



Проблемы интеллектуального анализа данных при моделировании транспортных потоков мегаполиса



Куфтинова Наталья Григорьевна – Московский автомобильно-дорожный государственный технический университет (МАДИ), Москва, Россия.*

Наталья КУФТИНОВА

Рассматриваются проблематика использования в транспортной модели интеллектуального анализа данных как цифровой платформы анализа данных транспортных потоков мегаполиса и предпосылки создания в будущем единых банков данных и интегрированной среды взаимодействия моделей разных уровней как кластеров цифровой экономики, которые будут учитывать все виды транспорта для оценки транспортного спроса и создания проектов организации дорожного движения мегаполиса.

Целью данной работы является исследование процессов получения количественных характеристик объектов транспортного моделирования при создании единой электронной среды с помощью вычисления производных параметров транспортной сети мегаполиса. Количественные пространственные характеристики объекта связаны с вычислением удалённости от центра города и магистральной улицы и определяются с помощью геоинформационных систем, вследствие чего возникает проблема унификации и эффективного хранения данных.

В рамках реализации поставленной цели показано, что для всех первичных транспортных данных необходимо создавать процедуру предобработки,

а также валидации, так как источники данных имеют различный формат и пространственную интерполяцию для трековых данных. Для этого рекомендуется использовать различные методы анализа данных на основе ГИС-технологий, цифрового моделирования рельефа, топологии дорожной сети и других объектов транспортной сети мегаполиса, при этом до момента использования интеллектуальных данных требуется проведение работы по форматированию и группировке исходных данных в режиме реального времени. Наиболее распространённые ошибки возникают на этапе итерационного процесса для получения количественных характеристик объектов транспортного моделирования и построения оптимальных с точки зрения времени проезда маршрутов движения по некоторой транспортной сети.

Тенденции роста городов требуют глобальной цифровизации всех объектов транспортной инфраструктуры, учитывающих изменение функций транспортной среды и интенсивности транспортных потоков. Это диктует необходимость дальнейшей разработки и применения новых информационных технологий для обработки данных с помощью нейронных сетей и других цифровых технологий.

Ключевые слова: транспортная система, транспортный поток мегаполиса, интеллектуальный анализ данных, информационные и коммуникационные технологии, транспортные расчётные районы, ГИС-технологии.

*Информация об авторе:

Куфтинова Наталья Григорьевна – кандидат технических наук, доцент кафедры автоматизированных систем управления Московского автомобильно-дорожного государственного технического университета, (МАДИ), Москва, Россия, nat.gk@mail.ru.

Статья поступила в редакцию 21.04.2020, актуализирована 23.10.2020, принята к публикации 30.10.2020.

For the English text of the article please see p. 33.

ВВЕДЕНИЕ

Процесс взрывного развития информационных и коммуникационных технологий, использование сетей 5G способствуют созданию и накоплению больших массивов данных, повышают актуальность вопросов не только качественного сбора информации, но и создания единого источника данных, содержащего полную, актуальную и достоверную информацию о любой транспортной системе. На практике очень часто возникают сложности с возможностью применения для моделирования транспортных потоков неструктурированных данных, например, касающихся ёмкости транспортных расчётных районов, информации, необходимой для составления матрицы межрайонных корреспонденций. Получение таких данных связано также с оформлением официального договора и другими административными вопросами. Применение геоинформационных систем (ГИС) помогает преодолеть эти и иные проблемы, связанные с использованием статистических данных о функционировании транспортных потоков и площади паркирования в центрах массового тяготения населения мегаполиса.

В настоящем исследовании поставлена *цель* построения моделей пространственного распределения показателей объектов в виде непрерывных поверхностей на основе дискретно заданной информации с помощью ГИС-технологий для исследования закономерностей территориальных структур, обладающих свойством сплошного распространения, континуальное приближение которых зависит от степени аппроксимации и того, насколько в этой дискретной информации отражены географические закономерности. При использовании графического средства моделирования искусственных полей — искусственных изолиний (псевдоизолиний) — следует выполнять тщательную оценку репрезентативности данных, достоверности получаемых выводов и пригодности результатов аппроксимации для использования с другими слоями баз данных ГИС (БД ГИС).

Для достижения поставленной цели в исследовании были использованы *методы* прогнозирования скоростей и анализа информации о топологии дорожной сети

с помощью временных рядов и интеллектуального анализа данных для определения пространственных координат. Методы мониторинга и прогнозирования характеристик транспортного потока дорожной сети на основе пространственно-временного подхода к прогнозированию производных параметров позволяют за счёт анализа предыдущих периодов снизить влияние неполноты информации о текущем состоянии транспортного потока на прогнозные значения.

