ДИФФУЗИОННАЯ СВЕРТОЧНАЯ РЕКУРРЕНТНАЯ НЕЙРОННАЯ СЕТЬ: ПРОГНОЗИРОВАНИЕ ТРАФИКА НА ОСНОВЕ ДАННЫХ

Ягуан Ли†, Роуз Юй, Сайрус Шахаби†, Янь Лю†

† Университет Южной Калифорнии, ‡ Калифорнийский технологический институт

† {ягуан, шахаби, yanliu.cs}@usc.edu, ‡ rose@caltech.edu

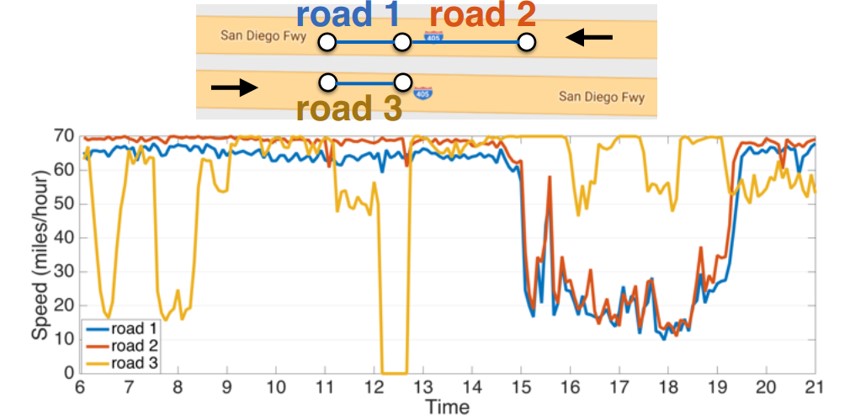
АБСТРАКТНЫЙ

Пространственно-временное прогнозирование имеет различные приложения в нейронауке, климате и транспортной сфере. Прогнозирование трафика является одним из канонических примеров такой задачи обучения. Задача сложна из-за (1) сложной пространственной зависимости от дорожной сети, (2) нелинейной временной динамики с изменяющимися дорожными условиями и (3) присущей сложности долгосрочного прогнозирования. Чтобы решить эти проблемы, мы предлагаем смоделировать транспортный поток как процесс диффузии на направленном графе и представить *диффузионную сверточную рекуррентную нейронную сеть* (DCRNN) — фреймворк глубокого обучения для прогнозирования трафика, который включает в себя как пространственную, так и временную зависимость в транспортном потоке. В частности, DCRNN фиксирует пространственную зависимость с помощью двунаправленных случайных блужданий на графе, а временную зависимость — с помощью архитектуры кодировщик-декодер с запланированной выборкой. Мы оцениваем структуру на двух реальных крупномасштабных наборах данных дорожной сети и наблюдаем последовательное улучшение на 12– 15% по сравнению с современными базовыми показателями.

# 1 ЗНАКОМСТВО

Пространственно-временное прогнозирование является важнейшей задачей для обучающейся системы, работающей в динамичной среде. Он имеет широкий спектр применений: от автономных транспортных средств до оптимизации энергетики и интеллектуальных сетей, логистики и управления цепочками поставок. В данной работе мы исследуем одну важную задачу: прогнозирование трафика на дорожных сетях, являющемся ключевым компонентом интеллектуальных транспортных систем. Целью прогнозирования дорожного движения является прогнозирование будущих скоростей движения сети датчиков с учетом исторических скоростей движения и основных дорожных сетей.

arXiv:1707.01926v3 [cs. LG] 22 фев 2018

Эта задача сложна в основном из-за сложных пространственно-временных зависимостей и присущих им трудностей долгосрочного прогнозирования. С одной стороны, временные ряды трафика демонстрируют сильную *временную динамику*. Повторяющиеся инциденты, такие как часы пик или несчастные случаи, могут привести к нестационарности, что затрудняет долгосрочное прогнозирование. С другой стороны, датчики на дорожной сети содержат сложные, но уникальные *пространственные корреляции*. На рисунке 1 показан пример. Дорога 1 и дорога 2 коррелированы, в то время как дорога 1 и

|  |  |
| --- | --- |
| Дорога 3 — нет. Хотя дороги 1 и 3 близки друг к другу в евклидовом пространстве, они демонстрируют совершенно разное поведение. Более того, | Рисунок 1: Пространственная корреляция доминирует в структуре дорожной сети. (1) Скорость движения на дороге 1 аналогична скорости движения на дороге 2, так как они расположены на одной и той же автомагистрали. |

Будущая скорость трафика в большей степени зависит от (2) Дороги 1 и 3 расположены в противоположном направлении вниз по течению, чем движение вверх по течению. ЭтотВ то же время Вы можете проехать по шоссе. Хоть и близко друг к другу означает, что пространственная структура в дорожном движении не является в евклидовом пространстве их протяженность по дорожной сети

Евклидовы и направленные. большой, и скорость их трафика существенно отличается.

Прогнозирование трафика изучается на протяжении десятилетий и делится на две основные категории: подход, основанный на знаниях, и подход, основанный на данных. В транспортных и операционных исследованиях методы, основанные на знаниях, обычно применяют теорию очередей и моделируют поведение пользователей в дорожном движении (Cascetta, 2013).

В сообществе временных рядов остаются популярными методы, основанные на данных, такие как модель авторегрессионного интегрированного скользящего среднего (ARIMA) и фильтрация Калмана (Liu et al., 2011; Lippi et al., 2013). Однако простые модели временных рядов обычно полагаются на предположение о стационарности, которое часто нарушается данными о трафике. Совсем недавно модели глубокого обучения для прогнозирования трафика были разработаны в Lv et al. (2015); Yu et al. (2017b), но без учета пространственной структуры. Wu & Tan (2016) и Ma et al. (2017) моделируют пространственную корреляцию с помощью сверточных нейронных сетей (CNN), но пространственная структура находится в евклидовом пространстве (например, двухмерные изображения). Bruna et al. (2014), Defferrard et al. (2016) изучали свертку графов, но только для неориентированных графов.

В данной работе мы представляем попарные пространственные корреляции между датчиками дорожного движения с помощью ориентированного графа, узлами которого являются датчики, а веса ребер обозначают близость между парами датчиков, измеряемую расстоянием от дорожной сети. Мы моделируем динамику транспортного потока как диффузионный процесс и предлагаем  *операцию диффузионной свертки* для учета пространственной зависимости. Кроме того, мы предлагаем *диффузионную сверточную рекуррентную нейронную сеть* (DCRNN), которая объединяет *диффузионную свертку*, архитектуру *последовательности к последовательности* и технику *запланированной выборки*. При оценке на реальных наборах данных о дорожном движении DCRNN неизменно значительно превосходит современные базовые показатели прогнозирования трафика. Вкратце:

* Исследуется задача прогнозирования трафика и моделируется пространственная зависимость трафика как диффузионный процесс на ориентированном графе. Мы предлагаем *диффузионную свертку*, которая имеет интуитивную интерпретацию и может быть эффективно вычислена.
* Мы предлагаем *диффузионную сверточную рекуррентную нейронную сеть* (DCRNN), целостный подход, который фиксирует как пространственные, так и временные зависимости между временными рядами с использованием *диффузионной свертки* и структуры обучения последовательности вместе с запланированной выборкой. DCRNN не ограничивается транспортом и легко применима для других задач пространственно-временного прогнозирования.
* Мы провели обширные эксперименты на двух крупномасштабных наборах реальных данных, и предложенный подход значительно улучшил современные базовые методы.

# 2 МЕТОДОЛОГИЯ

Формализуется задача обучения пространственно-временному прогнозированию трафика и описывается, как моделировать структуры зависимости с помощью *диффузионной сверточной рекуррентной нейронной сети*.

## 2.1 ПРОБЛЕМА ПРОГНОЗИРОВАНИЯ ТРАФИКА

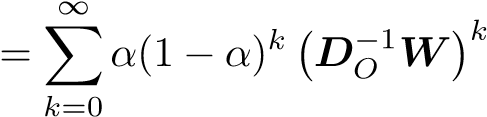
Целью прогнозирования дорожного движения является прогнозирование будущей скорости движения с учетом ранее наблюдаемого транспортного потока от *N* коррелированных датчиков на дорожной сети. Мы можем представить сенсорную сеть в виде взвешенного ориентированного графа *G = (V,E,W*), где V — набор узлов |V| *= N*, E — набор ребер, *а W* ∈ RN×N — взвешенная матрица смежности, представляющая близость узлов (например, функция расстояния от их дорожной сети). Обозначьте транспортный поток, наблюдаемый на G, в виде графа сигнала *X* ∈ RN×P, где *P* — количество характеристик каждого узла (например, скорость, объем). Пусть *X*(*t*) представляет сигнал графика, наблюдаемый в момент *времени t*, задача прогнозирования трафика направлена на обучение функции *h*(·) который сопоставляет *сигналы исторического графика* T0 с будущими сигналами *графика T*, заданным графом G:

[*X*(*t−T0+1),*··· *,X*(*t*); G] −−h(→·) [*X*(*t+1),*··· *,X*(*t+T*)]

## 2.2 МОДЕЛИРОВАНИЕ ПРОСТРАНСТВЕННЫХ ЗАВИСИМОСТЕЙ

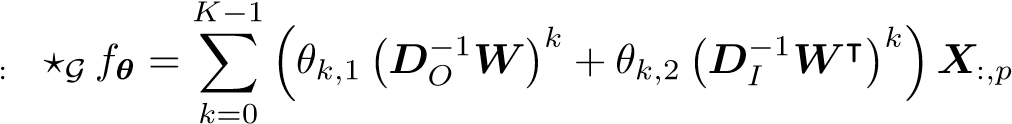
Мы моделируем пространственную зависимость, связывая транспортный поток с процессом диффузии, который явно отражает стохастическую природу динамики движения. Этот процесс диффузии характеризуется случайным блужданием на G с вероятностью перезапуска *α* ∈ [*0,1*] и матрицей переходов состояний *D**W*. Здесь *DO* = diag(*W1*) — диагональная матрица за пределами степени, а **1** ∈ RN обозначает весь вектор. После многих временных шагов такой марковский процесс сходится к стационарному распределению *P* ∈ RN×N, *i-я строка которого Pi,*: ∈ RN представляет вероятность диффузии от узла *vi* ∈ V, отсюда и близость к узлу *vi*. Следующая лемма предлагает решение закрытой формы для стационарного распределения.

Лемма 2.1. *(Teng et al., 2016) Стационарное распределение процесса диффузии может быть представлено в виде взвешенной комбинации бесконечных случайных блуждений на графике и рассчитано в замкнутой форме:*

*P* (1)

где *k* — шаг диффузии. На практике мы используем конечное K-шаговое усечение процесса диффузии и присваиваем обучаемый вес каждому шагу. Мы также включаем процесс диффузии в обратном направлении, так что двунаправленная диффузия обеспечивает модели большую гибкость для улавливания влияния как восходящего, так и нисходящего трафика.

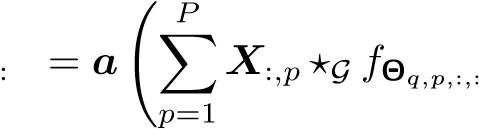
Диффузионная свертка Результирующая операция диффузионной свертки над графическим сигналом *X* ∈ RN×P и фильтром *fθ* определяется следующим образом:

*X* *,* для *p* ∈ {1*,*··· *,*} (2)

где *θ* ∈ RK×2 – параметры для фильтра*, а DO−1W,DI*−1W| представляют собой матрицы переходов диффузионного процесса и обратного соответственно. В целом, вычисление свертки может быть дорогостоящим. Однако, если G разрежено, уравнение 2 может быть эффективно вычислено с использованием  *O*(*K*) рекурсивного умножения разреженных матриц с общей временной сложностью. Более подробную информацию см. в Приложении В.

