

Veronika Matek, Karmen Frank, and Luka Mihelič

[illegible]

Domain Adaptation, TSDAE, GPL, SBERT, Sentence Classification

For the purpose of classifying Slovenian sentences based on their sentiment we fine-tune the SBERT model with TS-

DAE. We choose bert-base-uncased (TODO) for our base model to fine-tune. For the loss function we use the DenoisingAutoEncoderLoss as our loss function when training. We train the method and compare our results with the corpus.

TSDAE has been shown by Wang, Reimers and Gurevych [1] to outperform other unsupervised approaches and other supervised models, trained with a lot of labeled data. Many previous works were evaluated on Semantic Textual Similarity (STS) which might return good performance but it is unclear how it performs on specific domains.

GPL

GPL - opis metode, pri čemer se predvsem osredotočimo na domain adaptation. Povemo kateri model finetunamo. Ddamo skico arhitekture.

Data

Kakšne podatke uporabljamo, kako izgledajo, in what way did you prepare the data, delitev na množice (poudarimo, da se vse metode treniranje z enako učno množico). Pokažemo morda par primerov povedi v tabeli.

Testing approach

Naslov morda še ni ustrezen in se bo prilagodil. Katero metriko uporabimo za primerjavo rezultatov, kako iz sentence embedding pridemo do klasifikacije povedi.

Results

TO DO: Use the results section to present the final results of your work. Present the results in a objective and scientific

fashion. Use visualisations to convey your results in a clear and efficient manner. When comparing results between various techniques use appropriate statistical methodology.

Discussion

TO DO: Use the Discussion section to objectively evaluate your work, do not just put praise on everything you did, be critical and exposes flaws and weaknesses of your solution. You can also explain what you would do differently if you would be able to start again and what upgrades could be done on the project in the future.

Acknowledgments

Here you can thank other persons (advisors, colleagues ...) that contributed to the successful completion of your project.

References

- [1] Kexin Wang, Nils Reimers, and Iryna Gurevych. TSDAE: Using transformer-based sequential denoising auto-encoder for unsupervised sentence embedding learning. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 671–688, Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics.