



A Survey of Cross-lingual Sentiment Analysis: Methodologies, Models and Evaluations

Yuemei Xu¹ · Han Cao¹ · Wanze Du¹ · Wenqing Wang¹

Received: 23 February 2022 / Revised: 29 April 2022 / Accepted: 24 May 2022 / Published online: 8 June 2022
© The Author(s) 2022

Abstract

Cross-lingual sentiment analysis (CLSA) leverages one or several source languages to help the low-resource languages to perform sentiment analysis. Therefore, the problem of lack of annotated corpora in many non-English languages can be alleviated. Along with the development of economic globalization, CLSA has attracted much attention in the field of sentiment analysis and the last decade has seen a surge of researches in this area. Numerous methods, datasets and evaluation metrics have been proposed in the literature, raising the need for a comprehensive and updated survey. This paper fills the gap by reviewing the state-of-the-art CLSA approaches from 2004 to the present. This paper teases out the research context of cross-lingual sentiment analysis and elaborates the following methods in detail: (1) The early main methods of CLSA, including those based on Machine Translation and its improved variants, parallel corpora or bilingual sentiment lexicon; (2) CLSA based on cross-lingual word embedding; (3) CLSA based on multi-BERT and other pre-trained models. We further analyze their main ideas, methodologies, shortcomings, etc., and attempt to reach a conclusion on the coverage of languages, datasets and their performance. Finally, we look into the future development of CLSA and the challenges facing the research area.

Keywords Cross-lingual · Sentiment analysis · Bilingual word embedding

1 Introduction

Cross-lingual sentiment analysis (CLSA) leverages one or several source languages (usually resource-rich languages, such as English) to help the low-resource languages (termed as target language) perform sentiment analysis tasks. CLSA belongs to the field of sentiment analysis, which mines subjective information in texts to determine their sentimental tendencies (e.g., positive or negative). Most of the existing sentiment analysis methods are supervised learning methods, which rely on a large amount of annotated corpora to predict and analyze the sentiment polarity of unlabeled data. However, sentiment-annotated corpora are difficult to obtain, especially for non-English languages. The English language has accumulated rich sentiment resources, such as

annotated corpora and sentiment lexicon. However, for other languages, researches on sentiment analysis and sentiment-annotated resources are very scarce. In order to overcome the annotation problem in non-English languages, CLSA was first proposed by Yan et al. in 2004 who attempted to solve cross-lingual sentiment analysis problems through Machine Translation [1].

Many studies have shown that CLSA is capable of leveraging sentiment resources from resource-rich language to predict sentiment polarities of a resource-scarce language. For example, Wan et al. (2008) used sentiment-annotated data of English to achieve Chinese texts sentiment classification through Machine Translation [2]. Compared with monolingual sentiment analysis, the most imperative problem to be solved in CLSA is how to bridge the language gap. Although CLSA has achieved outstanding performance in some target languages, there are still many problems in practical applications, e.g., the methods based on Machine Translation cannot avoid the problem of Vocabulary Coverage, as the texts translated from a source language cannot cover all the words in the target language.

This paper reviews the researches of CLSA from 2004 to 2021, especially in the past decade, and discusses the

Han Cao, Wanze Du and Wenqing Wang have contributed equally to this work.

✉ Yuemei Xu
xuyuemei@bfsu.edu.cn

¹ School of Information Science and Technology, Beijing Foreign Studies University, Haidian, Beijing 100089, China

development of CLSA comprehensively. In order to obtain the CLSA studies thoroughly, we took the Web of Science database as the retrieval platform and constructed the retrieval formula $TS = \text{cross lingual sentiment OR cross lingual embedding}$ and selected 656 studies with high relevance to CLSA, 92 of which were selected as references after extended reading, as shown in Fig. 1.

This paper expounds the researches of CLSA according to their research timelines and methodologies. In the early stage of CLSA, four approaches are mainly adopted, which are CLSA based on Machine Translation and its improved variants, CLSA based on Parallel Corpora, CLSA based on Structural Correspondence Learning, CLSA based on Bilingual Sentiment Lexicon Construction, respectively. Since Mikolov et al. (2013) proposed the distributed word vectors model Word2Vec [3], CLSA has entered a new era. With the aid of word embedding representation, cross-lingual word embedding models, CLSA based on Generation Adversarial Network, and CLSA based on Pre-Trained Model have become the mainstream. And it has gradually developed from Supervised methods to Weakly Supervised and finally to Fully Unsupervised methods. The contributions of this paper are as follows:

- This paper summarizes the methods of CLSA systematically, classifies the existing researches according to their main ideas as well as methodologies, and summarizes their main ideas, methodologies, shortcomings, etc.
- This paper gives an overview of the source-target language pairs, the evaluated data sets and performance

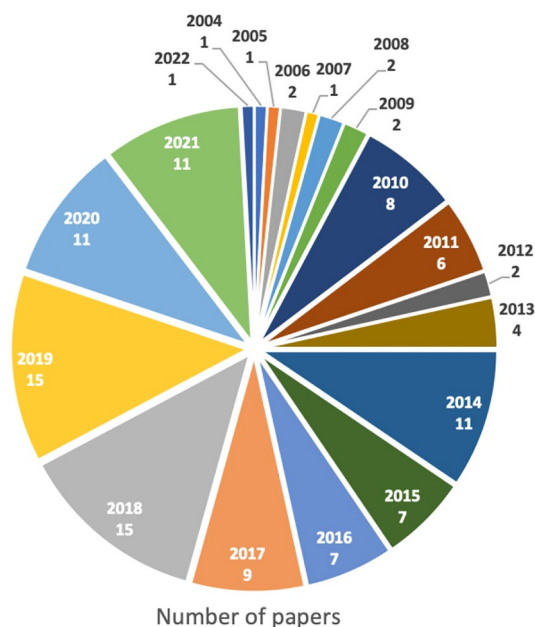


Fig. 1 Number of surveyed papers per year

evaluations of CLSA. To the best of our knowledge, the CLSA experiments are only conducted on a few language pairs, which limits the promotion of CLSA models to a certain extent. For example, it mainly takes English as the source language and French, German, Japanese, Chinese and Spanish as the target languages. However, the language-sensitive problem may exist when CLSA models perform on different target languages. Therefore, this paper gives an overview of the evaluated language pairs and the performance evaluations on these languages, so as to provide inspired ideas for the researches.

- Finally, this paper draws the conclusions of the two points above and analyzes the important challenges, problems and future directions for the research of CLSA.

The rest of the paper is organized as follows. In Sect. 2, we further expound the early stage of CLSA and analyze and compare several typical models. In Sect. 3, we deeply analyze the researches of cross-lingual word embedding as well as several state-of-the-art models. In Sect. 4, we give a thorough look on the advanced work adopted to CLSA - Generation Adversarial Network (GAN) and Pre-trained Model (PTM). In Sect. 5, we look into the future development of CLSA and the challenges facing the research area.

2 Early Cross-lingual Sentiment Analysis

2.1 CLSA based on Machine Translation and its Improved Variants

As a pioneering work, Yan et al. first used Machine Translation (MT) to perform CLSA tasks in 2004 [1]. In the following decade, MT has always been the mainstreaming method of CLSA. Its basic idea is to use MT system to directly translate one language into another [4]- [5], so as to predict unlabeled data in target language by using the annotated data in source language. As shown in Fig. 2(a), the present studies mostly used MT system to translate annotated training data from the source language into the target language [4, 6], and trained the sentiment classifiers for the target language by the translated annotated data, and then performed sentiment analysis on the target language. Some researchers used MT system in the opposite direction to translate unlabeled data from the target language into the source language and performed the sentiment analysis in the source language [7, 8] as shown in Fig. 2(b). Besides, some researches utilized bidirectional MT system, as shown in Fig. 2(b), simultaneously translating source language into target language and translating target language into source language, in order to eliminate translation limitations [9, 10].

However, as the language gap exists, the target language has a fixed intrinsic structure distinct from the source

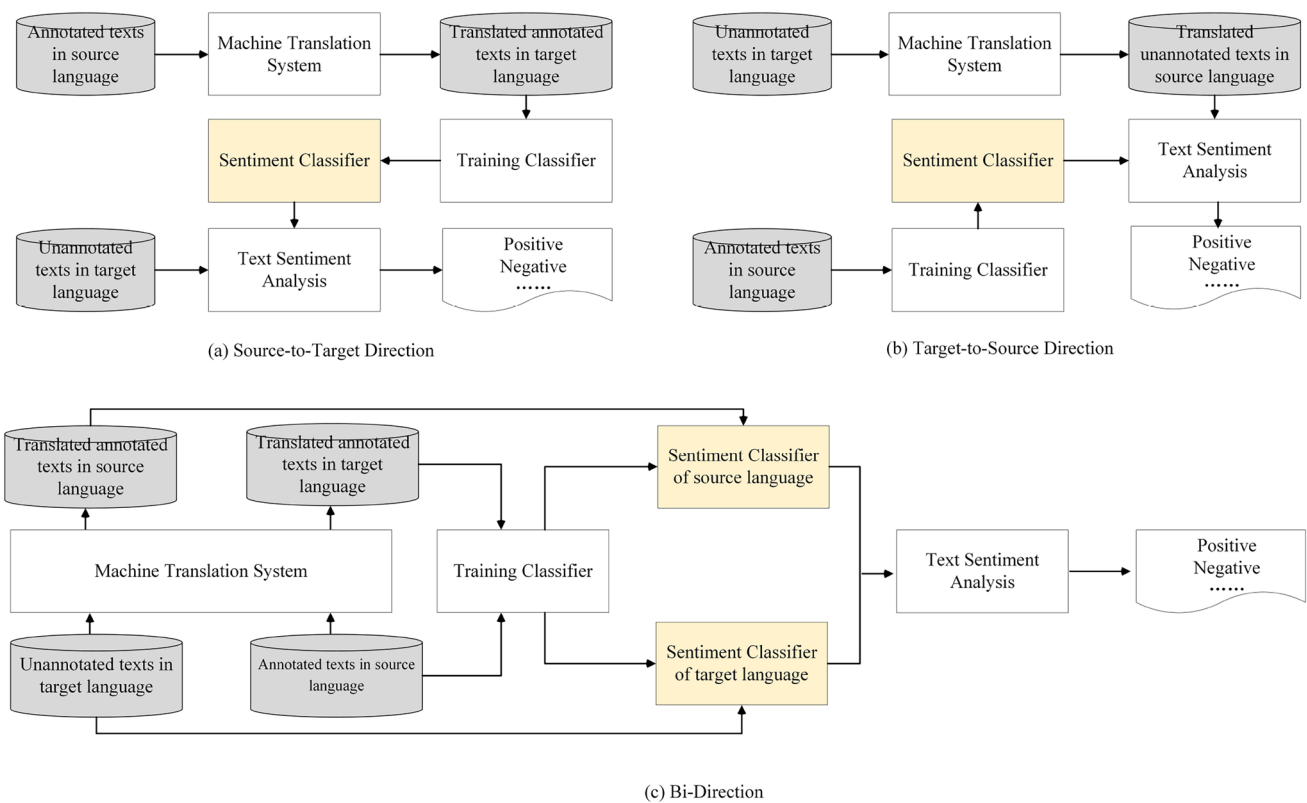


Fig. 2 Structure of Cross-lingual Sentiment Analysis based on Machine Translation

language. Previous studies found that even if the best MT system is adopted, about 10% distortion or reversal of sentiment is caused by MT [11]. In order to overcome the influence of the quality of MT on CLSA, relevant studies try to make some improvements on traditional MT system. The improvement methods mainly include: learning from the translation of source language sentiment lexicons [12], refining the training samples [13], finding the optimal baseline model for sentiment classification [14], using annotated data from multiple source languages [15], incorporating unlabeled texts from the target language [16] and incorporating Distortion Tolerance to balance the reversal of sentiment caused by MT system [5].