Диффузионный сверточный слой С помощью операции свертки, определенной в уравнении 2, мы можем построить диффузионный сверточный слой, который сопоставляет P-мерные характеристики с Q-мерными выходами. Обозначим тензор параметров как Θ ∈ RQ×P×K×2 = [*θ*]*q,p*, где **Θq,p,**:*,*: ∈ RK×2 параметризует сверточный фильтр для *p-го входа и q-го выхода. Таким образом, диффузионный сверточный слой представляет собой:*

!

 *H ,q* для *q* ∈ {1*,*··· *,Q*} (3)

где *X* ∈ RN×P — вход, *H* ∈ RN×Q — выход, {*fΘq,p,,*:} — фильтры и *a* — функция активации (например, ReLU, Sigmoid). Диффузионный сверточный слой изучает представления для структурированных данных графа, и мы можем обучить его с помощью метода, основанного на стохастическом градиенте.

Связь со сверткой спектрального графа Диффузионная свёртка определяется как на направленных, так и на неориентированных графах. Применительно к неориентированным графам показано, что многие существующие структурные операции свертки графов, включая популярную свертку спектральных графов, т.е. ChebNet (Defferrard et al., 2016), можно рассматривать как частный случай диффузионной свертки (вплоть до трансформационного сходства.

В то же время мы считаем, что перевод выполнен Пусть *D* обозначает матрицу степеней, а *L* — нормализованный график

Лапласиан, следующее утверждение демонстрирует эту связь. Предложение 2.2. *Свертка спектрального графа определяется как*

*X*:*,p ?*G *fθ* = **Φ** *F*(*θ*) **Φ|X**:*,p*

*с разложением на собственные значения L* = **ΦΛΦ|** *и* *эквивалентна свертке диффузии графа вплоть до преобразования подобия, когда граф* G *не направлен.*

*Доказательство.* Смотрите Приложение С.

## 2.3 МОДЕЛИРОВАНИЕ ВРЕМЕННОЙ ДИНАМИКИ

Мы используем рекуррентные нейронные сети (РНС) для моделирования временной зависимости. В частности, мы используем стробированные рекуррентные единицы (GRU) (Chung et al., 2014), которые являются простым, но мощным вариантом RNN. Мы заменяем умножение матриц в GRU на *диффузионную свертку*, что приводит к предложенной нами *диффузионной сверточной рекуррентной единице* (DCGRU).

*r*(*t*) = *s*(**Эр** *?*G [*X*(*t*)*, H*(*т−1)*] + *БР*) *u*(*t*) = *s*(**Θu** *?*G [*X*(*t*)*, H*(*т−1)*] + *бу*)

*С* *Ч*

где *X*(*t*)*,H*(*t*) обозначают вход и выход в момент *времени t*, *r*(*t*)*,u*(*t*) — сброшенный вентиль и обновленный вентиль в момент *t*, соответственно. *?*G обозначает *диффузионную свертку,* определенную в уравнении 2, а  **Θr,Θu,ΘC** являются параметрами для соответствующих фильтров. Как и ГРУ, DCGRU можно использовать для построения рекуррентных слоев нейронной сети и обучения с использованием обратного распространения во времени.

При прогнозировании на несколько шагов вперед мы используем архитектуру *Sequence to Sequence* (Sutskever et al., 2014). И энкодер, и декодер являются рекуррентными нейронными сетями с DCGRU. Во время обучения мы загружаем исторический временной ряд в энкодер и используем его конечные состояния для инициализации

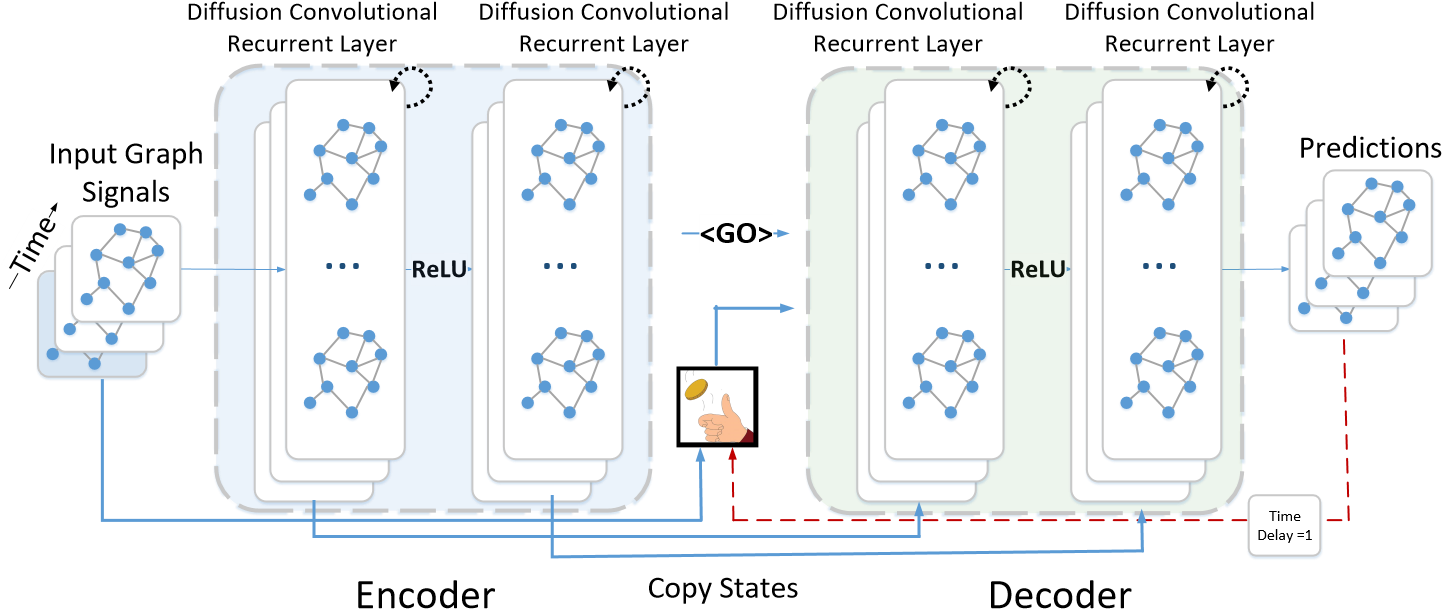


Рисунок 2: Архитектура системы для *диффузионной сверточной рекуррентной нейронной сети,* предназначенной для пространственно-временного прогнозирования трафика. Исторические временные ряды подаются в кодировщик, конечные состояния которого используются для инициализации декодера. Декодер делает прогнозы на основе либо предыдущей достоверной информации, либо выходных данных модели.

дешифратор. Декодер генерирует прогнозы на основе предыдущих *наземных наблюдений*. Во время тестирования наземные наблюдения заменяются прогнозами, сгенерированными самой моделью. Расхождение между распределениями входных данных обучения и тестирования может привести к снижению производительности. Чтобы смягчить эту проблему, мы интегрируем *в модель запланированную выборку* (Bengio et al., 2015), где мы вводим в модель либо наземное наблюдение с вероятностью *i*, либо прогноз модели с вероятностью  на *i-й итерации. В процессе обучения i* постепенно уменьшается до 0, чтобы позволить модели изучить тестовое распределение.

С помощью пространственного и временного моделирования мы строим *диффузионную сверточную рекуррентную нейронную систему*

*Сеть* (DCRNN). Архитектура модели DCRNN показана на рисунке 2. Вся сеть обучается путем максимизации вероятности создания целевого будущего временного ряда с использованием обратного распространения во времени. DCRNN способна фиксировать пространственно-временные зависимости между временными рядами и может быть применена к различным задачам пространственно-временного прогнозирования.

# 3 СВЯЗАННАЯ РАБОТА

Прогнозирование трафика является классической проблемой в транспортных и операционных исследованиях, которые в первую очередь основаны на теории и моделировании очередей (Drew, 1968). Подходы к прогнозированию трафика, основанные на данных, привлекли значительное внимание, и более подробную информацию можно найти в недавнем обзорном документе (Vlahogianni et al., 2014) и ссылках в нем. Однако существующие модели машинного обучения либо накладывают строгие стационарные предположения на данные (например, авторегрессионная модель), либо не учитывают крайне нелинейную временную зависимость (например, модель латентного пространства Yu et al. (2016); Deng et al. (2016)). Модели глубокого обучения открывают новые возможности для решения задачи прогнозирования временных рядов. Например, в Yu et al. (2017b); Laptev et al. (2017) авторы изучают прогнозирование временных рядов с помощью глубоких рекуррентных нейронных сетей (РНС). Сверточные нейронные сети (СНС) также применяются для прогнозирования трафика. Zhang et al. (2016; 2017) преобразуют дорожную сеть в обычную двухмерную сетку и применяют традиционную СНС для прогнозирования потока толпы. Cheng et al. (2017) предлагают DeepTransport, который моделирует пространственную зависимость путем явного сбора данных о дорогах вверх и вниз по течению для каждой отдельной дороги, а затем проводит свертку по этим районам соответственно.

В последнее время СНС была обобщена до произвольных графов, основанных на теории спектральных графов. Графовые сверточные нейронные сети (GCN) впервые представлены в работе Bruna et al. (2014), которая соединяет теорию спектральных графов и глубокие нейронные сети. Defferrard et al. (2016) предлагают ChebNet, который улучшает GCN с помощью быстрых локализованных фильтров сверток. Kipf & Welling (2017) упрощают ChebNet и достигают самых современных результатов в задачах полуконтролируемой классификации. Seo et al. (2016) объединяют ChebNet с рекуррентными нейронными сетями (RNN) для моделирования структурированных последовательностей. Yu et al. (2017a) моделируют сенсорную сеть в виде неориентированного графа и применяют ChebNet и модель сверточной последовательности (Gehring et al., 2017) для прогнозирования. Одним из ограничений упомянутых спектральных сверток является то, что они, как правило, требуют, чтобы график был ненаправленным для вычисления значимого

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 15 мин | МАЭ РМШЭ | 4.16  7.80 | 3.99  8.21 | 4.42  7.89 | 3.99  8.45 | 3.99  7.94 | 3.44  6.30 | 2.77 5.38 |
|  | МАПЕ | 13.0% | 9.6% | 10.2% | 9.3% | 9.9% | 9.6% | 7.3% |
| 30 мин | МАЭ РМШЭ | 4.16  7.80 | 5.15  10.45 | 5.41  9.13 | 5.05  10.87 | 4.23  8.17 | 3.77  7.23 | 3.15  6.45 |
|  | МАПЕ | 13.0% | 12.7% | 12.7% | 12.1% | 12.9% | 10.9% | 8.8% |
| 1 час | МАЭ РМШЭ | 4.16  7.80 | 6.90  13.23 | 6.52  10.11 | 6.72  13.76 | 4.49  8.69 | 4.37  8.69 | 3.60  7.59 |
|  | МАПЕ | 13.0% | 17.4% | 15.8% | 16.7% | 14.0% | 13.2% | 10.5% |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 15 мин | МАЭ РМШЭ | 2.88  5.59 | 1.62  3.30 | 1.74  3.16 | 1.85  3.59 | 2.20  4.42 | 2.05  4.19 | 1.38  2.95 |
|  | МАПЕ | 6.8% | 3.5% | 3.6% | 3.8% | 5.19% | 4.8% | 2.9% |
| 30 мин | МАЭ РМШЭ | 2.88 5.59 | 2.33 4.76 | 2.32 4.25 | 2.48 5.18 | 2.30 4.63 | 2.20 4.55 | 1.74  3.97 |
|  | МАПЕ | 6.8% | 5.4% | 5.0% | 5.5% | 5.43% | 5.2% | 3.9% |
| 1 час | МАЭ РМШЭ | 2.88  5.59 | 3.38  6.50 | 2.93  5.44 | 3.28  7.08 | 2.46  4.98 | 2.37  4.96 | 2.07  4.74 |
|  | МАПЕ | 6.8% | 8.3% | 6.5% | 8.0% | 5.89% | 5.7% | 4.9% |

Таблица 1: Сравнение производительности различных подходов к прогнозированию скорости движения. DCRNN достигает наилучшей производительности по всем трем метрикам для всех горизонтов прогнозирования, и это преимущество становится более очевидным с увеличением горизонта прогнозирования.

