Emotion Recognition in Social Network Texts Based on A Multilingual Architecture

1st Jiajun Zou
School of Humanities
Tsinghua University
Beijing, China
zjj21@mails.tsinghua.edu.cn

2nd Yexuan Zhang
Department of Electronic Engineering
Tsinghua University
Beijing, China
zhangyex21@mails.tsinghua.edu.cn

3rd Jinshuai Yang
Department of Electronic Engineering
Tsinghua University
Beijing, China
yjs20@mails.tsinghua.edu.cn

4thSixing Wu
School of Control and Computer Engineering
North China Electric Power University
Beijing, China
wusx@ncepu.edu.cn

5th Minghu Jiang
School of Humanities
Tsinghua University
Beijing, China
jiang.mh@mail.tsinghua.edu.cn

6th Yongfeng Huang
Department of Electronic Engineering
Tsinghua University
Zhongguancun Laboratory
Beijing, China
yfhuang@mail.tsinghua.edu.cn

***Abstract*—Languages are used by people to describe and categorize their emotional experiences and perspectives. For many applications, it is crucial to apply techniques like machine learning in social network texts to identify emotions. Most of these technologies now in use only detect a small number of emotion categories such as anger, happiness, sadness and so on, they do not distinguish more fine-grained levels of emotions. Additionally, they frequently concentrate on modeling the relationships between various emotions, ignoring the emotional semantic relations between different languages. Therefore, in this paper, we improve the Recognition of Emotion by utilizing a Multilingual architecture that combines machine Translation and Attention mechanism, enabling one language to provide additional emotional information for another language (REMTA). The experimental results on a fine-grained emotion dataset labeled with 28 categories show a performance improvement compared with other models, demonstrating the efficacy of our architecture.**

***Index Terms*—Social media text, Emotion recognition, Multilingual.**

I. Introduction

Thanks to the development of the Internet, people can now communicate their opinions, attitudes, and feelings through social media platforms such as Facebook and Twitter. This has led to a variety of languages being used to describe and categorize emotional experiences and perspectives. The task of analyzing the subjective information is essential to natural language processing (NLP), which has lately attracted the interest of numerous academics and industries [1]–[3]. Studying people's views, feelings, emotions, appraisals, and attitudes about things like goods, services, organizations, people, situations, events, themes, and their qualities is the goal of sentiment analysis or opinion mining [4].

Supported by the National Key R&D Program of China (Grant No. 2022YFC3302104) and the National Key Natural Science Foundation Project of China under Grants 62036001.

Yongfeng Huang is the corresponding author.

Generally speaking, emotions play an important role in successful and effective human-human communication [5]. Analyzing a text's sentiment polarity alone is usually insufficient; we also need to examine the emotions it expresses, such as happiness, sadness, anger, etc [6]. However, due to the annotated dataset [7]–[9], most of the existing methods for classifying emotions are based on basic emotions proposed by Ekman [10] or Plutchik [11] models, for example, these emotion models classify emotions into coarse-grained forms, which cannot describe a more detailed depiction of human emotions, and often exclude the self-conscious emotions (e.g. guilt, shame, embarrassment, social anxiety, pride, etc.) [12]. Therefore, it is important to develop methods that can recognize emotions with a more fine-grained level of detail.

Although there have been many researches here that have contributed to this work [13]–[15], they overlooked the impact of different languages on emotion recognition. There have been proved that the semantics of emotion concepts of different languages are related to each other [16]. And from this point of view, can one language help identify emotions in another language? If possible, what kind of model should we design to tap into this potential? In this paper, we propose our method.

The main contributions of our research are as follows:

(1) We propose a simple multilingual architecture which can make the emotional information between languages interact with each other.

(2) Experimental results on a dataset tagged with 28 emotion categories show the effectiveness of the architecture.

(3) The research may bring some thoughts to the study of emotion between different languages, such as why one language can provide some emotional information for another language, whether it is related to cultural transmission, or whether the results can further prove the theory of constructed emotion [17].

2375-9259/23/$31.00 ©2023 IEEE
DOI 10.1109/ICDMW60847.2023.00109

II. Related Work

There have been many works in sentiment analysis in recent years. In order to better infer polarity from text, SenticNet 7 [1] was proposed, it is a neurosymbolic AI system that leverages subsymbolic models, such as auto-regressive language models and kernel methods. In addition, there are new trends on neurosymbolic AI for explainable sentiment analysis. SKIER is [2] a symbolic knowledge integrated model for conversational emotion recognition, which surpasses baseline models on several indices. And pre-trained language models (PTLM) play important roles in sentiment analysis. The bias of these PTLMs on sentiment analysis and emotion detection are also studied [3]. These latest research developments all indicate that the field of sentiment analysis is attracting researchers' attention.

