Сверточные сети пространственно-временного графа: фреймворк глубокого обучения для прогнозирования трафика

Бин Юй ∗1, Хаотэн Инь∗*2,3*, Чжаньсин Чжу †*3,4*1 Школа математических наук, Пекинский университет, Пекин, Китай 2 Академия перспективных междисциплинарных исследований, Пекинский университет, Пекин, Китай 3 Центр науки о данных, Пекинский университет, Пекин, Китай 4

Пекинский институт исследований больших данных (BIBDR), Пекин, Китай

{byu, htyin, zhanxing.zhu}@pku.edu.cn

Абстрактный

Своевременный точный прогноз дорожного движения имеет решающее значение для управления и управления городским движением. Из-за высокой нелинейности и сложности транспортного потока традиционные методы не могут удовлетворить требования задач прогнозирования в среднесрочной и долгосрочной перспективе и часто пренебрегают пространственными и временными зависимостями. В этой статье мы предлагаем новую среду глубокого обучения, Spatio-Temporal Graph Convolutional Networks (STGCN), для решения проблемы прогнозирования временных рядов в области трафика. Вместо того, чтобы применять регулярные сверточные и рекуррентные единицы, мы формулируем задачу на графах и строим модель с полными сверточными структурами, которые обеспечивают гораздо более высокую скорость обучения с меньшим количеством параметров. Эксперименты показывают, что наша модель STGCN эффективно фиксирует всесторонние пространственно-временные корреляции за счет моделирования многомасштабных сетей трафика и постоянно превосходит современные базовые показатели на различных реальных наборах данных о дорожном движении.

# Знакомство

Транспорт играет жизненно важную роль в повседневной жизни каждого человека. Согласно опросу, проведенному в 2015 году, водители в США проводят за рулем в среднем около 48 минут в день.1 В этих условиях точный прогноз дорожной обстановки в режиме реального времени имеет первостепенное значение для участников дорожного движения, частного сектора и правительств. Широко используемые транспортные услуги, такие как управление потоком, планирование маршрутов и навигация, также в значительной степени зависят от высококачественной оценки состояния дорожного движения. В целом, многомасштабное прогнозирование дорожного движения является предпосылкой и основой управления и управления городским движением, что также является одной из основных функций интеллектуальной транспортной системы (ИТС).

arXiv:1709.04875v4 [cs. LG] 12 июл 2018

При исследовании трафика фундаментальные переменные транспортного потока, а именно скорость, объем и плотность, обычно выбираются в качестве индикаторов для мониторинга текущего состояния дорожных условий.

|  |
| --- |
| ∗  Равные взносы.  †  Автор, отвечающий за переписку.  1 https://aaafoundation.org/american-driving-survey-2014-2015/ |

для прогнозирования будущего. В зависимости от длины прогноза прогноз трафика обычно классифицируется на две шкалы: краткосрочный (5 ∼ 30 минут), среднесрочный и долгосрочный (более 30 минут). Наиболее распространенные статистические подходы (например, линейная регрессия) могут хорошо работать при прогнозировании коротких интервалов. Однако из-за неопределенности и сложности транспортного потока эти методы менее эффективны для относительно долгосрочных прогнозов.

Предыдущие исследования по прогнозированию трафика в среднесрочной и долгосрочной перспективе можно условно разделить на две категории: динамическое моделирование и методы, основанные на данных. Динамическое моделирование использует математические инструменты (например, дифференциальные уравнения) и физические знания для формулировки транспортных проблем с помощью вычислительного моделирования [Vlahogianni, 2015]. Для достижения устойчивого состояния процесс моделирования не только требует сложного систематического программирования, но и потребляет огромные вычислительные мощности. Непрактичные предположения и упрощения при моделировании также снижают точность прогнозирования. Таким образом, в связи с быстрым развитием методов сбора и хранения данных о дорожном движении, большая группа исследователей переключает свое внимание на подходы, основанные на данных.

Классические статистические модели и модели машинного обучения являются двумя основными представителями методов, управляемых данными. В анализе временных рядов авторегрессионное интегрированное скользящее среднее (ARIMA) и его варианты являются одним из наиболее консолидированных подходов, основанных на классической статистике [Ahmed and Cook, 1979; Williams and Hoel, 2003]. Однако этот тип модели ограничен стационарным предположением о временных последовательностях и не учитывает пространственно-временную корреляцию. Таким образом, эти подходы имеют ограниченную репрезентативность сильно нелинейного транспортного потока. В последнее время классические статистические модели были решительно оспорены методами машинного обучения в задачах прогнозирования трафика. С помощью этих моделей можно достичь более высокой точности прогнозирования и более сложного моделирования данных, таких как *алгоритм k-ближайших соседей (KNN), метод опорных векторов (SVM) и нейронные сети (NN).*

В настоящее время подходы глубокого обучения широко и успешно применяются для решения различных задач дорожного движения. Значительный прогресс был достигнут в смежных работах, например, в сети глубоких убеждений (DBN) [Jia *et al.*, 2016; Huang *et al.*, 2014], многоуровневый автоэнкодер (SAE) [Lv *et al.*, 2015; Chen *et al.*, 2016]. Однако для этих плотных сетей трудно извлекать пространственные и временные особенности из входных данных совместно. Более того, в условиях узких ограничений или даже полного отсутствия пространственных атрибутов репрезентативная способность этих сетей будет серьезно затруднена.

