

A Survey of Cross-Lingual Sentiment Analysis Based on Pre-Trained Models

Xudong Luo^{1,2,*}

Sihan Yin^{1,3}

Pingping Lin^{1,4}

*Corresponding author

¹ Guangxi Key Lab of Multi-Source Information Mining & Security, Guangxi Normal University, Guilin, China,

² Professor, luoxd@mailbox.gxnu.edu.cn

³ Student, 1469214793@qq.com

⁴ Student, 1041019568@qq.com

ABSTRACT

With the technology development of natural language processing, many researchers have studied Machine Learning (ML), Deep Learning (DL), monolingual Sentiment Analysis (SA) widely. However, there is not much work on Cross-Lingual SA (CLSA), although it is beneficial when dealing with low resource languages (e.g., Tamil, Malayalam, Hindi, and Arabic). This paper surveys the main challenges and issues of CLSA based on some pre-trained language models and mentions the leading methods to cope with CLSA. In particular, we compare and analyze their pros and cons. Moreover, we summarize the valuable cross-lingual resources and point out the main problems researchers need to solve in the future.

Keywords: multilingual, sentiment analysis, pre-trained language models, BERT, GTP.

INTRODUCTION

The rapid development and popularisation of the Internet inevitably lead to a significant increase in the amount of data (Coffman & Odlyzko, 2002). Moreover, with the massive network texts containing mixed data such as text, audio, and video in various languages, people face complex information.

In particular, SA detects, analyses, and extracts attitudes, opinions, and sentiments expressed by people in a given dataset (Liu & Zhang, 2012; Cambria *et al.*, 2017). SA is a subfield of Natural Language Processing (NLP) (Al-Saqqa & Awajan, 2019), and it is also called opinion mining, orientation analysis, sentiment classification, and subjective analysis. SA tasks involve many NLP problems, including entity recognition, word polarity disambiguation, satire detection, and aspect extraction. The number of issues in a SA task is directly proportional to the difficulties users face in their application.

The SA of online comments needs to consider the sentimental polarity, sentimental intensity, and sentimental polarity analysis. And online sentiment analysis has been applied to many fields to solve different problems, such as understanding the emotions of comments on various social platforms and understanding product reviews to promote product performance (Puranik *et al.*, 2021; Hegde *et al.*, 2021; Yasaswini *et al.*, 2021; Ghanghor *et al.*, 2021). The main task of the research is to identify the subjective attitudes or opinions expressed by people, and the purpose of sentimental intensity analysis is to define the intensity of the commenters' expressions range. Most studies treat SA as a simple classification problem, but SA is a massive suitcase research problem (Cambria *et al.*, 2018). The reason for this is that SA needs to solve many issues in natural language processing, including named entity recognition, word polarity disambiguation, satire detection and aspect extraction. Therefore, each subtask is extremely important and faces unresolved issues. For example, if there are multiple opinion targets for aspect-based SA, the aspect extraction sub-task is ignored, significantly reducing classification accuracy. Therefore, the number of subtasks included in SA tasks is directly proportional to the difficulties they face.

In recent years, with the development of NLP technology, it is gradually applied to various popular fields, such as translation systems, search engines, natural language assistants, and sentiment analysis, and it has also solved many urgent problems in society, which is of great significance (Acheampong, Nunoo-Mensah & Chen, 2021; Yue *et al.*, 2019; Chowdhury, 2003). Especially after combining NLP, ML technology and DL technology for sentiment analysis, monolingual SA research has gradually matured. However, although cross-lingual SA is a unique and essential research direction in SA, there is little work. Moreover, although there have been extensive studies on the polarity of SA, they mainly focus on language with rich resources. English is the most common language in the world. Therefore, most researchers focus on the SA task of English, so they have many ready-made research resources.

On the other hand, research on SA involving Hindi and Arabic is minimal, and the lack of resources has brought problems to researchers, and it has also attracted widespread attention from industry and academia (Yu, 2009). Therefore, to facilitate the research on CLSA, in this paper, we will survey the existing methods for CLSA. CLSA was called multilingual subjective analysis (Banea *et al.*, 2008).

Although there are some surveys on SA, including machine learning-based methods (Sasikala & Sukumaran, 2019), deep learning-based methods (Yadav & Vishwakarma, 2020), and aspect-level SA (Nazir *et al.*, 2020), these surveys are all aimed at SA. Unlike these surveys, in this paper, we survey CLSA based on pre-trained language models. Although some surveys on CLSA (e.g., (Pikuliak, Šimko & Bielikova, 2021)) do not focus on pre-trained model-based methods for CLSA, we do. And they do not include the latest research and techniques since 2019, but we do.

We organise the rest of this paper as follows. The second section briefs CLSA. The third section recaps pre-trained models. The fourth section reviews some cross-lingual pertained model-based methods for CLSA. The fifth section discusses this paper. Finally, the sixth section concludes this paper.

CROSS-LINGUAL SA

This section will give a general discussion of CLSA.

Motivation

The lack of reliable language sentiment resources is an important reason for the CLSA task. Sentiment vocabulary includes SentiWordNet (Esuli & Sebastiani, 2006) and MPQA (Blitzer, Dredze & Pereira, 2007). Similar to these public sentiment resources, most of the sentiment resources are English. There are far more sentiment resources in English with annotations than in other languages. This data imbalance problem has led to CLSA, and it is a time-consuming and laborious task to label the sentiment tags of each language manually. Researchers use “borrowing” a rich language resource to another scarce language for SA to solve the problem.

Thus, the problem of CLSA involves the source language and the target language. The source language refers to a language with many research results of SA and rich sentiment resources. The target language is a language with few research results, lacks sentiment resources, or has a few unreliable sentiment resources.

