Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20)
Sister Conferences Best Papers Track

Emoji-Powered Representation Learning for Cross-Lingual Sentiment Classification (Extended Abstract)*

Zhenpeng Chen[1], **Sheng Shen**[2], **Ziniu Hu**[3], **Xuan Lu**[4], **Qiaozhu Mei**[4] and **Xuanzhe Liu**[1]
[1]Key Lab of High-Confidence Software Technology, MoE (Peking University), Beijing, China
[2]University of California, Berkeley, USA
[3]University of California, Los Angeles, USA
[4]University of Michigan, Ann Arbor, USA
czp@pku.edu.cn, sheng.s@berkeley.edu, bull@cs.ucla.edu, {luxuan, qmei}@umich.edu,
xzl@pku.edu.cn

Abstract

Sentiment classification typically relies on a large amount of labeled data. In practice, the availability of labels is highly imbalanced among different languages. To tackle this problem, cross-lingual sentiment classification approaches aim to transfer knowledge learned from one language that has abundant labeled examples (i.e., the source language, usually English) to another language with fewer labels (i.e., the target language). The source and the target languages are usually bridged through off-the-shelf machine translation tools. Through such a channel, cross-language sentiment patterns can be successfully learned from English and transferred into the target languages. This approach, however, often fails to capture sentiment knowledge specific to the target language. In this paper, we employ emojis, which are widely available in many languages, as a new channel to learn both the cross-language and the language-specific sentiment patterns. We propose a novel representation learning method that uses emoji prediction as an instrument to learn respective sentiment-aware representations for each language. The learned representations are then integrated to facilitate cross-lingual sentiment classification.

1 Introduction

Sentiment analysis has been a critical component in many applications such as recommender systems [Sun *et al.*, 2019], personalized content delivery [Harakawa *et al.*, 2018], and online advertising [Qiu *et al.*, 2010]. However, existing work on sentiment analysis mainly deals with English texts [Deriu *et al.*, 2017]. Although some efforts have also been made with other languages, sentiment analysis for non-English languages is far behind. This creates a considerable inequality in the quality of the aforementioned Web services received by non-English users, especially considering that 74.8% of Internet users are non-English speakers[1].The cause of this inequality is quite simple: effective sentiment analysis tools are often built upon supervised learning techniques, and *there are way more labeled examples in English than in other languages*.

A straightforward solution is to transfer the knowledge learned from a label-rich language (i.e., the source language, usually English) to another language that has fewer labels (i.e., the target language), an approach known as *cross-lingual sentiment classification* [Chen *et al.*, 2017]. In practice, its biggest challenge is how to fill the linguistic gap between English and the target language. Most recent studies have been using off-the-shelf machine translation tools to generate pseudo parallel corpora and then learn bilingual representations for the downstream sentiment classification task [Xiao and Guo, 2013; Zhou *et al.*, 2016]. More specifically, many of these methods enforce the aligned bilingual texts to share a unified embedding space, and sentiment analysis of the target language is conducted in that space.

Although this approach looks sensible, the performance of these machine translation-based methods often falls short. Indeed, a major obstacle of cross-lingual sentiment analysis is the so-called *language discrepancy* problem [Chen *et al.*, 2017], which machine translation does not tackle well. More specifically, sentiment expressions often differ a lot across languages. Machine translation is able to retain the general expressions of sentiments that are shared across languages (e.g., "*angry*" or "怒っている" for negative sentiment), but it usually loses or even alters the sentiments in language-specific expressions [Mohammad *et al.*, 2016]. As an example, in Japanese, the common expression "湯水のように使う" indicates a negative sentiment, describing the excessive usage or waste of a resource. However, its translation in English, "*use it like hot water*," loses the negative sentiment.