Sentiment analysis is concentrated on sentiment polarity, which is different from emotion detection. Compared to sentiment polarity, recognizing fine-grained emotions is more complex and difficult, which can be divided into single-label recognition and multi-label recognition, and there have been many methods to deal with this task, such as [15], [18]–[21]. In fact, in real life, people tend to express multiple emotions in a single sentence, and carrying out single-label emotion classification would not align the practical needs, so we need to study the multi-label emotion recognition task. One technical route [13], [20] focuses on uncovering the correlations between emotions to improve the accuracy of model predictions. However, this approach is often constrained by predefined emotional categories and limited access to explicit and implicit relationships between emotions. Another approach relies on transfer learning [22], [23] and aims to boost the ability of neural network models to generalize by sharing specific features or parameters. Additionally, the cross-lingual approach [24] uses annotated emotion resources from one language (usually English) to classify the emotions in text documents written in another language, rather than the language itself. There is research using a multilingual method to realize this goal, which is based on Ekman's model, and just concatenates features from languages, such as words with information gain, unigrams, and bigrams. However, this method requires manual design [14] and may not effectively utilize the features of the second language. Further research is needed to improve cross-lingual emotion recognition.

Interestingly, Jackson et al. [16] estimate emotion semantics across a sample of 2,474 spoken languages using "colexification", they find that emotion concepts have different patterns of association in different language families and geographically closer language families tend to colexify emotion concepts in more similar ways than distant language families. This work also reveals that valence and activation serve as universal constraints to variability in emotion semantics, a common underlying structure in the meaning of emotion concepts across language. That is to say, the emotion semantics in different languages are related to each other, but differ from each other, which could be described by a cultural evolution framework [25]. Furthermore, using a foreign language could change our choice such as risk, inference, and morality [26], which may be explained by emotion, psychological distance and increased deliberation. From this perspective, we can consider that the emotion semantics in one language could provide additional emotional semantic information to another language. Therefore, it is important to not overlook the role of another language when recognizing emotions in one language.

In this paper, for the recognition of emotions, we propose a multilingual architecture that adopts machine translation and attention mechanism [27] (REMTA). Specifically, we apply a machine translation tool over corpora originally written in English, these texts are automatically translated into four languages. We employ attention mechanism to focus on and integrate the important information of these features from English and another language. In each language, we use stacked Long Short-Term Memory Networks (LSTM) [28] to fuse its contextual information. Furthermore, the GoEmotions dataset [29], which is tagged with 28 emotion categories, is also used in our research. We compare our approach with representative baselines and find that our models show improved performance for fine-grained emotion recognition, thus proving the effectiveness of our architecture. It should be emphasized that we do not aim to propose a very innovative model, but to point out that the interaction of emotional information between languages can be realized through our architecture. And this shows that the emotional information of one language can assist the emotion recognition of another language.

III. Architecture

In this section, we will introduce the proposed architecture.

A. Overall architecture

Considering a multi-label classification problem, suppose we have N predefined emotion labels, and we use $G = \{g_1, g_2, ..., g_N\}$ to denote the label set. The goal of this task is to predict a label subset belongs to G according to the input sentence $s = \{w_1, w_2, ..., w_L\}$, where L is the length of the sentence.

Fig.1 shows the overall architecture of REMTA. The architecture can be divided into several components and the details will be introduced in the following parts. In brief, we first use machine translation to automatically translate the original corpus into other languages; then we obtain the sentence representations of each language by a multilingual pre-trained language model; we use stacked LSTMs to fully fuse the contextual information in each language; and we also adopt the attention mechanism to fuse the representation vectors of the two languages; finally, the fully-connected layer is employed for emotion classification.

B. Machine translation module

Machine translation technology is now relatively mature and widely used, so in this module, we use a common machine translation method to obtain the corresponding translated texts.

