

Прогнозирование транспортного потока с использованием графовых нейронных сетей

А.А. Агафонов¹, А.С. Юмаганов¹

¹Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086

Аннотация. Прогнозирование транспортного потока остается одной из самых важных проблем в сфере транспортной инженерии. Сложность проблемы обусловлена тем фактом, что транспортные потоки имеют сложные нелинейные пространственно-временные зависимости. Развитие современных математических моделей и, в частности, аппарата нейронных сетей, позволяет применять сверточные нейронные сети для решения транспортных проблем. В статье исследуется архитектура графовой сверточной нейронной сети для решения задачи прогнозирования параметров транспортных потоков. Рассматриваемая нейронная сеть учитывает дневные и недельные паттерны распространения транспортных потоков. Экспериментальные исследования на транспортной сети города Самары показывают преимущества графовой нейронной сети по сравнению с другими рассмотренными алгоритмами прогнозирования.

1. Введение

Разработка интеллектуальных транспортных систем (ИТС) остается одной из наиболее актуальных задач в сфере транспортного планирования как с исследовательской, так и с практической точки зрения [1]. ИТС могут применяться как для управления транспортом (например, посредством регулирования светофорных тактов, диспетчеризации общественного транспорта и т.д.), так и для предоставления различных сервисов участникам дорожного движения (выдача прогноза времени прибытия общественного транспорта, нахождение кратчайших маршрутов движения и т.д. [2]). В свою очередь, точная и своевременная оценка и прогнозирование параметров транспортных потоков является ключевым элементом функционирования ИТС.

В статье [3] проведен анализ методов прогнозирования и их классификация по различным критериям: используемые технологии сбора данных, прогнозируемые характеристики транспортных потоков, используемые модели прогнозирования и т.д. В [4] представлен обзор состояния исследований в области краткосрочного прогнозирования параметров транспортных потоков, описаны последние технические достижения и нерешенные проблемы.

Методы прогнозирования условно можно классифицировать по двум категориям: статистические методы и методы машинного обучения. Большинство работ, использующих статистические методы прогнозирования, предполагают, что транспортные системы достаточно просты, либо работают с относительно небольшими объемами данных. Статистические методы включают в себя модели временных рядов [5, 6], векторные модели авторегре-

сии [7] и т.д. Однако возможности статистических моделей обрабатывать пространственно-временные данные высокой размерности достаточно ограничены.

В настоящее время, благодаря быстрому развитию вычислительных мощностей, а также росту объема данных о состоянии транспортных потоков, большинство работ фокусируется на использовании методов машинного обучения для решения задачи прогнозирования транспортных потоков. Во многих исследованиях было показано, что методы машинного обучения превосходят статистические методы благодаря своей способности моделировать сложные нелинейные отношения. В [8, 9] для решения задачи краткосрочного прогнозирования транспортных потоков использовался метод опорных векторов. Метод k ближайших соседей (kNN) с использованием технологии обработки больших данных изучался в [10, 11, 12]. Однако при увеличении объема выборки эффективность kNN падает из-за больших вычислительных затрат. Полносвязные нейронные сети применялись в работах [13, 14].

В последнее десятилетие в различных областях широкое распространение получили глубокие нейронные сети, включая сети с долгой краткосрочной памятью (LSTM) [15, 16], сверточные нейронные сети [17]. Традиционные сверточные нейронные сети способны эффективно извлекать локальные паттерны, но могут применяться только для структур данных, представляющих собой регулярные сетки. Свертка на графах обобщает традиционную свертку и может использоваться для обработки данных на графовых структурах. Графовые нейронные сети используют матрицу смежности графов или матрицу Кирхгофа для описания структуры графа [18]. Обобщенные графовые нейронные сети, основанные на использовании графового оператора Лапласа, были предложены в [19]. Их оптимизация с использованием полиномов Чебышева была исследована в [20]. Для задач классификации графовые сети использовались в работе [21]. В [22] графовые нейронные сети использовались для прогноза транспортных потоков, однако в работе не учитывались пространственно-временные корреляции в данных. В работах [23, 24] сети были расширены с использованием механизма внимания и узлов с долгой краткосрочной памятью. Однако стоит отметить, что в рассмотренных работах в качестве источников данных использовались данные с малого числа детекторов. В данной работе рассматривается дорожная сеть города Самары, в качестве данных используется среднее время прохождения дорожных сегментов.

