

Abstract

The article describes the models of signals scattered by atmospheric inhomogeneities. These models are used in simulation modeling to test the methods of signal processing in radar systems of atmosphere vertical sounding, as well as during the development of assessment methods of the scattered signals parameters and methods of inhibition of active and passive noises. The application of a more complex model, which describes the properties of the scattered signals, allows reducing the errors of assessment of meteorological parameters by radar systems of atmosphere vertical sounding. The complication of models and methods of processing put the problems of optimization of computational procedures of signal processing. One of the approaches that achieve the required performance under these conditions is the use of parallel computing systems - universal and graphics processor, and clusters. This approach is discussed in this article. The main purpose is to study the possibility of effective implementation of procedures for processing of radar signals based on parallel computation systems. In the study we used the method of simulation modeling and statistical processing of the results. The signals processing methods, which used during their development a physical model of scattering, provide lesser value of the error of assessment of the parameters of atmosphere. The research results can be used in development and modernization of processing systems of radar signal of atmosphere vertical sounding. The application of graphic processors for signal processing allows obtaining the results in real time using a physical model of scattering - the most detailed model of the formation of signals in radar systems of vertical sounding

Keywords: radar atmosphere sounding, digital processing of radar signals, models of scattered signals

Розглядається процес аналізу трафіків телефонних мереж в інтелектуальних системах підтримки прийняття рішень. Для дослідження обрана задача прогнозування зв'язків і її модифікації. Запропонований підхід до розв'язку задачі прогнозування тимчасових зв'язків, заснований на моделюванні мережі абонентів динамічним графом і виявленні знань про його динаміку, що використовує методи кластеризації, секвенціального аналізу, дерев рішень

Ключові слова: трафік телефонної мережі, прогнозування зв'язків, динамічний граф, виявлення знань

Рассматривается процесс анализа трафиков телефонных сетей в интеллектуальных системах поддержки принятия решений. Для исследования выбрана задача прогнозирования связей и ее модификации. Предложен подход к решению задачи прогнозирования временных связей, основанный на моделировании сети абонентов динамическим графом и обнаружении знаний о его динамике, использующий методы кластеризации, секвенциального анализа, деревьев решений

Ключевые слова: трафик телефонной сети, прогнозирование связей, динамический граф, обнаружение знаний

УДК 316.77:[004.82+004.89]

ПОСТАНОВКА ЗАДАЧИ ИССЛЕДОВАНИЯ ПРОГНОЗИРОВАНИЯ СВЯЗЕЙ В ТРАФИКАХ ТЕЛЕФОННЫХ СЕТЕЙ

О.О. Савельев

Аспирант*

Контактный тел.: 066-706-86-70

E-mail: savelyev.oleg@gmail.com

А.И. Шевченко

Член-корреспондент НАН Украины, доктор

технических наук, профессор, заведующий кафедрой*

Контактный тел.: (062) 304-60-82

E-mail: info@iai.donntu.edu.ua

*Кафедра программного обеспечения интеллектуальных систем

Институт информатики и искусственного интеллекта

Донецкий национальный технический университет

пр. Б. Хмельницкого, 84, г. Донецк, Украина, 83050

1. Введение

В настоящее время человечество пользуется большим количеством электронных средств, которые ведут протоколирование всех осуществляемых с их

помощью действий. Среди устройств, предназначенных для приема/передачи информации в сетях, такой протокол обозначается термином *трафик* или мониторинг. В области мобильной и стационарной телефонии существует понятие *трафик телефонной сети*

(ТТС). Зарубежный аналог термина ТТС – call detail records (CDR) [13]. ТТС – это данные о телефонных звонках и других событиях, которые автоматически создаются во время происхождения события и хранятся оператором связи [31].

Подобные данные включают номер вызывающего абонента (инициатора события) и номер принимающего абонента (участника события), IMEI аппаратов, дату, время, длительность события и идентификаторы базовых станций, обслуживающих абонентов во время события, но не включают содержимое сообщений [15, 19, 40].

ТТС можно рассматривать как информацию различной природы. Благодаря наличию в каждой записи идентификатора базовой станции в некотором приближении ТТС возможно интерпретировать как треки – записи маршрутов перемещений [23, 38]. С другой стороны трафики содержат точное время, длительность и порядок действий, т.е. являются отображением распорядка дня абонента [24, 27]. Кроме того, ТТС включают подмножество событий, генерируемое сообществом абонентов, а если рассматривать конкретного абонента – то подмножеством объектов его социальной сети, составляющих социальный граф по критерию использования услуг телефонии [14]. Поэтому ТТС могут рассматриваться как отображение коммуникаций социальной сети [25, 32]. В своем обобщении, учитывая происхождение и способ создания, трафики являются телекоммуникационными (телеметрическими) данными [1, 4].

Объектом исследования данной работы является процесс анализа трафиков телефонных сетей. “Анализ трафиков – это не криптоанализ, а основа коммуникационного интеллекта” [21]. Конечно, анализ трафиков по уровню ценности получаемых знаний находится уровнем ниже криптоанализа, но его конечная стоимость для аналитика также ниже, так как он поддается автоматизации в значительной степени. По этой причине анализ трафиков часто используется для “выбора цели” и дальнейшего тщательного наблюдения.

Непосредственное общение с экспертами предметной области показало, что анализ трафиков – одно из наиболее перспективных направлений решения специальных задач оперативных подразделений правоохранительных структур. Но существующее положение данного направления требует вовлечения технических и научных инвестиций. С позиции правовых оснований для анализа ТТС существует несколько сценариев использования: независимый анализ, самоанализ лицами – производителями ТТС либо доверенными лицами, анализ уполномоченными организациями. Отдельный сценарий – проведение научного эксперимента.

В области анализа ТТС существует ряд проблем:

- поиск связей между любыми объектами (событиями, людьми, группами людей, и т. п.) [42, 40];
- поиск латентных объектов; обнаружение аномальных ситуаций [4];
- профилирование абонентов [37], событий [27], географических локаций [38], временных сегментов [14];
- прогнозирование действий абонентов – установление связей [32, 36], перемещение [22];

– определение сообществ [39], ролей и позиций пользователей [27] в социальной сети;

– визуализация для поддержки анализа [2, 3, 9, 15, 16, 25, 43].

На настоящий момент эти проблемы частично решаются ручным либо частично автоматизированным способом с применением аналитического ПО, однако зачастую аналитику тяжело учитывать множество всех доступных данных, соответственно полученные решения часто являются не полными, а иногда – неверными, не всегда удается не только обнаружить новые, ранее не известные факты-тренды, а и достоверно обосновать проверяемые версии-гипотезы. “Невозможно принять правильное решение, не основываясь на теории” [34].

