Emoji-Powered Representation Learning for Cross-Lingual Sentiment Classification

Zhenpeng Chen
Peking University
Haidian District, Beijing, China

Sheng Shen
University of California, Berkeley
California, United States

Ziniu Hu
University of California, Los Angeles
California, United States

Xuan Lu
Peking University
Haidian District, Beijing, China

Qiaozhu Mei
University of Michigan
Ann Arbor, Michigan, United States

Xuanzhe Liu
Peking University
Haidian District, Beijing, China

ABSTRACT

Most existing sentiment classification approaches heavily rely on a large amount of labeled data. In practice, the distribution of available labeled data is highly imbalanced among different languages, e.g., more English texts are labeled than texts in other languages. To tackle this problem, cross-lingual sentiment classification approaches aim to transfer the knowledge learned from a language with abundant labeled examples (i.e., the source language, usually English) to another language with much 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 language. However, this approach often fails to capture sentiment knowledge specific to the target language, and thus compromises the accuracy of the downstream classification task. In this paper, we employ emojis, ubiquitous and emotional language units, as the instrument to learn both the cross-language and language-specific sentiment patterns in different languages. Specifically, we propose a representation learning approach through the emoji prediction task to learn respective text representations for both languages. The learned representations are then utilized to facilitate cross-lingual sentiment classification. We demonstrate the effectiveness and efficiency of our approach on representative benchmark data sets.

CCS CONCEPTS

• **Information systems → Sentiment analysis**; • **Computing methodologies → Natural language processing**;

KEYWORDS

Emoji; representation learning; cross-lingual; sentiment classification

WWW'19, May 2019, San Francisco, USA

ACM ISBN 123-4567-24-567/08/06…$15.00
https://doi.org/10.475/123_4

ACM Reference Format:
Zhenpeng Chen, Sheng Shen, Ziniu Hu, Xuan Lu, Qiaozhu Mei, and Xuanzhe Liu. 2019. Emoji-Powered Representation Learning for Cross-Lingual Sentiment Classification. In *Proceedings of The 2019 Web Conference (WWW'19).* ACM, New York, NY, USA, Article 0, 12 pages. https://doi.org/10.475/123_4

1 INTRODUCTION

In the past decade, sentiment analysis has become a critical research topic in various communities including natural language processing [29, 76], Web mining [32, 67], information retrieval [14, 78], ubiquitous computing [34, 69], and human-computer interaction [26, 80]. Due to its effectiveness of understanding user attitudes, emotions, and even potential psychological statuses, sentiment analysis is widely applied to various types of Web content such as blogs, Tweets, user reviews, and forum discussions, leading to many valuable applications such as customer feedback tracking [30], sales prediction [45], product ranking [50], stock market prediction [17], opinion polling [55], and online advertising [66].

Like many other text analysis tasks, existing work on sentiment analysis mainly focuses on English texts [14, 29, 44, 48, 76–78]. Although some efforts have also been made on other languages such as Japanese [58, 64, 68, 82], sentiment analysis for non-English languages are far from sufficient, given that 74.6% of Internet users are non-English speakers [9]. The cause of this imbalance is quite straightforward: effective sentiment analysis tools are often built upon supervised learning techniques, there are way more labeled examples in English than in other languages.

An intuitive treatment in this situation is to transfer the knowledge learned from the richly labeled English text (i.e., the source language) to other languages where labels are scarce (i.e., the target language), an approach known as *cross-lingual sentiment classification* in literature [63]. In practice, the biggest challenge of cross-lingual sentiment classification is the linguistic gap between English and the target language, and the gap may be bridged through standard NLP techniques such as machine translation. Indeed, most recent studies use machine translation tools to generate pseudo parallel texts and then learn bilingual representations for the downstream sentiment classification tasks [20, 28, 49, 60, 63, 71, 74, 79, 84, 87, 88]. 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 done in that space.

Although the idea looks sensible and is easily executable, the practical effectiveness of these machine translation based methods are far from satisfactory. Indeed, a major challenge of cross-lingual sentiment analysis is the language discrepancy problem [21], and machine translation does not tackle this problem well. More specifically, sentiment expressions often differ across languages. Machine translation is able to retain common expressions of sentiments that are shared across languages (e.g., "*angry*" or "怒っている" for negative sentiment), but usually falls short to carry through language-specific expressions without losing or even altering their sentiments [54]. As an example, in Japanese, the common expression "湯水のように使う" indicates a negative sentiment, describing the behavior of excess or waste. However, when directly translated to English, "*use it like hot water*," it not only loses the negative sentiment but also becomes an odd expression.

The reason behind is perhaps not hard to explain: machine translation tools are usually trained on parallel corpora that are built in the first place to reflect the common patterns across languages, instead of patterns specific to individual languages. In other words, the fundamental cause to the pitfall is the lack of language-specific sentimental information when unilaterally pursuing the common patterns through machine translation. A better bridge is needed beyond machine translation, that not only transfers "common sentiment knowedge" across languages, but also captures the language-specific knowledge of sentiments. Emojis provide such a bridge.

In this paper, we address the limitations of current cross-lingual sentiment analysis by using emojis as both the proxy of sentiment labels and the bridge between the source and target languages. The potential of expressing sentiment by emojis [24, 39, 61] motivates us to employ them as complementary labels for sentiments, while the ubiquitous characteristic [16, 23, 47] makes it feasible to learn emoji/sentiment representations for every single language. Utilized together with machine translation, the common patterns of emoji usage complement the pseudo parallel corpora and narrow the language gap, and the language-specific emoji usage helps address the language discrepancy problem in cross-lingual sentiment classification.

We propose ELSA, a novel emoji-powered representation learning framework for cross-lingual sentiment classification. In ELSA, language-specific representations are first derived based on how emojis are used among the texts in individual languages. 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 bilingual representations in existing studies, our representations can catch not only the inherently common sentiment patterns across languages, but also the language-specific knowledge. In this way, the joint representation is not dominated by the source language where the labels are.

The performance of ELSA is evaluated on a representative Amazon review data set that has been used in various cross-lingual sentiment classification studies [63, 79, 88], covering 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 these tasks in terms of classification accuracy. Experiments also show that the emoji-powered model still works well even when the volume of unlabeled and labeled data are rather limited. To evaluate the generalizability of ELSA, we also apply our approach to Tweets, which also achieves the state-of-the-art performance.

The major contributions of this paper are as follows:

- To the best of our knowledge, we are the first to leverage emojis both as surrogate sentiment labels and as the bridge to address the language discrepancy in cross-lingual sentiment classification.
- We propose a representation learning approach to incorporating the language-specific knowledge into the cross-lingual sentiment classification task, based on an attention-based LSTM model to capture the sentiment information implied in the ubiquitous emoji usage.[1]
- We demonstrate the effectiveness and efficiency of ELSA in improving the cross-lingual sentiment classification over extensive data sets of reviews and Tweets. Results demonstrate that ELSA can significantly outperform the state-of-the-art results on benchmark data sets.

The rest of this paper is organized as follows. Section 2 formulates the problem and presents our cross-lingual representation learning approach. Section 3 evaluates our approach on nine tasks and explores the power of emojis in the learning process. Section 4 discusses the data size sensitivity and generalizability of our approach. Section 5 presents related work, followed by concluding remarks in Section 6.

2 THE ELSA APPROACH

To better introduce the workflow of our approach, we first formulate our problem. The cross-lingual sentiment classification aims to use the labeled data in source language (i.e., English) to learn a model that can classify the sentiment of test data in target language. In our setting, besides labeled English texts (L_S), we also have large-scale unlabeled English (U_S) and target-language texts (U_T). Furthermore, we need some unlabeled texts containing emojis for both English (E_S) and the target language (E_T). In practice, these unlabeled data can be easily accessed in Twitter on which emojis are widely used. Our task is to build the model that can classify the sentiment polarity of target-language texts solely based on the labeled source-language texts (i.e., L_S) and these unlabeled data (i.e., U_S, U_T, E_S and E_T). Finally, we use the labeled target-language texts L_T to evaluate the model.

The framework of our approach is illustrated in Figure 1, with the following steps. In *step 1* and *step 2*, we learn representation models for source and target languages, respectively. More specifically, we employ large-scale Tweets to learn word embeddings of both languages through Word2Vec [51] in an unsupervised fashion. Then in a distant-supervised way, we use emojis as complementary sentiment labels to transfer the word embeddings into a higher-level sentence representation. In *step 3*, we translate the labeled English texts to the target language through *Google Translate*. We input each English text and its translation into source- and target-language representation model respectively and get representation for every single sentence in *step 4* and *step 5*. Then in *step 6* and

[1]We plan to release the code and pre-trained models for all languages when the work is published.