В работе исследуются возможности применения графовых нейронных сетей для краткосрочного прогнозирования параметров транспортных потоков. Исследуется структура сети, использующая пространственно-временные зависимости в данных. Работа организована следующим образом. Во втором разделе приводятся основные обозначения и постановка задачи. Описание структуры нейронной сети представлено в третьем разделе. В четвертом разделе представлены постановка и результаты экспериментальных исследований. В завершение работы представлены заключение и возможные направления дальнейших исследований.

2. Формулировка проблемы

Улично-дорожную сеть будем рассматривать в виде ориентированного графа $G = (V, E, \mathbf{A})$, где V - множество вершин графа, $|V| = N$ - количество вершин, E - множество ребер, $\mathbf{A} \in \mathbb{R}^{N \times N}$ - матрица смежности графа. Вершины графа представляют собой сегменты дорожной сети.

Будем считать, что с каждой вершиной графа связан вектор признаков $\mathbf{x}_t^i \in \mathbb{R}^F$, описывающий транспортный поток в вершине $i \in V$ в заданный момент времени t (временной интервал). $x_t^{j,i} \in \mathbb{R}$ определяет значение j -го элемента вектора признаков вершины $i \in V$ в момент времени t . В качестве элементов вектора признаков могут использоваться следующие параметры транспортного потока:

- средняя скорость транспортного потока;
- плотность потока;
- поток (собственно величина потока).

В настоящей работе в качестве элемента вектора признаков и прогнозируемого параметра транспортного потока в экспериментальных исследованиях используется средняя скорость движения.

Обозначим набор векторов признаков для всех вершин графа в момент времени t как

$$\mathbf{X}_t = (\mathbf{x}_t^1, \mathbf{x}_t^2, \dots, \mathbf{x}_t^N) \in \mathbb{R}^{N \times F}. \quad (1)$$

Набор векторов признаков для всех вершин графа за несколько временных интервалов τ обозначим как

$$\chi = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_\tau)^T \in \mathbb{R}^{N \times F \times \tau}. \quad (2)$$

Обозначим за $y_t^i \in \mathbb{R}$ искомое (прогнозируемое) значение вектора признаков в будущий момент времени t для вершины $i \in V$, $\mathbf{y}^i = (y_{\tau+1}^i, y_{\tau+2}^i, \dots, y_{\tau+T}^i) \in \mathbb{R}^T$ - искомые значения вектора признаков для вершины i для горизонта прогноза в T интервалов, $\mathbf{Y} = (\mathbf{y}^1, \mathbf{y}^2, \dots, \mathbf{y}^N)^T \in \mathbb{R}^{N \times T}$ - искомые значения вектора признаков для всех вершин графа.

Учитывая введённые обозначения, формальная постановка задачи краткосрочного прогнозирования транспортного потока может быть сделана следующим образом:

Имея заданный граф $G = (V, E, A)$ и последовательность χ наблюдаемых значений параметров транспортных потоков, рассчитать оценку параметров \mathbf{Y} для T временных интервалов.

3. Графовая нейронная сеть

3.1. Вектор признаков

Пусть данные для каждого сегмента поступают с частотой q записей в день. Обозначим текущий момент времени как t_0 .

Для учета текущего, дневного и недельного паттернов распространения транспортных потоков для каждого дорожного сегмента будем рассматривать три временных ряда:

1. Текущий временной ряд.

Для учета текущих характеристик транспортного потока будем рассматривать временной ряд, составленный из векторов признаков за предшествующие T_c временных интервалов:

$$\chi_c = (\mathbf{X}_{t_0-T_h+1}, \mathbf{X}_{t_0-T_h+2}, \dots, \mathbf{X}_{t_0}). \quad (3)$$

2. Временной ряд, учитывающий дневные паттерны.