Table 1 shows the representative researches in early CLSA, and those works marked with * are CLSA researches based on MT and its improved variants, mainly published on 2011-2017. CLSA based on MT is a supervised method, which suffers from the generalization problem, especially when there is a domain mismatch between source and target languages. In order to solve this problem, He et al. (2011) proposed a weakly supervised model called latent sentiment model (LSM) which was based on the latent Dirichlet allocation (LDA) model combined with prior sentiment knowledge learned from sentiment lexicons [12]. Compared with other models based on LDA, LSM model incorporates

prior sentiment knowledge into the process of sentiment preference computation using generalized expectation criteria. LSM model uses sentiment preferences to create an informed prior distribution for the sentiment labels, which allows the model to extract domain-specific sentiment polarity of words. Experiments show that LSM model performs comparably to supervised classifiers like support vector machine (SVM) trained with annotated corpora.

Cross-lingual sentiment learning is a challenging task due to the different distributions between source and target languages and the language gap. Therefore, Zhang et al. (2016) proposed the Similarity Discovery plus Training Data Adjustment (SD-TDA) model to refine the training data of the source language to eliminate the different distributions and the language gap between two languages [13]. SD-TDA model maps words from source and target language data into a common concept space through the aligned-translation topic model to alleviate the distribution discrepancy. After that, it utilizes a semi-supervised learning model to further refine the training data to reduce the language gap.

Al-Shabi et al. (2017) attempted to find an optimal baseline model [14]. The proposed model uses annotated Arabic data translated from English through MT to train several classifiers, such as K-Nearest Neighbor (KNN), Naive Bayes (NB) and SVM, and then uses the trained classifiers

Table 1 Representative Researches in early CLSA

Author	Model	Characteristics	Data sets	Languages	Accuracy /%
He et al. (2011)* [12]	LSM	Learn from the translation of source language sentiment lexicons	Chinese product review corpora	ch-en	81.41
Zhang et al. (2016))* [13]	ATTM	Refine the training samples	COAE2014	ch-de ch-en ch-fr ch-es	84.3 87.7 80.1 83.3
Al-Shabi et al. (2017))* [14]	SVM, NB, KNN	Find the optimal baseline model for sentiment classification	Amazon product reviews	en-ar	–
Hajmohammadi et al. (2014))* [15]	MLMV	Use annotated data from multiple source languages	Amazon product reviews; Pan Reviews dataset	en/de-fr en/fr-de en/fr-jp en/ja-ch	79.85 81.55 73.73 76.65
Hajmohammadi et al.(2015))* [16]	DBAST	Incorporate unlabeled texts from the target language	Amazon product reviews; Pan Reviews dataset	en-fr en-ch en-jp	78.63 71.36 70.04
Hajmohammadi et al. (2015))* [5]	Graph-based Semi-supervised Learning Model	Incorporate Distortion Tolerance to balance the reversal of sentiment caused by MT system	Amazon product reviews; Pan Reviews dataset	en-ch en-jp	73.81 73.81
Lu et al. (2011)) ⁻ [17]	Joint	Leverage unlabeled data of target language	MPQA; NTCIR-EN; NTCIR-CH; ISI Chinese-English parallel corpora	ch-en en-ch	83.54 79.29
Meng et al. (2012)) ⁻ [18]	CLMM	Learn sentiment words from unlabeled bilingual parallel corpora to expand the vocabulary coverage	MPQA; NTCIR-EN; NTCIR-CH; ISI Chinese-English parallel corpora	ch-en	83.02
Gao et al. (2014)) ⁻ [19]	BLP	Generate sentiment lexicons of target language through parallel corpora	General Inquirer lexicon; ISI Chinese-English parallel corpora; WordNet; NTCIR corpus	ch-en	78.90
Zhou et al. (2015)) ⁻ [20]	NMF	Leverage small amount of parallel data and large-scale non-parallel data	Amazon product reviews	en-fr en-de en-jp fr-en de-en jp-en	81.83 80.45 75.78 79.47 79.56 78.79

to predict the sentiment classification of the target language. They studied the effects on the quality of the Machine Translation so as to determine the extent to which the noise of translation in training data may affect the accuracy of sentiment classification.

For the improvement of Machine Translation methods, Hajmohammadi et al. have made a series of related explorations and put forward inspired ideas [5, 15, 16]. First, due to the existence of different linguistic terms and writing styles between different languages, translated data from source language cannot cover all vocabularies of target language, which would influence the performance of CLSA tasks if using the translated data through MT as the training data. In order to overcome the problem of Vocabulary Coverage,

they proposed a new model in 2014 which used annotated data from multiple source languages in a multi-view semi-supervised learning approach [15].

Later, Hajmohammadi et al. proposed a novel learning model in 2015 [16]. Unlabeled sentiment documents from the target language were incorporated to improve the performance of CLSA methods, and the density measures of unlabeled examples were considered in active learning part to avoid outlier selection. Finally, to overcome the different term distribution between source language and target language caused by MT, which means a term may be frequently used in source language to express an opinion, while the translation of that term is rarely used in the target language, they proposed a graph-based semi-supervised learning

model in 2015 [5]. This model used sentiment information of unlabeled data as well as annotated data in a graph-based semi-supervised learning approach, so as to incorporate the intrinsic structure of unlabeled data from the target language into the learning process.

It can be seen from the above researches that in order to overcome the problems of generalization, vocabulary coverage and language gap between source and target language, lots of explorations have been made. By translating target (source) language into source (target) language, CLSA based on MT and its improved variants have achieved good results. However, there is still not a generalized MT system which performs well on all of the targeted languages. Moreover, most of these improved variants use Amazon Product review data sets, which are not diverse enough to fully support and reflect the performance of the improved methods.

2.2 CLSA based on Parallel Corpora

A parallel corpus is a corpus composed of mutually translated texts. Instead of using MT system, the association between source language and target language is completely established by Parallel Corpora or Comparable Corpora [21], which is also one of the main methods for early stage of CLSA.

Parallel corpora contain a large number of parallel sentence pairs. By connecting two aligned words in parallel corpora, the mapping between languages can be quickly constructed. Figure 3 (b) shows an example of the parallel sentence pairs of Chinese and English, and the word-level alignment is easily obtained from the sentence pair. Figure 3(a) takes words of two languages in a parallel corpus as nodes and establishes connections between languages through word alignment, synonyms, antonyms and other information in the corpus.

In Table 1, related works marked with the symbol ⁻ are researches of CLSA based on parallel corpus in recent years. Their main ideas are as follows: leverage unlabeled data of

target language [17], expand vocabulary coverage by learning parallel data [18], generate sentiment lexicons of target language through parallel corpora [19], and leverage small amount of parallel data and large-scale non-parallel data [20].

Lu et al. (2011) first proposed to use unlabeled parallel corpora to improve the performance of monolingual sentiment classifiers [17]. They believed that the parallel sentences in the unlabeled corpus should also have the same sentiment polarity. Therefore, an unlabeled parallel corpus together with the labeled data available for each language are used to train the monolingual sentiment classifiers simultaneously. Experiment results showed that the proposed method was superior to the monolingual baseline method, and improved the accuracy of sentiment classification in both languages by 3.44%-8.12%.

CLSA based on Machine Translation suffers from the vocabulary coverage problem. Therefore, instead of relying on unreliable Machine Translation to obtain annotated data of target language, Meng et al. (2012) proposed a Generative Cross-lingual Mixture Model (CLMM) [18]. CLMM leverages unlabeled bilingual parallel data to learn sentiment words in source and target language and improves vocabulary coverage. In order to automatically generate sentiment lexicons for the target languages with available English sentiment lexicons, Gao et al. (2014) proposed a bilingual word graph method based on parallel corpora and word alignment derived from a large parallel corpus [19].

As large-scale document-aligned or sentence-aligned parallel data is difficult to obtain, there is usually only a few of parallel data but a large amount of non-aligned text in different languages. Zhou et al. proposed a subspace learning framework to learn from small-scale document-aligned data between source language and target language and large-scale non-parallel data simultaneously [20].

CLSA based on MT and CLSA based on Parallel Corpus have some similarities and there are overlaps between these two methods, but the main ideas of these two methods are different. The main idea of CLSA based on MT is translating

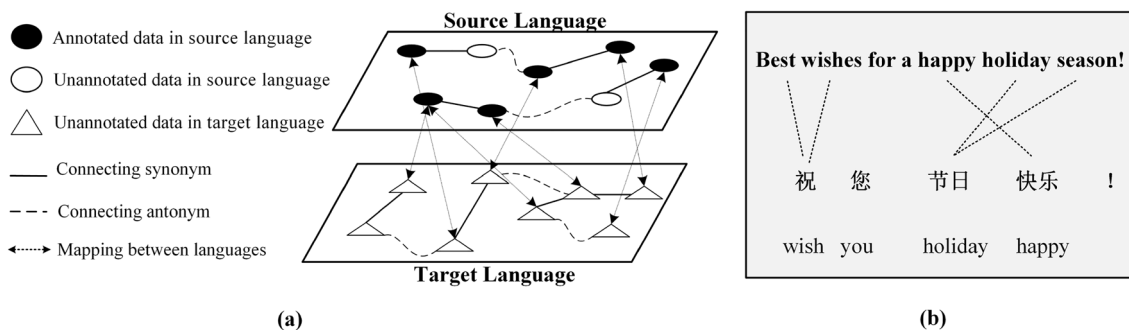


Fig. 3 Structure of CLSA based on parallel corpora

one language into another utilizing MT system and then CLSA tasks can be regarded as monolingual sentiment. In this process, parallel corpora are leveraged to train the MT system. However, the main idea of CLSA based on parallel corpus is to leverage bilingual parallel corpora to obtain the corresponding mapping relationship of words of source and target languages, which can avoid the noise caused by MT systems and bridge the differences of term distribution and intrinsic structure between source and target languages semantically and conceptually. At the same time, it is worth noticing that the traditional CLSA methods based on parallel corpus require a large scale of parallel or annotated data, which are often difficult to obtain, especially in low-resource languages. Therefore, many existing studies were proposed to reduce the dependence on parallel and annotated data, such as using comparable corpora, non-parallel or unlabeled data, etc.