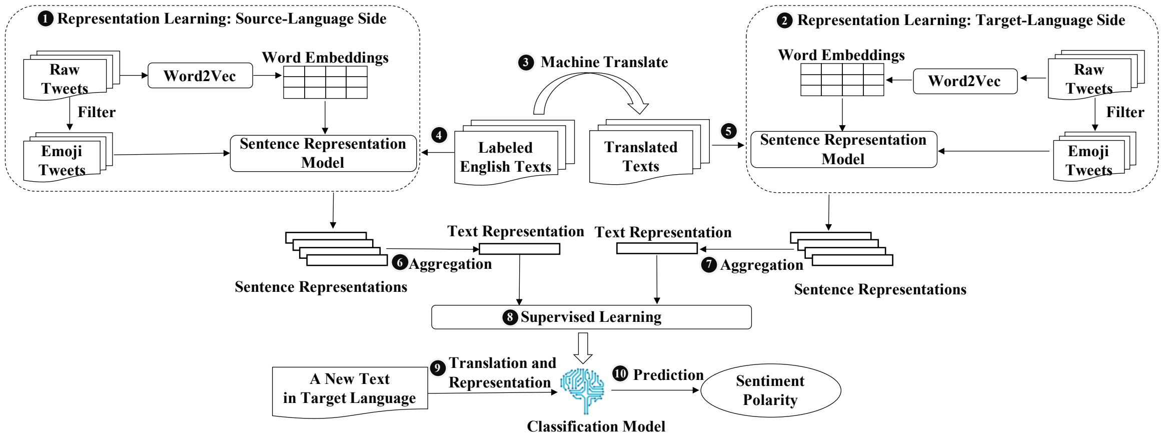


Figure 1: The overview of our ELSA approach.

step 7 we aggregate the sentence representations to form a compact representation for each source- and target-language text. In *step 8*, we use the two representations as features and the real sentiment labels as the ground-truth to learn the final sentiment classifier in a supervised way. In the test phase, for each target-language text, we translate it to English and then follow the previous steps to achieve representation features (*step 9*) for the classifier to predict sentiment polarity (*step 10*). We next introduce the details of this workflow.

2.1 Representation Learning

First, we define representation rules for the source and target languages. In practice, we can just use the existing word embedding techniques to create word representations and then average the word vectors in each text to represent it. However, as the goal of our study is cross-lingual sentiment classification, we call for a better representation approach that can capture both the sequential relationships of words and the sentiment signals in each language. Existing research efforts in ubiquitous computing have demonstrated that emojis can express the sentiment across languages [16, 47]. Thus, we choose emojis to incorporate sentiment information. More specifically, in a distant-supervised way, we use emojis as sentiment labels and represent texts co-used with the same emoji similarly through an emoji prediction task. The representation learning process of source and target languages is conducted separately in order to capture the language-specific knowledge.

The model architecture of representation learning is illustrated in Figure 2. First, we pre-train basic word embeddings based on large-scale unlabeled texts (refers to word embedding layer). Then, we can represent every single word as a unique vector and use the stacked bi-directional LSTM layers and one attention layer to encode these word vectors into sentence representations. The attention layer takes the outputs of both the embedding layer and the two LSTM layers as input by the skip-connection algorithm [36] and enables the unimpeded information flow in the whole training process. Finally, the model parameters can be learned by minimizing

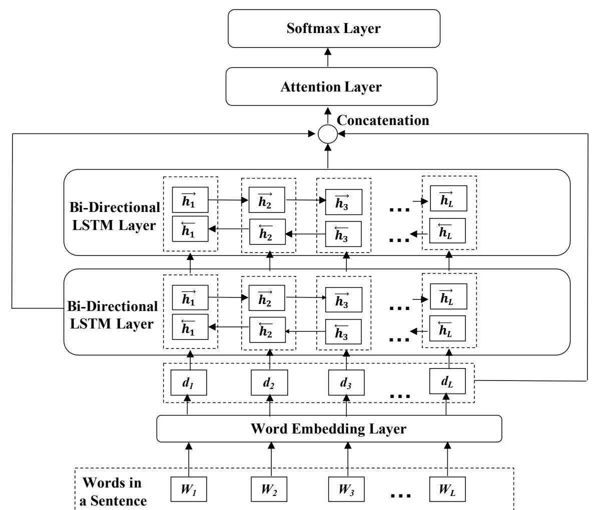


Figure 2: The architecture of representation learning network.

the output error of the softmax layer. The details of the architecture are elaborated below.

Word Embedding Layer. The word embeddings are pre-trained with skip-gram algorithm [51] based on U_S and U_T to encode every single word into a continuous vector space. By predicting the nearby words, the words that usually occur in the similar contexts are embedded closely in the vector space, which indicates the semantic information of the embeddings. As the word embedding approach is standard as the well-known Word2Vec, we do not describe the details of the whole process. Through this phase, each text in E_S or E_T can be denoted as (x, e), where $x = [d_1, d_2, ..., d_L]$ is a sequence of word vectors that represent the plain text (by removing emoji) and e is the emoji contained in the text.

Bi-directional LSTM layer. As a kind of recurrent neural network (RNN), the long short-term memory network (LSTM) [38] is suitable for processing the sequential data such as texts due to its recurrent nature. At each time step, LSTM combines the current input and knowledge from the previous time steps to update the state of the hidden layer. In addition, to tackle the gradient vanishing problem [37] of traditional RNNs, LSTM incorporates a gating mechanism to determine when and how the states of hidden layers can be updated. Each LSTM unit contains a memory cell and three gates (i.e., an input gate, a forget gate, and an output gate) [56]. The input and output gates control the input activations into the memory cell and the output flow of cell activations into the rest of the network, respectively. The memory cells in LSTM store the temporal states of the network. Each memory cell has a self-loop whose weight is controlled by the forget gate.

At time step t, the LSTM computes unit states of the network as follows:

$$\begin{aligned} i^{(t)} &= \sigma(U_i x^{(t)} + W_i h^{(t-1)} + b_i), \\ f^{(t)} &= \sigma(U_f x^{(t)} + W_f h^{(t-1)} + b_f), \\ o^{(t)} &= \sigma(U_o x^{(t)} + W_o h^{(t-1)} + b_o), \\ c^{(t)} &= f_t \odot c^{(t-1)} + i^{(t)} \odot tanh(U_c x^{(t)} + W_c h^{(t-1)} + b_c), \\ h^{(t)} &= o^{(t)} \odot tanh(c^{(t)}), \end{aligned}$$

where $x^{(t)}$, $i^{(t)}$, $f^{(t)}$, $o^{(t)}$, $c^{(t)}$, and $h^{(t)}$ denote the input vector, the state of the input gate, forget gate, output gate, memory cell and hidden layer at time step t. W, U, b respectively denote the recurrent weights, input weights, and biases. $\odot$ is the element-wise product. We can extract the latent vector for each time step t from LSTM. In order to capture the information from the past and future of a word in its current context, 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 as:

$$\begin{aligned} \overrightarrow{h_i} &= \overrightarrow{LSTM}(d_i), i \in [1, L] \\ \overleftarrow{h_i} &= \overleftarrow{LSTM}(d_i), i \in [L, 1] \\ h_i &= [\overrightarrow{h_i}, \overleftarrow{h_i}], \end{aligned}$$

Attention Layer. As we employ the skip-connection that concatenates the outputs of the embedding layer and the two bi-directional LSTM layers as the input of the attention layer, the i-th word of the input sentence can be represented as u_i:

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

where d_i, h_{i1}, and h_{i2} denote the encoded vector of word extracted in the word embedding layer, the first bi-directional LSTM, and the second bi-directional LSTM, respectively. Since not all words contribute equally to predicting emojis and classifying sentiment, we employ the attention mechanism [13] 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 for the attention layer. Then each sentence can be represented by the weighted sum of all words in it, using the attention scores as weights. It means that the sentence representation is calculated as:

$$v = \sum_{i=1}^{L} a_i u_i,$$

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. The i-th element of the probability vector is calculated as:

$$y_i = \frac{exp(v^T w_i + b_i)}{\sum_{j=1}^{K} exp(v^T w_j + b_j)},$$

where w_i and b_i are respectively the weight and bias of the i-th element. Finally, we learn the models' parameters by minimizing the cross entropy between the output probability vectors and the one-hot representations of the emojis contained in the texts. After learning the parameters, we can extract the output of the attention layer to represent the input sentence. Through this emoji-prediction phase, the words with distinctive sentiments can be identified, and the texts co-used with the same emoji will be represented similarly. Given the fact that the scale of labeled data is quite limited, we should avoid the possible over-fitting problem by freezing the sentence representation models.

