

Е. В. Долгушева¹, Д. Е. Пальчунов^{1,2}

¹ *Новосибирский государственный университет
ул. Пирогова, 2, Новосибирск, 630090, Россия*

² *Институт математики им. С. Л. Соболева СО РАН
пр. Акад. Коптюга, 4, Новосибирск, 630090, Россия*

palch@math.nsc.ru, katerina.v.shadrina@gmail.com

ТЕОРЕТИКО-МОДЕЛЬНЫЕ МЕТОДЫ ПОРОЖДЕНИЯ ЗНАНИЙ О ПРЕДПОЧТЕНИЯХ АБОНЕНТОВ МОБИЛЬНЫХ СЕТЕЙ *

Статья посвящена разработке методов порождения знаний о том, какие виды тарифов и услуг мобильного оператора могут быть полезными для данного абонента мобильной сети. Порождение знаний производится на основе анализа множества прецедентов – деперсонифицированных профилей абонентов данной мобильной сети. Разрабатываемые методы основаны на теоретико-модельном подходе к формализации предметной области и на методологии анализа формальных понятий. На основе интеграции знаний, извлеченных из профилей абонентов, при использовании описания услуг и тарифов, строится онтологическая модель данной предметной области. Для порождения знаний о тарифах и услугах, которые могут представлять интерес для данного абонента, используются анализ формальных понятий и алгоритмы поиска ассоциативных правил.

Ключевые слова: мобильные сети, абоненты мобильных сетей, онтологическая модель, порождение знаний, теоретико-модельные методы, анализ формальных понятий, ассоциативные правила.

Введение

В настоящее время одним из главных признаков качества работы как предприятия, так и сервиса является степень их ориентированности на клиента. Это, в частности, относится к сервисам и услугам, предоставляемым мобильными операторами. Для привлечения новых клиентов, а также для удержания старых проектируются и внедряются новые тарифные планы и сервисы, проводятся разнообразные акции. Успешные компании работают над тем, чтобы прогнозировать потребности и желания своих клиентов, поддерживать на высоком уровне качество оказываемых услуг.

Мобильные телефоны стали неотъемлемой частью жизни работающего населения, школьников, студентов. Операторы мобильной сети предоставляют своим абонентам воз-

* Исследование выполнено при частичной финансовой поддержке РФФИ в рамках научного проекта № 14-07-00903 а.

Долгушева Е. В., Пальчунов Д. Е. Теоретико-модельные методы порождения знаний о предпочтениях абонентов мобильных сетей // Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии. 2016. Т. 14, № 2. С. 5–16.

возможность быть на связи в любой точке мира, обеспечивают доступ к различным USSD-приложениям и полномасштабный доступ к ресурсам сети Интернет. Для абонентов важны стоимость и качество услуг, удовлетворяющих их индивидуальные запросы.

В условиях высокой конкуренции мобильные операторы предоставляют широкий спектр тарифных планов и сервисов, среди которых абоненту часто бывает сложно выбрать именно те, которые будут для него максимально выгодными и полезными. Более того, сложно следить за появлением новых тарифов и услуг, чтобы иметь возможность своевременно начать пользоваться нововведениями. Часто мобильные операторы производят информационные рассылки о новых сервисах и акциях, но при этом подходе возникают свои сложности. Например, для абонентов, которые не ездят за границу, информация о новых тарифах на звонки в роуминге будет абсолютно ненужной. Поэтому многие абоненты относятся к подобным рассылкам как к спаму.

Целью данной работы является разработка методов, направленных на выявление услуг и тарифных планов операторов мобильной связи, в наибольшей степени удовлетворяющих потребности конкретных абонентов.

В ходе решения данной задачи предложена формализация предметной области на основе теоретико-модельного подхода [1–9]. Разработаны методы выделения ассоциативных правил, позволяющих делать предсказания об услугах и тарифах, которые могут быть наиболее интересны абоненту, на основе множества значений параметров этого абонента.

Обзор существующих решений

Применение методов анализа данных к набору характеристик мобильных абонентов, как правило, преследует одну из следующих целей.

1. Поиск и предотвращение нарушений:

а) выявление характеристик профилей абонентов, которые не собираются оплачивать подключенные услуги;

б) выявление характеристик профилей абонентов, которые намереваются совершить противоправные действия, такие как рассылка спама, телефонное мошенничество и др.

2. Маркетинг:

а) улучшение качества обслуживания клиентов;

б) реклама и эффективное продвижение новых продуктов;

в) раннее выявление абонентов, которые могут уйти от данного мобильного оператора.

3. Оптимизация и поддержание работоспособности инфраструктуры мобильной сети:

а) предсказание распределения нагрузки на различные части сети мобильного оператора;

б) предсказание возможностей (рисков) сбоев в работе сети;

в) планирование направления развития сети для расширения зоны покрытия.

Рассмотрим некоторые работы, посвященные применению методов анализа данных для улучшения качества обслуживания клиентов.

В [10] для анализа рыночной корзины потребителя применялись методы анализа формальных понятий. Результаты, полученные в работе, были использованы для планирования и проведения маркетинговых акций.

