

Инструментальные средства для формирования рекомендаций при принятии управленческих решений по развитию IT-отрасли в регионах

О.Н. Сметанина

Факультет информатики и робототехники
Уфимский государственный авиационный
технический университет
Уфа, Россия
e-mail: smoljushka@mail.ru

Г.И. Медведев

Факультет информатики и робототехники
Уфимский государственный авиационный
технический университет
Уфа, Россия
e-mail: gleb_m94@mail.ru

Н.И. Юсупова

Факультет информатики и робототехники
Уфимский государственный авиационный
технический университет
Уфа, Россия
e-mail: yussupova@ugatu.ac.ru

А.Г. Ионис

Факультет информатики и робототехники
Уфимский государственный авиационный
технический университет
Уфа, Россия
e-mail: solomonarmkeys@gmail.com

А.Ю. Адельметова

Факультет информатики и робототехники
Уфимский государственный авиационный технический университет
Уфа, Россия
e-mail: alyonadelmetova@mail.ru

Аннотация¹

В статье рассматриваются вопросы разработки инструментальных средств для формирования рекомендаций при принятии управленческих решений по развитию IT-отрасли в регионе. Приведена характеристика предметной области и особенности используемых в задаче данных, математическая постановка задачи.

Подробно рассмотрены используемая для решения задачи концепция и инструменты анализа данных.

Результаты анализа интерпретированы в соответствии с семантикой предметной области, описаны средства формирования рекомендаций, обеспечивающие информационную поддержку принятия решений.

1. Введение

В настоящее время конкурентоспособность экономики любой страны в большей степени связана с развитием информационных технологий.

Труды шестой Всероссийской научной конференции "Информационные технологии интеллектуальной поддержки принятия решений", 28 - 31 мая, Уфа-Ставрополь, Россия, 2018

Согласно данным Всемирного экономического форума, индекс конкурентоспособности экономики государств имеет высокий уровень корреляции с индексом развития в странах информационно-коммуникационных технологий. Направления развития для IT-отрасли со стороны государства обозначены в «Стратегии развития отрасли информационных технологий в Российской Федерации на 2014 - 2020 годы и на перспективу до 2025 года». Согласно стратегии IT-отрасль является одной из наиболее динамично развивающихся отраслей, как в мире, так и в России. Средний темп роста российского рынка за последние 10 лет превосходит среднемировой. Каждый регион утверждает свои стратегии развития отрасли. В значительной части региональных стратегий рассматривает критерии, которые могут повлиять на уровень развития отрасли.

Вопросы составления рейтинга по развитию IT-отрасли, как отдельных стран, так и регионов РФ затрагивают различные компании, например, консалтинговая компания McKinsey, и Министерства РФ. В частности, Министерство цифрового развития, связи и массовых коммуникаций представило рейтинг регионов по уровню развития информационного общества. Ряд факторов, влияющих на развитие информационного

Шестая Всероссийская научная конференция "Информационные технологии интеллектуальной поддержки принятия решений", Уфа-Ставрополь, Россия, 2018

общества, непосредственно связан и с развитием ИТ-отрасли.

Авторами предложено выявить сходство между объектами (регионами) на основе наиболее значимых факторов, влияющих на развитие отрасли, и с использованием новых знаний сформулировать рекомендации по управлению отраслью. Для выявления сходства путем кластеризации применяется аппарат нейронных сетей. Выявленные новые знания в виде характеристик схожих объектов путем интерпретации данных позволяют сформулировать рекомендации по развитию отрасли в регионе. В дальнейшем авторы предлагают использование нейро-нечетких инструментов для решения задачи. На данном этапе решения задачи в дополнении к результатам кластерного анализа предлагается использовать знания эксперта.

Авторы статьи неоднократно обращались к использованию технологий искусственного интеллекта в своих исследованиях [6].