Main task

Shanahan *et al.* (2004) have explored the use of machine learning and translation technology to mine multiple languages' views as a pioneering work on CLSA. Specifically, this task first uses the rich source language resources to train the sentiment or opinion classifier, then uses the translation tool to translate the target language into the source language and apply the trained classifier to classify. Finally, it divides the review text among different sentiment polarities. A variant strategy based on this is to first translate the existing knowledge of various language databases into the target language, use the translated resources to train the classifier, and then the classifier analyses the sentiments or opinions of the target language. The difference between the two methods is whether to use the source language or the target language to train and build the classifier. The target language is translated into the source language using the source language; the source language resources are translated into the target language if using the target language.

As an essential topic of text analysis, CLSA uses the source language to assist SA in the target language by establishing a connection between a source language and a target language. Next, we should develop appropriate knowledge transfer methods to use annotation data from the source language to supervised or semi-supervised train the target language's sentiment classification model. These methods can automatically use and transform knowledge from the source language to generate and expand pseudo-training data in the target language.

Main methods

Researchers mainly divide CLSA methods into machine translation-based methods (Zhou *et al.*, 2016), machine learning-based methods (Shanahan *et al.*, 2004), structural correspondence learning-based methods (Blitzer, Dredze & Pereira, 2007; Li *et al.*, 2017), transfer learning-based methods (Xu, Xu & Wang, 2011), deep learning-based methods (Dong & De Melo, 2018), and so on. Early researchers mostly used machine translation-based methods and structural correspondence learning-based methods to solve CLSA. However, the in-depth research and practice of ML and DL in NLP brought new ideas to CLSA. As a result, more researchers use machine learning-based methods, transfer learning-based methods, deep learning-based methods, and hybrid methods to solve CLSA problems.

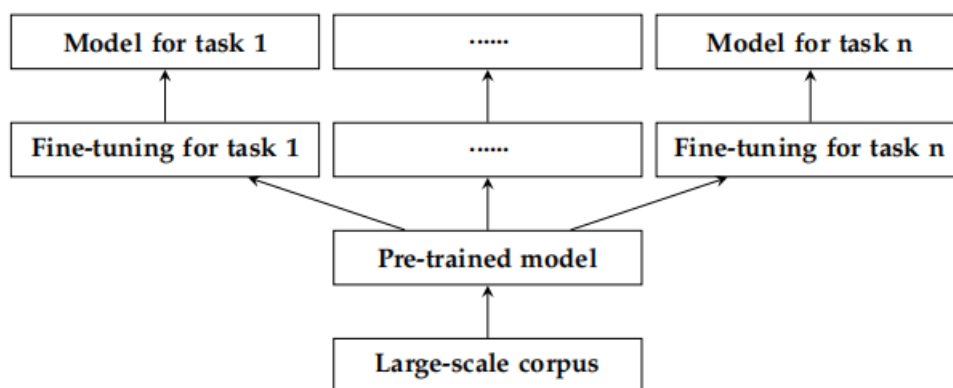
If using learning-based for CLSA, there are mainly three methods for training: (1) Monolingual training and direct conversion: Mikolov, Le, and Sutskever (2013) show that linear transformation can transform the word vector of a source language into the word vector of the target language. For example, when training, we enter a Chinese “angry” word vector; this model can find a word vector that looks similar to English “angry”. When testing, we enter a word vector for “happy” to get a word vector identical to the English one. (2) Monolingual training, respectively, conversion to third-party semantic space: Faruqui and Dyer (2014) do one of the well-known studies. It converts two-word vectors into a third-party space and hopes that the converted Chinese “angry” and converted English “angry” are together in the third-party space. The process of learning this model maximises the similarity between the translated words in our parallel corpus, that is, try to make Chinese “angry” and English “angry” close in the third-party space during the learning process. (3) Bilingual training together: Two representative studies

are (Gouws, Bengio & Corrado, 2015) and (Chandar *et al.*, 2014). The first two methods are performed separately by learning a single-word vector and a bilingual word vector, then find ways to make a relationship between the single-word learned vectors. In short, can we start bilingual learning directly from the beginning? That is to learn the word vector while learning the relationship between bilinguals at the same time. The translated words should be as much as possible, and the not translated words should not matter.

The straightforward way to apply knowledge of one language to another is to use machine translation (such as Google Trans and Yahoo Trans), which can translate the entire corpus into a feasible solution. Language translation is the first step of the machine translation-based method. The second step is to use the Transfer Learning (TL) method (e.g., (Yu & Jiang, 2016)) to transfer the learned knowledge in a source language to a target language for CLSA; TL is a strategy that develops from one language knowledge to another language. However, when applying TL into CLSA, we have to manually select domains to share sentiment words, which may be inefficient.

Challenges

Researchers have done much work on CLSA, but some unsolved problems still exist. (1) CLSA is to identify the sentiment polarity of the target language by using the sentiment knowledge of the source language. The effective sharing of sentiment knowledge across languages is a considerable challenge for CLSA tasks. Designing reliable sentiment knowledge sharing channels and strategies and constructing beneficial cross-lingual sentiment expression associations are the current research on CLSA. (2) Cross-lingual sentiment knowledge transfer and adaptability, cross-lingual sentiment expression correlation research, and inherent differences in sentiment polarity expression. Some researchers are also committed to studying some resources that can apply CLSA, including datasets (Chakravarthi *et al.*, 2020; Patra, Das & Das, 2018; Priyadarshini *et al.*, 2018), sentiment dictionaries, and bilingual word vectors (Ziser & Reichart, 2018; Kuriyozov, Doval & Gómez-Rodríguez, 2020; Kuriyozov *et al.*, 2019; Chen *et al.*, 2020).



Source: the study.

Figure 1: The process of pre-trained the model.

PRE-TRAINED MODELS

In this section, we will brief pre-trained models.

Pre-trained models are a new paradigm in NLP (Li *et al.*, 2021). As shown in Figure 1, the pre-trained model is a trained and saved network. The model is previously trained on a large-scale corpus through self-supervised learning. The pre-trained stage used self-supervised knowledge to learn common sense irrelevant to specific tasks from a large corpus; after fine-tuning, the final model is formed. Moreover, pre-trained can be regarded as regularisation to prevent the model from overfitting small data (Erhan *et al.*, 2010).