2.3 CLSA based on Structural Correspondence Learning

Structural Correspondence Learning (SCL) was proposed by Blitzer et al. in 2006 [22], which is also one of the main methods for early phase of CLSA. In this method, the correspondences between the source language and the target language can be discovered based on feature transfer and then texts from different languages are mapped into the same feature space. Finally, cross-lingual sentiment analysis can be achieved through this feature space projection.

Figure 4 illustrates the schematic diagram of CLSA methods based on SCL, which requires annotated as well as unannotated documents in source language and unlabeled

documents in target language. The first step is to select pivots. Words which may help sentiment prediction are selected from the annotated documents of source language. The translated pairs of these words are called Pivots. Then, a linear classifier is trained to model the correlations between each pivot and all other words, which can predict the occurrence of pivot words in documents based on other words. Finally, the projection function is obtained by Singular Value Decomposition (SVD) to realize the knowledge transfer between the two languages.

Prettenhofer et al. (2010) [8] proposed to use unlabeled documents along with little pivots to automatically induce the generation of cross-lingual word pairs. It can greatly reduce the cost of resource computation by requiring unlabeled documents and a few translated words. Based on the former approach, Prettenhofer et al. formally published the method based on Structural Correspondence Learning for CLSA in the same year [23]. The key idea is to capture the pivot features of the source language and the target language and generate a mapping matrix to represent their correspondence by using the correlation modeling of the pivot features.

Wang et al. (2017) proposed a Cross-lingual Structural Correspondence Learning (M-CLSCL) algorithm by using the selected word axis based on SCL and Laplacian Mapping algorithm for sentiment analysis [24]. The main idea is to use the diagonal matrix of the classifier to construct the Laplacian matrix and then use the constructed matrix to solve the eigenvalues to form a mapping function, so as to predict the sentiment analysis of the target language.

It can be seen from the above studies that the SCL method has high efficiency in resource acquisition and computation, and the accuracy of sentiment classification in different

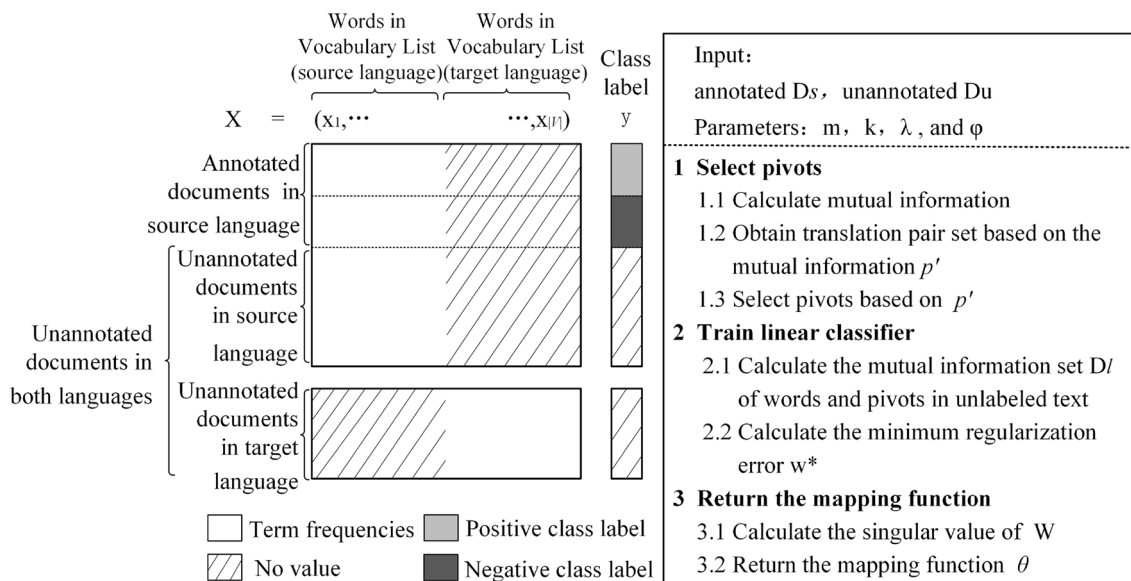


Fig. 4 Schematic diagram of CLSA method based on SCL

language data sets is mostly higher than that of traditional Machine Translation methods, especially in Japanese data sets. However, for simplicity, the SCL method only adopts one-to-one mapping for each pivot of the source language. Similarly, early MT system also simply replaced words in one language with words in another language. As we mentioned earlier, even if the best MT system is adopted, CLSA is still not comparable to the performance of monolingual sentiment classification [12].

Therefore, the restriction of one-to-one mapping between two languages in Structural Correspondence Learning method is too strict, and the accuracy of the sentiment analysis is severely affected. CLSA based on structural correspondence learning is no longer the mainstreaming method today.

2.4 CLSA based on Bilingual Sentiment Lexicon

Document-level or sentence-level Machine Translation (MT) is prone to introduce large translation errors, while word-level MT can overcome this shortage to some extent. Therefore, Bilingual Sentiment Lexicon was proposed for cross-language sentiment analysis. Compared with supervised methods, such as those based on Machine Learning, CLSA based on sentiment lexicon is an unsupervised method and does not rely on a large amount of annotated training data. The sentiment score of the text is obtained by constructing the bilingual sentiment lexicon and computing the sentiment score of each word in the target language text as an important basis to judge the polarity of the text sentiment.

In particular, if the bilingual sentiment lexicon can be established in advance, CLSA can be performed without any annotated data of source language and target language, e.g., SentiWordNet [25] is a well-established English sentiment lexicon. SentiWordNet lists each word with its sentiment polarity (positive/negative) and how strong the polarity is (on a score). For example, the sentiment score of the given sentence “I like this book” is computed by adding the sentiment score of each word to classify the sentence sentiment polarity.

Recently, some researchers mainly focus on the sub-task of bilingual sentiment lexicon construction to accomplish CLSA tasks. The existing methods for bilingual sentiment lexicon construction are mainly based on Machine Translation, Synset, or parallel corpus.

The bilingual sentiment lexicon construction based on MT is much simple through translating the existing monolingual sentiment lexicons of the source language into the target language. Darwich et al. (2016) mapped Indonesian WordNet and English WordNet through machine translation to obtain a sentiment lexicon of Malaysian. This method performs well for resource-rich languages, but for

resource-scarce languages, the accuracy is much low, with only 0.563 of accuracy after 5 iterations [26].

The bilingual sentiment lexicon construction based on Synset leverages the existing monolingual Synsets and obtains a cross-lingual sentiment lexicon through mapping methods. Nasharuddin (2017) proposed a cross-lingual sentiment lexicon acquisition to map Malaysian Sentiment lexicon to English Sentiment lexicon according to synset and Part of Speech (POS) [27]. Sazzed (2020) obtained Bengali synsets through English WordNet and Bengali comment corpus to generate Bengali sentiment lexicon [28].

The method based on parallel corpus is one of the most popular methods to construct bilingual sentiment lexicon in recent years. This method constructs bilingual sentiment lexicon by analyzing and extracting word alignment from parallel corpora of two languages. Vania et al., through parallel corpora of English and Indonesian, extracted sentiment words according to the extracted Senti-Pattern to construct a bilingual sentiment lexicon [29]. Chang et al. trained monolingual word vectors with contexts through skip-gram method, utilizing multilingual WordNet corpus to construct synonym and antonym relations for each language, which served as the pseudo contexts for skip-gram to help better distinguish the different sentiment polarity between words with similar contexts. Then, a translation matrix was learned through linear transformation and a small bilingual dictionary to map the word vector space of source language to that of the target language, so that the bilingual lexicon was constructed [30].

After the sub-task of bilingual sentiment lexicon accomplished, the researchers can easily conduct CLSA based on bilingual sentiment lexicon. For example, Gao et al. used the Chinese-English bilingual sentiment lexicon constructed by themselves to perform CLSA tasks [31]. He et al. used Convolutional Neural Network (CNN) to analyze sentiment polarity of Chinese-Vietnamese news based on Chinese-Vietnamese bilingual lexicon [32]. Zabha et al. leveraged Chinese-Malay bilingual sentiment lexicon and Term Counting method to make sentiment classification on Malay twitter data [33]. These researches show the feasibility of CLSA methods based on bilingual sentiment lexicon. In these methods, the performance of CLSA depends on the quality of the generated bilingual sentiment lexicon, as well as the model they adopted, such as CNN approach in [32] and sentiment score counting method in [33].

3 Cross-lingual Word Embedding

With the successively development of word vector representation models, such as Word2Vec, GloVe and ELMO, cross-lingual word embedding approaches were introduced into

the research of CLSA. Therefore, instead of sticking in how to improve the supervised methods based in Machine Translation or Parallel Corpora, researches on CLSA gradually developed from Supervised methods to Semi-Supervised and finally to fully unsupervised methods, which brought CLSA into a new era.

Researches of cross-lingual word embedding aim to represent the word vector of source and target language in the same semantic space. Thus, words with the same meaning and from different languages should have the same or similar vector representation.

Figure 5 shows the distribution of a group of English and Chinese words mapped to the same space, respectively, with English words marked in red and Chinese words marked in black. In this same semantic space, words with the same meaning or similar semantics from English and Chinese have similar distances. If different languages can be projected into the same semantic space, the annotated data of source language can be leveraged to predict sentiment polarity of target language.

In recent years, cross-lingual word embedding models have achieved fruitful results. Depending on whether bilingual parallel corpora are used or not, CLWE models can be categorized into three approaches: the Supervised Approach, the Semi-Supervised Approach, and the Unsupervised Approach. In its early stage, the Supervised Approach was mainly adopted, relying on expensive manually annotated corpora between source language and target language, such as cross-lingual parallel corpora or seed dictionary of word-level alignment [34], sentence-level alignment [35], and document-level alignment [36] served as cross-lingual supervised information. For most languages, however, such parallel corpora and seed dictionaries are not readily available. Therefore, the Semi-Supervised Approach was put

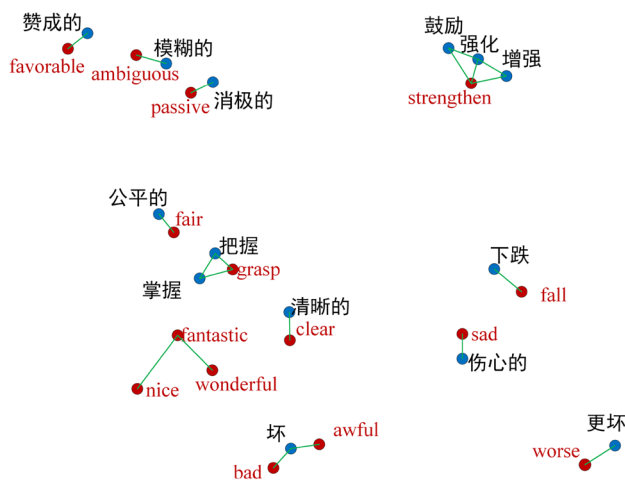


Fig. 5 Mapping of English and Chinese words to the same feature space

forward, trying to reduce the dependence on supervised information by using smaller corpora or seed dictionary (e.g., only 25 word pairs). The Semi-Supervised Approach has achieved good results on some language pairs. For example, 37.27% translation accuracy was achieved in English-French bilingual dictionary task, and nearly 40% in English-German bilingual dictionary task [37]. In recent years, Fully Unsupervised Approach has become popular in CLWE modeling [38]. The main reason is that Unsupervised Approach does not need any parallel corpora or seed dictionary and is applicable to a wider range of languages and has stronger portability.