2. Характеристика предметной области и особенности данных для анализа

Осенью 2017 года, Минкомсвязи России представило рейтинг регионов по развитию информационного общества на 2017 год. Среди приоритетных подындексов нового рейтинга: ИКТ-инфраструктура, электронное правительство, ИКТ в сфере образования, ИКТ в сфере здравоохранения и ИКТ в сфере транспорта. В целом перечень подындексов соответствует разделам Концепции развития региональной информатизации, утвержденной распоряжением правительства РФ от 29 декабря 2014 года №2769-р [2]. Первые три позиции в рейтинге 2017 года занимают Москва, Тюменская область и Ханты-Мансийский автономный округ. В первую десятку вошли также Республика Татарстан, Тульская область, Новосибирская область, Республика Башкортостан, Ямало-Ненецкий автономный округ, Томская область и Челябинская область. На последних позициях – Еврейская АО, Чеченская Республика, Республика Крым.

Практически любая стратегия развития ИТ-отрасли в регионе ориентируются на человеческий капитал; инновационный потенциал; институциональную и инфраструктурную среду; ИКТ-инфраструктуру и доступ; информационную индустрию; информационную безопасность; использование ИКТ для развития и др.

3. Математическая постановка задачи исследования

Постановка общей задачи исследования может быть сформулирована следующим образом: на основе анализа исходных данных, представленных множеством векторов, описывающих характеристики регионов РФ в области развития

ИТ-отрасли, сформулировать рекомендации с целью повышения уровня развития.

Математическая постановка общей задачи исследования будет сформулирована таким образом. Дано: X – множество векторов, представленных характеристиками регионов в области $X_i = \{X_{i1}, X_{i2}, \dots, X_{in}\}$, здесь

$X_{i1}, X_{i2}, \dots, X_{in}$ – информация о регионе (сведения собраны с *web*-ресурсов), где $i = 1 \div 85$, $j = \overline{1, n}$ – номер, учитываемого фактора, влияющего на уровень развития (всего n факторов). Найти:

$F: X_i \rightarrow Z_i \rightarrow Y_c$ – функцию, которая в результате анализа позволит отнести регион по его характеристикам к одному из кластеров, объединяющих регионы со схожими характеристиками, и правила $RULE: X_i \wedge Y_c \rightarrow R$ (где R – рекомендации) с использованием которых можно формулировать рекомендации по развитию отрасли.

Математическая постановка частной задачи факторного анализа может быть сформулирована следующим образом: дано: $X_i = \{X_{i1}, X_{i2}, \dots, X_{in}\}$,

где $j = \overline{1, n}$, n – количество факторов, влияющих на ИТ-отрасль в регионах. Найти: сокращенную систему факторов $Z_i = \{Z_{i1}, Z_{i2}, \dots, Z_{il}\}$, где l – сокращенное количество факторов, путем отображения $f': X_i \rightarrow Z_i$.

Математическая постановка частной задачи кластерного анализа может быть сформулирована следующим образом: дано: $Z_i = \{Z_{i1}, Z_{i2}, \dots, Z_{il}\}$ – входная переменная. Найти: множество кластеров Y , при этом $\cup Y_j = X$, где Y_j – j -ый кластер, $\emptyset \subset Y_j \subset X$, $j = \overline{1, c}$, $Y_i \cap Y_j = \emptyset$; $i, j = \overline{1, c}$; $i \neq j$, где c – количество кластеров.

$$V = \sum_{j=1}^k \sum_{x_i \in Y_j} (x_i - \mu_j)^2 \rightarrow \min$$
, где μ_j – центр кластера.

Таким образом, для получения решения необходимо проведение факторного и кластерного анализа.

4. Инструменты анализа данных

Для анализа данных используется комплексная аналитическая платформа Deductor Studio, включающая хранилище данных, консолидирующее информацию из разных источников (Warehouse); приложение, позволяющее пройти все этапы построения прикладного решения, рабочее место аналитика (Studio); рабочее место конечного пользователя, одно из средств тиражирования знаний (Viewer); служба, обеспечивающая удаленную

Инструментальные средства для формирования рекомендаций при принятии управленческих решений по развитию ИТ-отрасли в регионах

аналитическую обработку данных (Server); клиент доступа к Deductor Server (Client), обеспечивающий доступ к серверу и управление его работой из сторонних приложений.

Функции аналитика: создание в Deductor сценария – последовательности шагов, которую необходимо осуществить для получения нужного результата; построение, оценка и интерпретация моделей; настройка панели отчетов Deductor Viewer для пользователей; настройка сценария на пакетную обработку новых данных.

