

КОНЦЕПЦИЯ ПОСТРОЕНИЯ ЦИФРОВОГО ДВОЙНИКА ГОРОДА

© 2020 С.А. Иванов, К.Ю. Никольская, Г.И. Радченко,

Л.Б. Соколинский, М.Л. Цымблер

Южно-Уральский государственный университет

(454080 Челябинск, пр. им. В.И. Ленина, д. 76)

E-mail: saivanov@susu.ru, nikolskaiaki@susu.ru, gleb.radchenko@susu.ru,

leonid.sokolinsky@susu.ru, mzym@susu.ru

Поступила в редакцию: 15.09.2020

В данной статье представлена концепция построения цифрового двойника города. Городское хозяйство представляет собой сложную многовекторную систему, создание единого цифрового двойника которой в настоящее время является трудно решаемой задачей. Авторами предлагается эволюционный подход к решению этой задачи, в соответствии с которым на единой программно-аппаратной платформе последовательно строятся цифровые двойники отдельных элементов городской среды. Эти цифровые двойники связываются в единую кооперативную систему, позволяющую одним цифровым двойникам использовать данные, производимые другими цифровыми двойниками. В статье дается определение и архитектура такой системы. Описываются классы моделей, которые могут использоваться для создания цифровых двойников. Особое внимание уделяется нейросетевым моделям и моделям для анализа данных. Рассматривается информационная инфраструктура цифрового двойника города, включающая в себя сенсорные сети, механизмы очистки данных и туманные вычисления.

Ключевые слова: цифровой двойник, умный город, городское управление, нейронные сети, интеллектуальный анализ данных, сенсоры, туманные вычисления.

ОБРАЗЕЦ ЦИТИРОВАНИЯ

Иванов С.А., Никольская К.Ю., Радченко Г.И., Соколинский Л.Б., Цымблер М.Л. Концепция построения цифрового двойника города // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2020. Т. 9, № 4. С. 5–23. DOI: 10.14529/cmse200401.

Введение

Сегодня главной движущей силой социально-экономического развития становятся информационные технологии. Инновационные способы сбора и анализа данных постепенно занимают место устоявшихся механизмов управления городом. В отличие от статистических выборок, которые успевают устареть к моменту их анализа, «большие данные» могут обрабатываться в режиме реального времени, что повышает качество и скорость принятия решений. «Большие данные» в области городского управления дополняют традиционные типы информации о городе и расширяют сферу их применения. Так, благодаря «большим данным» стал возможен мониторинг поведенческих моделей и анализ городского образа жизни на пересечении таких привычных категорий как население, экономическое развитие, застройка и инфраструктура и др. Цифровая революция и повсеместное распространение интернета породили новый феномен — Data Driven City (город, управляемый данными) [3] и близкий ему термин Smart City (Умный город). Умный город можно определить как стратегический подход к интеграции данных и цифровых технологий для обеспечения устойчивости, благосостояния граждан и экономического развития городской среды [40]. Концепция Умного города определяет концепцию пространства, в котором ключевые компоненты городской инфраструктуры — окружающая среда, борьба

с чрезвычайными ситуациями, управление дорожным движением и энергетика — интегрированы таким образом, что их функции и возможности могут быть легко объединены как друг с другом, так и с новыми системами [33].

В Министерстве строительства России совместно с МГУ им. М.В. Ломоносова разработан индекс цифровизации городского хозяйства «IQ городов» в рамках проекта «Умный город», который реализуется в рамках двух национальных проектов — «Жилье и городская среда» и «Цифровая экономика». «IQ городов» рассчитывается по десяти направлениям (городское управление, умное ЖКХ, инновации для городской среды, умный городской транспорт, интеллектуальные системы общественной и экологической безопасности, туризм и сервис, интеллектуальные системы социальных услуг, экономическое состояние и инвестиционный климат, инфраструктура сетей связи) и содержит 47 показателей [2]. Одним из важных индикаторов «IQ городов» является наличие цифрового двойника города.

Концепция «Цифрового двойника» (ЦД) обеспечивает создание и поддержку виртуальных моделей объектов и процессов реального мира. Подход цифровых двойников ориентируется на возможность получения и эффективной обработки потоков данных, собираемых автоматически посредством распределенных сенсорных систем «Интернета вещей» (Internet of Things, IoT). «Цифровой двойник» города постепенно наполняется данными реального города, собираемыми в реальном времени из развернутой IoT-инфраструктуры и городских информационных систем. «Цифровой двойник», анализируя информацию о состоянии городской инфраструктуры, динамике движения людей и транспорта, их взаимозависимость, а также их колебания во времени и пространстве, способен прогнозировать изменения состояния систем городского хозяйства и предлагать оптимальные решения. Кроме того, независимо от текущего состояния, цифровой двойник позволит отвечать на вопросы вида «что, если», помогая аналитикам понять, как города, оснащенные интеллектуальными технологиями, будут функционировать в тех или иных экономических, экологических и социальных условиях, и определить факторы, способствующие возможным сбоям [32]. В статье [4] рассматриваются результаты анализа, проведенного Градостроительным институтом пространственного моделирования и развития «Гипрогор Проект», в котором были рассмотрены 10 наиболее комплексных решений по созданию цифровых двойников городов. В список вошли Сингапур, Амаравати, Бостон, Ньюкасл, Джайпур, Хельсинки, Роттердам, Стокгольм, Ренн и Антверпен. Как отмечается в работе, Россию также ждет неотвратимый процесс цифровизации городов, так как это ведет к существенному росту качества жизни граждан. Среди российских решений, приближающихся к видению «Умного города» как цифрового сервиса, обеспечивающего качественно новый уровень сервиса для жителей города, можно отметить проект «Умные города Росатома» [5]. В рамках этого проекта, на базе единой ГИС-платформы сформированы решения по предоставлению жителям сервисов по поиску актуальной информации, отслеживанию работы городских служб, сбора обратной связи о текущих городских проблемах и др.

Целью данной статьи является исследование существующих технологий построения цифровых двойников городов. Для этого нам необходимо определить понятие цифрового двойника города, проанализировать методы моделирования, которые могут быть применены для построения цифровых двойников, а также систематизировать информационную инфраструктуру, которая должна обеспечивать функционирование цифрового двойника города. Статья имеет следующую структуру. В разделе 1 дается определение цифрового двойника города, описывается его архитектура и функциональные требования. В раз-

деле 2 рассматриваются классы моделей, применяемые при создании цифровых двойников. В разделе 3 дано описание информационной инфраструктуры цифрового двойника города, включающее в себя сенсорные сети, очистку данных и туманные вычисления. В заключении суммируются полученные результаты.

1. Цифровой двойник города: определение, структура, функционал

Цифровой двойник (ЦД) — это интегрированная мульти-физическая, мульти-масштабная вероятностная симуляция сложного объекта, которая использует физические, математические, имитационные и другие модели, для того чтобы получить как можно более достоверное представление о соответствующем реальном объекте на основе анализа данных, поступающих из сенсорных сетей и других источников [21].

Цифровой двойник города — это комплекс взаимосвязанных цифровых двойников, представляющих определенные аспекты функционирования и развития городской среды. Эти цифровые двойники поддерживают возможности тонкой подстройки и синхронизации с реальным состоянием городской инфраструктуры посредством данных, поступающих из различных источников в режиме реального времени [32, 38]. Основой для эффективного функционирования цифрового двойника города является непрерывный поток данных, генерируемых различными источниками в цифровой инфраструктуре умного города. Как пример таких источников можно выделить следующие (см. рис. 1).