From 2013 to 2020, the pre-trained model is updated year by year. As early as 2013, Word2vec started the prelude of the pre-trained model, which Mikolov *et al.* (2013) proposed. It is a pre-trained model to generate word vectors. According to the corpus, the optimised training model quickly and effectively expressed a word into vector form. Word2vec is also a word embedding model widely used in emotion analysis, with excellent analysis performance (Alnawas & Arici, 2021). Word2vec has two algorithms for generating word vectors, Skip-Gram (SG) and Continuous Bag-of-Words (CBOW). According to the research of Mikolov *et al.*, SG has good performance in small training data, but CBOW is very efficient, and its accuracy for frequent words is also higher than SG.

Pennington, Socher, and Manning (2014) propose GloVe to overcome Word2vec's defect: it has a low vector dimension and cannot completely contain the data in the corpus. Meanwhile, GloVe takes advantage of the matrix factorisation method and the shallow window-based method. As a result, it can include the local or global information of specific words in the corpus, which is necessary for improving performance. Moreover, it can be more generalised in the process of word embedding than Word2vec.

Bahdanau, Cho, and Bengio (2015) first used the Attention Model (AM) in machine translation, but now it is more and more widely used in the field of NLP (Galassi, Lippi & Torroni, 2020). The first pre-trained model was developed in 2015, combining a mechanism with Long Short-Term Memory (LSTM) (Hochreiter & Schmidhuber, 1997), breaking the limitation that the traditional encoder-decoder structure relies on a fixed-length vector in its encoding and decoding. The AM + LSTM model retains the intermediate output results of the LSTM encoder on the input sequence and then trains a model to selectively learn these inputs and associate the output sequence with the model output.

In 2016, the pre-trained model based on self-attention came into being. Self-attention has another name, internal attention, which is one of the attention mechanisms and has successfully applied to reading comprehension, abstractive summarisation, text implication and learning task-independent sentence representation (Vaswani *et al.*, 2017). For example, word vector feature extraction is usually used in sentiment analysis, while general embedding methods, such as word2vec, need to combine context to clarify the semantics and lose the sequence information of sentences. Self-attention can effectively solve the above problems.

Vaswani *et al.* (2017) propose the Transformer model, which advantage is that it can solve the issues of the slow training speed and low efficiency of the RNN model by using self-attention to achieve a fast parallel effect. In addition, it can dig deep into the characteristics of Deep Neural Networks (DNN) and improve the efficiency and performance of the model. Beginning in 2018, pre-trained models have entered a prosperous stage. As a result, a large number of pre-trained models have emerged to solve various problems. Among them, GPT (Radford *et al.*, 2018), ELMo (Peters *et al.*, 2018), BERT (Devlin *et al.*, 2019), GPT-2 (Radford *et al.*, 2019), and GPT3 (Brown *et al.*, 2020) are typical representatives of pre-trained models.

The pre-trained model can embed common sense with nothing to do with the target task, such as grammatical and semantic knowledge (Su *et al.*, 2021). The pre-trained model can transfer the learned ability to downstream tasks, which solves resource shortages. The pre-trained model supports many NLP tasks, such as natural language understanding and generation tasks (Devlin *et al.*, 2019; Bao *et al.*, 2020; Yang *et al.*, 2019; Dong *et al.*, 2019). The pre-trained model provides scalable solutions for various applications, and it supports new functions that only need to be fine-tuned through task-specific labelled data. The pre-trained model's initial performance is very high, and the speed of performance improvement in the training phase is also breakneck, and the model's fusion is good. The disadvantage is that the pre-trained model needs to meet physical conditions before being used.

CROSS-LINGUAL PRE-TRAINED MODELS FOR CLSA

Typical cross-lingual pre-trained models include BERT, ELMo, ULMFiT, multilingual BERT, XLM-R, and MetaXL. This section will discuss the pre-trained models for CLSA.

Task-specific CLSA model based on BERT

In bilingual and multilingual countries, Code-Switching is a widespread phenomenon, that is, the use of multiple languages in a single utterance (Priyadarshini *et al.*, 2020; Sitaram *et al.*, 2019; Jose *et al.*, 2020). However, few annotation data sets are available in the Code-Mixed language in the Dravidian language, and there is no way to perform sentiment analysis. Previous work has shown that the BERT-based model achieves the most advanced performance when dealing with SA tasks. To this end, Gupta, Rallabandi and Black (Gupta, Rallabandi & Black, 2021) use BERT (Devlin *et al.*, 2019), mBERT (Devlin *et al.*, 2019), XLM-RoBERTa (Conneau *et al.*, 2020), and TweetEval (Barbieri *et al.*, 2020) for SA on two other code-switched Dravidian languages (Tamil-English and Malayalam English (Chakravarthi *et al.*, 2020)). Their experiments show that task-specific pre-training results in superior zero-shot and supervised performance, better than the performance achieved by leveraging cross-lingual transfer from multilingual BERT models.

Although the model has achieved good performance, its cross-lingual transmission effect on specific data sets is not very good, and the pre-training using cross-lingual and particular tasks does not significantly improve the model's performance, so we can explore how to solve these two problems in the future work.

