FID3018 - Paper review

Filip Cornell

May 2021

1 Introduction

In this course, I chose to present three papers, which all had a common theme; few-shot learning either using a graph-based learning approach [3, 11], or few-shot learning on graphs [10].

The few-shot learning paradigm, also known as the meta-learning paradigm, is based on quickly learning new, unseen classes given only a few examples of each class [5]. Each class is considered a task $\mathcal{T}_j = \{D^{tr}_{\mathcal{T}_j}, D^{test}_{\mathcal{T}_j}\}$ to learn, where $D^{tr}_{\mathcal{T}_j}$ is a support set and $D^{test}_{\mathcal{T}_j}$ is a query set which we want to correctly classify given the support set. Note that the training, validation and test set are disjoint sets of tasks, so that a class only exists in either the training, validation or the test set.

The meta-learning paradigm has seen great progress in the latest few years, since the emergence of similarity-based approaches such as matching networks [8] and optimization-based approaches such as Model-Agnostic Meta Learning [2]. In my three papers, all are in some way similarity-based, and two of them [11, 3] are their own versions of similarity-based meta-learning based on graph neural networks, while the other uses an LSTM-based [4] Matching Network [8].

2 Few-shot Learning with Graph Neural Networks

2.1 Approach

The first paper, originally made by Garcia et al. [3], introduces the concept of using graph neural networks on few-shot image classification. This is a highly flexible approach, in which the images are seen as nodes in a small, fully-connected network. Each image first obtains its features by first inputting the image into a Convolutional Neural Network (CNN) architecture, and these are then concatenated with a one-hot encoding vector of its label (if the label is known) and a uniform distribution of the labels if it is unknown.

The network then learns similarities between the images represented as scalar values represented as edges, which are produced typically through 3-layer Graph

Neural Network (GNN) blocks. The final output is then a vector of probabilities for each class, and the final loss is a cross-entropy loss considering the probabilities for each class.

2.2 Strong points

This paper was a completely new approach to perform few-shot learning when it was published, and a lot of papers have spun off from it since, proving it to be a useful framework. It is highly flexible as it can be used in both a few-shot, semi-supervised and active learning setup.

2.3 Weak points

One of the paper's weak points is the fact that labels need to be concatenated as features to each image. This requires us to know precisely how many labels are in all datasets. This therefore has a weakness in few-shot learning, as it requires us to know one before-hand exactly how many labels there are in all disjoint sets of tasks.

3 Few-shot audio classification with attentional graph neural networks

3.1 Approach

This paper is a spin-off application from Garcia et al. [3], in which the same approach is used but applied differently in a new domain; audio classification. Here, the task is to classify different sounds using a set of support examples.

There are a few minor differences and changes in the approach. First of all, instead of using a uniform distribution of the labels, they use a 0-vector for examples for which the label is unknown. Second of all, they use a pre-trained network, SoundNet [1] to arrive at efficient representations to input into GNN which later produces class predictions. They also use attention-based mechanisms, which is only used by Garcia et al. [3] in an active learning-based scenario. They use two types of attention; global attention, where attention is used in between all examples, and intra-class attention, where attention is only applied in-between classes, where the former is shown to be more effective.

3.2 Strong points

This paper is interesting because it makes use of a previously existing architecture and applies it onto a new domain, showing its effectiveness in different domains and therefore its generalizability. A strong point is that it shows that the uniform distribution over unknown labels is not crucial for good predictions; rather, one can instead use 0-vectors for this.

3.3 Weak points

A weak point is that it depends on using a highly complex pre-trained network SoundNet [1]. Moreover, it depends on knowing the labels on before-hand, much like Garcia et al. [3].

4 Few-shot Knowledge Graph Completion

4.1 Approach

The last paper, made by Zhang et al. [10], shows that knowledge graph completion can be made in an unsupervised manner. Inspired by Xiong et al. [9] who performs one-shot relational learning by using a single example in the support set, they improve upon this work by extending the few-shot scenario to up to five-shot relational learning.