Рис. 1. Архитектура цифрового двойника города

- *Данные о передвижении жителей города*, включая информацию о перемещении частного, коммерческого и общественного транспорта, а также загруженности улиц, собираемые посредством различных механизмов (данные о транзакциях «единого проездного билета», результаты мониторинга дорожного движения и др.).
- *Информация о физических параметрах городской среды*, получаемых в реальном времени с массивов интеллектуальных датчиков (как частных, так и общественных), позволяет осуществлять мониторинг и анализ таких параметров как температура и влажность воздуха, количество взвешенных частиц и химический состав воздуха, шумовая загрязненность, радиационный фон, химический состав воды и др. в привязке к географической позиции.

- *Данные с камер наружного наблюдения* позволяют проводить интеллектуальный анализ таких характеристик городской среды, которые невозможно или затруднительно собрать другими средствами (загруженность автомобильных и пешеходных дорог, загрязненность и качество дорожной сети, идентификация отдельных объектов и событий в общем потоке).
- *Данные из открытых источников* (такие как открытые государственные порталы и службы, данные о метеорологических условиях, открытая отчетная информация субъектов бизнеса и др.) позволяют обогатить модели интеллектуального анализа данных.

Собранные данные позволяют обеспечить функционирование комплекса цифровых двойников города. Примерами таких цифровых двойников могут служить [13]:

- *цифровой двойник городской инфраструктуры*, представляющий собой интерактивную 3D-модель зданий, сооружений, инженерных коммуникаций и другой городской инфраструктуры;
- *цифровой двойник транспортной сети*, обеспечивающий мониторинг и прогнозирование развития ситуации транспортной доступности, эффективности работы общественного транспорта и др.;
- *цифровой двойник городской экологии*, предоставляющий механизмы мониторинга и прогнозирования экологического состояния городской среды, включая качество почвы, воды, воздуха и др.;
- *цифровой двойник энергетики* и др.

Цифровой двойник города обеспечивает следующие возможности:

- мониторинг текущего состояния городской среды;
- оперативное реагирование на возникновение чрезвычайных ситуаций;
- оценку эффективности проектных решений;
- выявление источников потенциальных рисков;
- прогнозирование развития ситуации с учетом исторических данных.

Цифровой двойник города — это чрезвычайно сложное, комплексное решение, формирование которого может и должно реализовываться поэтапно, путем эволюционного развития и интеграции частных решений, направленных на решение наиболее острых («точечных») проблем. Анализ показывает, что такого рода решения сегодня начинают применяться повсеместно во всех странах мира, не только в передовых мегаполисах, но и в городах большого и среднего размера. Использование таких систем совместно с технологиями «интернета вещей» позволяет решать целый ряд проблем на качественно ином уровне [1]:

- контроль над уровнем загрязнения и анализ регулирующего воздействия путем обследования состояния окружающей среды;
- микроклиматические прогнозы погоды с опорой на городскую сеть датчиков;
- повышение эффективности и снижение расходов за счет вывоза и переработки мусора по необходимости, а не по графику;
- улучшение ситуации на дорогах и экономия топлива за счет умных светофоров и разметки
- рациональное потребление электроэнергии благодаря городскому освещению по необходимости;

- оптимизация снегоуборочных работ благодаря поступающим в реальном времени данным о ситуации на дорогах, погодных условиях и ближайших снегоуборочных машинах;
- умная система полива в парках и общественных местах, учитывающая погодные условия и текущее состояние;
- умные камеры наблюдения для отслеживания преступных деяний и автоматизированная система оповещений AMBER Alert в реальном времени;
- умные парковки, помогающие автоматически подобрать лучшее парковочное место;
- мониторинг износа и состояния мостов, улиц и городской инфраструктуры, направленный на своевременное обслуживание и продление срока службы.

Одним из примеров формирования цифрового двойника может служить пример города Такамацу (Япония, население 420 тыс. человек). В качестве пилотного проекта, власти города внедрили 2 цифровых двойника: цифровой двойник мониторинга и предотвращения чрезвычайных ситуаций (в частности, наводнений) и цифровой двойник туристической привлекательности города [26].

Цифровой двойник для мониторинга и предотвращения чрезвычайных ситуаций основывается на сборе и анализе данных с датчиков уровня воды, расположенных по всей площади города, и обеспечивает мониторинг риска наводнения каждого из районов города в реальном времени. Система также обеспечивает мониторинг состояния убежищ для жителей города путем сбора информации с датчиков влажности и потребления электроэнергии в каждом из убежищ. Мобильное приложение, доступное каждому жителю города, позволяет не только заблаговременно оповестить жителей района, находящихся в зоне риска, но и предоставить всю необходимую информацию на случай экстренной ситуации, включая путь до ближайшего функционирующего убежища.

В плане цифрового двойника туристической привлекательности города был внедрен мониторинг перемещения арендованных велосипедов (как наиболее привлекательного метода перемещения по городу для туристов). Сопоставление гео-трека перемещения арендованных велосипедов с данными о туристе (возраст, пол, гражданство и др.) позволяют сформировать карту наиболее важных туристических локаций города, спланировать и оценить эффективность маркетинговых кампаний по продвижению туристических достопримечательностей (включая поддержку мульти-язычной среды и др.).

2. Классы моделей

ЦД используют различные подходы для моделирования реальных объектов и технологических процессов, включая методы статистического и интеллектуального анализа данных, методы вычислительного моделирования, такие как метод конечных элементов и др. [29]. Каждый из этих методов предъявляет особые требования к необходимым вычислительным ресурсам. Например, методы интеллектуального анализа данных требуют хранилищ большого объема с высокой пропускной способностью для сбора и получения доступа к аналитическим данным, а также высокой масштабируемости вычислительной системы для их обработки; для применения методов машинного обучения требуются узлы с установленными графическими ускорителями, а для моделей, использующей метод конечных элементов, требуются центральные процессоры с высокой производительностью и большие объемы оперативной памяти [9]. В этом разделе рассматриваются различные классы моделей, применяемые для создания цифровых двойников.

Физическая модель обеспечивает компьютерное моделирование физических процессов, протекающих во времени. Построение таких моделей на практике связано с использованием современных пакетов моделирования, например, таких, как ANSYS. Основные физические модели, используемые для построения цифровых двойников:

- модели на основе уравнения Навье—Стокса (моделирование гидро-газодинамических процессов);
- модели на основе метода конечных элементов (моделирование конструкций из твердых материалов).

Пример применения физической модели для построения цифрового двойника можно найти в статье [31]. В данной статье представлены результаты численного моделирования рассеивания в атмосфере загрязнителей воздуха, образующихся при отводе дымовых газов, образующихся при сгорании топлива на ТЭС. Модель использует уравнение Навье—Стокса и реализована с использованием пакета ANSYS CFX-CFD.

Оптимизационная модель обеспечивает поиск оптимумов целевой функции при наличии ограничений с использованием математических методов. Структура оптимизационной модели состоит из целевой функции, области допустимых решений и системы ограничений, определяющих эту область. Главной задачей оптимизационного моделирования является нахождение экстремума функций при ограничениях в форме уравнений и неравенств.

Основные подходы:

- линейное программирование;
- смешанное целочисленное программирование (Mixed Integer Programming);
- нелинейное программирование.