Чтобы в полной мере использовать пространственные особенности, некоторые исследователи используют сверточную нейронную сеть (CNN) для захвата смежных отношений между сетью трафика, а также используют рекуррентную нейронную сеть (RNN) на оси времени. Объединив сеть долгосрочной кратковременной памяти (LSTM) [Hochreiter and Schmidhuber, 1997] и 1-D CNN, Ву и Тан [2016] представили объединенную архитектуру CLTFP на уровне функций для краткосрочного прогнозирования трафика. Несмотря на то, что CLTFP придерживалась простой стратегии, она все же предприняла первую попытку выровнять пространственные и временные закономерности. Впоследствии Ши *и др.* [2015] предложил сверточный LSTM, который представляет собой расширенный полносвязный LSTM (FC-LSTM) со встроенными сверточными слоями. Однако применяемая обычная сверточная операция ограничивает модель только обработкой сетчатых структур (например, изображений, видео), а не общих областей. Между тем, рекуррентные сети для обучения последовательностям требуют итерационного обучения, которое вводит пошаговое накопление ошибок. Кроме того, широко известно, что сети на основе RNN (включая LSTM) сложны в обучении и требуют больших вычислительных ресурсов.

Для решения этих проблем мы вводим несколько стратегий для эффективного моделирования временной динамики и пространственных зависимостей транспортного потока. Чтобы в полной мере использовать пространственную информацию, мы моделируем транспортную сеть с помощью общего графа, а не рассматриваем ее отдельно (например, в виде сеток или сегментов). Чтобы справиться с недостатками, присущими рекуррентным сетям, мы используем полностью сверточную структуру на оси времени. Прежде всего, мы предлагаем новую архитектуру глубокого обучения, сверточные сети пространственно-временного графа, для задач прогнозирования трафика. Эта архитектура включает в себя несколько пространственно-временных сверточных блоков, которые представляют собой комбинацию графовых сверточных слоев [Defferrard *et al.*, 2016] и сверточных слоев обучения последовательностей, для моделирования пространственных и временных зависимостей. Насколько нам известно, это первый случай, когда чисто сверточные структуры были применены для одновременного извлечения пространственно-временных особенностей из графически структурированных временных рядов в исследовании трафика. Мы оцениваем предложенную модель на двух реальных наборах данных о дорожном движении. Эксперименты показывают, что наша платформа превосходит существующие базовые показатели в задачах прогнозирования с несколькими предустановленными длинами прогнозирования и масштабами сети.

# Предварительный

## Прогнозирование трафика на графиках дорог

Прогнозирование трафика является типичной задачей прогнозирования временных рядов, т.е. прогнозированием наиболее вероятных измерений трафика (например, скорости или транспортного потока) на следующих *временных* шагах H с учетом предыдущих *M* наблюдений за трафиком в следующем виде:

*vˆt+1,...,vˆt+H* = argmax log *P*(*vt+1,...,vt+H|vt−M+1,...,vt*)*,* (1)

*vt+1,...,vt+H*

где *vt* ∈ Rn – вектор наблюдения *за n* сегментами дороги на временном шаге *t*, каждый элемент которого записывает исторические наблюдения для одного участка дороги.

В этой работе мы определяем сеть трафика на графе и фокусируемся на структурированных временных рядах трафика. Наблюдение *vt* равно

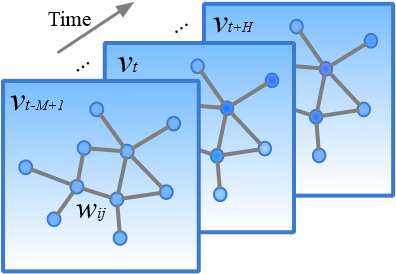


Рисунок 1: Данные о трафике с графической структурой. Каждый *vt* указывает кадр текущего состояния трафика на временном шаге *t*, который записывается в графическую матрицу данных.

не независимы, но связаны попарной связью в графе. Таким образом, точку данных *vt* можно рассматривать как сигнал графа, который определен на неориентированном (или направленном) графе G с весами *wij*, как показано на рисунке 1. На  *t-м временном шаге, в графе* Gt = (*Vt,E,W*), Vt — конечное множество вершин, соответствующее наблюдениям с *n* станций мониторинга в сети дорожного движения; E — набор ребер, указывающих на связанность между станциями, в то время как *W* ∈ Rn×n обозначает взвешенную матрицу смежности Gt.

## Извилины на графиках

Стандартная свертка для регулярных сеток явно неприменима к общим графикам. В настоящее время существует два основных подхода, изучающих способ обобщения СНС на структурированные формы данных. Один из них состоит в том, чтобы расширить пространственное определение свёртки [Niepert *et al.*, 2016], а другой — манипулирование в спектральной области с помощью графовых преобразований Фурье [Bruna *et al.*, 2013]. Первый подход перестраивает вершины в определенные формы сетки, которые могут быть обработаны обычными сверточными операциями. Последний представляет собой спектральную структуру для применения сверток в спектральных областях, часто называемую сверткой спектрального графа. Несколько последующих исследований делают свертку графов более перспективной за счет снижения вычислительной сложности с O(*n2*) до линейной [Defferrard *et al.*, 2016; Kipf and Welling, 2016]. Вводится понятие оператора свертки графа "∗G", основанного на концепции спектральной свертки графа, как умножения сигнала *x* ∈ Rn на ядро Θ,

Θ ∗G *x* = Θ(*L*)*x* = Θ(*UΛUT*)*x* = *UΘ*(Λ)*UTx,* (2) где граф Базис Фурье *U* ∈ Rn×n — матрица собственных векторов нормализованного графа Лапласиана *L* = *In* −  — тождественная матрица, *D* ∈ Rn×n — диагональная матрица степеней с *Dii* = σjwij); Λ ∈ Rn×n — диагональная матрица собственных значений *L*, а фильтр Θ(Λ) — также диагональная матрица. Согласно этому определению, графовый сигнал *x* фильтруется ядром Θ с умножением между Θ и графом преобразования Фурье *UTx* [Shuman *et al.*, 2013].

