Self Supervision in Graph Neural Networks for Text Classification.

Team name: HSC01
Members: Sahil Raina, Harshavardhan Battula, Cody Wang

University of Minnesota
Driven to Discover®

Overview

- Problem Definition and Motivation
- Literature Survey
- Proposed Methods
- Experimental Setups
- Final Results and Error Analysis
- Limitations and Ethical Issues
- Future Direction

Motivation

- Recent progress in Language Modeling with Self Supervision showed good improvements in capturing the underlying semantic meaning of sentences and documents.

- But we observed that some of these attention based Large Language Models fail in simple cases. Ex: So delicious was the noodles but terrible vegetables; P(negative) = 0.72 with BERT base.

- At first look, the word “terrible” is closer to “noodles” than “delicious”, and there could be “terrible <dish>” appearing in datasets that makes the model think these two words are closely associated.

- Some other examples, I like the recipe here; P(positive) = 0.62 when “like is used as a verb”, the recipe includes chinese food like dumplings; P(negative) = 0.92 when used as a preposition. This is a challenge when analyzing product reviews.

Motivation (continued)

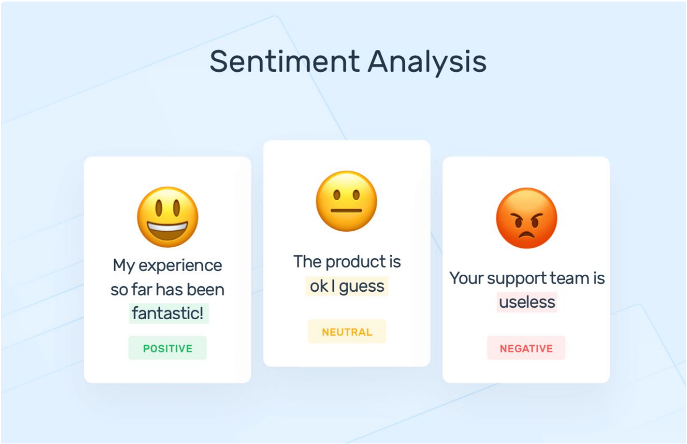
- Prior research suggests using syntax information to establish connections between words to solve such problems. Relative to semantic, very little work is done in using the structural information of these sentences for downstream tasks.

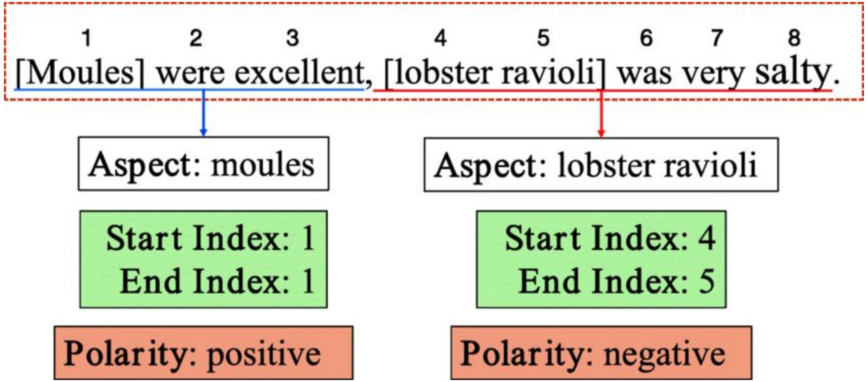
- Graphs can be used to express meaningful relationships between two entities(ex: words, patches in an image, random variables, etc.). We plan to exploit the strength of graph neural networks in capturing structural relationships on dependency trees.

- Prior research in this area by *Wang et al*. suggests combining structural embeddings from GNN along with semantic embeddings from Language models like BERT, gives a good performance improvement over baseline.

Problem Definition

- Trying to improve the performance of Graph Based models for Text Classification by using pre-training schemes in a hope that it helps the downstream classification task.