РЕЗУЛЬТАТЫ

Интеллектуальный анализ данных (Data Mining) — это современная концепция, изначально предполагающая, что данные могут быть неточными, разнородными, содержать пропуски и при этом иметь гигантские объёмы [1; 2]. Подготовка и точность таких данных зависят от качества процессов:

- определения и анализа требований к данным;
- сбора данных в хранилище;
- предварительной обработки данных для обеспечения качественного анализа.

Основными источниками информации для формирования позиционной части базы данных ГИС являются карты различного назначения. Общегеографические карты позволяют создавать информационные объекты, описывающие топографические особенности территории, в первую очередь, рельеф и гидрографическую сеть, которые используются практически во всех ГИС, независимо от их тематической направленности. Тематические карты (населения, экономики, природы) являются базовыми источниками для первоначального формирования и территориальной привязки соответствующих пространственных объектов. Атрибутивная информация по большей части добавляется позже из источников данных, которые не обязательно являются пространственно-координированными. Возможно, например, использование комплексных атласов. Материалы дистанционного зондирования Земли позволяют создавать пространственные объекты напрямую, без формирования карт как промежуточного информационного слоя. Источниками атрибутивной информации также являются многолетние гид-



рологические и метеорологические данные, статистические материалы в цифровой форме, прочие текстовые материалы. Главными условиями возможности использования таких данных являются существование, на момент их ввода, пространственных объектов, к которым они относятся, и возможность однозначного соотнесения атрибутивных данных с определённым объектом или их группой.

Всего можно выделить четыре основные концептуальные функции ГИС, из которых первые две (сбор данных и обработка данных) являются подготовительными и реализуются, чаще всего, однократно (за исключением ГИС, требующих постоянной актуализации своих данных, тогда их сбор и обработка выполняются неоднократно, с определенной периодичностью). В наибольшей степени возможность получения новых знаний проявляется при анализе пространственно-координированных данных, для чего средствами ГИС выполняется построение различных концептуальных и формальных моделей.

Главной функцией ГИС является информационная поддержка управленческих решений, принимаемых по результатам анализа и моделирования на основе пространственно-координированных данных. Прикладные функции (технологические процедуры), поддерживаемые соответствующими программными средствами ГИС, включают в себя широкий перечень возможностей. Ввод и редактирование данных позволяют создавать информационные объекты на основе как картографической информации, изначально основанной на пространственно-координированных данных, так и путём прямого ввода координат объекта. Поддержка различных пространственных моделей (регулярно-ячеистых, квадротомических, векторных) средствами ГИС позволяет на основе одинаковых исходных данных создавать различные виды информационных объектов. Хранение данных осуществляется с учётом наличия в них позиционной составляющей, для чего, как правило, используются достаточно узкоспециализированные, а для атрибутивных данных, наоборот, более широкие форматы. Для организации базы данных ГИС может использоваться иерархическая или сетевая

модель, но более распространёнными являются реляционные базы.

Преобразование систем координат является часто используемой прикладной функцией ГИС. Такая необходимость может возникать уже на стадии ввода данных при использовании нескольких источников с различными системами координат. В этом случае выполняется приведение данных к некоторой единой системе. Чаще всего производится преобразование географических координат в прямоугольные координаты (Гаусса—Крюгера), и наоборот. Для хранения позиционных данных пространственных объектов более удобна система географических координат; для их визуализации и выполнения на их основе пространственного анализа более подходящей является система координат Гаусса—Крюгера. Главная особенность преобразования координат из одной системы в другую заключается в том, что криволинейный участок на земной поверхности (поверхности сфероида) невозможно без погрешности расположить («растянуть») на плоскости. Величина этой погрешности нарастает по мере удаления точки от начала прямоугольных координат, поэтому система Гаусса—Крюгера всегда охватывает только относительно небольшую площадь, в пределах которой удаление от нулевой точки не превышает несколько сотен километров. При таких расстояниях погрешность координат отдельных точек, а значит, и определяемых на их основе длин и площадей остаётся на приемлемом уровне. Обычно все данные, используемые конкретной ГИС, хранятся в единой системе координат, и необходимость их преобразования при анализе данных возникает редко. В то же время достаточно частой операцией является трансформация картографических проекций. Это связано с неоднозначностью представления на плоскости картины, изначально находящейся на выпуклой поверхности земного шара. Наиболее часто используемыми моделями в ГИС являются растровая и векторная. Возможность их взаимного преобразования в автоматическом (или максимально близком к нему) режиме является ещё одной важной прикладной функцией ГИС. Возможности по анализу и преобразованию данных, обусловленные наличием в их составе по-