Примером применения оптимизационных моделей является оптимизация программы озеленения улиц города [43]. В данном исследовании предпринимается попытка определить модель планирования, которая определяет местоположение и тип зеленых насаждений на основе их множественных эффектов (например, охлаждение и улучшение связности), и рассчитывает стоимость реализации с использованием мета-эвристических алгоритмов оптимизации. Еще одним примером может служить оптимизация дорожного строительства [10]. В статье решается задача поиска наилучшего плана строительства с учетом предварительных условий и взаимозависимостей пользы от завершения проектов. Задача формулируется как двухуровневая проблема, где объективная функция заключается в минимизации обобщенных затрат, а нижний уровень учитывает выбор маршрута водителями.

Имитационное моделирование представляет собой метод исследования, при котором изучаемая система заменяется компьютерной моделью, имитирующей с достаточной точностью процессы, происходящие в реальной системе, с которой проводятся эксперименты с целью получения информации об этой системе [28]. Яркими представителями имитационных моделей могут служить модели распределения энергоресурсов конечных потребителей [25]. Еще одним ярким примером может служить прототип системы, представленный в статье [14]. Прототип включает в себя трехмерную модель застроенной среды, модель уличной сети с использованием теории и метода космического синтаксиса, моделирование городской мобильности, моделирование ветрового потока и ряд эмпирических количественных и качественных данных с использованием добровольной географической информации (volunteered geographic information). Кроме того, городской цифровой двойник был реализован в платформе визуализации для виртуальной реальности.

2.1. Модели на основе интеллектуального анализа данных

Модели на основе *интеллектуального анализа данных (Data Mining)* применяются для обнаружения в данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия стратегически важных решений в различных сферах человеческой деятельности [19]. В рамках построения цифрового двойника города модель на основе интеллектуального анализа данных позволяет решать следующие основные задачи:

- выявление влияния факторов друг на друга;
- выявление степени влияния факторов на показатель;
- прогнозирование значений факторов и показателя.

Таблица

Показания датчиков

Дата	Метеофакторы			Внутренние факторы			Внешние факторы		...	Показатель
	Сила ветра, м/с	Влажность воздуха, %	t° воздуха, $^{\circ}C$	Интенс-ть город. трафика, 1..10	Выбросы промышл. предпр-й города, тыс. тонн	Чистота улиц, 1..10	Выбросы промышл. предпр-й области, тыс. тонн	Въезд-й грязный транспорт, %		
10 апр	5	40	12	5	1,3	4	4	65		2
11 апр	7	60	14	4	0,6	8	2,1	40		1
12 апр	8	65	13	3	0,5	9	1,2	30		0,3
13 апр	10	80	16	10	1,6	3	3,9	85	...	3
...										

Рассмотрим методы решения указанных задач на примере анализа концентрации мелкодисперсных взвешенных частиц в городском воздухе (далее данный показатель обозначается как n и в качестве единиц измерения используется ПДК — предельно допустимая концентрация). Пусть имеется база данных, содержащая показания, приведенные в таблице.

Выявление влияния факторов друг на друга осуществляется с помощью методов *поиска шаблонов (pattern mining)* [6]. *Шаблон* представляет собой устойчиво выполняющееся правило вида «ЕСЛИ *антецедент*, ТО *консеквент*», где в качестве антецедента и консеквента фигурируют непустые и непересекающиеся между собой множества факторов. Пример двух таких правил представлен ниже:

- ЕСЛИ Сила ветра > 10 м/с И Влажность > 65% ТО Грязный транспорт > 75%;
- ЕСЛИ Дата = Понедельник ТО Выбросы пром. предприятий города > 1.5 тыс. тонн И Выбросы пром. предприятий области > 3 тыс. тонн.

Устойчивость выполнения правила определяется посредством двух основных мер: поддержка и достоверность. *Поддержка правила (rule support)* представляет собой вероятность наличия в базе данных записей, в которых присутствует как антецедент, так и консеквент. *Достоверность правила (rule confidence)* представляет собой условную вероятность наличия в базе данных записей, в которых присутствует антецедент, при условии

присутствия в этих записях консеквента. В качестве шаблонов на практике выбирают правила, поддержка и достоверность которых превышают пороговые значения ($minsup$ и $minconf$ соответственно), наперед заданные экспертом в данной предметной области. Для поиска шаблонов основными алгоритмами являются Apriori [7], FP-Growth [22, Eclat [45], а также их многочисленные модификации, и параллельные и распределенные версии.

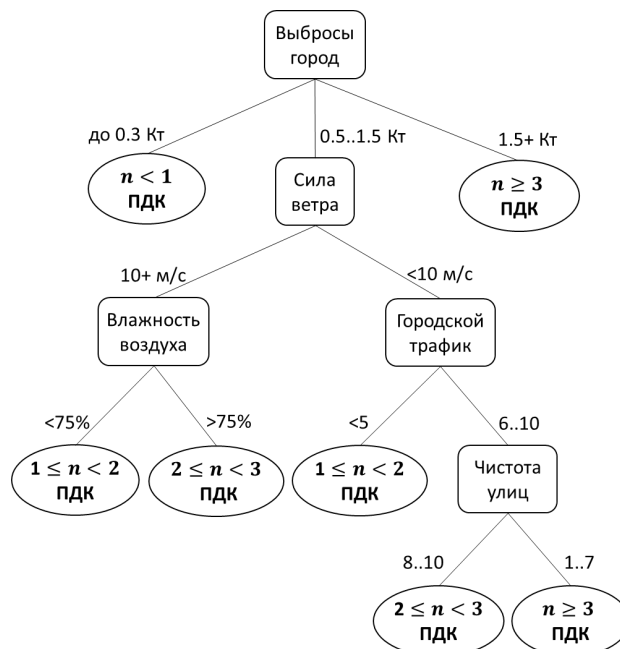


Рис. 2. Пример использования классификационной модели интеллектуального анализа данных

Выявление степени влияния факторов на показатель может быть выполнено как построение иерархии факторов, отражающей степень влияния фактора на показатель: чем выше фактор в иерархии, тем больше его влияние на показатель. Решение данной задачи связано с классификацией данных в базе данных сенсоров на основе деревьев решений [36]. На рис. 2 приведен пример построенного дерева решений, позволяющего классифицировать дни, в которые производились наблюдения, по значению показателя «Концентрация мелкодисперсных взвешенных частиц» со следующими классами значений: «не более нормы», «превышение нормы до одной ПДК», «превышение нормы до двух ПДК», «превышение нормы более двух ПДК».