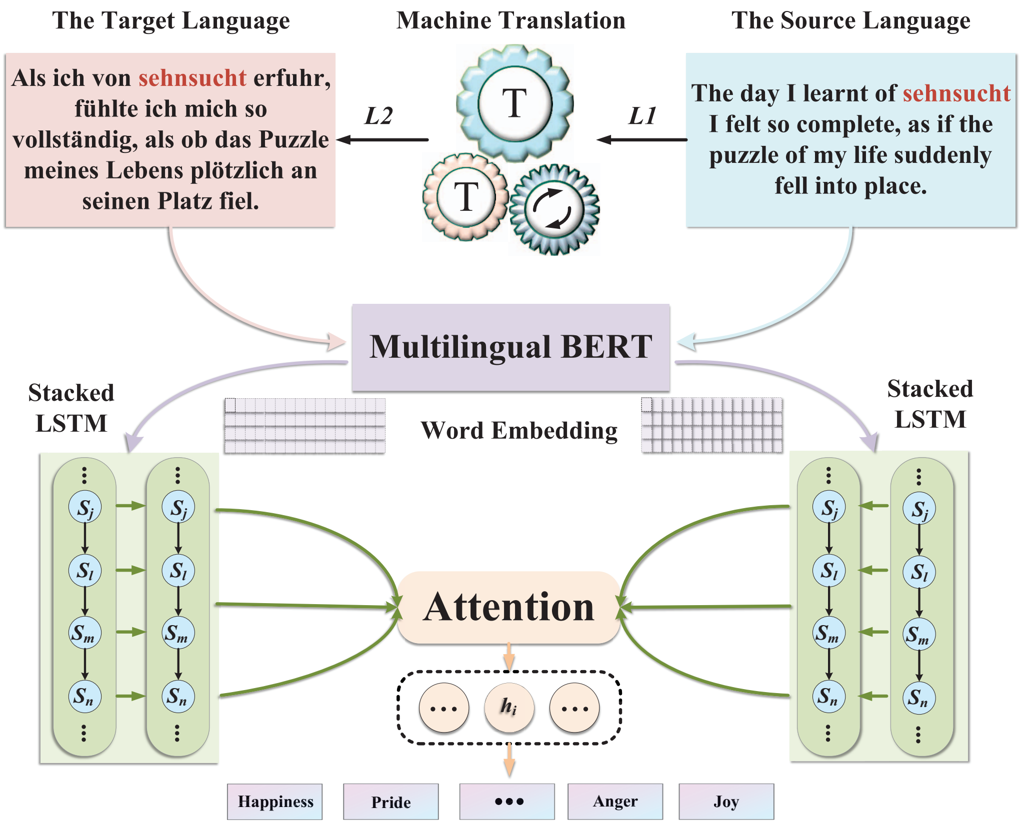

Fig. 1. The overall architecture of REMTA.

That is to say, for an original sentence s labeled with emotions g belongs to G, written as $\{(s,g)|g \in G\}$, we get $\{(s,t,g)|g \in G\}$, where t is the target language sentence translated by the machine.

C. Stacked LSTM module

We can get the representations of these sentences from different languages through a multilingual pre-trained model, we denoted them as follows:

$$E_s = \{w_{s1}, w_{s2}, ..., w_{sL}\}, \tag{1}$$

$$E_t = \{w_{t1}, w_{t2}, ..., w_{tL}\}, \tag{2}$$

where w_{si} and w_{ti} is the word embedding in source language and target language, $w \in \mathbb{R}^{d_B}$, d_B is the dimension of the word embedding.

For each language sentence, we use a stacked LSTM to encode the contextual information of words and the output of it is as follows:

$$\begin{aligned} H_s &= \{h_{s1}, h_{s2}, ..., h_{sL}\} \\ &= Stacked\_LSTM(E_s) \\ &= Stacked\_LSTM(\{w_{s1}, w_{s2}, ..., w_{sL}\}), \end{aligned} \tag{3}$$

$$\begin{aligned} H_t &= \{h_{t1}, h_{t2}, ..., h_{tL}\} \\ &= Stacked\_LSTM(E_t) \\ &= Stacked\_LSTM(\{w_{t1}, w_{t2}, ..., w_{tL}\}), \end{aligned} \tag{4}$$

where $h_{si} \in \mathbb{R}^{d_h}$ and $h_{ti} \in \mathbb{R}^{d_h}$, d_h is the dimension of the hidden state of the stacked LSTM.

D. Attention mechanism

After the stacked LSTM learns the features of the respective language sentences, we make the representations of the different language sentences interact by using attention mechanism, ensuring that the crucial information is attended to in the two languages. The attention mechanism generates the attention vector α_i by:

$$\alpha_i = \frac{exp(\gamma(h_{si}, h_{ti}))}{\sum_{i=0}^{L} exp(\gamma(h_{sj}, h_{tj}))}, \tag{5}$$

where $\gamma(\cdot)$ is a score function. We use a non-linear function *tanh* as the score function which calculates the importance of h_{si} in the original language sentence and we only calculate the attention once.