зиционной составляющей, реализуются группой соответствующих функций ГИС. К ней относятся измерительные операции и операции аналитической геометрии, такие как вычисление длин, объёмов, площадей, расстояний. Полигональные операции позволяют выполнять объединение и разъединение участков, а также выявлять факты их перекрытия и попадания на заданную площадь точек и линий. Пространственно-аналитические операции применяются для анализа близости объектов, построения буферных зон, сетевого анализа (определения связности или несвязности сети, выявления кратчайшего маршрута). Геомоделирование заключается в программной реализации законов изменения параметров объектов, в том числе во времени, с последующим моделированием поведения рассматриваемой системы объектов. Цифровое моделирование рельефа рассматривается как отдельная прикладная функция ГИС. С процедурой построения рельефа тесно связаны задачи его анализа, такие, как интерполяция высот и построение изолиний. Заключительной прикладной функцией ГИС является вывод данных (в том числе визуализация пространственных объектов, их атрибутов), экспорт данных в другие информационные системы [3–6].

Проблема использования интеллектуального анализа данных является обособленной не только в работах отечественных исследователей, но и в мировом научном сообществе, в практике использования распределённых вычислений и, в частности, облачных вычислительных сред, где применяется интеграция различных технологий интеллектуального анализа данных, обработки больших данных, распределённых и облачных вычислений.

Данная проблематика представлена в популярных и научных публикациях, в том числе и по транспортной политике. Так, например, газета New York Times обозначает девять проблем больших данных [7]. Анализ подобных данных очень сложен, и поэтому в последнее время большое количество исследований посвящено разработке необходимых для этого алгоритмов, включая алгоритмы интеллектуального анализа для параллельной обработки данных в системах облачных вычислений,

а также для моделирования транспортных процессов [8–12]. Например, методы интеллектуального анализа данных были использованы для создания моделей (классификаторов) для прогнозирования серьёзности травм в любой новой аварии с разумной точностью на основе 5973 записей о дорожно-транспортных происшествиях в Абу-Даби за шестилетний период с 2008 по 2013 год. Исследование было направлено на установление набора правил, которые могут использоваться транспортными организациями Объединённых Арабских Эмиратов (ОАЭ) для определения основных факторов, влияющих на тяжесть аварии [13].

В настоящее время одним из самых популярных средств Data Mining (например, по версии *KDnuggets*¹) является система RapidMiner, разработанная одноименной компанией. Данный продукт имеет открытый исходный код (Open Source), и версия с минимальными функциональными возможностями распространяется бесплатно. RapidMiner реализует клиент-серверную архитектуру. RapidMiner Server может использоваться отдельно, предоставляя возможности интеллектуального анализа в виде веб-сервисов, тем самым реализуя модель облачных вычислений — SaaS. RapidMiner реализует все необходимые операции для анализа: загрузку и преобразование данных (ETL), предобработку данных, их визуализацию, решение задач Data Mining [14]. Он имеет открытую архитектуру, предоставляя возможность расширять его новыми алгоритмами, в том числе и реализуемыми из библиотек Weka и R.

Исходными данными для рассматриваемого нами транспортного прогнозирования являются временные ряды скоростей и информация о топологии дорожной сети, представленная, например, в виде графа. На основе пространственно-координированных данных создаются пространственные информационные объекты (рис. 1). В показанном на рисунке примере реальный физический объект на земной поверхности (участок автодороги) представлен как совокупность прямолинейных сегментов. Точность такого

¹ См., напр.: [Электронный ресурс]: <https://www.kdnuggets.com/2017/05/poll-analytics-data-science-machine-learning-software-leaders.html>.