В Deductor реализовано большинство аналитических технологий – от ETL и хранилищ данных до алгоритмов Data Mining (табл. 1), в частности, консолидация, трансформация, визуализация, очистка и предобработка данных, задачи DataMining (ассоциации, кластеризации, классификации и регрессии), анализа и прогнозирования временных рядов, ансамбли и сравнение моделей. В качестве языка визуального моделирования используются структуры в виде деревьев.

Таблица 1. Технологии, реализованные в Deductor

Аналитические задачи	Характеристика
Консолидация данных	Реляционное хранилище данных POLAR, виртуальное хранилище данных
Трансформация данных	Фильтрация, группировка, разгруппировка и замена данных, скользящее окно, квантование, слияние, нормализация и др.
Визуализация данных	OLAP-кубы, диаграммы, графики, гистограммы, правила, деревья, многомерные карты, матрицы классификации и др.
Очистка и предобработка данных	Фильтр Калмана, вейвлеты, заполнение пропусков средним значением, корреляционный анализ, метод главных компонент и др.
Data Mining: задача ассоциации	Алгоритм a priori
Data Mining: кластеризация	Сети и карты Кохонена, алгоритм k-means
Data Mining: классификация и регрессия	Линейная и логистическая регрессия, дерево решений, многослойный перцептрон и др.
Анализ и прогнозирование временных рядов	Выделение тренда, регрессионные и нейросетевые модели и др.
Ансамбли моделей	Сценарный подход, поддерживающий создание

Аналитические задачи	Характеристика
	однородных ансамблей и ансамблей на основе различных типов моделей и др.
Сравнение моделей	ROC-анализ

Вся работа в Deductor Studio построена на создании сценариев обработки данных при помощи трех мастеров – импорта, обработки и экспорта. Для настройки подключений к разнообразным источникам данных (от СУБД до специализированных учетных систем и веб-серверов) используется мастер подключений.

Одним из критериев выбора числа факторов является их хорошая интерпретируемость. Интерпретация же фактора осуществляется, исходя из того, с какими исходными переменными он сильно коррелирует. Коэффициенты корреляции между факторами и исходными переменными равны факторным нагрузкам, на основе которых можно рассчитать исходные переменные, зная факторы.

Просматривая список исходных переменных, сильно коррелирующих с фактором, можно дать фактору название и содержательную интерпретацию.

При проведении факторного анализа используется метод варимакс. В его основе лежит сокращение числа переменных, имеющих высокие нагрузки на данные. Метод используется на практике чаще всего, так как он существенно облегчает интерпретацию факторов.

Интерпретация результатов факторного анализа позволяет сделать следующие выводы, что на уровень развития ИТ-отрасли в регионах существенное значение оказывают внутренние резервы организации (затраты на проводимые научные исследования и их доля к валовому региональному продукту; проведение дополнительного обучения сотрудников (повышение человеческого капитала); инвестиции в основной капитал; доля использования компьютеров в организации); удельный вес, занятых в ИКТ; развитие сетей связи и передачи данных; доля выпускников ИТ направлений (специальностей) к общему выпуску; уровень развития техники и технологий; индексом рынка труда (рис. 1).

Результаты кластерного анализа (рис. 2) позволяют выявить схожие объекты, как с высоким, так и с низким рейтингом. Поэтому для того, чтобы ориентировать регионы на более высокий рейтинг необходимо повышать/понижать значения некоторых параметров. Эти рекомендации заложены в производственные правила.

Переменные	Окончательные факторы (Варимакс метод)						
	Фактор 1	Фактор 2	Фактор 3	Фактор 4	Фактор 5	Фактор 6	Фактор 7
Уровень развития науки и технологий					0,5369		
Развитие сетей связи и передачи данных, %				0,9696			
Внутренние тек. затраты на научные исслед. и разраб.	0,8532						
% выпускников с ИТ-специальностей (на 10000 населения)							0,9213
Доля организаций, проводивших дополнительное обучение сотр...	0,8682						
Удельный вес занятых в секторе ИКТ в общей численности зан...		0,9409					
Объем финансирования в 2016 г. (на 10000 человек населения)...			0,9635				
Доля внут. затрат на исслед. и разраб., в % к валовому региона...						0,9224	
Инвестиций в основной капитал, направленных на приобрете...	0,8866						
Использование организациями компьютеров и компьютерных с...					0,8543		
Индекс рынка труда					0,8153		

