Instructions for COLING-2020 Proceedings

Anonymous COLING submission

Abstract

1 Introduction

2 Method Overview

3 Experiments

4 Related Work

4.1 Ben

Very similar to what we want to do: (Rupnik et al., 2016; Miranda et al., 2018; Wang et al., 2018; Germann and BBC, 2019; Seki, 2018; Seki, 2020; Linger and Hajaiej, 2020)

Brexit case study: (Peterlin and others, 2019)

Also similar, but involving search terms; probably just good for citations and not techniques: (Rupnik et al., 2016)

Multilingual BERT:

(K et al., 2020)

(Pires et al., 2019)

More multilingual models non-BERT from google:

https://ai.googleblog.com/2019/07/multilingual-universal-sentence-encoder.html

https://arxiv.org/abs/1807.11906 https://arxiv.org/abs/1810.12836 https://arxiv.org/abs/1902.08564 https://arxiv.org/abs/1906.08401 https://arxiv.org/abs/1907.04307

Old survey but good on non deep learning techniques of the era: (Oard and Dorr, 1998).

4.2 Stefanos

Datasets:

(Liu et al., 2019)

(Mohammad et al., 2018)

Pretrained models:

(Puri et al., 2018)

(Kant et al., 2018)

Examples of large scale sentiment analysis:

(Mohammad et al., 2015)

(Hemmatian and Sohrabi, 2017)

(Yang et al., 2015)

4.3 Nate

Look at the way these papers run experiments:
(Tifrea et al., 2018; Meng et al., 2019)
Datasets at: https://aclweb.org/aclwiki/Similarity_(State_of_the_art)
Use wikidata to automatically generate classes: https://www.wikidata.org/wiki/Q43689 https://pywikidata.readthedocs.io/en/latest/
Can we use wikidata to automatically generate good ?

5 Discussion

References

Reviewers Ulrich Germann and Chris Hernon BBC. 2019. Scalable understanding of multilingual media (summa).

Fatemeh Hemmatian and Mohammad Karim Sohrabi. 2017. A survey on classification techniques for opinion mining and sentiment analysis. *Artificial Intelligence Review*, pages 1–51.

Karthikeyan K, Zihan Wang, Stephen Mayhew, and Dan Roth. 2020. Cross-lingual ability of multilingual bert: An empirical study. In *International Conference on Learning Representations*.

Neel Kant, Raul Puri, Nikolai Yakovenko, and Bryan Catanzaro. 2018. Practical text classification with large pre-trained language models. *arXiv preprint arXiv:1812.01207*.

Mathis Linger and Mhamed Hajaiej. 2020. Batch clustering for multilingual news streaming. *arXiv preprint arXiv:2004.08123*.

Chen Liu, Muhammad Osama, and Anderson de Andrade. 2019. Dens: A dataset for multi-class emotion analysis. *EMNLP*.

Yu Meng, Jiaxin Huang, Guangyuan Wang, Chao Zhang, Honglei Zhuang, Lance Kaplan, and Jiawei Han. 2019. Spherical text embedding. In *Advances in Neural Information Processing Systems*, pages 8206–8215.

Sebastião Miranda, Artūrs Znotiņš, Shay B Cohen, and Guntis Barzdins. 2018. Multilingual clustering of streaming news. *arXiv preprint arXiv:1809.00540*.

Saif M Mohammad, Xiaodan Zhu, Svetlana Kiritchenko, and Joel Martin. 2015. Sentiment, emotion, purpose, and style in electoral tweets. *Information Processing & Management*, 51(4):480–499.

Saif M. Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018. Semeval-2018 Task 1: Affect in tweets. In *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, New Orleans, LA, USA.

Douglas W Oard and Bonnie J Dorr. 1998. A survey of multilingual text retrieval. Technical report.

Sebastian Peterlin et al. 2019. Detecting influence in media with brexit as case study.

Telmo Pires, Eva Schlinger, and Dan Garrette. 2019. How multilingual is multilingual bert? *arXiv preprint arXiv:1906.01502*.

Raul Puri, Robert Kirby, Nikolai Yakovenko, and Bryan Catanzaro. 2018. Large scale language modeling: Converging on 40gb of text in four hours. In *2018 30th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD)*, pages 290–297. IEEE.

Jan Rupnik, Andrej Muhic, Gregor Leban, Primoz Skraba, Blaz Fortuna, and Marko Grobelnik. 2016. News across languages-cross-lingual document similarity and event tracking. *Journal of Artificial Intelligence Research*, 55:283–316.

Kazuhiro Seki. 2018. Exploring neural translation models for cross-lingual text similarity. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 1591–1594.

Kazuhiro Seki. 2020. Cross-lingual text similarity exploiting neural machine translation models. *Journal of Information Science*, page 0165551520912676.

Alexandru Tifrea, Gary Bécigneul, and Octavian-Eugen Ganea. 2018. Poincar\'e glove: Hyperbolic word embeddings. *arXiv preprint arXiv:1810.06546*.

Zhouhao Wang, Enda Liu, Hiroki Sakaji, Tomoki Ito, Kiyoshi Izumi, Kota Tsubouchi, and Tatsuo Yamashita. 2018. Estimation of cross-lingual news similarities using text-mining methods. *Journal of Risk and Financial Management*, 11(1):8.

Steve Y Yang, Sheung Yin Kevin Mo, and Anqi Liu. 2015. Twitter financial community sentiment and its predictive relationship to stock market movement. *Quantitative Finance*, 15(10):1637–1656.