CLSA model based on ELMo, ULMFiT, and BERT

Sentiment analysis for English has made significant progress, but due to the sparseness of Japanese and its dependence on large data sets, there are little researches on sentiment analysis for Japanese. Today, the pre-trained transfer model has made significant progress. Therefore, Bataa and Wu (2019) focus on studying whether the existing methods are helpful in Japanese. Based on the idea of TL, they are committed to verifying the effectiveness of ELMo (Peters *et al.*, 2018), ULMFiT (Howard & Ruder, 2018), and BERT (Devlin *et al.*, 2019) in Japanese SA tasks. In their research, the model they fine-tune on the pre-trained model to apply to a specific study is called a TL model. Firstly, they use Japanese Wikipedia4 to pre-train ELMo, ULMFiT, and BERT for comparison under the same pre-trained conditions. The experiment's downstream task is to experiment with Rakuten product review and Yahoo movie review datasets and divide its sentiment polarity into two categories and five ones. The experiment uses the Rakuten product review dataset for pre-trained model comparison research and the Yahoo movie review dataset for ablation research. The purpose of ablation research is to understand better the effectiveness of TL models in Japanese SA. Secondly, they trained ELMo with Biattentiv Classification Network (BCN) and

ULMFiT on the Rakuten product review dataset. The training results show that the error rates of ELMo + BCN and ULMFiT in two and five categories are lower than other non-pre-trained models.

On the Rakuten product review dataset, ULMFiT achieved the lowest error percentages on the second classification task, and BERT achieved the lowest error percentages on the five classification tasks. In addition, BERT achieved the lowest error percentages on the Yahoo movie review dataset's double classification task error percentages. The experimental results show that the method based on TL performs better than the specific task model trained three times the data. ELMo and ULMFiT can also show promising results in the small corpus, which provides a feasible solution to low-resource language SA tasks. However, BERT's training performance on the small-scale training set is lacking. Therefore, if selecting BERT as the migration model, a large amount of training data is still needed for functional SA.

Nevertheless, they prove that TL can effectively solve Japanese SA problems. However, whether these pre-trained models are helpful in other natural language processing tasks is worth studying. In addition, if these models are practical, it is worth discussing whether we can get the same results and conclusions in the future.

CLSA model based on Multilingual BERT

The pre-trained model has made good progress in many English NPL tasks. However, there is a lack of training data in other languages, and the cost of generating new resources for a specific language is too high. Dong and de Melo (2019) propose a self-learning framework based on multilingual BERT to address the issues. Choosing multilingual BERT is that it performs well in the cross-lingual transfer of natural language tasks and can better solve sentiment analysis. It first learns from the available English samples and then predicts the unlabeled non-English samples. The samples with high confidence prediction scores are used as the marking examples for the next fine-tuning iteration until the model can implement cross-lingual tasks. For SA from English to Chinese, on datasets Yelp and Chinese hotel reviews, their experiments show that their model is better than the baseline in terms of accuracy.

In the future, it is worth extending the model by introducing a larger dataset with further fine-grained classification and content analysis.

CLSA model based on XLM-R

The emotions expressed by language may be diverse, but a positive tone can help those who attempt to harm themselves (Chakravarthi, 2020). Therefore, Hossain, Sharif, and Hoque (2021) use pre-trained model XLM-R model for CLSA on comments from the social media platform YouTube with 28,451, 20,198, and 10,705 words in English, Tamil and Malayalam, particularly to identify whether a social media comment or article contains a positive tone. Their experiments show that their XLM-R model outperforms ML and DL baselines in terms of F1-score, precision, and recall as evaluation indicators. In addition, the performance of the XML-R model is better than all baseline models. In the future, we can also try to integrate context embedding based on sentiment analysis and pre-training models to explore the feasibility of the research.

CLSA model based on MetaXL

The combination of the pre-trained model and TL can effectively solve NPL tasks. However, for monolingual corpora or languages with very low annotation, the effect of transfer learning is not very good, and multi-lingual representation has little correlation between different languages, which is also a big challenge for CLSA. Meta-learning can make the source language learn and transform under the guidance of the target task (Finn, Abbeel & Levine, 2017), and the pre-trained model based on BERT is the most advanced multi-lingual training model at present. Xia *et al.* (2021) propose a MetaXL model combining meta-learning, TL and pre-training, and evaluating the CLSA task model to solve the above problems. The experiment uses Multi-lingual Amazon Reviews Corpus (MARC) (Keung *et al.*, 2020), SentiPers (Hosseini *et al.*, 2018), and Sentiraama (Gangula & Mamidi, 2018) as data sets, English as the source language, and Persian and Telugu as target language. Xia *et al.* use F1-score as an index to evaluate the performance of the model. The experimental results show that MetaXL can effectively transmit representations in other languages. Combined with the pre-trained model, the source language can be closer to the target language, and the model's performance is good.

This model has achieved good performance by placing fewer representation conversion networks on multiple layers of the pre-trained model. In future work, we can study the effect of placing numerous networks on various layers and improving the transfer language performance.

DISCUSSION

Cross-lingual representation has played an essential role in overcoming language barriers in natural language processing. Pre-trained language models have succeeded in lingual classification problems. Therefore, we can view an SA task as an emotion classification problem to deal with pre-trained language models. Although SA tasks use the pre-trained model widely, it does not perform well when a pre-trained model faces aspect-level SA (e.g., fine-grained SA). Applying pre-trained models in fine-grained SA tasks, especially cross-lingual fine-grained SA tasks, is more challenging.

Table 1 compares different cross-lingual pre-trained models for SA, especially their pros and cons. The pre-training results of TweetEval in specific tasks have excellent zero-shot, and we can use supervision performance as a baseline for future research.

However, the cross-lingual transmission effect of this model on particular data sets is not very good. BERT_{base} uses TL technology and pre-trained models to solve Japanese sentiment classification and shows the possibility of this work in the future. However, the model does not perform K-fold validation, which may lead to inaccurate performance evaluation. Multilingual BERT improves classification accuracy without target language training data and uses unlabeled text to enhance the cross-language transmission of text classification in SA. Finally, XLM-R has cross-lingual synthesis capabilities at different language levels and performs well, but it cannot reflect the intrinsic meaning of sentences. MetaXL makes the target language and the source language closer in the representation space, with good transmission performance, but it has not yet explored placing multiple conversion networks in various layers of the pre-trained model.