Избежать подобной ситуации возможно применяя системы поддержки принятия решений (СППР), предназначенные для увеличения эффективности работы лица принимающего решения (ЛПР) (рис. 1). В данной предметной области СППР может быть основана на методах интеллектуального анализа данных и анализа социальных сетей, которые являются подмножеством методов искусственного интеллекта, что позволяет говорить об интеллектуальной СППР (ИСППР). Как правило, такая система обладает двумя свойствами: накопленные данные об исследуемом объекте используются для извлечения знаний о характере поведения объекта, а в процессе анализа новых данных используются ранее приобретенные знания [4].

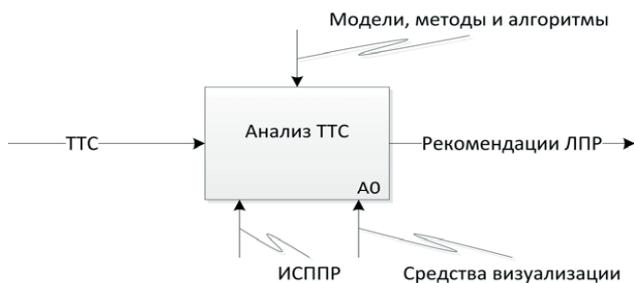


Рис. 1. Процесс анализа ТТС с использованием ИСППР

Соответственно в качестве предмета исследования выделим модели, методы и алгоритмы, которые могут быть использованы в основе ИСППР при анализе ТТС.

В работе [6] рассмотрены концептуальные проблемы разработки интеллектуальных СППР в области анализа телекоммуникационных данных.

Цель данной работы – увеличить качество рекомендаций, выдаваемых ИСППР при анализе ТТС через полное фокусирование на одной конкретной, но существенной задаче, постановку дальнейшего исследования которой требуется осуществить. Для достижения цели необходимо выполнить следующие задачи: провести обзор и анализ работ известных ученых в области анализа ТТС; на основе чего выбрать наиболее актуальную задачу, процесс решения которой возможно реализовать в качестве моделей, методов и алгоритмов СППР; дать формальную постановку выбранной задачи, определить вероятный способ и допустимые методы ее решения.

2. Обзор и анализ

В соответствии с перечисленными ранее проблемами проведем обзор существующих работ. На рис. 2 показана интеллект-карта проблематики предметной области. Отмечены основные проблемы объекта исследования, задачи их составляющие, а также ассоциативные связи между ними.

В проблеме *анализа связей* используется одноименный подход – анализ связей (link analysis). Анализ связей – графический инструмент визуализации ассоциаций между сущностями и событиями, применяется для первичного ознакомления с данными, позволяет графически устанавливать наличие связей между объектами и локациями по физическим контактам или сетевой коммуникации. Как правило, анализ связей начинается с вопроса “кто кого знает и где и когда они встречались?” [37]. Основой анализа связей является теория графов [13]. Анализ связей полезен при ручном поиске аномалий экспертом. Базовая задача анализа связей – классификация отношений между объектами на слабые и сильные (матрица отношений) по количеству событий коммуникации [34], на основе чего строятся диаграммы связей [8].

Другие задачи – минимизация графов для идентификации внешне скрытых связей, восстановление графа по его компонентам [42].

Проблема *обнаружения аномалий* заключается в поиске редких событий. В обнаружении аномалий существуют оптимизационные задачи – уменьшение ложных определений – негативных и позитивных. Часто решение о наличии аномалии принимается экспертом на основе визуального представления данных [16]. В работе [15] применяются методы визуализации для обнаружения типичных и аномальных

паттернов звонков. А в работе [25] рассмотрено структурное представление динамического графа социальной сети с помощью стандартного шаблона – компоненты максимально минимизирующей состояния исходного графа. Все компоненты, не вошедшие в стандартный шаблон, могут рассматриваться как потенциальные аномалии для каждого состояния сети.

Проблема *профилирования* заключается в возможности распознавать паттерны определенной активности, прогнозировать, когда и где вероятно произойдет определенное событие и идентифицировать участников [37]. В соответствии с задачами проблемы профилирования выделяют три типа анализа ТТС с целью построения профилей: географический, временной, социальный [16]. Большое внимание проблеме было уделено в крупном проекте Массачусетского технологического института – проекте майнинга реальности (MIT Reality Mining Project) [10, 23, 24, 26-29, 31]. Майнинг реальности – измерение человеческих отношений и поведения с применением интеллектуального анализа данных к ТТС [31]. Проект включал в себя эксперимент мониторинга активности более ста пользователей смартфонов (студенты и сотрудники Media Lab). Мониторинг заключался в записи данных трафиков протоколов GSM и Bluetooth, кроме того, пользователи давали словесное описание своему местоположению по запросу ПО эксперимента, а также отмечали характер отношений (друг или коллеги) с другими пользователями. В работах [23, 24, 27] предлагается осуществлять построение профилей пользователей, согласно времени – в некотором роде формирование распорядка дня. Выделяются группы схожих распорядков. Работы [10, 22, 29, 38] посвящены обнаружению паттернов перемещения. Обобщенно предлагаемые подходы заключаются в кластеризации смежных частей маршрутов. Для но-

