INTERSPEECH 2023
20-24 August 2023, Dublin, Ireland

Mix before Align: Towards Zero-shot Cross-lingual Sentiment Analysis via Soft-mix and Multi-view Learning

Zhihong Zhu[*], *Xuxin Cheng*[*], *Dongsheng Chen,*
Zhiqi Huang, Hongxiang Li, and Yuexian Zou[†]

School of ECE, Peking University, China
{zhihongzhu, chengxx, chends, lihongxiang}@stu.pku.edu.cn
{zhiqihuang, zouyx}@pku.edu.cn

Abstract

Due to the insufficient sentiment corpus in many languages, recent studies have proposed cross-lingual sentiment analysis to adapt sentiment analysis models from *rich-resource* languages to *low-resource* ones. However, existing models heavily rely on *code-switched* sentences to reduce the alignment discrepancy of cross-lingual embeddings, which could be limited by their inherent constraints. In this paper, we propose a novel method SOUL (short for Soft-mix and Multi-view learning) to enhance zero-shot cross-lingual sentiment analysis. Instead of using the embeddings of *code-switched* sentences directly, SOUL first mixes them softly with the embeddings of *original* sentences. Furthermore, SOUL utilizes multi-view learning to encourage contextualized embeddings to align into a refined language-invariant space. Experimental results on four cross-lingual benchmarks across five languages clearly verify the effectiveness of our proposed SOUL.

Index Terms: multi-lingual NLP, zero-shot cross-lingual sentiment analysis, soft-mix, multi-view learning.

1. Introduction

Recently, sentiment analysis methods demonstrate strong performance through the fine-tuning of large-scale pre-trained language models [1, 2, 3, 4]. However, their success largely depends on the presence of manual sentiment annotations. Therefore, such methods may not accurately recognize sentiment polarities in low-resource languages lacking annotated data.

To this end, recent studies have paid attention to cross-lingual sentiment analysis [5, 6, 7]. Multilingual language models such as `mBERT` [8], and `XLM-R` [9] have been prevalent, which are pre-trained under multilingual corpus and then fine-tuned by source language sentiment supervision. Despite their success, their performances are still unsatisfactory due to the large language discrepancy (*cf.* Figure 1).

Code-switched sentences are generated by substituting randomly selected words in the source languages with their counterparts in the target languages, which is generally adopted to improve model performance in cross-lingual transfer [10, 11, 12]. Some works have addressed the effectiveness of *code-switching* in improving the performance of multilingual models on zero-shot cross-lingual tasks [13, 14]. Though *code-switching* has shown great potential and strong generalization ability on the semantic representation, we discover two main issues remain: (1) A word embedding may incorporate language-specific cues whose loss is possible when substituting it with a word embedding from another language. (2) They solely rely on code-switched sentences to fine-tune multilingual language models, while completely ignoring the relationship between the *original* and *code-switched* sentences, which could potentially result in the loss of interactive information and impede the alignment of contextualized embeddings.

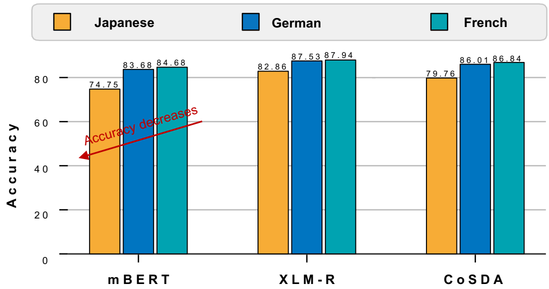

Figure 1: *`mBERT` [8], `XLMR` [9] based zero-shot and code-switching (`CoSDA` [13]) results on the Amazon Reviews dataset [15] (Average accuracy reported).*

To solve the aforementioned issues, we propose a novel method via Soft-mix and Multi-View learning (SOUL) for zero-shot cross-lingual sentiment analysis. For the first issue, we propose to softly mix [16, 17] source and target word embeddings in order to enable the model to leverage language-specific information from each language. In this manner, it is possible to alleviate the contextual space inconsistency of the *code-switched* sentences. To the best of our knowledge, we are the first to employ soft mixing for cross-lingual sentiment analysis.