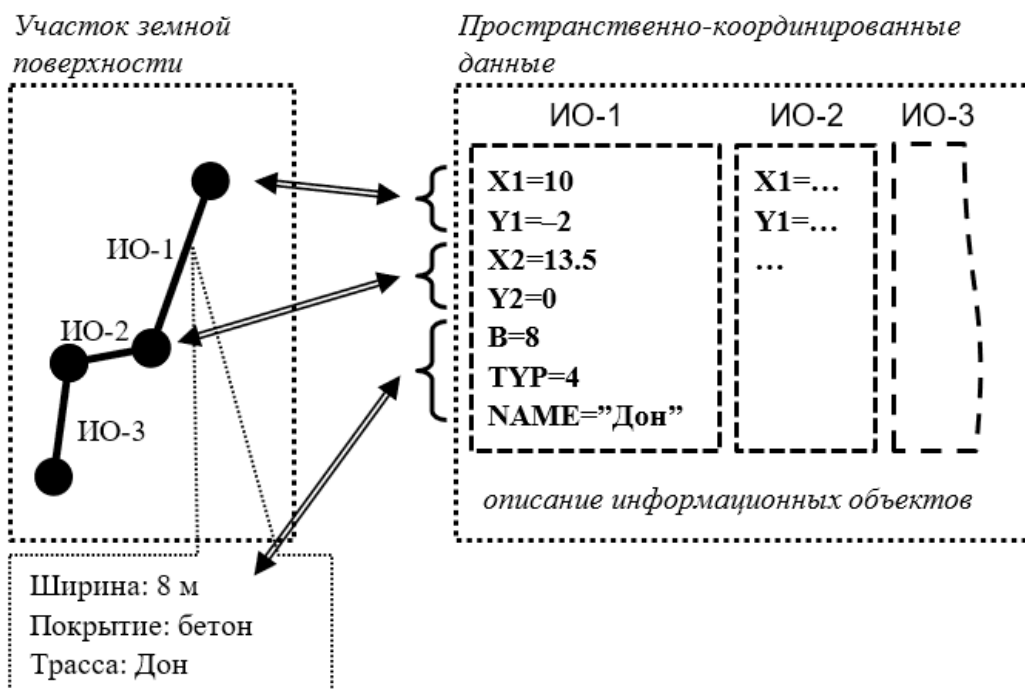


Рис. 1. Сущность пространственно-координированных данных.
(Авторский рисунок выполнен на основе работы [15]).

представления при необходимости может быть реализована сколь угодно высокой за счёт увеличения общего количества сегментов меньших размеров. Каждый сегмент описывается соответствующим ему информационным объектом (ИО-1, ИО-2 и т.д.). Для описания сегмента необходимо, во-первых, зафиксировать его положение на земной поверхности; для этого в рассматриваемом примере достаточно указать координаты его начальной и конечной точек (X_1, Y_1, X_2, Y_2). Во-вторых, каждый сегмент автодороги характеризуется определёнными эксплуатационными показателями (такими, как ширина проезжей части или тип покрытия), которые также отражаются при создании информационного объекта (с использованием атрибутов B, TYP и $NAME$).

Структура данных для создания информационных объектов разных типов будет, скорее всего, также неодинакова, но в любом случае данные будут оставаться пространственно-координированными. Все технологические процедуры при работе с пространственно-координированными данными (их ввод, редактирование, преоб-

разование, хранение, передача, отображение) имеют существенные особенности по сравнению с иными типами данных. Эти особенности обусловлены, прежде всего, необходимостью поддержания постоянной связи между позиционными и атрибутивными данными конкретных информационных объектов. Кроме того, наличие позиционной информации приводит к появлению принципиально новых возможностей, а значит, и технологических процедур для их реализации, включая отражение информационных объектов на экране (их визуализация), организацию поиска объектов в зависимости от их местоположения, в том числе относительно друг друга, и т.д. Другой часто используемой моделью пространственных объектов является квадротомическая (рис. 2). При её построении используется так называемое квадродерево, которое формируется начиная с самого высокого из возможного уровня агрегирования объектов. В примере на рис. 2 наивысший уровень агрегирования – третий; так могут быть объединены в единый элемент размером 3×3 девять объектов с атрибутом A .