Coming to the approach and the architecture, they take each head entity, relation and triplet in both the support and the query set and embed them along with their one-hop neighbors. When embedding the one-hop neighborhood into a compact representation for each entity, they apply attention to the different neighbors differently.

Once the entities from the different triplet examples are embedded, they use a Recurrent Neural Network (RNN) [6] based encoder-decoder to embed the support set into an aggregate representation of the support set. This is then compared to the query's representation in an LSTM-based Matching Network [4, 8] and outputs a similarity-based score, based on how similar the query representation is to the representation of the support set.

4.2 Strong points

The paper is strong in the sense that it improves and makes possible the use of several support examples. Furthermore, it does not make the assumption that Xiong et al. [9] makes, that each neighbor in the one-hop neighborhood of an entity contributes equally to the representation of it.

4.3 Weak points

There are a few weak points of the paper. First of all, the aggregation of the query representation is LSTM-based, which is not order-invariant. This creates a bias towards which order one inputs the support examples into the aggregation, and this could therefore be improved. Furthermore, the attention applied onto the one-hop neighbors is static throughout relational tasks, creating relation-independent representations of each entity. This is improved upon by Sheng et al. [7] later, but it is important to note that the representations of each head and tail entity are statically encoded independent from the relational task.

5 Conclusion

The three papers I presented all showed that few-shot learning either using a graph-based learning approach or applying the few-shot learning paradigm onto graphs can be efficient in scenarios where it is needed. While the first two papers are variants of the same approach, the latter shows that graph neural networks might not necessarily be the only approach in a graph-learning scenario. I hae not seen any paper in the few-shot knowledge graph completion domain making use of graph neural networks, but it can perhaps be future work to do by drawing inspiration from the two previous papers.

If I am to be critical towards my own presentation, I think the two former paper were highly related to each other, while the last one was not. I would probably have, in hindsight, chosen a paper more similar and comparable to the other two as the last paper to present. Nevertheless, it still provides a comparison in the sense that few-shot learning can be done on graphs without a graph-based approach, and non-graph data with a graph-based approach. I therefore deem them suitable in general, and I was happy with the discussion that followed.

References

[1] Yusuf Aytar, Carl Vondrick, and Antonio Torralba. Soundnet: Learning sound representations from unlabeled video. *arXiv preprint arXiv:1610.09001*, 2016.

[2] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*, pages 1126–1135. PMLR, 2017.

[3] Victor Garcia and Joan Bruna. Few-shot learning with graph neural networks. *arXiv preprint arXiv:1711.04043*, 2017.

[4] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[5] Mike Huisman, Jan N van Rijn, and Aske Plaat. A survey of deep meta-learning. *Artificial Intelligence Review*, pages 1–59, 2021.

[6] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning internal representations by error propagation. Technical report, California Univ San Diego La Jolla Inst for Cognitive Science, 1985.

[7] Jiawei Sheng, Shu Guo, Zhenyu Chen, Juwei Yue, Lihong Wang, Tingwen Liu, and Hongbo Xu. Adaptive attentional network for few-shot knowledge graph completion. *arXiv preprint arXiv:2010.09638*, 2020.

[8] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Koray Kavukcuoglu, and Daan Wierstra. Matching networks for one shot learning. *arXiv preprint arXiv:1606.04080*, 2016.

[9] Wenhan Xiong, Mo Yu, Shiyu Chang, Xiaoxiao Guo, and William Yang Wang. One-shot relational learning for knowledge graphs. *arXiv preprint arXiv:1808.09040*, 2018.

[10] Chuxu Zhang, Huaxiu Yao, Chao Huang, Meng Jiang, Zhenhui Li, and Nitesh V Chawla. Few-shot knowledge graph completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 3041–3048, 2020.

[11] Shilei Zhang, Yong Qin, Kewei Sun, and Yonghua Lin. Few-shot audio classification with attentional graph neural networks. In *INTERSPEECH*, pages 3649–3653, 2019.