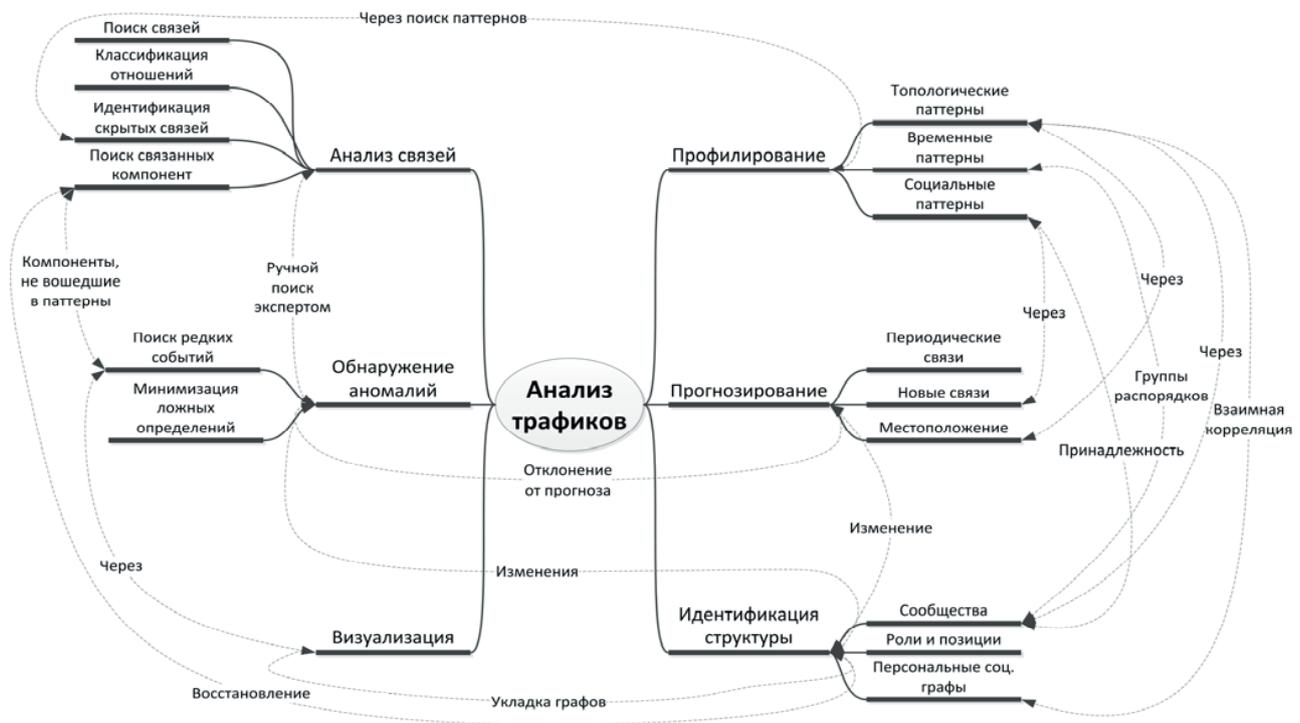


Рис. 2. Проблематика предметной области

вого неизвестного (текущего) маршрута возможно определение максимально соответствующего кластера и соответственно прогнозирование дальнейшего местоположения объекта. Особенностью работы [22] является учет социального графа отдельного пользователя. Показано, что при наличии информации о других пользователях из близкого круга (друзей) для отдельного пользователя возможно значительное увеличение качества прогноза. В работе [28] данные о паттернах физической близости между объектами сопоставляются с опросными данными. Показано, что возможно использовать ТТС для выявления социальной сети конкретного объекта (сети друзей) даже, если не инициируются события телефонии (звонки, сообщения), а по геолокационной составляющей. В работе [27] используется профилирование всех трех типов. Для каждого объекта определяются пространство его действий, и выделяются основные компоненты – *eigenbehaviors* (опорные вектора). По временному критерию эти компоненты отражают типовые паттерны распорядка дня пользователя, по критерию позиционирования – перемещения, по критерию коммуникации – позицию в социальной сети и принадлежность к определенному сообществу.

Проблема *прогнозирования* состоит из задач прогнозирования возникновения связей между объектами [32, 36] и географического положения объектов [22]. Также возможно другое применение – восстановление пропущенных связей [36]. В работе [23] предлагается осуществлять прогнозирование типов отношений через анализ времени и места общего пребывания объектов. В работе [27] показано как использование аппарата *eigenbehaviors* позволяет находить типовые паттерны поведения и производить интерполяцию – как прогнозирование, так и восстановление неизвестных действий. Отмечено, что использование традиционных моделей, таких как, например, марковские модели, не годится для данных, где временные паттерны могут быть по разному шкалированы. В работах [1, 4] описана методика, позволяющая осуществлять прогнозирование переходов и состояния общесистемных объектов по получаемой от них телеметрической информации, что также дает возможность установления нештатных аномальных состояний. В работе [30] рассмотрен фреймворк, позволяющий моделировать поведение объектов в структурах, представимых динамическими графами. В отличие от методов, основанных на майнинге паттернов, предложенный подход основан на представлении графа в многомерном непрерывном пространстве состояний, в котором можно получить траектории вершин как функции времени.

Проблема *идентификации структуры сети* объединяет задачи определения сообществ [39], ролей и позиций пользователей [2, 27]. Данные задачи междисциплинарные, затрагивают как социологию, так и компьютерные науки. В работах [20, 39] дается формальная постановка задачи определения сообществ. Дано определение критерия оценки качества методов определения сообществ как вычисление меры модульности. Выполнен обзор таких методов: спектральная бисекция, алгоритм Кернигана-Лина, иерархическая кластеризация. Отмечено, что эти методы плохо подходят для данных реального происхождения, поэтому

рассмотрены более современные: Newman & Girvan, Girvan & Newman, Fortunato, Radicchi, Newman, Donetti & Munoz, Eckmann & Moses, Zhou & Lipowski, Bagrow & Bollt, Duch & Arenas, Capocci, Wu & Huberman, Palla, Reichardt & Bornholdt, Guimera. Показана линейная зависимость вычислительной сложности от качества работы алгоритмов. Кроме того, наиболее быстрые методы требуют априорной информации о количестве и примерном размере сообществ.

Задача определения сообществ часто решается методами профилирования и кластеризации. В работах [23, 24, 26] производится определение сообществ через обучение гауссовской модели смещения паттернами близости пользователей, а затем производится определение типов отношений между ними. Проверка адекватности моделей производилась опросом пользователей. Показано, что можно отличить круги коллег, знакомых и друзей. В работе [27] производится кластеризация пользователей по их временным паттернам, что позволяет вычислять похожесть между индивидуальными и групповыми паттернами и определять наиболее близкое, представленное кластером, сообщество для пользователя. В работе [17] производится поиск сообществ в очень большой сети с использованием иерархической агломеративной кластеризации, разбивающей граф соцсети на кластеры сообществ по критерию модульности (*modularity*) – плотности ребер в подграфе сообщества. Иногда определение сообществ осуществляется экспертно на основе визуализации сети. Существует ряд алгоритмов укладок графов соцсетей: алгоритм Фрухтермана-Рейнгольда (*force-directed/силовой*), Барнеса-Хата, Ка Пинг Йи (*radial-tree*), Edge Betweenness Clusterer, алгоритм Ньюмана [16].

Сообщества в соответствии с природой социальной сети могут быть динамическими. В работе [33] рассматриваются вопросы определения статических и динамических сообществ. Ключевым фактором в определении отношений принимается длительность характерных контактов, их количество, которые коррелируются со знакомством и регулярностью. Предложены три алгоритма распределенного обнаружения сообществ: SIMPLE, k-CLIQUE, MODULARITY. Особенностью, которых является возможность использования для *scale-free* сетей.

