GCNs-Based Context-Aware Short Text Similarity Model

Xiaoqi Sun[†], Shaochun Wu[*], Yue Liu[‡]
[*†]School of Computer Engineering and Science, Shanghai University
[‡]School of Computer Engineering and Science, Shanghai University,
Shanghai Institute for Advanced Communication and Data Science,
Shanghai Engineering Research Center of Intelligent Computing System,
Shanghai, 200444, China
Email: {xiaoqisun, scwu, yueliu}@shu.edu.cn

Xiaoqi Sun
Shanghai University

Contents

Introduction

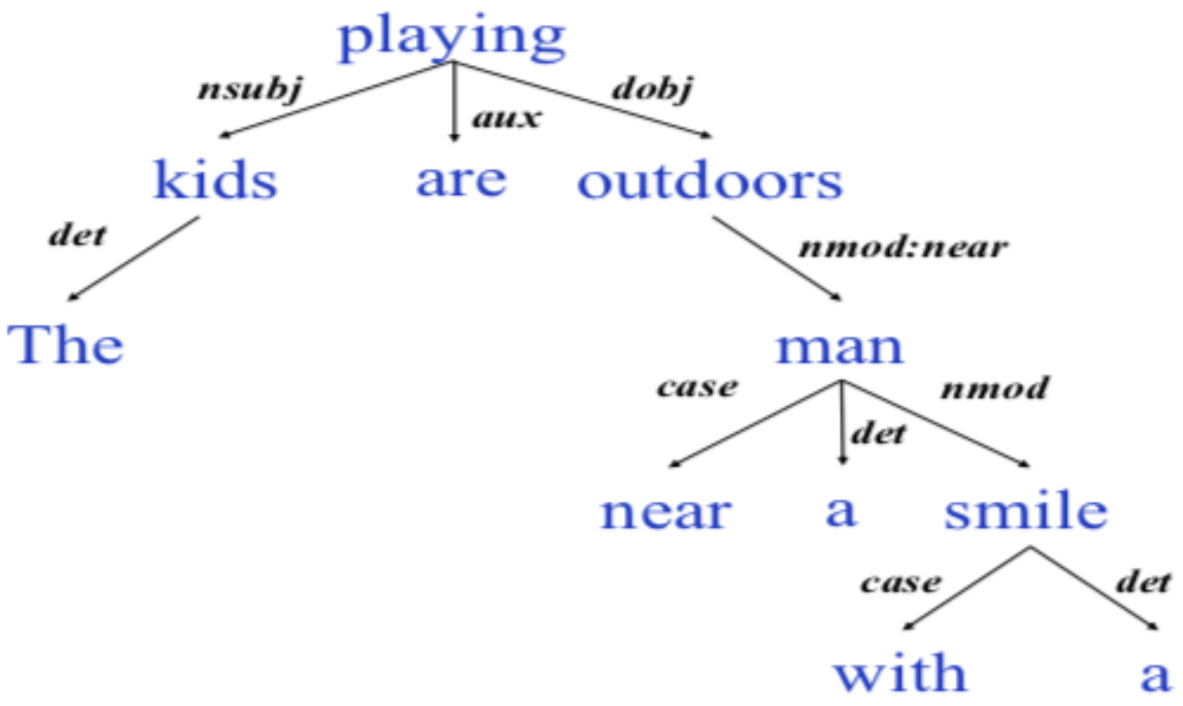

Fig. 1: The dependency tree of the sample sentence: The kids are playing outdoors near a man with a smile.

