

# Интеллектуальный анализ данных при проактивном управлении телекоммуникационными системами Умного города

Д.Ю. Воронин, В.П. Евстигнеев, А.И. Дрожжин, Д.Э. Боровский

*Аннотация* — В современных условиях городское сообщество все больше формируется по принципу гетерархии, включающей в себя множество относительно автономных, но взаимозависимых групп: органы государственного и муниципального управления, бизнес-сообщества и общественные организации, представляющие интересы различных слоев населения. Как результат – между группами формируются гетерогенные информационные и коммуникационные потоки, успешное управление которыми является одним из факторов оптимального развития современного города, формирования комфортной городской среды. Однако в академическом дискурсе не складывается единого подхода к проактивному управлению телекоммуникационными системами, являющимися техническим базисом при реализации концепции «Умный город». В статье представлен подход авторов к проактивному управлению телекоммуникационными системами, базирующийся на использовании технологий имитационного моделирования и методов интеллектуального анализа данных в контексте развития концепции «Умный город» (Smart City) с позиции прикладного системного анализа.

Предлагаемый подход ориентирован на решение следующих задач: преобработка большого объема данных при помощи комплекса методов и алгоритмов Data Science; выявление наиболее содержательных причинно-следственных связей в системе, проведение имитационных экспериментов для формирования эффективной стратегии проактивного управления телекоммуникационными сервисами, консолидация полученных результатов в рамках создания прототипа интеллектуальной интерактивной системы поддержки принятия решений в области проактивного управления телекоммуникационными системами. Представляется, что предлагаемый подход может служить основой для проведения комплексных исследований по реализации эффективных процедур проактивного управления инновационными объектами умной городской среды.

Воронин Дмитрий Юрьевич. Севастопольский государственный университет, зав. базовой кафедры "Программная инженерия интеллектуальных систем", канд. технич. наук (e-mail: voronin@sevsu.ru)

Евстигнеев Владислав Павлович. Севастопольский государственный университет, доцент кафедры Физика, канд. физ.-мат. наук (e-mail: VPEvstigneev@sevsu.ru)

Дрожжин Андрей Игоревич, Университет ИТМО, заместитель директора Института дизайна и урбанистики (e-mail: drozhzhin@itmo.ru)

Боровский Дмитрий Евгеньевич. Севастопольский государственный университет, магистрант, (e-mail: dima\_77@mail.ru)

*Ключевые слова* — Умный город, телекоммуникационная система, интеллектуальный анализ данных, проактивное управление, имитационное моделирование, система поддержки принятия решений, системный анализ, Big Data, AnyLogic, Data Science, R.

## I. ВВЕДЕНИЕ

В настоящее время человечество живет в информационной эре, характеризующейся глобальным сдвигом от концепции сосредоточенного производства – развитого вследствие индустриальной революции – к глобальной транснациональной цифровой индустрии, широко использующей информационные технологии для сбора, накопления и синтеза новых знаний. Одним из основных вызовов для современного общества является, так называемый, информационный бум, проявляющийся в постоянном увеличении объема цифровых данных. Например, за несколько предыдущих лет обществом было создано информации больше, чем за всю предшествующую историю [1].

Концепция «умного» города направлена на поиск новых инфраструктурных решений в градостроительстве и городском управлении на основе интеграции информационных и коммуникационных технологий, внедрения интеллектуальных систем распознавания ситуаций, интегрированных с интернет-технологиями [2, 3]. Таким образом, для противодействия вызовам и рискам информационной эры необходимо развивать технологии Big Data, связанные с критическим анализом и систематизацией данных, а также со способностью добывать необходимые знания путем построения логических зависимостей между цепочками слабоструктурированных данных большого объема [4, 5, 6].

Кроме того, быстрорастущий уровень современной компьютеризации общества приводит к необходимости все более широкого применения телекоммуникационных систем, которые в ряде случаев (например, при реализации проекта «Умный город») могут быть отнесены к объектам критического применения, для которых даже незначительные сбои при их функционировании могут приводить к существенным авариям или даже катастрофам [7, 8]. Негативные последствия таких инцидентов характеризуются существенными материальными и репутационными

издержками, а в некоторых случаях – даже непосредственным образом влияют на безопасность целой страны или региона, например, если говорить о функционировании мобильных ситуационных центров руководителей исполнительной власти и т.п.

На сегодняшний день активно применяются две альтернативные стратегии управления телекоммуникационными системами: реактивная и проактивная [9, 10]. Для реактивного подхода характерно наличие точных оценок последствий возникших инцидентов и основной целью является нейтрализация негативных последствий уже произошедших событий. Проактивный подход ориентирован, в основном, на использование превентивных мер и эффективен как для весьма сложных, так и для относительно динамичных систем.