The reason behind this pitfall is easy to explain: machine translation tools are usually trained on parallel corpora that are built in the first place to capture patterns shared across languages instead of patterns specific to individual lan-

*This paper is an abridged version of a paper with the same title, which won the Best Full Paper Award at the WWW 2019 conference [Chen *et al.*, 2019]. This work was supported by the Key-Area Research and Development Program of Guangdong Province under the grant No. 2020B010164002 and the Beijing Outstanding Young Scientist Program under the grant NO. BJJWZYJH01201910001004. Corresponding author: Xuanzhe Liu.

[1]https://www.internetworldstats.com/stats7.htm.

guages. In other words, the problem is due to the failure to retain language-specific sentiment knowledge when unilaterally pursuing generalization across languages. A new bridge needs to be built beyond machine translation, which not only transfers "general sentiment knowedge" from the source language but also captures "private sentiment knowledge" of the target language. *That bridge can be built with emojis*.

In this paper, we tackle the problem of cross-lingual sentiment analysis by employing *emojis* as an instrument. Emojis are considered an emerging ubiquitous language used worldwide [Chen *et al.*, 2018]; in our approach they serve both as a proxy of sentiment labels and as a bridge between languages. Their functionality of expressing emotions motivates us to employ emojis as complementary labels for sentiments, while their ubiquity makes it feasible to learn emoji-sentiment representations for almost every active language. Coupled with machine translation, the cross-language patterns of emoji usage can complement the pseudo parallel corpora and narrow the language gap, and the language-specific patterns of emoji usage help address the language discrepancy problem.

We propose ELSA, a novel framework of *Emoji-powered representation learning for cross-Lingual Sentiment Analysis* [Chen *et al.*, 2019]. Through ELSA, language-specific representations are first derived based on modeling how emojis are used alongside words in each language. These per-language representations are then integrated and refined to predict the rich sentiment labels in the source language, through the help of machine translation. Different from the mandatorily aligned bilingual representations in existing studies, the joint representation learned through ELSA catches not only the general sentiment patterns across languages, but also the language-specific patterns. In this way, the new representation and the downstream tasks are no longer dominated by the source language.

We evaluate the performance of ELSA[2] on a benchmark Amazon review dataset, which covers nine tasks combined from three target languages (i.e., Japanese, French, and German) and three domains (i.e., book, DVD, and music). Results indicate that ELSA outperforms existing approaches on all of these tasks in terms of classification accuracy.

2 The ELSA Approach

We first give a formulation of our problem. Cross-lingual sentiment classification aims to use the labeled data in a source language (i.e., English) to learn a model that can classify the sentiment of test data in a target language. In our setting, besides labeled English documents (L_S), we also have large-scale unlabeled data in English (U_S) and in the target language (U_T). Furthermore, there exist unlabeled data containing emojis, both in English (E_S) and in the target language (E_T). In practice, these unlabeled, emoji-rich data can be easily obtained from online social media such as Twitter. Our task is to build a model that can classify the sentiment polarity of document in the target language solely based on the labeled data in the source language (i.e., L_S) and the different kinds of unlabeled data (i.e., U_S, U_T, E_S and E_T). Finally, we use a held-out set of labeled documents in the target language (L_T), which can be small, to evaluate the model.

The workflow of ELSA is illustrated in Figure 1(a), with the following steps. In *step 1* and *step 2*, we build sentence representation models for both the source and the target languages. In *step 3*, we translate each labeled English document into the target language, sentence by sentence, through *Google Translate*. Both the English sentences and their translations are fed into the representation models learned in steps 1 and 2 to obtain their per-language representations (*step 4* and *step 5*). Then in *step 6* and *step 7* we aggregate these sentence representations back to form two compact representations for each training document, one in English and the other in the target language. In *step 8*, we use the two representations as features to predict the real sentiment label of each document and obtain the final sentiment classifier. In the test phase, for a new document in the target language, we translate it into English and then follow the previous steps to obtain its representation (*step 9*), based on which we predict the sentiment label using the classifier (*step 10*).