Table 2 compares the performance of various cross-lingual pre-trained models, and Figure 2 displays the comparison against different datasets and each evaluation criterion. For the performance comparison of accuracy, because some models do not use this indicator to evaluate model performance, we do not show the accuracy values of these models in the figure, and the same is true for other performance indicators. As shown in Table 2 and Figure 2, for accuracy, multilingual BERT achieves the best performance, reaching 83.10%; for precision, recall and F1-score, XLM-R achieves the best performance, reaching 0.931; and for error percentage, BERT_{base} achieves the best performance, only 4.68%.

Table 1: Cross-lingual pre-trained models for SA.

Reference	Model	Task	Advantage	Disadvantage	Dataset
(Gupta, Rallabandi & Black, 2021)	TweetEval	Task-specific pre-training 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 (Chakravarthi, 2020), Malayalam English (Chakravarthi, 2020), Sentimix Hinglish (Patwa <i>et al.</i> , 2020)
(Bataa & Wu, 2019)	BERT _{base}	Japanese sentiment classification	Use TL technology and pre-trained models to solve Japanese sentiment classification	Did not perform K-fold validation	Japanese Rakuten product review binary (Zhang & LeCun, 2017), five class Yahoo datasets
(Dong & de Melo, 2019)	Multilingual BERT	Sentiment classification	Improve classification accuracy without target language training data	No shortcomings for the time being	Yelp and Chinese hotel reviews (Zhang, Zhao & LeCun, 2015), (Chen <i>et al.</i> , 2018)
(Hossain, Sharif & Hoque, 2021)	XLM-R	Multi-classification text sentiment analysis	Have cross-lingual comprehension capabilities at different language levels	Cannot apprehend the inherent meaning of sentences	Hope speech corpus (Chakravarthi & Muralidaran, 2021)
(Xia <i>et al.</i> , 2021)	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 pre-trained model	Multi-lingual Amazon Reviews Corpus (MARC) (Keung <i>et al.</i> , 2020), SentiPers (Hosseini <i>et al.</i> , 2018), Sentiraama (Gangula & Mamidi, 2018).

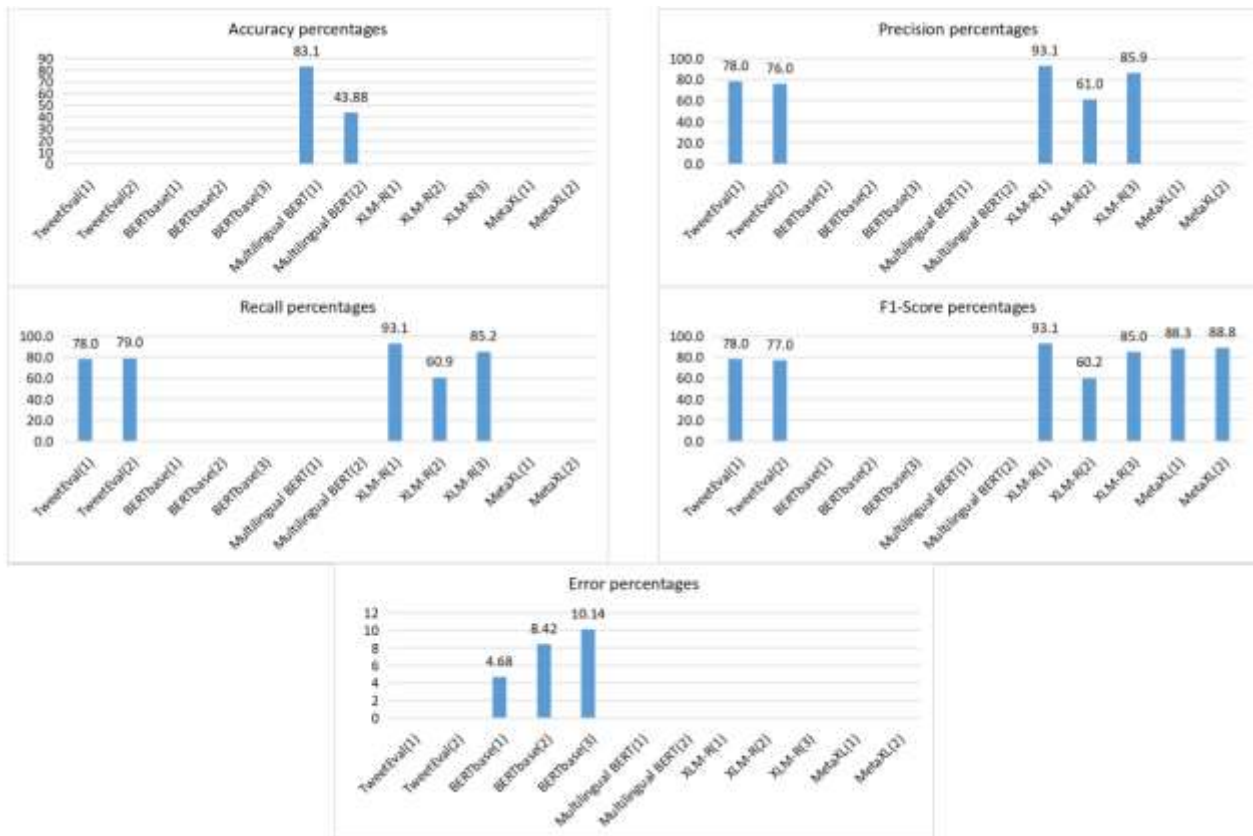
Source: This study.

Table 2: The performance comparison of various cross-lingual pre-trained models for SA.