Table 2 shows the representative researches of CLSA based on CLWE. The Supervised Approach, Semi-Supervised Approach, and Unsupervised Approach are discussed in the following sections, respectively.

3.1 Supervised Cross-lingual Word Embedding Model

Supervised Cross-Lingual Word Embedding (CLWE) model relies on a large number of bilingual parallel texts. The existing researches for CLWE have proposed methods of building word pairs with the aid of Machine Translation systems [39], modeling linguistic differences of sentiment expression [40], adding sentiment information into cross-lingual word embeddings [41], fine-grained aspect-level cross-lingual word embeddings [42], and studying the influence of word order on cross-lingual word embedding generation [44], transferring lexical information by translation dictionaries derived from parallel corpora [43], etc.

Abdalla et al. (2017) attempted to find whether the sentiment information could be highly reserved when the supervised information reduced. Therefore they conducted four cross-lingual experiments including lexicon induction, binary word sentiment classification, fine-grained sentiment analysis, and sentiment classification of reviews [39]. The results show that when the supervised information reduces, to some extent, the sentiment information is still highly preserved, which does not affect the quality of word vector transformation matrix and the performance of word vector when it is applied to CLSA tasks.

Inspired by the promising results of encoding sentiment information into word embedding representation in monolingual sentiment analysis [55], Chen et al. (2017) [40] and Dong (2018) [41] proposed to incorporate sentiment information into semantic word embedding vectors. In specific, Chen et al. believed that Language Discrepancy in the existing CLSA was mostly ignored and proposed to use Intrinsic Bilingual Polarity Correlations (IBPCs) to model language differences inherent in sentiment expression. A Relation-based Bilingual Sentiment Transfer (RBST) model was proposed to project the documents of source language and

Table 2 Representative CLSA Researches based on Cross-lingual Word Embedding

Author	Model	Characteristics	Data sets	Languages	Accuracy/%
Faruqi et al. (2014) [34]	Model based on CCA	Use LSA model to incorporate multilingual evidence found in language translations into the generation of CLWE	WMT-2011 and WMT-2012 monolingual news corpora; Europarl corpus; CEDC from WMT-2006-10 SemEval-2012 Task 4; OntoNotes (v4.0)	de-en fr-en es-en en-ch	81.90 81.30 78.20 -
Zou et al. (2013) [35]	Unsupervised neural model	Improve the objective function to embody both monolingual semantics and bilingual translation equivalence to generate CLWE	Wikipedia data from Vulić and Moens; standard 1000 ground truth one-to-one translation pairs from Vulić and Moens	es-en it-en nl-en	68.90 68.30 36.30
Vulić et al. (2015) [36]	BWESG	Use document-aligned corpus and merge the aligned documents to pseudo-bilingual documents to train the model and generate CLWE	Wikipedia corpora; Reuters corpora; News Commentary 2015 corpora	en-it en-de	- -
Barone et al. (2016) [38]	AAE	Use AAE to map source language word embeddings to target language word vector space for the first time	Google News dataset; Spanish Billion Word Corpus; Wikimedia; Google's Trillion Word Corpus; Hu & Liu list; ANEW, SAM; dataset from Lin et al.	en-ch en-es	F: 77.00 F: 81.00
Abdalla et al. (2017) [39]	SVM; LR classifier	Use word pairs obtained by machine translation to calculate the vector space translation matrix	Amazon product reviews; 5 millions Chinese posts from Weibo	en-ch	81.50
Chen et al. (2017) [40]	RBST	Model the language discrepancy as a fixed transfer vector between source language and target language in each particular polarity	SST; TA; AC; SE16-T5; AFF	en-es/ en-nl en-ru/ en-de en-cs/ en-it en-fr/ en-ja	85.93/ 79.30 93.26/ 92.31 93.69/ 96.48 92.97/ 88.08
Dong et al. (2018) [41]	DC-CNN	Encode latent sentiment information into CLWE based on annotated bilingual parallel corpus	ABSA dataset in Hindi; English dataset of SemEval-2014	en-hi	Multilingual: 76.29 Cross-lingual: 60.39
Akhtar et al. (2018) [42]	Bilingual-SGNS	Use SGNS to construct word embedding vector representations of two languages and map them to the same space for fine-grained aspect level sentiment analysis	Google universal treebank; Wikipedia data; Bible data; Europarl	MS-en / MS-de MS-es / MS-fr MS-it / MS-pt MS-sv	76.50 / 72.80 82.10 / 82.20 83.30 / 84.70 80.30
Rasooli et al. (2017) [43]	Density-driven projection	Derive cross-lingual clusters and transfer lexical information by translation dictionaries derived from parallel corpora	OpeNER copora; Catalan MultiBooked dataset	en-es en-ca	Bi: F=65.10 4-C: F=35.80 Bi: F=65.60 4-C: F=38.10
Atrio et al. (2019) [44]	SVM; SNN; BiLSTM	Adjust the word order of the target language to improve the performance of short text sentiment analysis	TiGer corpus; AMT	es-en de-en	47.00 48.00
Peirsman et al. (2010) [45]	Cross-lingual model of selectional preferences	Use a small sample seed lexicon of bilingual cognates as the initial solution of the projection matrix	Google's universal tagset; Princeton WordNet and DanNet; Wikipedia; Europarl corpus	en-es / en-de en-dk / en-se en-it / en-nl en-pt	80.60 / 80.40 63.00 / 71.60 80.10/ 74.50 76.90
Gouws et al. (2015) [46]	BARISTA	Learn task-specific bilingual embeddings based on existing monolingual word embedding algorithm and does not require parallel data.			

Table 2 (continued)

Author	Model	Characteristics	Data sets	Languages	Accuracy/%
Vulić et al. (2013) [47]	MuPTM	Use multilingual probabilistic topic modeling to generate a one-to-one seed lexicon of the one-to-many mapping relationship between words	Wikipedia articles	es-en it-en	89.10 88.20
Artetxe et al. (2017) [37]	Self-learning framework	Construct a seed lexicon based on the similarity between word embeddings of two languages	Public English-Italian dataset; ukWac+Wikipedi+BNC; itWAC; Europarl; OPUS; SdeWac; 2.8 billion word Common Crawl corpus; RG-65 and Word-Sim-353 cross-lingual dataset	en-it en-de en-fi	37.27 39.60 28.16
Chen et al. (2018) [48]	Ermes	Use Emoji as supplementary sentiment supervision information to obtain the sentiment-aware representations of words	Amazon review dataset; Tweets	en-ja en-fr en-de	80.17 86.50 86.60
Barnes et al. (2018) [49]	BLSE	Use a small bilingual dictionary and annotated sentiment data from source language to obtain a bilingual translation matrix with sentiment information	OpeNER; MultiBooked dataset	en-es en-ca en-eu	Bi: F=80.30 4-C: F=50.30 Bi: F=85.00 4-C: F=53.90 Bi: F=73.50 4-C: F=50.50
Gouws et al. (2014) [50]	BiBOWA	Use bilingual raw data to generate CLWE without supervision based on an optimized word similarity matrix	Reuters RCV1/RCV2 multilingual corpora; EuroParl	en-de de-en	86.50 75.00
Shen et al. (2020) [51]	TL-AAE-BiGRU	Use AAE to learn bilingual parallel text and map bilinguals to the same vector space by linear translation matrix	Product comments on Amazon	en-ch en-de	F=78.57
Artetxe et al. (2018) [52]	Vecmap	Use unsupervised model Vecmap to construct the initial solution, removing the dependence on small seed dictionaries	Public English-Italian dataset ; Europarl; OPUS	en-it en-de en-fi en-es	48.13 48.19 32.63 37.33
Rasooli et al. (2018) [53]	NBLR+POSwemb; LSTM	Use multiple source languages to bridge the language gap, use annotation projection and direct transfer to learn a more robust sentiment analysis model	Tweets; SentiPer; SemEval 2017 Task 4; BQ; EuroParl; LDC; GIZA++; Wikipedia articles	-	-
Ma et al. (2020) [54]	NBLR+POSwemb; LSTM	Use GAN to pre-train word embeddings first and then fine-tune	OpenNER; NLP&CC 2013 CLSC	en-es en-ca en-eu	76.10 75.60 72.10

F refers to F1 score; Bi refers to binary; 4-C refers to 4-class; MS refers to the selected multiple source languages

its translation into the common hybrid sentiment space. On the basis of Abdalla [39], Dong (2018) encoded the latent sentiment information into vectors by leveraging annotated bilingual parallel corpora. Then sentiment embeddings and semantic word embeddings are merged by using a dual-channel Convolutional Neural Network (DC-CNN) [41]. DC-CNN sets a separate sentiment channel to input sentiment information instead of simply concatenating the word embeddings and sentiment embeddings, because a simple concatenation will diminish the model's ability to exploit the intrinsic semantic relatedness information among words, especially when the dimension of sentiment embedding increases.

Most existing CLSA models only cover rough sentiment analysis, such as sentence-level or document-level sentiment analysis. Therefore, Akhtar et al. (2018) focused on a more fine-grained aspect-level of entity sentiment analysis [42]. Akhtar used Bilingual Skip-Gram with Negative Sampling (Bilingual-SGNS) to generate word embeddings for the two languages and then translated the words that could not generate word embeddings in Hindi into English. The accuracy of the model reached 76% in the multi-language sentiment analysis task at entity level, as well as more than 60% in the entity-level cross-language sentiment analysis task.

Atrio (2019) noticed the difference in word order between languages and studied the influence of word order on CLSA [44]. He used bilingual parallel corpora and adjusted the word order of the target language, including noun-adjective adjustment and reordered adjustment.

Rasooli et al. (2017) [43] attempted to optimize their previous density-driven annotation projection method (2015) [56]. Firstly, they used monolingual source language treebanks and translation dictionaries derived from parallel corpora to derive cross-lingual clusters, where words with a similar syntactic or semantic role were grouped in the same cluster, and which could be used as features in the parser. Then the translation dictionary was also used to add the translated form to the underlying sentences in the source treebank data, so that the parser trained on the source language treebanks could integrate the lexical features from the target language.

It can be seen that due to the supervised information, supervised CLWE models can guarantee the quality of the generated word embeddings and perform better in CLSA tasks. However, supervised information, such as bilingual parallel corpora and annotated data, is scarce in many non-English languages.

3.2 Semi-supervised Cross-lingual Word Embedding Model

CLWE based on semi-supervised approach mainly uses an heuristic bilingual seed dictionary of small

samples to generate cross-lingual word embeddings. Let $D = \{x_i, y_i\}_{i \in \{1, n\}}$ be the bilingual seed dictionary, where x_i is the embedding vector of each word in source language, y_i is the embedding vector of the corresponding translated word in target language, and n is the number of words in D . In semi-supervised approach, we need to find a projection matrix W , which can transfer word embeddings of different languages into a shared space. The optimal projection matrix is obtained by minimizing the mean squared error as:

$$\text{MSE} = \frac{1}{n} \sum_i (Wx_i - y_i)^2 \quad (1)$$

Given an initial projection matrix of W , iterative steps will be carried out to find the optimal projection matrix W^* to minimize objective function (1). After gaining W^* , every word in source language can be mapped to the vector space of target language by calculating the dot product of its word embedding and W^* .

