# Введение

Одним из приоритетных направлений модернизации экономики РоссийскойФедерации является реализация кластерного подхода. С 2011 года в России идет активная кластерная политика подэгидой Министерства экономического развития РФ, принято решение поддержать часть кластеров черезпредоставление субсидий из федерального бюджета.

Согласно определению, приведенному в«Методических рекомендациях по реализациикластерной политики» Минэкономразвития России, кластер – это «объединение предприятий, поставщиков оборудования,комплектующих, специализированных производственных и сервисныхуслуг, научно-исследовательских и образовательных организаций,связанных отношениями территориальной близости и функциональнойзависимости в сфере производства и реализации товаров и услуг».

Кластерные хозяйственные системы отличаются высокой конкурентоспособностью и стабильностью. В Ульяновской области развиваются ядерный, автомобильный, авиационный,пищевой, стекольный кластеры, кластер производства строительныхматериалов[22].

Одним из приоритетных направлений развития экономики является развитие информационных технологий, поэтому актуально формирование ИТ-кластера Ульяновской области, где происходит разработка новых информационных технологий и программного обеспечения, их реализация и внедрение в производство.В связи с этим актуальным является исследование текущего уровня ИТ-сектора, применение методов кластерного анализа, сравнение теоретического анализа данных с намечающимися кластерами ИТ-предприятий.

Этой тематикой занимается коллектив ученых УлГТУ: Н.Г. Ярушкина, Т.В. Афанасьева, О.В. Шиняева, К.В. Святов, Л.М. Валкин, Д.А. Ефремов, К.Г. Калимуллин и др. Результаты исследования представлены в монографии Н.Г. Ярушкиной, Т.В. Афанасьевой и др. «Исследование ИТ-кластера Ульяновской области» [22].

Социально-экономические системы в их динамическом развитии характеризуются многомерными данными. При анализе информации о сложных системах необходимы обработка и анализ огромных объёмов разнородных данных, важно упорядочить их по иерархическим уровням и произвести группировку, разбиение множества исследуемых объектов и признаков на однородные группы или кластеры. Это позволяет сделать кластерный анализ, впервые применённый в социологии Трионом[25]в 1939г.

Встаёт проблема автоматизации процесса обработки и анализа информации, необходимости создания простого и доступного инструмента (программного пакета) для группировки и распознавания данных, произвольно структурированных, и типов данных. Остро возникает проблема представления данных в виде удобном и понятном для пользователя. Необходим программный код, который бы мог выполняться на различномHardWare, без его предварительного переноса на «родной язык» устройства. Актуально использование методов кластерного анализа для численных и нечетких данных.

На сегодняшний день актуальным является развитие методов кластерного анализа (см. [9],[10], [11], [12], [15]и др.), среди которых метод полных связей, метод максимального локального расстояния, метод Ворда, центроидный метод и др. Актуально их применение к решению экономических задач (см. [6], [17], [18], [20]и др.), в частности, сегментации рынка (модель Уинда и Кардозы и др.), их программная реализация.

На сегодняшний день актуально сочетание кластерного анализа с теорией нечетких множеств (см. [2], [3], [5], [7], [13], [21], [23] и др.), основы которой заложены в 60-е годы американским математиком Латфи Заде [5]. В настоящее время Японией и США активно разрабатываются электронные системы с нечеткими управляющими алгоритмами.

Существует программное обеспечение для кластерного анализа в узких предметных областях, например, «ClusterDelta» позволяет производить анализ данных на бирже, существуют программы для прогноза конъюнктуры рынка и др.

Существуют пакеты программ, реализующие набор методов кластерного анализа данных («STATISTICA MultivariateExploratoryTechniques», «MatLab» и др.) В программном пакете «STATISTICA MultivariateExploratoryTechniques» реализованы методы k-средних, иерархической кластеризации и двухвходового объединения. Используются различные метрики расстояний. Этот пакет не ориентирован на решение задач кластеризации в условиях неоднородных показателей, на обработку нечетких данных.В «MatLab» имеются встроенные функции, позволяющие реализовать алгоритмы четкой и нечеткой кластеризации.Эти пакеты рассчитаны на использование профессиональными математиками и в этом их ограничение.

**Цель работы** – создать программный инструмент для кластеризации ИТ-предприятий.

**Задачи работы:**

– изучение методов кластерного анализа, элементов теории нечетких множеств;

– изучение возможностей, особенностей программной реализации в программных продуктах для кластерного анализа;

– программная реализация методов создания, сохранения, загрузки, обработки кластеров (данных);

– изучение методических, аналитических материалов, результатов статистических и социологических исследований ИТ-кластера;

– анализ объектов ИТ-рынка,анализ и выбор критериев и параметров, используемых при кластеризации ИТ-предприятий;

– разработка алгоритма кластеризации ИТ-предприятий; разработка программы кластеризации ИТ-предприятий;

– экспериментальная часть: применение программы при обработке данных ИТ-компаний Ульяновской области, кластеризация ИТ-предприятий Ульяновской области; анализ текущего уровня развития ИТ-предприятий в регионе.