2.2 Supervised Training

We next leverage pre-trained English and target-language representation models to conduct the cross-lingual sentiment classification.

First, for each English text $x_s \in L_S$, we use the pre-trained English representation model to generate representation for every single sentence in it. Second, we aggregate these sentence representations to derive a compact text representation. Because different parts of a text can have quite different importances for the overall sentiment, we still adopt attention mechanism here. Supposing the sentence vector as v, we calculate the text vector r_s as:

$$\beta_i = \frac{exp(W_b v_i)}{\sum_{j=1}^{T} exp(W_b v_j)},$$

$$r_s = \sum_{i=1}^{T} \beta_i v_i,$$

where W_b is the weight matrix of the attention layer and β_i is the attention score of the i-th sentence in the text. Next, we use *Google Translate* to translate x_s to the target language (x_t). We then leverage the pre-trained target-language representation model to form representation for each translation following the above-mentioned steps. Supposing the text representations of x_s and x_t are r_s and r_t respectively, we concatenate them as $r_c = [r_s, r_t]$. The concatenated representation contains sentiment knowledge from both English and the target language, ensuring our model is not dominated by English where the labels are. Finally, we input r_c into a softmax layer and minimize the cross entropy between the output and the real sentiment labels to learn the network parameters in a supervised way.

Table 1: The size of the Tweets and emoji-texts.

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

2.3 Test Phase

In the test phase, for each text in L_T, we first translate it to English. Based on the models trained above, the original text 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. The classifier will output the predicted sentiment polarity.

3 EVALUATION

We then use widely adopted benchmark data sets in cross-lingual sentiment classification and a large-scale corpus of Tweets to validate the power of our approach.

3.1 The Data Set

The labeled data (L_S for training and L_T for testing) used in our work are from the Amazon review data set [3] created by Prettenhofer and Stein [63]. This data set is representative and used in various cross-lingual sentiment classification work [49, 63, 79, 88]. 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 data set contains 1,000 positive reviews and 1,000 negative reviews that have been labeled. We select English as the source language and the rest three as the target languages. Therefore, we can evaluate our approach on nine tasks in total (i.e., combinations of three domains and three target languages). For each task, we use the 2,000 labeled English data in the corresponding domain for training and the 2,000 labeled target-language data for test. The translations of the test data are provided by this data set, so we need to translate only the English reviews to target languages.

To achieve unlabeled data (U_S and U_T), we crawl English, Japanese, French, and German Tweets from Twitter. As emojis are widely adopted on Twitter [46], we can extract emoji-texts directly from these unlabeled data. To fully exploit the characteristic properties of different languages, we extract the most frequently used 64 emojis, which cover about 70% of the total emoji usage, in each language. The Tweets do not contain any of these emojis are filtered out accordingly. As many Tweets contain multiple emojis, for each Tweet, we create separate samples for each unique emoji in it to make our emoji prediction task a single-label classification task instead of a complicated multi-label one. For example, "I love you ❤️❤️💋" can be separated into two samples, i.e., ("I love you", ❤️) and ("I love you", 💋). The reprocessed emoji-texts provide the E_S and E_T data for learning representation models.

We then conduct the following processing procedures for the text data. We remove the reTweets and Tweets that contain URLs to ensure all the words appear in their original contexts and the words' interpretations do not depend on external contents. Then we tokenize all the texts (including reviews and Tweets) into words, convert them into lowercase, and shorten the words with redundant characters into their canonical forms [18] (e.g., "cooooool" is converted to "cool"). As Japanese words are not separated by whitespace, we use a tokenization tool called *MeCab* [2] to segment Japanese reviews. In addition, we use special tokens to replace mentions and numbers. Finally, we present an overview of our unlabeled Tweet data in Table 1.

3.2 Implementation Details

We learn the initial word embeddings using the skip-gram model with the window-size of 5 on the raw Tweets. The word vectors are then fine-tuned during the sentence representation learning phase. In the sentence representation learning phase, to regularize our model, the L2 regularization of 10^{-6} is applied for embedding weights. The dropout with 0.5 rate is introduced before the softmax layer. The hidden units of bi-directional LSTM layers are set as 1,024 (512 in each direction). We randomly split the emoji-texts into the training, validation, and test sets in the proportion of 7:2:1. Accordingly, we use early stopping [19, 33] to tune parameters based on the validation performance through 50 epochs, with mini-batch size of 250. We used the Adam [40] for optimization, with the two momentum parameters set to 0.9 and 0.999, respectively. The initial learning rate was set to 10^{-3}. In the supervised training phase, for exhaustive parameter tuning, we randomly select 90% of the labeled data as the training set and the rest 10% as the validation set. The whole framework is implemented with TensorFlow [11].

3.3 Baselines and Results

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

MT-BOW uses the bag-of-words features to learn a linear classifier on the labeled English data [63]. Then it uses *Google Translate* to translate the test data to English and applies the pre-trained classifier to predict the sentiment polarity of the translated texts.

CL-RL is the word-aligned representation learning method proposed by Xiao and Guo [79]. It constructs a unified word representation that consists of both language-specific components and shared components, for the source and target languages. To establish connections between the two languages, it leverages *Google Translate* to create a set of critical parallel word pairs, then forces each parallel word pair to share the same word representation. The text representation is computed by taking average over all words in the text. Given the representation scheme, it trains a linear SVM model using the labeled English data.

BiDRL is the document-aligned representation learning method proposed by Zhou *et al.* [88] It uses *Google Translate* to create the labeled parallel documents and forces the pseudo parallel documents to share the same embedding space. It also uses constraints to make the document vectors associated with different sentiments fall into different positions in the embedding space. Furthermore, it forces the documents with large textual differences but the same sentiment to have similar representations. Based on the representation approach, it concatenates the vectors of one document in both languages to represent the document and trains a logistic regression sentiment classifier.

Table 2: The accuracy of the cross-lingual sentiment classifications for the nine benchmark tasks.

Language	Domain	MT-BOW	CL-RL	BiDRL	ELSA
Japanese	Book	0.702	0.711	0.732	**0.789**
	DVD	0.713	0.731	0.768	**0.800**
	Music	0.720	0.744	0.788	**0.820**
French	Book	0.808	0.783	0.844	**0.865**
	DVD	0.788	0.748	0.836	**0.859**
	Music	0.758	0.787	0.825	**0.864**
German	Book	0.797	0.799	0.841	**0.867**
	DVD	0.779	0.771	0.841	**0.862**
	Music	0.772	0.773	0.847	**0.874**

Table 3: The performance of our original ELSA and simplified models.

Language	Domain	N-ELSA	T-ELSA	S-ELSA	ELSA
Japanese	Book	0.527	0.742	0.757	0.789
	DVD	0.507	0.756	0.766	0.800
	Music	0.513	0.792	0.782	0.820
French	Book	0.505	0.821	0.851	0.865
	DVD	0.507	0.816	0.846	0.869
	Music	0.503	0.811	0.852	0.864
German	Book	0.513	0.804	0.853	0.867
	DVD	0.521	0.790	0.851	0.862
	Music	0.513	0.818	0.861	0.874

As the benchmark data sets have quite balanced positive and negative reviews, we follow aforementioned studies to use accuracy as evaluation metric. We summarize the performance of the baseline methods and our ELSA in Table 2. All the baseline methods have been evaluated on the same benchmark data sets in the previous literatures [63, 79, 88].

As illustrated in Table 2, the performance of Japanese tasks are worse than French and German tasks. According to the language systems defined by ISO 639 [10], English, French, and German belong to the same language family (i.e., Indo-European), while Japanese belongs to Japonic. In other words, French and German have more common sentiment patterns with English. It is easier to translate English texts to French and German and transfer the sentiment knowledge from English to them. Therefore, in fact, Japanese tasks are most difficult in these tasks and all the previous methods cannot achieve an accuracy above 0.8. However, it is encouraging to find that our approach achieves an accuracy of 0.8 in Japanese DVD task and an accuracy of 0.82 in Japanese music task. The 0.789 accuracy in book task is also non-negligible as it outperforms the previous best performance (0.732) with 0.057. In addition, although French and German tasks are a little easier than Japanese ones, no existing approaches can achieve an accuracy over 0.85 on any of the six tasks. However, our approach can achieve accuracy higher than 0.85 on all of the six tasks, which further indicates the classification power of our approach.