Для выявления особенностей поведения абонентов в [11] был применен метод визуализации графа звонков абонентов внутри сети. В результате изучения построенного графа было замечено большое число подгрупп, состоящих из часто взаимодействующих между собой абонентов. Такая особенность в поведении абонентов стала основанием для разработки и внедрения нового тарифного плана, предоставляющего скидки на стоимость разговоров внутри групп.

Целью работы [12] являлась разработка методов увеличения числа клиентов, которые пользуются сервисами мобильного оператора. Для выявления закономерностей в использовании сервисов был применен алгоритм Frequent Pattern-Growth Strategy [13]. В результате были выделены «частые» множества признаков, описывающие зависимости между возрастом абонентов, уровнем их дохода и используемыми сервисами. Варианты оптимизации предлагались экспертами на основе полученных наборов «частых» множеств.

Методы поиска ассоциативных правил, предназначенных для оптимизации структуры меню, предоставляющего доступ к сервисам мобильного оператора, разрабатывались в [14]. Основой этого исследования являлись данные о частоте доступа абонентов к различным сервисам мобильного оператора. С помощью алгоритма Apriori [15] были выделены ассоциативные правила, описывающие закономерности между сервисами, использованными в ходе одной сессии. На основе полученных данных правил были предложены стратегии оптимизации структуры меню.

Работы [16–18] были посвящены изучению понятий нечеткой транзакции, поддержки и достоверности нечеткого ассоциативного правила, а также разработке алгоритмов поиска нечетких ассоциативных правил. В исследовании [19] при помощи разработанного в [11] алгоритма были выделены наборы нечетких правил, описывающих взаимосвязи между популярными телекоммуникационными сервисами, предоставляемыми мобильными операторами Тайваня.

В [20] для определения шаблона поведения абонента был применен анализ формальных понятий. Основой для исследования являлись данные о звонках абонента. На первом этапе был проведен поиск формальных понятий для имеющихся данных, представленных в виде формального контекста. На втором этапе, на основе анализа полученных формальных понятий, производилась классификация абонентов. Полученные классы представляют значительный интерес для мобильного оператора – для оптимизации работы мобильной сети.

Каждый год публикуется большое количество работ, посвященных проблеме выделения высокоуровневых знаний на основе различных параметров абонентов мобильных операторов. Задача предсказания предпочтений пользователей представляет значительный интерес как для исследователей, занимающихся анализом данных, так и для мобильных операторов.

Получение знаний об абоненте мобильной сети происходит в два этапа. На первом этапе к множеству слабоструктурированных данных применяются методы анализа данных. Результатом первого этапа является набор закономерностей, полученных выбранными методами анализа данных. На втором этапе происходит экспертный анализ полученных на первом этапе закономерностей. Конечный результат в форме новых тарифов, сервисов, стратегий маркетинговых мероприятий достигается экспертом в полуавтоматическом режиме.

Описание исходных данных

Разработка метода поиска сервисов, представляющих интерес для абонента, производилась на основе анализа множества деперсонифицированных логов абонентов, содержащих 115 параметров.

Все параметры можно разделить на следующие группы:

1. Личные параметры абонента: пол, возраст, модель мобильного телефона и пр.
2. Параметры, описывающие совершенные абонентом звонки: начисления за звонки, количество исходящих звонков в различные части мира, количество звонков, совершенных в роуминге, и пр.
3. Параметры, описывающие использование WAP протокола (Wireless Application Protocol – беспроводной протокол передачи данных): количество потребленного трафика, начисления за трафик в домашнем регионе, России, СНГ, мире.
4. Параметры, описывающие использование SMS услуги: количество потребленного трафика, начисления за SMS в домашнем регионе, России, СНГ, мире.
5. Параметры, описывающие использование MMS услуги (Multimedia Messaging Service – система передачи мультимедийных сообщений в сетях сотовой связи): количество потребленного трафика, начисления за MMS в домашнем регионе, России, СНГ, мире.
6. Параметры, описывающие использование LBS услуги (Location Based Services – тип информационных и развлекательных услуг, основанных на определении текущего местоположения мобильного телефона): количество потребленного трафика, начисления за LBS трафик в домашнем регионе, России, СНГ, мире.
7. Список услуг, подключенных абонентом: «Конференц-звонки», «Видео-звонки», «Определитель номера», «Оповещение о абоненте в сети» и пр.

Описание используемых методов

Анализ формальных понятий. Анализ формальных понятий (АФП) – один из методов анализа данных. Основные идеи АФП были сформулированы Рудольфом Вилле [21]. Здесь мы кратко приведем основные понятия АФП, которые потребуются нам далее. Более подробно понятия и результаты АФП изложены в [22].

Определение 1. Формальным контекстом называется тройка $K = (G, M, I)$, где G – множество объектов, M – множество признаков, а отношение инцидентности $I \subseteq G \times M$ говорит о том, какие объекты какими признаками обладают.