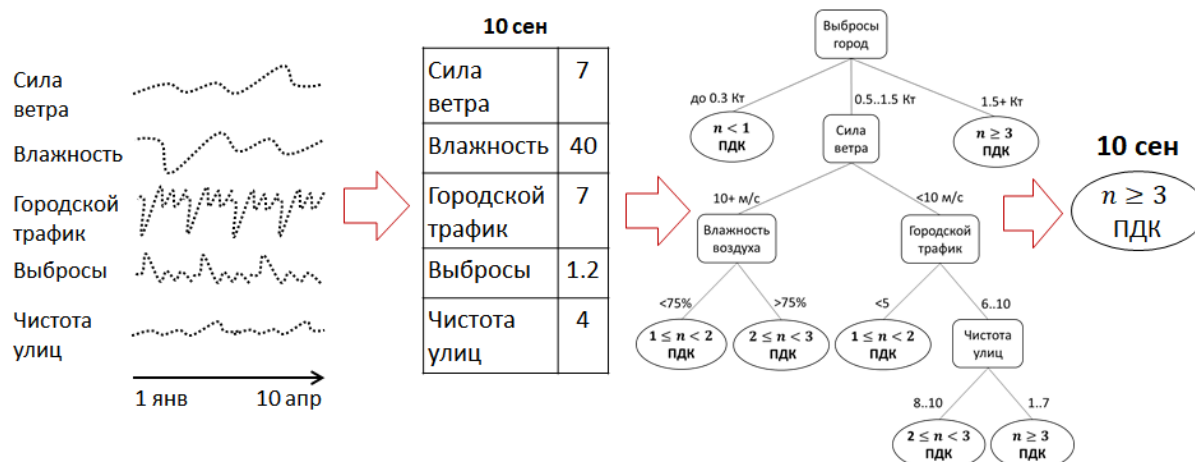


Рис. 3. Модель предсказания предельно допустимой концентрации

Задача прогнозирования значений факторов и показателя может быть решена как на базе нейросетевых моделей (см. раздел 2.2), так и в рамках моделей интеллектуального анализа данных. В последнем случае решение задачи предполагает два этапа (см. рис. 3). На первом этапе выполняется прогноз значений факторов на основе регрессионных моделей [15]. На втором этапе, используя полученные прогнозные значения показателей и ранее построенную классификационную модель, вычисляется прогнозное значение показателя.

2.2. Нейросетевые модели

Искусственная нейронная сеть (ИНС) — математическая модель, а также ее программное или аппаратное воплощение, построенные по принципу организации и функционирования биологических нейронных сетей [47]. ИНС представляет собой систему соединенных и взаимодействующих между собой искусственных нейронов. Каждый нейрон подобной сети имеет дело только с сигналами, которые он периодически получает, и сигналами, которые он периодически посылает другим нейронам. В результате обучения нейронная сеть способна выявлять сложные зависимости между входными данными и выходными, а также выполнять обобщение и сможет вернуть верный результат на основании данных, которые отсутствовали в обучающей выборке, а также неполных и/или «зашумленных», частично искаженных данных.

Сверточная нейронная сеть (СНС) — специальная архитектура искусственных нейронных сетей, нацеленная на анализ изображений. Сверточные нейронные сети состоят из нескольких слоев: сверточные слои, субдискретизирующие (subsampling, подвыборка) и слои «обычной» нейронной сети (полносвязные) — персептрона. Первые два слоя (сверточный и субдискретизирующий), чередуясь между собой, формируют входной вектор признаков для персептрона [8]. Сверточный слой включает в себя ядра свертки, которые являются матрицами весов, обрабатывающие предыдущие слои по фрагментам. На данный момент существует множество реализаций сверточных нейронных сетей для решения задач умного города. Приведем некоторые примеры. Detectron2 [42] — программная система Facebook AI Research [16], в которой реализованы современные алгоритмы обнаружения объектов такие, как Faster R-CNN, Mask R-CNN, RetinaNet, DensePose и другие. YOLOv3 [37] — система обнаружения объектов в реальном времени. Благодаря широкому диапазону доступных вариантов можно выбрать версию, наиболее подходящую для приложения. Например, Tiny YOLO [17] — это самый «компактный» вариант, который может работать даже на смартфонах.

Нейросетевые модели применяются для решения задач следующих классов.

- Детекция объектов на изображении. Например, детектирование типа автотранспорта и наличия грязи на нем или мусора, в том числе и нескольких объектов сразу на одном изображении.
- Выявления сложных нелинейных зависимостей между исходными данными и целевыми показателями. Например, построение моделей зависимости концентрации мелкодисперсных частиц от различных факторов, оценка степени влияния факторов.
- Построение рекомендательных систем. Например, систем поддержки принятия решений для управления городским трафиком с целью уменьшения вредных выбросов в атмосферу.

Примером нейросетевой модели для оценки городского трафика на перекрестке является «Интеллектуальная система мониторинга транспортных потоков и дорожно-транспортной инфраструктуры AIMS (Artificial Intelligence Monitoring System)» [39]. Целью проекта являлась разработка и внедрение системы оценки эффективности использования



Рис. 4. Пример работы нейронной сети для распознавания автомобилей

дорожной инфраструктуры, прогнозирования транспортных заторов и суммарных токсичных выбросов от автотранспорта. Использование нейронных сетей глубокого обучения позволяет в режиме реального времени осуществлять сбор, интерпретацию и агрегацию данных по интенсивности и классификации дорожного трафика.

Интеллектуальная система мониторинга транспортных потоков и дорожно-транспортной инфраструктуры позволяет решить сразу ряд сложных задач, таких как сбор, интерпретация и агрегация данных дорожного трафика, выявление недоиспользованных ресурсов в транспортной инфраструктуре, что позволяет сократить капитальные и эксплуатационные затраты. Система позволяет оценить эффективность решений по организации дорожного движения в режиме реального времени (изменения циклов работы светофорных объектов, перенос места нанесения разметки и др.), прогнозировать суммарные выбросы токсичных выхлопных газов от автотранспорта с учетом атмосферных и климатических условий. Также программа предупредит о превышении ПДК в узлах (перекрестках) улично-дорожной сети, что позволит не допустить негативного сценария (увеличив пропускную способность узла, ограничив движения для грузовиков).

Технология с использованием обученных нейросетей, не требует больших затрат на серверное оборудование и видеокамеры. Для мониторинга крупного перекрестка, зачастую, достаточно установки одной камеры уличного видеонаблюдения [39]. Пример работы такой системы приведен на рис. 4.

Еще одним примером использования нейросетевых технологий является задача детектирования мусора на улицах города [44]. Общая структура схемы работы модели показана на рис. 5.

Область переднего плана, совместные тепловые карты и совместные позиции получают путем вычитания фона и совместной оценки. Затем алгоритм отслеживает пешеходов и обнаруживает объект, который они несут. После обнаружения объекта он отслеживается с помощью трекера фильтра корреляции, который работает в режиме реального времени. Одновременно, расстояние постепенно моделируется между суставом каждого человека и отслеживаемым объектом. Если обнаружено изменение расстояния между объектом и человеком, алгоритм выдает предупреждение о сбросе мусора. Экспериментальные результаты показали высокую эффективность предложенного метода.