Next, we compare the results more thoroughly and further demonstrate the advantages of our approach. As is shown, the representation learning approaches (CL-RL, BiDRL, and ELSA) all outperform the naive MT-BOW on most tasks. It is reasonable as representation learning approaches represent words as high-dimensional vectors in a continuous space and thus overcome the feature sparsity of the traditional bag-of-words approaches. Furthermore, we observe the document-level representation approaches (BiDRL and ELSA) outperform the world-level CL-RL. It indicates that incorporating document-level information (including the sentiment knowledge) into representations is more effective than focusing on independent words. Finally, our ELSA outperforms the BiDRL in all tasks. In order to narrow the linguistic gap, BiDRL only leverages the pseudo parallel texts to learn the common sentiment patterns between languages. Besides the pseudo parallel texts, ELSA also learns from the emoji usage of both languages. On the one hand, as a ubiquitous emotional signal, emojis are adopted across languages with some common sentiment patterns, which can complement the pseudo parallel corpus. On the other hand, the language-specific patterns of emoji usage help incorporate the language-specific knowledge of sentiment into the representation learning, which can benefit the downstream target-language sentiment classification. Then we want to explore the power of emojis in the learning process more deeply.

3.4 Power of Emojis

To further evaluate the "emoji power", we conduct subsequent experiments to investigate the effects of emojis from three perspectives, i.e., overall performance, representation learning efficiency, and text comprehension.

3.4.1 Overall Performance

To understand how emojis affect our cross-lingual sentiment model, a straightforward idea is to remove the emoji-prediction phase and implement simplified models for comparisons:

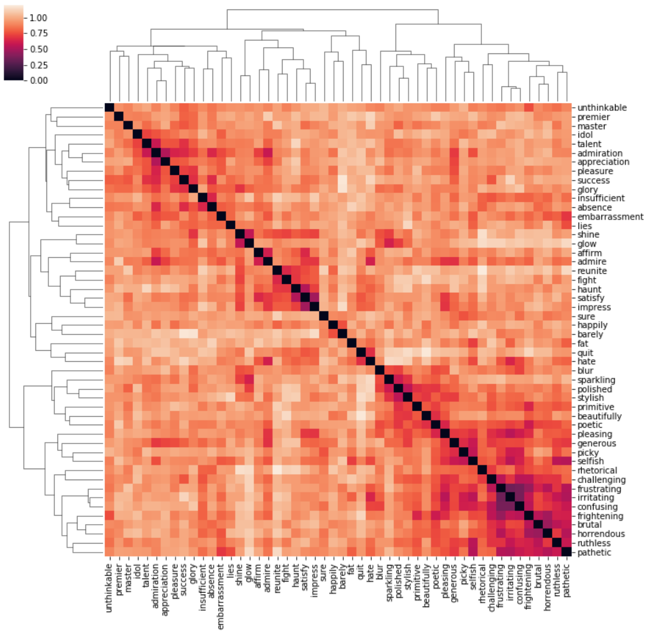
N-ELSA removes the emoji-prediction phase of both languages and directly uses two attention layers to realize the transformation from word vectors to the final text representation. There is no emoji data used in this model.

T-ELSA removes the English side representation learning. It uses the emoji-powered representations of the target language and the translations of English labeled data to train a sentiment classifier for the target language directly. This model leverages only the emoji data in the target language.

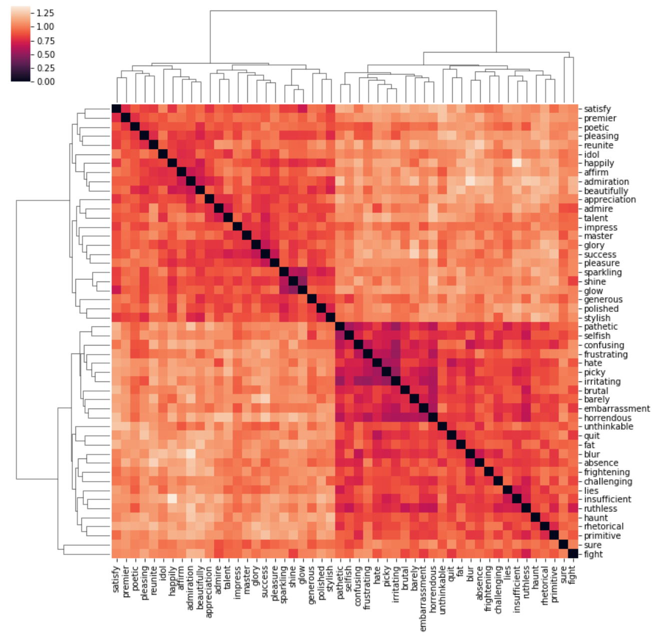
S-ELSA removes the target-language side representation learning. It uses the emoji-powered representations of English and the labeled English data to train a sentiment classifier and applies it to the translations of the target-language test samples. This model leverages only the emoji data in the source language (i.e., English).

Test results of these models are illustrated in Table 3. By comparison, we find that ELSA outperforms N-ELSA on all nine tasks. N-ELSA surpasses only quite a little compared to the uniform guess (50%) since it only learns the common patterns between languages from pseudo parallel texts and does not incorporate sentiment information effectively. Another alternative conjecture is that the 2,000 reviews are insufficient to train such a complex network model, which leads to the well-known "over-fitting" problem.

To alleviate such potential concern, we mix up the English and target-language labeled reviews and randomly select 2,000 samples from the mixed data as a training set and the remaining samples as a test set. The testing results are acceptable with an average



(a) Cluster map of the selected words without emoji information.



(b) Cluster map of the selected words with emoji information.

Figure 3: Comparison of the word representations *with* and *without* the emoji incorporation.

accuracy of 0.777 on all tasks. It indicates that "over-fitting" is not the major reason and N-ELSA can still work well if we effectively incorporate sentiment information into the training process. More specifically, the original N-ELSA is dominated by English sentiment information learned from pseudo parallel texts and thus fail to classify target-language samples correctly. When we input the sentiment information (labeled data) of both English and the target-language into the model, it can work well. However, in cross-lingual sentiment classification setting, we can not acquire the labeled target-language data for training. To alleviate this absence, we can leverage the emotional emoji texts, i.e., our ELSA approach.

In addition, ELSA also consistently achieves higher accuracy compared to T-ELSA and S-ELSA on all tasks, albeit often only slightly higher. We further use McNemar's test [27] to ensure the performance differences are statistically significantly at 5% level. The superiority of ELSA shows that only extracting sentiment information from one language is not enough for the cross-lingual sentiment task and the language-specific knowledge incorporated for both languages can benefit the model performance.

3.4.2 Representation Learning Efficiency

To better understand the sentiment information learned by the emoji usage, we then conduct an empirical experiment at the word representation level. Recall that after the word embedding phase, each single word can be represented by a unique vector and these word vectors are then fine-tuned in the emoji-prediction phase. Next, we would like to evaluate whether the sentiment information is captured by the new word representations under the effects of emojis. We sample 50 English words with distinct sentiments from MPQA subjectivity lexicon [1] based on their frequency in our corpus. These words are manually labeled in positive or negative polarity from MPQA, and we regard these labels as the ground-truth for further evaluation.

We expect an informative representation approach can embed the words with same sentiment polarity closely in the vector space. To measure and illustrate the similarity, we calculate the similarity score between every two words as the cosine distance of the corresponding representations. Based on the cosine distances, we perform a hierarchical clustering algorithm [15] and visualize the clustering results in Figure 3. The darkness of each cell indicates the similarity between the two words. The darker the cell is, the more similar representations the two words have.

In Figure 3(a), we use naive representations learned by Word2Vec and words with different sentiments can not be clearly separated. For example, in the bottom right part, the positive "generous" and the negative "picky" are considered to be similar. It indicates that these naive word embeddings fail to capture the sentiment information. Just because these words are all usually used to express sentiment, the representation mechanism represents them similarly.