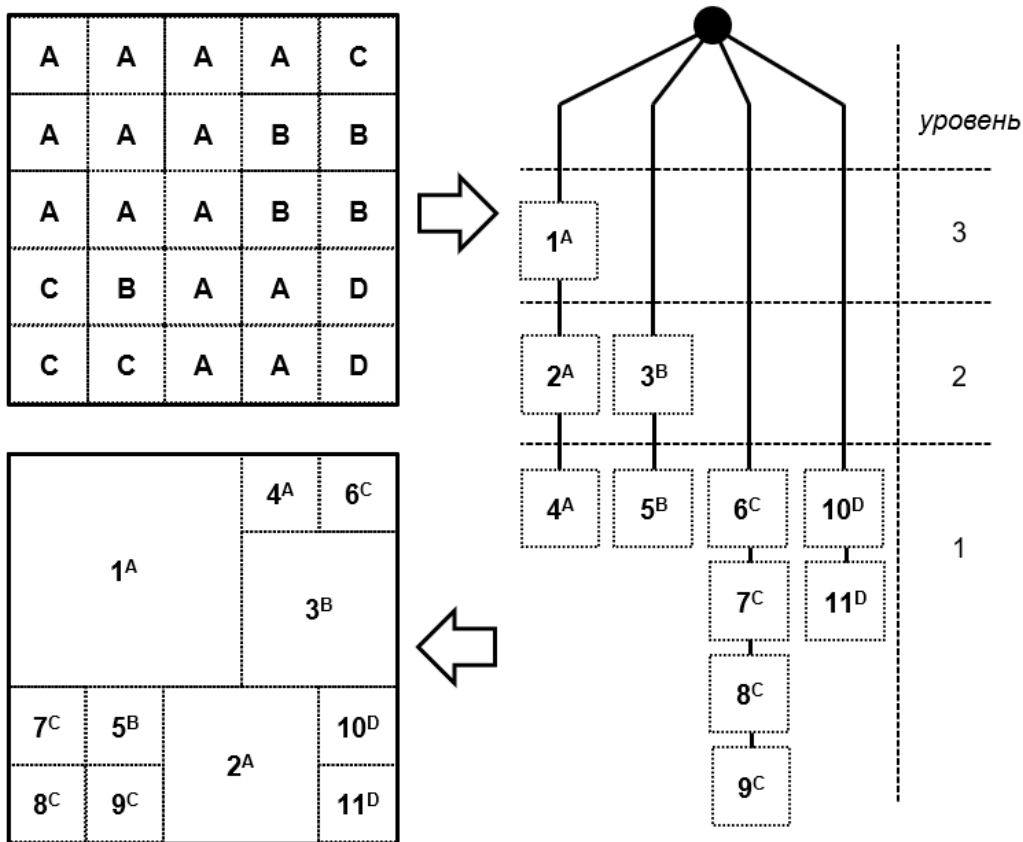


Рис. 2. Построение квадротомической модели данных.
(Авторский рисунок выполнен на основе работы [15]).

Укрупнённый объект маркируется как 1A (здесь совмещены номер объекта и значение атрибута), что отражается на квадродере и на итоговой схеме расположения объектов.

Больше возможностей агрегирования на уровне 3 нет, и производится переход к уровню 2. Здесь возможно создание двух укрупнённых объектов, маркируемых как 2A и 3B, после чего осуществляется переход к низшему уровню, на котором объекты, не имеющие возможности укрупнения, маркируются и отражаются на квадродере и на схеме расположения. В результате построения квадротомической модели на рис. 2 без потери атрибутивной информации удалось сократить общее количество объектов с 25 до 11. Уменьшенное количество объектов и обусловленная этим компактность хранения данных являются преимуществами квадротомической модели. Кроме того, квадродерево, созданное

на этапе построения модели, впоследствии может обеспечить максимально быстрый поиск объектов по заданному атрибуту. Модель отличается быстрым ростом разрешения при увеличении уровня квадродерева. Она также допускает переменный шаг при уменьшении размеров элементов по уровням.

Для описания трёхмерных объектов используется октотомическая модель, формирующая объекты в виде кубов переменного размера.

Векторная модель пространственных данных отличается максимальной гибкостью (поскольку в ней не присутствуют ограничения по форме объектов, их размерам и расположению), что одновременно приводит к необходимости полного указания всех позиционных данных для каждого объекта. Векторная модель может быть представлена в двух видах: нетопологическом и топологическом. Нетопологическая