Architecture

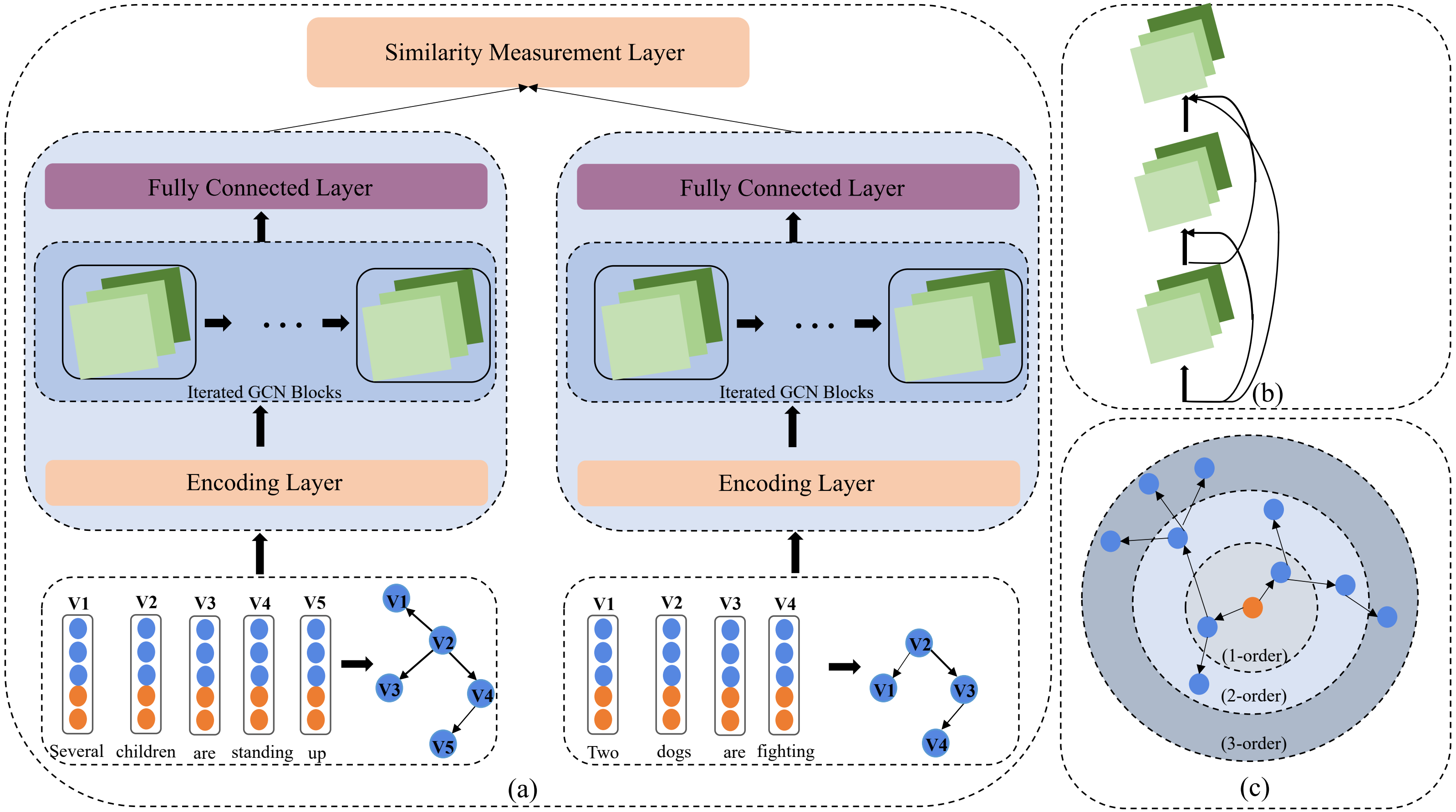

Fig. 2: (a): The overall architecture of GCSTS with an example sentence pair and the dependency trees. GCSTS includes encoding layer, iterated GCN block, fully connected layer and similarity layer. (b): The iterated GCN blocks with dense connections. (c): An example node and its 1-order, 2-order, and 3-order neighbors.

Experiments

上海大学 Shanghai University

➢ Comparison with State-of-the-Arts

Model	STS12	STS13	STS14	STS15	STS16	STS-B	Avg.
Avg. GloVe embeddings (Pennington et al., 2014))	0.547	0.711	0.585	0.653	0.630	0.575	0.617
Multi-Perspective-CNN (He et al., 2015)	0.571	0.725	0.627	0.702	0.666	0.594	0.647
Dependency Tree-LSTM (Tai et al., 2015)	0.583	0.759	0.649	0.731	0.685	0.616	0.671
Constituent Tree-LSTM (Tai et al., 2015)	0.577	0.746	0.645	0.727	0.673	0.607	0.663
Siamese LSTM (Mueller et al., 2016)	0.580	0.758	0.644	0.724	0.689	0.610	0.668
InferSent (Conneau et al., 2017)	0.599	0.763	0.650	**0.748**	0.694	0.625	0.679
Avg. BERT embeddings (Devlin et al., 2019)	0.520	0.693	0.584	0.650	0.615	0.562	0.604
Text-GNN (Huang et al., 2019)	0.603	0.771	0.656	0.736	0.710	0.624	0.683
GCSTS	**0.615**	**0.782**	**0.669**	0.742	**0.715**	**0.637**	**0.693**

TABLE I: STS12-STS16: SemEval 2012-2016 datasets. STS-B: STSbenchmark dataset. We report the Spearman correlation coefficient in this work. The best results are bold.

