размерности, число кластеров может быть не задано, формируют наглядное двумерное отображение множества. Метод анализа с использованием самоорганизующихся карт признаков (СКП) Кохонена [7], являющихся разновидностью неуправляемых нейросетей, позволяет автоматизировать все действия по поиску закономерностей. Главной причиной использования 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. 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 [6]. В результате, трудно анализируемые совокупности многомерных объектов 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 [20], что позволяет 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 получается так же, как и в обычной сети Кохонена, то есть 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 сильно. 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. Таким образом, есть два важных фактора: положение объекта на карте (расстояние до других объектов) и цвет ячейки [15].

Алгоритм работы с данными

Формат статьи не позволяет в полном объеме представить разработанную методику, но возможно привести краткое описание основных этапов работ. Прежде всего, отметим, что описываемая методика предполагает свою реализацию только с использованием сведений из производственных баз данных с процедурой импорта сведений из них по заданному алгоритму в среду аналитической платформы «Deductor» [15], и последующий запуск специального сценария обработки и анализа с применением описанных выше математических приемов. Аналитические платформы позволяют осуществлять широкий спектр полезных действий по предварительной предобработке импортированного массива данных. На описываемом подготовительном этапе анализа проводится проверка соответствия типов переменных (полей) заданным характеристикам и осуществляется подготовка меток полей к последующему анализу и интерпретации. Операция осуществляется путем импорта полного имени меток переменных из соответствующих справочников производственной базы данных, к которым относятся справочники, поясняющие весь диапазон возможных значений меток переменных (полей). Например, «Образование», «Отношение к занятости», «Категория незанятости» и другие.

Следующий этап – основной, или проведение кластерного анализа. Он предполагает следующие шаги:

- 1) действия по фильтрации данных, участвующих в сегментации за отчетный период;
- 2) осуществление предобработки данных (аудит данных);
- 3) расчет вспомогательных переменных для проведения сегментации рынка труда, подготовка массива данных;
- 4) построение карт Кохонена;
- 5) содержательный анализ и интерпретация результатов.

Аналогичная работа проводится с записями, представляющими спрос на рабочую силу (предлагаемые вакансии). Специфичным является лишь набор справочников, например: «Способ получения информации о вакансии», «Статус вакансии» и другие.

Полный текст частной методики сегментации рынка труда включает перечень приложений со значениями меток переменных (полей), участвующих в анализе, из Федеральных, Региональных и прочих справочников и регламентов. Также присутствует раздел математического описания алгоритма сегментации официальной части открытого рынка труда в для анализа сведений о клиентах и вакансиях, регистрируемых в службах занятости населения.

Применение представленной методики может служить инструментом анализа актуального состояния института рынка труда, а также основанием для порождения гипотез при осмыслении трансформационных процессов современного общества. Так, например, сочетание результатов применения методики с объяснительными возможностями дистрибутивного направления изучения социально-структурных изменений общества, существующего в рамках социологического знания, позволит:

- 1) при анализе данных о предложениях рабочей силы увидеть: а) уже произошедшие изменения социальных позиций; б) тенденции к изменению социальных позиций; в) вре-

менные изменения социальных позиций; г) нарушения социальных ожиданий в отношении смены социальных позиций и другие показатели изменения социальной структуры;

2) при анализе данных о спросе на рабочую силу диагностировать совокупности сложившихся требований, задающих систему взаимосвязанных позиций для граждан, выполнение которых обеспечивает приобретение заданного вакансией места в социальной структуре общества.

В свою очередь изыскания в этом направлении способны дать обоснованные выводы о функциональном состоянии социального института рынка труда, выявить признаки его дисфункциональности и увидеть позитивные или негативные изменения социальных позиций, оценить перспективы трансформации социальной структуры. Таким образом, рынок труда становится важным индикатором данной трансформации [13].

Материалы публикуются в рамках реализации гранта РГНФ "Рынок труда как индикатор трансформации социальной структуры современного российского общества (на примере исследований в Алтайском крае)" №12-13-22001.

