

УДК 519.254

ОБОБЩЕННОЕ ПРЕДСТАВЛЕНИЕ ПРОСТРАНСТВЕННО-ВРЕМЕННЫХ ТРАНЗАКЦИОННЫХ СЕТЕЙ

Богомолов Ю.И. (НИУ ИТМО)

Научный руководитель – к.т.н., дир. ИДУ Митягин С.А.
(НИУ ИТМО)

Консультант – д.ф.-м. н., Соболевский С.Л.
(Нью-Йоркский университет)

Доклад предлагает инновационное представление данных для пространственно-временных транзакционных сетей. Вторая часть работы посвящена описанию трансформаций, которые использует указанное представление и позволяют совместить временной, пространственный, потоковый и сетевой анализ внутри одной системы.

Введение. Многие процессы и события в современном мире могут быть представлены в виде транзакционных пространственно-временных сетей. Примеры включают общение по телефону, поездки на транспорте, операции покупки/продажи, взаимодействие с социальными сетями. Каждое из этих событий представляет собой транзакцию, которая совершается в пространстве и времени.

Развитие мобильных технологий и информационных систем привело к экспоненциальному росту объема пространственно-временных данных, что в свою очередь увеличило привлекательность исследований в этом направлении. Транзакции помогают описать и проанализировать социальные и экономические взаимодействия между людьми, либо между человеком и юридическим лицом. Дополнительные сложности анализа вызваны тем, что различные виды взаимодействия порождают многослойные сети.

Каждая из трех представленных подзадач: временной анализ, пространственный анализ и сетевой анализ хорошо изучена и описана в литературе. В то же время ни одна из известных нам систем не позволяет проводить анализ, который одновременно оперирует с сетевой структурой и пространственно-временными измерениями данных. Наша цель: разработать масштабируемое решение, которое позволяет автоматизировать стандартные методы обработки и анализа пространственно-временных сетей.

Основная часть. В основе нашей модели данных лежит ориентированный псевдограф и две функции меток. Первая функция возвращает метки для вершин псевдографа, а вторая определяет метки для дуг. Каждая метка в общем случае представляет собой вектор, который описывает атрибуты вершины либо дуги.

С каждой вершиной мы ассоциируем геопространственный объект, который не пересекается с геопространственными объектами других вершин. Векторы атрибутов для дуг в свою очередь содержат временное измерение. Заметим, что множество меток дуг представляет из себя временной ряд, в то время как множество меток вершин содержит пространственное разбиение.

Рассмотрим предложенную модель на конкретном примере данных: поездки на такси за определенный промежуток времени в исследуемом городе. Такие данные широко используются для изучения транспортной системы города и мобильности горожан.

В целях обеспечения анонимности данные для каждой из поездок на практике содержат не конкретные координаты, а блок начала и конца поездки (в качестве блока может выступать квартал или любое другое пространственное разбиение города, которое предоставляет достаточную точность). Каждый блок города в нашей модели удобно представить в виде вершины графа, а каждую поездку - в виде дуги между двумя блоками. Метка каждой дуги содержит вектор с атрибутами поездки (такими как время начала поездки, время конца поездки, стоимость и так далее). А метки вершин содержат геопространственные координаты

каждого блока и дополнительные атрибуты, необходимые для нашего анализа (например количество жителей в указанном блоке).

Предложенное представление данных объединяет в себе сетевую структуру данных и пространственно-временные измерения, что в свою очередь позволяет определить новые виды высокоуровневых операций. Работа с пространственно-временными сетями на практике часто требует переход от одной административно-территориальной единицы к другой, так как один набор данных может быть определен для каждого квартала, а другой только для целого района города.

Наша модель данных позволяет определить операцию трансформации для сети: `group_nodes(node_partition, label_agg_func)`, где `node_partition` - это разбиение вершин на непересекающиеся группы, а функция `label_agg_func` позволяет вычислить метку группы вершин на основании меток набора вершин, которые входят в группу. Для примера, описанного выше, `node_partition` определяет разбиение районов города на кварталы. А функция `label_agg_func` помогает вычислить метку района города на основании меток кварталов: если в векторе хранятся только данные о населении квартала - то население района будет можно посчитать как сумму показателей по каждому из кварталов района.

Заметим, что данная трансформация возвращает новую пространственно-временную сеть, в которой в качестве вершин выступают уже районы города. А дуги исходной сети между кварталами одного района становятся циклами. Геопространственные координаты районов также могут быть автоматически вычислены на основании объединения кварталов.

Описанная операция представляет очень простой интерфейс, который позволяет одновременно модифицировать структуру сети и ее пространственно-временные характеристики. По аналогии мы определяем операции трансформации сети для агрегации входящих/исходящих дуг и операции фильтрации. Прототип реализации указанной модели можно найти по адресу: <https://github.com/yuribogomolov/sttn>.

Выводы. Предложенная модель представления пространственно-временных сетей предлагает унифицированное представление и интерфейс для обработки, фильтрации и агрегации данных. Модель хорошо показала себя при применении к транспортным и социальным сетям, а также при анализе мобильности горожан на основе данных переписи населения. Короткий список предложенных трансформаций позволил автоматизировать большинство шагов обработки пространственно-временных данных, которые мы встретили на практике.

Богомоллов Ю.И. (автор)

Подпись

Митягин С.А. (научный руководитель)

Подпись