For the second issue, we aim to obtain comprehensive cross-lingual information from diverse perspectives and explore the consistency of multiple views using multi-view learning [18, 19, 20, 21]. More specifically, we leverage the competitive multilingual language model `XLM-R` [9] to construct two views: the encoded feature representation of the *original* sentence and the corresponding *code-switched* one. Our insight is that the key of cross-lingual transfer is to learn a language-invariant feature space [22, 23, 24]. Therefore, it is necessary for the two feature representations to exhibit as much similarity as possible. We employ multi-view learning to enforce a consensus between the two views, which encourages similar words in different languages to align into a shared latent space.

The contributions of our work are three-fold: (1) We propose soft mixing for zero-shot cross-lingual sentiment analysis to transfer language-specific information and alleviate the contextual space inconsistency of the *code-switched* sentences. (2) We leverage multi-view learning to enforce consistent representation of the *original* sentence and *code-switched* sentence. By doing so, we can build a refined, language-invariant space that is more robust to language shifts. (3) SOUL is extensively evaluated on four benchmarks across nine languages for zero-

[*]Equal contribution.
[†]Corresponding author.

10.21437/Interspeech.2023-69

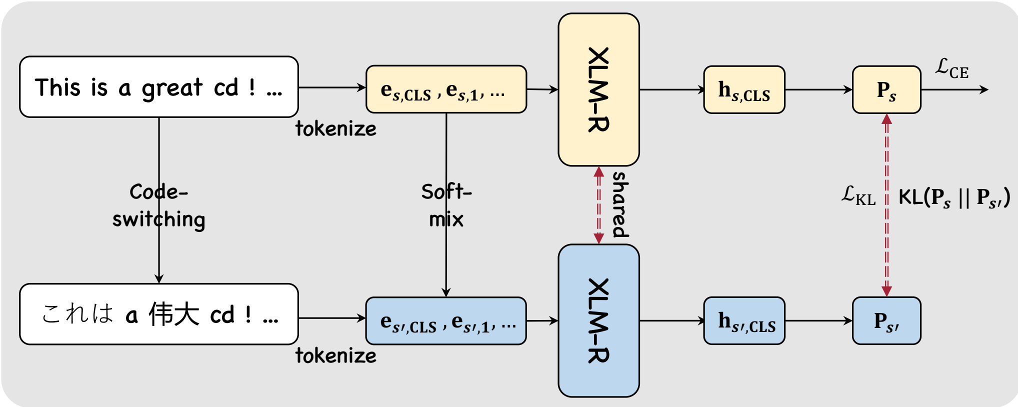

Figure 2: *The main architecture of* SOUL*, which consists of (1) soft-mix code-switching: softly mix the code-switched sentence with the original one to alleviate the contextual space inconsistency of the code-switched sentence; (2) multi-view learning: training the multilingual language model (i.e.,* `XLM-R`*) based on multi-view learning.*

shot cross-lingual sentiment analysis and achieves state-of-the-art results, which demonstrates the effectiveness of SOUL.

2. Method

The architecture of SOUL is shown in Figure 2, which consists of two main components: (1) Soft-Mix *Code-Switching*; (2) Multi-View Learning. We elaborate each part in this section.

2.1. Soft-Mix Code-Switching

Given a sentence $\mathbf{x} = [x_{s,1}, \ldots, x_{s,n}]$ in a source language s (*e.g.*, English), we follow [13, 25] to use bilingual dictionaries [26] to generate *code-switched* sentence s'. More specifically, we select a subset of words randomly to translate T to obtain s', where $x_{s,i} \in T$ are switched to $x_{t,i}$, which denotes a translated counterpart of $x_{s,i}$ in the target language t. Since multiple target languages are available, we randomly choose one of them and skip the substitution if there exists no corresponding word in the bilingual dictionary. For both sentences, we first prepended `[CLS]` and appended `[SEP]`, in order to match the input of `XLM-R` [9]. To note, $\mathbf{e}_s$ and $\mathbf{e}_{s'}$ are input hidden state vectors of the *original* and *code-switched* sentence, respectively.