# Предлагаемая модель

## Сетевая архитектура

В этом разделе мы подробно остановимся на предлагаемой архитектуре пространственно-временных графовых сверточных сетей (STGCN). Как показано на рисунке 2, STGCN состоит из нескольких пространственно-временных сверточных блоков, каждый из которых образован в виде «сэндвича» структуры с двумя стробируемыми последовательными слоями свертки и одним слоем свертки пространственного графа между ними. Подробная информация о каждом модуле описана ниже.

Блок ST-Conv

Блок ST-Conv

Выходной слой

(

*v*

*t-*

*М+1*

*, …*

*v*

*t*

)

*W*

Пространственный

Граф-Конв,

*C*

=16

Временный

Закрытый конв,

*C*

=64

Временный

Закрытый конв,

*C*

=64

GLU

1-

D

Конв

*W*

ˆ

*v*

(

*v*

*t-*

*М+1*

*, …*

*v*

*t*

)

*l*

*l*

Временный

Закрытый конв

Блок ST-Conv

(

*v*

*t-*

*М+К*

*, …*

*v*

*t*

)

*l*

*l*

*t*

*v*

*l*

*v*

*Л+1*

Рисунок 2: Архитектура сверточных сетей пространственно-временного графа. Фреймворк STGCN состоит из двух пространственно-временных сверточных блоков (блоков ST-Conv) и полностью подключенного выходного слоя в конце. Каждый блок ST-Conv содержит два слоя темпоральной стробированной свертки и один слой свертки пространственного графа посередине. Остаточное соединение и стратегия узких мест применяются внутри каждого блока. Входной *сигнал vt−M+1,...,vt* равномерно обрабатывается блоками ST-Conv для согласованного исследования пространственных и временных зависимостей. Комплексные функции интегрируются выходным слоем для генерации окончательного прогноза *vˆ*.

## Графические СНС для извлечения пространственных объектов

Сеть трафика обычно организуется в виде графовой структуры. Естественно и целесообразно математически формулировать дорожные сети в виде графов. Однако в предыдущих исследованиях пренебрегают пространственными атрибутами транспортных сетей: связность и глобальность сетей упускаются из виду, поскольку они разделены на несколько сегментов или сеток. Даже с двумерными свертками на сетках он может лишь приблизительно зафиксировать пространственную локальность из-за компромиссов в моделировании данных. Соответственно, в нашей модели свертка графа применяется непосредственно к графовым структурированным данным для извлечения высокозначимых шаблонов и признаков в космической области. Несмотря на то, что вычисление ядра Θ в свертке графа по уравнению (2) может быть дорогостоящим из-за умножения O(*n2*) с базисом Фурье графа, для решения этой проблемы применяются две стратегии аппроксимации.

Аппроксимация полиномов Чебышева Для локализации фильтра и уменьшения числа параметров ядро Θ может быть ограничено полиномом , где *θ* ∈ RK — вектор коэффициентов полинома. *K* — размер ядра свертки графа, который определяет максимальный радиус свертки от центральных узлов. Традиционно, полином Чебышева *Tk*(*x*) используется для аппроксимации ядер в виде усеченного разложения порядка *K−1* как  при перемасштабированном Λ = 2Λ ̃ */λmax* − *In* (*λmax* обозначает наибольшее собственное значение *L*) [Hammond *et al.*,

2011]. Свертка графа может быть затем переписана следующим образом:

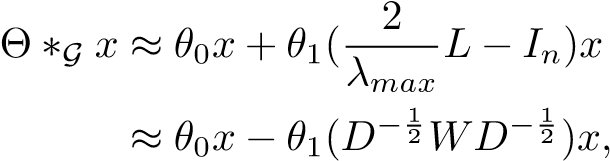
*К−1*

И ∗Г *x* = Θ(*L*)*x* ≈ X *θkTk*(*Л ̃*)*x,* (3)

*к=0*

где *Tk*(*L ̃*) ∈ Rn×n — многочлен Чебышева порядка *k,* оцененный в масштабном лапласиане *L ̃* = 2L/*λmax* − *In*. Путем рекурсивного вычисления K-локализованных сверток через полиномиальную аппроксимацию, стоимость уравнения (2) может быть уменьшена до O(*K|E|)* как показывает уравнение (3) [Defferrard *et al.*, 2016].

Послойная линейная формулировка может быть определена путем наложения нескольких локализованных сверточных слоев графа с аппроксимацией первого порядка графа Лапласа [Kipf and Welling, 2016]. Следовательно, может быть построена более глубокая архитектура для глубокого восстановления пространственной информации, не ограничиваясь явной параметризацией, задаваемой полиномами. За счет масштабирования и нормализации в нейронных сетях в дальнейшем можно предположить, что *λmax* ≈ 2. Таким образом, уравнение (3) может быть упрощено до:

 (4)

где *θ0*, *θ1* — два общих параметра ядра. Чтобы ограничить параметры и стабилизировать численные характеристики, *θ0* и *θ1* заменяются одним параметром *θ*, позволяя *θ* = *θ0* = −θ1; *W* и *D* перенормируются как *W ̃* = *W* + *In* и *D ̃ii* = ΣjW ̃ij отдельно. Тогда свертка графа может быть альтернативно выражена как:



(5)



Применение стека графовых сверток с аппроксимацией 1-го порядка по вертикали, что позволяет достичь того же эффекта, что  *и локализованные свертки по горизонтали, все из которых используют информацию из окрестности* (*K-1)*-порядка центральных узлов. В этом сценарии *K* — это количество последовательных операций фильтрации или сверточных слоев в модели. Кроме того, послойная линейная структура является параметрически экономичной и высокоэффективной для крупномасштабных графов, поскольку порядок аппроксимации ограничен единицей.