2.1 Representation Learning

Representations of documents need to be learned before we train the sentiment classifier. Since emojis are widely used to express sentiments across languages, we learn sentiment-aware representations of documents using emoji prediction as an instrument. Specifically, in a distantly supervised way, we use emojis as surrogate sentiment labels and learn sentence embeddings by predicting which emojis are used in a sentence. This representation learning process is conducted separately in the source and the target languages to capture language-specific sentiment expressions. The architecture of the representation learning model is illustrated in Figure 1(b).

Word Embedding Layer. The word embeddings are pre-trained with the skip-gram algorithm based on either U_S or U_T, which encode every word into a continuous vector space.

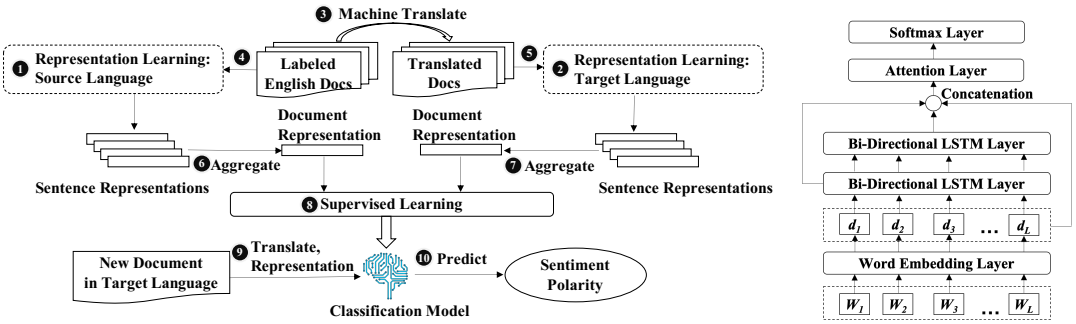
Bi-Directional LSTM Layer. LSTM is particularly suitable for modeling the sequential property of text data. At each step (e.g., word token), LSTM combines the current input and knowledge from the previous steps to update the states of the hidden layer. Let us denote each sentence in E_S or E_T as (x, e), where $x = [d_1, d_2, ..., d_L]$ as a sequence of *word* vectors representing the plain text (by removing emojis) and e as one emoji contained in the text. At each step t, we can extract the latent vector from LSTM. In order to capture the information from the context both preceding and following a word, we use the bi-directional LSTM. We concatenate the latent vectors from both directions to construct a bi-directional encoded vector h_i for every single word vector d_i, which is:

$$h_i = [\overrightarrow{h_i}, \overleftarrow{h_i}].$$

Attention Layer. The attention layer takes the outputs of both the embedding layer and the two LSTM layers as input, through a skip-connection, which enables unimpeded information flow in the whole training process. The i-th word of the input sentence can be represented as u_i:

$$u_i = [d_i, h_{i1}, h_{i2}],$$

[2]The benchmark datasets, scripts, and pre-trained models are available at https://github.com/sInceraSs/ELSA.



(a) The workflow of ELSA.

(b) Network architecture for representation learning through emoji prediction.

Figure 1: The ELSA approach.

where d_i, h_{i1}, and h_{i2} denote the encoded vectors of words extracted in the word embedding layer and the first and second bi-directional LSTMs, respectively. Since not all words contribute equally to predicting emojis or expressing sentiments, we employ the attention mechanism [Bahdanau *et al.*, 2014] to determine the importance of every single word. The attention score of the i-th word is calculated by

$$a_i = \frac{exp(W_a u_i)}{\sum_{j=1}^{L} exp(W_a u_j)},$$

where W_a is the weight matrix used by the attention layer. Then each sentence can be represented as the weighted sum of all words in it, using the attention scores as weights. We denote the sentence representation as v.