Транспортные потоки имеют дневную периодичность (например, утренние и вечерние часы пик). Для учета дневной периодичности будем рассматривать временной ряд следующего вида:

$$\chi_d = (\mathbf{X}_{t_0-T_d*q}, \mathbf{X}_{t_0-(T_d-1)*q}, \dots, \mathbf{X}_{t_0-q}), \quad (4)$$

где T_d - количество рассматриваемых дней.

3. Временной ряд, учитывающий недельные паттерны.

Часто распределение транспортного потока различается по дням недели (особенно в будние / выходные дни). Для учета недельной периодичности в данных будем рассматривать временной ряд следующего вида:

$$\chi_w = (\mathbf{X}_{t_0-7*T_w*q}, \mathbf{X}_{t_0-7*(T_w-1)*q}, \dots, \mathbf{X}_{t_0-7*q}), \quad (5)$$

где T_w - количество рассматриваемых недель.

3.2. Нормализация данных

Входной вектор признаков и результирующая оценка должны быть нормализованы, чтобы ускорить обучение нейронной сети, улучшить сходимость и снизить вероятность нахождения локального оптимума. Данные нормализуются следующим образом:

$$\hat{x} = \frac{(x - x_{min})}{(x_{max} - x_{min})}, \quad (6)$$

где x - исходные данные, \hat{x} - нормализованные данные в диапазоне $[0, 1]$.

3.3. Свертка на графе в пространственной области

Спектральная теория графов обобщает операцию свертки на графовые данные. Использование свертки на графах позволит учесть топологические свойства дорожной сети для решения задачи прогнозирования транспортного потока.

Для определения свертки на графовых данных будем описывать граф дорожной сети с помощью матрицы Кирхгофа, которая представляет собой дискретный оператор Лапласа для графа. Матрица Кирхгофа представима в виде:

$$\mathbf{L} = \mathbf{D} - \mathbf{A}, \quad (7)$$

или в нормализованном виде

$$\mathbf{L} = \mathbf{I} - \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{\frac{1}{2}}, \quad (8)$$

где \mathbf{A} - матрица смежности данного графа, \mathbf{I} - единичная матрица, \mathbf{D} - матрица, на главной диагонали которой степени вершин графа, а остальные элементы — нули:

$$D_{ii} = \sum_j A_{ij}. \quad (9)$$

Спектральное разложение матрицы Кирхгофа имеет вид:

$$\mathbf{L} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T, \quad (10)$$

где $\mathbf{\Lambda}$ - диагональная матрица, \mathbf{U} - Фурье-базис.

Операция свертки на графе определяется как результат перемножения сигнала $x \in \mathbb{R}^N$ на графе с ядром g_θ [25]:

$$g_\theta * x = g_\theta(\mathbf{L}) x = g_\theta(\mathbf{U} \mathbf{\Lambda} \mathbf{U}^T) x = \mathbf{U} g_\theta(\mathbf{\Lambda}) \mathbf{U}^T x. \quad (11)$$

В качестве сигнала $x \in \mathbb{R}^N$ может использоваться i -е значение вектора признаков \mathbf{x}_t для всех вершин графа в заданный момент времени t .

Однако, операция спектрального разложения матрицы Кирхгофа является вычислительно сложной. Для решения этой проблемы в работе [25] было предложено аппроксимировать спектральное разложение матрицы с помощью полиномов Чебышева:

$$g_\theta * x = g_\theta(\mathbf{L}) x = \sum_{k=0}^{K-1} \theta_k T_k(\tilde{\mathbf{L}}) x, \quad (12)$$

где параметр θ - вектор полиномиальных коэффициентов, $\tilde{\mathbf{L}} = \frac{2}{\lambda_{max}} \mathbf{L} - \mathbf{I}$, λ_{max} - максимальный собственный вектор матрицы Кирхгофа. Полиномы Чебышева рекурсивно определяются следующим образом:

$$\begin{aligned} T_k(x) &= 2xT_{k-1}(x) - T_{k-2}(x), \\ T_0(x) &= 1, T_1(x) = x. \end{aligned} \quad (13)$$

В качестве функции активации в сверточном слое нейронной сети используется функция ReLU.