Then we weight and sum the word representations in the original sentence according to the attention weights to obtain their representations by:

$$H_r = \sum_{i=0}^{L} \alpha_i h_{si} \tag{6}$$

E. Emotion classification

At last, the feature representation of the original language sentence is fed into a fully connected layer, and we use *sigmoid* as the activation function, thus predicting the emotion probability distribution of the sentence:

$$y^o = sigmoid(WH_r + b), \tag{7}$$

where W and b are the weight matrix and bias, respectively.

F. Model training

During the model training phase, we optimize the parameters from the networks. And we use binary cross-entropy loss to compute the loss between the model predictions and categories label values:

$$L = -\frac{1}{n}\sum_i y_i * log(y_i^o) + (1 - y_i) * log(1 - y_i^o), \tag{8}$$

where y_i is the target label.

IV. Experiment and results

A. Experimental setup

All experiments are conducted on an open source English dataset (GoEmotions) annotated from English Reddit[1] comments, labeled with 28 emotion categories [29], the train / validation / test split is 43,405 / 5,426 / 5,427 samples respectively. The statistics of the dataset labels are shown in Fig.2. We can see that the number of these emotional labels is unbalanced.

We use Baidu Fanyi API[2] to perform translation task over this dataset. The original English (En) texts are translated into four common languages in advance: Chinese (Zh), French (Fr), Russian (Ru) and German (De), it should be emphasized that there are no additional parameters from the branch of the translated language. The multilingual BERT[3] (MulBERT) is adopted to obtain the representations of these different language sentences. The hidden size of the stack LSTMs is 768, the number of recurent layers is set to 2. For the sake of fairness, we keep the same hyper-parameters as in [29] and [30], we set the learning rate to 5e-5 and batch size to 16. We train the model for 4 epochs. The set of random seeds is $\{42, 44, 47\}$. And the average of F1 score, precision and recall corresponding to each emotion are used as the evaluation metrics, respectively.

B. Baselines

We compare the performance of REMTA with several main categories of baselines.

Pre-trained language models: BERT, RoBERTa, DistllBERT, XLNet and ELECTRA. We take the results from [30].

Sequence-to-emotion approach: Seq2Emo, which implicitly models the emotion correlations in a bi-directional decoder [20].

[1]https://www.reddit.com/
[2]https://fanyi-api.baidu.com/
[3]https://huggingface.co/bert-base-multilingual-cased

Knowledge-enhanced models: KEA-ELECTRA and KEA-BERT adopt Knowledge-Embedded Attention (KEA) to use knowledge from emotion lexicons to augment the contextual representations from pre-trained models (ELECTRA and BERT), respectively [21], and we change its random seeds to match ours. We also adopt multilingual BERT (MulBERT) to conduct experiments.

Multi-task learning framework: it models definitions of emotions by using masked language modeling (MLM) and class definition prediction (CDP) tasks [19], and its parameters are set to be the same as ours. Multilingual BERT (MulBERT) is also used in this framework.

TABLE I
Results of different models from GoEmotions

Model	Macro-F1	Precision	Recall
BERT [29]	46.00	40.00	63.00
DistillBERT [30]	48.00	-	-
RoBERTa [30]	49.00	-	-
XLNet [30]	48.00	-	-
ELECTRA [30]	33.00	-	-
Seq2Emo [20]	47.28	-	-
KEA-ELEC†	49.89(0.7)	48.68(1.3)	53.39(0.1)
KEA-BERT†	**50.53(0.9)**	**51.31(1.9)**	**52.70(2.1)**
KEA-MulBERT†	47.11(0.8)	48.72(0.9)	48.3(1.2)
BERT+CDP†	46.84(0.2)	49.10(0.8)	48.24(0.2)
BERT+MLM†	46.54(0.7)	50.15(0.7)	47.03(0.7)
BERT+C+M†	46.24(1.4)	49.14(1.2)	47.20(1.1)
MulBERT+CDP†	46.91(0.2)	47.71(1.4)	48.50(0.5)
MulBERT+MLM†	47.71(0.7)	48.80(0.4)	49.96(0.1)
MulBERT+C+M†	46.28(1.3)	48.55(1.1)	47.69(1.5)
REMTA (Zh)	**51.72(0.4)**	**51.31(1.6)**	**53.94(0.1)**
REMTA (De)	**51.32(0.6)**	**51.53(1.6)**	**53.26(0.4)**
REMTA (Fr)	**51.58(0.1)**	**51.58(1.1)**	**53.92(0.6)**
REMTA (Ru)	**50.81(0.8)**	**51.25(1.6)**	**53.22(0.7)**

† denotes the results obtained from our experiments. The values in parenthesis represent the standard deviation.