Определение 2. Рассмотрим формальный контекст $K = (G, M, I)$, пусть $A \subseteq G$ и $B \subseteq M$. Определим операторы Галуа:

$$A' = \{m \in M \mid \text{для любого } g \in A \text{ выполнено } (g I m)\};$$

$$B' = \{g \in G \mid \text{для любого } m \in B \text{ выполнено } (g I m)\}.$$

Оператор $''$ (двукратное применение оператора $'$) является оператором замыкания: он идемпотентен ($A'''' = A''$), монотонен ($A \subseteq B$ влечет $A'' \subseteq B''$) и экстенсивен ($A \subseteq A''$).

Множество объектов $A \subseteq G$, такое что $A'' = A$, называется замкнутым. Множество признаков $B \subseteq M$, такое что $B'' = B$, также называется замкнутым.

Определение 3. Пусть $K = (G, M, I)$ – формальный контекст. Пара множеств (A, B) , такая что $A \subseteq G$ и $B \subseteq M$, называется формальным понятием контекста K тогда и только тогда, когда $A' = B$ и $B' = A$. Множества A и B замкнуты и называются, соответственно, объемом и содержанием формального понятия (A, B) .

Поиск ассоциативных правил. Это метод, часто применяемый совместно с анализом формальных понятий.

В одной из первых формулировок задача поиска ассоциативных правил была предложена для нахождения типичных шаблонов покупок, совершаемых в супермаркетах, поэтому иногда ее еще называют анализом рыночной корзины [15].

Основой для анализа рыночной корзины является множество, состоящее из покупательских транзакций. Каждая транзакция – это набор товаров, купленных покупателем за один визит. Говоря о задаче анализа рыночной корзины в терминах АФП, множество транзакций – это множество объектов, а множество товаров – множество признаков.

Определение 4. Пусть $K = (G, M, I)$ – формальный контекст, G – множество объектов и M – множество признаков. Ассоциативным правилом называется выражение вида $X \rightarrow Y$, где $X, Y \subseteq M$.

Наиболее важными характеристиками ассоциативных правил являются их поддержка и достоверность. Пусть $|X|$ – мощность множества X .

Определение 5. Поддержкой (*support*) ассоциативного правила $X \rightarrow Y$ называется величина $\text{supp}(X \rightarrow Y) = |(X \cup Y)'| / |G|$.

Определение 6. Достоверностью (*confidence*) ассоциативного правила $X \rightarrow Y$ называется величина $\text{conf}(X \rightarrow Y) = |(X \cup Y)'| / |X'|$.

Определение 7. Импликацией формального контекста $K = (G, M, I)$ называется признаковая зависимость вида $X \rightarrow Y$, где $X, Y \subseteq M$, такая, что все объекты, обладающие признаками из X , также обладают признаками из Y , т. е. $X' \subseteq Y'$. В анализе формальных понятий импликация является частным случаем ассоциативного правила: импликация – это в точности такое ассоциативное правило, достоверность которого равна 1.

Имеется ряд алгоритмов поиска ассоциативных правил для произвольных наборов признаков и объектов. В данной работе мы используем алгоритм Apriori [15]. В настоящее время он является одним из наиболее популярных алгоритмов поиска ассоциативных правил. Для данного исследования была выбрана реализация алгоритма Apriori, предоставляемая пакетом *arules* языка R. Согласно работам [23, 24], реализация алгоритма Apriori в пакете *arules* позволяет, в частности, эффективно решать задачи, поставленные в ходе данного исследования.

Методы представления знаний об абонентах мобильной связи

Онтологическая модель предметной области. В данном разделе мы рассмотрим теоретико-модельный подход к определению понятия онтологической модели. Этот подход является развитием теоретико-модельного подхода к разработке онтологий предметных областей [1–4].

Для структурирования знаний о предметной области мобильной связи мы используем четырехуровневую модель представления знаний, реализованную в виде онтологической модели предметной области. Онтологическая модель состоит из онтологии и еще трех уровней представления знаний о предметной области: общих (теоретических) знаний, эмпирических знаний и оценочных (вероятностных) знаний (см. рисунок) [8].



Четырехуровневое представление знаний в онтологической модели

По аналогии с работой [8] построим описание четырехуровневой модели представления знаний рассматриваемой предметной области.

Уровень 1. Онтология предметной области содержит набор ключевых понятий, на языке которых описывается данная предметная область, а также определения этих понятий. Онтология предметной области мобильной связи включает в себя множество терминов, специфичных для данной предметной области, таких как «роуминг», «звонок», «смс», «тариф» и др. названия услуг и сервисов, предоставляемых мобильным оператором, термины, используемые оператором для описания особенностей абонента.

Описание предметной области содержит еще три уровня представления знаний.

Уровень 2. Универсальные, общие утверждения – законы и постулаты предметной области. Это знания о предметной области: общих принципах, законах и закономерностях, которые на данный момент времени считаются полностью достоверными. Эти знания могут изменяться (т. е. истинные утверждения становиться ложными, а ложные – истинными) даже в том случае, когда смысл употребляемых терминов остался полностью неизменным. В рассматриваемом случае универсальное знание является истинным для всех абонентов данного мобильного оператора.