Исследование временной эволюции отношений в социальной сети важно. Отношения могут представлять события и ассоциации, которые важны в определенный момент времени. В работе [28] рассматриваются мгновенные микросоциальные графы объектов некоторым состояниям динамической сети. В работах [3, 9] предложено использовать динамический граф как модель социальной сети, которая позволяет исследовать рост и снижение интенсивностей связей между объектами. Интерпретация графических укладок серии состояний динамического графа позволяет отслеживать динамику ролей объектов, и управляющих иерархий. В работе [11] предлагается фреймворк, использующий динамический граф и его кластеризацию и позволяющий определять произошедшие структурные изменения, идентифицировать события и выявлять командные иерархии. Предлагаемый подход состоит из стадий: дискретизация динамического графа на состояния, кластеризация

каждого состояния в отдельности с формированием пересекающихся нечетких кластеров. Кластеры представляют сообщества.

На следующей стадии кластеры последовательных состояний сравниваются. Низкая степень соответствия свидетельствует о произошедших значительных структурных изменениях в промежутке времени между состояниями, а высокая – о стабильном промежутке времени. Последовательные состояния с монотонными стабильными кластерами образуют периоды устойчивости, а с различными кластерами – периоды высокой активности. Структура сети, представленная статическим графом в период устойчивости, анализируется методом подсчета метрики дельта-эффективности, что позволяет определять роли отдельных узлов (их влияние).

3. Постановка задачи

Обобщая выполненный обзор, можно сделать вывод, что малоизученными остаются аспекты рассмотрения динамики социальных сетей, в частности: анализ и прогнозирование связей, идентификация структуры сети. Рассмотренные проблемы широко взаимосвязаны (рис. 3), так например, анализ и прогнозирование связей требуют определения паттернов динамики объектов и отражают эволюцию структуры социальной сети, а отклонение от прогноза может толковаться как аномальная ситуация.

Поэтому в качестве задачи для дальнейшего исследования выберем задачу прогнозирования связей. Задача прогнозирования связей в зарубежной литературе формулируется как “проблема прогнозирования связей”. Это связано с некорректным переводом

ее первоначальной формулировки “link prediction problem” (LPP).

Для моделирования социальных сетей применяется математический аппарат теории графов. Графы широко используются и для представления ТТС [13, 25]. Классические методы социально- сетевого анализа применялись к конечным статическим графам ввиду того, что информация о сети собиралась социологическими методами [2, 9, 44]. Но возможности графа как модели ограничиваются указанием абонентов (вершины), связей между ними (ребра), направлений коммуникации (дуги), интенсивности коммуникаций (веса). Социальная сеть в свою очередь является динамическим объектом [36], и к сожалению, граф не может отразить динамику событий в социальной сети.

Для решения такой проблемы вводится понятие динамического графа [12]. В таком случае каждое состояние социальной сети представляется графом, а последовательность таких статических состояний составляет динамический граф. Использование динамического графа как модели социальной сети обосновано в работах [3, 9, 11, 12, 14, 25, 30, 35].

В общем виде задача прогнозирования связей ставится как прогнозирование появления новых ребер в динамически изменяющемся графе. Впервые задача сформулирована в работе [36] следующим образом: имеется социальная сеть в виде статического графа $G = \langle V, E \rangle$, где каждое ребро $e = \langle u, v \rangle$ представляет взаимную связь между объектами u и v , которая имела место быть в определенный момент времени $t(e)$. Будем понимать множественные связи между объектами как параллельные ребра, существующие в разные моменты времени. Для двух моментов времени $t < t'$, пусть $G[t, t']$ обозначает подграф графа G , который

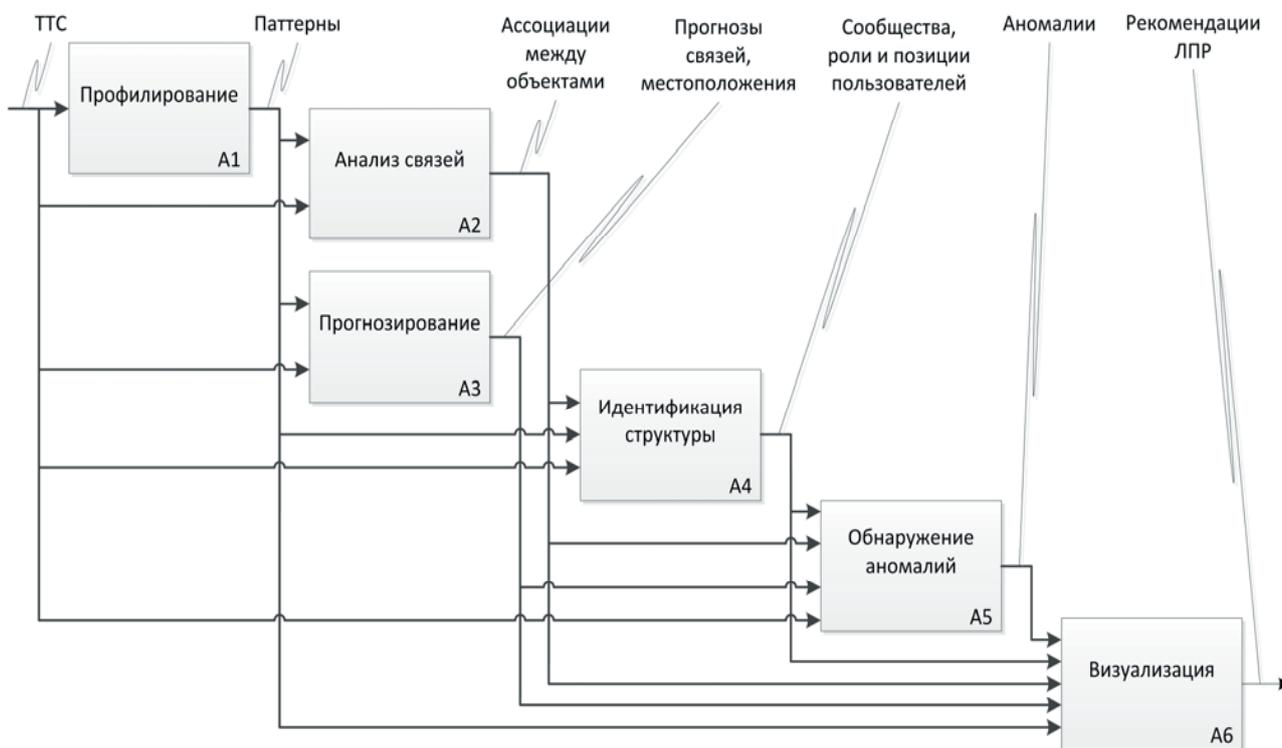


Рис. 3. Декомпозиция процесса анализа ТТС на связанные проблемы

содержит все ребра, существующие между моментами t и t' . Пусть имеется четыре момента времени $t_0 < t'_0 < t_1 < t'_1$ и граф $G[t_0, t'_0]$, необходимо определить список ребер, не принадлежащих $G[t_0, t'_0]$, но принадлежащих $G[t_1, t'_1]$. Интервал $[t_0, t'_0]$ обозначается как обучающий, а $[t_1, t'_1]$ – как тестовый.