Рисунок 1. – Результаты факторного анализа

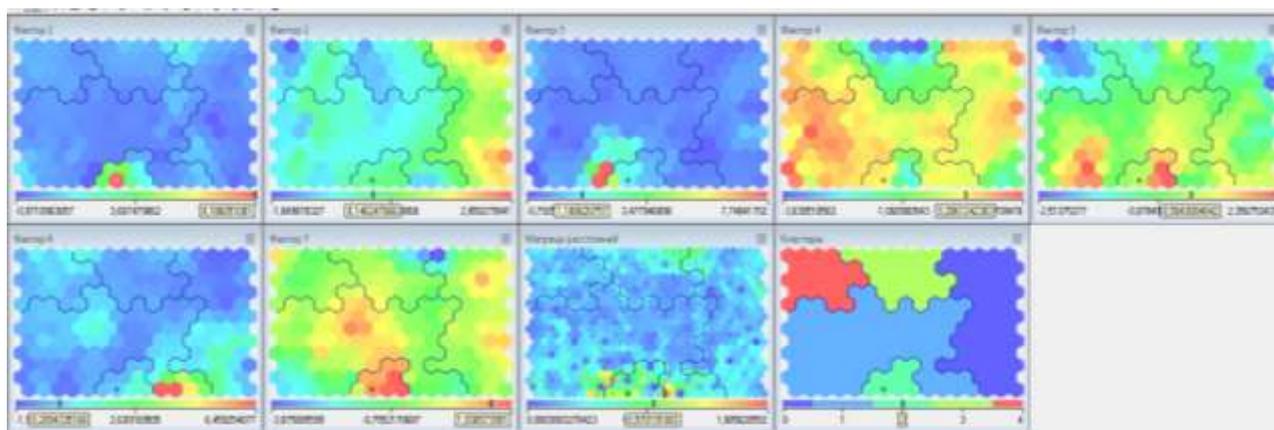


Рисунок 2. – Результаты кластеризации

5. Средства формирования рекомендаций

Системы обработки знаний, использующие продукционную модель получили название «продукционных систем». В состав экспертных систем продукционного типа входят база правил (знаний), рабочая память и интерпретатор правил (решатель), реализующий определенный механизм логического вывода. Любое продукционное правило, содержащееся в базе знаний, состоит из двух частей: антецедента и консеквента. Антецедент представляет собой посылку правила (условную часть) и состоит из элементарных предложений, соединенных логическими связками «и», «или». Консеквент (заключение) включает одно или несколько предложений, которые выражают либо некоторый факт, либо указание на определенное действие, подлежащее исполнению. Продукционные правила принято записывать в виде антецедент-консеквент.

Любое правило состоит из одной или нескольких пар «атрибут-значение». В рабочей памяти систем, основанных на продукционных моделях, хранятся пары атрибут-значение, истинность которых установлена в процессе решения конкретной задачи к некоторому текущему моменту времени. Содержимое рабочей памяти изменяется в

процессе решения задачи. Это происходит по мере срабатывания правил. Правило срабатывает, если при сопоставлении фактов, содержащихся в рабочей памяти, с антецедентом анализируемого правила имеет место совпадение, при этом заключение сработавшего правила заносится в рабочую память. Поэтому в процессе логического вывода объём фактов в рабочей памяти, как правило, увеличивается (уменьшается он может в том случае, если действие какого-нибудь правила состоит в удалении фактов из рабочей памяти). В процессе логического вывода каждое правило из базы правил может сработать только один раз.

С использованием полученных результатов кластеризации в комплексе со знаниями эксперта возможно формирование продукционных правил. В качестве примеров могут быть приведены следующие правила:

Правило 1: ЕСЛИ кластер = 3, то на предприятиях ИТ-отрасли нужно найти внутренние резервы для дополнительного обучения сотрудников, на научные исследования, обновление аппаратных средств.

Правило 2: ЕСЛИ кластер = 4 и рейтинговая оценка = 14, ТО на предприятиях ИТ-отрасли нужно найти внутренние резервы для

Инструментальные средства для формирования рекомендаций при принятии управленческих решений по развитию ИТ-отрасли в регионах

дополнительного обучения сотрудников, на научные исследования, обновление аппаратных средств И повысить уровень развития науки и технологий путем создания инновационных центров.

