View Reviews

Paper ID
7385

Paper Title
Adaptive Graph Convolutional Recurrent Network for Traffic Forecasting

Reviewer #1

Questions

1. Summary and contributions: Briefly summarize the paper and its contributions.
This paper proposes a GNN based model for traffic forecasting. Its backend model is a recurrent GNN model for sequential data. GNN is used to capture structure information between multiple time series and recurrent solution is used to adopt temporal modeling. The key contribution for this work lies in the first part, which can be summarized as follows

1. Different from general GNN that uses a shared model to learn embedding for all nodes, this model hints using unique parameters for each node.
2. Once the adjacency matrix is not provided, it introduces learnable parameters for each node, and its pairwise similarly are used to formulize the adjacency matrix. Such approach can be trained end-to-end.

2. Strengths: Describe the strengths of the work. Typical criteria include: soundness of the claims (theoretical grounding, empirical evaluation), significance and novelty of the contribution, and relevance to the NeurIPS community.
1. This paper is well motivated. Indeed for time series data, especially those from different resources, doing unique fitting for each data is reasonable. This paper further has a low rank assumption for time series data, which is a reasonable set up.
2. This papers provides a straight forward and easy to implement solution for the adjacency matrix when the graph structure is unknown.

3. Weaknesses: Explain the limitations of this work along the same axes as above.
It is not totally clearly where the superiority is coming from and it is not very convincing whether the experiments are fair comparisons.
a. There is no regularization for the E_g to be orthogonal, where is not very clear what is the underlying structure of E_g. Intuitively, a in-detail analysis should be provided, e.g, What is the rank of E_g, especially when sparse regularization is introduced.
b. Since E_a has parameter in size of Nxd_c, I am in doubt whether this model is over-parameterized for the embedding. It is important to compare the model size of the proposed one and those baselines; further, the analysis for the learned embeddings are missing. It would be interesting to understand what types of time series are picked to be connected. What is the relation between such time series, are they only highly correlated or do they have strong lead-lag relation or whatever? How could be model achieve sparsity in the graph? It is common some other time series are not informative to forecast an target time series, if so, how to avoid such link in the graph? Some insightful and essential analysis are missing.

4. Correctness: Are the claims and method correct? Is the empirical methodology correct?

The claims and method are correct

5. Clarity: Is the paper well written?

This paper is clearly writtten and easy to follow

6. Relation to prior work: Is it clearly discussed how this work differs from previous contributions?

This paper has properly refer previous relatede work.

7. Reproducibility: Are there enough details to reproduce the major results of this work?

Yes

9. Please provide an "overall score" for this submission.

5: Marginally below the acceptance threshold.

10. Please provide a "confidence score" for your assessment of this submission.

3: You are fairly confident in your assessment. It is possible that you did not understand some parts of the submission or that you are unfamiliar with some pieces of related work. Math/other details were not carefully checked.

11. Have the authors adequately addressed the broader impact of their work, including potential negative ethical and societal implications of their work?

Yes

Reviewer #2

Questions

1. Summary and contributions: Briefly summarize the paper and its contributions.

This paper investigates the problem of traffic flow forecasting, and proposes methods to capture node specific patterns as well as learn the inter-dependencies among nodes without a predefined graph. The investigated problem is important, and the empirical results on two real-world datasets show very promising results.

--

Update after rebuttal:

Thanks for the feedback. It will be helpful to add discussion in the paper about (1) Integrates existing graph structure information. (2) Multiple step prediction. (3) and scalability.

2. Strengths: Describe the strengths of the work. Typical criteria include: soundness of the claims (theoretical grounding, empirical evaluation), significance and novelty of the contribution, and relevance to the NeurIPS community.

S1: The investigated problem is important, and empirical results on real-world datasets are promising.

S2: The idea of using learning node-specific parameters with shared basis (or lower rank decomposition) is very interesting.

S3: The idea of learning the normalized adjacency matrix from scratch with random embedding is very interesting. Here E (E_G or E_A) can be seen the representation of the node, W (W_z, W_r or W_h) can be seen as transformations from the node embedding space to target spaces

S4: The proposed approaches do not seem to limit to traffic flow prediction, and can be potentially applied to a much broader domain wherever pair-wise correlation is important.

3. Weaknesses: Explain the limitations of this work along the same axes as above.

W1: The presentation can be improved. It seems that this paper is finished in a rush (or has last minute change).