Обобщение графовых сверток Оператор свертки графа "∗G", определенный на *x* ∈ Rn, может быть расширен до многомерных тензоров. Для сигнала с *каналами*  *Ci X* ∈ Rn×Ci свертка графа может быть обобщена следующим образом:

*Ке*

*yj* = XΘi,j(*L*)*кси* ∈ Рн*,*1 ≤ *j* ≤ *Со*(6)

*i=1*

с векторами *Ci* × *Co* коэффициентов Чебышева Θi,j ∈ RK (*Ci*, *Co* — размер входных и выходных карт признаков соответственно). Свертка графика для двумерных переменных обозначается как "Θ ∗G *X*" с Θ ∈ RK×Ci×Co. В частности, входные данные прогноза трафика состоят из *M* кадра дорожных графов, как показано на рисунке 1. Каждый кадр *vt* можно рассматривать как матрицу, столбец *которой* i является *Ci-мерным значением* vt *в i-м* узле графа Gt, как *X* ∈ Rn×Ci (в этом случае *Ci* = 1). Для каждого временного шага *t*  *из M* операция свертки равного графа с одним и тем же ядром Θ накладывается на *Xt* ∈ Rn×Ci параллельно. Таким образом, свертка графа может быть дополнительно обобщена в трехмерных переменных, обозначенных как "Θ ∗G X" с X ∈ RM×n×Ci.

## Закрытые СНС для извлечения временных признаков

Несмотря на то, что модели на основе RNN получили широкое распространение в анализе временных рядов, рекуррентные сети для прогнозирования трафика по-прежнему страдают от трудоемких итераций, сложных механизмов шлюзов и медленной реакции на динамические изменения. Напротив, СНС имеют преимущество в быстром обучении, простых структурах и отсутствии ограничений зависимости от предыдущих шагов. Вдохновленный [Gehring *et al.*, 2017], мы используем целые сверточные структуры на оси времени для фиксации временного динамического поведения транспортных потоков. Такая специфическая конструкция позволяет проводить параллельные и управляемые процедуры обучения с помощью многослойных сверточных структур, сформированных в виде иерархических представлений.

Как показано на рисунке 2 (справа), временной сверточный слой содержит одномерную причинную свертку с ядром шириной *Kt,* за которым следуют стробированные линейные единицы (GLU) в качестве нелинейности. Для каждого узла в графе G временная свертка исследует *Kt* соседей входных элементов без заполнения, что приводит к сокращению длины последовательностей на Kt-1 каждый раз. Таким образом, вход временной свертки для каждого узла можно рассматривать как последовательность длины-Mс каналами Ci как *Y* ∈ RM×Ci. Ядро свертки Γ ∈ RKt×Ci×2Co предназначено для отображения входного *Y* на один выходной элемент [*P Q*] ∈ R(*M−Kt+1)×(2Co*) (*P*, *Q* разделяется пополам при одинаковом размере каналов). В результате, темпоральная стробируемая свертка может быть определена как:

*,* (7)

где *P*, *Q* - вход вентилей в GLU соответственно; обозначает произведение Адамара по элементам. Сигмовидный вентиль *σ*(*Q*) определяет, какие входные *P* текущих состояний имеют значение для обнаружения композиционной структуры и динамических дисперсий во временных рядах. Нелинейные вентили также способствуют использованию полного ввода, подаваемого через стековые временные слои. Кроме того, между многослойными временными сверточными слоями реализуются остаточные связи. Аналогичным образом, временная свертка может быть также обобщена на трехмерные переменные, если использовать одно и то же ядро свертки Γ для каждого узла Yi ∈ RM×Ci (например, сенсорных станций) в G в равной степени, отмеченных как «Γ ∗T Y» с Y ∈ RM×n×Ci.

## Пространственно-височный сверточный блок

Чтобы объединить объекты как из пространственной, так и из временной областей, строится пространственно-временной сверточный блок (блок ST-Conv) для совместной обработки временных рядов, структурированных графами. Сам блок может быть уложен или расширен в зависимости от масштаба и сложности конкретных случаев.

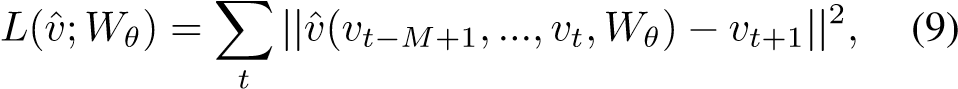
Как показано на рисунке 2 (в центре), пространственный слой в середине должен быть мостом между двумя временными слоями, которые могут обеспечить быстрое распространение в пространственном состоянии от свертки графа через временные свертки. Структура «сэндвича» также помогает сети в достаточной степени применять стратегию узких мест для достижения сжатия масштаба и сжатия функций за счет уменьшения и увеличения каналов *C* через сверточный слой графа. Кроме того, нормализация слоев используется в каждом блоке ST-Conv для предотвращения переобучения.

Вход и выход блоков ST-Conv являются трехмерными тензорами. Для входного *vl* ∈ RM×n×Cl блока *l* выходной *vl+1* ∈

R(*M−2(Kt−1))×n×Cl+1* вычисляется по формуле:

РеЛУ*,* (8)

где  — верхнее и нижнее височные ядра в блоке *l* соответственно; Θl — спектральное ядро свертки графа; ReLU(·) обозначает функцию выпрямленных линейных единиц. После наложения двух блоков ST-Conv мы присоединяем дополнительный слой временной свертки с полностью соединенным слоем в качестве выходного слоя в конце (см. левую часть рисунка 2). Слой темпоральной свертки сопоставляет выходные данные последнего блока ST-Conv с одношаговым прогнозом. Затем мы можем получить окончательный выход *Z* ∈ Rn×c из модели и рассчитать прогноз скорости для *n* узлов, применив линейное преобразование по c-каналам как *vˆ = Zw* + *b*, где w ∈ Rc — вектор веса, а *b* является предвзятостью. Мы используем потери L2 для измерения производительности нашей модели. Таким образом, функция потерь STGCN для прогнозирования трафика может быть записана как:



где *Wθ* — все обучаемые параметры в модели; *vt+1* — это основная истина, а *vˆ(·)* обозначает прогноз модели.