С другой стороны, управляющие воздействия при реактивном подходе более затратные, чем при проактивном, но и реализуются сравнительно реже. Данный подход использует прогнозные оценки развития рисков и ориентирован на реализацию превентивных мер по обеспечению эффективного, сбалансированного, надежного функционирования телекоммуникационной системы [11, 12]. Примерами управляющих воздействий такого рода могут служить управляющие решения о своевременном запуске стабилизирующих процессов, заключающихся в параметрических или структурных изменениях конфигурации телекоммуникационной системы [13]. Реактивный подход позволяет сэкономить на мониторинге, но при этом вероятность попасть в аварийное состояние становится существенно выше, что недопустимо для телекоммуникационных систем критического применения. Другими словами, при проактивном управлении основной целью принимаемых решений является предупреждение возникновения различных неблагоприятных событий. Они, в свою очередь, могут привести к деградации телекоммуникационной системы, что недопустимо при ее использовании в качестве интеграционного базиса в сложных ИТ-инфраструктурах, например, при реализации проекта «Умный город» Министерства строительства и жилищно-коммунального хозяйства Российской Федерации. Таким образом, эффективное проактивное управление такой сложной гетерогенной системой невозможно без разработки специальных интеллектуальных систем поддержки принятия решений, базирующиеся на использовании технологий аналитического и имитационного моделирования, технологии Big Data и комплекса методов и алгоритмов под обобщенным названием Data Science [14, 15].

## II. ЦЕЛЬ РАБОТЫ

Целью работы является описание авторского подхода, используемого при создании прототипа интеллектуальной интерактивной системы принятия решений в области проактивного управления телекоммуникационными системами, базирующейся на использовании имитационного моделирования и сквозных цифровых технологий обработки слабоструктурированных данных.

## III. АКТУАЛЬНОСТЬ

Анализ литературы позволяет сделать вывод о том, что, эффективность имплементации объектов умной городской среды во многом определяется качеством функционирования телекоммуникационной системы, которое может быть гарантированно обеспечено только при реализации процедур проактивного управления, требующих развитых методов сбора, хранения и аналитики гетерогенных потоков сложно организованных или неструктурированных данных большого объема, обработка которых должна вестись с использованием технологии BigData, а также комплекса методов и алгоритмов под обобщенным названием DataScience [2, 3, 5, 14, 18 – 23].

Особенностью таких систем управления является их региональная специфичность и отсутствие готового универсального технологического решения. Их разработка требует дополнительных усилий, заключающихся в необходимости сочетания традиционных и эвристических методов эмпирического описания данных, многомерного статистического анализа и оценивания случайных процессов с нестационарной структурой ряда, а также применения имитационного моделирования для описания различных сценариев функционирования телекоммуникационных сетей при использовании многоверсионного подхода к синтезу моделей на основе различных методологий: агентной, дискретно-событийной, системно-динамической.

Такой гибридный подход позволяет интегрировать различные решения, описывающие анализируемый процесс с различной степенью детализации, применяя принцип связности и иерархичности для парирования сложности анализируемых процессов. Все это в комплексе и обуславливает научную новизну проекта.

Необходимо отметить, что результаты, полученные при исследовании данных провайдеров телекоммуникационных услуг, планируется использовать при синтезе типового обобщающего решения, формирующего методологический подход по внедрению технологий BigData и DataScience при анализе городских процессов.

В России имеется положительный опыт создания территорий, имеющих признаки «умного города». По данным банка решений умного города (<https://russiasmartcity.ru>) только по направлению «Информационный город и системы» реализуется более 150 проектов, например: «Автоматизированная система обеспечения градостроительной деятельности «Мониторинг», «Информационно-аналитическая система ситуационного центра», базовая платформа «Умный город» Росатома и многие другие. В частности, реализуемый Cisco проект «Умный и безопасный город Казань» предусматривает единую городскую сеть Wi-Fi и видеонаблюдения, контроль городской среды и экологической обстановки, интеллектуальную транспортную систему – комплекс датчиков транспортных потоков, центр обработки и управляемые светофоры. Предполагается, что в итоге реализации проекта затраты

в сфере ЖКХ снизятся на 80%, на уличное освещение – на 40% и на 50 % повысится эффективность пользования общественным транспортом.

Применение смарт-технологий существенным образом позволяет повысить качество жизни граждан [16, 17]. Основными направлениями развития концепции «умный город» традиционно являются: жилищно-коммунальное хозяйство, образование, здравоохранение, техногенная и общественная безопасность. Все эти области являются как источником появления больших данных, так и активным потребителем тех знаний, которые могут быть добыты при их аналитической обработке и визуализации. Однако, в настоящее время не существует четкого перечня функциональных требований к программным продуктам, ориентированным на работу с большими данными. Их обработка и визуализация имеют ряд ограничений как методического, так и технического характера, что существенным образом ограничивает их развитие и повсеместное распространение. В современном цифровом обществе данные генерируются с огромной интенсивностью, причем форматы и форма их представления существенно разнятся. В ряде случаев конвертация может потребовать использования специальных процедур, содержащих сложный процесс прохождения промежуточных стадий или вовлечения сложных экспортирующих и импортирующих процессов перехода от одного формата к другому. Таким образом, задача преобразования форматов в сфере современных компьютерных технологий является нетривиальной, так как для больших данных, в ряде случаев, приходится проводить их агрегацию, кластеризацию и фильтрацию. Процессы обработки больших данных являются достаточно затратными, что требует внедрения адаптивных технологий реконфигурации вычислительных процессов, эффективных методов агрегации и визуализации больших данных.

