

УДК: 656.021.2

Использование сверточных нейронных сетей для прогнозирования скоростей транспортного потока на дорожном графе

Н. Г. Прокопцев^{1,a}, А. Е. Алексеенко^{1,2,b}, Я. А. Холодов^{1,2,c}

¹ Университет Иннополис,
Россия, 420500, г. Иннополис, ул. Университетская, д. 1

² Институт автоматизации проектирования РАН,
Россия, 123056, г. Москва, ул. 2-я Брестская, д. 19/18

E-mail: ^a n.prokoptsev@innopolis.ru, ^b a.alekseenko@innopolis.ru, ^c y.kholodov@innopolis.ru

*Получено 28.02.2018, после доработки — 22.05.2018.
Принято к публикации 24.05.2018.*

Краткосрочное прогнозирование потока трафика является одной из основных задач моделирования транспортных систем, основное назначение которой — контроль дорожного движения, сообщение об авариях, избежание дорожных пробок за счет знания потока трафика и последующего планирования транспортировки. Существует два типа подходов для решения этой задачи: математическое моделирование трафика и модель с использованием количественных данных трафика. Тем не менее большинство пространственно-временных моделей страдают от высокой математической сложности и низкой эффективности. Искусственные нейронные сети, один из видных подходов второго типа, показывают обещающие результаты в моделировании динамики транспортной сети. В данной работе представлена архитектура нейронной сети, используемой для прогнозирования скоростей транспортного потока на графе дорожной сети. Модель основана на объединении рекуррентной нейронной сети и сверточной нейронной сети на графе, где рекуррентная нейронная сеть используется для моделирования временных зависимостей, а сверточная нейронная сеть — для извлечения пространственных свойств из трафика. Для получения предсказаний на несколько шагов вперед используется архитектура encoder-decoder, позволяющая уменьшить накопление шума из-за неточных предсказаний. Для моделирования сложных зависимостей мы используем модель, состоящую из нескольких слоев. Нейронные сети с глубокой архитектурой сложны для тренировки; для ускорения процесса тренировки мы используем skip-соединения между каждым слоем, так что каждый слой учит только остаточную функцию по отношению к предыдущему слою. Полученная объединенная нейронная сеть тренировалась на необработанных данных с сенсоров транспортного потока из сети шоссе в США с разрешением в 5 минут. 3 метрики — средняя абсолютная ошибка, средняя относительная ошибка, среднеквадратическая ошибка — использовались для оценки качества предсказания. Было установлено, что по всем метрикам предложенная модель имеет более низкую погрешность предсказания по сравнению с ранее опубликованными моделями, такими как Vector Auto Regression, Long Short-Term Memory и Graph Convolution GRU.

Ключевые слова: нейронная сеть, граф транспортной сети, транспортный поток, предсказание трафика

Работа выполнена при поддержке гранта РФФИ 14-11-00877.

UDC: 656.021.2

Traffic flow speed prediction on transportation graph with convolutional neural networks

N. G. Prokoptsev^{1,a}, A. E. Alekseenko^{1,2,b}, Ya. A. Kholodov^{1,2,c}

¹Innopolis University,

1 Universitetskaya st., Innopolis, 420500, Russia

²Institute of Computer Aided Design of the Russian Academy of Sciences,
19/18 2-ya Brestskaya st., Moscow, 123056, Russia

E-mail: ^a n.prokoptsev@innopolis.ru, ^b a.alekseenko@innopolis.ru, ^c y.kholodov@innopolis.ru

Received 28.02.2018, after completion — 22.05.2018.

Accepted for publication 24.05.2018.

The short-term prediction of road traffic condition is one of the main tasks of transportation modelling. The main purpose of which are traffic control, reporting of accidents, avoiding traffic jams due to knowledge of traffic flow and subsequent transportation planning. A number of solutions exist — both model-driven and data driven had proven to be successful in capturing the dynamics of traffic flow. Nevertheless, most space-time models suffer from high mathematical complexity and low efficiency. Artificial Neural Networks, one of the prominent data-driven approaches, show promising performance in modelling the complexity of traffic flow. We present a neural network architecture for traffic flow prediction on a real-world road network graph. The model is based on the combination of a recurrent neural network and graph convolutional neural network. Where a recurrent neural network is used to model temporal dependencies, and a convolutional neural network is responsible for extracting spatial features from traffic. To make multiple few steps ahead predictions, the encoder-decoder architecture is used, which allows to reduce noise propagation due to inexact predictions. To model the complexity of traffic flow, we employ multilayered architecture. Deeper neural networks are more difficult to train. To speed up the training process, we use skip-connections between each layer, so that each layer teaches only the residual function with respect to the previous layer outputs. The resulting neural network was trained on raw data from traffic flow detectors from the US highway system with a resolution of 5 minutes. 3 metrics: mean absolute error, mean relative error, mean-square error were used to estimate the quality of the prediction. It was found that for all metrics the proposed model achieved lower prediction error than previously published models, such as Vector Auto Regression, LSTM and Graph Convolution GRU.