3.4. Свертка во временной области

Следующим оператором графовой нейронной сети после выполнения свертки в пространственной области является стандартная свертка во временной области. Тогда вычисления на r -м слое нейронной сети могут быть записаны в виде

$$\chi^r = ReLU(\Phi * (ReLU(g_\theta * \chi^{r-1}))), \quad (14)$$

где Φ - параметры ядра свертки, χ - исходный временной ряд / результат на выходе из предыдущего слоя нейронной сети.

3.5. Структура нейронной сети

Структура используемой нейронной сети показана на рисунке 1. Сеть состоит из трех независимых компонент с одинаковой структурой, которые учитывают текущие данные, дневные и недельные паттерны в характеристиках транспортного потока соответственно.

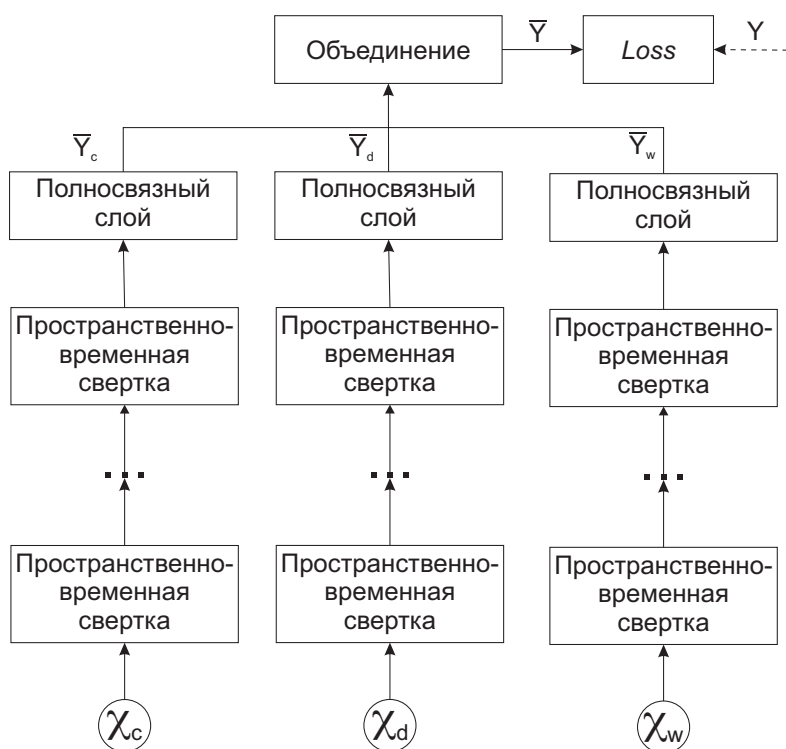


Рисунок 1. Структура нейронной сети.

Таким образом, каждая нейронная сеть состоит из нескольких модулей, выполняющих пространственно-временную свертку на графовых данных, способных учитывать пространственно-временные корреляции в характеристиках транспортного потока. Несколько слоев выполняются последовательно для выделения большего диапазона пространственно-временных корреляций. Выходной полносвязный слой необходим для получения требуемой размерности выходных данных нейронной сети.

Для объединения результатов прогнозирования по каждому из трех компонент (текущий, дневной и недельный временные ряды) используется следующая операция:

$$\hat{Y} = W_c \odot \hat{Y}_c + W_d \odot \hat{Y}_d + W_w \odot \hat{Y}_w, \quad (15)$$

где \odot - произведение Адамара, W_c, W_d, W_w - настраиваемые параметры.

4. Экспериментальные исследования

4.1. Набор данных

Экспериментальные исследования модели проводились для транспортной сети г. Самары. Для тестирования была выбрана область, включающая в себя 1760 дорожных сегментов. В качестве исходных данных для проведения экспериментальных исследований использовались значения средней скорости движения транспортного потока (в км/ч), полученные за 50 дней, начиная с 1 ноября 2019 г. Средняя скорость движения агрегирована за 10-минутный интервал времени. Пропуски в данных были заполнены с помощью линейной интерполяции. Используемый набор данных был разделен на три части: обучающая выборка (60% выборки), контрольная выборка (20% выборки) и валидационная выборка (20% выборки).