Reference	Model	Accuracy	Precision	Recall	F1-score	Error
(Gupta, Rallabandi & Black, 2021)	TweetEval(1)	N/A	78%	78%	78%	N/A
(Gupta, Rallabandi & Black, 2021)	TweetEval(2)	N/A	76%	79%	77%	N/A
(Bataa & Wu, 2019)	BERT _{base} (1)	N/A	N/A	N/A	N/A	4.68%
(Bataa & Wu, 2019)	BERT _{base} (2)	N/A	N/A	N/A	N/A	8.42%
(Bataa & Wu, 2019)	BERT _{base} (3)	N/A	N/A	N/A	N/A	10.14%

(Dong & de Melo, 2019)	Multilingual BERT(1)	83.10%	N/A	N/A	N/A	N/A
(Dong & de Melo, 2019)	Multilingual BERT(2)	43.88%	N/A	N/A	N/A	N/A
(Hossain, Sharif & Hoque, 2021)	XLM-R(1)	N/A	93.1%	93.1%	93.1%	N/A
(Hossain, Sharif & Hoque, 2021)	XLM-R(2)	N/A	61.0%	60.9%	60.2%	N/A
(Hossain, Sharif & Hoque, 2021)	XLM-R(3)	N/A	85.9%	85.2%	85.4%	N/A
(Xia <i>et al.</i> , 2021)	MetaXL(1)	N/A	N/A	N/A	88.3%	N/A
(Xia <i>et al.</i> , 2021)	MetaXL(2)	N/A	N/A	N/A	88.8%	N/A

Source: This study.



Source: This study.

Figure 2: The performance comparison of cross-lingual pre-trained models.

CONCLUSION

CLSA is an essential kind of SA. In this paper, we review the application of pre-trained models in CLSA, compare their similarities and differences, including pros and cons and point out the direction of future improvement. Besides, we also provide some available cross-lingual resources. One of CLSA's future works is the fine-grained CLSA, which is very challenging but significant in practice. Compared with text-based CLSA, multimodal CLSA is also a topic worthy of in-depth study. Besides, pre-trained language models are essential for NLP. Therefore, the application of pre-trained language models to CLSA deserves researchers' more attention.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (No. 61762016) and Graduate Student Innovation Project of School of Computer Science and Engineering, Guangxi Normal University (JXXYYJSCXXM-2021-001).

REFERENCES

Acheampong, F. A., Nunoo-Mensah, H., & Chen, W. (2021). Transformer models for text-based emotion detection: a review of BERT-based approaches. *Artificial Intelligence Review*, 54, 5789–5829. <https://doi.org/10.1007/s10462-021-09958-2>

- Alnawas, A., & Arici, N. (2021). Effect of word embedding variable parameters on Arabic sentiment analysis performance. *arXiv preprint arXiv:2101.02906*.
- Al-Saqqa, S., & Awajan, A. (2019). The use of word2vec model in sentiment analysis: A survey. In *Proceedings of the 2019 International Conference on Artificial Intelligence, Robotics and Control* (pp. 39-43). AIRC'19, Cario, Egypt, December 14-19.
- Bahdanau, D., Cho, K. H., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations*. ICLR 2015, San Diego, USA, May 7-9.
- Banea, C., Mihalcea, R., Wiebe, J., & Hassan, S. (2008). Multilingual subjectivity analysis using machine translation. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing* (pp. 127-135). EMNLP2008, Honolulu, Hawaii, USA, October 25-27.
- Bao, H., Dong, L., Wei, F., Wang, W., Yang, N., Liu, X., Wang, Y., Gao, J., Piao, S., Zhou, M. & Hon, H. W. (2020, November). Unilmv2: Pseudo-masked language models for unified language model pre-training. In *International Conference on Machine Learning* (pp. 642-652). PMLR.
- Barbieri, F., Camacho-Collados, J., Anke, L. E., & Neves, L. (2020). TweetEval: Unified Benchmark and Comparative Evaluation for Tweet Classification. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings* (pp. 1644-1650). <https://doi.org/10.18653/v1/2020.findings-emnlp.148>
- 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). Florence, Italy, July 28 - August 2. <https://doi.org/10.18653/v1/P19-1458>
- Blitzer, J., Dredze, M., & Pereira, F. (2007). Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *Proceedings of the 45th annual meeting of the association of computational linguistics* (pp. 440-447). Prague, Czech Republic.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.
- Cambria, E. , Poria, S. , Gelbukh, A. , & Thelwall, M. . (2018). Sentiment analysis is a big suitcase. *Intelligent Systems*, IEEE, 32(6), 74-80. <https://doi.org/10.1109/MIS.2017.4531228>
- Cambria, E., Das, D., Bandyopadhyay, S., & Feraco, A. (2017). Affective computing and sentiment analysis. In *A practical guide to sentiment analysis* (pp. 1-10). Springer, Cham. https://doi.org/10.1007/978-3-319-55394-8_1
- Chakravarthi, B. R. (2020). HopeEDI: A multilingual hope speech detection dataset for equality, diversity, and inclusion. In *Proceedings of the Third Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media* (pp. 41-53).
- Chakravarthi, B. R., & Muralidaran, V. (2021). Findings of the shared task on Hope Speech Detection for Equality, Diversity, and Inclusion. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion* (pp. 61-72).
- Chakravarthi, B. R., Jose, N., Suryawanshi, S., Sherly, E., & McCrae, J. P. (2020). A Sentiment Analysis Dataset for Code-Mixed Malayalam-English. In *Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)* (pp. 177-184). SLTU-CCURL 2020, Marseille, France, May 11-16.
- Chakravarthi, B. R., Muralidaran, V., Priyadharshini, R., & McCrae, J. P. (2020). Corpus Creation for Sentiment Analysis in Code-Mixed Tamil-English Text. In *Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)* (pp. 202-210).
- Chandar, A. S., Lauly, S., Larochele, H., Khapra, M. M., Ravindran, B., Raykar, V., & Saha, A. (2014). An autoencoder approach to learning bilingual word representations. In *Proceedings of the 27th International Conference on Neural Information Processing Systems*, 2, 1853-1861.
- Chen, X., Sun, Y., Athiwaratkun, B., Cardie, C., & Weinberger, K. (2018). Adversarial deep averaging networks for cross-lingual sentiment classification. *Transactions of the Association for Computational Linguistics*, 6, 557-570. http://dx.doi.org/10.1162/tacl_a_00039
- Chen, Z., Shen, S., Hu, Z., Lu, X., Mei, Q., & Liu, X. (2019, May). Emoji-powered representation learning for cross-lingual sentiment classification. In *The World Wide Web Conference* (pp. 251-262). WWW'19, San Francisco, CA, USA, May 13-17. <https://doi.org/10.24963/ijcai.2020/649>
- Chowdhury, G. G. (2003). Natural language processing. *Annual review of information science and technology*, 37(1), 51-89. <https://doi.org/10.1002/aris.1440370103>
- Coffman, K. G., & Odlyzko, A. M. (2002). Internet growth: Is there a "Moore's Law" for data traffic? In *Handbook of massive data sets* (pp. 47-93). Springer, Boston, MA. <http://dx.doi.org/10.2139/ssrn.236108>
- Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Luke Zettlemoyer, M. & Stoyanov, V. (2020). Unsupervised Cross-lingual Representation Learning at Scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 8440-8451). <https://doi.org/10.18653/v1/2020.acl-main.747>
- Devlin, J., Chang, M. W., Lee, K., & 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*, 1, 4171-4186. NAACL-HLT 2019, Minneapolis, Minnesota, June 2-7.