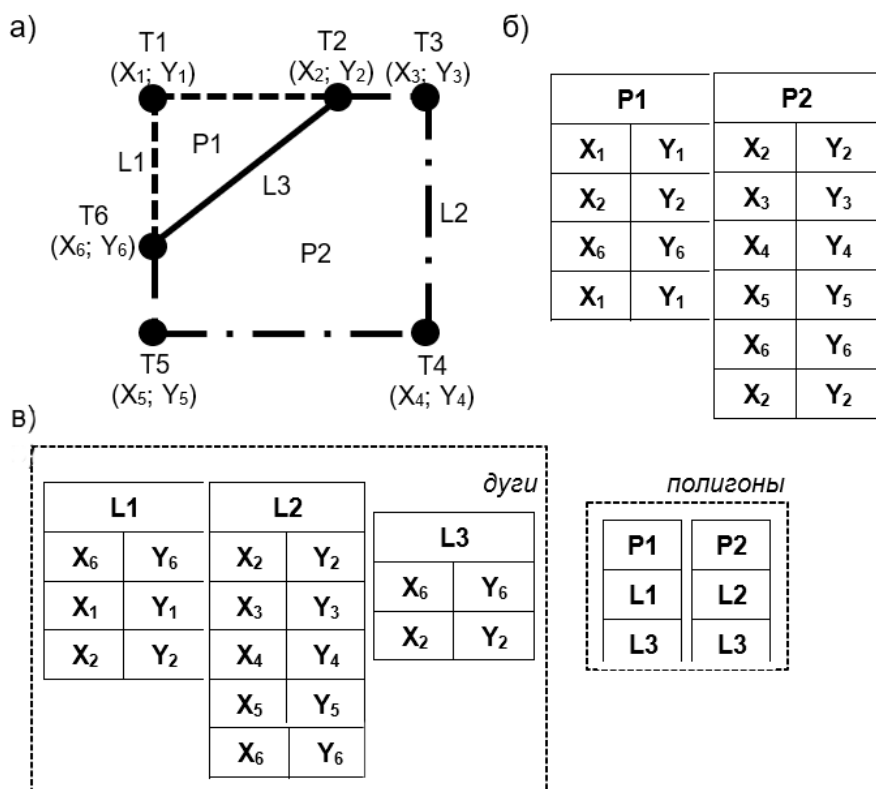


Рис. 3. Векторные модели данных: а) исходные полигоны; б) не топологическая модель; в) топологическая модель. (Авторский рисунок выполнен на основе работы [15]).

разновидность (модель «спагетти») подразумевает независимое указание границ для каждого объекта. Это приводит к тому, что данные о границах соседних объектов дублируются и, как следствие, общий объём используемых данных возрастает. Кроме того, при изменении границ какого-либо объекта необходимо выявить и синхронно изменить границы у всех соседних с ним объектов. Топологическая разновидность модели включает в себя набор отдельных полилиний (дуг), которые являются границами объектов, а также сведения о принадлежности этих дуг конкретным пространственным объектам. Дублирования информации при этом не происходит, и при редактировании границы смежных объектов изменения будут автоматически отражены для них обоих. Пример двух разновидностей векторной модели приведён на рис. 3, где показаны два полигона P1 и P2, имеющих смежную границу L3.

Нетопологическая модель содержит только указания о координатах граничных точек каждого полигона (три точки для P1

и пять точек для P2; при этом начальная точка второй раз указывается в конце списка, как знак того, что граничная полилиния замкнулась). В топологической модели сначала описаны три дуги L1, L2 и L3, после чего указано, из каких дуг состоит каждый из полигонов. Топологическая модель, по сравнению с нетопологической, отличается большей сложностью и по этой причине используется реже, в основном на этапе формирования пространственных объектов (ввода их границ). На этапе анализа и моделирования более высокое быстроедействие обеспечивает нетопологическая модель. Построение на базе элементарных пространственных элементов более сложных объектов производится в рамках определённой модели пространственных данных. Одной из простейших и, одновременно с этим, наиболее часто используемых является регулярно-ячеистая модель, пример которой показан на рис. 4.

При использовании регулярно-ячеистой модели территория разбивается на одинаковые участки (ячейки), каждый из

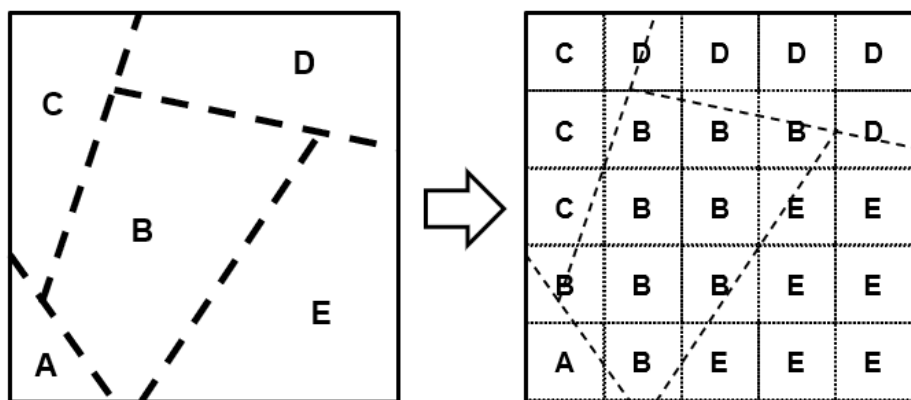


Рис. 4. Регулярно-ячеистая модель данных. (Авторский рисунок выполнен на основе работы [15]).