Нами разработан программный продукт «Кластеризация ИТ-предприятий».Программный продукт позволяет группировать данные иерархическим центроидным методом кластерного анализа, неиерархическим FCM-методом, который предполагает возможность работы с нечеткими данными. Выполнена программная реализация методов создания, сохранения, загрузки, обработки кластеров (данных).

Программный продукт позволяет: визуализировать данные на экране в виде дерева, позволяющего получить результат кластеризации в виде иерархической структуры, по которой можно судить о ходе выполнения кластеризации, выводить на экран результаты анализа данных в виде таблиц, графиков и диаграмм.

Используется программный код, который может выполняться на различномHardWare, без его предварительного переноса на «родной язык» устройства. Нами был выбран C#. Код на C# будет работать везде, где установлен «C# .NET». Сохранение и загрузка данных происходит в формате \*.SOF (SerializableObjectFiles).

Предусмотрена возможность использования кластерного анализа в сочетании с другими методами многомерного анализа, возможность использования различных метрик; возможность включения программного продукта в состав более сложной автоматической или автоматизированной системы для экономического анализа.

Программа может быть интегрирована в состав более сложной автоматической или автоматизированной вычислительной системы (автоматической системы управления, прогнозирования и т.д.), размеры которой могут быть как в пределах одной машины, в пределах локальной, глобальной сети, с оповещением через уже существующие средства связи (sms, Email, …).

**Новизна работы** в том, что:

– подобных программных продуктов для сегментации рынка IT нами не найдено;

– используется авторский подход программной реализации методов кластерного анализа;

– сохранение и загрузка данных происходит в формате \*.SOF (SerializableObjectFiles), пользователь имеет возможность редактировать данные вне программы;

– возможность использования кластерного анализа в сочетании с другими методами многомерного анализа.

– программное обеспечение ориентировано на обработку числовых и нечетких данных.

**Практическая значимость.**Программный продукт предназначен для кластерного анализа данных IT-предприятий, для обработки экономических данных и их анализа, группировки, распознавания и представления в удобном для пользователя виде.

Программное обеспечение предполагается использовать для сегментации ИТ-предприятий Ульяновской области по следующим группам критериев: область деятельности компании; финансово-экономические показатели; рынок сбыта; кадровые показатели; внутренние процессы компании; инновационная деятельность; социальные программы и образовательная деятельность. При вычислительном эксперименте используются результаты анкетировании руководителей ИТ-предприятий в 2013 г. [21].

Программное обеспечение адаптировано для экономической терминологии «Кластеризация IT-предприятий».

Потребителями данного программного продукта могут выступить экономисты, социологи, фирмы, лица, занимающиеся производством и продажей в сфере IT, которым полезно сегментирование рынка товаров и услуг, структурирование рынка, основанное на неоднородности и нечеткости показателей и критериев. Такой анализ необходим для дифференцированного маркетинга.

**Апробация.**Программный продукт был представлен на конкурсе инновационных проектов по программе Фонда содействия развитию малых форм предприятий в научно-технической сфере «Участник молодежного научно-инновационного конкурса 2013» («УМНИК»).

Работа представлена на XXIXмеждународной научно-практической конференции «Инновации в науке» (Новосибирск, 2014).

Опубликована статья Афанасьевой Т.В., Сибирева И.В. «Программа «Сегментации и кластеризация рынка IT» [4].

Работа состоит из введения, четырех глав, заключения, списка литературы и приложения.

Перваяглава посвящена истории возникновения и развития кластерного анализа и теории нечетких множеств, здесь рассмотрены основные понятия, приведен обзор методов кластерного анализа.

Во второй главеописана методика кластеризации ИТ-предприятий: выбор критериев и параметров, используемых при кластеризации ИТ-предприятий; анализ и предварительная обработка данных анкет руководителей ИТ-предприятий г. Ульяновска; методы и алгоритмы кластеризации, применяемые при написании программы «Кластеризация ИТ-предприятий».

В третьей главе произведено описание программного продукта «Кластеризация ИТ-предприятий», его функций,руководство пользователю, руководство программисту.

В четвертой главе произведенанализ результатов кластеризации ИТ-предприятий Ульяновской области.

Библиография содержит 47литературных источников, в том числе 22 – электронных ресурса.