**Emoji-Powered Representation Learning for Cross-Lingual Sentiment Classification.**

**1.Annotation**

This work was made by members of Key Lab of High-Confidence Software Technology of MoE (Peking University), University of California in Berkeley, University of California in Los Angeles in Beijing, School of Information of University of Michigan in Ann Arbor.

Their first idea was that there isn’t much of labeled data for sentiment analysis in different languages except English. Usually to escape the lack of labeled data people transferring knowledge from one language with abundant examples (usually English) to language without big amount of them. But this approach can’t capture some important sentiment knowledge specific to the target language.

So, they decide to make cross-lingual sentiment analyzer. For this aim they update text representation and classification with emojis which are unified for each language. The framework named ELSA (Emoji-powered representation learning for cross-Lingual Sentiment Analysis) became the result of their work.

**2. The workflow of ELSA**

Firstly, they train their Word2Vec models on raw tweets of two languages: source language (English) and target language. Then they take emoji tweets and make emoji predictor for both languages. Using RNN and attention layer they at first denote each sentence as (x, e) where x is a sequence of word vectors in this sentence and is emoji in this tweet. Then they make encoded bi-directional vector for every word vector in the sentence. After that they use attention layer to give weight for each word to make emoji prediction more accurate. Then data is going into the softmax layer where the whole sentence representation is transforming into the probability vector and each element of this vector is the probability of the specific emoji.

Training of classificatory is working in this way: after getting representation of sentence on both languages they get representation of documents on both languages and concatenate them into one representation to add this into softmax layer to make emoji-powered real sentiment label of source document.

The algorithm receives as input two representations of text: on target language and on English, that was get after translating target text with Google Translate.

**3. Comparison with other methods**

ELSA according to their research is better than MT-BOW, CL-RL, Bi-DRL. They tested all three algorithms on three languages: Japanese, German and French. All three methods that used translations for getting more accurate results translating using English, so more accurate results were reached with German and French because they are from the same language family, when Japanese from absolutely another. But on Japanese material it still gets the 80% accuracy which is still very good result. Then they compare ELSA with the same algorithm but without emoji-powering. In their paper we can see that ELSA outperforms not only another cross-lingual analyzers, but also itself. It proves that these results were reached not only by updating the algorithm of document embedding, but also because of using emojis as another feature that is important for cross-lingual analysis.

**4.Summary**

This research contains very fresh and interesting idea of versatility of emojis and using them in sentiment analysis. But I couldn’t find out from the paper how they combine emojis with certain labels, which of emojis represents which labels, how much labels there was in the research, was there two classes (for positive and negative), three classes (including neutral), or maybe five and more (make distributions for more positive and more negative). But in general, we can see how ELSA outperforms other methods and how emojis could positively change the accuracy of such cross-lingual sentiment analysis.