

Connecting the dots: MTGNN

Zonghan Wu, Shirui Pan, et al. *Connecting the Dots: Multivariate Time Series Forecasting with Graph Neural Networks*, 2020

20 мая 2024 г.

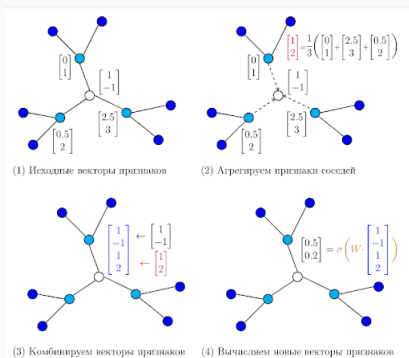
Постановка задачи

Пусть дан многомерный временной ряд $f = (f_1, f_2, \dots, f_T)$, $f_i \in \mathbb{R}^d$, d - количество отслеживаемых показателей. Нужно сделать предсказания $(f_{T+1}, f_{T+2}, \dots, f_{T+M})$, где M - выбранная длина предсказаний.

→ **Переходим к датасету.** Выберем параметр l (*sequence length*) - количество предыдущих точек, которые используются для предсказания. Одно наблюдение в датасете - тензор размерности $n \times d \times l$, n - получившееся количество наблюдений ($n = T - l + 1$).

Graph neural network

Пусть задан граф, где у каждой вершины v есть n -мерный вектор признаков. Будем агрегировать информацию о соседях каждой вершины, умножая их признаки на матрицу W и складывая (*Graph Convolution, GC*). Подробнее - [1].



Spatio temporal GNN

Теперь каждая вершина в графе обладает признаком $x_i \in \mathbb{R}^{T \times n}$. Фактически, мы добавили ось времени для каждой оси признака вершин. После применения GC слоя используется свертка по времени, например, CNN или LSTM, отдельно для каждой вершины (*Time series Convolution, TC*). Подробнее - [2]

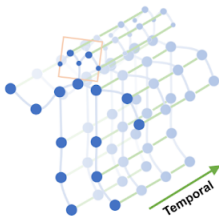


Figure 1: The spatial temporal graph of a skeleton sequence used in this work where the proposed ST-GCN operate on. Blue dots denote the body joints. The intra-body edges between body joints are defined based on the natural connections in human bodies. The inter-frame edges connect the same joints between consecutive frames. Joint coordinates are used as inputs to the ST-GCN.

Представление исходных данных

В нашей задаче будем воспринимать каждый временной ряд как отдельную вершину, которая обладает признаком $x_i \in \mathbb{R}^{l \times 1}$.

Проблема: в подходах выше структура графа была определена заранее - откуда её взять в нашей задаче?

Решение: будем обучать не только GC, TC модули, но и саму структуру графа [4].

Общий pipeline: исходные эмбединги вершин подаем в неструктурированном (табличном) виде, применяем последовательно GC и TC слои.

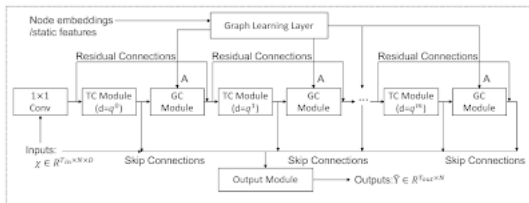


Figure 2: The framework of MTGNN. A 1×1 standard convolution first projects the inputs into a latent space. Afterward, temporal convolution modules and graph convolution modules are interleaved with each other to capture temporal and spatial dependencies respectively. The hyper-parameter, dilation factor d , which controls the receptive field size of a temporal convolution module, is increased at an exponential rate of q . The graph learning layer learns the hidden graph adjacency matrix, which is used by graph convolution modules. Residual connections and skip connections are added to the model to avoid the problem of gradient vanishing. The output module projects hidden features to the desired dimension to get the final results.

MTGNN: graph learning layer

Авторы обучают структуру ориентированного графа с несимметричной матрицей смежности A без петель. Такая структура удобна для модуля графовой свертки, куда одновременно подаются на вход A и A^T .

$$M_1 = \tanh(\alpha E_1 \Theta_1)$$

$$M_2 = \tanh(\alpha E_2 \Theta_2)$$

$$A = \text{ReLU}(\tanh((M_1 M_2^T - M_2 M_1^T)))$$

Далее, в каждой строке выбирается top-k наибольших значений, а остальные обнуляются, чтобы сделать матрицу более разреженной. E_1, E_2 - обучаемые эмбединги вершин, Θ_1, Θ_2 - обучаемые параметры, α - гиперпараметр, аналогичный температуре. Чем выше α , тем ближе положительные значения в M_1, M_2 ближе к 1, и тем меньше значений обнуляется с помощью $\text{ReLU}(\cdot)$ в последнем уравнении.

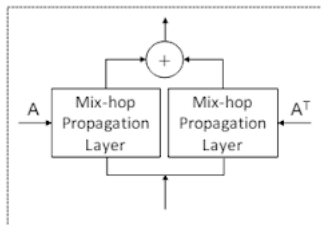
Пусть $D^* = 1 + \sum_j A_{ij}$, $A^* = D^{*-1}(A + I)$, $H_{(0)} = H_{in}$. Тогда на каждом шаге свертку по соседям можно написать как

$$H_{(k)} = \beta H_{in} + (1 - \beta) A^* H_{(k-1)}$$

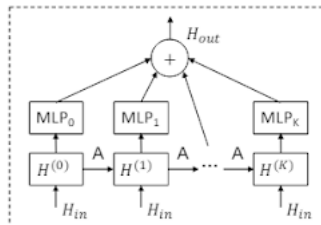
$$H_{out} = \sum_{i=0}^K H_{(i)} W_{(i)}$$

По существу, A^* - нормализованная по соседям матрица смежности, т.е. теперь для каждой вершины каждое ребро до её соседа отнормированно на сумму ребер до всех соседей этой вершины.