In contrast, in Figure 3(b), we can easily observe two obvious clusters generated by our fine-tuned emoji-prediction model. The top left corner cluster contains the positive words, while the bottom right corner contains the negative words. Only one positive word "sure" is incorrectly clustered with negative words. By checking the contexts of this word in our corpus, we find it is usually co-used with both positive and negative words, causing its ambiguous representation. The correct cluster of nearly all the words indicates that under the effects of emojis, the new word representations capture more sentiment information, which can be very informative and beneficial to our downstream sentiment classification.

3.4.3 Text Comprehension

We then explore how our emoji-powered model benefits text comprehension. We select a representative case that is incorrectly classified by N-ELSA but correctly classified by ELSA. This case is selected from the Japanese test samples and we use the segment of its translated English version for illustration in Figure 4. Although the whole document indicates the dissatisfaction to an album, it is not that easy to identify this intent directly from each single sentence due to the translation quality and the document's complex compositions. For example, if we consider only the third sentence without context, the author seems to express a positive attitude. However, in fact, the author expresses obviously negative attitude in the fourth and sixth sentences.

In Figure 4, we present the attention distribution of words and sentences generated by N-ELSA and ELSA respectively, which indicates how the two models comprehend this document. We use the darkness of the cells to indicate the attention scores of words in each sentence. The darker the word, the higher its attention score. For each sentence, we list its attention score in this document. In Figure 4(b), we also list the top 3 emojis ELSA predicts for each sentence with the highest probabilities, which may indicate its sentiment polarity predicted by ELSA.

We first turn to Figure 4(a) that demonstrates how N-ELSA processes the sentiment information. In the word level, N-ELSA tends to focus more on the neural words like "song" or "album" instead of other sentimental words. In the sentence level, the extreme attention is addressed on the fifth sentence. However, the fifth sentence describes how the album is different from the first one and it does not express the obviously negative sentiment.

On the contrary, due to the incorporation of emojis, ELSA is able to work with a proper logic (see Figure 4(b)). To induce the appropriate sentence representation with considering the role of each word in it, ELSA addresses its attention on the emotional adjectives, such as "interesting", "not possible", and disjunctive conjunctions such as "however". Thus, it manages to indicate the sentiment of each sentence as we expected, which can be further explained by the predicted emojis on the left. Besides the most popular 😂 in our corpus, 😭 and 😩 predicted for the fourth and sixth sentence indicate the negative intent of the author, while 👍 and 😊 foretold in the third sentence indicate positive sentiment. Then in the sentence level, ELSA addresses less attention on the positive third sentence, while centering upon the fourth and the sixth sentences. By comparison, we can find that the emojis bring in additional knowledge to the text comprehension and thus make the attention mechanism more effective.

4 DISCUSSION

So far, we have presented the performance of ELSA on the review data sets and demonstrated the power of emojis in our learning process. Indeed, there are some issues that could potentially affect the effectiveness and efficiency of our approach, and some further discussions are necessary.

4.1 Data Volume Sensitivity Analysis

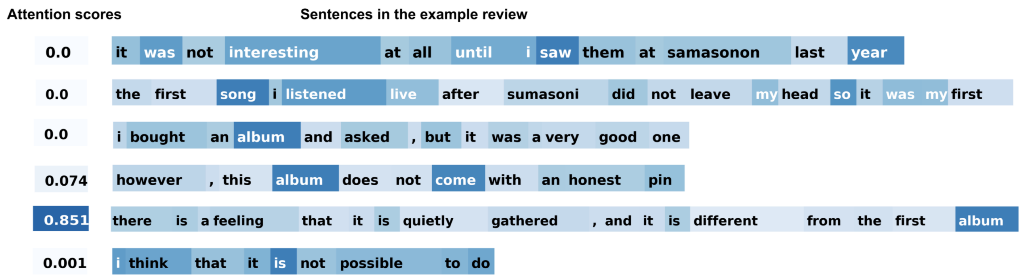
As we crawl large-scale Tweets for learning representations, we want to investigate whether our approach works well with less volume of data. First, we investigate the size of unlabeled data. The English-side model can be reusable to any other English-target language pair as we have it learned already. We need to scale down only the target-language Tweets and emoji Tweets simultaneously and observe the changes of performance on benchmarks. Detailedly, we use 80%, 60%, 40%, 20% of the collected data to re-train the target-side representation model and keep the final supervised training unchangeable. We summarize the results in Figure 5(a), 5(b), and 5(c). For the Japanese tasks, when we scale down the unlabeled data, the performance gets slightly worse. Comparing the results of 20% and 100% data, the accuracy differences in three domain tasks are 0.026, 0.023, and 0.029, respectively. For French and German, the performance fluctuates even less than 0.01. Most importantly, we find our approach can outperform the existing approaches on all the nine tasks even with the 20% unlabeled data. It indicates that when the target languages have limited unlabeled resources or the researchers have no patience to collect so many Tweets, our approach still works.

Furthermore, although the labeled English resources are relatively richer than that of other languages, it is still fairly limited. Hence, if our model can be robust even with less labeled English resources, it will be encouraging. To this end, we scale down the labeled data by 80%, 60%, 40%, and 20%. As shown in Figure 5(d), 5(e), and 5(f), the performance gets slightly worse with the decrease of the labeled English data, but our model with 20% labeled data (i.e., 400 labeled English samples) can outperform the existing approaches with the total 2,000 labeled samples on almost all the tasks. It shows that with large-scale emoji-texts applied in representation learning, the model relies less on the labeled data.

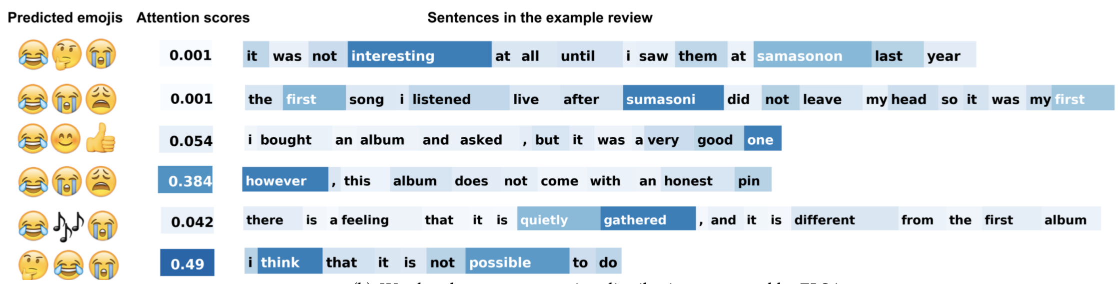
4.2 Generalizability Analysis

Most of the previous cross-lingual sentiment studies [60, 63, 79, 88] used the Amazon review data set for evaluation. For comparisons with them, we also adopt this data set in the main body of this paper. However, sentiment classification for other domains such as social media is also important. Can our approach still work well on this domain? For evaluating the generalizability of our approach, we next apply ELSA on a representative kind of social media data – Tweets. Due to the short and informal nature of Tweets, sentiment classification for them is considered to be a big challenge [31].

As the cross-lingual studies on Tweets are very limited, we take only the recent cross-lingual setting method (MT-CNN) proposed by Deriu *et al.* [25] for comparison. It also relies on large-scale unlabeled Tweets and translation technique. It first trains a sentiment classifier for English and then applies it to the translations of target-language test data. The training process for English contains three phases. First, it uses raw Tweets to create word embeddings just like us. Second, it leverages ":)" and ":(" as weak labels and applies a multi-layer CNN model to adapt the word embeddings. Finally, it trains the model on labeled English Tweets. This work and our work have a coverage of French and German, so we use them as the target languages for comparison.

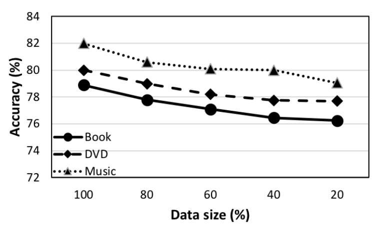


(a) Word and sentence attention distribution generated by N-ELSA.

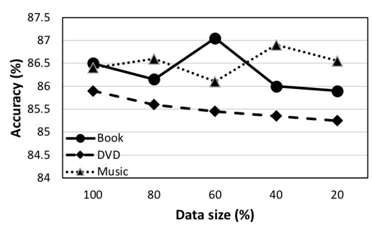


(b) Word and sentence attention distribution generated by ELSA.