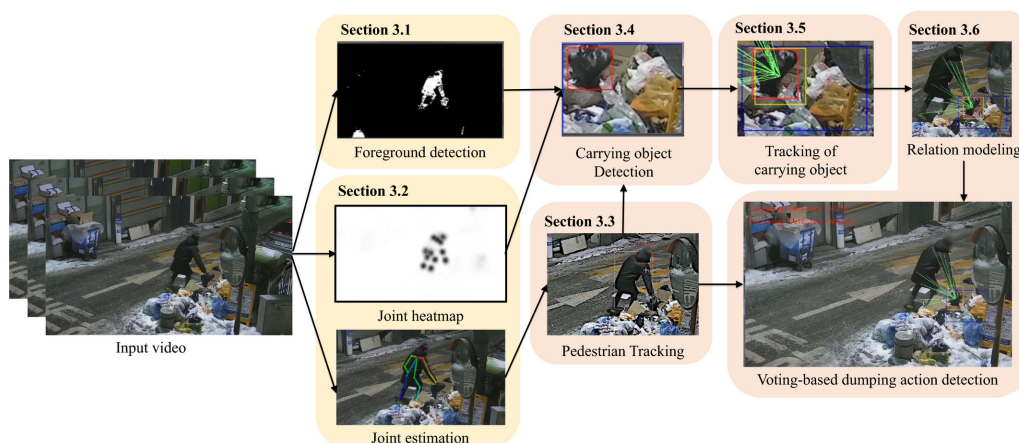


Рис. 5. Пример работы нейронной сети для определения мусора на изображении

3. Информационная инфраструктура цифрового двойника города

Под информационной инфраструктурой цифрового двойника города подразумеваются следующие компоненты:

- датчики, которые собирают информацию;
- сети, которые занимаются передачей данных;
- программные системы, которые занимаются очисткой данных;
- программные комплексы, которые занимаются интеллектуальным анализом данных.

Для того, чтобы ЦД мог обеспечить синхронизацию между состоянием процесса в реальном мире и его виртуальной копией, мы должны обеспечить ему возможность получать, передавать и анализировать поток данных от интеллектуальных конечных устройств Интернета вещей.

3.1. Сенсорные сети

Для сбора информации могут применяться как существующие инфраструктуры датчиков, так и разворачиваться новые сенсорные кластера. Можно выделить следующие типовые виды источников данных:

- видеокамеры;
- датчики давления;
- датчики влажности;
- датчики температуры;
- датчики загрязнения воздуха;
- геолокационные данные транспортных средств;
- информация с проездных карт пассажиров и др.

Выделяют несколько уровней связи, через которые проходят данные, собираемые с сенсорных сетей. Одним из вариантов такой классификации является разбиение каналов связи по зоне охвата на уровни полевых, локальных и глобальных вычислительных сетей [13].

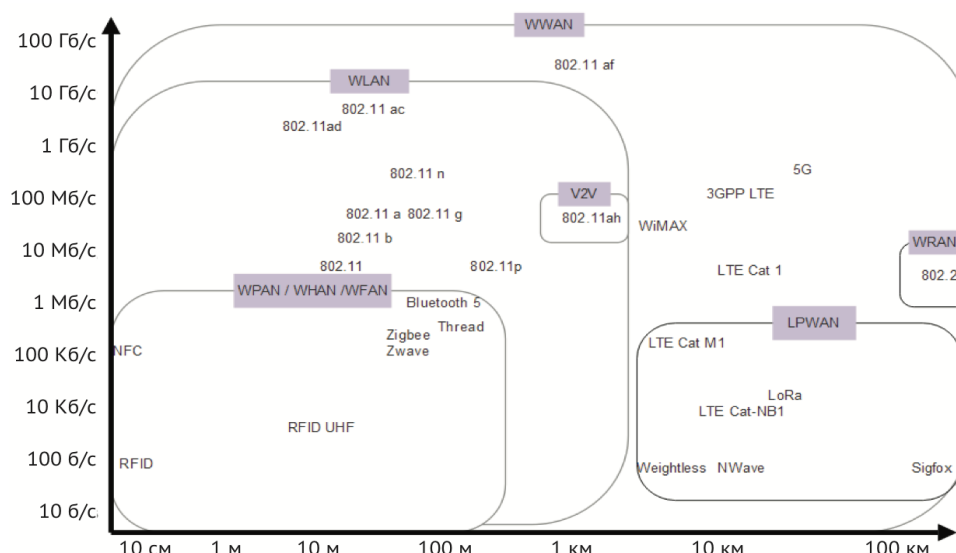


Рис. 6. График сравнения скорости и дальности передачи данных различными каналами связи [13]

Полевой уровень связи обеспечивает передачу данных на расстояния от десятков сантиметров до сотни метров, от места их генерации до первого узла на краю сети, отвечающего за их сбор и дальнейшую ретрансляцию. Для обмена данными на этом уровне применяются низкоскоростные маломощные информационные каналы. Среди коммуникационных протоколов данного уровня можно отметить протоколы для промышленных сетей Modbus и HART, которые ориентированы на работу на базе проводных сетей типа RS-232/RS-485 и «Токовая петля». В последнее десятилетие также значительно расширился набор технологий, обеспечивающих возможности беспроводной передачи данных от сенсорных систем. Среди наиболее известных беспроводных протоколов полевого уровня можно назвать: Bluetooth; WirelessHART; ZigBee (стандарт IEEE 802.15.4); Z-Wave; NFC; RFID и другие (см. рис. 6).

Уровень локальных вычислительных сетей обеспечивает передачу данных в рамках локальной сети, развернутой в непосредственной близости от источников данных (на расстояниях от сотен метров до километра). Сегодня на этом уровне чаще всего используются стандартные сетевые решения, базирующиеся на стеке коммуникационных протоколов TCP/IP. Для организации проводных сетей используются решения на базе семейства технологий Ethernet, а беспроводные сети реализуются на базе семейства протоколов Wi-Fi (IEEE 802.11).

На уровне глобальных вычислительных сетей обеспечивается передача данных в рамках сети Интернет. За функционирование данного уровня отвечают телекоммуникационные операторы, в том числе операторы сетей сотовой связи, спутниковой связи, маломощных глобальных сетей (Low-Power Wide-Area Network, LPWAN, см. рис. 6). Для организации передачи и эффективной обработки данных в глобальных сетях обычно применяются транспортные протоколы интернета для IoT и сетевых устройств, обеспечивающие возможности синхронной либо асинхронной обработки данных, такие как MQTT, CoAP, HTTP [1]. Для проводной передачи данных применяются решения магистральных телекоммуникационных линий, такие как Gigabit Ethernet, EPON и GPON. Для беспроводной передачи данных IoT могут быть организованы маломощные глобальные сети, в том числе на базе протокола LoRa. Отдельно стоит отметить возможности использования сетей сотовой связи для обеспечения прямого соединения устройств IoT с узлами в сети Интернет на основе технологий 2G (GSM), 3G, LTE, 5G.

3.2. Очистка данных

На протяжении жизненного цикла сенсоров данные, выдаваемые ими, могут «загрязняться». Загрязнение поступающих данных проявляется в виде некорректных, аномальных или пропущенных значений. Причиной загрязнения могут быть человеческий фактор, отказ сенсора, сбои в линиях связи, плановое техобслуживание сенсора и др. Загрязненные данные обесценивают их последующее применение в моделях на основе интеллектуального анализа данных и нейронных сетей. Для использования в цифровом двойнике города данные нуждаются в очистке [41]. Очистка данных, поступающих с сенсоров, подразумевает следующие основные операции: поиск аномалий и восстановление пропущенных значений.

Поиск аномалий предполагает нахождение промежутков и соответствующих значений временного ряда, которые существенно не похожи на значения во всех остальных промежутках этого ряда [12]. *Восстановление пропущенных значений* предполагает генерацию синтетических значений временного ряда вместо отсутствующих или ошибочных показаний на основе ретроспективного анализа значений временного ряда данного сенсора и/или значений временных рядов сенсоров, географически/логически близких к данному [11]. Решение указанных задач может быть выполнено как методами интеллектуального анализа данных, так и на основе нейросетевых моделей [27, 35].