МЕТР-ЛА

ПЕМС-БЭЙ

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *T* | Метрический | ХА | АРИМАКал | Реселлерная система | СВР | ФНН | ФК-ЛСТМ | *DCRNN* |

спектральное разложение. Переходя от спектральной области к вершинной, Этвуд и Таусли (2016) предложили диффузионно-сверточную нейронную сеть (DCNN), которая определяет свертку как процесс диффузии по каждому узлу в графово-структурированном входе. Hechtlinger et al. (2017) предлагают GraphCNN обобщить свертку на граф путем свертывания каждого узла с его ближайшими соседями p. Однако оба эти метода не учитывают временную динамику и в основном имеют дело со статическими настройками графика.

Наш подход отличается от всех этих методов как постановкой задачи, так и формулировкой свертки на графике. Мы моделируем сенсорную сеть в виде взвешенного ориентированного графа, который более реалистичен, чем сетка или неориентированный граф. Кроме того, предложенная свертка определена с помощью случайного блуждания двунаправленного графа и в дальнейшем интегрирована со структурой обучения последовательности, а также с запланированной выборкой для моделирования долгосрочной временной зависимости.

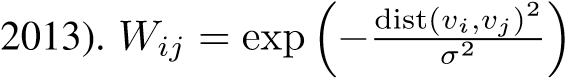
# 4 ЭКСПЕРИМЕНТЫ

Мы проводим эксперименты на двух реальных крупномасштабных наборах данных: (1) METR-LA Этот набор данных о дорожном движении содержит информацию о дорожном движении, собранную с помощью петлевых детекторов на шоссе округа Лос-Анджелес (Jagadish et al., 2014). Мы выбрали 207 датчиков и собрали данные за 4 месяца с 1 марта 2012 года по 30 июня 2012 года для эксперимента. (2) PEMS-BAY Этот набор данных о дорожном движении собирается Системой измерения эффективности (PeMS) Калифорнийских транспортных агентств (CalTrans). Мы выбрали 325 датчиков в районе залива Сан-Франциско и собрали данные за 6 месяцев с 1 января 2017 года по 31 мая 2017 года для эксперимента. Распределения датчиков обоих наборов данных визуализированы на рисунке 8 в приложении.

В обоих этих наборах данных мы агрегируем показания скорости движения в 5-минутные окна и применяем

Нормализация Z-Score. 70% данных используется для обучения, 20% — для тестирования, а оставшаяся часть

10% за валидацию. Для построения графа датчиков мы вычисляем расстояния между датчиками по парной дорожной сети и строим матрицу смежности с использованием порогового гауссова ядра (Shuman et al.,

 если dist(*vi,vj*) ≤ *K*иначе 0*,* где *Вий* представляет собой ребро

Вес между датчиком *VI* и датчиком *VJ*, dist(*vi,vj*) обозначает расстояние по дорожной сети от датчика *VI* до датчика *VJ*. *σ* — стандартное отклонение расстояний, *а κ* — порог.

## 4.1 ЭКСПЕРИМЕНТАЛЬНЫЕ НАСТРОЙКИ

Базовых линий Сравниваем DCRNN[[1]](#footnote-1) с широко используемыми регрессионными моделями временных рядов, в том числе (1)

HA: Historical Average, который моделирует транспортный поток как сезонный процесс и использует взвешенные

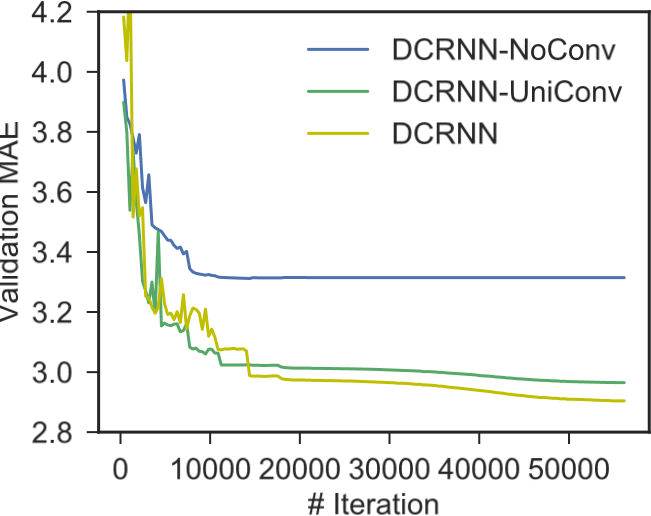


Рисунок 3: Кривая обучения для DCRNN и DCRNN без диффузионной свертки. Устранение свертки диффузии приводит к гораздо более высокой ошибке валидации. Более того, DCRNN с двунаправленным случайным блужданием достигает наименьшей ошибки валидации.

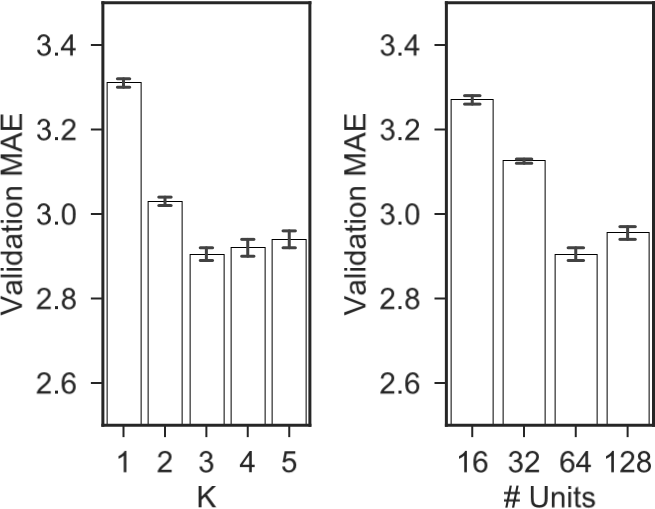


Рисунок 4: Эффекты K и количество единиц в каждом слое DCRNN. K соответствует ширине поля приема фильтра, а количество единиц соответствует количеству фильтров.

среднее значение предыдущих сезонов в качестве прогноза; (2) ARIMAkal: Авторегрессионная интегрированная модель скользящей средней с фильтром Калмана, которая широко используется для прогнозирования временных рядов; (3) VAR: Вектор

Авторегрессия (Hamilton, 1994). (4) SVR: Регрессия опорных векторов, которая использует метод линейных опорных векторов для задачи регрессии; Также включены следующие подходы, основанные на глубоких нейронных сетях: (5) Нейронная сеть с прямой связью (FNN): нейронная сеть с прямой связью с двумя скрытыми слоями и регуляризацией L2. (6) Рекуррентная нейронная сеть с полностью связными скрытыми единицами LSTM (FC-LSTM) (Sutskever et al., 2014).

Все подходы, основанные на нейронных сетях, реализованы с использованием Tensorflow (Abadi et al., 2016) и обучены с помощью оптимизатора Адама с отжигом скорости обучения. Наилучшие гиперпараметры выбираются с помощью древовидной оценки Парзена (TPE) (Bergstra et al., 2011) на валидационном наборе данных. Подробные настройки параметров для DCRNN, а также базовые показатели доступны в Приложении E.

## 4.2 СРАВНЕНИЕ ЭФФЕКТИВНОСТИ ПРОГНОЗИРОВАНИЯ ТРАФИКА

В таблице 1 показано сравнение различных подходов для прогнозирования на 15 минут, 30 минут и 1 час вперед по обоим наборам данных. Эти методы оцениваются на основе трех часто используемых метрик в прогнозировании трафика, включая (1) среднюю абсолютную ошибку (MAE), (2) среднюю абсолютную процентную ошибку (MAPE) и (3) среднеквадратичную ошибку (RMSE). Отсутствующие значения исключаются при расчете этих метрик. Подробные формулировки этих метрик приведены в Приложении E.2. В обоих наборах данных мы наблюдаем следующее явление. (1) Методы, основанные на RNN, включая FC-LSTM и DCRNN, в целом превосходят другие базовые показатели, что подчеркивает важность моделирования временной зависимости. (2) DCRNN достигает наилучшей производительности по всем метрикам для всех горизонтов прогнозирования, что говорит об эффективности моделирования пространственно-временных зависимостей. (3) Методы, основанные на глубоких нейронных сетях, включая FNN, FC-LSTM и DCRNN, как правило, имеют более высокую производительность, чем линейные базовые показатели для долгосрочного прогнозирования, например, на 1 час вперед. Это связано с тем, что временная зависимость становится все более нелинейной с увеличением горизонта. Кроме того, поскольку метод исторического среднего не зависит от краткосрочных данных, его эффективность инвариантна к небольшому увеличению горизонта прогнозирования.

Обратите внимание, что прогнозирование трафика в наборе данных METR-LA (Лос-Анджелес, который известен своими сложными условиями дорожного движения) является более сложным, чем в наборе данных PEMS-BAY (Bay Area). Таким образом, мы используем METR-LA в качестве набора данных по умолчанию для последующих экспериментов.

## 4.3 ЭФФЕКТ МОДЕЛИРОВАНИЯ ПРОСТРАНСТВЕННЫХ ЗАВИСИМОСТЕЙ

Для дальнейшего исследования эффекта моделирования пространственной зависимости мы сравниваем DCRNN со следующими вариантами: (1) DCRNN-NoConv, который игнорирует пространственную зависимость, заменяя матрицы переходов в диффузионной свертке (уравнение 2) на идентичные матрицы. По сути, это означает, что прогнозирование датчика может быть выведено только на основе его собственных исторических показаний; (2) DCRNN-UniConv, Таблица 2: Сравнение производительности DCRNN и GCRNN на наборе данных METRA-LA.