C. Model performance

Table I presents the results in terms of overall performance. As we can see, in terms of Macro-F1 score, when the source English texts are translated into Chinese, German, French, and Russian, REMTA reaches 51.72, 51.32, 51.58 and 50.81, respectively. And the best of other models reaches 50.53. That is to say, regardless of what languages the source English texts are translated into, REMTA consistently surpasses the second-best model (KEA-BERT), not to mention other models. Additionally, the performance of REMTA exceeds the original baseline (BERT) by even more than 5 scores on Macro-F1 score. And our work, unlike other approaches for emotion classification that use transfer learning or knowledge embedding, achieves the best performance.

To demonstrate the details of performance, we show the F1 score for each emotion in Table II, which can also be regarded as a case study. We can observe that our method performs competitively across a range of emotions, on par with other models. It should be clearly noted that REMTA (Zh) can still recognize even the least frequent emotions such as *grief*, *realization* and *relief* shown in Fig.2, with F1 scores

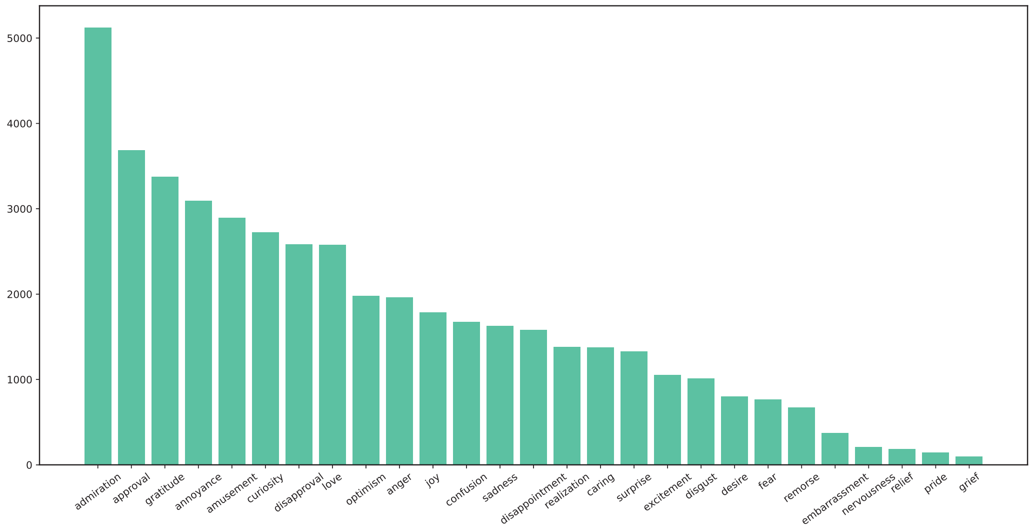

Fig. 2. The statistics of the dataset emotional labels. The neutral emotion labels are not included.

TABLE II
F1 SCORE OF REPRESENTATIVE MODELS ON EACH EMOTION

Emotion	KEA-BERT	BERT+CDP	KEA-ELEC	REMTA (Zh)
admiration	68.15(0.9)	68.65(0.1)	69.93(0.4)	66.54(1.1)
amusement	81.16(1.4)	82.45(0.4)	81.04(1.0)	81.70(0.3)
anger	49.80(0.8)	45.40(1.0)	49.12(0.9)	48.45(1.3)
annoyance	35.79(1.3)	37.63(1.0)	37.67(0.5)	36.98(0.8)
approval	37.82(2.1)	42.18(0.6)	38.96(0.8)	39.03(1.0)
caring	40.57(1.4)	41.40(1.6)	43.32(1.7)	38.38(0.5)
confusion	44.20(1.4)	44.15(0.2)	43.58(2.1)	43.95(1.8)
curiosity	56.57(2.2)	55.85(0.2)	55.05(1.7)	55.14(0.5)
desire	52.78(2.3)	52.94(2.8)	52.76(1.1)	47.25(1.9)
disappointment	31.76(1.4)	31.74(0.9)	31.87(1.4)	29.42(1.4)
disapproval	40.37(2.8)	38.11(0.1)	39.65(1.1)	40.67(0.3)
disgust	**46.95(4.1)**	**45.11(0.7)**	**46.86(1.5)**	**50.35(1.2)**
embarrassment	48.36(1.4)	50.81(3.9)	45.47(1.0)	42.36(3.4)
excitement	46.35(0.8)	43.61(0.3)	40.17(1.8)	42.65(1.6)
fear	68.73(4.3)	68.29(0.3)	65.77(1.3)	66.81(0.4)
gratitude	90.90(0.7)	91.6(0.9)	91.05(0.6)	91.57(0.7)
grief	**0.00(0.0)**	**0.00(0.0)**	**0.00(0.0)**	**54.81(7.3)**
joy	58.99(1.9)	59.95(0.9)	62.67(1.0)	59.72(2.7)
love	80.02(0.6)	80.83(0.7)	80.18(0.4)	79.25(0.3)
nervousness	41.18(2.8)	21.79(4.5)	43.12(2.7)	32.22(1.6)
optimism	55.28(3.0)	59.16(0.6)	54.38(1.6)	56.27(0.6)
pride	46.72(5.0)	0.00(0.0)	43.28(5.6)	43.98(5.1)
realization	21.61(1.2)	15.69(1.2)	25.19(4.4)	**26.00(1.5)**
relief	**28.15(20.0)**	**0.00(0.0)**	**13.65(10.9)**	**32.85(8.4)**
remorse	67.15(0.4)	65.34(0.2)	66.95(1.2)	68.2(1.2)
sadness	53.72(2.4)	51.9(1.0)	55.16(0.5)	51.94(1.0)
surprise	55.61(1.0)	53.97(1.2)	55.97(2.3)	54.62(1.1)
neutral	67.07(1.8)	67.79(0.5)	65.14(1.0)	66.98(0.7)

