Research Statement

Shubhangi Ghosh

For my undergraduate thesis project, I worked on a Multi-hop reasoning based Question Generation task. Multi-hop reasoning assesses the ability of the model to reason with information taken from more than one document to arrive at the answer. The dataset used, HotpotQA (Yang et al. [2018]), provides supporting facts as documents. These documents are concatenated to form the input sequence.

Derivatives of Seq-to-seq models have long been the state of the art for the Question Generation task. Seq-to-seq models have been found to recognize and generate short-range syntactical patterns in an elegant end-to-end trainable fashion. However, since they process information sequentially, long-range dependencies have been the bottleneck of seq-to-seq models.

To facilitate multi-hop reasoning, one of the methods I used, reads long-range dependencies in the form of an input graph. The input graph was created by linking the multiple instances of named entities across documents, while all named entities within each document were also linked. Graph Convolutional Networks (GCNs) were used in conjugation with seq-to-seq models to incorporate the dependency information. The results obtained upon using GCNs were unconvincing, however. Working on this problem introduced me to GCNs and got me to think about the suitability of the structural aspects of GCNs to the problem at hand.

1 Adapting GCNs to be robust to variable input graphs for Seq-to-seq tasks in Natural Language Processing

GCNs, as proposed in Kipf and Welling [2016], have been primarily used on undirected, partially-labelled graphs for semi-supervised node classification. The introductory work in Marcheggiani and Titov [2017] presents a generalization of GCNs on directed, labelled graphs for computing improved node representations that capture neighborhood information from input graphs (commonly depicting syntactic relationships among words). GCNs are since increasingly being used to improve word representations for natural language processing tasks, including semantic role labelling, machine translation, question answering, and question generation. (Marcheggiani and Titov [2017], Marcheggiani et al. [2018]). Performance improvements using GCNs, however, have not been highly satisfactory.

In semi-supervised node classification tasks, the input graphs were clearly shared across all the data points to be classified. In NLP tasks, however, the input graph that the GCN operates on varies across data points. While analysing the shortcomings of GCNs forNLP tasks for my undergraduate thesis, a possible explanation that I came up with is that, the variability of input graphs across data points increases the model complexity and makes the optimisation of the GCN parameters an exceedingly hard task. Under your mentorship, I would like to develop a measure to quantify sentence variability/graph variability and evaluate the performance of GCNs on NLP tasks on different scales of variability of input graphs. Based on the validity of the proposed justification, I would also like to build a model capable of encoding input graphical relationships among words while being robust to input graph variability.

2 Adapting GCNs to learn arbitrary neighbourhood filters for identifying graph features

GCNs were originally proposed as a generalization of Convolutional Neural Networks (CNNs) from regular grids (e.g. images) to irregular graphs defined on non-Euclidean domains (e.g. social networks). CNNs were thus an obvious prior model to compare with for me to understand GCNs better. The seminal work by Defferrard et al. [2016] formulates the convolution/filtering operation in the spatial domain by reducing it to the multiplication operation in the spectral domain in accordance with the Convolution Theorem (Mallat [1999]). The fundamental spectral modes are identified as the eigenvectors of the Laplacian operator. (The Laplacian operator is the equivalent of the differentiation operator in linear, time-domain sequences, which has exponential functions as its Fourier modes.)

Defferrard et al. [2016] formulates *localized* filters as polynomial coefficients of the graph Laplacian. A K^{th} order polynomial represents a filter localized to the K-hop neighbourhood. Since polynomial filters weigh each entire m-distance neighbourhood with a constant multiplier, in the equivalent spatial domain, nodes belonging to an m-distance neighbourhood are only weighed proportionately to the neighbourhood weights defined in the input graph. In comparison to the spatial formulation of CNNs, where filters can learn arbitrary weights for local neighbours, this polynomial formulation of filters in the spectral domain appears to be restrictive. I would like to work on modelling a broader class of localized filters while maintaining the same linear computational complexity and constant learning complexity.

Correspondence to Shubhangi Ghosh: shubhangisghosh@gmail.com

References

Zhilin Yang, Peng Qi, Saizheng Zhang, Yoshua Bengio, William W Cohen, Ruslan Salakhutdinov, and Christopher D Manning. Hotpotqa: A dataset for diverse, explainable multi-hop question answering. *arXiv preprint arXiv:1809.09600*, 2018.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

Diego Marcheggiani and Ivan Titov. Encoding sentences with graph convolutional networks for semantic role labeling. *arXiv preprint arXiv:1703.04826*, 2017.

Diego Marcheggiani, Joost Bastings, and Ivan Titov. Exploiting semantics in neural machine translation with graph convolutional networks. *arXiv preprint arXiv:1804.08313*, 2018.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in neural information processing systems*, pages 3844–3852, 2016.

Stéphane Mallat. *A wavelet tour of signal processing*. Elsevier, 1999.

Correspondence to Shubhangi Ghosh: shubhangisghosh@gmail.com