которых является самостоятельным пространственным объектом. Размер ячейки предопределяет пространственное разрешение модели. Атрибутивные данные для каждого объекта определяются независимо друг от друга. На рис. 4 можно видеть определение одного из пяти возможных значений атрибута. Удобной особенностью этого вида моделей является то, что пространственное положение ячейки определяется только её номером, что избавляет от необходимости указывать полные позиционные данные каждой ячейки (координаты всех угловых точек). Это даёт возможность компактного хранения информации. Например, на рис. 4 объекты могут быть описаны в виде последовательности из 25 значений атрибута (от А до Е каждое). При этом, при необходимости, координаты всех угловых точек каждого объекта могут быть легко восстановлены по его номеру. За счёт наличия идущих подряд ячеек с одинаковым значением атрибута становится возможным и ещё более компактное описание, например, C4DC3BDC2B2E3B2EAB3E (где вместо DDDD использовано обозначение 4D с экономией 2 знаков). В регулярно-ячеистых моделях возможно использование ячеек любой правильной формы, в том числе криволинейной. Также допускается применение разного разрешения и разной формы ячеек для разных атрибутов. В случае прямоугольных ячеек, единственным атрибутом которых является их цвет, модель превращается в растровую. Например, для построения оптимальных с точки зрения времени проезда

маршрутов движения по некоторой транспортной сети требуется прогнозирование скоростей транспортных средств на участках. Горизонт прогнозирования при этом должен быть не меньше типичного времени одной поездки, чтобы при построении маршрута перед выездом можно было учитывать скорости движения на участках транспортной сети, близких к цели поездки.

ЗАКЛЮЧЕНИЕ

Вместе с тем, кроме использования описанных выше алгоритмов, для целей транспортного моделирования необходимо решить целый ряд проблем, связанных со сбором и обработкой первичных данных.

Для достижения максимальной производительности оцифровки необходимо учитывать ошибки, связанные с интерпретацией результатов анализа данных, например, считывания регистрационных знаков ТС. Очевидно, что прогноз зависит от времени суток, дня недели и сезона поездки. К внешним факторам можно отнести погодные условия. Кроме этого, существуют нетривиальные закономерности между скоростями движения на различных рёбрах графа дорожной сети [16; 17]. Например, очень низкая скорость на определённом участке (пробка) может вызывать снижение скорости на одних рёбрах (распространение пробки) и одновременно повышение скорости на других рёбрах за счёт того, что резко уменьшается количество попадающих на них транспортных средств (экранирование). Количество постов для сбора



данных и исследования для создания транспортной модели назначается исходя из возможности одного учётчика или видеокамеры фиксировать максимальное возможное число событий по мониторингу транспортных и пешеходных потоков. Но существует ряд проблем в отношении организации работы учётчиков, например, связанных с фиксированием разнородной информации (одновременным фиксированием входящих и выходящих потоков ТС или пешеходов). В отношении применения видеосъёмки как средства фиксирования центров массового тяготения, можно выделить такие проблемы, как проведение съёмки в тёмное время суток, обработка видеоданных вручную (оцифровка данных), недоступность восприятия числа пассажиров в ИТ даже при лёгкой «тонировке» ТС. Перечисленные проблемы приводят к резкому снижению качества получаемых и исследуемых данных транспортного моделирования.

Такие проблемы могут быть решены не только за счёт совершенствования методов и технических средств получения первичных данных, но и за счёт разработки новых информационных технологий их обработки, в частности, с помощью нейронных сетей.

ЛИТЕРАТУРА

1. ГОСТ Р 56670-2015 Интеллектуальные транспортные системы. Подсистема мониторинга параметров транспортных потоков на основе анализа телематических данных городского пассажирского транспорта. [Электронный ресурс]: <http://docs.cntd.ru/document/1200125977>. Доступ 20.07.2020.
2. Ваксман С. А. Информационные технологии в управлении городским общественным пассажирским транспортом (задачи, опыт, проблемы) / Под ред. С. А. Ваксмана — Екатеринбург: Изд-во АМБ, 2012. — 250 с.
3. Куфтинова Н. Г. Моделирование динамики автотранспортных потоков с помощью кластерного анализа // Сб. научных трудов по итогам IV международной научно-практ. конференции «Транспортное планирование и моделирование». — СПб., 2019. — С. 106—108.
4. Куфтинова Н. Г. Математическое моделирование транспортных потоков на основе макро- и микроподходов городской транспортной системы // Сб. научных трудов по итогам III международной научно-практ. конференции «Транспортное планирование и моделирование. Цифровое будущее управления транспортом» / Под ред. д-ра техн. наук, профессора С. В. Жанказиева. — М.: МАДИ, 2018. — С. 67—76.
5. Куфтинова Н. Г. Интеллектуальная транспортная инфраструктура мегаполиса на основе геоанализа и геомоделирования автотранспортных систем // Логистический аудит транспорта и цепей поставок:

материалы международной научно-практической конференции. — Тюмень: ТИУ, 2018. — С. 76—82.