- Dong, L., Yang, N., Wang, W., Wei, F., Liu, X., Wang, Y., Gao, J., Zhou, M. & Hon, H. W. (2019). Unified language model pre-training for natural language understanding and generation. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems* (pp. 13063-13075).
- Dong, X. L., & de 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 (EMNLP-IJCNLP)* (pp. 6306-6310). Hong Kong, China, November 3–7. <https://doi.org/10.18653/v1/D19-1658>
- Dong, X., & De Melo, G. (2018). Cross-lingual propagation for deep sentiment analysis. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*. Hilton New Orleans Riverside, New Orleans, Louisiana, USA, February 2-7.
- Erhan, D., Courville, A., Bengio, Y., & Vincent, P. (2010). Why does unsupervised pre-training help deep learning?. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics* (pp. 201-208). JMLR Workshop and Conference Proceedings. AISTATS 2010, Chia La-guna Resort, Sardinia, Italy.
- 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). Gothenburg, Sweden, April 26-30.
- Finn, C., Abbeel, P., & Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning* (pp. 1126-1135). PMLR.
- Galassi, A., Lippi, M., & Torrioni, P. (2020). Attention in natural language processing. *IEEE Transactions on Neural Networks and Learning Systems*, 32(10), 4291-4308. <https://doi.org/10.1109/TNNLS.2020.3019893>
- Gangula, R. R. R., & Mamidi, R. (2018). Resource creation towards automated sentiment analysis in telugu (a low resource language) and integrating multiple domain sources to enhance sentiment prediction. In *Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018)*. Miyazaki, Japan.
- Ghanghor, N., Ponnusamy, R., Kumaresan, P. K., Priyadharshini, R., Thavareesan, S., & Chakravarthi, B. R. (2021). IITK@LT-EDI-EACL2021: Hope Speech Detection for Equality, Diversity, and Inclusion in Tamil, Malayalam and English. In *Proceedings of the 1st Workshop on Language Technology for Equality, Diversity and Inclusion* (pp. 197-203).
- Gouws, S., Bengio, Y., & Corrado, G. (2015, June). Bilbowa: Fast bilingual distributed representations without word alignments. In *International Conference on Machine Learning* (pp. 748-756). PMLR.
- Gupta, A., Rallabandi, S. K., & Black, A. W. (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. 73-79).
- Hegde, S. U., Hande, A., Priyadharshini, R., Thavareesan, S., & Chakravarthi, B. R. (2021). UVCE-IIT@DravidianLangTech-EACL2021: Tamil Troll Meme Classification: You need to Pay more Attention. In *Proceedings of the 1st Workshop on Speech and Language Technologies for Dravidian Languages* (pp. 180-186).
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hossain, E., Sharif, O., & Hoque, M. M. (2021). NLP-CUET@LT-EDI-EACL2021: 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. 168-174).
- Hosseini, P., Ramaki, A. A., Maleki, H., Anvari, M., & Mirroshandel, S. A. (2018). SentiPers: a sentiment analysis corpus for Persian. *arXiv preprint arXiv:1801.07737*.
- 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*, 1, 328-339. <https://doi.org/10.18653/v1/P18-1031>
- Jose, N., Chakravarthi, B. R., Suryawanshi, S., Sherly, E., & McCrae, J. P. (2020). A survey of current datasets for code-switching research. In *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)* (pp. 136-141). Coimbatore, India, March 6-7. <https://doi.org/10.1109/ICACCS48705.2020.9074205>
- Keung, P., Lu, Y., Szarvas, G., & Smith, N. A. (2020). The Multilingual Amazon Reviews Corpus. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 4563-4568).
- Kuriyozov, E., Doval, Y., & Gómez-Rodríguez, C. (2020). Cross-Lingual Word Embeddings for Turkic Languages. In *Proceedings of the 12th Language Resources and Evaluation Conference* (pp. 4054-4062). LREC 2020, Marseille, France, May 11-16.
- Kuriyozov, E., Matlatipov, S., Alonso, M. A., & Gómez-Rodríguez, C. (2019). Deep learning vs. classic models on a new Uzbek sentiment analysis dataset. *Human Language Technologies as a Challenge for Computer Science and Linguistics*, 258-262.
- Li, J., Tang, T., Zhao, W. X., & Wen, J. R. (2021). Pretrained Language Models for Text Generation: A Survey. *arXiv preprint arXiv:2105.10311*.
- Li, N., Zhai, S., Zhang, Z., & Liu, B. (2017). Structural correspondence learning for cross-lingual sentiment classification with one-to-many mappings. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence* (pp. 3490-3496). AAAI-17, San Francisco, California USA, February 4-9.
- Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In *Mining text data* (pp. 415-463). Springer, Boston, MA. https://doi.org/10.1007/978-1-4614-3223-4_13
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