MTGNN: GC module



(a) GC module

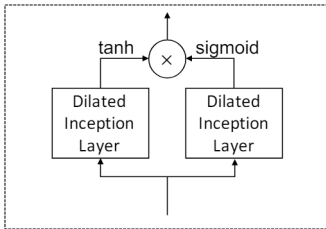


(b) Mix-hop propagation layer

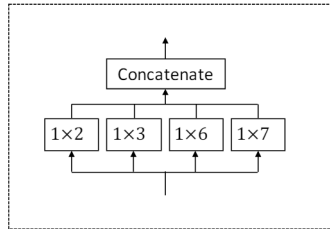
Figure 4: Graph convolution and mix-hop propagation layer.

MTGNN: TC module

На вход подается последовательность состояний каждой вершины и независимо применяется $1d$ свертка одновременно для 4 фильтров. Выходы этих фильтров конкатенируются в один вектор. 4 фильтра с размерностью ядер 2,3,6 и 7 помогают проходить сверткой по часто встречающимся сезонностям: 7, 12, 24, 28 и 60.



(a) TC module








(b) Dilated inception layer

Figure 5: The temporal convolution and dilated inception layer.

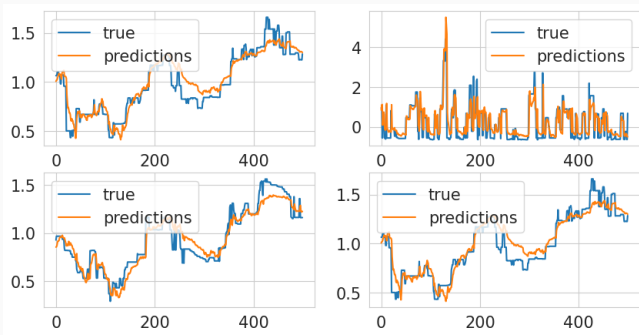
Результаты

Более подробно про архитектуру можно почитать в оригинальной работе [4] или в [письменной версии доклада](#).

Task	Dataset	Model	Metric Name	Metric Value	Global Rank	Result	Benchmark
Univariate Time Series Forecasting	Electricity	MTGNN (3 step)	RRSE	0.0745	# 1		Compare
Univariate Time Series Forecasting	Electricity	MTGNN (6 step)	RRSE	0.0878	# 4		Compare
Univariate Time Series Forecasting	Electricity	MTGNN (12 step)	RRSE	0.0916	# 5		Compare
Univariate Time Series Forecasting	Electricity	MTGNN (24 step)	RRSE	0.0953	# 8		Compare
Traffic Prediction	NE-BJ	MTGNN	12 steps MAE	4.90	# 4		Compare

Применение на реальных данных: high frequency

1 группа рядов - высокочастотные данные по курсу биткоина (грануляция ≈ 2 миллисекунды). Включают такие показатели, как bid price, bid amount, ask price, ask amount.



Применение на реальных данных: high frequency

Код можно найти по [ссылке](#). Данные - [здесь](#). При подсчете метрик использовалось макро-усреднение.

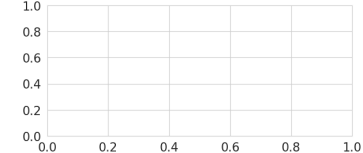
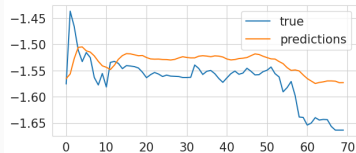
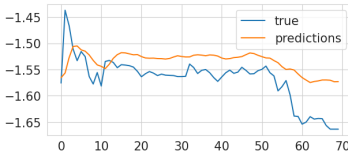
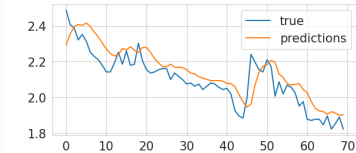
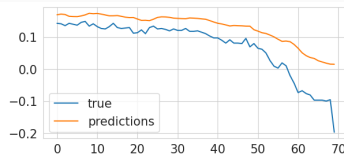
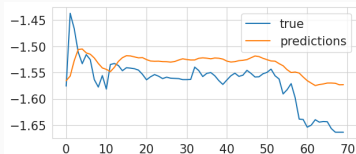
```
MAPE: 0.46296686  
MSE: 0.2627155  
RMSE: 0.5125578  
RRSE: 0.5259136  
Correlation coef: 0.8278372007923586
```





Применение на реальных данных: котировки

Дневные данные по акциям: ОБК, Фармсинтез, Иркут, Диод, Аптечная сеть 36.

```
MAPE: 0.3117344  
MSE: 0.0043272967  
RMSE: 0.06578219  
RRSE: 0.9740782  
Correlation coef: 0.7988964085151267
```

Применение на реальных данных: котировки



-  Jie Zhou, Ganqu Cui et al. Graph neural networks: A review of methods and applications // AI Open. - 2020
-  Sijie Y., Yuanjun X., Dahua L. Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition // 2018
-  Zahraa A. Spatio-Temporal Graph Neural Networks: A Survey // 2023
-  Zonghan W., et al. Connecting the Dots: Multivariate Time Series Forecasting with Graph Neural Networks // 2020