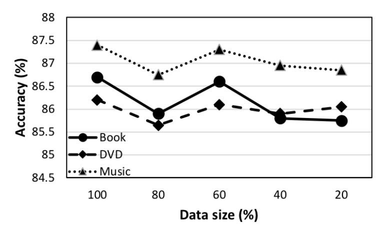
Figure 4: Case study: Effect of emojis in text comprehension.



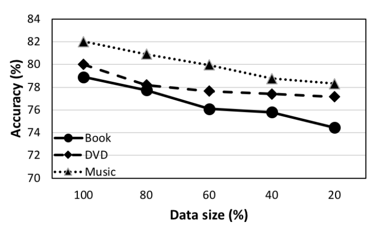
(a) Unlabeled data: Japanese



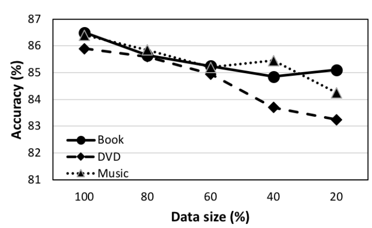
(b) Unlabeled data: French



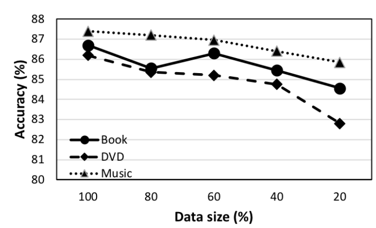
(c) Unlabeled data: German



(d) Labeled data: Japanese



(e) Labeled data: French



(f) Labeled data: German

Figure 5: Results obtained by varying the size of unlabeled and labeled data.

As the labeled Tweets are released in forms of Twitter IDs and some of them can not be crawled now, we can not apply our model to the crawled data and then compare our results with the reported results in [25] directly. For fair comparison, we reproduce their method on the crawled data. Based on the pre-trained representation models of MT-CNN [5] and ELSA, we use the same labeled English Tweets to train and validate the two classifiers and then test them on the same data (i.e., labeled French and German Tweets crawled by us). We list the data size of our used labeled English, French and German Tweets in Table 4. From the data distribution, we can get naive accuracy baselines for French and German by uniform guess. We may simply guess the sentiment labels as positive, neutral, or negative, and choose the highest accuracy in the three conditions as the baseline result, i.e., 0.451 for French and 0.628 for German.

Results are summarized in Table 5. The two approaches both outperform the uniform guess, and our ELSA outperforms the MT-CNN by 0.161 on French task and 0.155 on German task. Although we

Table 4: The data size of crawled labeled Tweets.

Dataset	Language	Positive	Neutral	Negative
Training	English [8]	5,101	3,742	1,643
Validation	English [7]	1,038	987	365
Test	French [4]	987	1,399	718
	German [6]	1,057	4,441	1,573

Table 5: The accuracy on French and German Tweets.

Language	ELSA	MT-CNN	Uniform Guess
French	0.696	0.535	0.451
German	0.809	0.654	0.628

use the same training, validation, and test set for both approaches, we are still concerned whether the pre-trained representation models bring in unfairness. More specific, if we use more unlabeled Tweets for representation learning than MT-CNN, our outstanding performance may attribute to the advantage of data size. To answer this question, we refer to [25] to check their data size. We find that they use 300M raw Tweets and 60M Tweets containing ":)" and ":(" for representation learning. By contrast, We use only 81M raw Tweets and 13.7M emoji-Tweets of English, French, and German in total. Considering emoticons are significantly less used than emojis on Twitter [57], although they use about 4.4 times weak-labeled Tweets more than us, they need to spare more than 4.4 times efforts to collect them. To sum up, our approach can obviously outperform the existing MT-CNN with significantly less data.

5 RELATED WORK

We then present the background and literature related to our study. Our work is inspired by three streams of literature: prevalence of emojis, textual sentiment analysis, and cross-lingual text classification.

Prevalence of Emojis. Emojis, also known as ideograms or smileys, can be used as compact expressions of objects, topics, and emotions. Being encoded in Unicode, they have no language barriers and are diffused on the Internet rapidly [23, 47]. The prevalence of emojis has attracted researchers from various research communities such as ubiquitous computing, human-computer interaction, computer-mediated communication, multimedia, and Web mining [12, 16, 22–24, 39, 47, 52, 53, 61, 62, 72, 86]. Many research efforts have been devoted to studying their usage across apps [72], across platforms [52, 53], across genders [23], and across cultures [47]. The various non-verbal functions of emojis are an important factor of their wide adoption. Emojis are proposed to replace content words, provide the situational and additional emotional information, adjust tone, express intimacy, etc [24, 39, 61]. Especially, expressing sentiment is demonstrated as the most popular intention for using emojis [39], so it is believed that emojis can be benign proxies of representing sentiment [29, 83]. Considering the ubiquitous usage of emojis across languages and their sentiment expression function, we make the first effort to use emojis as sentiment labels to improve the cross-lingual sentiment analysis.

Textual Sentiment Analysis. Textual sentiment analysis is explored as an NLP task to study the opinions, sentiments, evaluations, appraisals, attitudes, and emotions of people [42]. Many well-known tools simply leverage the polarity of single words to determine the overall sentiment score of text, such as SentiStrength [73] and LIWC [59]. Recently, as the emergence of deep learning, many researchers attempt to use advanced neural network algorithms to solve the sentiment analysis tasks [65, 70, 75, 76, 81]. Supervised machine-learning or deep-learning based methods usually need a large volume of labeled data to train the model. However, it is time-consuming and error-prone to label sentences manually [29]. To counter the scarcity of labeled data, many researchers adopt distant-supervised learning and use emotional hashtags and emoticons as weak sentiment labels [25, 29, 43, 83]. However, the hashtags are too language-specific to be general, and the emoticons have been gradually replaced by increasingly popular emojis [57]. In this study, we propose a distant-supervised learning framework with emojis as labels to learn informative representations for different languages.

Cross-Lingual Text Classification. Due to the the imbalanced labeled resources among different languages, there is an urgent need to design algorithms in a cross-lingual view, so as to tackle various text classification tasks such as Web page classification [41], topic categorization [85], and sentiment analysis [88] in different languages. Cross-lingual text classification aims to exploit knowledge in source language (usually refers to English) with relatively sufficient labeled data to assist classifications in the target language with limited annotated texts [41]. Many researchers divided the learning process into two stages as encoding texts in source and target language into continuous representations, and leveraging these representations for the final classification task [49, 85]. Furthermore, to bridge the linguistic gap between the source and target language, most efforts introduced the translation oracle at different levels to project different languages' representation into a unified space, including word, sentence, or document level [20, 28, 35, 49, 71, 74, 84, 88]. The performance of these methods heavily relies on the generated pseudo parallel texts from machine translation tools. However, the task-specific information like sentiment which may hold language-specific characteristic cannot be easily transferred in this way. In this study, we use easily-accessed emoji-texts to incorporate both common and language-specific knowledge into the representation learning models for the source and target languages. The implicit sentiment information contained in the diverse emotional emoji usage of the two languages can benefit the cross-lingual sentiment classification.

6 CONCLUSION

As a ubiquitous emotional signal, emojis are widely adopted across languages to express sentiment. Based on this nature, we leverage emojis both as surrogate sentiment labels and as the bridge to address the language discrepancy in cross-lingual sentiment classification. To be more specific, we have presented ELSA, an emoji-powered representation learning framework, to capture both common and language-specific sentiment knowledge of the source and target languages for cross-lingual sentiment classification. We first apply Word2Vec to large-scale Tweets to learn word representations for both languages, respectively. Then based on the word representations, we use an attention-based bi-directional LSTM model to learn informative sentence representations via an emoji prediction task. The representations can catch not only the common sentiment patterns across languages, but also the language-specific knowledge. Finally, based on the learned representation models, we employ the labeled English data to train a sentiment classifier for the target language. Our ELSA has been comprehensively evaluated on various benchmarks and outperforms the existing cross-lingual sentiment classification methods.