The initial solution is very important in constructing bilingual word vector space. Instead of using bilingual parallel corpora or large-sample bilingual dictionary to induce the initial solution, Peirsman et al. (2010) [45] proposed to use bilingual cognates to form a small-sample seed dictionary. Therefore, the initial solution can be induced with this seed dictionary.

Since weakly supervised embedding algorithms have made huge improvements for tasks like sentiment analysis [55], Gouws et al. (2015) [46] proposed a simple wrapper method based on the existing monolingual word embedding algorithm to learn task-specific bilingual embeddings and applied it to cross-lingual tasks e.g., cross-language POS tagging or CLSA. Instead of relying on parallel data, they only assumed small dictionary seeds to produce mixed context-target pairs that were used to train monolingual embedding models.

Vulic et al. (2013) made two hypotheses [47]: one was that words in two languages were mapped one-to-one; the other was that there were words in two languages mapped one-to-many. Based on the first hypothesis, Vulic directly constructed a seed dictionary of one-to-one mapping as the initial solution. Based on the second hypothesis, Vulic used Multilingual Probabilistic Topic Modeling to generate a one-to-one mapping sub-dictionary, and reserved only the symmetric translation pairs as initial solutions for the generation of bilingual word vector space.

Artetxe et al. (2017) used a different way to construct the seed dictionary [37]. He did it based on the similarity between monolingual word vectors of two languages. He took the two words closest to each other as corresponding translations and added them to the seed dictionary.

Chen et al. (2018) proposed a representation learning framework, named Ermes, to classify cross-lingual sentiment polarity of texts based on expressions. Emoji information was used as a new bridge for text sentiment analysis and was encoded into the generated word embeddings [48].

Similar to works [40] [41], Barnes proposed a semi-supervised method to incorporate sentiment information into word embedding representations [49]. A Bilingual Sentiment Embedding (BLSE) was proposed to learn the projection matrices of source language and target language, which were jointly optimized to represent both semantic information and sentiment information.

Semi-supervised CLWE approaches suppose that word embeddings of the same meaning in different languages possess isometry. Based on this hypothesis, a large number of parallel corpora are discarded and an heuristic seed dictionary is used to generate cross-lingual word embedding vectors. However, this assumption is not valid in some cases, especially in semantically distant languages, such as English and Japanese, where the assumption is weak. As semi-supervised approaches essentially leverage the aligned seed dictionary to learn the mapping matrix of the whole space, they ignore the information contained in the word embeddings. Besides, it will bring large errors if the mapping matrix learned from the seed dictionary is taken as the mapping matrix of the whole space, especially for the language pairs with a long semantic distance.

3.3 Unsupervised Cross-lingual Word Embedding

CLWE based on unsupervised approach was formally proposed in 2014 [50] and has gradually become a mainstream in the field of cross-lingual sentiment analysis. It is capable of mining the relationship between the two languages from large-scale non-parallel corpus resources with the help of generative adversarial learning models, such as the Generation Adversarial Networks (GAN) model, the Auto-Encoder-Decoder model and other models. By using these models, the transformation matrix between the two languages is learned, so as to map the word representations of the two languages into the same space, as shown in Fig. 6.

As the first unsupervised CLWE approach, Gouws et al. (2014) [50] proposed an optimized word similarity matrix calculation method to generate cross-lingual word embedding vectors by using raw bilingual data. This method achieved 85% and 75% accuracy in English-German and German-English cross-lingual text classification tasks, and 39% and 44% accuracy in English-Spanish translation tasks, far higher than other models [51].

Barone (2016) first attempted to map source language word embedding vectors to target language word embedding vector space using Adversarial Auto-Encoder (AAE) [38].

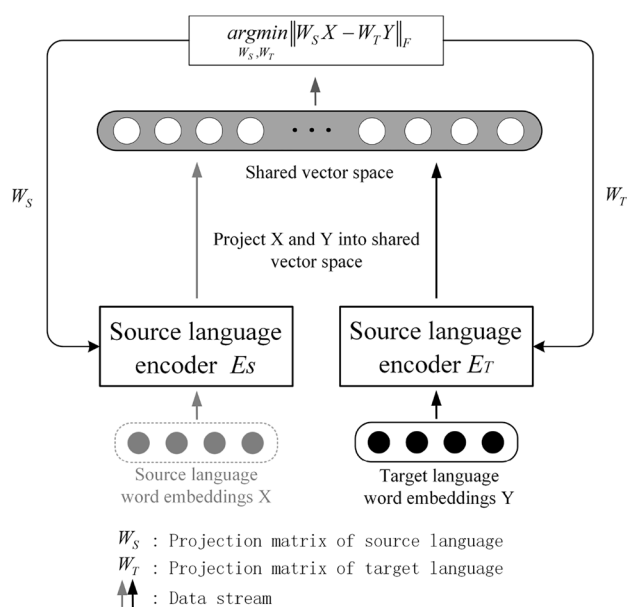


Fig. 6 Structure of cross-lingual word embedding based on unsupervised approach

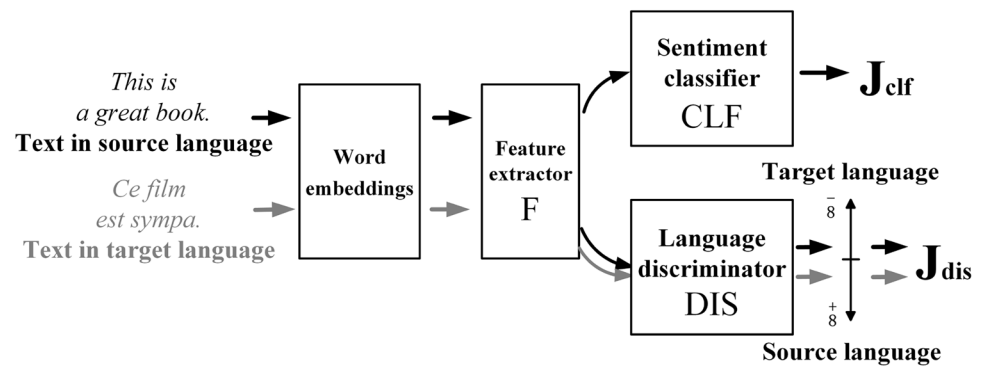
However, when training without parallel texts, the results were not so good. Shen (2020) used AAE to learn bilingual parallel texts and mapped the two languages to the same shared vector space through a linear transformation matrix, which was used as the input of BiGRU model to obtain the final sentiment prediction results [51].

Artetxe et al. (2018) proposed a method different from the above researches based on the semi-supervised method with 25 word pairs, replacing the AAE Model with a Self-Learning Model [52]. He constructed an initial solution from a bilingual non-parallel corpus, and then initiated a self-learning model to generate cross-lingual word vectors. This method achieves 48% accuracy in English-Italian and English-German bilingual dictionary generation tasks, and 37% accuracy in English-Spanish bilingual dictionary generation tasks [52].

Rasooli et al. (2018) [53] considered the influence of language families on CLWE and introduced a multi-source approach to bridge the gap between source language and target language. Annotation Projection and Direct Transfer were used to create a robust sentiment analysis system for languages with minimal machine translation capability and without annotated data.

Motivated by integrating more sentiment information into bilingual word embeddings, Ma et al. (2020) proposed an Unsupervised Bilingual Sentiment word Embedding (UBSE) [54]. UBSE used Generative Adversarial Nets to pre-train word embeddings first and then fine-tune the embedding projection matrix according to a sentiment lexicon of source language. Compared with the existing unsupervised CLWE

Fig. 7 Structure of cross-lingual sentiment analysis based on GAN



methods without sentiment information (e.g., MUSE), UBSE shows comparable results, a better performance on F1 metric but a worse performance on Precision and Recall metrics.

Compared with supervised and semi-supervised methods, unsupervised methods abandon bilingual supervised information (bilingual parallel text or corpora), thus reduce the time cost and manpower cost of data preprocessing. At the same time, most of the target languages in CLSA are resource scarce languages, which further reflects the importance of unsupervised methods. At present, most of the unsupervised CLWE approaches have achieved excellent performance. For example, BiBOWA model has high accuracy in English-German and English-Spanish cross-lingual sentiment analysis tasks, with the former reaching more than 85% [38]. TL-AE-BiGRU model achieves more than 78% F1 metric in English-Chinese and English-German CLSA tasks [52].

Although unsupervised CLWE approaches have no longer heavily relied on bilingual parallel corpora, challenges still exist. Sogaard et al. found that unsupervised cross-lingual word embedding method was exceedingly sensitive to the choice of source and target languages (language pairs) [46]. It will be difficult to gain high-quality CLWE without any bilingual supervised information. Besides, based on the assumption that the embedding vectors of corresponding words in different languages are similar, the unsupervised method only uses monolingual word embedding vectors to generate cross-lingual word embeddings. However, the above hypothesis is not always valid when the two languages have large semantic and structural differences, such as English-Japanese and Spanish-Chinese. Therefore, in the absence of supervised information, unsupervised CLWE methods tend to fall into the local optimal solution, or even the worse solution instead of the global optimal solution, thus affecting the quality of the generated cross-lingual word embeddings.

4 Cross-lingual Sentiment Analysis Combined with CLWE

Generation Adversarial Network (GAN) was first proposed by Goodfellow et al. in 2014 [57], which has achieved great success in image generation task and has been successfully applied in natural language processing, especially in domain transfer and language transfer. In recent years, it has been applied in the field of CLSA. Cross-lingual sentiment analysis based on GAN doesn't rely on annotated data of target languages, but transfers knowledge learned from annotated data of resource-rich source languages to low-resource languages with unannotated data through GAN.

4.1 CLSA based on Generation Adversarial Network

As shown in Fig. 7, the core idea of cross-lingual sentiment analysis based on GAN is adversarial training. Feature extractor is used as a generator to extract features from texts, and language discriminator is used as discriminator to identify the language to which the features belong. In each iteration, the discriminator first promotes the language recognition ability, and then the extractor tries to confuse the discriminator. Finally, the feature extractor makes the discriminator completely unable to identify language, so that it can extract language-independent features. If so, the structure can be applied to the target language sentiment classification, and then the sentiment classifier trained with corpora of source language will take these features as input to predict the sentiment polarity of target language. Table 3 shows the representative researches of CLSA based on GAN.

Chen et al. (2018) first proposed a Adversarial Deep Averaging Network (ADAN) model [58] that can automatically learn and extract language-independent features from annotated data of source languages through multiple iterations of Feature Extractor and Language Discriminator. During the adversarial learning, ADAN tries to minimize Wasserstein distance between source language and target language to ensure high-quality language-independent feature extraction.

