

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

Anonymous NAACL submission

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.

- 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

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

- Should I introduce in this section the specifics? That our source language is English, our targets Spanish and Catalan and that we will be using embeddings?

2 Related Work

- Cross-lingual Sentiment Approaches that are relevant here: under-resourced langs

I am not sure what do you mean here. Abdallah and hirst 2017, chen et al 2016 and barnes et al acl18?

- Bilingual Word Embeddings: Artetxe and why we use these: SOTA and low-resource

Copying the same section in the acl18 paper?

- Word order in sentiment

Do you have any specific paper in mind? Because I don't think I found any literature in this in my thesis.

- Reordering for machine translation

Because of some problems we may find in MT, sometimes it is considered best to preprocess the source language by reordering it and then carrying out the translation. Transformation (re-ordering) 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.

2.1 Cross-lingual Sentiment Analysis

(Mohammad et al., 2015)

2.2 Bilingual Word Embeddings

What is the difference between this subsection and the last subsection on blse?

3 Methodology

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.

3.2 Corpora and Datasets

- OpeNER, Multibooked

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.

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

- Europarl, Tatoeba

I am not sure if we have to mention these. I used them for statistical data, but we are not using it, since the only reordering is CREGO.

- motivation for using these resources

What exactly do you mean by this?

3.3 Experimental Setup

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

(Maybe next item should be before this one?)

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

(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?)

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 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?

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, ...?

4 Results

(Falta omplir-les)

			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*.

5 Analysis

6 Conclusion and Future Work

References

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

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.