- For example, in Table 1, Metr-LA is shown, while it is actually means PeMSD8.
- In Table 1, "*" seems to be missing from? It seems that the paper originally compares with baselines on METR-LA (citing performance from other paper), but decide to switch to PeMSD8 for some reason.
- In Table 2, the model names are NAGG-x while the ones in the corresponding paragraphs are DAGG-x.

W2: The experimental are not quite convincing.
- In Table 1, the "*"sign is not present in any numbers, does it mean all the baseline methods in the Table implemented by the author?
- Most baselines, e.g., DCRNN, STGCN, ASTGCN, have their performance evaluated on PeMSD4, PeMSD7, PeMSD8, METR-LA, PEMS-BAY (w.r.t. traffic speed rather than traffic flow). Thus, it will be more convincing to compare with baselines with numbers achieved in the corresponding papers rather than regenerating the results on a different setting.
- It will be very helpful to see the comparison on the Metr-LA dataset (it seems the lib/dataloadr.py also has METR-LA as an option) It will be helpful to show whether the proposed approach still able to achieve superior performance.

W3: More justification of the proposed approach.

- Learning graph structure. It is very surprising to see that the proposed approach achieves better results than baselines without using graph information. Does it mean the pair-wise distance or similarity cannot provide any extra information? It is also interesting to see if it is possible to achieve improved result if these information is utilized in AGCRN (e.g., as regularization of the embedding).
- Multiple step prediction. AGCRN generates the output in one-shot, instead of sequential manner (without positional embedding). It will be interesting to see if sequentially generates the output or adding positional embedding will further improve the result.
- It seems that the proposed method need O(N^2) computation, comparing to some baseline whose time complexity is O(|E|) where |E| is the number of edges. It will be helpful to discuss how to scale the AGCRN to a large scale graphs. Will graph partition applicable? Will mini-batch training applicable?

4. Correctness: Are the claims and method correct? Is the empirical methodology correct?

Both the idea of using learning node-specific parameters with shared basis (or lower rank decomposition) and the idea of learning the normalized adjacency matrix from scratch with random embedding is intuitive. Besides, the experimental results in Table 1, 2 also support the claims in the paper.

5. Clarity: Is the paper well written?

The paper is generally easy to follow, and the proposed approach is described in a relatively clear way. However, it seems that this paper is finished in a rush (or has last minute changes).
- For example, in Table 1, Metr-LA is shown, while it is actually means PeMSD8.
- In Table 1, "*" seems to be missing from? It seems that the paper originally compares with baselines on METR-LA (citing performance from other paper), but decide to switch to PeMSD8

for some reason.
- In Table 2, the model names are NAGG-x while the ones in the corresponding paragraphs are DAGG-x.

6. Relation to prior work: Is it clearly discussed how this work differs from previous contributions?
Yes, important related works are discussed and differentiated in Section 2.

7. Reproducibility: Are there enough details to reproduce the major results of this work?
Yes

9. Please provide an "overall score" for this submission.
7: A good submission; accept.

10. Please provide a "confidence score" for your assessment of this submission.
5: You are absolutely certain about your assessment. You are very familiar with the related work.

11. Have the authors adequately addressed the broader impact of their work, including potential negative ethical and societal implications of their work?
Yes

Reviewer #3

Questions

1. Summary and contributions: Briefly summarize the paper and its contributions.
This paper proposes two adaptive modules: 1) Node Adaptive Parameter Learning (NAPL) 2) Data Adaptive Graph Generation (DAGG) to improve the capability of traditional Graph Neural Networks. Further, they combine both modules to propose a new variant of Graph Neural Networks. The experiments show some promising results for proposed methods on two traffic datasets.

2. Strengths: Describe the strengths of the work. Typical criteria include: soundness of the claims (theoretical grounding, empirical evaluation), significance and novelty of the contribution, and relevance to the NeurIPS community.
1. Two proposed modules can work separately or jointly to improve the traditional graph neural network.
2. The evaluation results are solid and the improvements are significant.
3. The proposed framework can be extended to other time-series applications.

3. Weaknesses: Explain the limitations of this work along the same axes as above.
1. In section 3.5, I wonder why the authors particularly choose L1 loss to optimize and why not using other loss functions, e.g., L2 loss (square loss). Can the authors elaborate more on this?
2. In section 4.1, the authors formulate the problem as a 12-step prediction. Does it suggest that 0-12 hrs are inputs and 13-24 hrs are targets?
3. In section 4.4, I am a little bit confused about the setup for short-term (e.g., 5mins) and long-term predictions (e.g., 60 mins). Are they also multi-step predictions? I would like the authors to clarify this part.

4. Correctness: Are the claims and method correct? Is the empirical methodology correct?
The proposed method looks correct and the evaluation results seem reasonable to me.