ЛИТЕРАТУРА

1. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Классификация и снижение размерности. - М.: Финансы и статистика, 1989. - 607с.
2. Аралбаев А.Т. Управление развитием рынка труда в регионе: маркетинговый подход : дис. канд. экон. наук: 08.00.05/ А.Т. Аралбаев. - Оренбург: Пресса, 2007. 212с.
3. Аралбаев И.М. Теоретический анализ особенностей рынка труда малых городов : дис. канд. экон. наук: 08.00.05/ И.М. Аралбаев. - Уфа, 2000. - 165с.
4. Ассэль Г. Маркетинг: принципы и стратегия. - М.: ИНФРА-М, 1999. - 803с.
5. Бритвихин В.А., Красина Ф.А., Симонцев С.Н. Использование кластерного анализа для типологизации признаков // Социологические исследования. -1994. - № 2. - 100-105.
6. Горбунов А. Применение самоорганизующихся карт в бизнесе и финансах // Банковские технологии. -1999. - N 4. - <http://it2b2.view2.page103.html>.
7. Дебок Г., Кохонен Т. Анализ финансовых данных с помощью самоорганизующихся карт. - М.: Альпина Паблишер, 2001. - 317с.
8. Дуда Р., Харт П. Распознавание образов и анализ сцен. - М.: Мир, 1976. - 512с.
9. Зайченко Ю.П. Нечеткие модели и методы в интеллектуальных системах. - Киев: Слово, 2008. - 344с.
10. Котлер Ф., Армстронг Г., Сондерс Д., Вонг В. Основы маркетинга. - М., СПб.: Питер, 1998. - 448с.
11. Круглов В.В., Дли М.И., Голунов Р.Ю. Нечеткая логика и искусственные нейронные сети. - М.: Физматлит, 2000. - 224с.
12. Кузнецов Д.Ю., Трошина Т.Л. Кластерный анализ и его применение // Ярославский педагогический вестник. - 2006. - Вып. 4. - 103-107.
13. Мальцева А.В. Индикативная роль социального института рынка труда в трансформации социальной структуры современного российского общества : автореф. дис. ... докт. социол. наук : 22.00.04 / А.В. Мальцева. - Барнаул: Изд-во Алтайского гос. ун-та, 2012. - 48 с.

14. Мальцева А.В. Современные информационные технологии в обеспечении прикладных исследовательских задач в социально-культурной сфере // Известия Алтайского государственного университета, Серия «Математика и механика, управление, вычислительная техника и информатика». - 2009. - 1(61). - 143-144.
15. Мальцева А.В., Чудова О.В., Шилкина Н.Е. Сегментация рынка труда: теория и методика. - Барнаул: АзБука, 2010. - 116с. Режим доступа: http://www.basegroup.ru/library/books/asu_monograph/.
16. Маркетинг рабочей силы: Учеб. Пособие / Томилов В.В., Семеркова Л.Н. - СПб.: Гос. Ун-т экономики и финансов, 1997. - 184с.
17. Методы количественного анализа в маркетинге / Под ред. Т.П. Данько, И.И. Скоробогатых. - СПб.: Питер, 2005. - 384с.
18. Никифорова А.А. Рынок труда: занятость и безработица. - М.: Международные отношения, 1991. - 184с.
19. Оссовский С. Нейронные сети для обработки информации / Пер. с польского И.Д. Рудинского. - М.: Финансы и статистика, 2002. - 365с.
20. Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных к знаниям. - СПб.: Питер, 2009. - 624с.
21. Паршин А.А. Методы и принципы сегментного анализа в маркетинге // <http://www.marketing.spb.ru/read/article/a54.htm>.
22. Резниченко Б.А. Критический анализ критериев сегментирования // <http://www.antema.ru>.
23. Berry M.J.A., Linoff G.S. Data Mining Techniques for Marketing, Sales and Customer Relationship Management. – New York: Wiley Publ. Inc., 2004. - 672p.
24. Berthold M., Hand D.J. Intelligent Data Analysis. - Berlin: Springer-Verlag, 2007. - 514p.