Change or Sustain Talk Classification with Language Models and Graph Neural Networks

Ahson Saiyed, MS[1], Maksim Tsvetovat, PhD[1], Tatyna Kanzaveli, MS[1], John Layton, BA[2], Joannalyn Delacruz[3], Brian Borsari[3,1], Jason M. Satterfield, PhD[2]

1 Open Health Network, Mountain View, CA, USA, 2 University of California, San Francisco, San Francisco, CA, USA, 3 San Francisco Va Medical Center, San Francisco, CA, USA.

Abstract

Change or Sustain Talk classification performance plays a central role in a technology-assisted MI system's ability to support end-to-end automated MI-consistent conversation.

This study leverages advancements in graph neural networks and large scale pertained language models to achieve SOTA Change and Sustain talk classification on smoking cessation related dialogue.

Introduction

“Motivational Interviewing is a counseling strategy designed to elicit behavior change by strengthening personal desire for a specific goal” (Miller & Rollnick, 2012). A key component to improving the efficacy of technology delivered adaptations of MI counseling is the efficient and accurate recognition of Change or Sustain talk from text.

Classical machine learning methods for text classification, such as those reliant on handcrafted feature engineering, have been out performed by deep learning techniques centered around leveraging machine-generated embeddings for representations of text (Minaee). Fine-tuned variants of large-scale transformer based Pre-Trained Language Models have demonstrated the ability to produce high quality embeddings and SOTA performance on many downstream NLP tasks, including text classification (Minaee). More recently, studies have shown that the utility of PLM embeddings can be further enhanced as input to various Graph Neural Networks architectures (Wu).

This study evaluates the performance of PLM embeddings with and without further augmentation generated from convolutions over an utterance similarity graph on the task of Change or Sustain classification using “BertGCN" (Liu).

Data

Table 1. Dataset summary statistics

# Utters.	# Train	# Test	# Words	# Nodes	# Classes	Average Length
51,942	44812	7702	4123	61,616	3	19.83

51,942 utterances were collected from expert annotated MI session transcripts discussing smoking cessation. Each utterance was labeled one of three categories, “Change”, “Sustain” or “Follow/Neutral”.

Methodology

Language Model Encoding

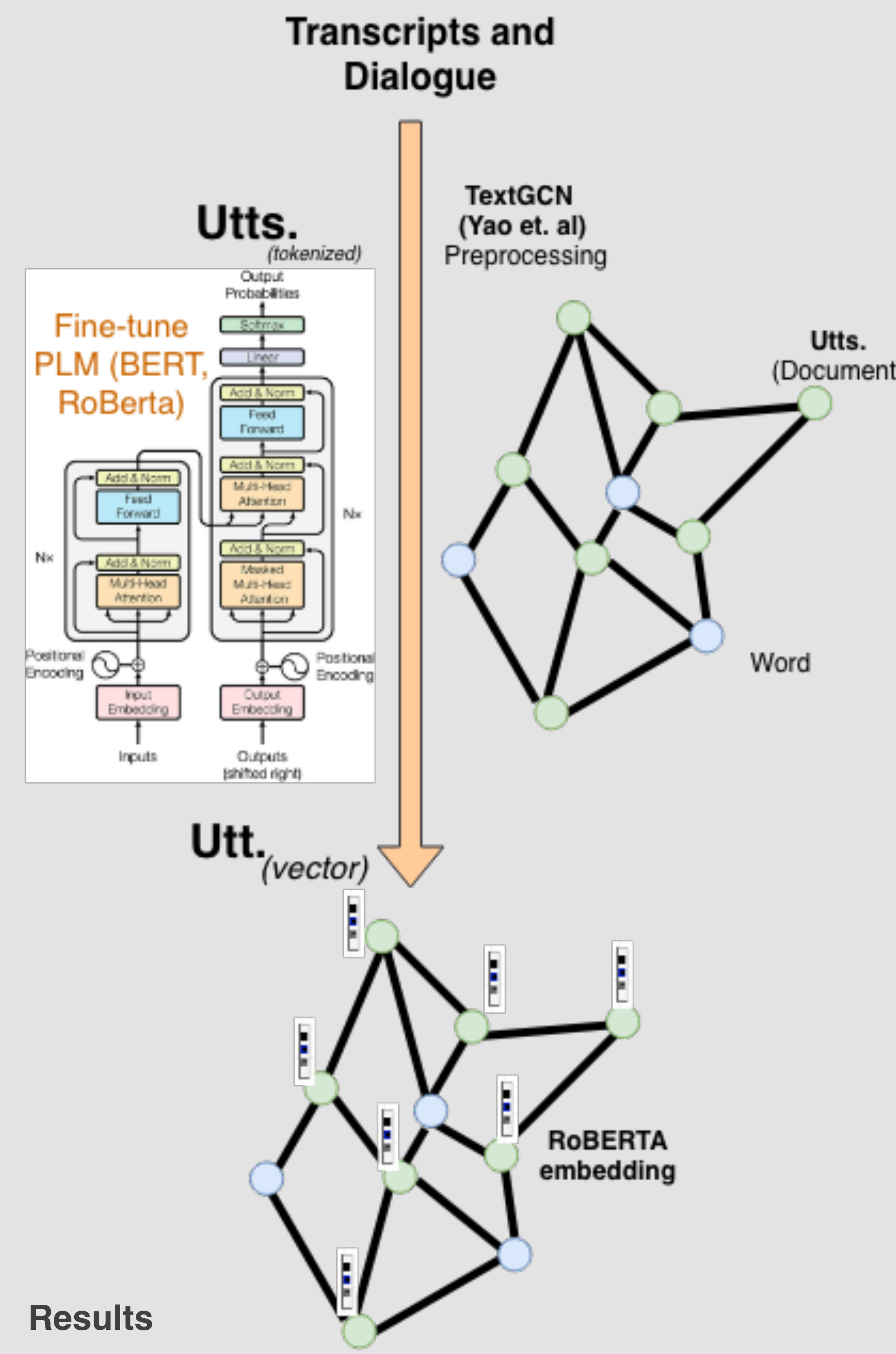
Utterance level representation are generated from fine-tuning large scale pre-trained language models, BERT, RoBERTa, and XLNet. Language models were trained with a learning rate of 4e-7 with a batch size of 8.