Примером такого рода знаний является детализированное описание тарифов и сервисов. Эти знания общедоступны, их актуальное состояние может быть получено на сайте мобильного оператора. Например, в 2015 г. тариф N студенческий предоставлял 2 Гб интернет-трафика и 500 смс в месяц за ежемесячную абонентскую плату 100 руб. Начиная с 1 января 2016 г., тариф изменился, и теперь предоставляется 5 Гб трафика и 200 смс за ежемесячную абонентскую плату 300 руб.

Описание сервиса состоит из набора {название услуги, объем предоставляемой услуги, период действия сервиса, абонентская плата за данный период, абонентская плата за подключение услуги}. Описание тарифов состоит из набора {множество наборов {название услуги, объем предоставляемой услуги}, период действия тарифа, абонентская плата за период, абонентская плата за подключение тарифа}.

Уровень 3. Эмпирические данные содержат описание конкретных прецедентов предметной области. В нашем случае прецедентом является профиль абонента. Формальное описание каждого прецедента представляется в виде фрагмента атомарной диаграммы алгебраической системы. При этом, кроме сигнатурных символов, соответствующих реальным понятиям предметной области, вводятся дополнительные, служебные сигнатурные символы [9]. Эти символы обозначают неизвестные объекты, действия и отношения, т. е. это специальные символы констант и предикатов. Таким образом, на третьем уровне представлено множество профилей абонентов.

Уровень 4. Вероятностные и оценочные знания. Эти знания либо берутся из внешних источников, либо порождаются на основе знаний, представленных в онтологической модели. Таким образом, на основе анализа имеющихся в онтологической модели эмпирических данных и сопоставления их с универсальными знаниями и онтологическими (аналитическими) знаниями, также представленными в онтологической модели, порождаются вероятностные и оценочные знания и закономерности.

Ассоциативные правила и импликации, полученные нами на этапе анализа данных, попадают на четвертый уровень онтологической модели. Следует отметить, что получаемую на основе результатов анализа информацию можно поделить на два подуровня.

1. Уровень данных. Закономерности, полученные полностью автоматическим способом, например, наборы ассоциативных правил с соответствующими значениями поддержки и достоверности.

2. Уровень знаний. Здесь закономерности получаются в полуавтоматическом режиме. Окончательные выводы делаются экспертами предметной области, исходя из семантики, на основе информации, представленной на уровне данных (1).

Формальный контекст. По аналогии с [9] построим модель предметной области мобильной связи.

В данной работе рассматривается предметная область мобильной связи, содержащая сведения об абонентах с фиксированным набором параметров. В формализации данной предметной области используются только свойства, описывающие конкретных абонентов. Таким образом, мы рассматриваем сигнатуру, состоящую из конечного числа одноместных предикатных символов. Сигнатуру предметной области будем обозначать через σ_M . Предикаты $P(x) \in \sigma_M$ будем называть сигнатурными.

Для определения сигнатуры σ_M предметной области $M = \text{«Сети мобильной связи»}$ мы рассматриваем два набора признаков: σ_P – индивидуальные показатели абонента и σ_Q – набор тарифных планов и услуг. Набор признаков σ_P состоит из двух частей: σ_{P_1} – «трафики» и σ_{P_2} – «начисления». Набор σ_Q состоит из двух частей, имеющих иерархическую структуру: σ_{Q_1} – «тарифные планы» и σ_{Q_2} – «услуги и опции». σ_{Q_1} состоит из символов одноместных предикатов, каждый из которых описывает наличие либо отсутствие у абонента подключения к тому или иному тарифу. σ_{Q_2} состоит из символов одноместных предикатов, каждый из которых описывает наличие или отсутствие у абонента подключения к определенной услуге и опции этой услуги.

Данными, лежащими в основе исследования, является конечное множество $\{a_1, \dots, a_n\}$ профилей абонентов. Каждый профиль a_i – это множество (набор) значений параметров абонента. Следовательно, для каждого профиля мы можем описать множество признаков (сигнатурных предикатов), истинных для этого абонента. Пусть c_a – персонифицированный идентификатор абонента. Таким образом, по каждому профилю абонента a_i мы можем

построить одноэлементную модель $a_i = (\{a_i\}, \sigma \cup c_{a_i})$, которую будем называть прецедентом предметной области (здесь $\sigma = \sigma_{\mathbb{M}}$). Класс всех таких одноэлементных моделей сигнатуры σ обозначим как

$$K_1(\sigma) = \{ \langle \{a\}, \sigma \rangle \mid a \in |\mathfrak{A}| \}.$$

На множестве прецедентов $A = \{a_1, \dots, a_n\}$ мы определяем структуру онтологической модели $\mathfrak{A} = \langle A, \sigma \rangle$. В онтологической модели \mathfrak{A} для каждого сигнатурного предиката $P(x) \in \sigma$ и для каждого прецедента $a_i \in A$ выполнено $\mathfrak{A} \models P(a_i)$ тогда и только тогда, когда $a_i \models P(a_i)$.