Анализ литературы [2, 3, 5, 14, 18 – 23] позволяет сделать вывод о том, что, к сожалению, вопросам развития и применения технологий анализа Больших Данных и имитационного моделирования для повышения эффективности функционирования телекоммуникационной системы уделяется недостаточное внимание. А эта система, по сути, является «кровеносной системой» любой умной городской среды. По результатам выполненного обзора литературных источников можно сделать ряд выводов:

1. Задачи по обработке больших данных в литературе рассматриваются, в основном, с точки зрения их практического использования в узкой предметной области.

2. Существующие подходы к управлению телекоммуникационными системами не реализуют в полной мере гибридного подхода, базирующегося на обработке Больших Данных, методике проактивного управления, использовании имитационного моделирования и учете вопросов интеграции управляющих стратегий с проектом «Умный город».

3. Существующие на рынке программные решения по управлению телекоммуникационными системами ориентированы, в основном, на использование

реактивного подхода и не имеют, в подавляющем большинстве, детального описания внутренних алгоритмов их функционирования, что не дает возможности оценить эффективность реализуемых в них подходов.

#### IV. ПРЕДЛАГАЕМЫЙ ПОДХОД

В соответствии с [14, 15], функциональная деятельность современных телекоммуникационных систем описывается одновременно структурированными и неструктурированными данными больших объемов, характеризующимися разнообразием и высокой скоростью роста – что соответствует понятию Big Data. Именно в этой информации скрыты наиболее качественные и содержательные связи и причинно-следственные цепочки (инсайты), которые могут быть использованы компанией для принятия эффективных стратегически важных решений, связанных со структурной и параметрической оптимизацией сетевой инфраструктуры, внедрением новых стратегий для привлечения клиентов, поддержания интереса существующих пользователей и прочее.

На рисунке 1 приведена обобщенная схема описания предлагаемого гибридного подхода к синтезу прототипа интеллектуальной интерактивной системы поддержки принятия решений в области проактивного управления телекоммуникационными системами. Предлагаемый гибридный подход базируется на сочетании процедур проактивного управления (для минимизации вероятностей реализации неблагоприятных событий), технологий обработки Big Data (для парирования априорной неопределенности), интерактивных технологий обеспечения комфорта ЛПР и многоверсионного описания различных сценариев функционирования телекоммуникационных сетей при использовании различных методологий: дискретно-событийной, системно-динамической, агентной.

Таким образом, предлагаемая концепция взаимодействия с Big Data состоит из следующих этапов:

1. Пре-процессинг: сжатие размерностей, фильтрация грубых выбросов, прочие подготовительные вычисления.

2. Работа с инсайтами: выявление, классификация, определение оптимальных способов применения.

3. Пост-процессинг: фильтрация основных «паттернов», визуализация, интерпретация.

4. Агентное моделирование акторных взаимодействий в системе.

Далее описываются инструментальные средства, используемые при реализации предлагаемого подхода.

Для сбора, обработки и анализа данных используется среда статистического программирования R Software [15], которая является лидером среди некоммерческих систем статистического анализа и постепенно становится незаменимой при проведении научно-технических расчетов.

Свободно распространяемая система R является наиболее полной, надежной и динамично развивающейся статистической средой, объединяющей язык программирования высокого уровня и мощные библиотеки программных модулей для вычислительной и графической обработки данных.



Рис. 1. Обобщенная схема описания предлагаемого гибридного подхода

В рамках среды модули SparkR, H2O, Radoop используются для создания инфраструктуры анализируемых данных и выработки типовых решений обращения и манипуляции мультиструктурными данными. Работа с инсайтами, их выявление, классификация и определение наиболее оптимальных способов применения, проводится с использованием междисциплинарного подхода, базирующего на комплексном использовании технологий аналитического и имитационного моделирования и методах DataScience, включающих как элементы многомерного статистического анализа и шкалирования, так и методы машинного обучения.

Одной из важных областей применения Больших Данных является прогнозирование, применяемое в тех случаях, когда проведение экспериментов над реальной системой невозможно или нецелесообразно, например, из-за высокой стоимости или неподходящей длительности проведения эксперимента в реальном масштабе времени. В настоящее время благодаря возрастающей сложности моделируемых объектов с одной стороны и повышению вычислительной способности современных ЭВМ с другой, особую популярность приобретает имитационное моделирование. Имитационная модель представляет собой компьютерную программу, которая описывает структуру и воспроизводит поведение реальной системы во времени. Целью синтеза моделей является проведение имитационных экспериментов, результаты которых могут быть использованы при принятии обоснованных, целесообразных управленческих решений. Современные системы моделирования поддерживают весь арсенал новейших информационных технологий, включая развитые графические оболочки для целей конструирования моделей и интерпретации выходных результатов моделирования, мультимедийные средства, анимацию в реальном масштабе времени, объектно-ориентированное программирование, Internet-решения и др. В имитационном моделировании существует несколько парадигм – постановок проблем и

подходов к их решению, используемых в качестве “каркаса” при построении и анализе моделей.