REFERENCES

[1] 2005. MPQA opinion corpus. https://mpqa.cs.pitt.edu/lexicons/subj_lexicon/. (2005). Retrieved on October 22, 2018.
[2] 2006. MeCab: Yet Another Part-of-Speech and Morphological Analyzer. http://taku910.github.io/mecab. (2006). Retrieved on October 22, 2018.
[3] 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.
[4] 2015. DEFT 2015: Test Corpus. https://deft.limsi.fr/2015/corpus.fr.php?lang=en. (2015). Retrieved on April 28, 2018.
[5] 2017. Deep-mlsa. https://github.com/spinningbytes/deep-mlsa. (2017). Retrieved on October 22, 2018.
[6] 2017. SB-10k: German Sentiment Corpus. https://www.spinningbytes.com/resources/germansentiment/. (2017). Retrieved on April 28, 2018.
[7] 2017. Twitter-2015test-A. http://alt.qcri.org/semeval2017/task4/index.php?id=download-the-full-training-data-for-semeval-2017-task-4. (2017). Retrieved on April 28, 2018.
[8] 2017. Twitter-2015train-A, Twitter-2016train-A, Twitter-2016dev-A, and Twitter-2016devtest-A. http://alt.qcri.org/semeval2017/task4/index.php?id=download-the-full-training-data-for-semeval-2017-task-4. (2017). Retrieved on April 28, 2018.
[9] 2018. Internet world users by language. https://www.internetworldstats.com/stats7.htm. (2018). Retrieved on October 22, 2018.
[10] 2018. ISO 639. https://www.iso.org/iso-639-language-codes.html. (2018). Retrieved on October 22, 2018.
[11] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. 2016. Tensorflow: A system for large-scale machine learning.. In *OSDI*, Vol. 16. 265–283.
[12] Wei Ai, Xuan Lu, Xuanzhe Liu, Ning Wang, Gang Huang, and Qiaozhu Mei. 2017. Untangling emoji popularity through semantic embeddings. In *Proceedings of the Eleventh International Conference on Web and Social Media, ICWSM 2017*. 2–11.
[13] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. In *Proceedings of the 3rd International Conference on Learning Representations, ICLR 2014*.
[14] Georgios Balikas, Simon Moura, and Massih-Reza Amini. 2017. Multitask learning for fine-grained Twitter sentiment analysis. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2017*. 1005–1008.
[15] Ziv Bar-Joseph, David K. Gifford, and Tommi S. Jaakkola. 2001. Fast optimal leaf ordering for hierarchical clustering. In *Proceedings of the Ninth International Conference on Intelligent Systems for Molecular Biology, July 21-25, 2001, Copenhagen, Denmark*. 22–29.
[16] Francesco Barbieri, Germán Kruszewski, Francesco Ronzano, and Horacio Saggion. 2016. How cosmopolitan are emojis?: Exploring emojis usage and meaning over different languages with distributional semantics. In *Proceedings of the 2016 ACM Conference on Multimedia Conference, MM 2016*. 531–535.
[17] Johan Bollen, Huina Mao, and Xiao-Jun Zeng. 2011. Twitter mood predicts the stock market. *Journal of Computational Science* 2, 1 (2011), 1–8.
[18] Samuel Brody and Nicholas Diakopoulos. 2011. Cooooooooooooooollllllllllllll!!!!!!!!!!!!!! Using word lengthening to detect sentiment in microblogs. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, EMNLP 2011*. 562–570.
[19] Rich Caruana, Steve Lawrence, and C. Lee Giles. 2000. Overfitting in neural nets: backpropagation, conjugate gradient, and early stopping. In *Proceedings of advances in neural information processing systems 13, NIPS 2000*. 402–408.
[20] A. P. Sarath Chandar, Stanislas Lauly, Hugo Larochelle, Mitesh M. Khapra, Balaraman Ravindran, Vikas C. Raykar, and Amrita Saha. 2014. An autoencoder approach to learning bilingual word representations. In *Advances in Neural Information Processing Systems 27, NIPS 2014*. 1853–1861.
[21] Qiang Chen, Chenliang Li, and Wenjie Li. 2017. Modeling language discrepancy for cross-lingual sentiment analysis. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017*. 117–126.
[22] Yuxiao Chen, Jianbo Yuan, Quanzeng You, and Jiebo Luo. 2018. Twitter sentiment analysis via bi-sense emoji embedding and attention-based LSTM. In *2018 ACM Multimedia Conference on Multimedia Conference, MM 2018*. 117–125.
[23] Zhenpeng Chen, Xuan Lu, Wei Ai, Huoran Li, Qiaozhu Mei, and Xuanzhe Liu. 2018. Through a gender lens: learning usage patterns of emojis from large-scale Android users. In *Proceedings of the Web Conference 2018, WWW 2018*. 763–772.
[24] Henriette Cramer, Paloma de Juan, and Joel R. Tetreault. 2016. Sender-intended functions of emojis in US messaging. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI 2016*. 504–509.
[25] Jan Deriu, Aurélien Lucchi, Valeria De Luca, Aliaksei Severyn, Simon Müller, Mark Cieliebak, Thomas Hofmann, and Martin Jaggi. 2017. 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*. 1045–1052.
[26] Nicholas Diakopoulos and David A. Shamma. 2010. Characterizing debate performance via aggregated Twitter sentiment. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems, CHI 2010*. 1195–1198.
[27] Thomas G Dietterich. 1998. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural computation* 10, 7 (1998), 1895–1923.
[28] Manaal Faruqui and Chris Dyer. 2014. Improving vector space word representations using multilingual correlation. In *Proceedings of the 14th conference of the european chapter of the association for computational linguistics, EACL 2014*. 462–471.
[29] Bjarke Felbo, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann. 2017. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017*. 1615–1625.
[30] Michael Gamon. 2004. Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis. In *Proceedings of 20th International Conference on Computational Linguistics, COLING 2004*.
[31] Anastasia Giachanou and Fabio Crestani. 2016. Like it or not: a survey of Twitter sentiment analysis methods. *Comput. Surveys* 49, 2 (2016), 28:1–28:41.
[32] Lin Gong, Benjamin Haines, and Hongning Wang. 2017. Clustered model adaption for personalized sentiment analysis. In *Proceedings of the 26th International Conference on World Wide Web, WWW 2017*. 937–946.
[33] Alex Graves, Abdel-rahman Mohamed, and Geoffrey E. Hinton. 2013. Speech recognition with deep recurrent neural networks. In *IEEE International conference on acoustics, speech and signal processing, ICASSP 2013*. 6645–6649.
[34] Kiraz Candan Herdem. 2012. Reactions: Twitter based mobile application for awareness of friends' emotions. In *The 2012 ACM Conference on Ubiquitous Computing, UbiComp 2012*. 796–797.
[35] Karl Moritz Hermann and Phil Blunsom. 2014. Multilingual models for compositional distributed semantics. In *Proceedings of the 52nd annual meeting of the association for computational linguistics, ACL 2014*. 58–68.
[36] M. Hermans and B. Schrauwen. 2013. Training and analysing deep recurrent neural networks. *Proceedings of advances in Neural Information Processing Systems* (2013), 190–198.
[37] Sepp Hochreiter. 1998. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 6, 2 (1998), 107–116.
[38] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9, 8 (1997), 1735–1780.
[39] Tianran Hu, Han Guo, Hao Sun, Thuy-vy Thi Nguyen, and Jiebo Luo. 2017. Spice up your chat: the intentions and sentiment effects of using emojis. In *Proceedings of the Eleventh International Conference on Web and Social Media, ICWSM 2017*. 102–111.
[40] Diederik P. Kingma and Jimmy Ba. 2014. Adam: a method for stochastic optimization. *CoRR* abs/1412.6980 (2014).
[41] Xiao Ling, Gui-Rong Xue, Wenyuan Dai, Yun Jiang, Qiang Yang, and Yong Yu. 2008. Can Chinese Web pages be classified with English data source?. In *Proceedings of the 17th International Conference on World Wide Web, WWW 2008*. 969–978.
[42] Bing Liu. 2012. *Sentiment analysis and opinion mining*. Morgan & Claypool Publishers.
[43] Kun-Lin Liu, Wu-Jun Li, and Minyi Guo. 2012. Emoticon smoothed language models for Twitter sentiment analysis. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2012*.
[44] Qiao Liu, Haibin Zhang, Yifu Zeng, Ziqi Huang, and Zufeng Wu. 2018. Content attention model for aspect based sentiment analysis. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web, WWW 2018*. 1023–1032.
[45] Yang Liu, Xiangji Huang, Aijun An, and Xiaohui Yu. 2007. ARSA: a sentiment-aware model for predicting sales performance using blogs. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2007*. 607–614.
[46] Nikola Ljubesic and Darja Fiser. 2016. A global analysis of emoji usage. In *Proceedings of the 10th Web as Corpus Workshop, WAC@ACL 2016*. 82–89.