В рамках теоретико-модельного подхода формальный контекст мы задаем при помощи класса моделей $K_{\mathfrak{A}} \subseteq K_1(\sigma)$ фиксированной сигнатуры $\sigma = \sigma_{\mathbb{M}}$ и некоторого множества предложений $\Delta \subseteq S(\sigma)$ этой же сигнатуры. Формальный контекст представляется как упорядоченная тройка $(K_{\mathfrak{A}}, \Delta, \models)$ [9].

В качестве множества Δ начально мы рассматриваем множество Δ_{σ} атомарных формул вида $P(c_a)$, где $P(x) \in \sigma$. Таким образом, в качестве базового мы рассматриваем формальный контекст $(K_{\mathfrak{A}}, \Delta_{\sigma}, \models)$.

Порождение ассоциативных правил для определения тарифов и услуг, которые могут представлять интерес для абонента

Нахождение ассоциативных правил для исходного контекста $(K_{\mathfrak{A}}, \Delta_{\sigma}, \models)$ дает немного правил с высокой достоверностью. Это связано в том числе с тем, что абоненту может быть полезна какая-то одна из нескольких схожих по содержанию услуг. То, какую именно услугу выберет абонент, зависит от многих обстоятельств. Некоторые из этих обстоятельств могут быстро меняться, поэтому они не могут быть отслежены в рамках формального контекста $(K_{\mathfrak{A}}, \Delta_{\sigma}, \models)$, который отражает достаточно длительную историю поведения абонентов.

При этом оператор может предложить абоненту 2–3 услуги, из которых абонентом будет сделан конкретный выбор после анализа всех деталей. Поэтому к множеству параметров абонента Δ_{σ} целесообразно добавить все дизъюнкции сигнатурных предикатов из Δ_{σ} – а именно, множество предложений Δ_{σ}^{\vee} . Таким образом, от формального контекста $(K_{\mathfrak{A}}, \Delta_{\sigma}, \models)$ мы переходим к формальному контексту $(K_{\mathfrak{A}}, \Delta_{\sigma}^{\vee}, \models)$.

При работе с ассоциативными правилами, найденными в расширенном формальном контексте $(K_{\mathfrak{A}}, \Delta_{\sigma}^{\vee}, \models)$, возникают две проблемы. Во-первых, выявляются ассоциативные правила, имеющие высокую достоверность, но заключения которых – дизъюнкции содержательно никак не связанных между собою услуг. Такие ассоциативные правила нельзя использовать для выдачи рекомендаций абонентам – подобные рекомендации будут восприниматься ими как спам. Становятся необходимыми весьма затратная ручная отбраковка и выделение «разумных» ассоциативных правил. Во-вторых, обработка всего формального контекста $(K_{\mathfrak{A}}, \Delta_{\sigma}^{\vee}, \models)$ является очень трудоемкой вычислительной процедурой.

Для решения этих проблем к сигнатуре $\sigma = \sigma_{\mathbb{M}}$ мы добавляем новые одноместные предикаты, с помощью которых можно выразить смысл тарифов и услуг. При помощи этих новых предикатов, представленных сигнатурой $\sigma_{\mathbb{R}}$, происходит автоматическое порождение содержательных дизъюнкций исходных предикатов из сигнатуры $\sigma_{\mathbb{M}}$. Мы рассматриваем множество $\Delta = \Delta_{\sigma_{\mathbb{M}} \cup \sigma_{\mathbb{R}}}$ и формальный контекст $C_{\mathfrak{A}}^{\Delta} = (K_{\mathfrak{A}}, \Delta, \models)$.

В формальном контексте $C_{\mathfrak{A}}^{\Delta} = (K_{\mathfrak{A}}, \Delta, \models)$ мы ищем ассоциативные правила, отвечающие следующим условиям:

- 1) множество посылок правила содержится в множестве $\Delta \upharpoonright \sigma_{\mathfrak{M}}$;
- 2) заключение правила является элементом множества $\Delta \upharpoonright \sigma_{\mathfrak{R}}$;
- 3) поддержка и достоверность правила выше заданных порогов.

Далее происходит автоматическая обработка полученных ассоциативных правил. Ассоциативное правило с одноэлементным заключением P , принадлежащим множеству $\sigma_{\mathfrak{R}}$, преобразуется в ассоциативное правило с той же посылкой, заключением которого будет формула от одной переменной, принадлежащая множеству Δ_{σ}^{\vee} , – дизъюнкция предикатов $Q_i \in \sigma_{\mathfrak{M}}$, которые являются следствиями предиката P .

Таким образом, программная система позволяет автоматически производить поиск ассоциативных правил для контекста $(K_{\mathfrak{A}}, \Delta_{\sigma}^{\vee}, \models)$. В результате экспериментов было установлено, что расширение содержания контекста предикатами из $\sigma_{\mathfrak{R}}$ позволяет порождать ассоциативные правила с высокой достоверностью и поддержкой. Заключение таких правил затем преобразуется в дизъюнкции предикатов из $\sigma_{\mathfrak{M}}$. В результате получаются ассоциативные правила, представляющие практический интерес для оператора мобильной связи.