В современном имитационном моделировании выделяют три достаточно обособленные системы взглядов: системная динамика, дискретно-событийное моделирование, мультиагентные системы. Агентное (мультиагентное) моделирование (“agent-based modeling”) возникло достаточно недавно и имеет свои особенности. Примечательно, что среда имитационного моделирования AnyLogic поддерживает все перечисленные выше парадигмы имитационного моделирования, что дает возможность свободно экспериментировать с уровнем абстракции в очень широких пределах, оставаясь в рамках одного инструмента. AnyLogic используется для разработки имитационных исполняемых моделей и последующего их прогона для анализа. Разработка модели выполняется в графическом редакторе AnyLogic, с использованием многочисленных средств поддержки, упрощающих работу. Построенная модель затем компилируется встроенным компилятором AnyLogic и запускается на выполнение. В процессе выполнения модели пользователь может наблюдать ее поведение, изменять параметры модели, выводить результаты моделирования в различных формах и выполнять разного рода компьютерные эксперименты с моделью. Для реализации специальных вычислений и описания логики поведения объектов AnyLogic позволяет использовать мощный современный язык Java.

Общую совокупность данных провайдера можно разделить на два типа – «статические» данные (обезличенные характеристики пользователей услуг, а также интегральные характеристики использования телекоммуникационных услуг) и «хронологические» данные, отображающие временную динамику нагрузок сети и использования сервисов компании. Применительно к группе «статических» данных обработка проводилась по схеме «Препроцессинг данных» - «Анализ данных и поиск инсайтов» - «Постпроцессинг» - «Аналитика».

Этап «препроцессинга» предполагает выявление главных признаков и факторов, сжатие размерностей

мультиструктуры данных и прочие подготовительные вычисления, включая проверку данных на качество и наличие грубых выбросов. На этом этапе предполагается использование методов главных компонент и многомерного шкалирования.

На этапе анализа данных скрытые инсайты выявляются по взаимной коррелированности (количественных и качественных) признаков телекоммуникационной инфраструктуры и целевых показателей пользователя. Для этого используется метод канонической корреляции. Факторный анализ позволяет определять доминирующие факторы, имеющие основной вклад в дисперсию многомерного массива данных. Важной частью исследования является кластеризация совокупности данных во многомерном пространстве признаков. Для этого использованы неиерархические (k-средних, k-медоид и пр.) и иерархические методы кластерного анализа.

На этапе постпроцессинга предполагается фильтрация основных «паттернов» признаков, визуализация и интерпретация полученных групп или классов признаков, определяющих основные скрытые закономерности в данных (инсайты).

На этапе аналитики предполагается деятельность по работе с инсайтами и определение наиболее оптимальных способов их применения. Целью этого этапа является определение вида закономерностей и причинно-следственных связей между совокупностями признаков телекоммуникационной инфраструктуры и целевых показателей пользователя для создания «моделей» поведения клиента или нагрузок в сети, важных для принятия управленческих решений компании и оптимизации процесса предоставления услуг. Для этого использован метод многомерной линейной регрессии, в том числе, с применением обобщенных линейных моделей. Также на этом этапе планируется выработка дискриминирующих правил для отбора признаков инсайтов того или иного выявленного на предыдущем этапе типа. Также установленные правила представляются в виде дерева решений или регрессионного дерева, широко используемое в машинном обучении и интеллектуальном анализе данных и являющееся средством предсказания целевых показателей пользователя по совокупности признаков телекоммуникационной инфраструктуры.

Применительно к группе «хронологических» данных использованы специальные методы диагноза и прогноза временных рядов. В частности, такой анализ опирается на объективное разделение рядов на трендовые, квазипериодические и стохастические («шумовые») компоненты. Часто детерминированные составляющие ряда маскируются «шумовой» компонентой и проявляют себя через автокоррелированную структуру ряда. Выделение и анализ таких компонент должен проводиться разными методами. В проекте предлагается использовать новый способ эмпирического разложения временных рядов на конечное число ортогональных нерегулярных компонент – метод EMD. Метод EMD не зависит от строго заданного базиса функций разложения (Фурье-анализ) или от необходимости предварительного определения материнского вейвлета (вейвлет-анализ), что, по сути, больше подходит для временного ряда с

регулярной структурой. Метод обладает высокой степенью локализации во временной области разложения и позволяет эффективно выделять трендовую составляющую, квазипериодические и шумовые компоненты из нестационарных нерегулярных временных рядов. Применение метода EMD позволяет построить существенно лучшую модель анализируемых временных рядов. В качестве модели трендовых и квазипериодических компонент могут послужить полиномы разных степеней и периодические функции, модель же стохастической компоненты строится на основе авторегрессии остатков временных рядов после удаления детерминированных компонент. Считается, что наилучшими с точки зрения воспроизведения структуры временного ряда с длинной «памятью» являются дробно-интегрированные модели авторегрессии – скользящего среднего (FARIMA). Применение FARIMA-моделей позволяет адекватно определять доверительные интервалы статистических оценок случайного процесса.