Instead of using $\mathbf{e}_{s'} = [\mathbf{e}_{s',\texttt{CLS}}, \mathbf{e}_{s',1}, \ldots, \mathbf{e}_{s',n}, \mathbf{e}_{s',\texttt{SEP}}]$ directly as an input of `XLM-R`, we mix the hidden states of two sentences as follows:

$$\mathbf{e}_{s'} = \lambda_{s'}\mathbf{e}_{s'} + (1-\lambda_{s'})\mathbf{e}_s, \quad (0 \le \lambda_{s'} \le 1), \tag{1}$$

where $\lambda_{s'}$ is a trade-off hyperparameter. Note that if $\lambda_{s'}$ is 1 only the *code-switched* sentence is used, and $\lambda_{s'}$ is 0 only the *original* sentence is used. In each batch, we sample $\lambda_{s'}$ from a beta distribution $\beta(\alpha, \alpha)$ following [16].

Remark 1 *Since many language models employ subword (e.g., WordPiece) for tokenization, there may be a difference in the length of source and target word embeddings. we use weighted sum of n subword embeddings in the source language to normalize the imbalance of the two embeddings.*

2.2. Multi-View Learning

After obtaining $\mathbf{e}_s$ and $\mathbf{e}_{s'}$, we first feed them into a shared `XLM-R` model separately:

$$\begin{aligned} \mathbf{h}_s &= \mathrm{Encoder}(\mathbf{e}_s), \\ \mathbf{h}_{s'} &= \mathrm{Encoder}(\mathbf{e}_{s'}), \end{aligned} \tag{2}$$

where $\mathbf{h}_s$ and $\mathbf{h}_{s'}$ are the encoded feature representation for the *original* and *code-switched* sentence, respectively.

For the sentiment analysis tasks, `XLM-R` takes $\mathbf{h}_{s,\texttt{CLS}}$ and $\mathbf{h}_{s',\texttt{CLS}}$ into a classification layer:

$$\begin{aligned} \mathbf{p}_s &= \mathrm{softmax}(\mathbf{W}\mathbf{h}_{s,\texttt{CLS}} + \mathbf{b}), \\ \mathbf{p}_{s'} &= \mathrm{softmax}(\mathbf{W}\mathbf{h}_{s',\texttt{CLS}} + \mathbf{b}), \end{aligned} \tag{3}$$

where $\mathbf{W}$ and $\mathbf{b}$ are learnable parameters.

Our main learning objective is to train the classifier to match the predicted labels with the ground truth ones, which can be achieved by minimizing the cross-entropy loss function between the predicted probability distribution $\mathbf{p}_s$ and the ground truth label distribution $\mathbf{p}$:

$$\mathcal{L}_{\mathrm{CE}} = \mathrm{CE}(\mathbf{p}_s, \mathbf{p}). \tag{4}$$

On the other hand, previous works employ only *code-switched* sentences, neglecting the rich information associations between the *original* and *code-switched* one. In contrast, we leverage multi-view learning to obtain a comprehensive representation from multiple views which generally contain complementary information. Concretely, we take two views into consideration: (1) the *original* sentence feature representation $\mathbf{h}_s$; (2) the *code-switched* sentence feature representation $\mathbf{h}_{s'}$. We enforce a consensus between the two views, where the predicted distributions of both views should be as similar as possible:

$$\mathcal{L}_{\mathrm{KL}} = \mathrm{KL}(\mathbf{p}_s || \mathbf{p}_{s'}), \tag{5}$$

where KL denotes *Kullback-Leibler* (KL) divergence to measure the difference between two distributions.

Thus, the final objective, combing the cross-entropy loss (Eq. 4) and the KL divergence loss (Eq. 5) is written as follows:

$$\mathcal{L} = \lambda_{\mathrm{CE}}\mathcal{L}_{\mathrm{CE}} + \lambda_{\mathrm{KL}}\mathcal{L}_{\mathrm{KL}}, \tag{6}$$

in which λ_{CE} and λ_{KL} are trade-off hyper-parameters.

