On the Effect of Word Order on Cross-lingual Sentiment Analysis

Àlex Ramrez Atrio[1], **Toni Badia**[1] **Jeremy Barnes**[2]
[1]Universitat Pompeu Fabra
[2]Universitetet i Oslo
Informacin de contacto

Abstract

This document contains the instructions for preparing a camera-ready manuscript for the proceedings of NAACL-HLT 2019. The document itself conforms to its own specifications, and is therefore an example of what your manuscript should look like. These instructions should be used for both papers submitted for review and for final versions of accepted papers. Authors are asked to conform to all the directions reported in this document.

1 Introduction

- intro to task: Why is sentiment analysis cool/useful/difficult?

  When facing a relatively simple text such as an hotel review, we can ask for its general sentiment. Is it positive, or negative? Is it better to ask for more gray areas in between? Automatizing the process of classifying the sentiment of a text is called Sentiment Analysis (SA), and it can allow us to get a good understanding of how the author(s) of these texts feel about the topics that are discussed in them. Of course, there are issues that we may find in this process. For instance, given a certain text, its sentiment may be ambiguous: two independent human annotators may disagree in it. Or maybe the text does not have an overall sentiment, and we should focus on individual sentences.

Jeremy
Could you add a bit more specifics here? For example, why it's hard, why we need annotated data, possible differences between languages? Try to think of some examples. If you can't, don't worry :).

- motivation for cross-lingual approaches: We often have no annotated data for Language X, especially for specific domains.

  Cross-lingual Sentiment Analysis (CLSA) consists on using resources such as labeled data of a high-resource language (our source language) to train a sentiment classifier in order for it to classify low-resource languages (our target languages). This can be important when our target language lacks plentiful labeled data, particularly when considering specific domains.

- why it's interesting to use no MT: under-resourced languages, MT requires too much parallel data

  This process can be carried out using Machine Translation (MT) of the source language and training a classifier using these translated texts. The main problem with this approach is its high requirements of parallel annotated data, which may be difficult to find for some low-resource languages. Bilingual Sentiment Embeddings (should I use 'BLSE' as acronym?), on contrast, have been shown to be competitive while requiring less amount of parallel annotated data.

- what problem that might introduce

Alex
I am not sure what do you have in mind here. Problems introduced by not using MT in general? Problems introduced by using blse?

Jeremy
I was thinking specifically to include the problem of differences in word order between languages. You could explain that current state-of-the-art sentiment models are heavily influenced by word order, as they are usually pretrained language models which are then fine-tuned for sentiment, *i. e.* (Peters et al., 2018; Howard and Ruder, 2018; Devlin et al., 2018)

2 Related Work

- Cross-lingual Sentiment Approaches that are relevant here: under-resourced langs
- Bilingual Word Embeddings: Artetxe and why we use these: SOTA and low-resource
- Word order in sentiment
- Reordering for machine translation

Alex
Should I introduce in this section the specifics? That our source language is

Because of some problems we may find in MT, sometimes it is considered best to preprocess the source language by reordering it and then carryig out the translation. Transformation (reordering) rules can be determined manually or with data-driven approaches. Their application can be deterministic or non-deterministic. Some hybrid techniques exist as well, where long-range transformations are deterministic and the rest non-deterministic.

Jeremy
Could you cite relevant work for deterministic, non-deterministic and hybrid and give some idea of the relevance to our work?

2.1 Cross-lingual Sentiment Analysis

(Mohammad et al., 2015)

2.2 Bilingual Word Embeddings

3 Methodology

Alex
What is the difference between this subsection an the last subsection on blse?

Jeremy
Not too much, but if you could use your words to describe it, it would be very useful.

3.1 Models

- LSTM, CNN, SVM
- Differences between how models handle word order

For our experiment we will compare the results of three different classifiers: a Long Short Term Memory Network (LSTM), a Convolutional Neural Network (CNN), and a Support Vector Machine (SVM) with Bag-of-Embeddings. The SVM does not take into account word order, the CNN considers only short-range word order, and the LSTM considers both short-term and long-term word order.