Обобщенная организационно-функциональная структура интеллектуальной системы поддержки принятия решений по проактивному управлению телекоммуникационными системами представлена на рисунке 2. Целевое назначение данной системы – решение задачи синтеза оптимального управления телекоммуникационной системой. Ее формализация описана при помощи модели КР (1). Исходными данными являются параметры, характеризующие процесс функционирования телекоммуникационной системы. Результатом функционирования – найденное оптимальное проактивное управление.

$$\left\{ \begin{array}{l} KP: I \rightarrow u^{opt}; \\ I = (\Omega, \Theta, \Psi, \Phi, \Xi); \\ M_1: Dat \rightarrow KP; \\ M_2: \Omega \rightarrow \Theta; \\ M_3: \Theta \rightarrow \Psi; \\ M_4: \Psi \rightarrow \Phi; \\ M_5: \Phi \rightarrow \Xi; \\ M_6: (\Phi, \Xi) \rightarrow u^{opt}. \end{array} \right. \quad (1)$$

где  $I$  – кортеж, описывающий информационное обеспечение модели КР;

$u^{opt}$  – кортеж, описывающий параметры оптимального проактивного управления коммуникационной системой;

$Dat$  – множество, описывающее данные о предметной области, которые используются для постановки задачи проактивного управления телекоммуникационной системой;

$\Omega$  – множество, описывающее «сырые» данные, используемые для интеллектуального анализа;

$\Theta$  – множество, описывающее, данные, успешно прошедшие предварительную обработку;

$\Psi$  – множество, описывающее найденные инсайты;

$\Phi$  – множество, описывающее результаты интеллектуального анализа данных;

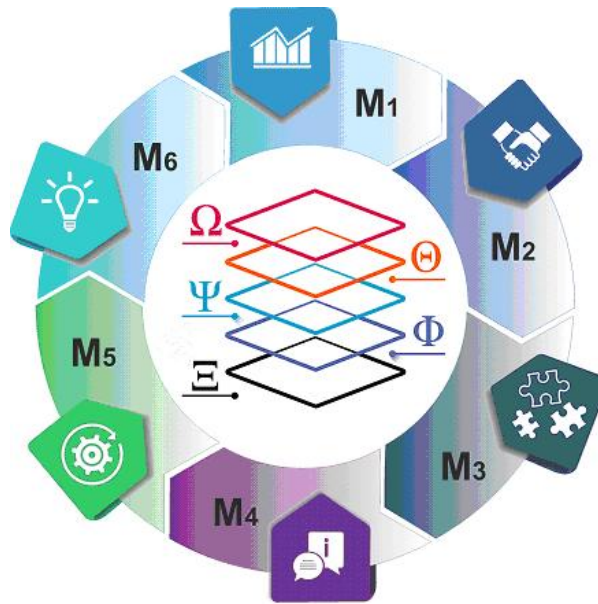


Рис. 2. Обобщенная организационно-функциональная структура интеллектуальной системы поддержки принятия решений по проактивному управлению телекоммуникационными системами

$\Xi$  – множество, описывающее модели «типовых» сценариев поведения клиента телекоммуникационной сети;

где  $I$  – кортеж, описывающий информационное обеспечение модели КР;

$u^{opt}$  – кортеж, описывающий параметры оптимального проактивного управления коммуникационной системой;

$Dat$  – множество, описывающее данные о предметной области, которые используются для постановки задачи проактивного управления телекоммуникационной системой;

$\Omega$  – множество, описывающее «сырые» данные, используемые для интеллектуального анализа;

$\Theta$  – множество, описывающее, данные, успешно прошедшие предварительную обработку;

$\Psi$  – множество, описывающее найденные инсайты;

$\Phi$  – множество, описывающее результаты интеллектуального анализа данных;

$\Xi$  – множество, описывающее модели «типовых» сценариев поведения клиента телекоммуникационной сети;

$M_1$  – функциональный блок, осуществляющий анализ информации для постановки корректной задачи исследования;

$M_2$  – функциональный блок, описывающий пре-процессинг при проведении интеллектуального анализа данных: сжатие размерностей, фильтрация грубых выбросов, прочие подготовительные вычисления;

$M_3$  – функциональный блок, осуществляющий выявление наиболее содержательных причинно-следственных связей в системе (инсайтов);

$M_4$  – функциональный блок, реализующий пост-процессинг результатов интеллектуального анализа данных: фильтрацию основных «паттернов», визуализацию, интерпретацию и т.п.;

$M_5$  – функциональный блок, реализующий построение имитационных моделей для описания «типовых» сценариев поведения клиента телекоммуникационной сети;

$M_6$  – функциональный блок, формирующий оптимальное проактивное управление телекоммуникационной системой.

Предлагаемый подход ориентирован на решение следующих задач:

1. Предобработка большого объема данных (полученных от провайдера телекоммуникационных услуг) при помощи комплекса методов и алгоритмов Data Science.

2. Выявление наиболее содержательных причинно-следственных связей (инсайтов), которые могут быть использованы для принятия эффективных стратегически важных решений по управлению телекоммуникационной системой. Используемые методы: многомерная линейная регрессия, метод канонических корреляций, дерево решений, синтез дискриминирующих правил и др.

3. Построение комплекса имитационных моделей, содержащих алгоритмы систематизации и обобщения выявленных скрытых закономерностей, например, синтез поведенческих моделей клиента телекоммуникационной системы.

4. Проведение имитационных экспериментов для формирования эффективной стратегии проактивного управления сервисами провайдера, анализ гарантоспособности телекоммуникационной инфраструктуры.

5. Консолидация полученных результатов в рамках создания прототипа интеллектуальной интерактивной системы поддержки принятия решений в области проактивного управления телекоммуникационными системами.

## V. ЗАКЛЮЧЕНИЕ И ВЫВОДЫ

Подытоживая можно сформулировать следующие выводы.