† denotes the results obtained from our experiments. The values in parenthesis represent the standard deviation.

of 54.81, 26.00 and 32.85, respectively. However, it is difficult for other models, and they either perform poorly or cannot distinguish these emotions at all, for example, the F1 scores of these model on *grief* are all zero. The reason for the baseline models that achieves 0 F1 score for the *grief* emotion category is that the number of this emotion is the least, causing the model to be unable to effectively recognize this emotion. And REMTA recognizes this emotion, the reason may be the *grief*'s colexification is significant [16], allowing other languages to provide additional information.

TABLE III
ABLATION STUDY OF REMTA

Variant	Macro F1	Precision	Recall
REMTA w Eh	51.19	51.89	53.31
REMTA w S-LSTM	51.21	50.51	53.29
REMTA w concatenate	21.85	31.08	21.14
REMTA w/o modules	50.04	57.39	49.53
REMTA (Zh)	**52.27**	**52.12**	**54.08**
REMTA (De)	**51.86**	**53.14**	**52.74**
REMTA (Fr)	**51.54**	**52.36**	**53.12**
REMTA (Ru)	**51.60**	**53.33**	**52.47**

And we think the reason for the above results could be that our architecture learns emotion information from the target language, which can benefit emotion recognition in the source language. And different target languages have different effects on the performance of the architecture, which we think may be related to the emotional information of the corresponding language. One example of the difficulty in translation is the German word *sehnsucht*, which expresses a strong desire for an alternative life [16], encompassing a wide range of emotions and experiences that lack an exact equivalent in English. In most cases, *sehnsucht* expresses a state of sadness, and sometimes it can be used to describe the increasing joy of something before it happens. Therefore, it may provide

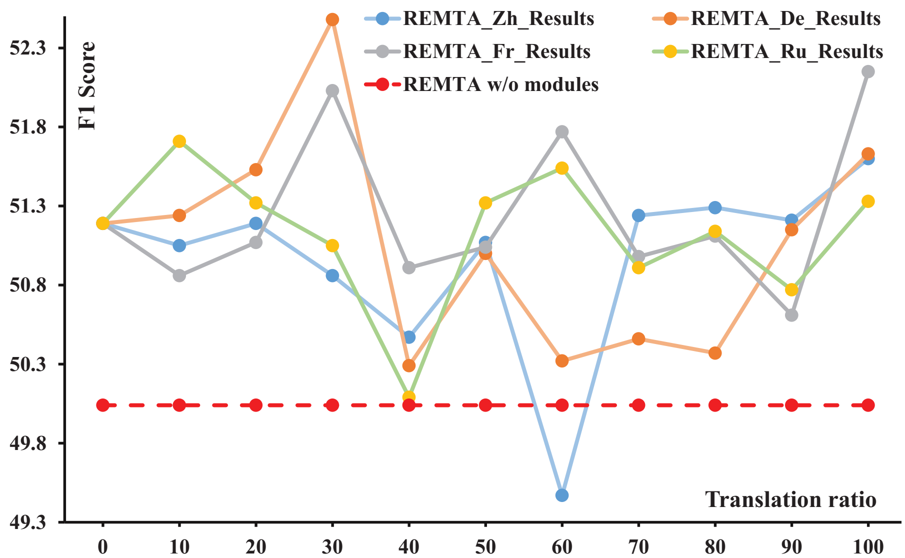

Fig. 3. The effect of translation ratio on the model.

more emotion information than English words with similar meanings, such as *longing*, *desire* and *yearning*.