Keywords: neural network, traffic network flow, traffic flow, traffic prediction

Citation: *Computer Research and Modeling*, 2018, vol. 10, no. 3, pp. 359–367 (Russian).

This research has been supported by Russian Science Foundation (grant ID 14-11-00877).

Введение

Предсказание скорости и интенсивности транспортного потока является одной из ключевых задач в области интеллектуальных транспортных систем (ИТС). Надежная и точная система предсказания загруженности транспортной сети необходима для надежного функционирования интеллектуальной транспортной инфраструктуры. С математической точки зрения данная задача может быть поставлена как задача прогнозирования некоторой наблюдаемой величины, например скорости или интенсивности потока, в заданном узле транспортной сети для выбранного горизонта прогнозирования (краткосрочный прогноз — не более 1 часа).

Изначально для этой задачи были использованы два разных подхода. В 1950-х годах появились первые математические модели потока трафика [Richards, 1956; Lighthill, Whitham, 1955]. Эти модели пытаются симулировать поток трафика с использованием метода кинематических волн. В качестве альтернативы этому подходу экспоненциальное увеличение вычислительной мощности современных компьютерных систем позволило использовать подход, основанный на количественных данных трафика. Некоторыми представителями данного подхода являются ARIMA, KNN, Support Vector Regression, искусственные нейронные сети.

Семейство моделей ARMA сыграло ключевую роль в развитии данной области исследования. Оно включает в себя такие одномерные модели, как ARIMA [Ahmed, Cook, 1979] и Seasonal ARIMA [Williams, Hoel, 2003], которая является модификацией ARIMA, дополнительно использующей сезонность в данных. Эти модели показывают хорошие результаты, однако они рассматривают каждый сенсор независимо, игнорируя зависимости между соседними сенсорами.

В многомерном случае самой используемой моделью, благодаря ее простоте, является векторная авторегрессия (Vector autoregression, VAR). Эта модель также учитывает пространственные зависимости. Однако с увеличением количества анализируемых временных рядов сложность VAR-моделей растет квадратично, так как она моделирует зависимости сенсоров всех со всеми, что затрудняет их применение к крупным транспортным сетям. Общее ограничение авторегрессивных моделей — моделирование только линейных зависимостей.

Из-за стохастической и нелинейной природы данных исследователи обращали много внимания на непараметрические методы. k-NN-метод был использован [Davis, Nihan, 1991] для прогнозирования трафика и показал результаты на уровне линейных моделей, но не лучше их.

Модель Online Support Vector Regression (Online SVR) [Castro-Neto et al., 2009] успешно применялась для прогнозирования времени в пути в транспортной сети. Данная модель обновляет веса функции предсказания в режиме реального времени с помощью пошагового обучения (incremental learning). Этот метод показал себя немного лучше по сравнению с бейзлайнами и превзошел их при нетипичных неповторяющихся условиях трафика.

Исследователи давно стремились использовать потенциал искусственных нейронных сетей для предсказания трафика [Jun, Ying, 2008], однако эффективность нейронных сетей с одним внутренним слоем уступала SVR с RBF ядром [Lippi et al., 2013].

Методы глубокого обучения широко и успешно применяются для ряда задач, таких как классификация и распознавание объектов. Для прогнозирования трафика глубокие сети доверия (Deep Belief Network) показали способность к пониманию стохастических свойств дорожного движения в задаче обучения без учителя [Huang et al., 2014]. Исследователи также применяли многоярусные автоэнкодеры (Stacked Autoencoders) [Lv et al., 2014] и многоярусные шумоподавляющие автоэнкодеры (Stacked Denoising Autoencoders) [Chen et al., 2016] для предсказания характеристик потока трафика и выделения его скрытых свойств. Такие нейронные сети могут получать пространственные и временные свойства, однако их трудно обучать из-за их полной взаимосвязанности.