1) Предлагаемый подход ориентирован на решение задач, связанных с предобработкой большого объема данных при помощи комплекса методов и алгоритмов Data Science, а также на выявление наиболее содержательных причинно-следственных связей в системе, для последующего использования этой информации при проведении имитационных экспериментов по формированию эффективной стратегии проактивного управления телекоммуникационными сервисами. В частности, он может быть использован при разработке «типовых» сценариев поведения клиента телекоммуникационной сети с целью синтеза для каждого из сценариев граничных условий при оптимизации процесса предоставления услуг.

2) Предлагаемые технологии консолидации гибридного имитационного моделирования и методики проактивного управления телекоммуникационными системами являются базисом для формирования масштабируемых эффективных решений при решении ряда важных вопросов интеграции существующей ИТ-инфраструктуры с инновационными объектами урбанизированной среды, внедряемыми в рамках проекта «Умный город».

3) Планируется развить данный подход при решении задачи синтеза интеллектуальной интерактивной системы поддержки принятия решений в области проактивного управления телекоммуникационными системами.

Работа выполнена при поддержке Севастопольского государственного университета в рамках внутреннего гранта №513/06-31 по проекту «Технологии Больших Данных при проактивном управлении телекоммуникационными системами».

## БИБЛИОГРАФИЯ

- [1] Graham M., Dutton W. Society and the internet: How networks of information and communication are changing our lives. – Oxford University Press, 2019. DOI: 10.1093/acprof:oso/9780199661992.001.0001
- [2] Albino V., Berardi U., Dangelico R.M. Smart cities: Definitions, dimensions, performance, and initiatives // *Journal of Urban Technology*. 2015. Vol. 22 (1). P. 3-21. DOI: 10.1080/10630732.2014.942092
- [3] Chourabi H., Nam T., Walker S., Gil-Garcia J., Mellouli S., Nahon K., Pardo T., Scholl H. Understanding smart cities: An integrative framework // *System Science (HICSS): 45th Hawaii International Conference on System Sciences*. 2012. P. 2289-2297. DOI: 10.1109/HICSS.2012.615
- [4] Bhagya Nathali Silva, Murad Khan, and Kijun Han. “Big Data Analytics Embedded Smart City Architecture for Performance Enhancement through Real-Time Data Processing and Decision-Making.” *Wireless Communications and Mobile Computing*, vol. 2017, Article ID 9429676, 12 pages, 2017. DOI: 10.1155/2017/9429676.
- [5] Abella A., Ortiz-de-Urbina-Criado M., De-Pablos-Heredero C. A model for the analysis of data-driven innovation and value generation in smart cities’ ecosystems // *Cities*. 2017. Vol. 64. P. 47–53. DOI: 10.1016/j.cities.2017.01.011
- [6] Torrecilla J.L., Romo J. Data learning from big data (2018) *Statistics and Probability Letters*, 136, P. 15-19. DOI: 10.1016/j.spl.2018.02.038
- [7] Воронин Д.Ю. Системное моделирование акторных взаимодействий для облачных сервисов (монография) / А.В. Скатков, В.И. Шевченко, А.А. Брюховецкий, Е.Н. Машенко, О.В. Ченгарь. — Симферополь: Изд-во ИТ «АРИАЛ», 2018. — 416 с. ISBN 978 – 5 – 907032 –64 – 4.
- [8] Воронин Д.Ю., Скатков А.В., Машенко Е.Н. Информационные технологии для критических инфраструктур. — Севастополь: Изд-во «СевНТУ», 2012. — 306 с.
- [9] Охтилев М. Ю., Мустафин Н. Г., Миллер В. Е., Соколов Б. В. Концепция проактивного управления сложными объектами: теоретические и технологические основы // *Известия вузов. Приборостроение*. 2014. Том 57. №11. С. 7–14. URL: <https://pribor.ifmo.ru/file/journal/518.pdf>
- [10] Воронин Д.Ю., Скатков А.В., Шевченко В.И., Ключарев А. А. Проактивный и реактивный риск-менеджмент IT-сервисов облачных сред // *Информационно-управляющие системы — Санкт-Петербург: Изд-во РИЦ "ГУАП"*, 2017. – №3 (88). — С. 25 – 33. URL: <https://rucont.ru/efd/633482>
- [11] Воронин Д.Ю., Скатков А.В., Сосновский Ю.В., Зганяйко Д.О. Методология и выбор перспективных направлений реализации системы сбора характеристических данных сетевого трафика / *Вестник СевНТУ. Сер. Автоматизация процессов и управление: сб. науч. тр. – Севастополь: изд-во СевНТУ*, 2014. – Вып. 154. – С. 117 – 120.
- [12] Voronin D., Skatkov A., Moiseev D., Shevchenko V. Modeling of monitoring processes of structurally heterogeneous technological objects // *MATEC Web of Conferences – International Conference on Modern Trends in Manufacturing Technologies and Equipment (ICMTMTE 2017)*. – № 129, 03022 (2017). – pp. 1 – 6. DOI: 10.1051/mateconf/201712903022
- [13] Voronin D., Skatkov A., Bryukhovetskiy A., Shevchenko V. Monitoring of Qualitative Changes of Network Traffic States Based on the Heteroscedasticity Effect. // *IEEE 10th International Conference on Application of Information and Communication Technologies (AICT)*. 2016. pp.562-565. DOI: 10.1109/ICAICT.2016.7991765
- [14] Zaheer Khan, Ashiq Anjum, Kamran Soomro, Muhammad Atif Tahir. Towards cloud based big data analytics for smart future cities. *Journal of Cloud Computing Advances, Systems and Applications*. 2015. DOI: 10.1186/s13677-015-0026-8
- [15] Rajeswari C., Basu D., Maurya N. Comparative Study of Big data Analytics Tools: R and Tableau (2017) *IOP Conference Series: Materials Science and Engineering*, 263 (4), статья № 042052, DOI: 10.1088/1757-899X/263/4/042052
- [16] Кабанов Ю.А., Чугунов А.В. Концептуализация понятий, используемых в исследованиях «умного города» и «электронного управления»: опыт наукометрического анализа // *International Journal of Open Information Technologies*. 2018. Т. 6, № 11. С. 54-58. URL: <http://injoit.org/index.php/j1/article/view/653>
- [17] Ghosal A., Halder S. Building Intelligent Systems for Smart Cities: Issues, Challenges and Approaches. 2018. DOI: 10.1007/978-3-319-76669-0\_5.
- [18] Chui K. T., Vasant P., Liu R. W. Smart city is a safe city: information and communication technology-enhanced urban space monitoring and surveillance systems: the promise and limitations // *Smart Cities: Issues and Challenges*. – Elsevier, 2019. – С. 111-124. DOI: 10.1016/B978-0-12-816639-0.00007-7
- [19] Idwan S. et al. Optimal Management of Solid Waste in Smart Cities using Internet of Things // *Wireless Personal Communications*. – 2019. – С. 1-17. DOI: 10.1007/s11277-019-06738-8
- [20] Savaglio C. et al. Agent-based Internet of Things: State-of-the-art and research challenges // *Future Generation Computer Systems*. – 2020. – Т. 102. – С. 1038-1053. DOI: 10.1016/j.future.2019.09.016