- Mikolov, T., Le, Q. V., & Sutskever, I. (2013). Exploiting similarities among languages for machine translation. *arXiv preprint arXiv:1309.4168*.
- Nazir, A., Rao, Y., Wu, L., & Sun, L. (2020). Issues and challenges of aspect-based sentiment analysis: a comprehensive survey. *IEEE Transactions on Affective Computing*. <https://doi.org/10.1109/TAFFC.2020.2970399>
- Patra, B. G., Das, D., & Das, A. (2018). Sentiment analysis of code-mixed indian languages: An overview of sail_code-mixed shared task@ icon-2017. *arXiv preprint arXiv:1803.06745*.
- Patwa, P., Aguilar, G., Kar, S., Pandey, S., Srinivas, P. Y. K. L., Gambäck, B., Chakraborty, T., Solorio, T. & Das, A. (2020). SemEval-2020 Task 9: Overview of Sentiment Analysis of Code-Mixed Tweets. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation* (pp. 774-790).
- Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1532-1543). Doha, Qatar, October 25-29.
- Peters, M., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018, June). Deep Contextualized Word Representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1, 2227-2237. NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6.
- Pikuliak, M., Šimko, M., & Bielikova, M. (2021). Cross-lingual learning for text processing: A survey. *Expert Systems with Applications*, 165, 113765. <https://doi.org/10.1016/j.eswa.2020.113765>
- Priyadharshini, R., Chakravarthi, B. R., Vegupatti, M., & McCrae, J. P. (2020). Named entity recognition for code-mixed Indian corpus using meta embedding. In *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)* (pp. 68-72). Coimbatore, India, March 6-7. <https://doi.org/10.1109/ICACCS48705.2020.9074379>
- Puranik, K., Hande, A., Priyadharshini, R., Thavareesan, S., & Chakravarthi, B. R. (2021). IIIT@ LT-EDI-EACL2021-Hope Speech Detection: There is always hope in Transformers. In *Proceedings of the 1st Workshop on Language Technology for Equality, Diversity and Inclusion* (pp. 98-106).
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. Retrieved from <https://www.cs.ubc.ca/~amuham01/LING530/papers/radford2018improving.pdf>.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8), 9.
- Sasikala, D., & Sukumaran, S. (2019). A survey on lexicon and machine learning based classification methods for sentimental analysis. *IJRAR-International Journal of Research and Analytical Reviews (IJRAR)*, 6(2), 256-259. <http://doi.org/10.1729/Journal.22468>
- Sebastiani, F., & Esuli, A. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th International Conference on Language Resources and Evaluation* (pp. 417-422). LREC'06, Genoa, Italy.
- Shanahan, J. G., Grefenstette, G., Qu, Y., & Evans, D. A. (2004). Mining multilingual opinions through classification and translation. In *Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text*.
- Sitaram, S., Chandu, K. R., Rallabandi, S. K., & Black, A. W. (2019). A survey of code-switched speech and language processing. *arXiv preprint arXiv:1904.00784*.
- Su, Y., Han, X., Zhang, Z., Lin, Y., Li, P., Liu, Z., Zhou, J. & Sun, M. (2021). Cokebert: Contextual knowledge selection and embedding towards enhanced pre-trained language models. *AI Open*, 2, 127-134. <https://doi.org/10.1016/j.aiopen.2021.06.004>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. & Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998-6008). NIPS 2017, Long Beach, CA, USA.
- Xia, M., Zheng, G., Mukherjee, S., Shokouhi, M., Neubig, G., & Awadallah, A. H. (2021). MetaXL: Meta representation transformation for low-resource cross-lingual learning. *arXiv preprint arXiv:2104.07908*.
- Xu, R., Xu, J., & Wang, X. (2011). Instance level transfer learning for cross lingual opinion analysis. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis* (pp. 182-188). WASSA 2011, Portland, Oregon, USA, June 24.
- Yadav, A., & Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: a review. *Artificial Intelligence Review*, 53(6), 4335-4385. <https://doi.org/10.1007/s10462-019-09794-5>
- Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, 32.
- Yasaswini, K., Puranik, K., Hande, A., Priyadharshini, R., Thavareesan, S., & Chakravarthi, B. R. (2021). IIIT@ DravidianLangTech-EACL2021: Transfer Learning for Offensive Language Detection in Dravidian Languages. In *Proceedings of the 1st Workshop on Speech and Language Technologies for Dravidian Languages* (pp. 187-194).
- Yu, C. M. (2009). Mining opinions from product review: Principles and algorithm analysis. *Information Studies: Theory & Application*, 32(7), 124-128. doi: 10.16353/j.cnki.1000-7490.2009.07.023 (in Chinese).
- Yu, J., & Jiang, J. (2016). Learning sentence embeddings with auxiliary tasks for cross-domain sentiment classification. *Association for Computational Linguistics* (pp. 236-246). Stroudsburg, PA, November 1-5..
- Yue, L., Chen, W., Li, X., Zuo, W., & Yin, M. (2019). A survey of sentiment analysis in social media. *Knowledge and Information Systems*, 60(2), 617-663. Springer. <https://doi.org/10.1007/s10115-018-1236-4>

- Zhang, X., & LeCun, Y. (2017). Which encoding is the best for text classification in chinese, english, japanese and korean? *arXiv preprint arXiv:1708.02657*.
- Zhang, X., Zhao, J., & LeCun, Y. (2015). Character-level convolutional networks for text classification. *Advances in neural information processing systems*, 28, 649-657.
- Zhou, G., Zhu, Z., He, T., & Hu, X. T. (2016). Cross-lingual sentiment classification with stacked autoencoders. *Knowledge and Information Systems*, 47(1), 27-44. Springer, Verlag London. <https://doi.org/10.1007/s10115-015-0849-0>
- Ziser, Y., & Reichart, R. (2018). Deep pivot-based modeling for cross-language cross-domain transfer with minimal guidance. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 238-249). Brussels, Belgium, October 31 - November 4.