3. Experiments

3.1. Datasets

As shown in Table 2, we employ data in six languages (*i.e.*, English, German, French, Japanese, Chinese and Arabic) from:

Amazon Review [15]. This dataset comprises four languages allotted for a binary sentiment classification task, wherein each language encompasses three domains (*i.e.*, Books, DVD, and

Table 1: *Prediction accuracy of binary classification in the test set for three language pairs. The highest performance is in* ***bold****, while the highest performance within the method group is* underlined*. '–' denotes missing results from the published work.*

Approach	German (2)				French (2)				Japanese (2)			
	Books	DVD	Music	Avg	Books	DVD	Music	Avg	Books	DVD	Music	Avg
w/ explicit cross-lingual resources												
LR+MT	79.68	77.92	77.22	78.27	80.76	78.83	75.78	78.46	70.22	71.30	72.02	71.18
CR-RL	79.89	77.14	77.27	78.10	78.25	74.83	78.71	77.26	71.11	73.12	74.38	72.87
Bi-PV	79.51	78.60	82.45	80.19	84.25	79.60	80.09	81.31	71.75	75.40	75.45	74.20
CLDFA	83.95	83.14	79.02	82.04	83.37	82.56	83.31	83.08	77.36	80.52	76.46	78.11
w/ implicit cross-lingual resources												
UMM	81.65	81.27	81.32	81.41	80.27	80.27	79.41	79.98	71.23	72.55	75.38	73.05
PBLM	78.65	79.90	80.10	79.50	77.90	75.65	75.95	76.50	-	-	-	-
w/o cross-lingual supervision												
mBERT	84.35	82.85	83.85	83.68	84.55	85.85	83.65	84.68	73.35	74.80	76.10	74.75
XLM	86.85	84.20	85.90	85.65	88.10	86.95	86.20	87.08	80.95	79.20	78.02	79.39
XLM-R	88.10	86.54	87.95	87.53	88.55	88.30	87.37	87.94	81.15	**83.93**	83.50	82.86
CoSDA	88.25	86.05	86.10	86.01	86.22	87.60	86.70	86.84	80.65	77.73	80.90	79.76
MVEC	88.41	87.32	89.97	88.61	89.08	88.28	88.50	88.62	79.15	77.15	79.70	78.67
ELSA	86.40	86.10	87.80	86.77	86.00	85.70	86.00	85.90	78.30	79.10	80.80	79.40
SOUL (Ours)	**91.97**	**90.15**	**91.08**	**91.07**	**91.67**	**91.04**	**89.17**	**90.62**	**83.32**	83.48	**85.30**	**84.03**

Table 2: *Statistics for datasets used in our experiments. Note that as Amazon Review is a multilingual dataset, we show the number of samples for each task in each language.*

Dataset	Language	#Classes	#Samples
Amazon Review	Multi	2	4, 000
Yelp	English	5	700, 000
Hotel Review	Chinese	5	20, 000
Social Media Posts	Arabic	3	1, 000

Music). Each cross-lingual task comprises 2,000 samples assigned for train and 2,000 for test.

Yelp [27]. It contains 700K reviews from 5 classes. We keep the original class tags for the English-Chinese pair while converting them into 3 sentimental levels (*i.e.*, positive, negative, and neutral) for the English-Arabic pair.

Hotel Review [28]. It contains 170k Chinese hotel reviews from 5 classes as in the Yelp dataset and is used for the English-Chinese pair in the experiments. Following [5], we randomly sample 20K reviews for test.

Social Media Posts [29]. This is an Arabic sentiment dataset with 3 sentimental labels for the English-Arabic pair (*i.e.*, positive, neutral and negative). We randomly sample 1000 sentences and use them for test.

3.2. Model Zoo

We compare our proposed SOUL with a group of baseline methods under different categories: (1) methods with explicit cross-lingual resources (*i.e.*, LR+MT, CR-RL [30], Bi-PV [31], and CLDFA [32]); (2) methods with implicit cross-lingual supervision (*i.e.*, UMM [33], PBLM [34], DAN [35], mSDA [36], and ADAN [35]); (3) methods with no cross-lingual supervision (*i.e.*, mBERT [8], XLM [37], XLM-R [9], CoSDA [13], MVEC [5], and ELSA [38]). It is worth noting that the proposed SOUL belongs to the third category. In addition, the compared baselines are validated on different benchmarks. Hence, we use different baselines in Table 1 and Table 3 for fairness.