Теперь мы резюмируем основные характеристики нашей модели STGCN в следующем:

* STGCN — это универсальная среда для обработки структурированных временных рядов. Он способен не только решать проблемы моделирования и прогнозирования сети трафика, но и применяться для более общих задач пространственно-временного обучения последовательностей.
* Пространственно-временной блок сочетает в себе графовые свертки и стробированные временные свертки, которые могут извлекать наиболее полезные пространственные признаки и когерентно захватывать наиболее существенные временные признаки.
* Модель полностью состоит из сверточных структур и поэтому обеспечивает распараллеливание по входным данным с меньшим количеством параметров и более высокой скоростью обучения. Что еще более важно, такая экономическая архитектура позволяет модели работать с крупномасштабными сетями с большей эффективностью.

# Эксперименты

## Описание набора данных

Мы проверяем нашу модель на двух реальных наборах данных о дорожном движении.

BJER4 и PeMSD7, собранные Пекинской муниципальной дорожной комиссией и Калифорнийским департаментом транспорта соответственно. Каждый набор данных содержит ключевые атрибуты наблюдений за дорожным движением и географическую информацию с соответствующими временными метками, как подробно описано ниже.

BJER4 был собран с основных участков маршрутов восточного кольца No 4 в Пекине с помощью двухпетлевых детекторов. Для нашего эксперимента выбрано 12 дорог. Данные о трафике агрегируются каждые 5 минут. Используется период с 1 июля по 31 августа 2014 года, за исключением выходных. В качестве тренировочного набора мы выбираем первый месяц исторических рекордов скорости, а остальное служит валидационным и тестовым набором соответственно.

PeMSD7 был собран из системы измерения производительности Caltrans (PeMS) в режиме реального времени с помощью более чем 39 000 сенсорных станций, развернутых в крупных мегаполисах системы автомагистралей штата Калифорния [Chen *et al.*, 2001]. Набор данных также агрегируется в 5-минутный интервал из 30-секундных выборок данных. Мы случайным образом выбираем средний и крупный масштаб среди округа 7 штата Калифорния, содержащего округа 228 и 1 026

Рисунок 3: Сеть датчиков PeMS в округе 7 штата Калифорния (слева), каждая точка обозначает сенсорную станцию; Тепловая карта взвешенной матрицы смежности в PeMSD7(M) (справа).

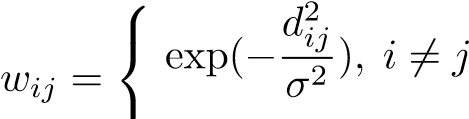
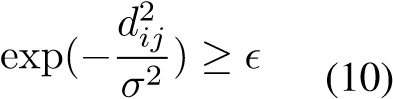
станции, обозначенные как PeMSD7(M) и PeMSD7(L) соответственно, в качестве источников данных (показаны слева на рисунке 3). Временной диапазон набора данных PeMSD7 — это май и июнь 2012 года. Мы разделили обучающий и тестовый наборы по тому же принципу, что и выше.

## Предварительная обработка данных

Стандартный интервал времени в двух наборах данных равен 5 минутам. Таким образом, каждый узел дорожного графа содержит 288 точек данных в день. Метод линейной интерполяции используется для заполнения пропущенных значений после очистки данных. Кроме того, ввод данных нормализуется по методу Z-Score.

В BJER4 топология графа дорог в системе кольцевых маршрутов No 4 «Восток Пекина» построена по схеме развертывания сенсорных станций. Сопоставляя аффилиацию, направление и точки отправления-назначения каждой дороги, система кольцевых маршрутов может быть оцифрована в виде ориентированного графа.

В PeMSD7 матрица смежности графа дорог вычисляется на основе расстояний между станциями в сети дорожного движения. Взвешенная матрица смежности *W* может быть сформирована как:

 и 

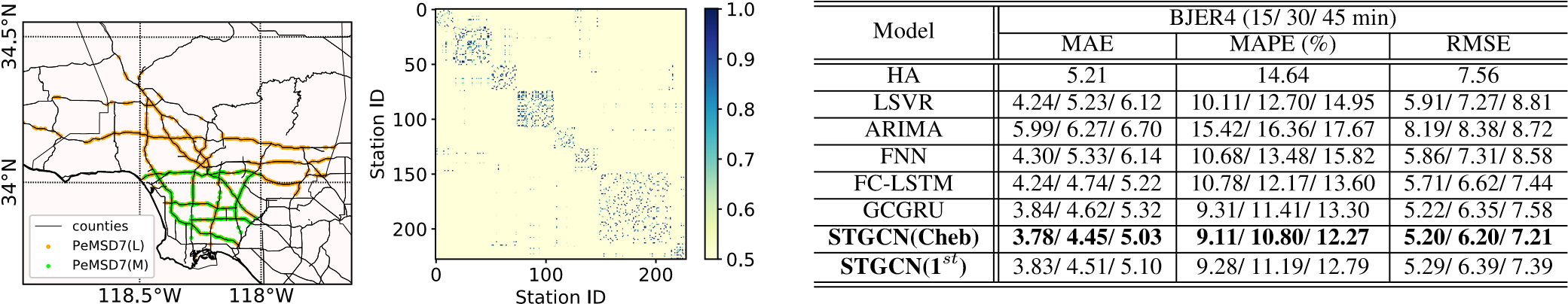
0 *,* иначе*.*

где *wij* — вес ребра, который определяется *как dij* (расстояние между станциями *i* и *j*). *σ2* и являются пороговыми значениями для управления распределением и разреженностью матрицы *W*, присвоенной значениям 10 и 0*.*5 соответственно. Визуализация *W* представлена справа на рисунке 3.