Так как решение задачи прогнозирования связей позволяет получить прогноз только для новых связей, но не ответить на вопрос какие связи будут поддерживаться, а какие и вовсе исчезнут, то вводится родственная задача – задача прогнозирования временных связей – time series link prediction problem (TSLPP). Задача заключается в генерации будущего состояния для исходного динамического графа. Рассматривалась в работах [32, 41], более общий случай в [30]. В виду формулировки задачи, социальная сеть представляется не статическим, а динамическим графом.

Динамический граф (ДГ) представим как множество

$$DG = \{G_{t_1}, \dots, G_{t_i}, \dots, G_{t_n}\}, \tag{1}$$

где n – количество разбиений времени существования ДГ на равные промежутки времени длиной $t_i - t_{i-1}$; $t_1, \dots, t_i, \dots, t_n$ – конечные значения промежутков времени $t_i - t_{i-1}$; $G_{t_1}, \dots, G_{t_i}, \dots, G_{t_n}$ – статические графы, представляющие состояние сети в моменты времени $t_1, \dots, t_i, \dots, t_n$.

Каждый статический граф определяется как

$$G_{t_i} = (V, E_{t_i}), \tag{2}$$

где V – постоянное множество вершин v_i , общее для всех G_{t_i} , $V = \{v_1, \dots, v_i, \dots, v_n\}$; E_{t_i} – множество ребер.

Каждое ребро e определяется как

$$e = \{v_i, v_j, w\}, \tag{3}$$

где v_i, v_j – вершины, образующие ребро; w – вес, который определяется по количеству событий коммуникации между v_i, v_j во время $t_i - t_{i-1}$.

Задача прогнозирования временных связей заключается в том, что требуется найти статический граф

$$G_{t_{n+1}} = \{V, E_{t_{n+1}}\} \tag{4}$$

такой, что

$$\forall e \in E_{t_{n+1}}, e \in \bigcup_{i=1}^n E_{t_i}. \tag{5}$$

Т.е. задача TSLPP не решает LPP в частном случае и наоборот. Но объединяя обе задачи можно получить задачу прогнозирования гибридных связей – hybrid link prediction problem (HLPP). Задача заключается в прогнозировании не только будущего состояния для исходного динамического графа, но и новых связей. Впервые была предложена в работе [32].

Ранее исследователями был предложен ряд методов для решения задач прогнозирования связей. В работе [36] авторами сделан обзор методов, использующих меры близости узлов. Каждый метод вычисляет оценку близости для всех возможных пар вершин, в соответствии с которой множество пар вершин упорядочивается. Для вычисления оценки могут использоваться различные метрики: длина кратчайшего пути; анализ соседних вершин – методы Common neighbors, Jaccard's coefficient, Adamic/Adar, Preferential attachment; анализ путей – методы Katz, Hitting time, PageRank, SimRank; другие – методы Low-rank approximation, Unseen bigrams, кластеризация. В работе [32] для TSLPP использовалась модель авторегрессионного интегрированного скользящего среднего (ARIMA), а для LPP применялись методы Common Neighbor, Preferential Attachment, Adamic/Adar, Katz, Generative Model, Spreading Activation. В свою очередь оба полученных прогноза объединялись и усреднялись, что явилось решением GLPP. В работе [27] вводится термин – eigenbehaviors – типовые повторяющиеся структуры в основе поведения – принципиальные компоненты из набора данных о действиях объекта. Использование кластеризации и классификации eigenbehaviors позволяет прогнозировать действие объектов, в т.ч. и установление связей. В работе [30] рассмотрены регрессионный, спектральный подходы, ARMA модель, используется метод мультипространственного шкалирования MDS (Multi Dimensional Scaling), состоящий из четырех стадий: встраивание графа в многомерное пространство, выравнивание вариации траекторий, прогнозирование траекторий, восстановление графа. В работе [41] используется прогнозирование временных рядов для описания индивидуального поведения узлов, вычисления их временных показателей для определения развития поведения и улучшения прогнозирования связей. Вводится ряд метрик для узла: activeness, recensu, понятие локального паттерна. В работе [5] предлагается комплексный метод, использующий в качестве признакового пространства коэффициенты Жаккара, Adamic/Adar, Katz, PageRank и их линейную комбинацию, настроенную с помощью алгоритма LENKOR [18]. Данный подход позволил задействовать два критерия для формирования решения о возникновении нового ребра – наличие общих соседей и профилирование вершин.

4. Способ и методы решения

Для решения задачи прогнозирования временных связей предложим следующий подход (рис. 4). Для исходного множества транзакционных данных ТТС строится динамический граф. Для ТТС характерно точное и полное отражение взаимодействия объектов с детализацией времени и продолжительности. В данном случае ТТС соответствуют “панельным данным” в терминах социологических наук [14], что позволяет интерпретировать ТТС как динамическую социальную сеть, а для каждого абонента в отдельности – динамический социальный граф по критерию использования услуг телефонии. Вопросы представления исходных данных и алгоритм постро-

ения динамического графа рассмотрены в работе [7]. Полученный динамический граф подвергается анализу с целью извлечения знаний о характере поведения анализируемого объекта, в результате чего осуществляется заполнение базы знаний. Предлагается использовать продукционную модель представления знаний о динамике объекта. Далее по базе знаний осуществляется прямой логический вывод, результатом которого должно быть прогнозируемое состояние сети.

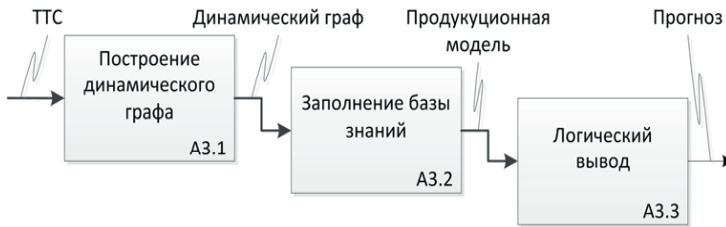


Рис. 4. Предлагаемый подход к решению задачи TSLPP

Наибольший интерес представляет процесс заполнения базы знаний. Вопрос перехода от данных к знаниям рассматривался в работах [1, 4], поэтому предлагаемый метод (рис. 5) адаптирует уже существующую методику в рамках стадий к новой предметной области и задаче.