При создании цифрового двойника города важной компонентой системы является *модуль очистки данных*, который обеспечивает поиск аномалий и восстановление пропущенных значений в потоках данных, поступающих с сенсоров системы. Работа модуля очистки данных организуется в соответствии со следующими основными принципами. Во-первых, перед поступлением в ЦОД поток данных каждой группы сенсоров должен обрабатываться в соответствии со своими правилами очистки. Во-вторых, очищенный поток данных каждой группы сенсоров должен регулярно контролироваться в соответствии со своими правилами верификации.

Правило очистки определяет область корректных значений для каждой группы сенсоров (например, минимальное и максимальное значения показаний, зависимость показаний сенсора данной группы от показаний сенсоров других групп и др.). *Правило верификации* для данной группы сенсоров определяет периодичность и способ оценки правдоподобия значений, генерируемых синтетически, вместо пропущенных или аномальных. Для оценки правдоподобия используются различные меры (например, нормализованная средняя абсолютная ошибка, средняя квадратическая ошибка и др. [23]) и соответствующие группе сенсоров предельные значения этих мер. Низкая степень правдоподобия синтетических значений, выявленная в результате верификации, является поводом для пересмотра соответствующих методов и алгоритмов, реализованных в модуле очистки данных.

3.3. Туманные вычисления

Для обработки информации, поступающей от множества источников интернета вещей (IoT) целесообразно применение концепции туманных вычислений.

Туманные вычисления (ТВ, англ. *fog computing*) — это многоуровневая модель-расширение облачных вычислений (англ. *cloud computing*), облегчающая развертывание распределенных приложений и услуг, учитывающих сетевые задержки, на, так называемых, *туманных узлах* (англ. *fog nodes*) (физических или виртуальных), располагающихся между умными конечными устройствами и централизованными (облачными) сервисами. *Туманные узлы* являются контекстно-зависимыми и поддерживают единую систему управления данными и организации связи. Они могут быть организованы в кластеры

вертикально (для поддержки изоляции), горизонтально (для поддержки федераций сервисов) или в привязке к сетевой близости от конечных умных устройств. Туманные вычисления минимизируют время сетевого отклика поддерживаемых приложений, а также обеспечивают конечные устройства локальными вычислительными ресурсами и, при необходимости, сетевым подключением к централизованным сервисам [24].

В то время как *облачные вычисления* обеспечивают потенциально безграничные ресурсы для решения задач, требующих существенных вычислительных ресурсов, туманные вычисления обеспечивают:

- предоставление вычислительных ресурсов в непосредственной близости от конечных устройств;
- предобработку данных перед их отправкой в облако;
- решение задач, которые требуют минимального времени отклика.

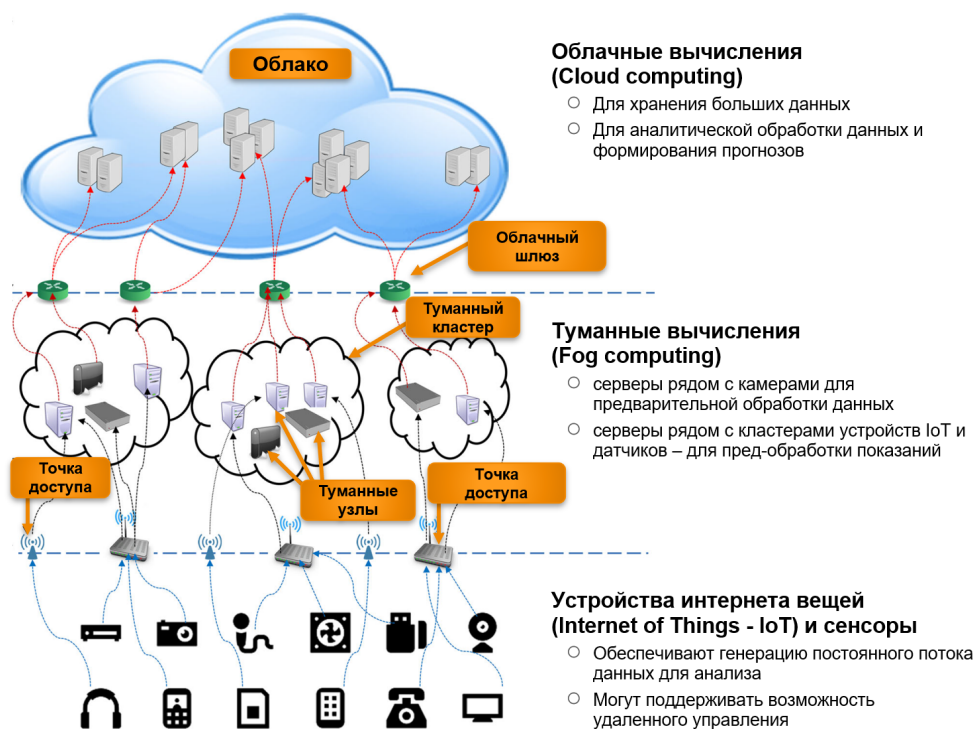


Рис. 7. Модель туманных вычислений

Примером использования туманных вычислений для организации эффективной обработки данных, может служить задача предобработки данных с камер видеонаблюдения. Сегодня в городе могут размещаться сотни камер наружного наблюдения, обеспечивающие покрытие сервисами видеомониторинга большой площади объектов, имеющих особый интерес для управления городским хозяйством. Если для их анализа применяются ресурсы единого центра обработки данных (облака), то задача передачи потока видеоданных по магистральным каналам от такого большого массива камер может быть значительно затруднена. Для решения этой проблемы приходится прибегать к определенным компромиссам, включающим в себя:

- сильное сжатие видеопотока;
- уменьшение разрешения передаваемой видеoinформации;
- ограничение количества передаваемых кадров в секунду и др.

Такие решения приводят к тому, что интеллектуальная модель, применяемая в облачной системе для анализа поступающих данных, может испытывать недостаток качественной

информации для обеспечения эффективной работы. Возможным решением этой проблемы является развертывание сети туманных узлов в непосредственной близости с источниками видеоданных. Эти узлы потребляют и проводят предварительный анализ видеоданных в оригинальном качестве, обеспечивая решение таких задач предобработки видеоданных, как:

- сегментация и выделение объектов интереса;
- удаление фона;
- интеллектуальное сжатие видеопотока.

Это позволяет значительно снизить объем и повысить качество данных, передаваемых в облако, повышая качество работы системы интеллектуального анализа данных [20, 30, 34].

На рис. 7 изображена общая модель передачи и обработки данных от сенсоров и устройств интернета вещей (IoT) при реализации концепции туманных вычислений. На нижнем уровне модели располагаются устройства IoT, обеспечивающие генерацию потоков данных. Это данные различной природы, различного формата и различной скорости генерации. Устройства IoT напрямую связаны с точками доступа, обеспечивающими их соединение с сетью Интернет.

В непосредственной близости к точкам доступа развернута инфраструктура кластеров туманных узлов, обеспечивающих предобработку поступающих потоков данных. Согласно [18], порядка 40% вычислительной нагрузки по обработке данных систем интернета вещей будет обеспечено на туманных узлах. Также, в силу географической близости к узлам IoT, туманные узлы обеспечивают низкую латентность при необходимости двусторонней коммуникации с конечными узлами.

Предобработанные данные передаются для дальнейшей обработки в облако. В отличие от туманных узлов, инфраструктура облачных вычислений подразумевает предоставление по требованию практически безграничных ресурсов для реализации вычислительных задач.