Table 3 Representative CLSA Researches based on Generation Adversarial Network

Author	Model	Characteristics	Data sets	Languages	Accuracy /%
Chen et al. (2018) [58]	Adversarial Deep Averaging Network (ADAN)	A language-adversarial training approach has feature extractor, language discriminator and sentiment classifier	Yelp reviews; Chinese hotel reviews; BBN Arabic Sentiment Analysis data set;	ch-en en-ar	42.49 54.54
Antony et al. (2020) [59]	Language Invariant Sentiment Analyzer (LISA)	Adversarial training on multiple resources-rich languages to infer sentiment information on a low resource language	Amazon Review Dataset; Sentiraama Dataset	en-de en-fr en-jp	85.63 85.73 81.13
Wang et al. (2018) [60]	Attention-based CNN model with adversarial cross-lingual learning	Extract language-independent features to alleviate the problem of insufficiency of personal data on Sina Weibo and Twitter datasets	Twitter; Weibo	en-ch (Twitter) en-ch (Weibo)	F: 81.24 F: 82.03
Kandula et al. (2021) [61]	Conditional Language Adversarial Network (CLAN)	Optimize model by the conditional adversarial training of the language model and discriminator	Websis-CLS-10 Dataset	en-de en-fr en-jp	86.30 88.30 83.70
Pelicon et al. (2020) [62]	Model based on Multi-BERT	A multilingual model to overcome BERT's inefficiency in long documents	SentiNews dataset; Croatian dataset	Slovenian-Croatian	F: 54.77

F refers to F1 score

Inspired by ADAN [58], Antony (2020) proposed the framework of Language Invariant Sentiment Analyzer (LISA) trained with monolingual data sets of multiple resource-rich languages [59]. The architecture first uses the unsupervised method in MUSE (Multilingual Unsupervised and Supervised Embeddings) [63] to align the spaces of other languages to the semantic space of English, so as to construct multi-language word embeddings. The results show that although LISA model is not suitable for zero-sample learning, it can achieve optimal performance with limited data.

Even though the work of Chen et al. [58] does not require the annotated data of the target language, it still relies heavily on CLWE. Studies have found that the quality of CLWE is very sensitive to different language pairs, and there are studies on how to obtain CLWE through unsupervised methods [59, 64]. For example, the generated CLWE between English and Japanese is not ideal, and additional cross-lingual supervision information may be required, such as adding seed words between language pairs to provide more supervision information. Feng (2019) improved the unsupervised generation of CLWE by mining large amount of unlabeled data of target language with the help of Auto Encoder-Decoder model.

In order to solve the problem that most of existing works on microblog sentiment classification suffer from the insufficiency of personal data in monolingual data set, Wang (2018) proposed a Personalized Microblog Sentiment Classification model via Adversarial Cross-lingual Multi-task Learning to exploit users' posts on different microblogging platforms (e.g., Sina Weibo and Twitter) in different languages [60]. The proposed model is composed of three components: a language discriminator, two feature generators to extract language-specific features and language-independent features to enrich the users' posts representations, and two sentiment classifiers in individual languages.

Kandula et al. derived inspiration from DANN (Ganin et al., 2016) [65] and CDAN (Long et al., 2018) [66] and proposed an end-to-end neural network, named Conditional Language Adversarial Network (CLAN), for cross-lingual sentiment analysis without cross-lingual supervision information [61]. Unlike the prior work, the adversarial training in CLAN is conditioned on both the extracted features and the sentiment prediction performance, so as to increase the discriminativity for learned representation. Pelicon et al. (2020) trained a news sentiment classification model based on multilingual BERT using Slovenian data sets [62]. In this model, an intermediate processing step, which jointly trains the model with the masked language modeling task

and sentiment classification task, is added before fine-tuning. At the same time, the model tests the method of generating document representation with the beginning, the end and full text respectively, in order to overcome the defect that BERT model cannot effectively deal with long documents.

Cross-lingual sentiment analysis based on GAN skillfully implements cross-lingual sentiment knowledge transfer with the help of generation adversarial network and achieve an accuracy of more than 83% in English-German and English-French language pairs. However, when applied to different languages, the performance of this method varies greatly and the parameters need to be adjusted, resulting in insufficient support for generalization of languages.

4.2 CLSA based on Pre-Trained Model

Pre-Trained Model (PTM) is a new paradigm in Natural Language Processing (NLP) that is being developed quite rapidly. Since it has been proposed successively and applied to the field of CLSA represented by ELMo (2018) [67], BERT (2019) [68], and GPT-3 (2020) [69], related researchers attempt to build a model which is proficient in any languages based on PTM. PTM is essentially a form of transfer learning, and the training process can be divided into two steps: pre-training and fine-tuning. The pre-training stage uses self-supervised knowledge to learn the initial model (hundreds of millions of parameters) irrelevant to specific tasks from a large corpus; and the fine-tuning stage fits the model for the downstream tasks (target tasks).

The advantages of PTM in CLSA can be summarized into three aspects [70]: (1) Pre-training on the huge text corpus can learn universal language representations and help with the downstream tasks. (2) Pre-training provides a better model initialization, which usually leads to a better generalization performance and speeds up convergence on the target task. (3) Pre-training can be regarded as a kind of regularization to avoid overfitting on small data. Table 4 summarizes the representative researches of CLSA based on PTM since 2019, including Multilingual BERT [71], XLM [72], XLM-RoBERTa [73], MetaXL [74] and analyzes their application on CLSA, pros and cons as well as the experiment datasets.

Multilingual BERT (Multi-BERT) proposed by Devlin et al. in 2018 [68], consisting of 12 Transformer layers, is trained with the monolingual Wikipedia text corpora in 104 languages. During the training step, Multi-BERT doesn't use any annotated data nor language representations computed by MT. It is pre-trained through Masked Language Modeling. Pires et al. (2019) [71] found Multi-BERT performed excellently in zero-sample cross-lingual tasks, especially when the source and target languages were similar to each other. However, systematic deficiencies exist when

Multi-BERT is applied to the multilingual representation tasks of some language pairs.

In order to improve cross-lingual text representations, Lample et al. (2019) [72] proposed three methods to learn Cross-lingual Language Models (XLMs), including two unsupervised models, Causal Language Modeling (CLM) and Masked Language Modeling (MLM), and one supervised model, Translation Language Modeling (TLM). CLM and MLM were proposed for monolingual tasks, while TLM was proposed for cross-lingual tasks. TLM leverages parallel data to improve cross-lingual pre-training instead of relying on monolingual text streams. During the training step of TLM, several words in the source sentence and the target sentence will be masked randomly, and TLM utilizes the context of target words or the translation of the corresponding source sentences to predict words masked in target sentences, so as to guide the model to align representations of two languages. Experiments showed that the supervised TLM outperformed the previous state-of-the-art on cross-lingual natural language inference (XNLI) by 4.9% accuracy on average.

On the basis of XLM [72], Conneau et al. (2020) [73] proposed a transformer-based multilingual masked language model XLM-Roberta in the following year and demonstrated that pre-training multilingual language models led to significant performance gained for a wide range of cross-lingual transfer tasks. Compared with XLM and Multi-Bert, XLM-Roberta improves mainly three aspects: (1) Enlarge the scales of languages and training datasets to 2.5TB in 100 languages; (2) Use multilingual annotated data to improve the performance; (3) Adjust the parameters to offset the inevitable problem that when cross-lingual transfer is used to fit the model for more languages, the model's understanding ability of each language may be limited. XLM-Roberta was pre-trained in 100 languages and obtained state-of-the-art performance on cross-lingual classification, sequence labeling and question answering, especially in low-resource languages. However, there may be a large number of code-mixed words in the model, which makes the system unable to comprehend the inherent meaning of sentences [77].

CLSA based on PTM requires sufficient labeled data for fine-tuning, which leads it to a bad performance for low-resource languages. Besides, the representation gap between languages will make the transfer difficult. To solve this problem, Xia et al. (2021) [74] proposed MetaXL model based on Meta-learning framework to bridge the representation gap between languages, made the representation space of the source language and the target language closer and improved the performance of cross-language transfer learning. Experiments show that, compared with Multi-BERT and XLM-RoBERTa, MetaXL in both Cross-Language Sentiment Analysis and Named Entity Recognition improved by 2.1% on average. In future work, we can study the effect of

Table 4 Representative Researches on Cross-lingual Pre-Trained Model

Reference	Model	Task	Advantage	Disadvantage	Data set /%
Pires et al. (2019) [71]	Multi-lingual BERT	Zero-shot cross-lingual model transfer	Performs excellently in zero-sample cross-lingual tasks, especially when source and target are similar	Systematic deficiencies exist when applied to multilingual representation tasks	Code-switching Hindi, English Universal Dependencies corpus
Lample et al. (2019) [72]	XLM	cross-lingual text representations	leverages parallel data to guide the model to align representations of two languages and improve cross-lingual pre-training	Scale of training data is small, especially for languages in low resources	MultiUN, IIT Bombay corpus, EBookshop corpus
Conneau et al. (2020) [73]	XLM-RoBERTa	cross-lingual classification, sequence labeling and question answering	Perform well at large scale for a wide range of cross-lingual transfer tasks and low resource languages	Cannot apprehend the inherent meaning of sentences with a large number of code-mixed words	Common Crawl Corpus in 100 languages, Wikipedia Corpus
Xia et al. (2021) [74]	MetaXL	Multilingual transfer for CLSA	Make the target language and the source language closer in the representation space, with good transmission performance	Have not yet explored placing multiple conversion networks on multiple layers of the PTM	Amazon product reviews, Sentipers, Sentiraama
Bataa et al. (2019) [75]	ELMo ULMFIT BERT	Japanese sentiment classification	Use TL technology and pre-trained models to solve Japanese sentiment classification	Did not perform K-fold validation	Japanese Rakuten review binary, five class Yahoo datasets
Gupta et al. (2021) [76]	BERT Multi-BERT.etc	Task-specific pretraining and cross-lingual transfer for sentiment analysis	Strong pertinence good performance, can be used as a baseline for future sentiment analysis work	Poor cross-lingual transmission	Tamil-English, Malayalam English, Sentimix Hinglish

placing numerous networks on various layers and improve the transfer language performance.

Bataa and Wu (2019) [78] focused on English-Japanese CLSA tasks and conducted their experiments on ELMo [67], ULMFiT [79], BERT [68], respectively. Experiments show that PTMs perform better than task-specific models represented by RNN, LSTM, KimCNN, Self-Attention and RCNN. For the task of Multi-language Dialogue System Identification, Gupta et al. (2021) [76] compared the effect of code switching in two language pairs (Tamil-English and Malayalam-English) based on four PTMs: BERT [68], Multi-BERT [71], XLM-RoBERTa [73], and TweetEval [80]. TweetEval was proposed by Barbieri et al. and the main idea is to solve seven classification tasks of Tweet media data based on XLM-Roberta, e.g., sentiment analysis, emotion recognition. Experiments show that TweetEval outperforms BERT [68], Multi-BERT [71], XLM-RoBERTa [73].