Рис. 5. Метод извлечения знаний о динамике объекта

Метод состоит из трех стадий: кластеризация состояний динамического графа, построение шаблонов для полученных кластеров и поиск паттернов следования шаблонов, генерация деревьев решений. Методы кластеризации широко использовались в [1, 4] для выделения групп схожих сегментов телеметрических сигналов.

В работе [27] кластеризация похожим образом позволяет снизить размерность пространства состояний объектов. В работах [23, 30] кластеризация используется для формирования профилей пользователей по социальному, временному и поведенческому критериям. В предложенном методе кластеризация также преследует цель снизить пространство состояний анализируемого объекта.

Кластеризация состояний позволит получить множество кластеров

$$K = \{K_1, \dots, K_x, \dots, K_{n^k}\}, \quad (6)$$

каждый из которых в свою очередь содержит множество близких состояний

$$K_x = \{G_1^{K_x}, \dots, G_{n^{K_x}}^{K_x}\}. \quad (7)$$

В качестве критерия близости можно использовать евклидово расстояние на основе весов соответствующих ребер

$$d_E(G_x, G_y) = \sqrt{\sum_i \sum_j (w_{\{v_i, v_j\} \in E_x} - w_{\{v_i, v_j\} \in E_y})^2}. \quad (8)$$

На следующей стадии осуществляется построение усредненных шаблонов для кластеров через определение их центров.

Центр также является графом

$$GC^{K_x} = (V, E^{K_x}), \quad (9)$$

где

$$E^{K_x} = \bigcup_{i=1}^{n^x} E_i^{K_x}, \quad (10)$$

но вес каждого ребра пересчитывается по формуле

$$w_{\{v_i, v_j\} \in E^{K_x}} = \frac{\sum_{k=1}^{n^{K_x}} w_{\{v_i, v_j\} \in E_k^{K_x}}}{\sum_{i=1}^{n^v} \sum_{j=1}^{n^v} \sum_{k=1}^{n^{K_x}} w_{\{v_i, v_j\} \in E_k^{K_x}}}. \quad (11)$$

Изначальные состояния в динамическом графе заменяются соответствующими шаблонами. Модифицированный динамический граф анализируется методами секвенциального анализа.

В работах [1, 4] секвенциальный анализ применяется для построения паттернов состояний объекта.

В работе [22] используется поиск шаблонов в истории объекта для характеристики текущего состояния. Аналогично этим работам результатами секвенциального анализа являются паттерны следования состояний объекта

$$S = \{S_1, S_2, \dots, S_{n^s}\}, \quad (12)$$

$$S_x = \langle GC_1^x, \dots, GC_{n^{S_x}}^x \rangle. \quad (13)$$

На последней стадии по полученным паттернам строятся деревья решений, с помощью которых получают продукционные правила – правила динамики объекта, типа

$$GC^x \rightarrow GC^y. \quad (14)$$

Полученные правила составляют продукционную модель знаний о характере поведения исследуемого объекта и могут быть использованы в процессе логического вывода для построения прогноза или идентификации аномальной ситуации.

5. Выводы

Рассмотрена проблема анализа трафиков телефонных сетей в интеллектуальных системах поддержки принятия решений. Проведен системный анализ процесса анализа трафиков и перспективы его практического применения, выделены проблемы и задачи предметной области, проведен обзор посвященным им работ. В качестве задачи для дальнейшего исследования выбрана задача прогнозирования связей, дана ее формальная постановка, проанализированы ее модификации – задачи прогнозирования временных и гибридных связей. Проанализирова-

ны существующие методы решения задач, отмечена недостаточная разработка методов для задачи прогнозирования временных связей. Предложен подход к ее решению, основанный на моделировании сети абонентов динамическим графом и обнаружении знаний о его динамике, использующий методы кластеризации, секвенциального анализа, деревья решений. В дальнейшем стоит рассмотреть возможность применения аппарата ассоциативных правил для прогнозирования событий в зависимости от географического положения объектов и времени. Остается необходимой экспериментальная проверка предложенного подхода.