Заключение

В рамках данной статьи были проанализированы существующие технологии построения цифровых двойников городов. Было рассмотрено определение и ключевые компоненты, составляющие цифровой двойник города, примеры решений, которые уже сегодня можно считать первыми шагами к построению полноценного цифрового двойника. Нами были рассмотрены ключевые методы моделирования, реальных объектов и технологических процессов, которые находят свое применение при создании цифровых двойников. Отдельно, на практических примерах, были рассмотрены модели на основе интеллектуального анализа данных, и нейросетевые модели. Также, нами было дано описание ключевых элементов информационной инфраструктуры цифрового двойника города и были рассмотрены ключевые элементы этой инфраструктуры, такие как сенсорные сети, совмещенные с системами очистки сырых данных. Отдельно было отмечено, что для успешного построения такой системы необходимо формирование информационной среды на основе концепции модели туманных вычислений, так как задачи предварительной обработки потоков данных часто требуют существенных вычислительных ресурсов на предобработку данных, расположенных в непосредственной близости от источников.

Исследование выполнено при финансовой поддержке Министерства науки и высшего образования РФ (государственное задание FENU-2020-0022) и Российского фонда фундаментальных исследований (проекты №№ 20-07-00092-а, 20-07-00140-а и 18-07-01224-а).

Литература

1. Ли П. Архитектура интернета вещей. М.: ДМК Пресс, 2019. 454 с.
2. Минстрой России. Индекс IQ городов по итогам 2018 года. URL: <https://minstroyrf.gov.ru/docs/57570/> (дата обращения: 14.09.2020).
3. От концепции до прикладных решений. Города, управляемые данными. 2016. URL: https://www.pwc.ru/ru/government-and-public-sector/assets/ddc_rus.pdf (дата обращения: 14.09.2020).
4. Повх Е. Десять цифровых двойников городов. 2020. URL: <https://realty.rbc.ru/news/5e297b079a79478024d54ff6> (дата обращения: 14.09.2020).
5. Умные города Росатома. URL: <https://rosatom.city/> (дата обращения: 14.09.2020).
6. Aggarwal C.C., Han J. Frequent pattern mining. Springer, 2014. 480 p. DOI: 10.1007/978-3-319-07821-2.
7. Agrawal R., Srikant R. Fast algorithms for mining association rules in large databases // Proceedings of 20th International Conference on Very Large Data Bases, VLDB'94 (Santiago de Chile, Chile, September, 12–15, 1994). 1994. P. 487–499.
8. Albawi S., Mohammed T.A., Al-Zawi S. Understanding of a convolutional neural network // The International Conference on Engineering and Technology 2017 (Antalya, Turkey, August, 21–23, 2017). 2017. P. 1–6. DOI: 10.1109/ICEngTechnol.2017.8308186.
9. Al-Jarrah O.Y., Yoo P.D., Muhaidat S., et al. Efficient machine learning for Big Data: A review // Big Data Research. 2015. Vol. 2, no. 3. P. 87–93. DOI: 10.1016/j.bdr.2015.04.001.
10. Bagloee S.A., Sarvi M., Patriksson M., et al. Optimization for roads' construction: selection, prioritization, and scheduling // Computer-Aided Civil and Infrastructure Engineering. 2018, Vol. 33, no. 10. P. 833–848. DOI: 10.1111/mice.12370.
11. Batista G.E.A.P.A., Monard M.C. An analysis of four missing data treatment methods for supervised learning // Appl. Artif. Intell. 2003. Vol. 17, no. 5-6. P. 519–533. DOI: 10.1080/713827181.
12. Chandola V., Banerjee A., Kumar V. Anomaly detection: A survey // ACM Comput. Surv. 2009. Article no. 15. DOI: 10.1145/1541880.1541882.
13. Data-driven smart cities: Big Data, analytics and security. 2018. URL: <https://skelia.com/articles/data-driven-smart-cities-big-data-analytics-and-security/> (дата обращения: 14.09.2020).
14. Dembski F., Wössner U., Letzgus M., et al. Urban digital twins for smart cities and citizens: the case study of Herrenberg, Germany // Sustainability. 2020. Vol. 12, art. 2307. DOI: 10.3390/su12062307.
15. Draper N.R., Smith H. Applied regression analysis. Wiley, 1981. 707 p.
16. Facebook AI Research. URL: <https://ai.facebook.com/> (дата обращения: 20.05.2020).
17. Fang W., Wang L., Ren P. Tinier-YOLO: A real-time object detection method for constrained environments // IEEE Access. 2019. Vol. 8. P. 1935–1944. DOI: 10.1109/ACCESS.2019.2961959.
18. Fernández-Cerero D., Fernández-Montesa A., Ortega F.J., et al. Sphere: simulator of edge infrastructures for the optimization of performance and resources energy consumption // Simulation Modelling Practice and Theory. 2020. Vol. 101. P. 101966. DOI: 10.1016/j.simpat.2019.101966.
19. Frawley W.J., Piatetsky-Shapiro G., Matheus C.J. Knowledge Discovery in databases: an overview // Knowledge Discovery in Databases. 1991. P. 1–30.

20. Gardašević G., Berbakov L., Mastilović A. Cybersecurity of Industrial Internet of Things // *Cyber Security of Industrial Control Systems in the Future Internet Environment*. 2020. P. 47–68. DOI: 10.4018/978-1-7998-2910-2.ch003.
21. Glaessgen E., Stargel D. The digital twin paradigm for future NASA and U.S. air force vehicles // *Proceedings of the 53rd AIAA/ASME/ASCE/AHS/ASC conference on structures, structural dynamics and materials conference* (Honolulu, Hawaii, USA, April, 23–26, 2012). 2012. P. 1818. DOI: 10.2514/6.2012-1818.
22. Han J., Pei J., Yin Y. Mining frequent patterns without candidate generation // *ACM SIGMOD Record*. 2000. Vol. 29, no. 2. P. 1–12.
23. Hyndman R.J.; Koehler A.B. Another look at measures of forecast accuracy // *International Journal of Forecasting*. 2006. Vol. 22, no. 4. P. 679–688. DOI: 10.1016/j.ijforecast.2006.03.001.
24. Iorga M., Feldman L., Barton R., et al. Fog computing conceptual model. 2018. URL: <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.500-325.pdf> (дата обращения: 14.09.2020).
25. Iqbal M.N., Kütt L. End-user electricity consumption modelling for power quality analysis in residential building // *19th Int. Scientific Conf. on Electric Power Engineering, EPE. IEEE*, 2018. P. 1–6. DOI: 10.1109/EPE.2018.8396030.
26. Kazuhiko I., Atsush Y. Building a common smart city platform utilizing FIWARE (case study of Takamatsu city) // *NEC Tech. J.* 2018. Vol. 13, no 1. P. 28–31.
27. Kim J., Tae D., Seok J. A survey of missing data imputation using generative adversarial networks // *Proc. of the 2020 Int. Conf. on Artificial Intelligence in Information and Communication, ICAIIC 2020*. P. 454–456. DOI: 10.1109/ICAIIIC48513.2020.9065044.
28. Kin W., Chan V. *Foundations of simulation modeling*. Wiley Encyclopedia of Operations Research and Management Science. Wiley, 2011. P. 6408.
29. Korambath P., Wang J., Kumar A., et al. A smart manufacturing use case: furnace temperature balancing in steam methane reforming process via Kepler workflows // *Procedia Computer Science*. 2016. Vol. 80. P. 680–689. DOI: 10.1016/j.procs.2016.05.357.
30. Kuplyakov D.A., Shalnov E.V., Konushin V.S., et al. A distributed tracking algorithm for counting people in video // *Programming and Computer Software*. 2019. Vol. 45, no. 4. P. 163–170. DOI: 10.1134/S0361768819040042.
31. Mărunțălu O., Lăzăroiu G., Manea E.E., et al. Numerical simulation of the air pollutants dispersion emitted by CHP using ANSYS CFX // *World Academy of Science, Engineering and Technology, International Journal of Environmental, Chemical, Ecological, Geological and Geophysical Engineering*. 2015. Vol. 9. P. 1058–1064. DOI: 10.5281/zenodo.1108252.
32. Mohammadi N., Taylor J.E. Smart city digital twins // *Proceedings of the 2017 IEEE Symposium Series on Computational Intelligence* (Honolulu, HI, USA, Nov., 27 – Dec., 1, 2017). IEEE, 2017. P. 1–5. DOI: 10.1109/SSCI.2017.8285439.
33. Nam T., Pardo T.A. Conceptualizing smart city with dimensions of technology, people, and institutions // *Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times* (College Park, MD, USA, Jun, 12–15, 2011). ACM, 2011. P. 282–291. DOI: 10.1145/2037556.2037602.
34. Nikouei S.Y., Chen Y., Aved A., et al. I-ViSE: interactive video surveillance as an edge service using unsupervised feature queries // *IEEE Internet of Things Journal* (Early Access). 2020. DOI: 10.1109/JIOT.2020.3016825.
35. Osman M.S., Abu-Mahfouz A.M., Page P.R. A survey on data imputation techniques: water distribution system as a use case // *IEEE Access*. 2018. Vol. 6. P. 63279–63291. DOI: 10.1109/ACCESS.2018.2877269.