В качестве примера рассмотрим набор ассоциативных правил, заключениями которых являются тарифы, предоставляющие доступ к сети Интернет. Предикат $P(x) \in \sigma_{\mathfrak{R}}$ обозначает наличие в тарифе абонента безлимитного Интернета определенного типа. Такой тип безлимитного Интернета имеют только тарифные планы $Q_1, Q_2, Q_3 \in \sigma_{\mathfrak{Q}_1} \subseteq \sigma_{\mathfrak{M}}$.

Примеры ассоциативных правил

	Правило	Поддержка, %	Достоверность, %
Пример 1	$\{P_1'', \dots, P_l''\} \rightarrow P$	1	94
	$\{P_1'', \dots, P_l''\} \rightarrow Q_1$	0,1	11
	$\{P_1'', \dots, P_l''\} \rightarrow Q_2$	0,2	25
	$\{P_1'', \dots, P_l''\} \rightarrow Q_3$	0,6	60
Пример 2	$\{P_1''', \dots, P_l'''\} \rightarrow T$	2	91
	$\{P_1''', \dots, P_l'''\} \rightarrow T_1$	1	51
	$\{P_1''', \dots, P_l'''\} \rightarrow T_2$	1	39
	$\{P_1''', \dots, P_l'''\} \rightarrow T_3$	0,8	48

В ходе эксперимента найдены ассоциативные правила с посылками, имеющими различные наборы индивидуальных характеристик пользователя из $\sigma_{\mathfrak{P}} \subseteq \sigma_{\mathfrak{M}}$, и заключением $P(x)$. Выбраны правила, имеющие достаточно высокие достоверность и поддержку (см.

таблицу, пример 1)¹. Затем в заключении ассоциативных правил предикат $P(x)$ заменяется на эквивалентную ему дизъюнкцию $(Q_1 \vee Q_2 \vee Q_3)$. Из таблицы видно, что замена дизъюнкции предикатов $(Q_1 \vee Q_2 \vee Q_3)$ на любой из этих предикатов Q_i существенно уменьшает как достоверность, так и поддержку ассоциативных правил.

Пример правила, имеющего высокое значение достоверности, но не подходящего для применения в качестве рекомендации, дает Пример 2 из таблицы: $T = (T_1 \vee T_2 \vee T_3) \in \Delta_\sigma^\vee$. При этом признаки T_i представляют собой разнородные сервисы, предоставляющие, соответственно: T_1 – скидки на доступ в сеть Интернет, T_2 – скидки на звонки на избранные номера, а T_3 – возможность звонков при отрицательном балансе. Формально правило $\{P_1'', \dots, P_l''\} \rightarrow T$ соответствует всем критериям. Однако оно таковым не является, поскольку, если им воспользоваться, абоненту будет предложен выбор из совершенно не связанных друг с другом услуг, что будет воспринято абонентом как спам.

Таким образом, новые ассоциативные правила, порождаемые нашим алгоритмом в расширенном формальном контексте $(K_\Delta, \Delta_\sigma^\vee, \models)$, характеризуются более высокими значениями поддержки и достоверности, по сравнению с правилами с той же посылкой, порождаемыми в исходном формальном контексте $(K_\Delta, \Delta_\sigma, \models)$. При этом, что принципиально важно, заключения полученных таким образом ассоциативных правил являются полностью осмысленными как для оператора мобильной связи, так и для абонентов.

Заключение

В результате проведенного исследования была построена онтологическая модель, описывающая поведение абонентов мобильной связи. Разработаны методы порождения ассоциативных правил, позволяющих делать выводы о том, какие из тарифов и сервисов, уже предоставляемых оператором мобильной связи, являются наиболее предпочтительными для абонентов. В дальнейшем предполагается расширить методы анализа поведения абонентов мобильной связи за счет введения в рассмотрение более сложных признаков, описываемых бескванторными формулами логики предикатов первого порядка.

Список литературы

1. *Palchunov D.* Lattices of Relatively Axiomatizable Classes. // ICFCA 2007. 2007. Vol. LNAI 4390. P. 221–239.
2. *Пальчунов Д. Е.* Решение задачи поиска информации на основе онтологий // Бизнес-информатика. 2008. № 1. С. 3–13.
3. *Пальчунов Д. Е.* Моделирование мышления и формализация рефлексии. II: Онтологии и формализации понятий // Философия науки. 2008. № 2 (37). С. 62–99.
4. *Palchunov D. E.* Virtual catalog: the ontology-based technology for information retrieval // Knowledge Processing and Data Analysis. LNAI 6581. Berlin; Heidelberg: Springer-Verlag, 2011. P. 164–183.
5. *Пальчунов Д. Е., Степанов П. А.* Применение теоретико-модельных методов извлечения онтологических знаний в предметной области информационной безопасности // Программная инженерия. 2013. № 11. С. 8–16.
6. *Махасоева О. Г., Пальчунов Д. Е.* Автоматизированные методы построения атомарной диаграммы модели по тексту естественного языка // Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии. 2014. Т. 12, № 2. С. 64–73.