- [21] Chamoso P. et al. Tendencies of technologies and platforms in smart cities: a state-of-the-art review //Wireless Communications and Mobile Computing. – 2018. – T. 2018. DOI: 10.1155/2018/3086854
- [22] Al Nuaimi E. et al. Applications of big data to smart cities //Journal of Internet Services and Applications. – 2015. – T. 6. – №. 1. – C. 25. DOI: 10.1186/s13174-015-0041-5
- [23] Elhoseny H. et al. A framework for big data analysis in smart cities //International Conference on Advanced Machine Learning Technologies and Applications. – Springer, Cham, 2018. – C.405–414. DOI: 10.1007/978-3-319-74690-6\_40



# Data Mining for Proactive Management of Telecommunication Systems in Smart City

Dmitri Yu. Voronin, Vladislav P. Evstigneev, Andrei I. Drozhzhin, Dmitri E. Borovsky

**Abstract** — In modern conditions, the urban community is increasingly being formed on the principle of hierarchy, which includes many relatively autonomous, but interdependent groups: state and municipal governments, business communities and public organizations representing the population various segments interests. As a result, heterogeneous information and communication flows are formed between groups. Successful management of these flows is one of the optimal modern city's development factors – the comfortable urban environment formation. However, the academic discourse has no approach to the proactive telecommunication systems management, suitable for technical basis for the Smart City concept implementation. The article presents the authors' approach to the telecommunication systems proactive management, based on the simulation technology and data mining use in the Smart City concept's development context using the applied system analysis perspective. The proposed approach is aimed at solving the following problems: pre-processing a large amount of data using a complex of methods and algorithms Data Science; identifying the most meaningful causal relationships in the system, conducting simulation experiments to formulate an effective strategy for telecommunication services proactive management, consolidating the results as part of the prototype creation of an intelligent interactive decision support system for telecommunication systems proactive management. It seems that the proposed approach can be successfully used as the basis for conducting comprehensive research on the effective proactive procedures implementation for the innovative management of urban smart objects.

**Keywords** — Smart city, telecommunication system, Data Mining, proactive management, simulation, decision support system, systems analysis, Big Data, AnyLogic, Data Science, R.