36. Quinlan J.R. Induction of decision trees // Machine Learning. 1986. Vol. 1, no. 1. P. 81–106. DOI: 10.1023/A:1022643204877.
37. Redmon J., Farhadi A. You only look once: unified: real-time object detection // IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016 (Las Vegas, NV, USA, June, 27–30, 2016). P. 779–789. DOI: 10.1109/CVPR.2016.91.
38. Ruohomaki T., Airaksinen E., Huuska P., et al. Smart city platform enabling digital twin // 2018 International Conference on Intelligent Systems (Funchal-Madeira, Portugal, September, 25–27, 2018). IEEE, 2018. P. 155–161. DOI: 10.1109/IS.2018.8710517.
39. Shepelev V., Aliukov S., Glushkov A., et al. Identification of distinguishing characteristics of intersections based on statistical analysis and data from video cameras // Journal of Big Data. 2020. Vol. 7, no. 1. P. 1–23. DOI: 10.1186/s40537-020-00324-7.
40. Smart cities readiness: smart cities maturity model and self-assessment tool, Scottish cities alliance. 2014. URL: https://www.scottishcities.org.uk/site/assets/files/1103/smart_cities_readiness_assessment_-_guidance_note.pdf (дата обращения: 14.09.2020)
41. Wang X., Wang C. Time series data cleaning: A survey // IEEE Access. 2020. Vol. 8. P. 1866–1881. DOI: 10.1109/ACCESS.2019.2962152.
42. Wu Y., Kirillov A., Massa F., et al. Detectron2. URL: <https://github.com/facebookresearch/detectron2> (дата обращения: 25.02.2020).
43. Yoon E.J., Kim B., Lee D.K. Multi-objective planning model for urban greening based on optimization algorithms // Urban Forestry & Urban Greening. 2019. Vol. 40. P. 183–194. DOI: 10.1016/j.ufug.2019.01.004.
44. Yun K., Kwon Y., Oh S., et al. Vision-based garbage dumping action detection for real-world surveillance platform // ETRI Journal. 2019. Vol. 41, no. 4. P. 494–505. DOI: 10.4218/etrij.2018-0520.
45. Zaki M.J. Scalable algorithms for association mining // IEEE Transactions on Knowledge and Data Engineering. 2000. Vol. 12, no. 3. P. 372–390.
46. Zanella A., Bui N., Castellani A., et al. Internet of things for smart cities // IEEE Internet of Things Journal. 2014. Vol. 1, no. 1. P. 22–32. DOI: 10.1109/JIOT.2014.2306328.
47. Zupan J. Introduction to artificial neural network methods: What they are and how to use them // Acta Chimica Slovenica. 1994. Vol. 41, no. 3. P. 327–352.

Иванов Сергей Александрович, к.ф.-м.н., доцент, кафедра системного программирования, Южно-Уральский государственный университет (национальный исследовательский университет) (Челябинск, Российская Федерация)

Никольская Ксения Юрьевна, ст. преподаватель, кафедра системного программирования, Южно-Уральский государственный университет (национальный исследовательский университет) (Челябинск, Российская Федерация)

Радченко Глеб Игоревич, к.ф.-м.н., доцент, директор Высшей школы электроники и компьютерных наук, Южно-Уральский государственный университет (национальный исследовательский университет) (Челябинск, Российская Федерация).

Соколинский Леонид Борисович, д.ф.-м.н., профессор, проректор по информатизации, Южно-Уральский государственный университет (национальный исследовательский университет) (Челябинск, Российская Федерация)

Цымблер Михаил Леонидович, д.ф.-м.н., доцент, начальник отдела интеллектуального анализа данных и виртуализации Лаборатории суперкомпьютерного моделирования, Южно-Уральский государственный университет (национальный исследовательский университет) (Челябинск, Российская Федерация)

DIGITAL TWIN OF A CITY: CONCEPT OVERVIEW

© 2020 S.A. Ivanov, K.Yu. Nikolskaya, G.I. Radchenko,
L.B. Sokolinsky, M.L. Zymbler

South Ural State University (pr. Lenina 76, Chelyabinsk, 454080 Russia)

E-mail: saivanov@susu.ru, nikolskaiaki@susu.ru, gleb.radchenko@susu.ru,

leonid.sokolinsky@susu.ru, mzym@susu.ru

Received: 15.09.2020

The urban economy is a complex multi-vector system. Creating a single digital twin of such a system is now a difficult task to solve. This article presents the concept of a digital twin of a city. The authors propose an evolutionary approach to this problem, according to which digital twins of individual elements of the urban environment are consistently built on a single hardware and software platform. These digital twins are linked in a single cooperative system that allows one digital twin to use data produced by other digital twins. The article gives the definition and architecture of such a system. It describes the classes of models that can be used to create digital twins. Special attention is paid to neural network models and models for data analysis. The information infrastructure of the digital twin of a city, including sensory networks, data cleansing mechanisms and nebulous calculations, is considered.

Keywords: digital twin, smart city, urban management, neural networks, data mining, sensors, fog computing.

FOR CITATION

Ivanov S.A., Nikolskaya K.Yu., Radchenko G.I., Sokolinsky L.B., Zymbler M.L. Digital Twin of a City: Concept Overview. *Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering*. 2020. Vol. 9, no. 4. P. 5–23. (in Russian) DOI: 10.14529/cmse200401.

This paper is distributed under the terms of the Creative Commons Attribution-Non Commercial 3.0 License which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is properly cited.