[47] Xuan Lu, Wei Ai, Xuanzhe Liu, Qian Li, Ning Wang, Gang Huang, and Qiaozhu Mei. 2016. Learning from the ubiquitous language: an empirical analysis of emoji usage of smartphone users. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp 2016*. 770–780.

[48] Washington Luiz, Felipe Viegas, Rafael Odon de Alencar, Fernando Mourão, Thiago Salles, Dárlinton Carvalho, Marcos André Gonçalves, and Leonardo C. da Rocha. 2018. A feature-oriented sentiment rating for mobile app reviews. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web, WWW 2018*. 1909–1918.

[49] Thang Luong, Hieu Pham, and Christopher D. Manning. 2015. Bilingual word representations with monolingual quality in mind. In *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing, VS@NAACL-HLT 2015*. 151–159.

[50] Mary McGlohon, Natalie S. Glance, and Zach Reiter. 2010. Star quality: aggregating reviews to rank products and merchants. In *Proceedings of the Fourth International Conference on Weblogs and Social Media, ICWSM 2010*.

[51] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *Computer Science* (2013).

[52] Hannah Jean Miller, Daniel Kluver, Jacob Thebault-Spieker, Loren G. Terveen, and Brent J. Hecht. 2017. Understanding emoji ambiguity in context: the role of text in emoji-related miscommunication. In *Proceedings of the Eleventh International Conference on Web and Social Media, ICWSM 2017*. 152–161.

[53] Hannah Jean Miller, Jacob Thebault-Spieker, Shuo Chang, Isaac L. Johnson, Loren G. Terveen, and Brent J. Hecht. 2016. "Blissfully happy" or "ready to fight": varying interpretations of emoji. In *Proceedings of the Tenth International Conference on Web and Social Media, ICWSM 2016*. 259–268.

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

[55] Brendan O'Connor, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. 2010. From Tweets to polls: linking text sentiment to public opinion time series. In *Proceedings of the Fourth International Conference on Weblogs and Social Media, ICWSM 2010*.

[56] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. 2013. On the difficulty of training recurrent neural networks. In *Proceedings of the 30th International Conference on Machine Learning, ICML 2013*. 1310–1318.

[57] Umashanthi Pavalanathan and Jacob Eisenstein. 2015. Emoticons vs. emojis on Twitter: a causal inference approach. *CoRR* abs/1510.08480 (2015).

[58] Stefania Pecore and Jeanne Villaneau. 2018. Complex and precise movie and book annotations in French language for aspect based sentiment analysis. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation, LREC 2018*.

[59] J. W. Pennebaker, L. E. Francis, and R. J. Booth. 1999. Linguistic inquiry and word count: LIWC. *Lawrence Erlbaum Associates Mahwah Nj* (1999).

[60] Hieu Pham, Thang Luong, and Christopher D. Manning. 2015. Learning distributed representations for multilingual text sequences. In *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing, VS@NAACL-HLT 2015*. 88–94.

[61] Henning Pohl, Christian Domin, and Michael Rohs. 2017. Beyond just text: semantic emoji similarity modeling to support expressive communication. *ACM Transactions on Computer-Human Interaction, TOCHI* 24, 1 (2017), 6:1–6:42.

[62] Henning Pohl, Dennis Stanke, and Michael Rohs. 2016. EmojiZoom: emoji entry via large overview maps. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI 2016*. 510–517.

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

[64] Michal Ptaszynski, Rafal Rzepka, Kenji Araki, and Yoshio Momouchi. 2012. Automatically annotating a five-billion-word corpus of Japanese blogs for affect and sentiment analysis. In *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis, WASSA@ACL 2012*. 89–98.

[65] Qiao Qian, Minlie Huang, Jinhao Lei, and Xiaoyan Zhu. 2017. Linguistically regularized LSTM for sentiment classification. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017*. 1679–1689.

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

[67] Phillipe Rodrigues, Ismael Santana Silva, Glívia Angélica Rodrigues Barbosa, Flávio Roberto dos Santos Coutinho, and Fernando Mourão. 2017. Beyond the stars: towards a novel sentiment rating to evaluate applications in Web stores of mobile apps. In *Proceedings of the 26th International Conference on World Wide Web, WWW 2017*. 109–117.

[68] Anna Rogers, Alexey Romanov, Anna Rumshisky, Svitlana Volkova, Mikhail Gronas, and Alex Gribov. 2018. RuSentiment: An enriched sentiment analysis dataset for social media in Russian. In *Proceedings of the 27th International Conference on Computational Linguistics, COLING 2018*. 755–763.

[69] Koustuv Saha, Larry Chan, Kaya de Barbaro, Gregory D. Abowd, and Munmun De Choudhury. 2017. Inferring mood instability on social media by leveraging ecological momentary assessments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, IMWUT* 1, 3 (2017), 95:1–95:27.

[70] Asif Salekin, Zeya Chen, Mohsin Y. Ahmed, John Lach, Donna Spruijt-Metz, Kayla de la Haye, Brooke Bell, and John A. Stankovic. 2017. Distant emotion recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, IMWUT* 1, 3 (2017), 96:1–96:25.

[71] Duyu Tang, Bing Qin, and Ting Liu. 2015. Document modeling with gated recurrent neural network for sentiment classification. In *Proceedings of the 2015 conference on empirical methods in natural language processing, EMNLP 2015*. 1422–1432.

[72] Channary Tauch and Eiman Kanjo. 2016. The roles of emojis in mobile phone notifications. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp Adjunct 2016*. 1560–1565.

[73] Mike Thelwall, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas. 2010. Sentiment in short strength detection informal text. *JASIST* 61, 12 (2010), 2544–2558.

[74] Ivan Vulic and Marie-Francine Moens. 2015. Bilingual word embeddings from non-parallel document-aligned data applied to bilingual lexicon induction. In *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international Joint Conference on natural language processing of the asian federation of natural language processing, ACL 2015*. 719–725.

[75] Jin Wang, Liang-Chih Yu, K. Robert Lai, and Xue-Jie Zhang. 2016. Dimensional sentiment analysis using a regional CNN-LSTM Model. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016*.

[76] Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao. 2016. Attention-based LSTM for aspect-level sentiment classification. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016*. 606–615.

[77] Yequan Wang, Aixin Sun, Jialong Han, Ying Liu, and Xiaoyan Zhu. 2018. Sentiment analysis by capsules. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web, WWW 2018*. 1165–1174.

[78] Fangzhao Wu, Jia Zhang, Zhigang Yuan, Sixing Wu, Yongfeng Huang, and Jun Yan. 2017. Sentence-level sentiment classification with weak supervision. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2017*. 973–976.

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

[80] Jiang Yang, Lada A. Adamic, Mark S. Ackerman, Zhen Wen, and Ching-Yung Lin. 2012. The way I talk to you: sentiment expression in an organizational context. In *Proceedings of the 2012 International Conference on Human Factors in Computing Systems, CHI 2012*.

[81] Min Yang, Wenting Tu, Jingxuan Wang, Fei Xu, and Xiaojun Chen. 2017. Attention based LSTM for target dependent sentiment classification. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, AAAI 2017*. 5013–5014.

[82] Taras Zagibalov and John Carroll. 2008. Automatic seed word selection for unsupervised sentiment classification of Chinese text. In *Proceedings of the 22nd International Conference on Computational Linguistics, COLING 2008*. 1073–1080.

[83] Jichang Zhao, Li Dong, Junjie Wu, and Ke Xu. 2012. MoodLens: an emoticon-based sentiment analysis system for Chinese Tweets. In *Proceedings of the 18th ACM International Conference on Knowledge Discovery and Data Mining, KDD 2012*. 1528–1531.

[84] Huiwei Zhou, Long Chen, Fulin Shi, and Degen Huang. 2015. Learning bilingual sentiment word embeddings for cross-language sentiment classification. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015*. 430–440.

[85] Joey Tianyi Zhou, Sinno Jialin Pan, Ivor W. Tsang, and Shen-Shyang Ho. 2016. Transfer learning for cross-language text categorization through active correspondences construction. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, AAAI 2016*. 2400–2406.

[86] Rui Zhou, Jasmine Hentschel, and Neha Kumar. 2017. Goodbye text, hello emoji: mobile communication on WeChat in China. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI 2017*. 748–759.

[87] Xinjie Zhou, Xiaojun Wan, and Jianguo Xiao. 2016. Attention-based LSTM network for cross-lingual sentiment classification. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016*. 247–256.

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