15

Мин

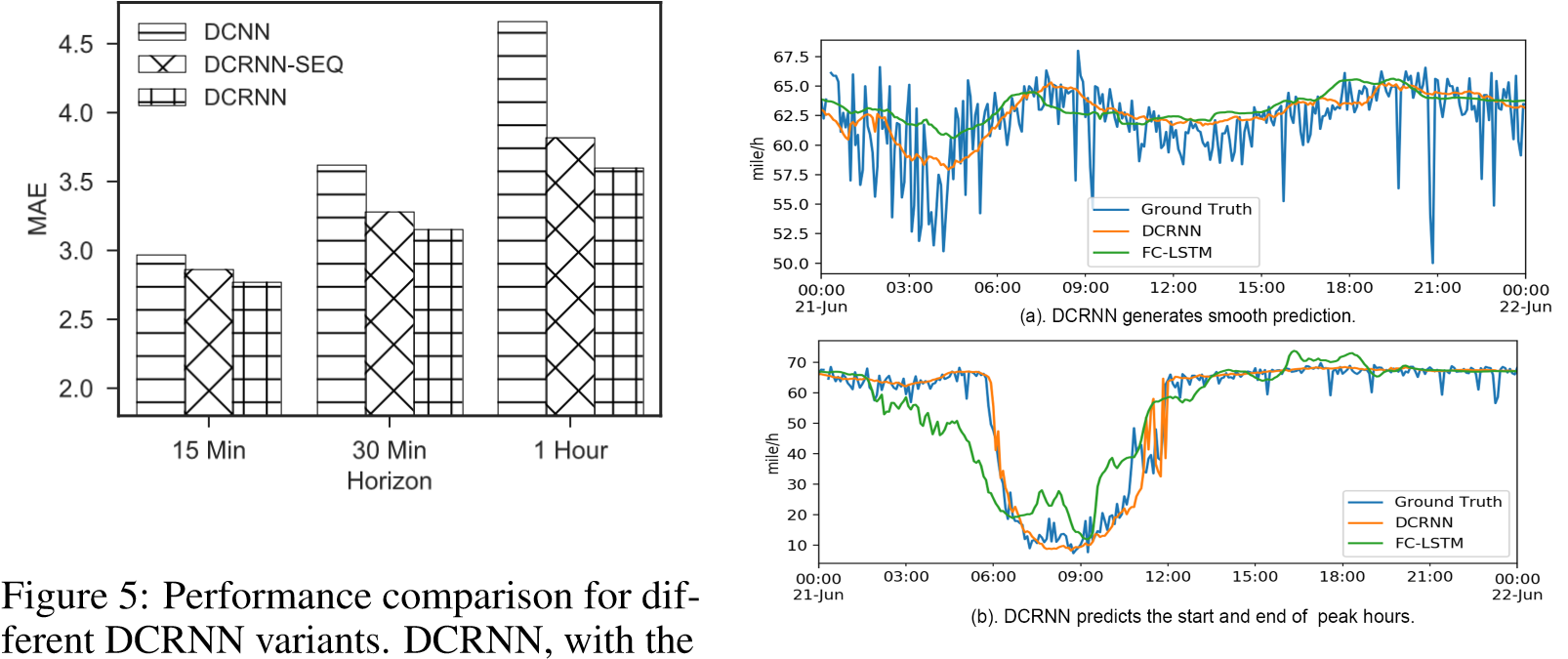
30

Мин

1

час

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | МАЕ | RMSE | МАПЕ | МАЕ | RMSE | МАПЕ | МАЕ | RMSE | МАПЕ |
| DCRNN | 2.77 | 5.38 | 7.3% | 3.15 | 6.45 | 8.8% | 3.60 | 7.60 | 10.5% |
| ГКРНН | 2.80 | 5.51 | 7.5% | 3.24 | 6.74 | 9.0% | 3.81 | 8.16 | 10.9% |



Последовательность к последовательностям и запланированная выборка достигают самого низкого MAE на рисунке 6: Визуализация прогнозирования временных рядов трафика. Валидационный набор данных. Преимущество DCRNN заключается в том, что DCRNN генерирует плавное прогнозирование и обычно становится более ясным с увеличением лучшего прогноза начала и окончания часов пиковой нагрузки. горизонт прогнозирования.

которая использует только прямую матрицу перехода случайного блуждания для диффузионной свертки; На рисунке 3 показаны кривые обучения этих трех моделей с примерно одинаковым количеством параметров. Без диффузионной свертки DCRNN-NoConv имеет гораздо более высокую ошибку валидации. Кроме того, DCRNN достигает наименьшей погрешности валидации, что показывает эффективность использования двунаправленного случайного блуждания. Интуиция заключается в том, что двунаправленное случайное блуждание дает модели возможность и гибкость для улавливания влияния как восходящего, так и нисходящего трафика.

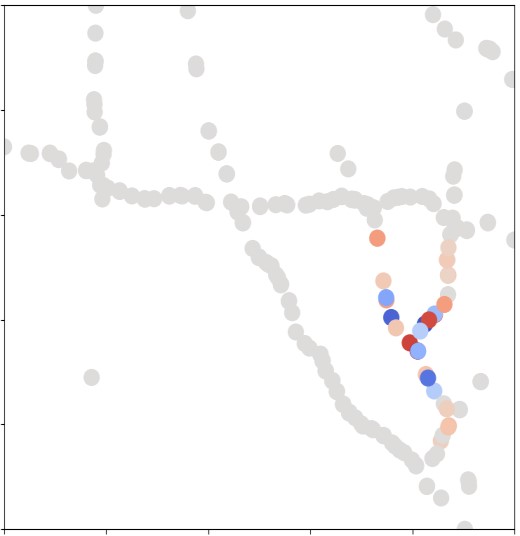
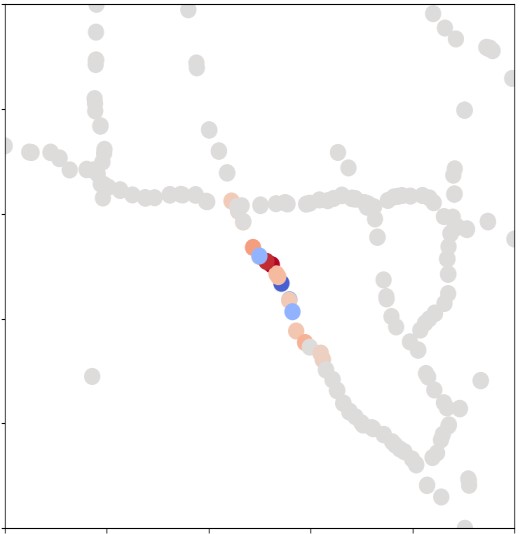
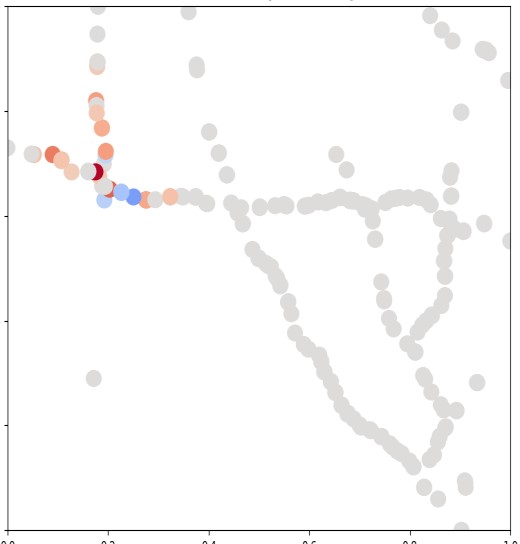
Чтобы исследовать эффект построения графа, мы строим неориентированный граф путем установки

, где *Wc* — новая симметричная весовая матрица. Затем мы разрабатываем вариант

DCRNN обозначает GCRNN, который использует обучение последовательности с *помощью свертки графа ChebNet* (уравнение 5) с примерно таким же количеством параметров. В таблице 2 показано сравнение между DCRNN и GCRNN в наборе данных METR-LA. DCRNN стабильно превосходит GCRNN. Интуиция подсказывает, что ориентированный граф лучше отражает асимметричную корреляцию между датчиками дорожного движения. На рисунке 4 показано влияние различных параметров. *K* примерно соответствует размеру полей приема фильтров, а количество единиц соответствует количеству фильтров. Большее *K* позволяет модели охватить более широкую пространственную зависимость за счет увеличения сложности обучения. Мы наблюдаем, что с увеличением *K* погрешность на валидационном наборе данных сначала быстро уменьшается, а затем немного увеличивается. Аналогичное поведение наблюдается и при изменении количества единиц.

## 4.4 ЭФФЕКТ МОДЕЛИРОВАНИЯ ВРЕМЕННЫХ ЗАВИСИМОСТЕЙ

Чтобы оценить эффект временного моделирования, включая структуру последовательности к последовательности, а также механизм запланированной выборки, мы разрабатываем три варианта DCRNN: (1) DCNN: в котором мы объединяем исторические наблюдения в виде вектора фиксированной длины и подаем его в сложенные диффузионные сверточные слои для прогнозирования будущих временных рядов. Мы обучаем одну модель для прогнозирования на один шаг вперед и вводим предыдущий прогноз в модель в качестве входных данных для выполнения прогнозирования на несколько шагов вперед. (2) DCRNN-SEQ: который использует среду обучения последовательности к последовательности кодировщика-декодера для прогнозирования на несколько шагов вперед. (3) DCRNN: аналогичен DCRNN-SEQ, за исключением добавления выборки по расписанию.



центр



Макс

Мин

0

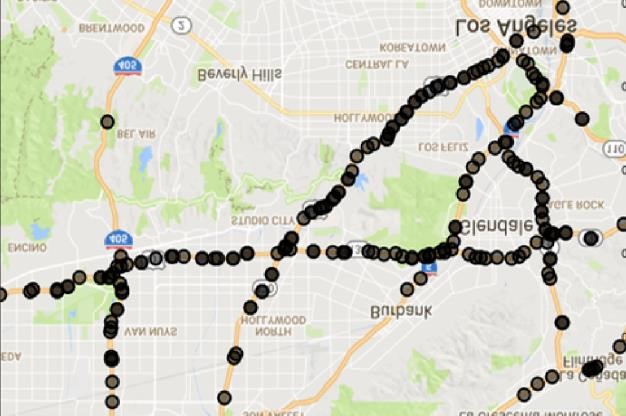


Рисунок 7: Визуализация обученных локализованных фильтров, центрированных в разных узлах с *K* = 3 на

Набор данных METR-LA. Звезда обозначает центр, а цвета — веса. Мы видим, что веса локализованы вокруг центра и рассеиваются вдоль дорожной сети.

На рисунке 5 показано сравнение этих четырех методов применительно к MAE для разных горизонтов прогнозирования. Мы видим, что: (1) DCRNN-SEQ превосходит DCNN с большим отрывом, что соответствует важности моделирования временной зависимости. (2) DCRNN достигает наилучшего результата, и его превосходство становится более очевидным с увеличением горизонта прогнозирования. В основном это связано с тем, что модель обучена справляться со своими ошибками во время прогнозирования на несколько шагов вперед и, таким образом, меньше страдает от проблемы распространения ошибок. Мы также обучаем модель, которая всегда получала выходные данные в качестве входных данных для прогнозирования на несколько шагов вперед. Однако его показатели значительно хуже, чем у всех трех вариантов, что подчеркивает важность планового отбора проб.

## 4.5 ИНТЕРПРЕТАЦИЯ МОДЕЛЕЙ

Чтобы лучше понять модель, мы визуализируем результаты прогнозирования, а также изученные фильтры. На рисунке 6 показана визуализация прогнозирования на 1 час вперед. У нас есть следующие наблюдения: (1) DCRNN генерирует плавное предсказание среднего при наличии небольших колебаний в скоростях движения

(Рисунок 6(а)). Это отражает надежность модели. (2) DCRNN с большей вероятностью точно предсказывает резкие изменения скорости трафика, чем базовые методы (например, FC-LSTM). Как показано на рисунке 6(b), DCRNN прогнозирует начало и окончание часов пиковой нагрузки. Это связано с тем, что DCRNN фиксирует пространственную зависимость и может использовать изменения скорости в датчиках окрестностей для более точного прогнозирования. На рисунке 7 представлены примеры изученных фильтров, центрированных в разных узлах. Звезда обозначает центр, а цвета обозначают веса. Мы можем наблюдать, что (1) веса хорошо локализованы вокруг центра, и (2) веса диффундируют в зависимости от расстояния от дорожной сети. Дополнительные визуализации представлены в Приложении F.