¹ В силу конфиденциальности информации мы не можем указать, какие именно характеристики P_i абонентов были выбраны.

7. Деревянко Д. В., Пальчунов Д. Е. Формальные методы разработки вопросно-ответной системы на естественном языке // Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии. 2014. Т. 12, № 3. С. 34–47.
8. Найданов Ч. А., Пальчунов Д. Е., Сазонова П. А. Теоретико-модельные методы интеграции знаний, извлеченных из медицинских документов // Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии. 2015. Т. 13, № 3. С. 29–41.
9. Пальчунов Д. Е., Яхьяева Г. Э., Ясинская О. В. Применение теоретико-модельных методов и онтологического моделирования для автоматизации диагностирования заболеваний // Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии. 2015. Т. 13, № 3. С. 42–51.
10. Ustundag A., Bal M. Evaluating Market Basket Data with Formal Concept Analysis // Proc. Chaos, Complexity and Leadership, 2012.
11. Han J., Pregibon D., Mannila H., Kumar V., Altman R. B. Emerging scientific applications in data mining // Communications of the ACM – Evolving Data Mining into Solutions for Insights. 2002. Vol. 45, No. 8. P. 54–58.
12. Surendiran R., Rajan K. P., Sathish Kumar M. Study on the Customer targeting using Association Rule Mining. // International Journal on Computer Science and Engineering. 2010. Vol. 2, No. 7. P. 2483–2484.
13. Han J., Pei J., Yin Y. Mining Frequent Patterns without Candidate Generation // Data Mining and Knowledge Discovery. 2004. Vol. 8, No. 1. P. 53–87.
14. Pravin A. P., Aggarwal A. K. Associative Rule Mining of Mobile Data Services Usage for Preference Analysis, Personalization & Promotion // Proc. WSEAS. 2004.
15. Agrawal R., Imieliński T., Swami A. Mining association rules between sets of items in large databases // Proc. ACM SIGMOD International Conference on Management of Data. 1993. P. 207–216.
16. Bosc P., Pivert O., Prade H. On fuzzy association rules based on fuzzy cardinalities // Proc. the 10th IEEE International Conference. 2001. P. 461–464.
17. Chueh H.-E., Lin N. P., Jan N.-Y. Mining Target-oriented Fuzzy Correlation Rules // Proc. International Conference on Advances in Social. 2009.
18. Lin N. P., Chueh H.-E. Fuzzy Correlation Rules Mining // Proc. 6th WSEAS International Conference on Applied Computer Science. 2007.
19. Chueh H.-E. Mining target-oriented fuzzy correlation rules to optimize telecom service management // International Journal of Computer Science & Information Technology. 2011. Vol. 3, No.1. P. 74–83.
20. Furletti B., Gabrielli L., Renso C., Rinzivillo S. Analysis of GSM calls data for understanding user mobility behavior // IEEE Big Data International Conference. 2013. P. 550–555.
21. Wille R. Restructuring Lattice Theory: an Approach Based on Hierarchies of Concepts / Ed. by I. Rival // Ordered Sets. Dordrecht. Boston: Reidel, 1982. P. 445–470.
22. Ganter B., Wille R. Formal Concept Analysis: Mathematical Foundations. Heidelberg: Springer, 1999.
23. Borgelt C. Efficient Implementations of Apriori and Eclat // Proc. 1st IEEE ICDM Workshop on Frequent Item Set Mining Implementations. 2003.
24. Agrawal R., Srikant R. Fast Algorithms for Mining Association Rules in Large Databases // Proc. 20th International Conference on Very Large Data Bases. 1994. P. 487–499.

E. V. Dolgusheva¹, D. E. Palchunov²

¹ *Novosibirsk State University*

2 Pirogov Str., Novosibirsk, 630090, Russian Federation

² *Sobolev Institute of Mathematics Siberian Branch of the Russian Academy of Sciences*

4 Acad. Koptug Ave., Novosibirsk, 630090, Russian Federation

palch@math.nsc.ru, katerina.v.shadrina@gmail.com

MODEL-THEORETIC METHODS OF GENERATION OF KNOWLEDGE ABOUT MOBILE SUBSCRIBERS' PREFERENCES

The article is devoted to methods of generation of knowledge about types of tariffs and services of mobile operator that might be useful for a given mobile network subscriber. We provide knowledge generation on the base of analysis of the set of precedents – impersonal mobile network subscriber profiles. These methods are based on the model-theoretic approach to domain formalization and on Formal Concept Analysis. The Ontological Model of the domain is constructed on the base of integration of knowledge extracted from users' profiles and descriptions of existing tariffs and services. Formal concept analysis and association rules mining are using for generation of knowledge about tariffs and services that might be interesting for mobile network subscribers.

Keywords: mobile networks, subscribers of mobile networks, ontology model, generation of knowledge, model-theoretic methods, formal concept analysis, association rules.