4.2. Настройки модели

Разработанная модель была реализована с использованием фреймворка MXNet. Каждый сверточный слой содержал 64 нейрона. В качестве горизонта прогноза было выбрано $T = 6$ временных интервалах, т.е. прогнозирование транспортного потока проводилось на 1 час.

Параметры модели настраивались с использованием обучающей выборки, лучшая модель была выбрана по наименьшему значению критерия среднеквадратичного отклонения на валидационной выборке. Сравнение моделей проводилось на контрольной выборке.

4.3. Базовые методы

Сравнение графовой нейронной сети было проведено с двумя базовыми моделями:

- Линейная регрессия;
- Полносвязная нейронная сеть с двумя скрытыми слоями, 24 нейрона в каждом слое (MLP).

Кроме того, было проведено сравнение моделей графовых нейронных сетей (GCNN) для различных значений степеней полиномов Чебышева $K \in \{1, 2, 3\}$.

Для оценки качества прогноза были использованы две стандартные метрики: средняя абсолютная ошибка (MAE) и средняя абсолютная ошибка в процентах (MAPE).

4.4. Результаты

В таблице 4.4 представлено сравнение результатов прогнозирования выбранными моделями при горизонте прогноза в 1 час.

Таблица 1. Среднее время задержки.

	MAE, км/ч	MAPE, %
Линейная регрессия	2.16	10.46
MLP	1.89	9.04
GCNN, K=1	1.7295	8.352
GCNN, K=2	1.7315	8.336
GCNN, K=3	1.821	8.773

Анализ результатов показывает, что представленная в работе графовая нейронная сеть дает более качественный результат прогноза, чем базовые модели прогнозирования. Лучший результат по совокупному критерию был показан моделью, использующей полиномы Чебышева второго порядка ($K = 2$).

Далее было проведено сравнение моделей для различных значений горизонта прогноза. Результаты сравнения по критериям средней абсолютной ошибки (MAE) и средней абсолютной ошибки в процентах (MAPE) показаны на рисунках 2 и 3 соответственно.

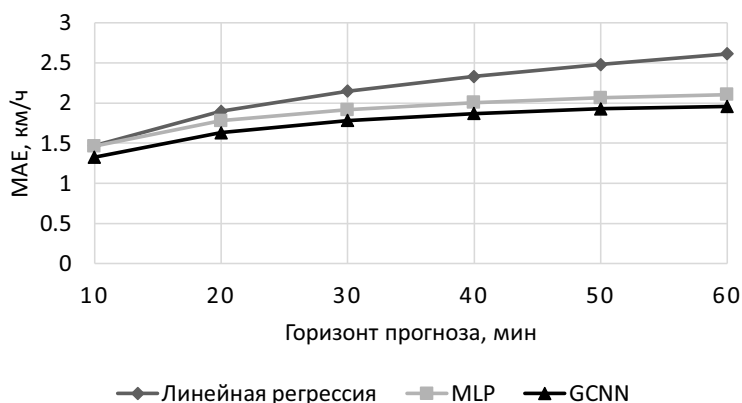


Рисунок 2. Средняя абсолютная ошибка.

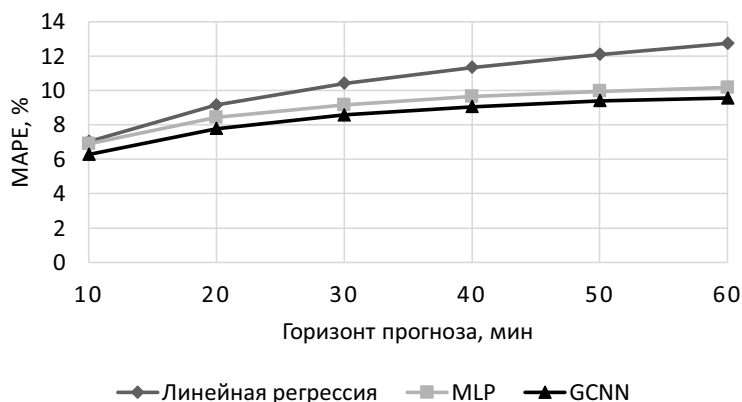


Рисунок 3. Средняя абсолютная ошибка в процентах.