Рекуррентные нейронные сети (RNN) становятся все более популярными для моделирования временных рядов и естественного языка и достигают передовых результатов на популярных наборах данных [Greff et al., 2017]. В то время как сети прямого распространения учитывают фиксированное количество временных данных для прогнозирования следующего, рекуррентные нейронные сети могут принимать во внимание все предыдущие временные шаги. Обычные RNN не могут моделировать долговременные зависимости из-за проблемы исчезающего градиента. Эта проблема решается разновидностью RNN-архитектуры Long Short-Term Memory (LSTM) [Hochreiter, Schmidhuber, 1997] и Gate Recurrent Unit (GRU) [Cho et al., 2014]. Они улавливают не только кратковременные зависимости, но и долговременные. Модули LSTM и GRU состоят из нескольких полностью связанных компонентов, которые улавливают пространственно-временные зависимости в данных. Сложность обработки каждого временного шага составляет $O(n^2)$, где n — это число детекторов.

В GRU и LSTM гиперболический тангенс и сигмоида используются как функции активации, что приводит к проблеме исчезания градиента на нижних слоях, так как эти функции активации быстро насыщаются. Independent RNN [Li et al., 2018] является модификацией RNN, где нейроны в каждом слое независимы друг от друга. Многослойная архитектура позволяет обеспечивать связь между нейронами. По сравнению с LSTM эта архитектура позволяет обрабатывать гораздо более длительные временные ряды и строить более глубокие сети.

Сверточные нейронные сети (CNN) являются основным инструментом для работы с данными, имеющими пространственную структуру. Они успешно применяются для извлечения пространственных свойств потока трафика. Эти сети не страдают от проклятия размерности, поскольку они используют локальные фильтры. Более того, объединение нескольких слоев позволяет моделировать глобальные зависимости [LeCun et al., 1998]. Обычные сверточные нейронные сети хорошо работают с однородными данными, такими как изображения, видео, речь, но они не предназначены для таких сложных топологических систем, как граф.

Было предложено обобщение CNN для графов с использованием методов обработки сигналов на графе [Defferrard et al., 2016], где свертка на графе аппроксимируется с помощью метода разложения Чебышёва со свободными параметрами, которые изучаются во время обучения. Этот метод использует быстрые локальные спектральные фильтры для извлечений пространственных свойств в окрестности вершины порядка K со сложностью $O(K|\mathcal{E}|)$, где $|\mathcal{E}|$ — это количество ребер в графе.

Graph Convolutional Networks (GCN) [Kipf, Welling, 2017] упрощают предыдущий метод, используя аппроксимацию разложения Чебышёва первого порядка. Каждый фильтр применяется только к окрестности каждого узла. Во многих экспериментах эта архитектура превосходит другие методы по точности прогноза.

Объединение рекуррентных нейронных сетей и сверточных нейронных сетей позволяет хорошо моделировать пространственно-временные последовательности, так была представлена convLSTM-, LSTM-сеть со встроенной сверткой на двумерной сетке. В этом методе матричные умножения заменяются на свертку для входящего сигнала и сигнала скрытого состояния [Shi et al., 2015]. Позже он был расширен на случай графа [Seo et al., 2016]. Для решения проблемы накопления шума было предложено использовать архитектуру encoder-decoder с модулем Graph Convolutional GRU (GCGRU) [Li et al., 2017].

Постановка задачи

Исторические данные о транспортном потоке, измеренные на N сенсорах в M предыдущих моментах времени, можно рассматривать как матрицу $V = [v_1, v_2, \dots, v_M]^T$ размером $M \times N$, где v_i — это вектор измерений с сенсора на шаге i .

Связи между сенсорами описывается неориентированным взвешенным графом транспортной сети $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$, где \mathcal{V} — это множество вершин (сенсоров), \mathcal{E} — множество ребер (дорог), \mathcal{A} — взвешенная матрица смежности графа, которая показывает связность сенсоров в транспортной сети.

Следовательно, проблема прогнозирования скорости транспортного потока может быть сформулирована следующим образом в терминах максимизации вероятности:

$$v_{M+T} = \operatorname{argmax} p(v_{M+T} | v_1, v_2, \dots, v_M, \mathcal{G}). \quad (1)$$

Предлагаемый метод

Детекторы потока трафика формируют взвешенный граф, где мы определяем вес ребер в зависимости от расстояния между детекторами. Мы предполагаем, что влияние между различными частями дороги уменьшается по мере увеличения расстояния между ними.