# 5 ЗАКЛЮЧЕНИЕ

В данной работе мы сформулировали прогнозирование трафика на дорожной сети как пространственно-временную задачу прогнозирования и предложили *диффузионную сверточную рекуррентную нейронную сеть* , которая фиксирует пространственно-временные зависимости. В частности, мы используем случайное блуждание двунаправленного графа для моделирования пространственной зависимости и рекуррентную нейронную сеть для захвата временной динамики. Мы дополнительно интегрировали архитектуру кодера-декодера и метод запланированной выборки для повышения производительности долгосрочного прогнозирования. При оценке на двух крупномасштабных наборах данных о реальном трафике наш подход показал значительно более точный прогноз, чем базовые показатели. В будущей работе мы исследуем следующие два аспекта: (1) применение предложенной модели к другим задачам пространственно-временного прогнозирования; (2) моделирование пространственно-временной зависимости при развитии базовой структуры графа, например, графа K ближайших соседей для движущихся объектов.

## ПОДТВЕРЖДЕНИЯ

Это исследование было частично профинансировано грантами NSF CNS-1461963, IIS-1254206, IIS-1539608, Caltrans-65A0533, Центром интегрированных медиасистем Университета Южной Калифорнии (IMSC) и Университетом Южной Калифорнии METRANS

Транспортный центр. Любые мнения, выводы, выводы или рекомендации, выраженные в этом материале, принадлежат авторам и не обязательно отражают точку зрения любого из спонсоров, таких как NSF. Кроме того, авторы хотели бы поблагодарить Шан-Хуа Тэна, Дэхуа Ченга и Сиян Ли за полезные обсуждения и комментарии.

# ССЫЛКИ

Мартин Абади и др. Tensorflow: Крупномасштабное машинное обучение в гетерогенных распределенных системах. *Препринт arXiv arXiv:1603.04467*, 2016.

Джеймс Этвуд и Дон Таусли. Диффузионно-сверточные нейронные сети. В *книге «Достижения в системах обработки нейронной информации*», стр. 1993–2001, 2016.

Сами Бенджио, Ориол Виньялс, Навдип Джейтли и Ноам Шазир. Выборка по расписанию для прогнозирования последовательности с помощью рекуррентных нейронных сетей. В *NIPS,* стр. 1171–1179, 2015.

Джеймс С. Бергстра, Реми Барденет, Йошуа Бенджио и Бал azs K' egl. Алгоритмы оптимизации гиперпараметров. В *Достижения в области нейронных систем обработки информации*, с. 2546–2554, 2011.

Джоан Бруна, Войцех Заремба, Артур Слам и Ян Лекун. Спектральные сети и локально связанные сети на графах. В *ICLR*, 2014.

Пинлун Цай, Юньпэн Ван, Гуанцюань Лу, Пэн Чэнь, Чуань Дин и Цзяньпин Сунь. Пространственно-временная коррелятивная k-модель ближайшего соседа для краткосрочного многоступенчатого прогнозирования трафика. *Транспортные исследования, часть C: Новые технологии*, 62:21–34, 2016.

Эннио Каскетта. *Инженерия транспортных систем: теория и методы*, том 49. Springer Science & Business Media, 2013.

Дэхуа Чэн, Юй Чэн, Янь Лю, Ричард Пэн и Шан-Хуа Тэн. Эффективная выборка для гауссовских графических моделей с помощью спектральной разреженности. Конференция *по теории обучения*, стр. 364–390, 2015.

Синъи Чэн, Жуйцин Чжан, Цзе Чжоу и Вэй Сюй. Deeptransport: Изучение пространственно-временной зависимости для прогнозирования состояния дорожного движения. *Препринт arXiv arXiv:1709.09585*, 2017.

Джунён Чунг, Чаглар Гульчере, Кён Хён Чо и Йошуа Бенджио. Эмпирическая оценка стробированных рекуррентных нейронных сетей при моделировании последовательностей. *Препринт arXiv arXiv:1412.3555,* 2014.

Михаэль Дефферрар, Ксавье Брессон и Пьер Вандергейнст. Сверточные нейронные сети на ̈ графах с быстрой локализованной спектральной фильтрацией. В *NIPS,* стр. 3837–3845, 2016.

Динсюн Дэн, Сайрус Шахаби, Угур Демирюрек, Линьхун Чжу, Роуз Юй и Янь Лю. Модель латентного пространства для дорожных сетей для прогнозирования изменяющегося во времени трафика. В *SIGKDD,* стр. 1525–1534, 2016. Дональд Р. Дрю. Теория и управление транспортными потоками. Технический отчет, 1968 год.

Гаэтано Фуско, Кьяра Коломбарони и Наталья Исаенко. Краткосрочные прогнозы скорости с использованием больших данных о крупных городских дорожных сетях. *Транспортные исследования, часть C: Новые технологии*, 73:183–201, 2016.

Йонас Геринг, Майкл Аули, Давид Гранжье, Денис Яратс и Ян Н Дофин. Обучение от сверточной последовательности к последовательности. В *ICML,* 2017.

Адитья Гровер и Юре Лесковец. node2vec: Масштабируемое обучение функций для сетей. В *материалах 22-й международной конференции ACM SIGKDD по обнаружению знаний и интеллектуальному анализу данных*, стр. 855–864. ACM, 2016.

Джеймс Дуглас Гамильтон. *Анализ временных рядов*, том 2. Издательство Принстонского университета, Принстон, 1994.

Йотам Хехтлингер, Пурваша Чакраварти и Цинин Цинь. Обобщение сверточных нейронных сетей на графово-структурированные данные. *Препринт arXiv arXiv:1704.08165*, 2017.

Х. В. Джагадиш, Йоханнес Герке, Александрос Лабринидис, Яннис Папаконстантину, Джигнеш М. Патель, Рагху Рамакришнан и Сайрус Шахаби. Большие данные и связанные с ними технические проблемы. *Коммуна. ACM*, 57(7):86–94, июль 2014.

Томас Н. Кипф и Макс Веллинг. Полуконтролируемая классификация с графовыми сверточными сетями. В *Международной конференции по учебным представлениям (ICLR),* 2017.

Николай Лаптев, Джейсон Йосински, Ли Эрран Ли и Славек Смыл. Прогнозирование экстремальных событий временных рядов с помощью нейронных сетей в Uber. На *Международной конференции по временным рядам машинного обучения*, 2017 г.

Марко Липпи, Марко Бертини и Паоло Фраскони. Краткосрочное прогнозирование транспортных потоков: экспериментальное сравнение анализа временных рядов и контролируемого обучения. *ITS, IEEE Transactions on*, 14(2): 871–882, 2013.

Вэй Лю, Юй Чжэн, Санджай Чавла, Цзин Юань и Се Син. Обнаружение пространственно-временных причинно-следственных взаимосвязей в потоках данных о дорожном движении. В *SIGKDD,* стр. 1010–1018. ACM, 2011.

Ишэн Лю, Яньцзе Дуань, Вэньвэнь Кан, Чжэнси Ли и Фэй-Юэ Ван. Прогнозирование транспортных потоков с помощью больших данных: подход к глубокому обучению. *ITS, IEEE Transactions on*, 16(2):865–873, 2015.

Сяолэй Ма, Чжуан Дай, Чжэнбин Хэ, Цзихуэй Ма, Юн Ван и Юньпэн Ван. Обучение трафика в виде изображений: глубокая сверточная нейронная сеть для прогнозирования скорости крупномасштабной транспортной сети. *Датчики*, 17(4):818, 2017.

Брайан Пероцци, Рами Аль-Рфу и Стивен Скина. Deepwalk: Онлайн-обучение социальным представлениям. В *материалах 20-й международной конференции ACM SIGKDD по обнаружению знаний и интеллектуальному анализу данных*, стр. 701–710. ACM, 2014.

Ёнджу Со, Михаэль Дефферрар, Пьер Вандергейнст и Ксавье Брессон. Структурированное моделирование последовательностей ̈ с помощью графовых сверточных рекуррентных сетей. *Препринт arXiv arXiv:1612.07659*, 2016.

Давид I Шуман, Сунил К. Наранг, Паскаль Фроссар, Антонио Ортега и Пьер Вандергейнст. Новая область обработки сигналов на графах: расширение многомерного анализа данных на сети и другие нерегулярные домены. *Журнал IEEE Signal Processing Magazine*, 30(3):83–98, 2013.

Илья Суцкевер, Ориол Виньялс и Куок в Ле. Последовательное обучение с помощью нейронных сетей. В *NIPS,* стр. 3104–3112, 2014.

Шан-Хуа Тенг и др. Масштабируемые алгоритмы для анализа данных и сетей. *Основы и направления* R *в теоретической информатике*, 12(1–2):1–274, 2016.

Элени и Влахоянни, Мэтью Г. Карлафтис и Джон К. Голиас. Краткосрочное прогнозирование трафика: где мы находимся и куда направлялись. *Транспортные исследования, часть C: Новые технологии*, 43:3–19, 2014.

Юанькай Ву и Хуачунь Тань. Краткосрочное прогнозирование транспортных потоков с пространственно-временной корреляцией в гибридной среде глубокого обучения. *Препринт arXiv arXiv:1612.01022*, 2016.

Юаньчан Се, Кайгуан Чжао, Ин Сунь и Давэй Чэнь. Гауссовы процессы для краткосрочного прогнозирования объемов трафика. *Отчет о транспортных исследованиях: Журнал Совета по транспортным исследованиям*, (2165):69–78, 2010.

Бин Юй, Хаотэн Инь и Чжаньсин Чжу. Сверточная нейронная сеть пространственно-временного графа: фреймворк глубокого обучения для прогнозирования трафика. *Препринт arXiv arXiv:1709.04875*, 2017a.

Сян-Фу Юй, Нихил Рао и Индерджит С. Диллон. Факторизация временной регуляризованной матрицы для прогнозирования многомерных временных рядов. В *книге «Достижения в системах обработки нейронной информации*», стр. 847–855, 2016.

Роуз Юй, Ягуан Ли, Сайрус Шахаби, Угур Демирюрек и Ян Лю. Глубокое обучение: общий подход к прогнозированию дорожного движения в экстремальных условиях. На *Международной конференции SIAM по интеллектуальному анализу данных (SDM),* 2017b.

Цзюньбо Чжан, Юй Чжэн, Декан Ци, Жуйюань Ли и Сювэнь И. Модель прогнозирования на основе Dnn для пространственно-временных данных. В *материалах 24-й Международной конференции ACM SIGSPATIAL по достижениям в области географических информационных систем*, стр. 92. ACM, 2016.

Цзюньбо Чжан, Юй Чжэн и Дэкан Ци. Глубокие пространственно-временные остаточные сети для прогнозирования потоков толпы в масштабах города. В *AAAI,* стр. 1655–1661, 2017.

# ПРИЛОЖЕНИЕ

## A НОТАЦИЯ

Таблица 3: Нотация

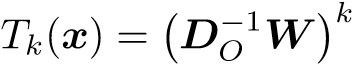
Имя

|  |  |
| --- | --- |
| G | График |
| V,vi | узлы графа, |V| = *N* и i-й узел. |
| E | Ребра графика |
| *W,Wij,* | Весовая матрица графа и его элементы |
| *D,of, два* | ненаправленная матрица степеней, матрица в степени/вне степени |
| *L* | нормализованный граф Лапласа |
| *Ф,Л* | матрица собственных векторов и матрица собственных значений *L* |
| X,Xˆ ∈ RN×P | сигнал графика и прогнозируемый сигнал графика. |
| *X*(*t*) ∈ RN×P | Сигнал графика в момент времени *t*. |
| *Г* ∈ РН×Q | вывод диффузионного сверточного слоя. |
| *фи,и* | сверточный фильтр и его параметры. |
| *f**I,Th* | сверточный слой и его параметры. |

В таблице 3 приведены основные обозначения, использованные в статье.