References

1. Pal'chunov D. Lattices of Relatively Axiomatizable Classes // ICFCA 2007, Vol. LNAI 4390, 2007, p. 221–239.
2. Palchunov D. E. The solution of the problem of information retrieval based on ontologies // *Bisnes-informatika*, 2008, no. 1, p. 3–13 (in Russ.).
3. Palchunov D. E. Modeling of reasoning and formalization of reflection II: Ontologies and formalization of concepts // *Filosofiya nauki*, 2008, no. 2 (37), p. 62–99 (in Russ.).
4. Palchunov D. E. Virtual catalog: the ontology-based technology for information retrieval // *Knowledge Processing and Data Analysis*. LNAI 6581. Springer-Verlag Berlin Heidelberg. 2011, p. 164–183.
5. Palchunov D. E., Stepanov P. A. The use of model-theoretic methods for extracting ontological knowledge in the domain of information security // *Programnaya ingeneriya*, 2013, no. 11, p. 8–16. (in Russ.)
6. Makhasoeva O. G., Palchunov D. E. Semi-automatic methods of a construction of the atomic diagrams from natural language texts // *Vestnik NSU, series: Informacionnye tehnologii*, 2014, vol. 12, no. 2, p. 64–73. (in Russ.)
7. Derevyanko D. V., Palchunov D. E. Formal methods of development of the question-answering system on natural language // *Vestnik NSU, series: Informacionnye tehnologii*, 2014, vol. 12, no. 3, p. 34–47 (in Russ.).
8. Naidanov C. A., Palchunov D. E., Sazonova P. A. Model-theoretic methods of integration of knowledge extracted from medical documents // *Vestnik NSU Series: Information Technologies*. 2015, vol. 13, Iss. 3, p. 29–41. ISSN 1818-7900. (in Russ.)
9. Palchunov D. E., Yakhyyayeva G. E., Yasinskaya O. V. Application of model-theoretic methods and ontological modeling to automate the diagnosis of diseases // *Vestnik NSU Series: Information Technologies*, 2015, vol. 13, Issue 3, p. 42–51. ISSN 1818-7900. (in Russ.)
10. Ustundag A., Bal M. Evaluating Market Basket Data with Formal Concept Analysis // *Proc. Chaos, Complexity and Leadership*, 2012.
11. Han J., Pregibon D., Mannila H., Kumar V., Altman R. B. Emerging scientific applications in data mining // *Communications of the ACM – Evolving data mining into solutions for insights*. 2002, vol. 45, no. 8, p. 54–58.

12. Surendiran R., Rajan K. P., Sathish Kumar M. Study on the Customer targeting using Association Rule Mining // *International Journal on Computer Science and Engineering*, 2010, vol. 2, no. 7, p. 2483–2484.
13. Han J., Pei J., Yin Y. Mining Frequent Patterns without Candidate Generation // *Data Mining and Knowledge Discovery*, 2004, vol. 8, no. 1, p. 53–87.
14. Pravin A. P., Aggarwal A. K. Associative Rule Mining of Mobile Data Services Usage for Preference Analysis, Personalization & Promotion // *Proc. WSEAS*. 2004.
15. Agrawal R., Imieliński T., Swami A. Mining association rules between sets of items in large databases // *Proc. ACM SIGMOD International conference on Management of data*. 1993, p. 207–216.
16. Bosc P., Pivert O., Prade H., On fuzzy association rules based on fuzzy cardinalities // *Proc. The 10th IEEE International Conference*. 2001, p. 461–464.
17. Chueh H.-E., Lin N. P., Jan N.-Y. Mining Target-oriented Fuzzy Correlation Rules // *Proc. International Conference on Advances in Social*. 2009.
18. Lin N. P., Chueh H.-E. Fuzzy Correlation Rules Mining // *Proc. 6th WSEAS International Conference on Applied Computer Science*. 2007.
19. Chueh H.-E. Mining target-oriented fuzzy correlation rules to optimize telecom service management // *International Journal of Computer Science & Information Technology*, 2011, vol. 3, no. 1, p. 74–83.
20. Furletti B., Gabrielli L., Renso C., Rinzivillo S. Analysis of GSM calls data for understanding user mobility behavior // *IEEE Big Data International Conference*. 2013, p. 550–555.
21. Wille R. Restructuring Lattice Theory: an Approach Based on Hierarchies of Concepts // *Ordered Sets* / Ed. by I. Rival. Dordrecht; Boston: Reidel. 1982. P. 445–470.
22. Ganter B., Wille R. Formal Concept Analysis: Mathematical Foundations // Heidelberg: Springer, 1999.
23. Borgelt C. Efficient Implementations of Apriori and Eclat // *Proc. 1st IEEE ICDM Workshop on Frequent Item Set Mining Implementations*. 2003.
24. Agrawal R., Srikant R. Fast Algorithms for Mining Association Rules in Large Databases // *Proc. 20th International Conference on Very Large Data Bases*, 1994, p. 487–499.