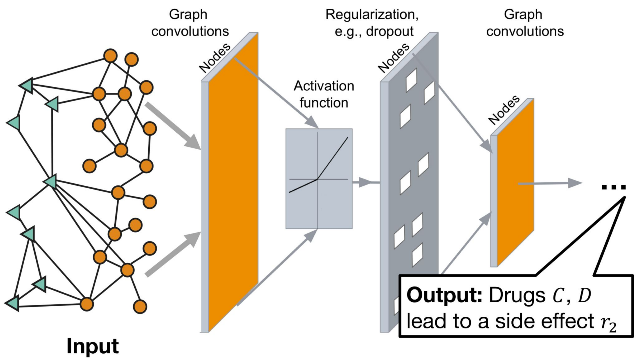




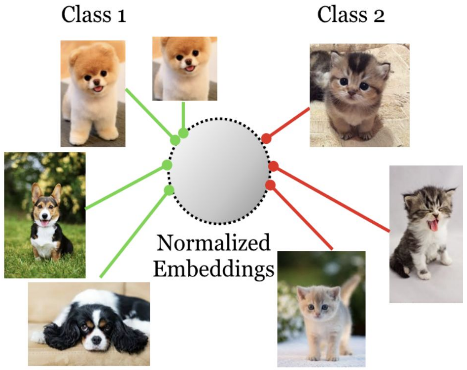
Inspiration

- Self-Supervision on graphs by *You et al.* reported better results on downstream classification tasks after pretraining GNN on citations, cite-ceer, and other datasets.

- Therefore, we implemented the Contrastive Learning Framework on Graph Neural Networks to improve the performance of the model.



GNN for Multi Relational Link Prediction



Contrast between pairs of images from different classes.

Datasets

- Laptop and Restaurant review datasets from the SemEval 2014 Task for pre-training and fine-tuning.

Dataset	Positive		Neutral		Negative	
	Train	Test	Train	Test	Train	Test
Laptop	994	341	870	128	464	169
Restaurant	2164	728	807	196	637	196

Literature Survey

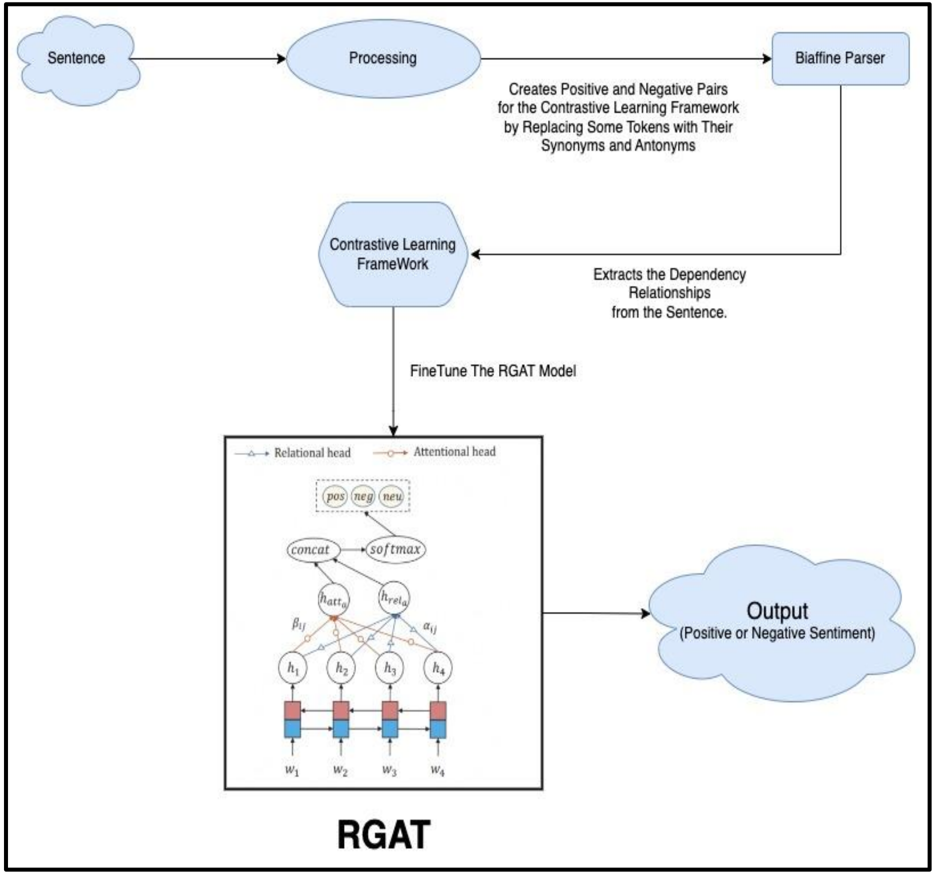
- ATAE-LSTM by *Wang et al., 2020* uses an attention-based LSTM to classify sentiment surrounding a target aspect.
- TNET by *Li et al.* employs a similar approach, using an LSTM in addition to Target-Specific Transformations to integrate target aspect information into the word representations.
- BERT by *Devlin, Jacob et al.* introduces a new language representation model that pre-trains deep bi-directional representations using the left and right context together.
- T-GCN by *Tian et al.* uses Graph Neural Networks on dependency trees to solve text classification problems.