Литература

1. Васильев, А. В. Применение алгоритмов кластеризации и классификации в задачах обработки и интерпретации телеметрической информации [Текст] / Васильев В. А., Геппенер В. В, Жукова Н. А., Клионский Д. М., Тристанов А. Б. // Доклады 9-й международной конференции «Цифровая обработка сигналов и ее применение», 28-30 марта 2007 г. – М : ИПУ РАН. – 2007. С. 389-392.
2. Давыденко, В. А. Моделирование социальных сетей [Текст] / В. А. Давыденко, Г. Ф. Ромашкина, С. Н. Чуканов // Вестник Тюменского государственного университета. – 2005. – № 1. – С. 68-79.
3. Долинина, О. Н. Модель графической визуализации динамической социальной сети с локальными ограничениями для образовательного учреждения [Текст] / Долинина О. Н., Тарасова В. В., Печенкин В. В. // Труды вольного экономического общества России. – М : Российский экономический университет им. Г. В. Плеханова. – 2010. – Т. 143. – С. 252-258.
4. Жукова, Н. А. Методы и модели оперативного контроля состояния сложных динамических объектов на основе измерительной информации с использованием алгоритмов интеллектуального анализа данных [Текст] : автореф. дис. на соискание уч. степени канд. техн. наук : спец. 05.13.01 “Системный анализ, управление и обработка информации (технические системы)” / Жукова Наталья Александровна. – СПб., 2008. – 16 с.
5. Кириллов, А. Н. Предсказание связности графа [Текст] / А. Н. Кириллов // Материалы XIX Международной научной конференции студентов, аспирантов и молодых ученых «Ломоносов-2012»: секция «Вычислительная математика и кибернетика», 9-13 апреля 2012 г. – М. : МГУ. – 2012. – С.101-102.
6. Савельев, О. О. О концепции создания информационной системы интеллектуального анализа данных телекоммуникационных компаний в рамках разработки интеллектуальной системы поддержки принятия решений [Текст] / О. О. Савельев // Искусственный интеллект. – 2010. – № 3. – С. 535-539.
7. Савельев, О. О. Построение динамического социального графа по транзакционным данным трафиков телефонных сетей [Текст] / О. О. Савельев // Материалы доповідей VI міжнародної науково-практичної конференції молодих учених, аспірантів, студентів «Сучасна інформаційна Україна: інформатика, економіка, філософія», 26 квітня 2012 р. – Донецьк : Наука і освіта. – 2012. – С. 79-83.
8. Соколова, А. Н. Инструменты расследования. Анализ социальных сетей [Электронный ресурс] / А. Н. Соколова. – 2011. – Режим доступа : <http://www.securityinfowatch.ru/view.php?section=articles&item=3>.
9. Тарасова, В. В. Применение динамической графовой модели для построения и анализа социальной сети образовательной организации [Текст] / В. В. Тарасова // Труды XVIII Всероссийской научно-методической конференции «Телематика'2011», 20-23 июня 2011 г. – СПб : НИУ ИТМО. – 2011. С. 192-194.
10. Bayir, M. A. Discovering Spatiotemporal Mobility Profiles of Cellphone Users [Электронный ресурс] / Murat Ali Bayir, Murat Demirbas, Nathan Eagle // Proceedings of 10th IEEE International Symposium on a “World of Wireless, Mobile and Multimedia Networks”, 15-19 June, 2009 – Kos, Greece. – 2009. – 9 pp. – Режим доступа : <http://reality.media.mit.edu/pdfs/bayir.pdf>.
11. Bourqui, R. Detecting structural changes and command hierarchies in dynamic social networks [Текст] // R. Bourqui, F. Gilbert, P. Simonetto, F. Zaidi, U. Sharan, F. Jourdan // Proceedings of the 2009 International Conference on Advances in Social Network Analysis and Mining ASONAM '09, 20-22 July 2009 – Athens, Greece : IEEE Computer Society. – 2009. – PP. 83-88.
12. Borgwardt, K. M. Pattern Mining in Frequent Dynamic Subgraphs [Текст] / Borgwardt K. M., Kriegel H.-P., Wackersreuther P. // Proceedings of the Sixth International Conference on Data Mining ICDM '06, 18-22 December 2006, Hong Kong, China. – Washington, DC, USA : IEEE Computer Society. – 2006. – PP. 818-822.
13. Berry, M. J. A. Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management [Текст] / Michael J. A. Berry, Gordon S. Linoff. – 2nd ed. – Indianapolis : Wiley Publishing, Inc, 2004. – 643 pp.

14. Clauset, A. Persistence and periodicity in a dynamic proximity network [Электронный ресурс] / Aaron Clauset, Nathan Eagle // DIMACS/DyDAn Workshop on Computational Methods for Dynamic Interaction Networks, September 24 - 25, 2007. – Rutgers, N.J. : DIMACS Center. – 2007. – 5 PP. – Режим доступа : <http://reality.media.mit.edu/pdfs/Clauset.pdf>.
15. Cox, K. C. Visual Data Mining: Recognizing Telephone Calling Fraud [Текст] / Kenneth C. Cox, Stephen G. Eick, Graham J. Wills, Ronald J. Brachman // Data Mining and Knowledge Discovery – 1997. Volume 1, № 2. – PP. 225-231.
16. Catanese, S. A. A Visual Tool for Forensic Analysis of Mobile Phone Traffic [Текст] / Salvatore A. Catanese, Giacomo Fiumara // Proceedings of the 2nd ACM workshop on Multimedia in forensics, security and intelligence, October 25 - 29, 2010, Firenze, Italy. – New York : ACM. – 2010. – PP. 71-76.
17. Clauset, A. Finding community structure in very large networks [Электронный ресурс] / Aaron Clauset, M. E. J. Newman, and Cristopher Moore // Physical Review E. – 2004. Volume 70, № 6. – 6 pp. – Режим доступа : <http://arxiv.org/pdf/cond-mat/0408187v2>.
18. D'yakonov, A. Two Recommendation Algorithms Based on Deformed Linear Combinations [Текст] / A. D'yakonov // Proceedings of the ECML PKDD 2011 Workshop on Discovery Challenge, September 5th, 2011, – Athens, Greece : Rudjer Boskovic Institute – 2011. – PP. 21-27.
19. Danezis, G. Introducing Traffic Analysis [Текст] / George Danezis and Richard Clayton // Digital Privacy: Theory, Technologies, and Practices. [ed. Alessandro Acquisti]. – New York : Auerbach Publications, 2007. – PP. 95-116.
20. Danon, L. Comparing community structure identification [Электронный ресурс] / L. Danon, A. Diaz-Guilera, J. Duch, A. Arenas // Journal of Statistical Mechanics: Theory and Experiment – 2005. – Issue 9. – 10 PP. – Режим доступа : <http://deim.urv.cat/~aarenas/publicacions/pdf/jstat05.pdf>.
21. Diffie, W. Privacy on the Line: The Politics of Wiretapping and Encryption [Текст] / Whitfield Diffie, Susan Landau. – Cambridge : MIT Press, 1998. – 352 pp.
22. De Dominicis, M. Interdependence and Predictability of Human Mobility and Social Interactions [Электронный ресурс] / M. De Dominicis, A. Lima, M. Musolesi // Proceedings of the Nokia Mobile Data Challenge Workshop, June 18th 2012. – Newcastle, UK : Pervasive. – 2012. – 6 PP. – Режим доступа : <http://www.cs.bham.ac.uk/~musolesm/papers/mdc12.pdf>.
23. Eagle, N. Can Serendipity Be Planned? [Текст] / Nathan Eagle // MIT Sloan Management Review – 2004. – Volume 46, № 1. – PP. 10-14.
24. Eagle, N. Using Mobile Phones to Model Complex Social Systems [Электронный ресурс] / Nathan Eagle // O'Reilly Network – 2005. – Режим доступа : <http://www.oreillynet.com/pub/a/network/2005/06/20/MITmedialab.html>.
25. Eberle, W. Analyzing Catalano/Vidro Social Structure Using GBAD [Электронный ресурс] / W. Eberle, L. Holder // Proceedings of IEEE Symposium on Visual Analytics Science and Technology, 19-24 October 2008. – Columbus, Ohio, USA : Eurographics Association. – 2008. – 2 PP. – Режим доступа : <http://users.csc.tntech.edu/~7Eweberle/VAST2008.pdf>.
26. Eagle, N. Social Serendipity: Mobilizing Social Software [Текст] / Nathan Eagle, Alex Pentland // Pervasive Computing – 2005. Volume 4, Issue 2. – PP. 28-34.
27. Eagle, N. Eigenbehaviors: identifying structure in routine [Текст] / Nathan Eagle, Alex Sandy Pentland // Behavioral Ecology and Sociobiology – 2009. Volume 63, Issue 7. – PP. 1057-1066.
28. Eagle, N. Inferring friendship network structure by using mobile phone data [Текст] / N. Eagle, A. S. Pentland, D. Lazer // Proceedings of the National Academy of Sciences (PNAS) – 2009, Volume 106, № 36. – PP. 15274-15278.
29. Farrahi, K. What Did You Do Today? Discovering Daily Routines from Large-Scale Mobile Data [Текст] / K. Farrahi, D. Gatica-Perez // Proceeding of the 16th ACM International Conference on Multimedia, October 26-31, 2008, Vancouver, Canada. – New York, NY, USA : ACM. – 2008. – PP. 849-852.
30. Fang, C. Graph Embedding Framework for Link Prediction and Vertex Behavior Modeling in Temporal Social Networks [Электронный ресурс] / C. Fang, M. Kohram, X. Meng, A. Ralescu // Proceedings of the fifth SNA-KDD Workshop, August 21, 2011, San Diego, CA, USA. – 2011. – 7 pp. – Режим доступа : <http://www.cs.uc.edu/~fangcg/Publications/FrameworkLinkPredKDDSSNA20110525Final.pdf>.
31. Greene, K. Reality Mining [Электронный ресурс] / Kate Greene // Technology Review – 2008. – № 2. – Режим доступа : <http://www.technologyreview.com/article/409598/tr10-reality-mining/>.
32. Huang, Z. The Time Series Link Prediction Problem with Applications in Communication Surveillance [Текст] / Zan Huang, Dennis K.J. Lin // INFORMS Journal on Computing – 2009. – Volume 21, Issue 2. – PP. 286-303.
33. Hui, P. Distributed Community Detection in Delay Tolerant Networks [Электронный ресурс] / P. Hui, E. Yoneki, S.-Y. Chan, J. Crowcroft // Proceedings of 2nd ACM/IEEE international workshop on Mobility in the evolving internet architecture, August 27, 2007, Kyoto, Japan. – New York, NY, USA : ACM – 2007. – 8 pp. – Режим доступа : <http://reality.media.mit.edu/pdfs/Hui.pdf>.
34. Klerks, P. The network paradigm applied to criminal organisations [Текст] / Peter Klerks // Connections – 2001. – № 24(3). – PP. 53-65.
35. Leskovec, J. Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations [Текст] / Jure Leskovec, Jon Kleinberg, Christos Faloutsos // Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery and data mining (KDD), August 21-24, 2005, Chicago, IL, USA. – New York : ACM. – 2005. – PP. 177-187.