PTMs, such as Multi-BERT, XLM and MetaXL, have been widely used in CLSA tasks and gained remarkable performance. However, there are still several problems to be solved. Firstly, the steps of pre-training, fine-tuning and inference are all costly because of the enormous number of parameters they have [75], e.g., the number of parameters of GPT-3 reaches 175 billion, and that of Gopher also reaches 280 billion, which makes PTMs difficult to fit to the online services and the projects on resource-restricted devices [70]. To solve this problem, a future direction is to design more effective model structures, self-supervised pre-trained tasks, optimizers and training skills under the existing software and hardware conditions for improvement. Secondly, the best CLSA result based on PTM is Multi-BERT which achieves an accuracy of 90.0 % on English-German language pairs in MLDoc (Schwenk and Li, 2018 [81]), while the worst result is on English-Chinese language pairs, with an accuracy of only 43.88 % [82], which shows the great differences of PTMs applied to different language pairs. Although the PTM can learn language-independent features from large-scale data and perform well in zero sample CLSA tasks, especially in those among closer languages, it still cannot be used as a generalization model for different language pairs. Since each language pair has its own fine-tuning parameters, applying PTM to different languages pairs requires fine-tuning according to language transfer which is very inefficient. One of the solutions is to fix the original parameters and add a small adaptive module for specific tasks [83]. More efficient methods are worth exploring in the future work. Generally speaking, PTM has achieved remarkable performance in CLSA tasks since 2019, but it is still a brand-new technique, and a synthesis of the recent advances has not yet been done. The application of PTM to CLSA deserves researchers' more attention.

5 Conclusion and Future Work

In this paper, we review the presentative researches of CLSA and systematically expound its development process. The researches can be divided into two periods according to the context: the early phase of CLSA and the modern CLSA based on Word Embedding, both of which can be further divided into sub-methods. Though CLSA has developed rapidly, there are still some problems to be solved: (1) Is there a generalization CLSA model which is adaptive to all the target languages? (2) Is it possible to find the best range of source languages for a given target language? (3) What is the relation between all of the CLSA methods? In view of these three problems, we make the following analysis and outlook based on the existing researches on CLSA.

Q1: Amounts of methods have been proposed but the state-of-the-art researches still cannot find a generalization model which performs well in every situation. For instance, ADAN outperforms other models in the English-French CLSA task, but the results of English-Japanese CLSA task aren't satisfactory [58]; 110 CLSA tasks, including 45 languages, can be solved by MUSE model, but its performance varies greatly among different languages [63]. The discrepancy between languages is the main challenge. Most researches set a fixed language, such as English, as the source language, which aggravates the imbalance of the discrepancy between languages.

To avoid it, some researchers set multiple source languages simultaneously to bridge the gap and obtain better performance. MAD-X model uses adapters to adjust the parameters of CLSA models for a specific target language. It shows that the CLSA model with the adaptation of MAD-X outperforms other models [84]. This research also shows that the discrepancy will be minimized if two languages belong to the same language family, which can significantly improve the model performance.

Since Pre-Trained Models (PTM) obtain remarkable achievements in the field of natural language processing, such as BERT, GPT-2, GPT-3, some researchers attempt to build a generalization model based on these PTM models. Multi-BERT is the representative PTM model applied on CLSA tasks, which has been trained with monolingual Wikipedia text corpora in 104 languages [85]. Multi-BERT performs excellently in zero-sample CLSA tasks, especially when the source and target languages are similar to each other [86]. However, systematic deficiencies exist when Multi-BERT is applied to the multilingual representation tasks of some language pairs. Besides, its training needs a large-scale data and PTM has numerous parameters, which makes PTMs training costly and time-consuming [87].

Q2: Most of the existing researches choose English as the source language mainly because: 1) English has rich

resources, which makes it suitable for Transfer Learning; 2) there are lots of available models for sentiment analysis for English. However, once the source language is fixed to the very language, it must cause the imbalance of language gap between the source and target languages and further affect the performance of CLSA models. In recent years, some researchers expand the choices of source language, such as Japanese and German [88]. Rasooli hypothesized that taking the languages belonging to the same language family as the source and target languages respectively could improve the performance of CLSA, and he conducted experiments on Multi-Source languages to confirm his prediction [53]. Besides, Farra (2019) explored the choice of source languages in CLSA and tried to select the best source language for 15 languages. They came to the following conclusions: Firstly, languages from similar language families can transfer sentiment information well from each other; secondly, languages with large amount of available parallel resources and evenly distributed sentiment-annotated datasets are generally good choice of source languages; thirdly, languages with similar morphological complexity and vocabulary sizes can transfer sentiment information well from each other [89].

Therefore, the accuracy and credibility of CLSA will significantly increase if it is possible to find the best range of source languages for a given target language. However, due to the limitation of lacking models and available data, the number of languages that can be easily leveraged as source languages without much preprocessing is relatively small, such as Hindi and Slovak [37, 90, 91].

Q3: On one hand, CLSA based on PTMs, such as CLSA based on Multi-BERT [71], becomes the most popular and dominant research direction nowadays, which contributes to expand the choice of target language to Chinese, Hindi, Malay and those with few available resources. Nonetheless, CLSA based on PTMs has strict requirements on computation ability and it needs fine-tuning steps to fit the model for a specific CLSA task, which may affect its performance on different language pairs and limit the large-scale application of CLSA based on PTMs.

On the other hand, those classic methods of CLSA will still influence the future development of CLSA, though a few years have passed since they were proposed, e.g., CLSA based on PTMs requires a large amount of data to pre-train the model which could leverage the MT method to obtain sufficient data sets, even though the MT method was proposed in 2004 [1]; unsupervised CLWE methods can utilize the main idea of SCL in the step to obtain the initial solution [22] [37]. Hence, it must be the future development tendency that CLSA will be mainly based on PTMs combined with different methods to solve the problems of source language selection and different performance on different language pairs.

Finally, it should be noted that the ultimate goal of CLSA is to help target languages achieve sentiment analysis with the help of source languages. Due to the large amounts of languages with scarce resources, CLSA is a direction worth exploring. If the cost of knowledge transfer required by the CLSA model is too expensive, or even far exceeds the cost of manpower and material resources required by the monolingual sentiment analysis, the original intention of CLSA will be violated. At the same time, it is also one of the important indicators to test whether the CLSA model can be applied in large-scale languages in the future.

Acknowledgements Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Author Contributions Authors contributed equally to this work.

Funding This work was supported by the Fundamental Research Funds for the Central Universities (No.2022JJ006).

Data Availability Statement Not applicable.

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Yan Q, David A, Evans JG, Gergory G (2004) Mining multi-lingual options through classification and translation. In: Proceeding of AAAI. Menlo Park, pp. 354–362. AAAI
2. Wan X (2008) Using bilingual knowledge and ensemble techniques for unsupervised chinese sentiment analysis. In: Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing, pp. 553–561. ACL
3. Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. In: Proceedings of ACM SIGIR Conference on RDIR, pp. 1301–37813. Proceedings of Workshop at ICLR
4. Balahur A, Turchi M (2014) Comparative experiments using supervised learning and machine translation for multi-lingual sentiment analysis. *Comput Speech Lang* 28(3):56–75

5. Hajmohammadi MS, Ibrahim AR, Selamat (2015) Graph-based semi-supervised learning for cross-lingual sentiment classification. In: Proceedings of ACIIDS, pp. 97–106. Springer
6. Banea C, Mihalcea R, Wiebe J (2010) Multilingual subjectivity: are more languages better? In: Proceedings of the 23rd International Conference on Computational Linguistics, pp. 28–36. ACL
7. Hajmohammadi MS, Ibrahim R, Selamat A (2014) Density based active self-training for cross-lingual sentiment classification. In: Advanced in Computer Science and Its Applications, pp. 1053–1059. Springer
8. Prettenhofer P, Stein B (2010) Cross-language text classification using structural correspondence learning. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 1118–1127. ACL
9. Hajmohammadi MS, Ibrahim R, Selamat A (2014) Bi-view semi-supervised active learning for cross-lingual sentiment classification. *Inf Process* 50:718–732
10. Pan J, Xue GR, Yu Y, Wang Y (2011) Cross-lingual sentiment classification via bi-view non-negative matrix tri-factorization. In: Proceedings of the 15th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining, pp. 289–300. Springer
11. Wan X (2011) Bilingual co-training for sentiment classification of chinese product reviews. *Comput Linguist* 37:587–616
12. He Y (2011) Latent sentiment model for weakly-supervised cross-lingual sentiment classification. In: Proceedings of ECIR, pp. 214–225. Springer
13. Zhang P, Wang SG, Li DY (2016) Cross-lingual sentiment classification: Similarity discovery plus training data adjustment. Elsevier B.V., Knowledge-Based Systems 107, 129–141
14. Al-Shabi A, Adel A, Omar N, Al-Moslmi T (2017) Cross-lingual sentiment classification from english to Arabic using machine translation. *Int J Adv Computer Sci Appl* 8(12):434–440
15. Hajmohammadi MS, Ibrahim R, Selamat A (2014) Cross-lingual sentiment classification using multiple source languages in multi-view semi-supervised learning. *Eng Appl Artif Intell* 36:195–203
16. Hajmohammadi MS, Ibrahim R, Selamat A, Fujita H (2015) Combination of active learning and self-training for cross-lingual sentiment classification with density analysis of unlabelled samples. *Inf Sci* 317:67–77
17. Lu B, Chenhao T, Claire C, Benjamin KT (2011) Joint bilingual sentiment classification with unlabeled parallel corpora. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pp. 320–330. ACL
18. Meng X, Furu W, Xiaohua L et al. (2012) Cross-lingual mixture model for sentiment classification. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers, pp. 572–581. ACL
19. Gao D, Furu W, Wenjie L et al. (2014) Cross-lingual sentiment lexicon learning with bilingual word graph label propagation. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, pp. 21–40. ACL
20. Zhou G, Tingting H, Jun Z, Wensheng W (2015) A subspace learning framework for cross-lingual sentiment classification with partial parallel data. In: Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI), pp. 1426–1432. AAAI / IJCAI
21. Yingfan G, Huilin W, Hongjiao X (2010) Research progress of cross language text classification technology. *Inf Stud: Theor Appl* 11:126–128
22. Blitzer RJ, McDonald, Pereira F (2006) Domain adaptation with structural correspondence learning. In: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, pp. 120–128. SIGDAT
23. Prettenhofer BP, Stein W (2010) Cross-lingual adaptation using structural correspondence learning. *ACM Trans Intell Syst Technol* 3(10):127–233
24. Wang P, Zhao X, Li N, Dong Y (2017) Research on cross language emotion classification based on structural correspondence learning. *J Nanjing Univ(Nat Sci)* 53(6):1133–1233
25. Baccianella S, Esuli A, Sebastiani F (2010) Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: Proceedings of the International Conference on Language Resources and Evaluation, pp. 17–23. LREC
26. Darwich M, Noah SAM, Omar N (2016) Automatically generating a sentiment lexicon for the Malay language. *Asia-Pacific J Inf Technol Multim* 5(1):49–59
27. Nasharuddin NA, Abdullah MT, Azman A, Kadir RA (2017) English and malay cross-lingual sentiment lexicon acquisition and analysis. In: Proceedings of ICISA: Information Science and Applications, pp. 467–475. Springer
28. Sazzed S (2020) Development of sentiment lexicon in bengali utilizing corpus and cross-lingual resources. In: Proceedings of 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI), pp. 237–244. IEEE
29. Vania C, Moh I, Adriani M (2014) Sentiment lexicon generation for an under-resourced language. *Int J Comput Linguistics Appl* 5(1):59–72
30. Chang CH, Wu ML, Hwang SY (2019) An approach to cross-lingual sentiment lexicon construction. In: Proceedings of 2019 IEEE International Congress on Big Data (BigDataCongress), pp. 129–131. IEEE
31. Gao D, Wei F, Li W et al (2015) Cross-lingual sentiment lexicon learning with bilingual word graph label propagation. *Comput Linguist* 41(1):21–40
32. He X, Gao S, Yu Z, Liu S, Hong X (2018) Sentiment classification method for chinese and vietnamese bilingual news sentence based on convolution neural network. In: Proceedings of International Conference on Mechatronics and Intelligent Robotics, pp. 1230–1239. Springer
33. Zabha NI, Ayop Z, Anawar S et al (2019) Developing cross-lingual sentiment analysis of Malay twitter data using lexicon-based approach. *Int J Adv Computer Sci Appl* 10(1):346–351
34. Faruqui M, Dyer C (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, pp. 462–471. ACL
35. Zou WY, Socher R, Cer D et al. (2013) Bilingual word embeddings for phrase-based machine translation. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pp. 1393–1398. ACL
36. Vulic I, Moens MF (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 (Volume 2: Short Papers), pp. 719–725. ACL
37. Artetxe M, Labaka G, Agirre E (2017) Learning bilingual word embeddings with (almost) no bilingual data. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 451–462. ACL
38. Barone AVM (2016) Towards cross-lingual distributed representations without parallel text trained with adversarial autoencoders. In: Proceedings of the 1st Workshop on Representation Learning for NLP, pp. 121–126. ACL
39. Abdalla M, Hirst G (2017) Cross-lingual sentiment analysis without (good) translation. In: Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 462–471. ACL