D. Analysis of the model

In order to further demonstrate the effectiveness of our architecture and the role of modules in REMTA, we carry out ablation experiments. Taking the case of translation into four languages and one round of experimental results as examples, we compare REMTA with its variants under the same parameter circumstances:

REMTA w En: we replace translated texts with the original English texts, retaining the stacked LSTM and attention mechanism.

REMTA w S-LSTM: we remove the attention mechanism and the translated texts, keeping only the stacked LSTM module as well as the original English texts.

REMTA w/o modules: we remove translated texts, stacked LSTM as well as the attention mechanism.

REMTA w concatenate: we remove the attention mechanism from REMTA but keep the translated texts and implement the fusion of the two language representations in a concatenate manner.

The results are shown in Table III. When we replace translated texts with the original English texts, that is to say, there is no additional information from other language, the performance of REMTA will be degraded. If we remove the attention module and S-LSTM module as well as translated texts, F1 score of REMTA show a significant decrease of approximately 2.23 for Chinese, 1.84 for German, 1.5 for French and 1.56 for Russian, respectively. From these results, we think that the absence of another language will lead to poor model performance, thus indicating that another language can bring useful information for emotion recognition. In addition, our architecture can realize information exchange between two languages, which is also indispensable.

E. Impact of translation ratio

Translation is an important part of our proposed framework. Due to the limitation of not being able to accurately determine the quality of the translated texts, so we observe the impact of the number of the translated language on REMTA through the ratio of the translated texts. We randomly select a certain percentage of translated texts and put them into REMTA to see how REMTA performs. The results are shown in Fig.3.

The red dashed line represents the performance of *REMTA w/o modules*, and the performance variance for different languages with various translation ratios is shown by the solid lines, accordingly. We can see that the performance improves significantly when another language is added. Furthermore, we should note that performance does not always get better when more original texts are translated. For example, when the percentage of translations in these languages is 40%, the performance of the model decreases in all cases. When the percentage of Chinese translation is 60%, the performance of the model is not as good as when no Chinese is added. The reason for this phenomenon could be that when the translated texts may be confused or have no corresponding words, these noises from machine translation have an impact on the results. Therefore, we should ensure the quality of the translated texts as much as possible when using a multi-lingual method to recognize emotions.

V. Conclusion

Although emotion recognition in social texts has attracted many researchers, the relation of emotional semantics between different languages are ignored. In this paper, we present a multilingual architecture with machine translation and attention mechanism to recognize fine-grained emotions (REMTA) in social network texts, which performs better than other baselines, and achieves a significant improvement in F1 score. The results prove the effectiveness of REMTA, and our architecture can integrate information from different languages. In the future, we will investigate how a greater variety of languages affect the emotion recognition.

Acknowledgements

This work was supported by by the National Key R&D Program of China (Grant No. 2022YFC3302104) and the National Key Natural Science Foundation Project of China under Grant 62036001.

References

[1] E. Cambria, Q. Liu, S. Decherchi, F. Xing, and K. Kwok, "SenticNet 7: A commonsense-based neurosymbolic AI framework for explainable sentiment analysis," in *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, Jun. 2022, pp. 3829–3839.

[2] W. Li, L. Zhu, R. Mao, and E. Cambria, "Skier: A symbolic knowledge integrated model for conversational emotion recognition," in *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*, ser. AAAI'23/IAAI'23/EAAI'23. AAAI Press, 2023.

[3] R. Mao, Q. Liu, K. He, W. Li, and E. Cambria, "The biases of pre-trained language models: An empirical study on prompt-based sentiment analysis and emotion detection," *IEEE Transactions on Affective Computing*, vol. 14, no. 3, pp. 1743–1753, 2023.

[4] B. Liu, *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge university press, 2020.

[5] E. Cambria, "Affective computing and sentiment analysis," *IEEE Intelligent Systems*, vol. 31, no. 2, pp. 102–107, 2016.

[6] M. Munezero, C. S. Montero, E. Sutinen, and J. Pajunen, "Are they different? affect, feeling, emotion, sentiment, and opinion detection in text," *IEEE Transactions on Affective Computing*, vol. 5, no. 2, pp. 101–111, 2014.

[7] C. Strapparava and R. Mihalcea, "Semeval-2007 task 14: Affective text," in *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, 2007, pp. 70–74.

[8] S. Poria, D. Hazarika, N. Majumder, G. Naik, E. Cambria, and R. Mihalcea, "MELD: A multimodal multi-party dataset for emotion recognition in conversations," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, Jul. 2019, pp. 527–536.

