Bilingual Sentiment Embeddings: Joint Projection of Sentiment Across Languages

Jeremy Barnes¹, Roman Klinger¹, Sabine Schulte im Walde¹
University of Stuttgart¹, Universitat Pompeu Fabra²

Motivation
Goal: How can you create bilingual word representations for cross-lingual sentiment analysis when you have little parallel data?
Problem: There is no labeled sentiment data or large amounts of parallel data available in most languages in the world.
Approach: We jointly train the projection and sentiment classifier.

Experimental Setup
• Sentence-level Cross-lingual Sentiment Analysis on hotel reviews: the OpeNER and MultiBooked Corpora. This avoids additional domain problems.
• Three language pairs:
  – English → Spanish (ES)
  – English → Catalan (CA)
  – English → Basque (EU)
• Binary and 4-class.
• Train on the English annotated sentiment corpus only. Test on the target language test set.
• Compare to two bilingual embedding methods with similar data requirements (ARTETXE, BARISTA) and Machine Translation (MT).

Results for Cross-lingual Sentiment Classification
• BLSE outperforms other projection methods on macro F₁ ($p < 0.05$ highlighted) and is close to MT.
• ENSEMBLE of BLSE and MT performed better than other combinations (not shown here).

Future Work
• Move our approach to aspect-level.
• Include multi-word phrases into projection to improve negation and modifiers.