## Экспериментальные настройки

Все эксперименты скомпилированы и протестированы на кластере Linux (CPU: Intel(R) Xeon(R) CPU E5-2620 v4 @ 2.10GHz, GPU: NVIDIA GeForce GTX 1080). Чтобы исключить нетипичный трафик, в нашем эксперименте используются только данные о трафике за рабочие дни [Li *et al.*, 2015]. Мы применяем стратегию поиска по сетке, чтобы найти лучшие параметры на валидациях. Во всех тестах используется 60 минут в качестве исторического временного окна, т.е. 12 наблюдаемых точек данных (*M* = 12) используются для прогнозирования условий дорожного движения в следующие 15, 30 и 45 минут (*H =* 3,6,9).

Метрика оценки и базовые показатели Для измерения и оценки производительности различных методов, Средние абсолютные ошибки (MAE), Средние абсолютные процентные ошибки (MAPE) и

Таблица 1: Сравнение производительности различных подходов на наборе данных BJER4.

Принимаются среднеквадратичные ошибки (RMSE). Мы сравниваем нашу систему STGCN со следующими базовыми показателями: 1). Историческое среднее значение (HA); 2). Регрессия Виктора с линейной поддержкой (LSVR); 3). Авторегрессионная интегрированная скользящая средняя (ARIMA); 4). Нейронная сеть с прямой связью (FNN); 5). Полносвязный LSTM (FC-LSTM) [Sutskever *et al.*, 2014]; 6). Граф сверточного ГРУ (GCGRU) [Li *et al.*, 2018].

Модель STGCN Для BJER4 и PeMSD7 (M/L) каналы трех слоев в блоке ST-Conv составляют 64, 16, 64 соответственно. Размер ядра свертки графа *K* и размер ядра темпоральной свертки *Kt*  в модели STGCN(Cheb) с приближением полиномов Чебышева равны 3, а в модели STGCN(1st) с аппроксимацией 1-го порядка K равен 1. Мы обучаем наши модели, минимизируя среднеквадратичную ошибку с помощью RMSprop для 50 эпох с размером партии 50. Начальная скорость обучения равна 10−3 со скоростью распада 0,7 через каждые 5 эпох.

## Результаты эксперимента

В таблицах 1 и 2 представлены результаты STGCN и исходные показатели на наборах данных BJER4 и PeMSD7(M/L). Предложенная нами модель обеспечивает наилучшую производительность со статистической значимостью (двусторонний Т-критерий, *α* = 0*).*01, *<* 0*.*01) по всем трем метрикам оценки. Мы можем легко заметить, что традиционные методы статистики и машинного обучения могут хорошо работать для краткосрочного прогнозирования, но их долгосрочные прогнозы неточны из-за накопления ошибок, проблем с запоминанием и отсутствия пространственной информации. Модель ARIMA показывает худшие результаты из-за своей неспособности обрабатывать сложные пространственно-временные данные. Подходы к глубокому обучению обычно достигают лучших результатов прогнозирования, чем традиционные модели машинного обучения.

Преимущества пространственной топологии

Предыдущие методы не включали пространственную топологию и моделировали временные ряды грубым образом. С другой стороны, благодаря моделированию пространственной топологии датчиков, наша модель STGCN достигла значительного улучшения в краткосрочном, среднесрочном и долгосрочном прогнозировании. Преимущество STGCN более очевидно на наборе данных PeMSD7, чем на BJER4, поскольку сенсорная сеть PeMS более сложна и структурирована (как показано на рисунке 3), и наша модель может эффективно использовать пространственную структуру для более точных прогнозов.

Чтобы сравнить три метода на основе свертки графа, выполните следующие действия:

GCGRU, STGCN(Хеб) и STGCN(1-й), мы показываем их

10:00

14:00

18:00

22:00

20

30

40

50

60

70

Скорость (км/ч)

ХА

ГЧГРУ

STGCN(1-я)

STGCN (Хеб)

Наземная достоверность

04:00

10:00

16:00

22:00

20

30

40

50

60

70

Скорость (км/ч)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Модель |  | PEMSD7(M) (15/ 30/ 45 мин) | | |  |  | PEMSD7(L) (15/ 30/ 45 мин) | | |  | | МАЕ |  | MAPE (%) |  | RMSE | МАЕ |  | MAPE (%) |  | RMSE | | ХА | 4.01 | | 10.61 | 7.20 | | 4.60 | | 12.50 | 8.05 | | | ЛСВР | 2.50/ 3.63/ 4.54 | | 5.81/ 8.88/ 11.50 | 4.55/ 6.67/ 8.28 | | 2.69/ 3.85/ 4.79 | | 6.27/ 9.48/ 12.42 | 4.88/ 7.10/ 8.72 | | | АРИМА | 5.55/ 5.86/ 6.27 | | 12.92/ 13.94/ 15.20 | 9.00/ 9.13/ 9.38 | | 5.50/ 5.87/ 6.30 | | 12.30/ 13.54/ 14.85 | 8.63/ 8.96/ 9.39 | | | ФНН | 2.74/ 4.02/ 5.04 | | 6.38/ 9.72/ 12.38 | 4.75/ 6.98/ 8.58 | | 2.74/ 3.92/ 4.78 | | 7.11/ 10.89/ 13.56 | 4.87/ 7.02/ 8.46 | | | ФК-ЛСТМ | 3.57/ 3.94/ 4.16 | | 8.60/ 9.55/ 10.10 | 6.20/ 7.03/ 7.51 | | 4.38/ 4.51/ 4.66 | | 11.10/ 11.41/ 11.69 | 7.68/ 7.94/ 8.20 | | | ГЧГРУ | 2.37/ 3.31/ 4.01 | | 5.54/ 8.06/ 9.99 | 4.21/ 5.96/ 7.13 | | 2.48/ 3.43/ 4.12 ∗ | | 5.76/ 8.45/ 10.51 ∗ | 4.40/ 6.25/ 7.49 ∗ | | | STGCN (Хеб) | 2.25/ 3.03/ 3.57 | | 5.26/ 7.33/ 8.69 | 4.04/ 5.70/ 6.77 | | 2.37/ 3.27/ 3.97 | | 5.56/ 7.98/ 9.73 | 4.32/ 6.21/ 7.45 | | | STGCN(1-я) | 2.26/ 3.09/ 3.79 | | 5.24/ 7.39/ 9.12 | 4.07/ 5.77/ 7.03 | | 2.40/ 3.31/ 4.01 | | 5.63/ 8.21/ 10.12 | 4.38/ 6.43/ 7.81 | |   Таблица 2: Сравнение производительности различных подходов на наборе данных PeMSD7.   |  |  |  |  | | --- | --- | --- | --- | | Набор данных | Затраты времени (с) | | | | STGCN (Хеб) | STGCN(1-я) | ГЧГРУ | | ПЭМСД7(М) | 272.34 | 271.18 | 3824.54 | | ПеМСД7(Л) | 1926.81 | 1554.37 | 19511.92 | |