В этой статье мы представляем свой метод решения проблемы предсказания трафика на основе Independent RNN (IndRNN) [Li et al., 2018] и Graph Convolutional Network (GCN) [Kipf, Welling, 2017]. Свертке на графе отведена задача моделирования пространственных зависимостей, в то время как IndRNN отвечает за временные зависимости. Каждый слой предлагаемой модели состоит из IndRNN, в котором матричное умножение заменяется сверткой графа с последующей активацией ReLU. Таким образом, свертка в каждом слое моделирует пространственные зависимости между соседними узлами, тогда как IndRNN самостоятельно моделирует временные зависимости нейронов. Объединение k таких слоев будет моделировать пространственно-временные зависимости в окрестности k -го порядка каждого узла.

Свертка на графе для сигнала $X \in \mathbb{R}^{N \times C}$ с C входными каналами и F фильтрами задана как [Kipf, Welling, 2017]

$$Z = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} X \Theta, \quad (2)$$

где $Z \in \mathbb{R}^{N \times F}$ — сигнал, прошедший свертку, $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$ — диагональная матрица, содержащая степени вершин, $\tilde{A} = A + I_N$, A — матрица смежности графа, I_N — единичная матрица, $\Theta \in \mathbb{R}^{C \times F}$ — матрица параметров фильтров.

Вместо обработки входных данных на каждом временном шагу с помощью полносвязного слоя мы используем графовую свертку. Мы используем skip-соединения в нашей модели, чтобы ускорить обучение, так что каждый слой учит только остаточную функцию по отношению к предыдущему слою [Kim et al., 2017]. Чтобы использовать skip-соединение (residual-соединение) в первом слое, мы проецируем данные в пространство нужной размерности. Это преобразование выполняется с той же матрицей преобразования для каждого детектора. После последнего слоя мы должны уменьшить размерность до одного. Мы связываем веса матриц, используемых для этих преобразований чтобы уменьшить количество параметров.

Традиционные модели прогнозирования трафика, такие как ARIMA и LSTM, имеют проблему накопления шума. Это происходит, когда мы используем предсказанные значения с предыдущего шага для прогнозирования следующих. Для решения этой проблемы мы используем модель sequence to sequence с архитектурой encoder-decoder [Cho et al., 2014], изображенную на рис. 1.

В модели sequence to sequence обычно используется teacher forcing во время обучения. Обучение состоит из максимизации вероятности следующего значения временного ряда на основании истинного предыдущего значения. Во время предсказания мы заменяем реальные значения на предсказанные. Такое поведение не дает нейронной сети учитывать накапливающийся

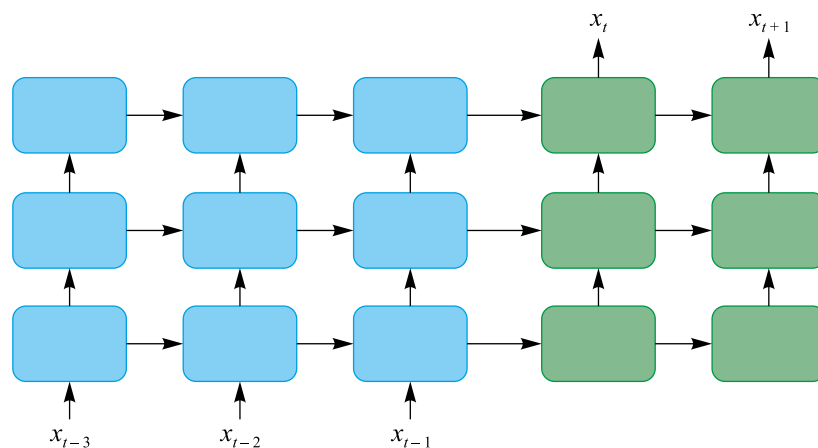


Рис. 1. Архитектура encoder-decoder. Синие ячейки представляют encoder, зеленые — decoder. Цветная версия рисунка доступна на сайте журнала

шум во время тренировочной фазы. Чтобы сделать плавный переход от подачи реального значения к предсказанному, мы используем *scheduled sampling*, которая показала высокую точность в долгосрочном предсказывании [Bengio et al., 2015].