## B ЭФФЕКТИВНОЕ ВЫЧИСЛЕНИЕ УРАВНЕНИЯ 2

Уравнение 2 можно разложить на две части с одинаковой временной сложностью, т.е. одну часть с *DO−1W,* а другую с *DI−1W|*. Таким образом, мы покажем только временную сложность первой части.

Пусть*x*, Первая часть уравнения 2 может быть переписана как

*К−1*

X *ЯкТк*(*X*:*,*) (4)

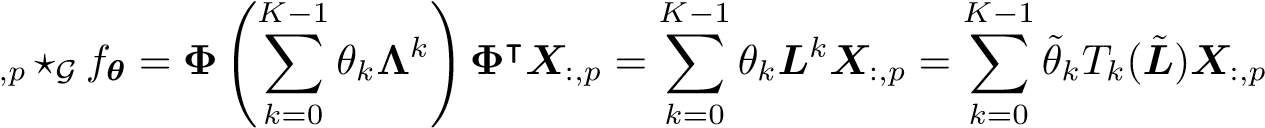
*к=0*

Поскольку *Tk+1*(*x*) = *DO−1W Tk*(*x*) и *DO−1W* разрежено, легко увидеть, что уравнение 4 может быть вычислено с помощью  *O*(*K*) рекурсивного умножения разреженных плотных матриц с временной сложностью *O*(|E|). Следовательно, временные сложности уравнения 2 и уравнения 4 равны *O*(*K|E|)*. Для плотного графа мы можем использовать спектральную разреженность (Cheng et al., 2015), чтобы сделать его разреженным.

## C СВЯЗЬ СО СВЕРТКОЙ СПЕКТРАЛЬНОГО ГРАФА

*Доказательство.* Свертка спектрального графа использует концепцию нормализованного графа Лапласа *L* =

*D*|. ChebNet параметризует *fθ* как многочлен K порядка **Λ** и вычисляет его с использованием стабильного полинома Чебышева.

*X*: (5)

где *T0*(*x*) = *1,T*1(*x*) = *x,Tk*(*x*) = *xTk−1*(*x*) − *Tk−2*(*x*) являются базисом полинома Чейшева. Пусть *λmax* обозначает наибольшее собственное значение *L**I*, представляющее собой перемасштабирование графа Лапласа, отображающего собственные значения от [*0,λmax*] до [−*1,1*], поскольку полином Чебышева образует ортогональный базис в [−*1,1*]. Уравнение 5 можно рассматривать как многочлен *L ̃*, и мы покажем, что выход свертки ChebNet *аналогичен* выходу диффузионной свертки с точностью до постоянного коэффициента масштабирования. Предположим*, что λmax* = 2 и *DI* = *DO* = *D* для неориентированного графа.

*Л ̃* *WD* *D−1Вт* (6)

*L ̃*  *аналогичен* отрицательной матрице перехода случайного блуждания, таким образом, выход уравнения 5 также аналогичен выходу уравнения 2 с точностью до постоянного коэффициента масштабирования.

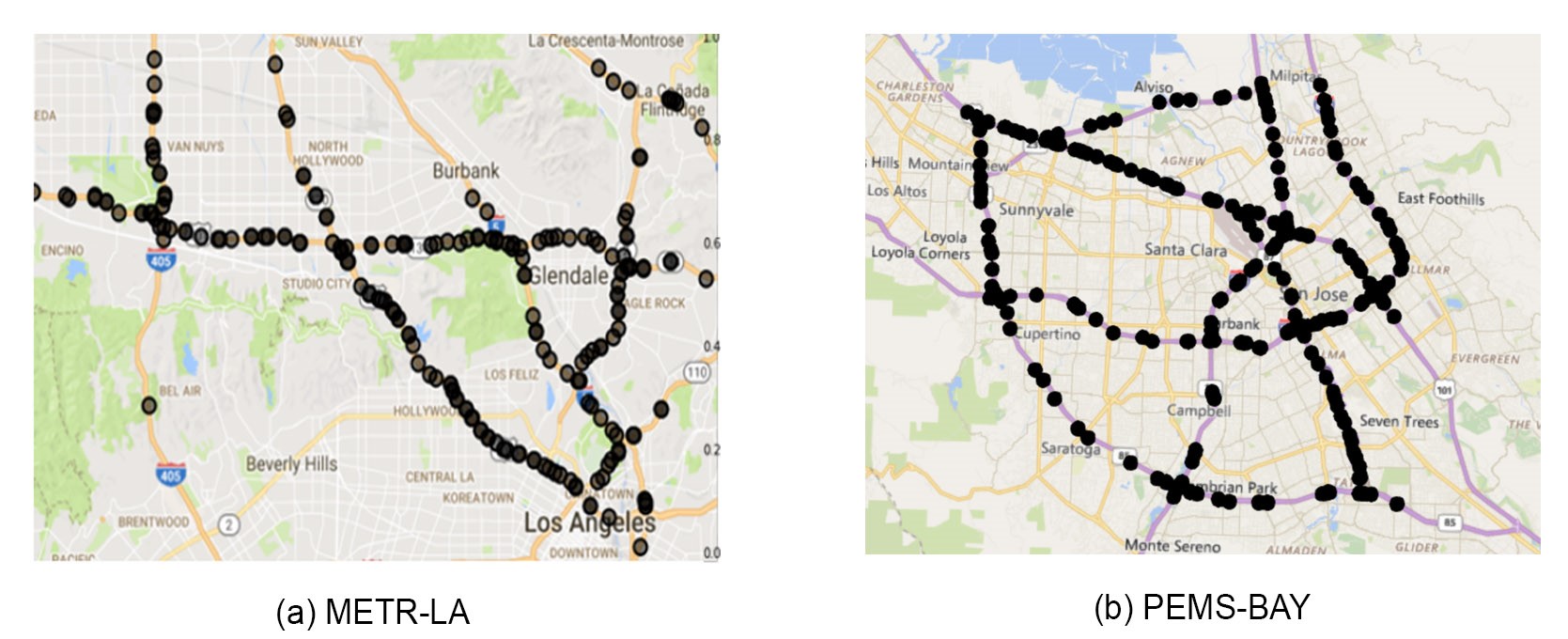


Рисунок 8: Распределение датчиков наборов данных METR-LA и PEMS-BAY.

## D БОЛЬШЕ СВЯЗАННОЙ РАБОТЫ И ОБСУЖДЕНИЯ

Xie et al. (2010) вводят метод, основанный на гауссовских процессах (ГП). GP трудно масштабировать до большого набора данных и, как правило, они не подходят для относительно долгосрочного прогнозирования трафика, например, на 1 час (т.е. на 12 шагов вперед), поскольку дисперсия может накапливаться и становится чрезвычайно большой.

Cai et al. (2016) предлагают использовать пространственно-временной ближайший сосед для прогнозирования трафика (ST-KNN). Несмотря на то, что ST-KNN учитывает как пространственные, так и временные зависимости, у него есть следующие недостатки. Как показано в работе Fusco et al. (2016), ST-KNN выполняет независимое прогнозирование для каждой отдельной дороги. Прогнозирование дороги представляет собой взвешенную комбинацию ее собственных исторических скоростей движения. Это затрудняет для ST-KNN полное использование информации от соседей. Кроме того, ST-KNN является непараметрическим подходом, и каждая дорога моделируется и рассчитывается отдельно (Cai et al., 2016), что затрудняет обобщение на невидимые ситуации и масштабирование на большие наборы данных. Наконец, в ST-KNN все сходства рассчитываются с помощью разработанных вручную метрик с небольшим количеством изучаемых параметров, и это может ограничить его репрезентативную силу.

Cheng et al. (2017) предлагают DeepTransport, который моделирует пространственную зависимость путем явного сбора определенного количества дорог вверх и вниз по течению для каждой отдельной дороги, а затем проводит свертку по этим дорогам соответственно. По сравнению с Cheng et al. (2017), DCRNN моделирует пространственную зависимость более систематическим образом, т.е. обобщая свертку на графе датчика трафика на основе диффузионного характера трафика. Кроме того, мы выводим DCRNN из свойства случайного блуждания и показываем, что популярная спектральная свертка ChebNet является частным случаем нашего метода.

Предлагаемый подход также связан с методами встраивания графов, например, Deepwalk (Perozzi et al., 2014), node2vec (Grover & Leskovec, 2016), которые изучают представление низкой размерности для каждого узла графа. DCRNN также изучает представление для каждого узла. Изученные представления учитывают как пространственную, так и временную зависимость и в то же время оптимизируются с учетом цели, например, будущих скоростей движения.

## E ДЕТАЛЬНО ПРОРАБОТАННЫЕ ЭКСПЕРИМЕНТАЛЬНЫЕ НАСТРОЙКИ

HA Historical Average, который моделирует транспортный поток как сезонный процесс и использует средневзвешенное значение предыдущих сезонов в качестве прогноза. Используемый период составляет 1 неделю, а прогноз основан на агрегированных данных за предыдущие недели. Например, прогноз на эту среду представляет собой усредненную скорость трафика за последние четыре среды. Поскольку метод исторического среднего не зависит от краткосрочных данных, его эффективность инвариантна к небольшому увеличению горизонта прогнозирования

АРИМАкал: Авторегрессионная интегрированная модель скользящего среднего с фильтром Калмана. Это ордера (3, 0, 1), а модель реализуется с помощью метода *Модель состояния* пакет Python.

Векторная авторегрессионная модель VAR (Hamilton, 1994). Количество лагов устанавливается равным 3, а модель реализуется с помощью  *пакета Python statsmodel*.

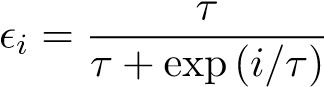
Линейная опорная векторная регрессия СВР, штрафной член *C* = 0*.*1, количество исторических наблюдений равно 5.

Также включены следующие подходы, основанные на глубоких нейронных сетях.

FNN Feedforward нейронная сеть с двумя скрытыми слоями, каждый слой содержит 256 единиц. Начальная скорость обучения равна 1e−3 и уменьшается до  каждых 20 эпох, начиная с 50-х эпох. Кроме того, для всех скрытых слоев используется дропаут с соотношением 0,5 и распад веса L2 1e−2. Модель обучается с размером пакета 64 и MAE в качестве функции потерь. Досрочная остановка выполняется путем мониторинга ошибки валидации.

ФК-ЛСТМ Фреймворк Encoder-Decoder с использованием LSTM с глазком (Sutskever et al., 2014). Как энкодер, так и декодер содержат два повторяющихся слоя. В каждом рекуррентном слое находится 256 единиц LSTM, распад веса L1 равен 2э−5, L2 снижение веса 5э−4. Модель обучается с размером пакета 64 и функцией потерь MAE. Начальный коэффициент обучения равен 1e-4 и снижается до  каждые 10 эпох, начиная с 20-й эпохи. Досрочная остановка выполняется путем мониторинга ошибки валидации.