Рисунок 4: Прогнозирование скорости в утренние пиковые и вечерние часы пик набора данных PeMSD7.

0

500

1000

1500

2000

2500

3000

3500

4000

Время тренировки (с)

6

7

8

9

10

11

12

ТестRMSE

STGCN (Хеб)

STGCN(1-я)

ГЧГРУ

0

10

20

30

40

50

Эпоха тренировок

3

4

5

6

7

ТестМАЭ

STGCN (Хеб)

STGCN(1-я)

ГЧГРУ

Рисунок 5: Сравнение RMSE между тестом и временем обучения (слева); Протестируйте MAE в зависимости от количества эпох обучения (справа). (ПеМСД7(М))

прогнозы в утренние часы пик и вечерние часы пик, как показано на рисунке 4. Нетрудно заметить, что наше предложение STGCN отражает тенденцию часов пик более точно, чем другие методы; И он обнаруживает окончание часов пик раньше других. Основываясь на эффективной свертке графов и структурах свертки времени с накоплением, наша модель способна быстро реагировать на динамические изменения в сети трафика, не полагаясь на среднее историческое значение, как это делает большинство рекуррентных сетей.

Эффективность и обобщение тренинга

Чтобы увидеть преимущества свертки по оси времени в нашем предложении, мы суммируем сравнение времени обучения между STGCN и GCGRU в таблице 3. С точки зрения справедливости, GCGRU состоит из трех слоев с 64, 64, 128 единицами соответственно в эксперименте для PeMSD7(M), а STGCN использует настройки по умолчанию, описанные в разделе 4.3. Наша модель STGCN потребляет всего 272 секунды, в то время как RNN-тип модели GCGRU тратит 3 824 секунды на PeMSD7(M). Это 14-кратное ускорение скорости обучения в основном выигрывает от применения временной свертки вместо рекуррентных структур, которые могут достичь полностью параллельного обучения, а не полагаться исключительно на цепные структуры в качестве RNN

Таблица 3: Затраты времени на обучение на наборе данных PeMSD7.

делать. Для PeMSD7(L) GCGRU должен использовать половину размера партии, так как потребление графического процессора превысило объем памяти одной карты (результаты помечены как "\*" в таблице 2); в то время как STGCN нужно только удвоить каналы в середине блоков ST-Conv. Даже несмотря на то, что наша модель все еще потребляет менее одной десятой времени обучения модели GCGRU при таких обстоятельствах. Между тем, преимущества аппроксимации 1-го порядка проявились, так как она не ограничивается параметризацией полиномов. Модель STGCN(1st) ускоряется примерно на 20% на большем наборе данных с удовлетворительной производительностью по сравнению с STGCN(Cheb).

Для дальнейшего исследования производительности сравниваемых моделей глубокого обучения мы построили график RMSE и MAE тестового набора PeMSD7(M) в процессе обучения (см. рисунок 5). Эти цифры также свидетельствуют о том, что наша модель может достичь гораздо более быстрой процедуры обучения и более легкой сходимости. Благодаря специальной конструкции блоков ST-Conv наша модель обладает превосходными характеристиками в балансировке времени потребления и настроек параметров. В частности, количество параметров в STGCN (4*.*54 × 105) составляет всего около двух третей GCGRU и сохраняет более 95% параметров по сравнению с FC-LSTM.

# Связанные работы

В последнее время существует несколько исследований глубокого обучения, которые также мотивированы сверткой графов в пространственно-временных задачах. Seo *et al.* [2016] Введена графовая сверточная рекуррентная сеть (GCRN) для совместной идентификации пространственных структур и динамических вариаций из структурированных последовательностей данных. Ключевой задачей данного исследования является определение оптимальных комбинаций рекуррентных сетей и свертки графов при определенных условиях. Основываясь на вышеизложенных принципах, Li *et al.* [2018] успешно применил стробированные рекуррентные блоки (ГРУ) со сверткой графов для долгосрочного прогнозирования трафика. В отличие от этих работ, мы строим нашу модель полностью из сверточных структур; Блок ST-Conv специально разработан для равномерной обработки структурированных данных со стратегией остаточной связи и узких мест внутри; В нашей модели также используются более эффективные ядра свертки графов.