## REFERENCES

- [1] Graham M., Dutton W. Society and the internet: How networks of information and communication are changing our lives. – Oxford University Press, 2019. DOI: 10.1093/acprof:oso/9780199661992.001.0001
- [2] Albino V., Berardi U., Dangelico R.M. Smart cities: Definitions, dimensions, performance, and initiatives // Journal of Urban Technology. 2015. Vol. 22 (1). P. 3-21. DOI: 10.1080/10630732.2014.942092
- [3] Chourabi H., Nam T., Walker S., Gil-Garcia J., Mellouli S., Nahon, K., Pardo T., Scholl H. Understanding smart cities: An integrative framework // System Science (HICSS): 45th Hawaii International Conference on System Sciences. 2012. P. 2289-2297. DOI 10.1109/HICSS.2012.615
- [4] Bhagya Nathali Silva, Murad Khan, and Kijun Han. "Big Data Analytics Embedded Smart City Architecture for Performance Enhancement through Real-Time Data Processing and Decision-Making," Wireless Communications and Mobile Computing, vol. 2017, Article ID 9429676, 12 pages, 2017. DOI: 10.1155/2017/9429676.
- [5] Abella A., Ortiz-de-Urbina-Criado M., De-Pablos-Heredero C. A model for the analysis of data-driven innovation and value generation in smart cities' ecosystems // Cities. 2017. Vol. 64. P. 47–53. DOI 10.1016/j.cities.2017.01.011
- [6] Torrecilla J.L., Romo J. Data learning from big data (2018) Statistics and Probability Letters, 136, pp. 15-19. DOI: 10.1016/j.spl.2018.02.038
- [7] Voronin D. System modeling of actor interactions for cloud services / Skatkov A., Shevchenko V., Shevchenko V., Mashchenko E., Chengar O. Simferopol: IT «ARIAL», 2018. 416 p. ISBN 978 – 5 – 907032 –64 – 4. (In Russian)
- [8] Voronin D., Skatkov A. Mashchenko E. Information Technology for Critical Infrastructures. Sevastopol: «SevNTU», 2012. 306 p. (In Russian)
- [9] Ohtilev M., Mustafin N. Miller V. Sokolov B. The concept of proactive management of complex objects: theoretical and technological foundations // Izvestija vuzov. Priborostroenie. 2014. Vol. 57. №11. pp. 7–14. URL: https://pribor.ifmo.ru/file/journal/518.pdf (In Russian)
- [10] Voronin D., Skatkov A., Shevchenko V., Kljucharev A. Proactive and reactive risk management of cloud IT services // Informacionno-upravljajushhie sistemy — St. Petersburg: RIC "GUAP", 2017. – №3 (88). — pp. 25 – 33. URL: https://rucont.ru/efd/633482 (In Russian)
- [11] Voronin D., Skatkov A., Sosnovskij Ju., Zganajko D. Methodology and selection of promising areas for implementing a system for collecting characteristic data of network traffic // Vestnik SevNTU. Ser. Avtomatizacija processov i upravlenie: sb. nauch. tr. –Sevastopol: «SevNTU», 2014. – № 154. – pp. 117 – 120. (In Russian)
- [12] Voronin D., Skatkov A., Moiseev D., Shevchenko V. Modeling of monitoring processes of structurally heterogeneous technological objects // MATEC Web of Conferences – International Conference on Modern Trends in Manufacturing Technologies and Equipment (ICMTMTE 2017). – № 129, 03022 (2017). – pp. 1 – 6. DOI: 10.1051/mateconf/201712903022
- [13] Voronin D., Skatkov A., Bryukhovetskiy A., Shevchenko V. Monitoring of Qualitative Changes of Network Traffic States Based on the Heteroscedasticity Effect. //IEEE 10th International Conference on Application of Information and Communication Technologies (AICT). 2016. pp.562-565. DOI: 10.1109/ICAICT.2016.7991765
- [14] Zaheer Khan, Ashiq Anjum, Kamran Soomro, Muhammad Atif Tahir. Towards cloud based big data analytics for smart future cities. Journal of Cloud Computing Advances, Systems and Applications. 2015. DOI: 10.1186/s13677-015-0026-8
- [15] Rajeswari C., Basu D., Maurya N. Comparative Study of Big data Analytics Tools: R and Tableau (2017) IOP Conference Series: Materials Science and Engineering, 263 (4), статья № 042052, DOI: 10.1088/1757-899X/263/4/042052
- [16] Kabanov Yu., Chugunov A. Conceptualization of the concepts used in the studies of the "smart city" and "electronic control": the experience of scientometric analysis // International Journal of Open Information Technologies. 2018. C. 6, № 11. pp. 54-58. URL: http://injoit.org/index.php/j1/article/view/653 (In Russian)
- [17] Ghosal A., Halder S. Building Intelligent Systems for Smart Cities: Issues, Challenges and Approaches. 2018.DOI: 10.1007/978-3-319-76669-0\_5.
- [18] Chui K. T., Vasant P., Liu R. W. Smart city is a safe city: information and communication technology–enhanced urban space monitoring and surveillance systems: the promise and limitations //Smart Cities: Issues and Challenges. – Elsevier, 2019. – pp. 111-124. DOI: 10.1016/B978-0-12-816639-0.00007-7

- [19] Idwan S. et al. Optimal Management of Solid Waste in Smart Cities using Internet of Things //Wireless Personal Communications. – 2019. – pp. 1-17. DOI: 10.1007/s11277-019-06738-8
- [20] Savaglio C. et al. Agent-based Internet of Things: State-of-the-art and research challenges //Future Generation Computer Systems. – 2020. – T. 102. – C. 1038-1053. DOI: 10.1016/j.future.2019.09.016
- [21] Chamoso P. et al. Tendencies of technologies and platforms in smart cities: a state-of-the-art review //Wireless Communications and Mobile Computing. – 2018. – T. 2018. DOI: 10.1155/2018/3086854
- [22] Al Nuaimi E. et al. Applications of big data to smart cities //Journal of Internet Services and Applications. – 2015. – V. 6. – №. 1. – p. 25. DOI: 10.1186/s13174-015-0041-5
- [23] Elhoseny H. et al. A framework for big data analysis in smart cities //International Conference on Advanced Machine Learning Technologies and Applications. – Springer, Cham, 2018.C.405–414. DOI: 10.1007/978-3-319-74690-6\_40