DCRNN : диффузионная сверточная рекуррентная нейронная сеть. И энкодер, и декодер содержат два повторяющихся слоя. В каждом рекуррентном слое есть 64 единицы, начальная скорость обучения равна 1э−2, и сводится к  Используются каждые 10 эпох, начиная с 20-й эпохи и досрочно останавливаясь на валидационном наборе данных. Кроме того, максимальные шаги случайных блужданий, т.е. *K*, имеет значение 3. Для запланированной выборки в качестве снижения вероятности используется пороговая обратная сигмоидальная функция:



где *i* — количество итераций, а *τ* — параметры для управления скоростью сходимости. *В* экспериментах τ установлено равным 3 000. Реализация доступна в [https://github.com/ liyaguang/DCRNN.](https://github.com/liyaguang/DCRNN)

## Е.1 НАБОР ДАННЫХ

Мы проводим эксперименты на двух реальных крупномасштабных наборах данных:

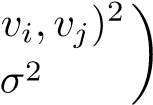
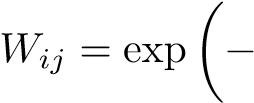
* METR-LA Этот набор данных о дорожном движении содержит информацию о дорожном движении, собранную с помощью петлевых детекторов на шоссе округа Лос-Анджелес (Jagadish et al., 2014). Мы выбрали 207 датчиков и собрали данные за 4 месяца с 1 марта 2012 года по 30 июня 2012 года для эксперимента. Общее количество наблюдаемых точек данных о трафике составляет 6 519 002.
* PEMS-BAY Этот набор данных о дорожном движении собран Калифорнийским транспортным агентством (CalTrans) Performance Measurement System (PeMS). Мы выбрали 325 датчиков в районе залива Сан-Франциско и собрали данные за 6 месяцев с 1 января 2017 года по 31 мая 2017 года для эксперимента. Общее количество наблюдаемых точек данных о трафике составляет 16 937 179.

Распределения датчиков обоих наборов данных визуализированы на рисунке 8.

В обоих этих наборах данных мы агрегируем показания скорости движения в 5-минутные окна и применяем

Нормализация Z-Score. 70% данных используется для обучения, 20% — для тестирования, а оставшаяся часть

10% за валидацию. Чтобы построить граф датчиков, мы вычисляем расстояния попарной дорожной сети между датчиками и строим матрицу смежности с использованием порогового гауссова ядра (Shuman et al., 2013).

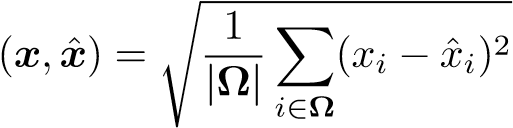
dist( если dist(*vi,vj*) ≤ *K*иначе 0

где *Wij* представляет собой вес ребер между датчиком *vi* и датчиком *vj*, dist(*vi,vj*) обозначает расстояние по дорожной сети от датчика *vi* до датчика *vj*. *σ* — стандартное отклонение расстояний, *а κ* — порог.

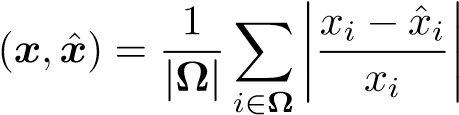
## Е.2 МЕТРИКА

Предположим*, x* = *x1,*··· *,xn* представляет основную истину, *xˆ = xˆ1,*··· *,xˆn* представляет прогнозируемые значения, а Ω обозначает индексы наблюдаемых выборок, метрики определяются следующим образом.

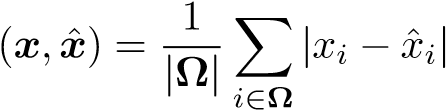
Среднеквадратичная ошибка (RMSE)

RMSE

Средняя абсолютная процентная ошибка (MAPE)

МАПЕ

Средняя абсолютная погрешность (MAE)