# Заключение и будущая работа

В этой статье мы предлагаем новую среду глубокого обучения STGCN для прогнозирования трафика, интегрирующую графовую свертку и закрытую временную свертку через пространственно-временные сверточные блоки. Эксперименты показывают, что наша модель превосходит другие современные методы на двух реальных наборах данных, что указывает на ее большие возможности для исследования пространственно-временных структур на основе входных данных. Кроме того, он обеспечивает более быстрое обучение, более легкую сходимость и меньшее количество параметров благодаря гибкости и масштабируемости. Эти особенности весьма перспективны и практичны для научных разработок и широкомасштабного промышленного развертывания. В дальнейшем мы будем дополнительно оптимизировать структуру сети и настройки параметров. Более того, предложенная нами структура может быть применена к более общим сценариям пространственно-временного структурированного прогнозирования последовательностей, таким как развитие социальных сетей, прогнозирование предпочтений в рекомендательных системах и т.д.

# Ссылки

[Ахмед и Кук, 1979] Мохаммед С. Ахмед и Аллен Р. Кук. *Анализ данных временных рядов о движении на автомагистралях с использованием методов Бокса-Дженкинса*. 1979 год.

[Bruna *et al.*, 2013] Джоан Бруна, Войцех Заремба, Артур Слам и Ян Лекун. Спектральные сети и локально связанные сети на графах. *Препринт arXiv arXiv:1312.6203*, 2013.

[Chen *et al.*, 2001] Чао Чэнь, Карл Петти, Александр Скабардонис, Правин Варайя и Чжаньфэн Цзя. Система измерения производительности автомагистрали: данные детектора контуров горных работ. *Отчет о транспортных исследованиях: Журнал Совета по транспортным исследованиям*, (1748):96–102, 2001.

[Chen *et al.*, 2016] Цюаньцзюнь Чэнь, Сюань Сун, Харутоси Ямада и Рёсукэ Сибасаки. Изучение глубокого представления больших и разнородных данных для вывода о дорожно-транспортных происшествиях. В *AAAI,* стр. 338–344, 2016.

[Defferrard *et al.*, 2016] Михаэль Дефферрар, сын Ксавье Бре и Пьер Вандергейнст. Сверточные нейронные сети на графах с быстрой локализованной спектральной фильтрацией. В *NIPS,* страницы 3844–3852, 2016.

[Gehring *et al.*, 2017] Йонас Геринг, Майкл Аули, Давид Гранжье, Денис Яратс и Ян Н Дофин. Обучение от сверточной последовательности к последовательности. *Препринт arXiv arXiv:1705.03122*, 2017.

[Hammond *et al.*, 2011] Дэвид К. Хаммонд, Пьер Вандергейнст и Реми Грибонваль. Вейвлеты на графиках с помощью теории спектральных графов. *Прикладной и вычислительный гармонический анализ*, 30(2):129–150, 2011.

[Hochreiter and Schmidhuber, 1997] Зепп Хохрайтер и Юрген Шмидхубер. Длительная кратковременная память. ̈ *Нейронные вычисления*, 9(8):1735–1780, 1997.

[Huang *et al.*, 2014] Вэньхао Хуан, Гоцзе Сун, Хайкунь Хун и Куньцин Се. Глубокая архитектура для прогнозирования транспортных потоков: сети глубоких убеждений с многозадачным обучением. *Труды IEEE по интеллектуальным транспортным системам*, 15(5):2191–2201, 2014.

[Jia *et al.*, 2016] Юйхань Цзя, Цзяньпин Ву и Иман Ду. Прогнозирование скорости движения с использованием метода глубокого обучения. В *ITSC*, стр. 1217–1222. IEEE, 2016.

[Кипф и Веллинг, 2016] Томас Н. Кипф и Макс Веллинг. Полуконтролируемая классификация с графовыми сверточными сетями. *Препринт arXiv arXiv:1609.02907*, 2016.

[Li *et al.*, 2015] Есинь Ли, Юй Чжэн, Хуэйчу Чжан и Лэй Чэнь. Прогнозирование трафика в системе велошеринга. В *SIGSPATIAL,* стр. 33. ACM, 2015.

[Li *et al.*, 2018] Ягуан Ли, Роуз Юй, Сайрус Шахаби и Янь Лю. Диффузионная сверточная рекуррентная нейронная сеть: прогнозирование трафика на основе данных. В *ICLR*, 2018.

[Lv *et al.*, 2015] Ишэн Лю, Яньцзе Дуань, Вэньвэнь Кан, Чжэнси Ли и Фэй-Юэ Ван. Прогнозирование транспортных потоков с помощью больших данных: подход глубокого обучения. *Труды IEEE по интеллектуальным транспортным системам*, 16(2):865–873, 2015.

[Niepert *et al.*, 2016] Матиас Ниперт, Мохамед Ахмед и Константин Куцков. Обучение сверточных нейронных сетей для графов. В *ICML*, страницы 2014–2023, 2016.

[Seo *et al.*, 2016] Ёнджу Сео, Михаэль Дефферрар, Пьер Вандергейнст и Ксавье Брессон. Структурированное моделирование последовательностей с помощью графовых сверточных рекуррентных сетей. *Препринт arXiv arXiv:1612.07659*, 2016.

[Shi *et al.*, 2015] Синцзянь Ши, Чжужун Чэнь, Хао Ван, Дит-Ян Юнг, Вай-Кин Вонг и Ван-чунь Ву. Сверточная сеть lstm: подход к машинному обучению для прогнозирования осадков. В *NIPS,* страницы 802–810, 2015.

[Шуман *и др.*, 2013] 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.

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

[Vlahogianni, 2015] Eleni i Vlahogianni. Вычислительный интеллект и оптимизация для больших данных на транспорте: проблемы и возможности. В книге « *Оптимизация инженерных и прикладных наук*», стр. 107–128. Springer, 2015.

[Уильямс и Хоэл, 2003] Билли М. Уильямс и Лестер А. Хоэл. Моделирование и прогнозирование транспортных потоков как сезонного процесса: теоретическая база и эмпирические результаты. *Журнал транспортной инженерии*, 129(6):664–672, 2003.

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