Experiments

➢ Comparison with State-of-the-Arts

Model	Accuracy	F1
Avg. GloVe embeddings (Pennington et al., 2014))	0.721	0.749
Multi-Perspective-CNN (He et al., 2015)	0.752	0.817
Dependency Tree-LSTM (Tai et al., 2015)	0.769	0.824
Constituent Tree-LSTM (Tai et al., 2015)	0.760	0.803
Siamese LSTM (Mueller et al., 2016)	0.767	0.820
InferSent (Conneau et al., 2017)	0.762	0.831
Avg. BERT embeddings (Devlin et al., 2019)	0.718	0.737
Text-GNN (Huang et al., 2019)	0.776	0.835
GCSTS	**0.785**	**0.854**

TABLE II: Experimental results on the MRPC dataset. The best results are bold.

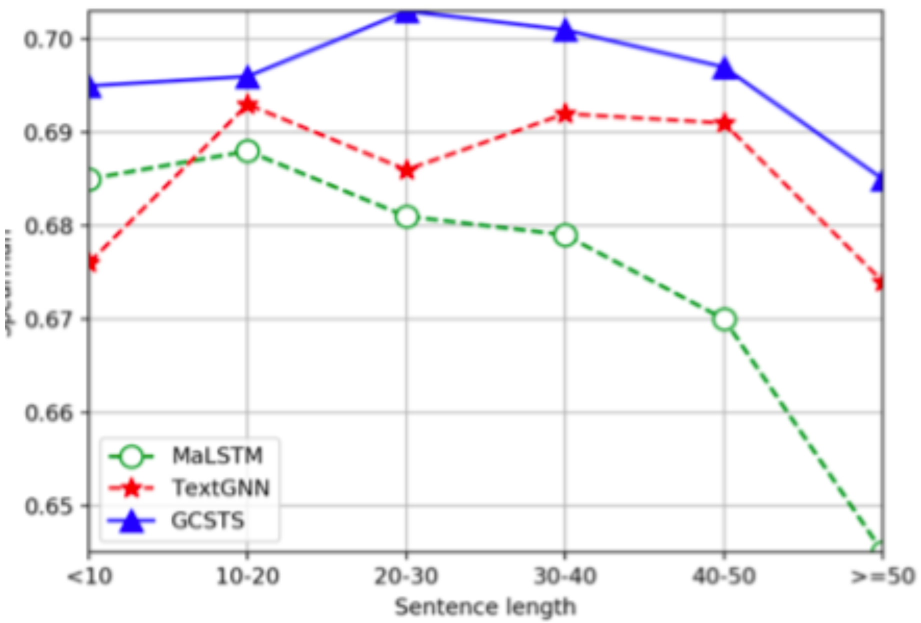

Fig. 3: Comparison of GCSTS, MaLSTM, and Text-GNN against different sentence lengths.

Experiments

上海大学 Shanghai University

Ablation Studies

GCSTS	STS15(p)	MRPC(Acc.)
(1)- Iterated GCN blocks	0.723	0.765
(2)- Dynamic graph structures	0.730	0.769
(3)- BiLSTM encoding layer	0.720	0.774
(4)- POS embedding	0.742	0.784

TABLE III: An ablation study for the proposed model.

- In (1), removing iterated GCN blocks means we use one block (2-layer GCN).
- In (2), we use a fixed graph structure in each layer of the GCN block.
- In (3), we remove the encoding layer from our model.
- In (4), we remove the part-of-speech embeddings.

Conclusion

- **Context Aware Hierarchical Feature Attention Network**

  - A novel GCSTS that apply iterated GCN blocks with dynamic graph structures, which learns better sentence representations in short text similarity. Com- bined with dense connections, GCSTS is able to capture local and non-local interactions of texts.
  - A new way to train deeper GCNs successfully.
  - Extensive experiments on the on seven challenging semantic textual similarity datasets that include different domains demonstrate that proposed model can learn better text representations.

Thanks for listening ！

Xiaoqi Sun

Shanghai University