МАЕ

## F ВИЗУАЛИЗАЦИЯ МОДЕЛИ

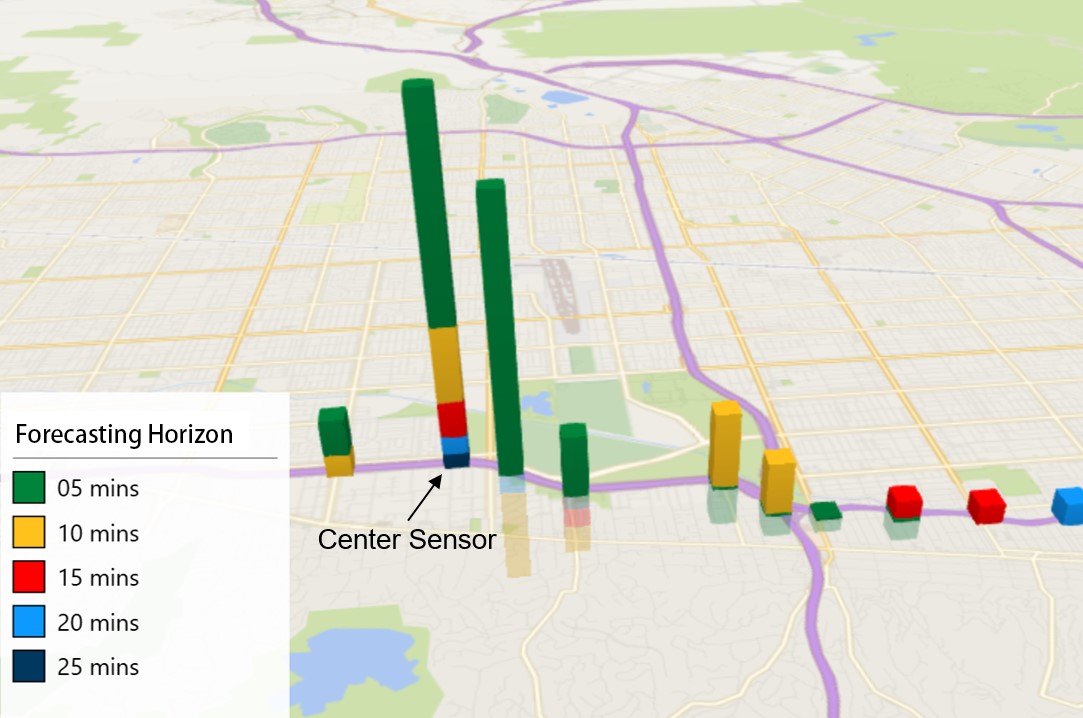
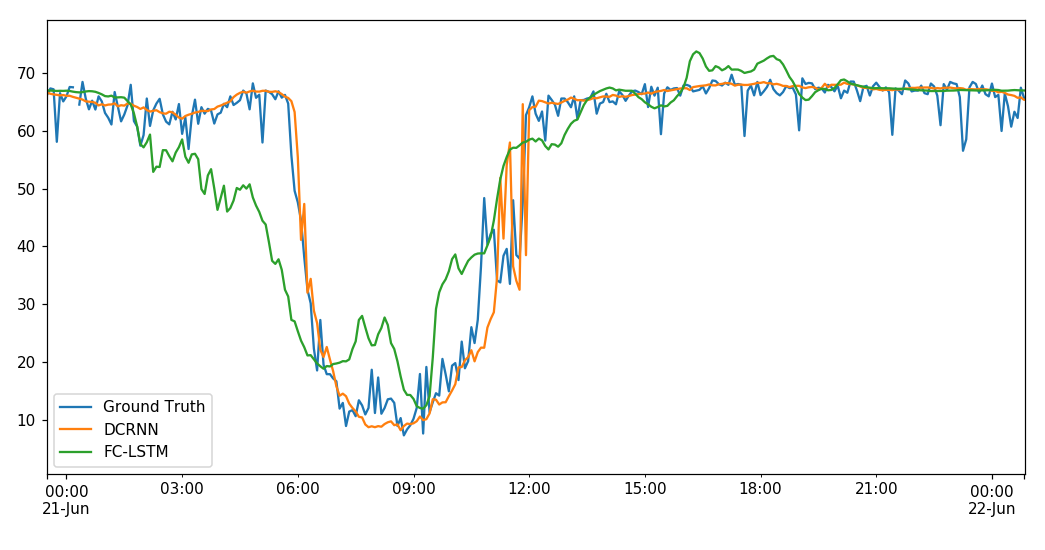
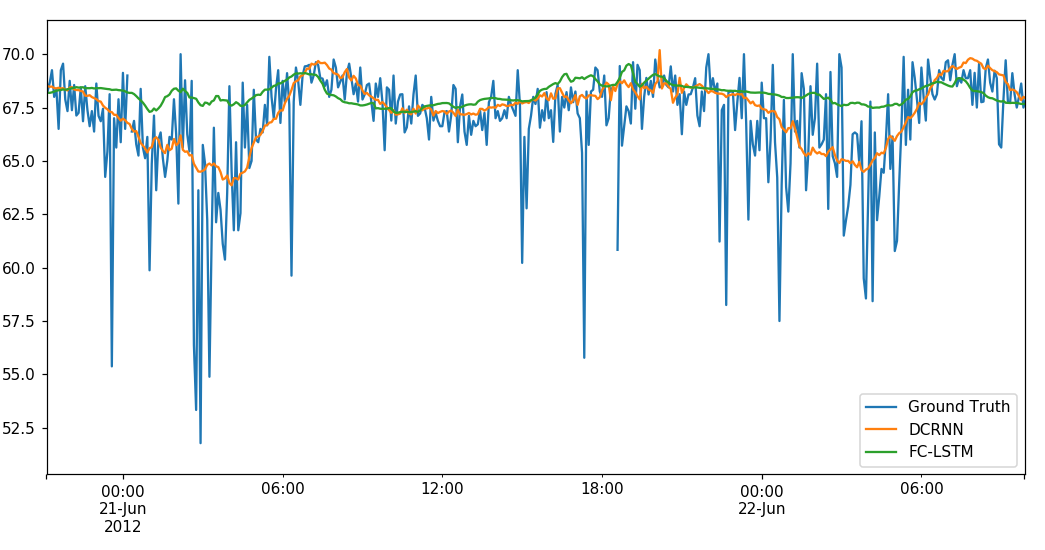
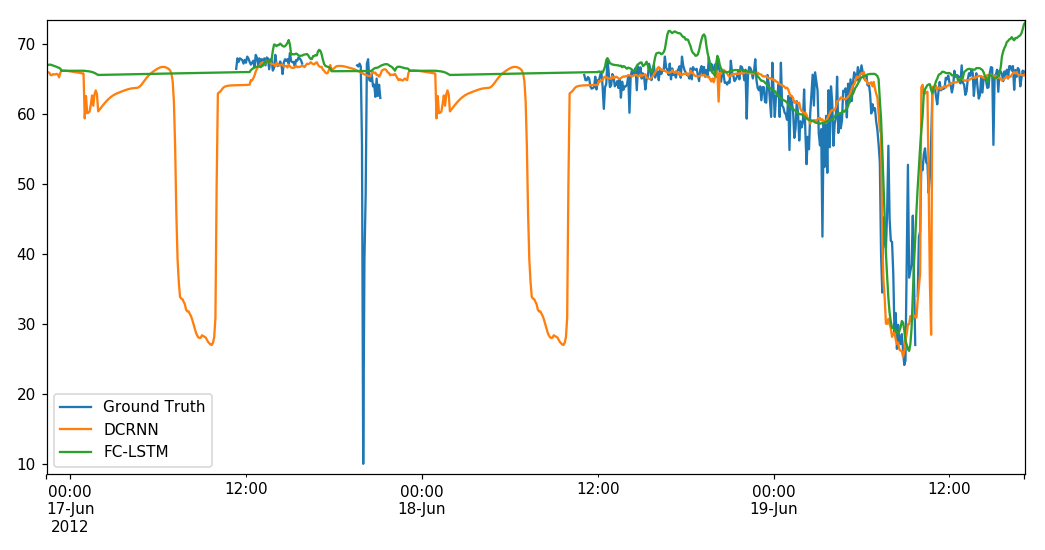
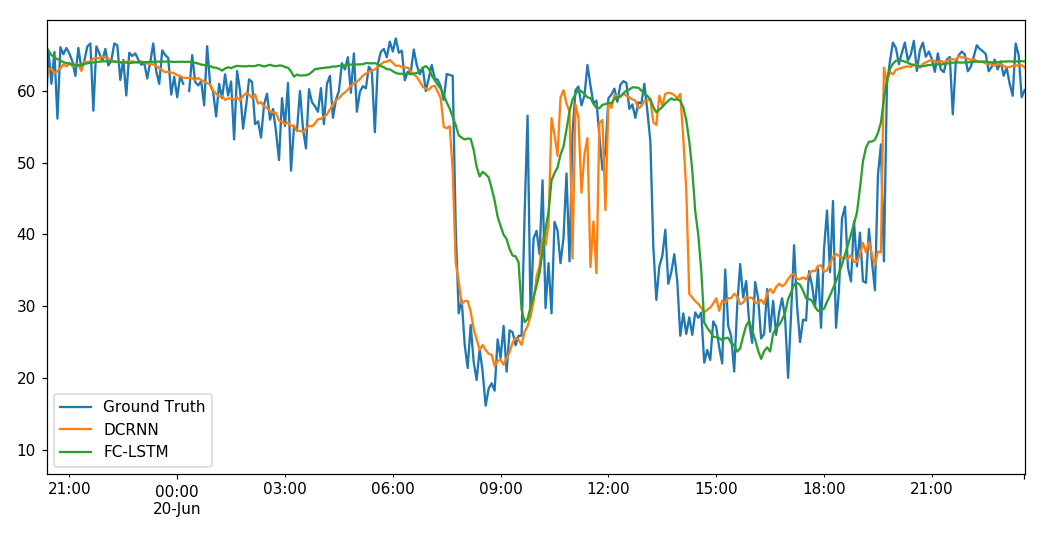
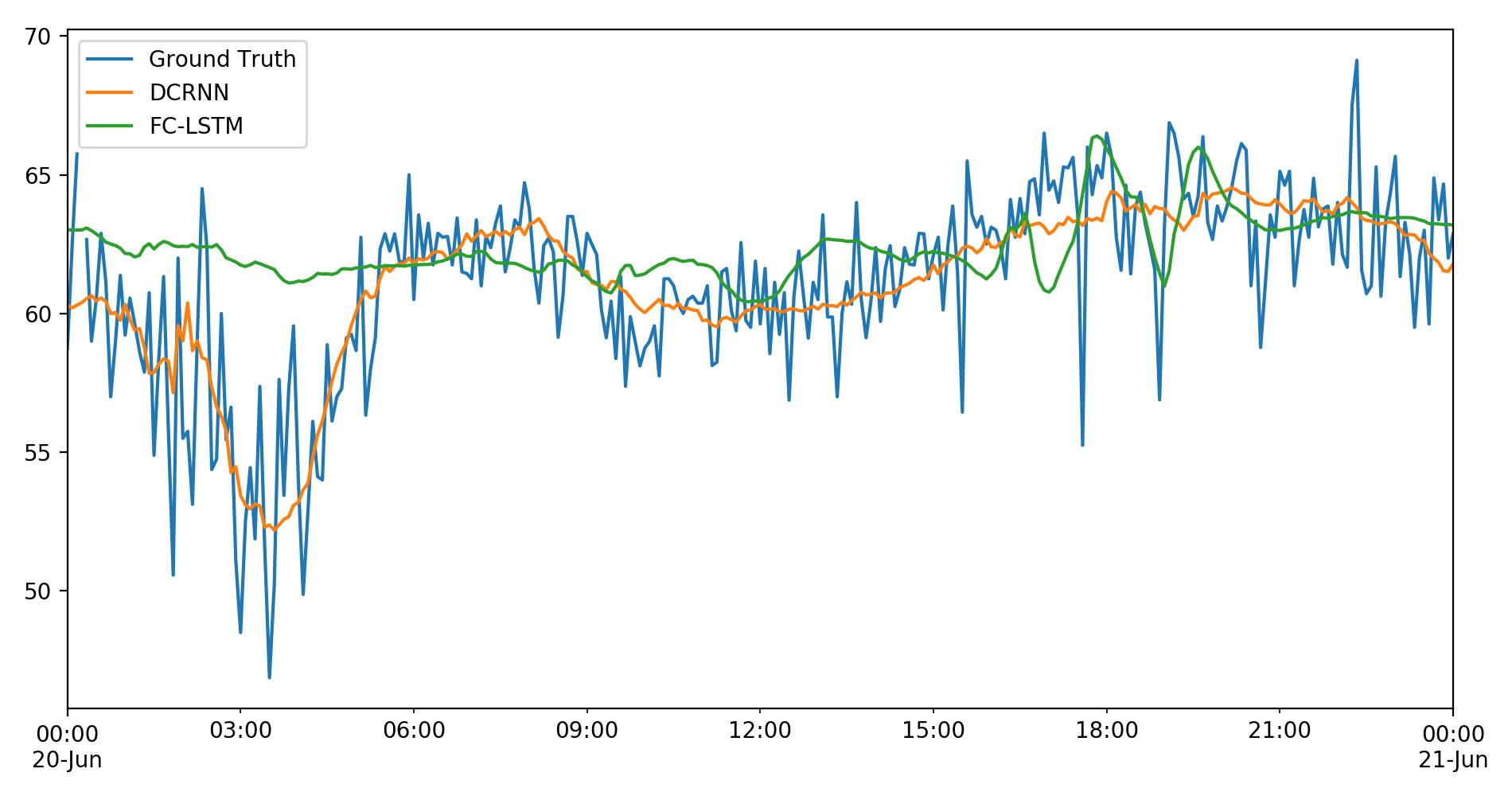
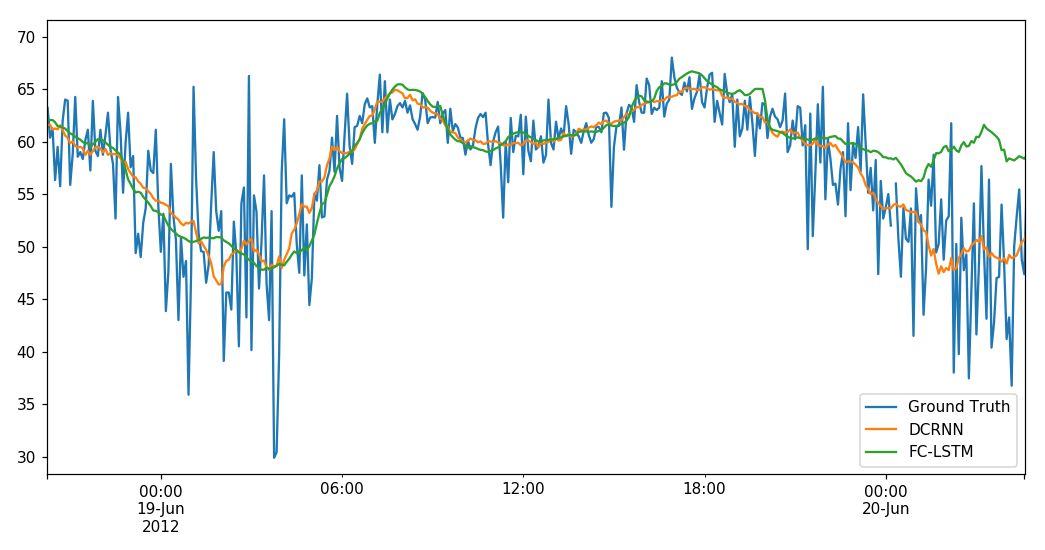
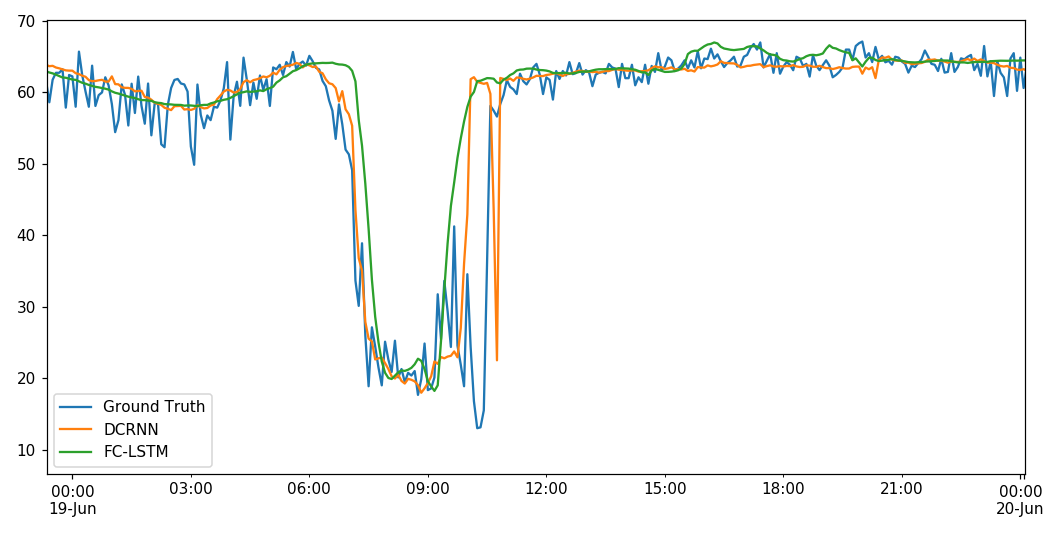
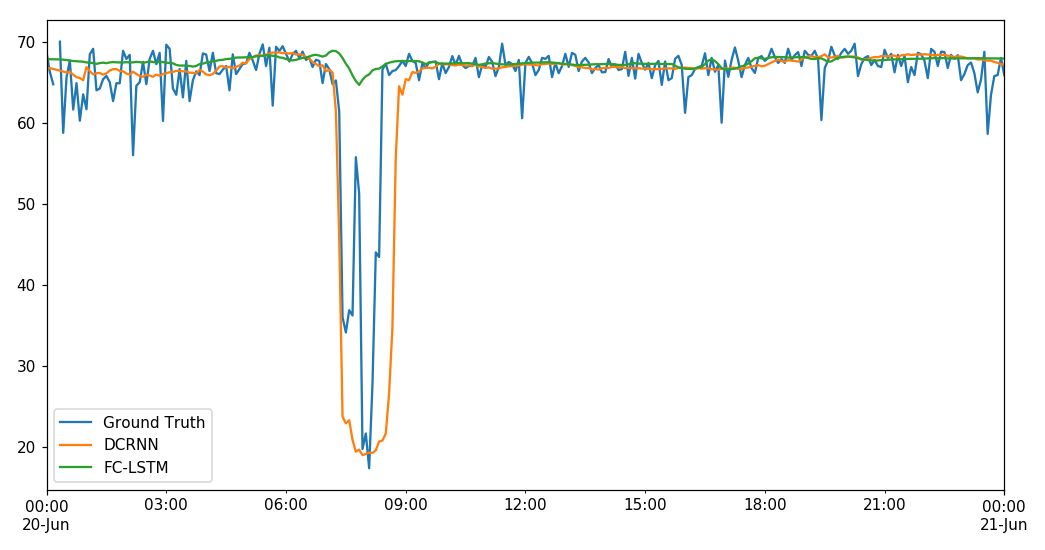


Рисунок 9: Корреляции датчиков между центральным сенсором и его окрестностями для различных горизонтов прогнозирования. Корреляции оцениваются с помощью регуляризованной VAR. Мы видим, что корреляции локализованы и более близкие окрестности обычно имеют большую релевантность, а величина корреляции быстро уменьшается с увеличением расстояния, что согласуется с процессом диффузии на графике.



15

Рисунок 10: Визуализация прогнозирования временных рядов трафика.



16

Рисунок 11: Визуализация прогнозирования временных рядов трафика.

1. Исходный код доступен по адресу [https://github.com/liyaguang/DCRNN.](https://github.com/liyaguang/DCRNN) [↑](#footnote-ref-1)