Анализируя полученные результаты, можно сделать вывод, что представленная модель показывает лучшие результаты для всего горизонта прогноза, при возрастании горизонта прогноза точность прогнозирования ухудшается.

На заключительном этапе была проведена оценка времени работы графовой нейронной сети в сравнении с базовыми моделями. Оценивалось время прогнозирования параметров транспортных потоков для всех сегментов сети на 1 час вперед. Вычисления проводились с использованием GPU nVidia GeForce GTX2080 Ti. Результаты представлены в таблице 4.4.

Таблица 2. Время работы моделей.

	MLP	GCNN, K=1	GCNN, K=2	GCNN, K=3
Время, мс	6.1	12.6	16.7	20.2

Все рассмотренные модель имеют очень малое время вычислений и могут быть использованы для краткосрочного прогнозирования параметров транспортного потока в крупномасштабной транспортной сети.

5. Заключение

В работе было проведено исследование модели графовой сверточной нейронной сети для решения задачи краткосрочного прогнозирования параметров транспортного потока. Представленная модель учитывает пространственно-временные характеристики транспортного потока и топологию дорожной сети.

Экспериментальные исследования, проведенные на данных о средней скорости движения транспорта в г. Самаре, показали, что представленная модель обладает высокой точностью прогнозирования (средняя ошибка 8.3%) и очень низким временем работы (порядка 16-20 миллисекунд), достаточным для прогнозирования в режиме реального времени.

Дальнейшие исследования могут быть направлены на усложнение структуры модели и учет дополнительных факторов, влияющих на распространение транспортного потока, например, погодных условий или общественных событий.

6. Благодарности

Работа выполнена при частичной финансовой поддержке грантов РФФИ № 18-29-03135-мк, № 18-07-00605 А.

7. Литература

- [1] Zhang, J. Data-driven intelligent transportation systems: A survey // *IEEE Transactions on Intelligent Transportation Systems*. – 2011. – Vol. 12(4). – P. 1624-1639.
- [2] Агафонов, А.А. Исследование численного метода резервирования маршрутов в геоинформационной задаче маршрутизации автономных транспортных средств / А.А. Агафонов, В.В. Мясников // *Компьютерная оптика*. – 2018. – Т. 42, № 5. – P. 912-920. DOI: 10.18287/2412-6179-2018-42-5-912-920.
- [3] Vlahogianni, E.I. Short-term traffic forecasting: Where we are and where we're going / E.I. Vlahogianni, M.G. Karlaftis, J.C. Golias // *Transportation Research Part C: Emerging Technologies*. – 2014. – Vol. 43. – P. 3-19.
- [4] Lana, I. Road Traffic Forecasting: Recent Advances and New Challenges // *IEEE Intelligent Transportation Systems Magazine*. – 2018. – Vol. 10(2). – P. 93-109.
- [5] Williams, B.M. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results / B.M. Williams, L.A. Hoel // *Journal of Transportation Engineering*. – 2003. – Vol. 129(6). – P. 664-672.
- [6] Shekhar, S. Adaptive seasonal time series models for forecasting short-term traffic flow / S. Shekhar, B.M. Williams // *Transportation Research Record*. – 2007. – Vol. 2024. – P. 116-125.
- [7] Chandra, S.R. Predictions of freeway traffic speeds and volumes using vector autoregressive models / S.R. Chandra, H. Al-Deek // *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*. – 2009. – Vol. 13(2). – P. 53-72.
- [8] Su, H. Short-term traffic flow prediction based on incremental support vector regression / H. Su, L. Zhang, S. Yu // *Computer Science*. – 2007. – Vol. 1. – P. 640-645.
- [9] Jeong, Y.-S. Supervised weighting-online learning algorithm for short-term traffic flow prediction / Y.-S. Jeong // *IEEE Transactions on Intelligent Transportation Systems*. – 2013. – Vol. 14(4). – P. 1700-1707.
- [10] Xia, D. A distributed spatial-temporal weighted model on MapReduce for short-term traffic flow forecasting // *Neurocomputing*. – 2016. – Vol. 179. – P. 246-263.
- [11] Агафонов, А.А. Анализ больших данных в геоинформационной задаче краткосрочного прогнозирования параметров транспортного потока на базе метода k ближайших соседей / А.А. Агафонов, А.С. Юмаганов, В.В. Мясников // *Компьютерная оптика*. – 2018. – Vol. 42(6). – P. 1101-1111. DOI: 10.18287/2412-6179-2018-42-6-1101-1111.
- [12] Agafonov, A. Short-term traffic flow forecasting using a distributed spatial-temporal k nearest neighbors model / A. Agafonov, A. Yumaganov // *IEEE International Conference Computational Science and Engineering*, 2018. – P. 91-98.