5. Clarity: Is the paper well written?
This paper is well-written and easy to follow.

6. Relation to prior work: Is it clearly discussed how this work differs from previous contributions?
The authors have demonstrated the differences between the proposed work and previous works.

7. Reproducibility: Are there enough details to reproduce the major results of this work?
Yes

8. Additional feedback, comments, suggestions for improvement and questions for the authors:
Please try to address my concerns in "Weaknesses".

==============post-rebuttal==============
Thanks for the response. I will keep my previous rating after reading the rebuttal.

9. Please provide an "overall score" for this submission.
6: Marginally above the acceptance threshold.

10. Please provide a "confidence score" for your assessment of this submission.
4: You are confident in your assessment, but not absolutely certain. It is unlikely, but not impossible, that you did not understand some parts of the submission or that you are unfamiliar with some pieces of related work.

11. Have the authors adequately addressed the broader impact of their work, including potential negative ethical and societal implications of their work?
Yes

Reviewer #4

Questions

1. Summary and contributions: Briefly summarize the paper and its contributions.
This paper proposed a deep learning model Adaptive Graph Convolutional Recurrent Network (AGCRN) for traffic forecasting while the pre-defined graph is avoidable. The proposed model consisted of three main parts: 1) a Node Adaptive Parameter Learning (NAPL) module to capture node-specific patterns; 2) a Data Adaptive Graph Generation (DAGG) module to infer the inter-dependencies among different traffic series automatically. 3) recurrent network. AGCRN captured fine-grained spatial and temporal correlations in traffic series automatically and showed good performance on two real-world traffic datasets.

2. Strengths: Describe the strengths of the work. Typical criteria include: soundness of the claims (theoretical grounding, empirical evaluation), significance and novelty of the contribution, and relevance to the NeurIPS community.
Existing GCN-based methods require pre-defining an inter-connection graph by similarity or distance measures to capture the spatial correlations. That further requires substantial domain knowledge and is sensitive to the graph quality. This paper proposed a GCN-based traffic forecasting model while the pre-defined graph is avoidable. Experiments on two real-world traffic datasets show the proposed model outperformed state-of-the-art models by a significant margin without pre-defined graphs about spatial connections.

3. Weaknesses: Explain the limitations of this work along the same axes as above.
1. Relationship to previous work. First and foremost, the paper lacks a discussion on previous related work on graph convolution network (GCN). As per my knowledge, the technologies used in 3.2 NAPL and 3.3 DAGG have already been proposed/discussed in [1, 2, 3]. These papers are very famous for improving GCN, but not mentioned in this paper.

As a reader, I am not sure what contributions exactly this paper has made. I encourage the authors to present which papers the technologies that they use to improve GCN are from, and what this paper is adding to the literature or how this paper modifying these technologies to adapt traffic forecasting task.

[1] Graph Attention Network, ICLR 2018.
[2] Self-Attention Graph Pooling, ICML 2019.
[3] Hierarchical Graph Representation Learning with Differentiable Pooling, NeurIPS 2018.

2. Technical Novelty: as explained in first item, due to lacks of a discussion on previous related work on graph convolution network, the proposed model in this paper seems replace the GCN part of previous traffic forecasting models, with a more advanced existing one. The technical novelty of the paper is limited. But I understand that technical novelty is not the aim of the paper, so I list this item in the bottom.

4. Correctness: Are the claims and method correct? Is the empirical methodology correct?
The claims and method in this paper seem correct.

The experimental settings seem reasonable. Code for the model was submitted.

5. Clarity: Is the paper well written?
This paper is well written.

6. Relation to prior work: Is it clearly discussed how this work differs from previous contributions?
This paper clearly discussed how this work differs from previous traffic forecasting work. But it lacks a discussion on previous related work on graph convolution network (GCN).

7. Reproducibility: Are there enough details to reproduce the major results of this work?
Yes

8. Additional feedback, comments, suggestions for improvement and questions for the authors:
1. What's Metr-LA in table 1? Metr-LA has not been introduced in the paper.
2. STSGCN is shown from reference [4], which should be from refernce [11].

9. Please provide an "overall score" for this submission.
6: Marginally above the acceptance threshold.

10. Please provide a "confidence score" for your assessment of this submission.
4: You are confident in your assessment, but not absolutely certain. It is unlikely, but not impossible, that you did not understand some parts of the submission or that you are unfamiliar with some pieces of related work.

11. Have the authors adequately addressed the broader impact of their work, including potential negative ethical and societal implications of their work?
Yes