6. Куфтинова Н. Г. Общая характеристика транспортных моделей для оценки дорожной сети на городских территориях // Сб. научных трудов по итогам III международной научной конференции «Scientific Discoveries», Карловы Вары — Москва. — М., 2018. — С. 47—60. ISBN 978-5-00090-133-5.

7. Marcus, G., Davis, E. Eight (No, Nine!) Problems With Big Data. [Электронный ресурс]: https://www.nytimes.com/2014/04/07/opinion/eight-no-nine-problems-with-big-data.html?_r=0. Доступ 20.07.2020.

8. Chenyang Xu; Changqing Xu; Trieu-Kien Truong. Mining the spatio-temporal pattern using matrix factorisation: A case study of traffic flow. IET Intelligent Transport Systems, 2020, Vol. 14, Iss. 10, pp. 1328—1337. DOI: <http://dx.doi.org/10.1049/iet-its.2019.0705>. Доступ 23.10.2020.

9. Alam, O., Kush, A., Emami, A., Pouladzadeh, P. Predicting irregularities in arrival times for transit buses with recurrent neural networks using GPS coordinates and weather data. Journal of Ambient Intelligence and Humanized Computing, 2020. DOI: <https://doi.org/10.1007/s12652-020-02507-9>. Доступ 23.10.2020.

10. Guerreiro, G., Figueiras, P., Silva, R., Costa, R., Jardim-Goncalves, R. An architecture for big data processing on intelligent transportation systems. An application scenario on highway traffic flows. 2016 IEEE 8th International Conference on Intelligent Systems (IS), 2016, pp. 65—72. DOI: 10.1109/IS.2016.7737393. Доступ 20.07.2020.

11. Kohan, M., Ale, J. M. Discovering Traffic Congestion through Traffic Flow Patterns Generated by Moving Object Trajectories. Computers, Environment and Urban Systems, March 2020, Vol. 80, Article 101426. DOI: 10.1016/j.compenvurbsys.2019.101426. Доступ 20.07.2020.

12. Kumar, B. A., Vanajakshi, L., Subramanian, S. C. A Hybrid Model Based Method for Bus Travel Time Estimation. Journal of Intelligent Transportation Systems, 2018, Vol. 22, Iss. 5, pp. 390—406. DOI: 10.1080/15472450.2017.1378102. Доступ 20.07.2020.

13. Taamneh, M., Alkheder, S., Taamneh, S. Data-Mining Techniques for Traffic Accident Modeling and Prediction in the United Arab Emirates. Journal of Transportation Safety & Security, 2017, Vol. 9, Iss. 2, pp. 146—166. DOI: <https://doi.org/10.1080/19439962.2016.1152338>. Доступ 20.07.2020.

14. Norris, D. RapidMiner — a potential game changer. November 15, 2013. [Electronic resource]: <https://www.bloorresearch.com/2013/11/rapidminer-a-potential-game-changer/>. Доступ 17.07.2020.

15. Якубович А. Н., Куфтинова Н. Г., Рогова О. Б. Информационные технологии на автотранспорте: Учеб. пособие. Московский автомобильно-дорожный государственный технический университет (МАДИ). — М., 2017. — 252 с. [Электронный ресурс]: <http://www.lib.madi.ru/fel/fel1/fel17E429.pdf>. Доступ 17.07.2020.

16. Mousa, S. R., Mousa, R. M., Ishak, S., Radwand, L. Modeling Speed-Density Relation for Highways in Developing Countries with No Lane Discipline: A Case Study in Egypt. 2017 ASCE India Conference, 2018, pp. 725—735. DOI: 10.1061/9780784482025.074. Доступ 17.07.2020.

17. Kong, X., Das, S., Jha, K., Zhang, Y. Understanding Speeding Behavior From Naturalistic Driving Data: Applying Classification Based Association Rule Mining. Accident Analysis and Prevention, September 2020, Vol. 144. DOI: <https://doi.org/10.1016/j.aap.2020.105620>. Доступ 23.10.2020.