- [13] Moretti, F. Urban traffic flow forecasting through statistical and neural network bagging ensemble hybrid modeling // *Neurocomputing*. – 2015. – Vol. 167. – P. 3-7.
- [14] Zheng, W. Short-term freeway traffic flow prediction: Bayesian combined neural network approach / W. Zheng, D.-H. Lee, Q. Shi // *Journal of Transportation Engineering*. – 2006. – Vol. 132. – Vol. 2. – P. 114-121.
- [15] Yu, R. Deep learning: A generic approach for extreme condition traffic forecasting, 2017. – P. 777-785.
- [16] Agafonov, A. Bus Arrival Time Prediction with LSTM Neural Network / A. Agafonov, A. Yumaganov // *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. – 2019. – Vol. 11554 LNCS. – P. 11-18.
- [17] Визильтер, Ю.В. Структурно-функциональный анализ и синтез глубоких конволюционных нейронных сетей / Ю.В. Визильтер, В.С. Горбацевич, С.Ю. Желтов // *Компьютерная оптика*. – 2019. – Т. 43, № 5. – P. 886-900. DOI: 10.18287/2412-6179-2019-43-5-886-900.
- [18] Zhang, S. Graph convolutional networks: a comprehensive review // *Computational Social Networks*. – 2019. – Vol. 6(1).
- [19] Bruna, J. Spectral networks and deep locally connected networks on graphs, 2014.
- [20] Defferrard, M. Convolutional neural networks on graphs with fast localized spectral filtering / M. Defferrard, X. Bresson, P. Vandergheynst, 2016. – P. 3844-3852.
- [21] Kipf, T.N. Semi-supervised classification with graph convolutional networks / T.N. Kipf, M. Welling, 2019.
- [22] Yu, B. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting / B. Yu, H. Yin, Z. Zhu. – 2018. – Vol. 2018. – P. 3634-3640.
- [23] Guo, S. Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting // *Proceedings of the AAAI Conference on Artificial Intelligence*. – 2019. – Vol. 33. – P. 922-929.
- [24] Cui, Z. Traffic Graph Convolutional Recurrent Neural Network: A Deep Learning Framework for Network- Scale Traffic Learning and Forecasting // *IEEE Transactions on Intelligent Transportation Systems*, 2019. – P. 1-12.
- [25] Hammond, D.K. Wavelets on graphs via spectral graph theory / D.K. Hammond, P. Vandergheynst, R. Gribonval // *Applied and Computational Harmonic Analysis*. – 2011. – Vol. 30(2). – P. 129-150.

Traffic flow prediction using graph neural networks

A.A. Agafonov¹, A.S. Yumaganov¹

¹Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

Abstract. Traffic flow prediction remains one of the most important issues in transport engineering. The complexity of the problem is explained by the fact that traffic flows have complex nonlinear space-time relationships. The development of modern mathematical models, and, in particular, the deep neural networks, allows using convolutional neural networks to solve transportation problems. The article investigates the architecture of a graph convolutional neural network to solve the short-term traffic flow prediction problem. The proposed neural network takes into account the daily and weekly patterns of traffic flows distribution. Experimental studies on the transport network of the Samara city demonstrate the advantages of a graph neural network in comparison with the other considered prediction algorithms.