Table 3: *Prediction accuracy of 5-class and 3-class classification on the test set. The highest performance is in* ***bold****.*

Approach	Chinese (5)	Arabic (3)
LR+MT	34.01	51.67
DAN	29.11	48.00
mSDA	31.44	48.33
ADAN	42.49	**52.54**
mBERT	38.85	50.40
XLM-R	46.60	51.16
MVEC	43.36	49.70
SOUL (Ours)	**48.75**	52.03

3.3. Experimental Setting

We leverage the $\text{XLM-R}_{\text{base}}$ [1] as Encoder in Eq.2. We select the best hyper-parameters by searching a combination of batch size, and learning rate with the following ranges: learning rate $\{1\times10^{-6}, 3\times10^{-6}, 5\times10^{-6}, 7\times10^{-6}\}$; batch size $\{4, 8, 16, 32\}$; mix weights $\lambda_{s'}$ $\{0.0, 0.1, \cdots, 1.0\}$. λ_{CE} and λ_{KL} in Eq.6 are set to 0.8 and 1, respectively. When we apply SOUL in training, we set α in beta distribution to be 0.75. Our approach is implemented with PyTorch [2] and all experiments are conducted on a single NVIDIA Tesla A100. All experiment results are the average score over 5 runs with random seeds.

3.4. Main Results

The classification results on Amazon Review Dataset, English-Chinese, and English-Arabic pairs are shown in Table 1 and Table 3, respectively (where (*) indicates the number of sentiment polarity). From the results, we can conclude: As for language model pre-trained approaches, *i.e.*, mBERT and XLM-R achieve impressive performances, which shows the effectiveness of multilingual language models in zero-shot cross-lingual

[1]https://huggingface.co/xlm-roberta-base
[2]https://pytorch.org

Table 4: *Ablation study for the contribution of each design over the Amazon Review dataset. "SO" and "UL" are short for soft-mix code-switching and multi-view learning, respectively.*

Variant	German (2)				French (2)				Japanese (2)			
	Books	DVD	Music	Avg	Books	DVD	Music	Avg	Books	DVD	Music	Avg
SOUL	**91.97**	**90.15**	**91.08**	**91.07**	**91.67**	**91.04**	**89.17**	**90.62**	**83.32**	**83.48**	**85.30**	**84.03**
w/o SO	91.41	89.53	88.42	88.61	91.21	90.14	86.69	89.34	82.66	82.80	85.13	83.53
w/o UL	88.87	87.98	86.07	87.64	89.24	88.89	84.15	85.90	81.64	82.05	82.68	82.12

Table 5: *Prediction accuracy over the Amazon Review dataset in different alignment strategies. The performances are the average results for the three domains, Books, DVD, and Music.*

Variant	German (2)	French (2)	Japanese (2)
KL	**91.07**	**90.62**	**84.03**
SBA	88.56	89.21	82.05
DBA	88.63	88.07	82.40

Table 6: *Prediction accuracy of 5-class and 3-class classification using different views for multi-view learning.*

Approach	Chinese (5)	Arabic (3)
ORG + CS (SOUL)	48.75	52.03
ORG + TRANS	44.51	48.95
ORG + CS + TRANS	**50.02**	**52.98**

transfer. Compared with strong baselines, SOUL leads to significant performance and achieves consistent improvements on 9/11 tasks by a clear margin. Particularly, SOUL significantly improves binary sentiment classification 2.46%, 2.00%, and 4.28% in *de*, *fr* and *jp*, respectively. As for multi-class sentiment classification, SOUL obtains slightly better accuracy in *ch*. We attribute this to the fact that different languages have distinct linguistic structures. During *code-switching*, substituting words *word-for-word* can interrupt this structure. Therefore, sentence representations cannot be directly mapped to the same space. Moreover, training with both the *original* and *code-switched* sentence enables better exploitation of their relationship.

3.5. Further Analysis

Ablation Study. As shown in Table 4, we conduct ablation studies to evaluate the contributions of each component in SOUL. The *first* line represents the model trained with all proposed components. The *next two* lines represent the models without *soft-mix code-switching* and without *multi-view learning*, respectively. It is clear that all the key parts of SOUL generally make good contributions to promote cross-lingual sentiment knowledge transfer. Specifically, from the *second* line, *soft-mix code-switching* can alleviate inconsistencies in the context space of *code-switched* sentences, resulting in superior performance compared to random *hard* substitution. For the *third* line, we can see that the performance of SOUL *w/o* UL is dramatically degraded compared to SOUL *w/* UL. This indicates that *multi-view learning* is capable of exploiting the rich associations between *original* and *code-switched* sentences, thereby enhancing the performance of cross-lingual sentiment analysis.