Softmax Layer. The sentence representation is then transferred into the softmax layer, which returns a probability vector Y. Each element of this vector indicates the probability that this sentence contains a specific emoji. Finally, we learn the model parameters by minimizing the cross entropy between the output probability vectors and the one-hot vectors of the emoji contained in each sentence. After learning the parameters, we can extract the output of the attention layer to represent each input sentence. Through this emoji-prediction process, words with distinctive sentiments can be identified, and the plain text surrounding the same emojis will be represented similarly. Given the fact that the sentiment labels are limited, once the emoji-powered sentence representations are trained, they are locked in the downstream sentiment prediction task to avoid over-fitting.

2.2 Training the Sentiment Classifier

Based on the pre-trained, per-language sentence representations, we then learn document representations and conduct cross-lingual sentiment classification. First, for each English document $D_s \in L_S$, we use the pre-trained English representation model to embed every single sentence in it. Second, we aggregate these sentence representations to derive a compact document representation. Because different parts of a document contribute differently to the overall sentiment, we once again adopt the attention mechanism here. Specifically, we calculate each document vector r_s as the weighted sum of all sentence vectors in it. Next, we use *Google Translate* to translate D_s into the target language (D_t). We then leverage the pre-trained target-language representation model to form representations for each translated document following the same process above. Supposing the text representations of D_s and D_t are r_s and r_t respectively, we concatenate them into a joint representation $r_c = [r_s, r_t]$, which contains sentiment knowledge from both English and the target language, ensuring that our model is not dominated by the labeled English documents. Finally, we input r_c into an additional softmax layer to predict the real sentiment label of D_s.

Language	English	Japanese	French	German
Raw Tweets	39.4M	19.5M	29.2M	12.4M
Emoji-Tweets	6.6M	2.9M	4.4M	2.7M

Table 1: The sizes of the Tweets and emoji-Tweets.

2.3 Sentiment Classification for Target Language

When we receive an unlabeled document in L_T, we first translate it into English. Based on the representation models trained above, the original document and its English translation can be represented as r_t and r_s. We represent this document as [r_s, r_t] and input it into the classifier, which outputs a predicted sentiment polarity.

3 Evaluation

3.1 The Dataset

The labeled data (L_S and L_T) used in our work are from the Amazon review dataset [Web, 2010]. It covers four languages (i.e., English, Japanese, French, and German) and three domains (i.e., book, DVD, and music). For each combination of language and domain, the dataset contains 1,000 positive reviews and 1,000 negative reviews. We select English as the source language and the other three as the target languages. Therefore, we can evaluate our approach on nine tasks in total (i.e., combinations of the three domains and three target languages). For each task, we use the 2,000 labeled English reviews in the corresponding domain for training and the 2,000 labeled reviews in the target language for evaluation.

Language	Domain	MT-BOW	CL-RL	BiDRL	ELSA
Japanese	Book	0.702	0.711	0.732	**0.783 (0.003)**
	DVD	0.713	0.731	0.768	**0.791 (0.004)**
	Music	0.720	0.744	0.788	**0.808 (0.005)**
French	Book	0.808	0.783	0.844	**0.860 (0.002)**
	DVD	0.788	0.748	0.836	**0.857 (0.002)**
	Music	0.758	0.787	0.825	**0.860 (0.002)**
German	Book	0.797	0.799	0.841	**0.864 (0.001)**
	DVD	0.779	0.771	0.841	**0.861 (0.001)**
	Music	0.772	0.773	0.847	**0.878 (0.002)**

Table 2: The accuracy of ELSA (standard deviations in parentheses) and baseline methods on the nine benchmark tasks.

Language	Domain	N-ELSA	T-ELSA	S-ELSA	ELSA
Japanese	Book	0.527*	0.742*	0.753*	0.783
	DVD	0.507*	0.756*	0.766*	0.791
	Music	0.513*	0.792*	0.778*	0.808
French	Book	0.505*	0.821*	0.850*	0.860
	DVD	0.507*	0.816*	0.843*	0.857
	Music	0.503*	0.811*	0.848*	0.860
German	Book	0.513*	0.804*	0.848*	0.864
	DVD	0.521*	0.790*	0.849*	0.861
	Music	0.513*	0.818*	0.863*	0.878

* indicates the difference between ELSA and its simplified versions is statistically significant ($p < 0.05$) by McNemar's test.