Jeremy
Could you cite the relevant papers for each of these and give more details on their assumptions regarding word order? It would also be useful to explain the training process and hyperparameter optimization.

3.2 Corpora and Datasets

We use a subset of the English and Spanish OpeNER corpora of hotel reviews. The corpora are annotated for Part-of-Speech tags and sentiment with 4 classes. We use the English subset for training our classifiers and the Spanish for testing different reorderings.

We also use MultiBooked, an annotated corpus of hotel reviews in Catalan. The corpus is also annotated for POS tags and we will use it to test different reorderings.

Alex
Here I am saying "reordering", but it is not only reorderings. Do you have a better word for the tests texts?

Jeremy
We explain the different setups and

3.3 Experimental Setup

- Test all models on two cross-lingual setups (en-es, en-ca)

Alex
Maybe next item should be before this one?

Jeremy
I agree.

The experiment consists on training the LSTM, CNN, and SVM with English data, and test them on different reorderings of both Spanish and Catalan corpora.

- Compare: No reordering, Random Reordering, N-ADJ, reordering-crego, No lexicon, Only lexicon

Alex
Here we should also talk about the Machine Translation parts? Here we introduce the lexicon thing, but we have not talked about it yet, right?

Jeremy
Yes, let's introduce all of that here.

We compare the original texts of the Catalan MultiBooked and Spanish OpeNER, a random reordering of these, a simple reordering consisting of the application of the the rule N-ADJ to ADJ-N, a reordering resulted of the application of 15 transformation rules extracted from Crego and Mario 2006a and 2006b, a version of the corpora with all the words appearing in their respective lexicon deleted, and a version of the corpora with exclusively the words appearing in their respective lexicon.

- What are the competing hypotheses for each of these setups?

Alex
Here do you mean to list different prediction for every of these setups, like: we expect n-adj to perform slightly better than the original, random to be worse, ...?

Jeremy
Yes I do

4 Results

5 Analysis

Alex
Falta omplir-les

Jeremy
It would be nice to show that the noise introduced by bilingual embeddings leads to LSTMs not being able to pick up on word order in the target language. We could train monolingual models for Spanish and Catalan and use the random reordering to see the difference.

Todo
It would also be interesting to look at particular examples of errors that

6 Conclusion and Future Work

References

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805.

Jeremy Howard and Sebastian Ruder. 2018. Universal language model fine-tuning for text classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 328–339. Association for Computational Linguistics.

Saif M Mohammad, Mohammad Salameh, and Svetlana Kiritchenko. 2015. How Translation alters sentiment. *Journal of Art*, 54:1–20.

			Binary	4-class
Bilingual Word Embeddings	EN-ES	Original		
		Reordered		
		N-ADJ		
		Random		
		Only Lexicon		
		No Lexicon		
	EN-CA	Original		
		Reordered		
		N-ADJ		
		Random		
		Only Lexicon		
		No Lexicon		
Mono	EN	Original		
		Random		
		Only Lexicon		
		No Lexicon		
Machine Translation	EN	Original		
		Random		
		Only Lexicon		
		No Lexicon		
	EN	Original		
		Random		
		Only Lexicon		
		No Lexicon		

Table 1: Macro F_1 results for all corpora and techniques. We denote the best performing projection-based method per column with a *blue box* and the best overall method per column with a **green box**.

Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237. Association for Computational Linguistics.

A Appendices

Appendices are material that can be read, and include lemmas, formulas, proofs, and tables that are not critical to the reading and understanding of the paper. Appendices should be **uploaded as supplementary material** when submitting the paper for review. Upon acceptance, the appendices come after the references, as shown here. Use `\appendix` before any appendix section to switch the section numbering over to letters.