7 Essential Principles to Make Multimodal Sentiment Analysis Work in the Wild

Björn W. Schuller[1,2]
[1]Department of Computing, Imperial College London, United Kingdom
[2]Chair of Complex & Intelligent Systems, University of Passau, Germany
bjoern.schuller@imperial.ac.uk

Abstract

Sentiment Analysis is increasingly carried out on multimodal data such as videos taken in everyday environments. This requires robust processing across languages and cultures when aiming for mining of opinions from the 'many'. Here, seven key principles are laid out to ensure a high performance of an according automatic approach.

1 Introduction

Sentiment Analysis (SA) recently found its way beyond pure text analysis [Cambria *et al.*, 2015], as sentiment is increasingly expressed also via video 'micro blogs', short clips, or other forms. Such multimodal data is usually recorded 'in the wild' thus challenging today's automatic analysers. For example, one's video-posted opinion on a movie may contain scenes of this movie, requiring subject tracking, and music in the background may need to be overcome for speech recognition and voice analysis. Here, I provide 'essential principles' to make a multimodal SA work despite such challenges.

2 The Principles

Seven selected recommendations to make a multimodal SA system 'ready for the wild' are given with a short statement :

Make it Multimodal – But Truly. Multimodal SA is often carried out in a late fusion manner, as e. g., (spoken) language, acoustics, and video analysis operate on different time levels and monomodal analysers prevail. However, recent advances in synergistic fusion allow for further exploitation of heterogeneous information streams such as analysis of cross-modal behaviour-synchrony to reveal, e. g., regulation.

Make it Robust. Robustness is a obviously key handling real-world data. Effective denoising and dereverberation can these days be reached by data-driven approaches such as (hierarchical) deep learning. Beyond, recognition of occlusions, background noises, and alike should be used in the fusion to dynamically adjust weights given to the modalities

Train it on Big Data. A major bottleneck for SA beyond textual analysis is the lack of suited 'big' (ideally multimodal) training data. While data is usually 'out there' (such as videos on the net), it is the labels that lack. Recent cooperative learning approaches such as by dynamic active learning and semi-supervised learning combined with (gamified) crowdsourcing can help to efficiently build a large training corpus. Smart pre-selection of suited material from large resources can further improve efficiency.

Make it Adapt. It has repeatedly been shown that in-domain training improves SA [Cambria *et al.*, 2015]. Recent transfer learning provides a range of solutions to adapt to a new domain even if only little and/or unlabelled data should exist from it. Subject adaptation is another key aspect.

Make it Context-aware. Temporal context modelling can be learnt, e. g., by LSTM recurrent networks. Additional exploitation of knowledge sources can help resolve ambiguities. Also, automatic recognition of further traits and states of the subject expressing sentiment such as age, gender, ethnicity, personality, or emotion, and health state can add important information regarding the sentiment expressed.

Make it Multilingual. It seems obvious that multilingualism is an issue for text-based SA. However, it is as well for acoustic analysis, and can in principle even influence video-based SA due to, e. g., varying lip-movements. In fact, languages are often even mixed in real-world statements.

Make it Multicultural. Cross-cultural SA has been researched comparably little, albeit it is clearly of crucial importance. Just as for multilingualism, a key requirement will be sufficient learning data. Then, models can be switched, and transferred across languages and cultures.

3 Conclusion

Seven major requirements were highlighted on the way to truly robust multimodal sentiment analysis in adverse conditions in today's 'big data' era. Beyond these, a number of further issues need to be addressed to best exploit automated sentiment analysis such as provision of meaningful confidence measures and optimal exploitation of 'internal' confidences.

Acknowledgments

The author was supported by the EU's H2020 Programme via the IAs # 644632 (MixedEmotions) and # 645094 (SEWA).

References

[Cambria *et al.*, 2015] E. Cambria, B. Schuller, and Y. Xia. New Avenues in Opinion Mining and Sentiment Analysis. In *Proc. IJCAI 2015*, Buenos Aires, 2015.

Proceedings of the 4th Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2016), IJCAI 2016, page 1, New York City, USA, July 10, 2016.