Table 3: Performance of ELSA and its simplified versions.

To achieve unlabeled data (U_S and U_T), we collect a sample of English, Japanese, French, and German Tweets between September 2016 and March 2018. All collected Tweets are used to train the word embeddings. As emojis are widely used on Twitter, we can extract emoji-labeled Tweets, which are used to learn emoji-powered sentence representations. For each language, we extract Tweets containing the top 64 emojis used in this language. As many Tweets contain multiple emojis, for each Tweet, we create separate examples for each unique emoji in it. The emoji-Tweets provide the E_S and E_T datasets, whose statistics are presented in Table 1.

3.2 Baselines and Accuracy Comparison

To evaluate the performance of ELSA, we employ three representative baseline methods for comparison:

MT-BOW uses the bag-of-words features to learn a linear sentiment classifier on the labeled English data [Prettenhofer and Stein, 2010]. It uses *Google Translate* to translate the test data into English and then applies the learned classifier.

CL-RL is a word-aligned representation learning method [Xiao and Guo, 2013]. It first uses *Google Translate* to create a set of parallel word pairs. Then it forces each word pair to share the same representation and constructs a unified word representation for English and the target language. The document representation is computed by averaging all words in it. Given the representation as features, it trains a linear SVM model using the labeled English data.

BiDRL is a document-aligned representation learning method [Zhou *et al.*, 2016]. It uses *Google Translate* to create labeled parallel documents and forces the pseudo parallel documents to share the same embedding space. It also enforces constraints to make the document vectors associated with different sentiments fall into different positions in the embedding space. Finally, it concatenates the vectors of one document in both languages and trains a logistic regression sentiment classifier.

As the benchmark datasets have quite balanced positive and negative reviews, we follow the aforementioned studies to use accuracy as an evaluation metric. All the baseline methods have been evaluated with exactly the same training and test data sets used in previous studies [Zhou *et al.*, 2016], so we make direct comparisons with their reported results. Unfortunately, we cannot obtain the individual predictions of these methods, so we are not able to report the statistical significance (such as McNemar's test [Dietterich, 1998]) of the difference between these baselines and ELSA. To alleviate this problem and get robust results, we run ELSA 10 times with different random initiations and summarize its average accuracy and standard deviation in Table 2, as well as the reported performance of the baselines. As illustrated, ELSA outperforms all three baseline methods on all nine tasks.

3.3 The Power of Emojis

To understand how emojis affect cross-lingual sentiment classification, a straightforward idea is to remove the emoji-prediction phase and compare simplified versions of ELSA:

N-ELSA removes the emoji-prediction phase of both languages and directly uses two attention layers to realize the transformation from word vectors to the final document representation. There is no emoji data used in this model.

T-ELSA removes the emoji-based representation learning on the English side. It uses the emoji-powered representations for the target language and translates labeled English documents into the target language to train a sentiment classifier for the target language.

S-ELSA removes the emoji-based representation learning in the target language. It uses the emoji-powered representations of English and trains a sentiment classifier based on labeled English documents. Documents in the target language are first translated into English and then classified.

Test accuracy of these models is illustrated in Table 3. We find that ELSA consistently achieves better accuracy compared to N-ELSA, T-ELSA, and S-ELSA on all tasks (McNemar's test is performed and the differences are all statistically significant at the 5% level). The superiority of ELSA shows that incorporating language-specific knowledge for both languages is critical to the model's performance.