Effects of Alignment Strategy. We explore two other strategies to align the embeddings of *original* and *code-switched* sentences: (1) Distance-Based Alignment. (2) Similarity-Based Alignment. That is to say, the *KL divergence* loss is replaced by the *Euclidean Distance* and *Cosine Similarity* loss, respectively. From Table 5, it can be inferred that the employment of minimizing the *KL divergence* between two probability distributions for multi-view learning yields better performance compared to the other two alignment strategies. This is because enforcing identical encoded features can impede the model's representation ability owing to the diverse semantic structures and translation biases discernible in different languages. Conversely, the leveraging of *KL divergence* for multi-view learning engenders two optimal predictions that are proximal to each other, which bestows upon the model a more pliable methodology to acquire language-invariant representations.

Effects of Learning Views. We further explore the potential of SOUL for cross-lingual sentiment analysis by changing the view. Concretely, we add the third view called TRANS, which is the translation of the *original* sentences by a Machine Translation system [3] trained on Europarl [4] corpus. From the results in Table 6, we find that ORG+CS+TRANS further improves the performance of SOUL by performing an additional view. By doing so, the model could learn more robust cross-lingual representations from these complementary views.

4. Conclusion

In this paper, we propose a novel method via **So**ft-mix and M**ul**ti-View learning (SOUL) for zero-shot cross-lingual sentiment analysis. Unlike existing *code-switching* sentences mixed at the word level, SOUL first softly mixes them with the embeddings of the *original* sentence. Based on this, we further leverage multi-view learning to construct a language-invariant feature space. Experiments show that SOUL significantly outperforms previous methods for most cross-lingual sentiment analysis tasks in the zero-shot setting.

5. Acknowledgements

We thank all anonymous reviewers for their constructive comments. This paper was partially supported by Shenzhen Science & Technology Research Program (No:GXWD20201231165807007-20200814115301001) and NSFC (No: 62176008).

6. References

[1] S. Lin, W. Su, P. Chien, M. Tsai, and C. Wang, "Self-attentive sentimental sentence embedding for sentiment analysis," in *ICASSP*, 2020.

[3] https://github.com/facebookresearch/fairseq

[4] https://statmt.org/europarl/

[2] C. Potts, Z. Wu, A. Geiger, and D. Kiela, "Dynasent: A dynamic benchmark for sentiment analysis," in *ACL*, 2021.

[3] W. Yang, S. Fukayama, P. Heracleous, and J. Ogata, "Exploiting fine-tuning of self-supervised learning models for improving bi-modal sentiment analysis and emotion recognition," in *INTERSPEECH*, 2022.

[4] Z. Zhu, W. Xu, X. Cheng, T. Song, and Y. Zou, "A dynamic graph interactive framework with label-semantic injection for spoken language understanding," in *ICASSP*, 2023.

[5] H. Fei and P. Li, "Cross-lingual unsupervised sentiment classification with multi-view transfer learning," in *ACL*, 2020.

[6] W. Zhang, R. He, H. Peng, L. Bing, and W. Lam, "Cross-lingual aspect-based sentiment analysis with aspect term code-switching," in *EMNLP*, 2021.

[7] J. Zhang, T. Liang, M. Wan, G. Yang, and F. Lv, "Curriculum knowledge distillation for emoji-supervised cross-lingual sentiment analysis," in *EMNLP*, 2022.

[8] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," in *NAACL*, 2019.

[9] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov, "Unsupervised cross-lingual representation learning at scale," in *ACL*, 2020.

[10] X. Zhou, X. Tian, G. Lee, R. K. Das, and H. Li, "End-to-end code-switching TTS with cross-lingual language model," in *ICASSP*, 2020.

[11] S. Zhao, T. H. Nguyen, H. Wang, and B. Ma, "Towards natural bilingual and code-switched speech synthesis based on mix of monolingual recordings and cross-lingual voice conversion," in *INTERSPEECH*, 2020.

[12] C. Whitehouse, F. Christopoulou, and I. Iacobacci, "Entitycs: Improving zero-shot cross-lingual transfer with entity-centric code switching," in *EMNLP Findings*, 2022.