Правило 3: ЕСЛИ кластер = 1 и $9 \leq$ рейтинговая оценка ≤ 12 , ТО необходимо повысить удельный вес занятых в ИТ-сфере, например путем, создания новых ИТ-компаний.

Правило n: ЕСЛИ кластер = 1 и $9 \leq$ рейтинговая оценка ≤ 12 , ТО нужно увеличить выпуск ИТ – специалистов в ВУЗах за счет привлечения коммерческих студентов, работы с уже имеющимися студентами, усилением профориентационной работы, увеличения плана приёма (бюджет).

Заключение

Одной из основных характеристика предметной области является то, что ИТ-отрасль - наиболее динамично развивающаяся отрасль, как в мире, так и в России. Средний темп роста российского рынка за последние 10 лет превосходит среднемировой.

Практически любая стратегия развития ИТ-отрасли в регионе ориентируются на человеческий капитал; инновационный потенциал; институциональную и инфраструктурную среду; ИКТ-инфраструктуру и доступ; информационную индустрию; информационную безопасность; использование ИКТ для развития и др.

Предложенная концепция ориентирована на выявление схожих объектов путем проведения кластерного анализа. Результаты анализа интерпретированы в соответствии с семантикой предметной области, описаны средства формирования рекомендаций, обеспечивающие информационную поддержку принятия решений. Подробно рассмотрены инструменты анализа данных.

Результаты исследований, приведенные в статье, частично поддержаны грантами РФФИ 16-07-00773, 18-07-00193.

Список используемых источников

1. Стратегия развития отрасли информационных технологий в Российской Федерации на 2014 - 2020 годы и на перспективу до 2025 года [Электронный ресурс]: утверждена распоряжением Правительства Российской Федерации от 1 ноября 2013 г. № 2036-р. Режим доступа: <http://minsvyaz.ru/ru/documents/4084/#documentcontent>(25.04.2017).

2. Концепции региональной информатизации» [Электронный ресурс]: утверждена распоряжение Правительства РФ от 29.12.2014 №2769-р. Режим доступа: <http://minsvyaz.ru/ru/documents/4924/>(25.04.2017).

3. Методика выявления сходства объектов и интерпретация результатов (на примере развития ИТ-отрасли в регионах) / О.Н. Сметанина, Г.И. Медведев. // Мавлютовские чтения: XI Всероссийская молодежная научная конференция: сб. тр. в 7 т. Том4/ Уфимск. гос. авиац. техн. ун-т. - Уфа: УГАТУ, 2013, с. 246-250.

4. Проблемы информатизации. [Электронный ресурс]: Режим доступа [http://www.tadviser.ru/index.php/Статья:Информатизация_регионов_\(рынок_России\)#.D0.A2.D0.B5.D0.BD.D0.B4.D0.B5.D0.BD.D1.86.D0.B8.D0.B8](http://www.tadviser.ru/index.php/Статья:Информатизация_регионов_(рынок_России)#.D0.A2.D0.B5.D0.BD.D0.B4.D0.B5.D0.BD.D1.86.D0.B8.D0.B8)(25.04.2017).

5. Гаврилова ТА., Хорошевский В.Ф., 2000. Базы знаний интеллектуальных систем / Учебник для вузов. – СПб, Изд-во «Питер», 2000.

6. Методические аспекты искусственного интеллекта Гузаиров М.Б., Юсупова Н.И., Кудряшев А.Ф., Елхова О.И., Попов Д.В., Ризванов Д.А., Еникеева К.Р., Доломатов М.Ю., Ярцев Р.А., Сметанина О.Н., Верхотуров М.А., Верхотурова Г.Н., Макеев Г.А., Гаянова М.М., Бойко М.В., Богданова Д.Р., Климова А.В., Галямов А.Ф., Орехов Ю.В., Орехов Э.Ю. и др. монография / Уфимский государственный авиационный технический университет, Научный совет РАН по методологии искусственного интеллекта, Рецензенты: профессор В.М. Картак, профессор Д.Е. Сорокин. Москва, 2014.