4 Conclusion

We have presented ELSA, a novel emoji-powered representation learning framework, to capture both general and language-specific sentiment knowledge in the source and the target languages for cross-lingual sentiment classification. The representations learned by ELSA capture not only sentiment knowledge that generalizes across languages, but also language-specific patterns. We evaluate ELSA with comprehensive experiments on representative benchmark datasets, which outperforms the state-of-the-art cross-lingual sentiment classification methods. The promising results indicate that emojis may be used as an a general instrument for text mining tasks that suffer from the scarcity of labeled examples, especially in situations where an inequality among different languages presents.

References

[Bahdanau *et al.*, 2014] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In *Proceedings of the 3rd International Conference on Learning Representations, ICLR 2014*, 2014.

[Chen *et al.*, 2017] Qiang Chen, Chenliang Li, and Wenjie Li. Modeling language discrepancy for cross-lingual sentiment analysis. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017*, pages 117–126, 2017.

[Chen *et al.*, 2018] Zhenpeng Chen, Xuan Lu, Wei Ai, Huoran Li, Qiaozhu Mei, and Xuanzhe Liu. Through a gender lens: learning usage patterns of emojis from large-scale android users. In *Proceedings of the 2018 World Wide Web Conference, WWW 2018*, pages 763–772, 2018.

[Chen *et al.*, 2019] Zhenpeng Chen, Sheng Shen, Ziniu Hu, Xuan Lu, Qiaozhu Mei, and Xuanzhe Liu. Emoji-powered representation learning for cross-lingual sentiment classification. In *Proceedings of the 2019 World Wide Web Conference, WWW 2019*, pages 251–262, 2019.

[Deriu *et al.*, 2017] Jan Deriu, Aurélien Lucchi, Valeria De Luca, Aliaksei Severyn, Simon Müller, Mark Cieliebak, Thomas Hofmann, and Martin Jaggi. Leveraging large amounts of weakly supervised data for multi-language sentiment classification. In *Proceedings of the 26th International Conference on World Wide Web, WWW 2017*, pages 1045–1052, 2017.

[Dietterich, 1998] Thomas G Dietterich. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural computation*, 10(7):1895–1923, 1998.

[Harakawa *et al.*, 2018] Ryosuke Harakawa, Daichi Takehara, Takahiro Ogawa, and Miki Haseyama. Sentiment-aware personalized tweet recommendation through multimodal FFM. *Multimedia Tools Appl.*, 77(14):18741–18759, 2018.

[Mohammad *et al.*, 2016] Saif M. Mohammad, Mohammad Salameh, and Svetlana Kiritchenko. How translation alters sentiment. *J. Artif. Intell. Res.*, 55:95–130, 2016.

[Prettenhofer and Stein, 2010] Peter Prettenhofer and Benno Stein. Cross-language text classification using structural correspondence learning. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, ACL 2010*, pages 1118–1127, 2010.

[Qiu *et al.*, 2010] Guang Qiu, Xiaofei He, Feng Zhang, Yuan Shi, Jiajun Bu, and Chun Chen. DASA: dissatisfaction-oriented advertising based on sentiment analysis. *Expert Systems with Applications*, 37(9):6182–6191, 2010.

[Sun *et al.*, 2019] Lihua Sun, Junpeng Guo, and Yanlin Zhu. Applying uncertainty theory into the restaurant recommender system based on sentiment analysis of online chinese reviews. *World Wide Web*, 22(1):83–100, 2019.

[Web, 2010] Webis-cls-10. https://www.uni-weimar.de/en/media/chairs/computer-science-department/webis/data/corpus-webis-cls-10/, 2010. Retrieved on October 22, 2018.

[Xiao and Guo, 2013] Min Xiao and Yuhong Guo. Semi-supervised representation learning for cross-lingual text classification. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP 2013*, pages 1465–1475, 2013.

[Zhou *et al.*, 2016] Xinjie Zhou, Xiaojun Wan, and Jianguo Xiao. Cross-lingual sentiment classification with bilingual document representation learning. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016*, pages 1403–1412, 2016.