[9] Y. Li, H. Su, X. Shen, W. Li, Z. Cao, and S. Niu, "DailyDialog: A manually labelled multi-turn dialogue dataset," in *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Taipei, Taiwan: Asian Federation of Natural Language Processing, Nov. 2017, pp. 986–995.

[10] P. Ekman, "Are there basic emotions?" 1992.

[11] R. Plutchik, "A general psychoevolutionary theory of emotion," in *Theories of emotion*. Elsevier, 1980, pp. 3–33.

[12] M. R. Leary, "Motivational and emotional aspects of the self," *Annual review of psychology*, vol. 58, p. 317, 2007.

[13] H. Alhuzali and S. Ananiadou, "SpanEmo: Casting multi-label emotion classification as span-prediction," in *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. Online: Association for Computational Linguistics, Apr. 2021, pp. 1573–1584.

[14] K. Becker, V. P. Moreira, and A. G. dos Santos, "Multilingual emotion classification using supervised learning: Comparative experiments," *Information Processing & Management*, vol. 53, no. 3, pp. 684–704, 2017.

[15] J. Deng and F. Ren, "Multi-label emotion detection via emotion-specified feature extraction and emotion correlation learning," *IEEE Transactions on Affective Computing*, 2020.

[16] J. C. Jackson, J. Watts, T. R. Henry, J.-m. List, R. Forkel, P. J. Mucha, S. J. Greenhill, R. D. Gray, and K. A. Lindquist, "Emotion semantics show both cultural variation anf universal structure," *Science*, vol. 1522, no. December, pp. 1517–1522, 2019.

[17] L. F. Barrett, "The theory of constructed emotion: an active inference account of interoception and categorization," *Social cognitive and affective neuroscience*, vol. 12, no. 1, pp. 1–23, 2017.

[18] M. Jabreel and A. Moreno, "A deep learning-based approach for multi-label emotion classification in tweets," *Applied Sciences*, vol. 9, no. 6, p. 1123, 2019.

[19] G. Singh, D. Brahma, P. Rai, and A. Modi, "Fine-grained emotion prediction by modeling emotion definitions," in *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 2021, pp. 1–8.

[20] C. Huang, A. Trabelsi, X. Qin, N. Farruque, L. Mou, and O. R. Zaiane, "Seq2emo: a sequence to multi-label emotion classification model," in *Proceedings of the 2021 conference of the North American chapter of the association for computational linguistics: human language technologies*, 2021, pp. 4717–4724.

[21] V. Suresh and D. C. Ong, "Using knowledge-embedded attention to augment pre-trained language models for fine-grained emotion recognition," in *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 2021, pp. 1–8.

[22] C. Baziotis, A. Nikolaos, A. Chronopoulou, A. Kolovou, G. Paraskevopoulos, N. Ellinas, S. Narayanan, and A. Potamianos, "NTUA-SLP at SemEval-2018 task 1: Predicting affective content in tweets with deep attentive RNNs and transfer learning," in *Proceedings of the 12th International Workshop on Semantic Evaluation*. New Orleans, Louisiana: Association for Computational Linguistics, Jun. 2018, pp. 245–255.

[23] J. Yu, L. Marujo, J. Jiang, P. Karuturi, and W. Brendel, "Improving multi-label emotion classification via sentiment classification with dual attention transfer network," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Brussels, Belgium: Association for Computational Linguistics, Oct.-Nov. 2018, pp. 1097–1102.

[24] Z. Ahmad, R. Jindal, A. Ekbal, and P. Bhattachharyya, "Borrow from rich cousin: transfer learning for emotion detection using cross lingual embedding," *Expert Systems with Applications*, vol. 139, p. 112851, 2020.

[25] K. A. Lindquist, J. C. Jackson, J. Leshin, A. B. Satpute, and M. Gendron, "The cultural evolution of emotion," *Nature Reviews Psychology*, vol. 1, no. 11, pp. 669–681, 2022.

[26] S. Hayakawa, A. Costa, A. Foucart, and B. Keysar, "Using a Foreign Language Changes Our Choices," *Trends in Cognitive Sciences*, vol. 20, no. 11, pp. 791–793, nov 2016.

[27] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[28] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[29] D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, and S. Ravi, "GoEmotions: A dataset of fine-grained emotions," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics, Jul. 2020, pp. 4040–4054.

[30] D. Cortiz, "Exploring transformers models for emotion recognition: A comparision of bert, distilbert, roberta, xlnet and electra," in *Proceedings of the 2022 3rd International Conference on Control, Robotics and Intelligent System*, ser. CCRIS '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 230–234.