Supervisor's Review of Diploma Thesis

Title of the thesis: Multilingual Multimodal Detection of Humour in Stand-Up Comedy

Author: Anna Kuznetsova

---

1. Task specification

The submitted thesis focuses on automatic multimodal humour detection in the specific domain of stand-up comedy. The assigned goal was to explore humour detection in the given domain, and also to experiment with multilingual settings to investigate potential differences in humour modalities across cultures. The student was expected to collect a new dataset for multimodal humour analysis consisting of stand-up comedy videos, not only in English, but also in Russian, and then to automatically annotate the dataset on humorous messages using laughter from the audience. Such a labeled collection should serve for further experiments with both unimodal and multimodal humour detection, which should be conducted as well.

Certainly, multimodal humour detection is a challenging task that has recently been studied in different domains, mainly on English datasets. To gain a better understanding of the specific modalities and features that contribute to humour detection and to explore the cross-cultural aspects of humour in different languages has been a good motivation for this work [p. 12].

2. Thesis structure

The submitted thesis is well and clearly structured. Its structure, containing Introduction, five inner chapters, and Conclusion, is quite standard. In Introduction the author explains the scope and the structure of her work. In the following Chapter 1 she reviews related literature, gives a detailed overview of related research, and at the end she defines the objectives of the thesis. Then Chapter 2 provides a description of her new datasets including data preprocessing procedures like manual annotation, alignment, segmentation, and algorithms for laughter detection. Chapter 3 quite briefly introduces models for humour detection, their implementation, and the conducted experiments. In Chapter 4 the author discusses her experimental results, and finally in Chapter 5 she gives short discussions on several topics related to her research findings: the quality of labelling, laughter detection approaches, advanced multimodal modelling, multilinguality, and ethical considerations. In her Conclusion, she summarizes main findings and contributions of the thesis, and possible future work.

The thesis contains an extensive list of used literature and two Appendices: (A) list of sourced Russian and English videos, and (B) description of electronic attachments. The newly created data collection, whose structure is clearly described in Appendix B.1, is available at Google Drive repository. The scripts used for the implemented experiments are available and documented at GitHub, and their functionality is briefly described in Appendix B.2.

3. Formal aspects

The submitted thesis fulfills all general formal requirements, and is written in good, fluent English with only few and minor flaws.

4. Evaluation and questions for the defense

The submitted work is definitely interesting. In my opinion, the research parts, especially in Chapter 1 and discussions in Chapter 5, are a strong point of the thesis. Also, I appreciate author's complex approach to the task. She devoted a lot of time and work to collecting and annotating the novel multilingual dataset for humour detection, with primary focus on the Russian language. The preparation of the experimental dataset belongs to main contributions of the thesis. Her inchoative experiments with humour detection in stand-up comedy videos are certainly an interesting contribution to the field, as well.

Very detailed experiments were done with procedures for automatic laughter detection (in Chapter 2.5), which is still a very challenging task. However, the laughter detection optimization and evaluation process is not described clearly. To me it is unclear which data were used for parameter optimization, and which data then used for evaluation (F-score measuring)?

On the other hand, the experimental part focused on the main goal of the thesis (in Chapters 3 and 4) is quite brief. Missing description of (hyper)parameters tuning seems to be a weak spot. Unfortunately, the multilingual experiments, which belongs to substantial parts of the assignment, are only rudimentary (only half a page in Chapter 4.3).

Questions for the defense:

- The relatively low performance of the automatic labeling of laughter presence (F-score 71% for Russian and 77% for English data) seems to be a crucial point. — Could you discuss its consequences for further machine learning process, in which in fact you use the relatively unreliable labels as a gold standard data set, both for training and for testing?

- Would it make sense to compare your performance with the TBBT dataset (which is not reported in 1.3.1)?

5. Conclusion

I highly appreciate that the student has contributed to quite a new research area, namely multimodal humor detection. My overall evaluation is that this thesis, as a whole, is definitely a nice piece of work, with rich potential for further research.

Although the imperfections mentioned above are certainly not negligible, I recommend the submitted thesis for the defense and I suggest accepting this work as a diploma thesis.

Práci doporučuji k obhajobě.
Práci nenavrhuji na zvláštní ocenění.

Prague, August 2023

RNDr. Martin Holub, Ph.D.

Institute of Formal and Applied Linguistics
Charles University in Prague