40. Chen Q, Li C, Li W et al. (2017) Modeling language discrepancy for cross-lingual sentiment analysis. In: Proceedings of CIKM 2017, pp. 117–126. ACM
41. Dong X, Melo GD (2018) Cross-lingual propagation for deep sentiment analysis. In: Proceedings of The Thirty-Second AAAI Conference on Artificial Intelligence(AAAI-18), pp. 5771–5778. AAAI
42. Akhtar MS, Sawant P, Sen S et al (2018) Improving word embedding coverage in less-resourced languages through multi-linguality and cross-linguality: a case study with aspect-based sentiment analysis. *ACM Trans Asian Low-Resour Language Information Process* 18(2):1–22
43. Rasooli MS, Collins M (2017) Cross-lingual syntactic transfer with limited resources. *Trans Assoc Comput Linguist* 58:279–293
44. Atrio AR, Badia T, Barnes J (2019) On the effect of word order on cross-lingual sentiment analysis. In: SEPLN 2019, pp. 5771–5778.
45. Peirsman Y, Pado S (2010) Cross-lingual induction of selectional preferences with bilingual vector spaces. In: Proceedings of Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp. 921–929 . ACL
46. Gouws S, Sigaard A (2015) Simple task-specific bilingual word embeddings. In: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 1386–1390. ACL
47. Vulic I, Moens MF (2013) A study on bootstrapping bilingual vector spaces from non-parallel data (and nothing else). In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pp. 1613–1624. ACL
48. Chen Z, Shen S, Hu Z et al. (2018) Hermes: Emoji-powered representation learning for cross-lingual sentiment classification. In: SEPLN 2018, pp. 1806–02557.
49. Barnes J, Klinger R, Schulte S, Walde S (2018) Bi-lingual sentiment embeddings: joint projection of sentiment across languages. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 2483–2493. ACL
50. Gouws S, Bengio Y, Corrado G (2014) Bilbowa: Fast bi-lingual distributed representations without word alignments. In: SEPLN 2014, pp. 1806–02557.
51. Shen J, Liao X, Lei S (2020) Cross-lingual sentiment analysis via aae and bigru. In: Proceedings of 2020 Asia-Pacific Conference on Image Processing Electronics and Computers (IPEC), pp. 237–241. IEEE
52. Artetxe M, Labaka G, Agirre E (2018) A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings. In: <http://arxiv.org/abs/1805.06297>, pp. 1806–02557.
53. Rasooli MS, Farra N, Radeva A et al (2018) Cross-lingual sentiment transfer with limited resources. *Mach Trans* 32:143–165
54. Ma C, Xu W (2020) Unsupervised bilingual sentiment word embeddings for cross-lingual sentiment classification. In: Proceedings of the 2020 the 4th International Conference on Innovation in Artificial Intelligence, pp. 180–183. ICI AI
55. Tang WFQBD, et al. (2016) Sentiment embeddings with application to sentiment analysis **28**, 496–509. IEEE
56. Rasooli MS, M., C. (2015) Parameter-efficient transfer learning for nlp. In: Density-driven Cross-lingual Transfer of Dependency Parsers, pp. 328–338. Association for Computational Linguistics
57. Goodfellow I, Pouget-Abadie J, Mirza M et al. (2014) Generative adversarial nets. In: Proceedings of the 27th International Conference on Neural Information Processing Systems, Volume 2, pp. 2672–2680. MIT
58. Chen X, Sun Y, Athiwaratkun B et al (2018) Adversarial deep averaging networks for cross-lingual sentiment classification. *Trans Assoc Comput Linguist* 6:557–570
59. Antony A, Bhattacharya A, Goud J, Mamidi R (2020) Leveraging multilingual resources for language invariant sentiment analysis. In: Proceedings of the 22nd Annual Conference of the European Association for Machine Translation, pp. 71–79 . EAMT
60. Wang W, Feng S, Gao W et al. (2018) Personalized microblog sentiment classification via adversarial cross-lingual multi-task learning. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 338–348. ACL
61. Kandula H, Min B (2021) Improving cross-lingual sentiment analysis via conditional language adversarial nets. In: Proceedings of the Third Workshop on Computational Typology and Multilingual NLP, pp. 32–37. ACL
62. Pelicon A, Pranjić M, Miljković D et al (2020) Domain-adversarial training of neural networks. *Zero-Shot Learn Cross-lingual News Sentiment Classif* 10(17):59–93
63. Conneau A, Lample G, Ranzato MA (2017) Word translation without parallel data. In: <http://arxiv.org/abs/1710.04087>, pp. 1806–02557.
64. Chen X, Awadallah AH, Hassan H et al. (2019) Zero-resource multilingual model transfer: Learning what to share. In: <http://arxiv.org/abs/1810.03552v1>, pp. 1806–02557.
65. Ganin Y, Ustinova E, Ajakan H et al (2016) Domain-adversarial training of neural networks. *J Mach Learn Res* 17(1):2030–2096
66. Long M, Cao Z, Wang J, Jordan MI (2017) Conditional adversarial domain adaptation. In: <http://arxiv.org/abs/1705.10667>, pp. 1806–02557.
67. Peters NMIM M., et al. (2018) Deep contextualized word representations. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), pp. 2227–2237 . ACL
68. Devlin CMWLKJ, Toutanova K (2019) Bert: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 4171–4186 . ACL
69. Brown MBRNTB et al. (2020) Language models are few shot learners. In: <http://arxiv.org/abs/2005.14165>.
70. Qiu STXXYG X P, et al. (2022) Pre-trained models for natural language processing: A survey. In: <http://arxiv.org/abs/2003.08271v2>.
71. Pires SET, Garrette D (2019) How multilingual is multilingual bert? In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, pp. 4996–5001. ACL
72. Lample G, Conneau A (2019) Cross-lingual language model pre-training. In: <http://arxiv.org/abs/1901.07291v1>.
73. Conneau KKGNA et al. (2020) Unsupervised cross-lingual representation learning at scale. In: <http://arxiv.org/abs/1911.02116>.
74. Xia ZGMS M et al. (2021) MetaXl: Meta representation transformation for low-resource cross-lingual learning. In: <http://arxiv.org/abs/2104.07908>.
75. Pikuliak SMM, Bielikova M (2021) Cross-lingual learning for text processing: A survey **165**
76. Gupta RSK A, Black AW (2021) Task-specific pre-training and cross lingual transfer for sentiment analysis in dravidian code-switched languages. In: Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages, pp. 328–339
77. Hossain SOE, Hoque MM (2021) Multilingual code-mixed hope speech detection using cross-lingual representation learner. In: Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion, pp. 4996–5001
78. Bataa E, Wu J (2019) An investigation of transfer learning-based sentiment analysis in japanese. In: Proceedings of the 57th Annual

- Meeting of the Association for Computational Linguistics, pp. 4652–4657. ACL
79. Howard J, Ruder S (2018) Universal language model fine-tuning for text classification. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, pp. 4652–4657. ACL
 80. Barbieri CJNLF, Espinosa L (2020) Tweeteval: Uni?ed benchmark and comparative evaluation for tweet classi?cation. In: <http://arxiv.org/abs/2010.12421>.
 81. Holger S, Xian L (2018) A corpus for multilingual document classi?cation in eight languages. In: <http://arxiv.org/abs/1805.09821>.
 82. Dong XL, Melo G (2019) A robust self-learning framework for cross-lingual text classification. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, pp. 6306–6310 . EMNLP-IJCNLP
 83. Neil AGSJ H, et al. (2019) Parameter-efficient transfer learning for nlp. In: International Conference on Machine Learning, pp. 2790–2799. ICML
 84. Pfeiffer J, Vulic I, Gurevych I, et al. (2020) Mad-x: An adapt-er-based framework for multi-task cross-lingual transfer. In: <http://arxiv.org/abs/2005.00052>, pp. 1806–02557.
 85. Jacob D, Ming-Wei C, Kenton L, Kristina T (2018) Bert: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics, pp. 4171–4186. ACL
 86. Pires T, Schlinger E, Garrette D (2019) How multilingual is multilingual bert? In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, pp. 4996–5001. ACL
 87. Qiu X, Sun T, Xu Y et al (2020) Pre-trained models for natural language processing: A survey. *Sci China Technol Sci* 63:1872–1897
 88. Duh K, Fujino A, Nagata M (2011) Is machine translation ripe for cross-lingual sentiment classification? In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pp. 429–433. ACL
 89. Farra N (2019) Cross-Lingual and Low-Resource Sentiment Analysis, pp. 1–240
 90. Lachraf R, Ayachi Y, Abdelali A et al. (2019) Arbengvec: Arabic-english cross-lingual word embedding model. In: Proceedings of the Fourth Arabic Natural Language Processing Workshop, pp. 40–48. ACL
 91. Khalid U, Beg MO, Arshad MU (2021) Rubert: A bilingual roman urdu bert using cross lingual transfer learning. <http://arxiv.org/abs/2102.11278> 63, 1872–1897