Результаты и обсуждение

Мы использовали данные трафика, которые были получены из Caltrans Performance Measurement System [PeMS] District 4. Данные были собраны 64 датчиками на шоссе Лос-Анджелеса в течение 4.5 месяцев в 2016 году с частотой в 5 минут. Мы разделяем данные на равные тренировочный, валидационный и тестовый наборы данных. Чтобы сконструировать граф сенсоров, мы считаем матрицу смежности следующим образом:

$$a_{ij} = \begin{cases} \exp\left(-\frac{d_{ij}^2}{\sigma^2}\right), & \text{если } \exp\left(-\frac{d_{ij}^2}{\sigma^2}\right) > \varepsilon, i \neq j, \\ 0, & \text{иначе,} \end{cases} \quad (3)$$

где d_{ij} — это расстояние между датчиками, σ и ε — параметры, отвечающие за распределение весов матрицы и ее разреженность.

Мы сравниваем нашу модель со следующими моделями: 1) VAR; 2) многослойная LSTM с полносвязными соединениями (FC-LSTM) [Sutskever et al., 2014]; 3) Graph Convolutional Gated Recurrent Unit (GCGRU). Все модели реализованы в Tensorflow [Abadi et al., 2015] и обучаются, минимизируя среднеквадратичную ошибку с помощью метода оптимизации Adam [Kingma, Ba, 2015]. Мы используем batch size 32 и learning rate 0.001, уменьшая его в 10 раз, когда производительность на валидационном наборе данных перестает улучшаться. Мы принудительно ограничиваем градиенты глобальной нормой 0.25. Кроме того, мы выполняем стандартизацию входных данных и заполняем отсутствующие значения нулями.

Модель VAR берет во внимание 12 предыдущих временных шагов. FC-LSTM состоит из трех слоев LSTM с функцией активации «гиперболический тангенс» и размером вектора скрытого состояния 512. GCGRU состоит из 2 слоев, содержащих 64 блока GCGRU, и моделирует зависимости в окрестности 2 порядка вершины.

Предлагаемая модель содержит 7 слоев Residual Graph Convolutional IndRNN с 64 каналами. Для *scheduled sampling* мы выбираем истинный предыдущий токен на шаге i с вероятностью ϵ_i :

$$\epsilon_i = k_0 / (k_0 + \exp(\max(0, i - k_1) / k_0)), \quad (4)$$

где константы $k_0 \geq 1$ и $k_1 \geq 0$ зависят от скорости сходимости. Мы эмпирически установили, что $k_0 = 30$ и $k_1 = 5000$ дают более быструю сходимость.

Для оценивания мы используем три метрики: среднюю абсолютную ошибку (MAE), среднюю относительную ошибку (MAPE) и среднеквадратичное отклонение (RMSE):

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |y_t - \widehat{y}_t|, \quad (5)$$

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \widehat{y}_t}{y_t} \right| * 100\%, \quad (6)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \widehat{y}_t)^2}, \quad (7)$$

где n — количество предсказаний, y — это истинные значения предсказываемой величины (скорость), \widehat{y} — это предсказываемое значение.

Таблица 1 показывает сравнения различных подходов для прогнозирования скорости дорожного потока на 15, 30, 60 минут вперед. Результаты показывают, что методы, основанные на рекуррентных сетях, работают лучше линейной модели, и наш метод по точности прогнозирования является наилучшим, учитывая все метрики для всех горизонтов предсказания.

Таблица 1. Сравнение качества предсказаний различных подходов

	MAE (mph)	MAPE (%)	RMSE (mph)
15 минут вперед			
VAR	4.29	8.29	6.30
FC-LSTM	2.92	6.22	5.40
GCGRU	2.98	6.15	4.89
Proposed Method	2.54	5.45	4.56
30 минут вперед			
VAR	5.72	11.15	8.51
FC-LSTM	3.20	7.00	6.10
GCGRU	3.67	7.57	5.97
Proposed Method	2.78	6.14	5.10
60 минут вперед			
VAR	7.26	14.15	10.83
FC-LSTM	3.52	7.86	6.87
GCGRU	4.92	10.01	7.58
Proposed Method	3.15	7.07	5.81

Заключение

В этой статье мы представили архитектуру нейронной сети для прогнозирования скоростей транспортного потока на графе дорожной сети. Мы представили архитектуру нейронной сети, способную отображать динамику и сложность потока трафика. Модель основана на объединении рекуррентной нейронной сети и сверточной нейронной сети на графе. Архитектура encoder-decoder в сочетании с scheduled sampling успешно устраняет проблему накопления шума

и позволяет прогнозировать одновременно на несколько временных шагов вперед. Полученная объединенная нейронная сеть тренировалась на необработанных данных с сенсоров транспортного потока и имеет более низкую погрешность предсказания по сравнению с ранее опубликованными моделями.