Graph Representation

Utterances and words are treated as nodes in a heterogeneous graph, and for utterances, node attributes are populated with 128-dimensional utterance level representations generated from fine-tuned language models. Edges between nodes are assigned and weighted considering semantic similarity between nodes. (Liu)

Evaluation

Multi-label node classification is conducted implementing various Graph Neural Networks architectures. Each model is run 5 times and the mean and standard deviation for accuracy, F1-score, MCC are recorded.



Results

Table 3. Performance metrics across models

Model	Accuracy in %
BERT	70.0
RoBERTa	72.1
TextGCN	65.6
RoBERTaGCN	81.2

Conclusion

Text classification using TextGCN was performed to demonstrate the difference between graph convolutional considering one-hot encoded representations of utterances compared to the deep bidirectional representations generated by the fine-tuned PLMs.

Fine-tuned PLM based architectures outperformed TextGCN. Initial results indicate that RoBertaGCN is the most performant architecture for Change or Sustain classification, and out performs classification on embeddings generated from only fine-tuned PLM models.

Augmenting fine-tuned utterance level representations considering information aggregated from utterance node neighborhood through graph convolutions has demonstrated superior performance on downstream Change or Sustain classification.

References

Lin, Yuxiao, et al. “BERTGCN: Transductive Text Classification by Combining GNN and Bert.” *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, 2021, doi:10.18653/v1/2021.findings-acl.126.

Miller, William R., and Stephen Rollnick. *Motivational Interviewing: Helping People Change*. Guilford Publ., 2012.

Minaee, Shervin, et al. “Deep Learning--Based Text Classification.” *ACM Computing Surveys*, vol. 54, no. 3, 2021, pp. 1–40., doi:10.1145/3439726.

Tanana, Michael, et al. “A Comparison of Natural Language Processing Methods for Automated Coding of Motivational Interviewing.” *Journal of Substance Abuse Treatment*, vol. 65, 2016, pp. 43–50., doi:10.1016/j.jsat.2016.01.006.

Wu, Lingfei, et al. “Graph Neural Networks for Natural Language Processing: A Survey.” *ArXiv*, 2021, pp. 207–221., doi:10.1017/9781108924184.015.

Yao, Liang, et al. “Graph Convolutional Networks for Text Classification.” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 7370–7377., doi:10.1609/aaai.v33i01.33017370.