Literature Survey (Contrastive Methods)

- *Wang et al. (2022)* combined three types of negatives for sampling on knowledge graphs to improve training efficiency and without much significant computational and memory overhead.
- *Wang et al. (2021)* proposes the CLINE framework as a way to generate positive and negative samples for the text by replacing some tokens with their synonyms and antonyms as this increases the robustness of pre-trained models against adversarial attacks and increases sensitivity to semantic changes.
- *Gao et al.* proposes the SimCSE framework that has two versions, a supervised and unsupervised one, to advance the state-of-the-art sentence embeddings through its simple contrastive learning framework.

Proposed Idea and Novel Contributions

- Developed a new framework, as shown on the side, with the objective of getting better performance for sentiment analysis.
- Modified the RGAT from the last time such that it can handle well the newly generated sentences from the original sentence to generate their dependency trees.
- Used the RGAT to combine word embeddings that were initialized using Glove algorithm and refined using LSTM based sequence model and information from dependency tags for text classification.
- Scrapped the Idea of Using Subgraph Augmentation and DistilledBert for pre-training as understood that it's not needed for our project.



Contrastive Learning

- Our intention was to generate contrastive pairs such that sentences with similar sentiment are closer together while pairs with opposite sentiment are far apart.
- The CLINE framework to generate our contrastive pairs.
- To create positive and negative pairs, tokens from a sentence were replaced with synonyms and antonyms.
- Roughly 40% of tokens were replaced for positive pairs and 20% for negative pairs.

Original: “Great laptop that offers many great features !”
Synonyms: “nifty laptop that offers many not bad features !”
Antonyms: “Great laptop that offers few great features !”

Limitations

- Parsing
  - Performance of the algorithm depends on the parser.
- Contrastive learning
  - Synonyms/antonyms did not always make sense.
  - Currently no method of selecting the proper synonym/antonym based on the context.

Original: "I charge it at night and skip taking the cord with me because of the good battery life ."
Synonym: "I charges it at nights and skips taking the cords with me because of the good batteries lives ."
Antonym: "I pay cash it at day and skip give the cord with me because of the evil battery life ."

Ethical Issues

- Bias
  - The researchers who maintain WordNet inject their biases into the database - this affects the synonyms/antonyms used to generate contrastive pairs.
- Abuse
  - Sentiment analysis can be misused by powerful groups to negatively impact minority groups.
  - Examples of this include automatically managing insurance premiums, automatic censorship, etc.