Список литературы (References)

- Abadi M., Agarwal A., Barham P. et al.* TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. — 2015. — Software available from tensorflow.org
- Ahmed M.S., Cook A.R.* Analysis of freeway traffic time-series data by using Box–Jenkins techniques // *Transportation Research Record*. — 1979. — No. 722. — P. 116.
- Bengio S., Vinyals O., Jaitly N. et al.* Scheduled Sampling for Sequence Prediction with Recurrent Neural Networks. — 2015, jun. — P. 1–9.
- Caltrans Performance Measurement — State of California. — URL: <http://pems.dot.ca.gov/>
- Castro-Neto M., Jeong Y.S., Jeong M.K. et al.* Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions // *Expert Systems with Applications*. — 2009. — Vol. 36, No. 3. — Part 2. — P. 6164–6173.
- Chen Q., Song X., Yamada H. et al.* Learning Deep Representation from Big and Heterogeneous Data for Traffic Accident Inference // *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI-16)*. — 2016. — P. 338–344.
- Cho K., van Merriënboer B., Gulcehre C. et al.* Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. — 2014.
- Davis G.A., Nihan N.L.* Nonparametric regression and short-term freeway traffic forecasting. — 1991. — Vol. 117, No. 2. — P. 178–188.
- Defferrard M., Bresson X., Vandergheynst P.* Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering. — 2016. — No. Nips.
- Greff K., Srivastava R.K., Koutnik J. et al.* LSTM: A Search Space Odyssey // *IEEE Transactions on Neural Networks and Learning Systems*. — 2017. — Vol. 28, No. 10. — P. 2222–2232.
- Hochreiter S., Schmidhuber J.* Long Short-Term Memory // *Neural Computation*. — 1997. — Vol. 9, No. 8. — P. 1735–1780.
- Huang W., Song G., Hong H. et al.* Deep architecture for traffic flow prediction: Deep belief networks with multitask learning // *IEEE Transactions on Intelligent Transportation Systems*. — 2014. — Vol. 15, No. 5. — P. 2191–2201.
- Jun M., Ying M.* Research of Traffic Flow Forecasting Based on Neural Network // *2008 Second International Symposium on Intelligent Information Technology Application*. — 2008. — Vol. 2, No. 973. — P. 451–456.
- Kim J., El-Khamy M., Lee J.* Residual LSTM: Design of a deep recurrent architecture for distant speech recognition // *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*. — 2017. — Vol. 2017, Augus. — P. 1591–1595.
- Kingma D.P., Ba J.L.* Adam: A Method for Stochastic Optimization. — 2015. — P. 1–15.
- Kipf T.N., Welling M.* Semi-Supervised Classification with Graph Convolutional Networks // *International Conference on Learning Representations (ICLR)*. — 2017.
- LeCun Y., Bottou L., Bengio Y. et al.* Gradient-based learning applied to document recognition // *Proceedings of the IEEE*. — 1998. — Vol. 86, No. 11. — P. 2278–2323.

-
- Li S., Li W., Cook C. et al.* Independently Recurrent Neural Network (IndRNN): Building A Longer and Deeper RNN. — 2018. — No. 1.
- Li Y., Yu R., Shahabi C. et al.* Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. — 2017. — P. 1–12.
- Lighthill M.J., Whitham G.B.* On Kinematic Waves. II. A Theory of Traffic Flow on Long Crowded Roads // Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences. — 1955. — Vol. 229, No. 1178. — P. 317–345.
- Lippi M., Bertini M., Frasconi P.* Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning // IEEE Transactions on Intelligent Transportation Systems. — 2013. — Vol. 14, No. 2. — P. 871–882.
- Lv Y., Duan Y., Kang W. et al.* Traffic Flow Prediction with Big Data: A Deep Learning Approach // IEEE Transactions on Intelligent Transportation Systems. — 2015. — Vol. 16, No. 2. — P. 865–873.
- Richards P.I.* Shock Waves on the Highway // Operations Research. — 1956. — Vol. 4, No. 1. — P. 42–51.
- Seo Y., Defferrard M., Vandergheynst P. et al.* Structured Sequence Modeling with Graph Convolutional Recurrent Networks. — 2016. — No. 2013. — P. 1–10.
- Shi X., Chen Zh., Wang H. et al.* Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. — 2015. — P. 1–12.
- Sutskever I., Vinyals O., Le Quoc V.* Sequence to Sequence Learning with Neural Networks. — 2014. — P. 1–9.
- Williams B.M., Hoel L.A.* Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results // Journal of Transportation Engineering. — 2003. — Vol. 129, No. 6. — P. 664–672.