36. Liben-Nowell, D. The Link Prediction Problem for Social Networks [Текст] / D. Liben-Nowell, J. Kleinberg // Proceedings of the twelfth international conference on Information and knowledge management, November 03-08, 2003, New Orleans, LA, USA. – New York : ACM. – 2003. – PP. 556-559.
37. Mena, J. Investigative Data Mining for Security and Criminal Detection [Текст] / Jesus Mena. – New York : Butterworth Heinemann, 2003. – 452 pp.
38. Marketos, G. Mobility Data Warehousing and Mining [Электронный ресурс] / Gerasimos Marketos, Yannis Theodoridis // Proceedings of 35th International Conference on Very Large Data Bases PhD Workshop (VLDB'09), 24-28 August 2009, Lyon, France. – Lyon, 2009. – 6 pp. – Режим доступа : <http://infolab.cs.unipi.gr/pubs/confs/VLDB09PhDWorkshop.pdf>.
39. Newman, M. E. J. Detecting community structure in networks [Текст] / M. E. J. Newman // The European Physical Journal B – Condensed Matter and Complex Systems. – 2004. – Volume 38, № 2. – PP. 321-330.
40. Pur, A. The Telephone Traffic Data Analysis [Текст] / Aleksander Pur, Igor Belic // Proceedings of the fifth international criminal justice conference «Policing in Central and Eastern Europe: Dilemmas of Contemporary Criminal Justice», September 23-25 2004, Ljubljana, Slovenia. – Ljubljana, Slovenia : Faculty of Criminal Justice, University of Maribor – 2004. – PP. 779-784.
41. Qiu, B. Evolution of Node Behavior in Link Prediction [Текст] / B. Qiu, Q. He, J. Yen // Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence (AAAI-11), August 7-11, 2011. – San Francisco, CA, USA : AAAI, 2011. – PP. 1810-1811.
42. Thuraisingham, B. Web Data Mining and Applications in Business Intelligence and Counter-Terrorism [Текст] / Bhavani Thuraisingham. – New York : Auerbach Publications 2003. – 516 pp.
43. du Toit, S. F. A Model for the Visual Data Mining of Call Patterns [Электронный ресурс] / Stephanus Francois du Toit , Andre Calitz // Proceedings of South African Telecommunications Networks and Applications Conference, 3-6 September 2006, Cape Town, South Africa. – Stellenbosch. – 2006. – 2 PP. – Режим доступа : <http://coe.nmmu.ac.za/coe/media/Store/documents/Distributed%20Multimedia%20Unit/Publications/SFduToit.pdf>.
44. Wasserman, S. Social Network Analysis: Methods and Applications (Structural Analysis in the Social Science) [Текст] / Stanley Wasserman, Katherine Faust. – Cambridge : Cambridge University Press, 1994. – 825 pp.

Abstract

The article concerns the process of analyzing the traffics of telephone networks in intelligence decision-making support systems (IDMSS). The main objective of the study is to select the new unexplored problem, the solution of which can be expressed as models, methods and algorithms of IDMSS. The review of existing works on the problems of the subject area has made it possible to select the problem of communication prediction. This article provides a formal statement of the problem, where the social network is designed by the dynamic graph, each state of which is a snapshot of the network. The solution of the problem is expressed in the form of the following condition for the initial dynamic graph. There are separated, independent components of the problem – the prediction of temporary and new communications. The analysis of existing methods of problems solution showed the inadequate designing of methods for the temporary communications prediction. We propose a novel approach to its solution, based on the detection of knowledge about the rules of transition and following of states of dynamic graph, using the clustering techniques, sequential analysis, and decision trees. The discovered knowledge will allow prediction as a logical conclusion of future state. The results can be implemented as a separate module of IDMSS, which is intended for use by researchers of social networks, employees of law-enforcement agencies while investigating offenses. There is a necessity of experimental verification of the suggested approach

Keywords: *telephone network traffic, communications prediction, dynamic graph, discovery of knowledge*