Preliminary Results

- We implemented the framework shown before and got the following results

Dataset	Accuracy	F1 Score
SemEval14-Restaurant	0.74	0.63
SemEval14-Laptop	0.71	0.70
SemEval14-Twitter	0.73	0.73

Table 3: Performance of Relational Graph Attention Network on SemEval14 datasets.

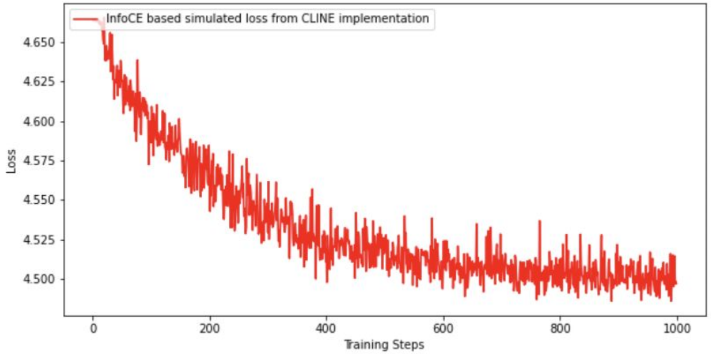


Figure 4: Contrastive Loss Metric from CLINE Paper

Dataset	Accuracy	F1 Score
SemEval14-Restaurant	0.75	0.68
SemEval14-Laptop	0.71	0.70
SemEval14-Twitter	0.69	0.65

Table 4: Performance of Relational Graph Attention Network Pre-trained with CLINE followed by fine tuning.

Error Analysis

- In some places, the replacement of synonyms and antonyms isn’t ideal or perfect.
- Contrastive learning framework is basic but not much sophisticated and needs improvement.

Future Steps

- Working on Improving the Accuracy of the Sentiment Analysis
- Try Developing better Contrastive Learning Framework to Learn and Pre-Train the Models Better
- Experiment with the Hyperparameters and Configurations of the Model and the Graph to improve Performance.

Broader Impact

- We think the framework can be transferred to most representation based problems, and it can also applied to diverse set of problems like Knowledge-Base problems, Anomaly Detection in sensors, etc.

References

1. Wang, Kai, Weizhou Shen, Yunyi Yang, Xiaojun Quan, and Rui Wang. "Relational graph attention network for aspect-based sentiment analysis." *arXiv preprint arXiv:2004.12362* (2020).
2. You, Yuning, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. "Graph contrastive learning with augmentations." *Advances in Neural Information Processing Systems* 33 (2020): 5812-5823
3. Khosla, Prannay, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. "Supervised contrastive learning." *Advances in Neural Information Processing Systems* 33 (2020): 18661-18673.
4. Wang, Yequan, Minlie Huang, Xiaoyan Zhu, and Li Zhao. "Attention-based LSTM for aspect-level sentiment classification." In Proceedings of the 2016 conference on empirical methods in natural language processing, pp. 606-615. 2016.
5. Li, Xin, Lidong Bing, Wai Lam, and Bei Shi. "Transformation networks for target-oriented sentiment classification." arXiv preprint arXiv:1805.01086 (2018).
6. Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).
7. Tian, Yuanhe, Guimin Chen, and Yan Song. "Aspect-based sentiment analysis with type-aware graph convolutional networks and layer ensemble." In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 2910-2922. 2021.

References

7. Liang Wang, Wei Zhao, Zhuoyu Wei, and Jingming Liu. 2022. Simkgc: Simple contrastive knowledge graph completion with pre-trained language models. arXiv preprint arXiv:2203.02167.
8. Dong Wang, Ning Ding, Piji Li, and Hai-Tao Zheng. 2021. Cline: Contrastive learning with semantic negative examples for natural language understanding. arXiv preprint arXiv:2107.00440.
9. Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. Simcse: Simple contrastive learning of sentence embeddings. arXiv preprint arXiv:2104.08821.

University of Minnesota

Driven to Discover®

Crookston Duluth Morris Rochester Twin Cities