[13] L. Qin, M. Ni, Y. Zhang, and W. Che, "Cosda-ml: Multi-lingual code-switching data augmentation for zero-shot cross-lingual NLP," in *IJCAI*, 2020.

[14] Y. Feng, F. Li, and P. Koehn, "Toward the limitation of code-switching in cross-lingual transfer," in *EMNLP*, 2022.

[15] P. Prettenhofer and B. Stein, "Cross-language text classification using structural correspondence learning," in *ACL*, 2010.

[16] J. Chen, Z. Yang, and D. Yang, "Mixtext: Linguistically-informed interpolation of hidden space for semi-supervised text classification," in *ACL*, 2020.

[17] D. Lee, J. Lee, G. Lee, B. Chun, and S. Hwang, "SCOPA: soft code-switching and pairwise alignment for zero-shot cross-lingual transfer," in *CIKM*, 2021.

[18] C. Yang, Z. An, and Y. Xu, "Multi-view contrastive learning for online knowledge distillation," in *ICASSP*, 2021.

[19] Y. Wang and Y. Lan, "Multi-view learning based on non-redundant fusion for icu patient mortality prediction," in *ICASSP*, 2022.

[20] X. Cheng, Z. Zhu, H. Li, Y. Li, and Y. Zou, "SSVMR: Saliency-based self-training for video-music retrieval," in *ICASSP*, 2023.

[21] X. Cheng, B. Cao, Q. Ye, Z. Zhu, H. Li, and Y. Zou, "ML-LMCL: Mutual learning and large-margin contrastive learning for improving asr robustness in spoken language understanding," in *ACL*, 2023.

[22] J. M. Dalhouse and K. Itou, "Cross-lingual transfer learning approach to phoneme error detection via latent phonetic representation," in *INTERSPEECH*, 2022.

[23] Z. Zhu, X. Cheng, Z. Huang, D. Chen, and Y. Zou, "Towards unified spoken language understanding decoding via label-aware compact linguistics representations," in *ACL*, 2023.

[24] P. Guo, S. Huang, P. Jiang, Y. Sun, M. Zhang, and M. Zhang, "Curriculum-style fine-grained adaption for unsupervised cross-lingual dependency transfer," *IEEE ACM Trans. Audio Speech Lang. Process.*, 2023.

[25] X. Cheng, W. Xu, Z. Yao, Z. Zhu, Y. Li, H. Li, and Y. Zou, "FC-MTLF: A fine- and coarse-grained multi-task learning framework for cross-lingual spoken language understanding," in *INTERSPEECH*, 2023.

[26] G. Lample, A. Conneau, M. Ranzato, L. Denoyer, and H. Jégou, "Word translation without parallel data," in *ICLR*, 2018.

[27] X. Zhang, J. J. Zhao, and Y. LeCun, "Character-level convolutional networks for text classification," in *NeurIPS*, 2015.

[28] Y. Lin, H. Lei, J. Wu, and X. Li, "An empirical study on sentiment classification of chinese review using word embedding," in *Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation, PACLIC 29, Shanghai, China, October 30 - November 1, 2015*, 2015.

[29] S. M. Mohammad, M. Salameh, and S. Kiritchenko, "How translation alters sentiment," *J. Artif. Intell. Res.*, 2016.

[30] M. Xiao and Y. Guo, "Semi-supervised representation learning for cross-lingual text classification," in *EMNLP*, 2013.

[31] H. Pham, T. Luong, and C. D. Manning, "Learning distributed representations for multilingual text sequences," in *AACL*, 2015.

[32] R. Xu and Y. Yang, "Cross-lingual distillation for text classification," in *ACL*, 2017.

[33] K. Xu and X. Wan, "Towards a universal sentiment classifier in multiple languages," in *EMNLP*, 2017.

[34] Y. Ziser and R. Reichart, "Deep pivot-based modeling for cross-language cross-domain transfer with minimal guidance," in *EMNLP*, 2018.

[35] X. Chen, Y. Sun, B. Athiwaratkun, C. Cardie, and K. Q. Weinberger, "Adversarial deep averaging networks for cross-lingual sentiment classification," *Trans. Assoc. Comput. Linguistics*, 2018.

[36] M. Chen, Z. E. Xu, K. Q. Weinberger, and F. Sha, "Marginalized denoising autoencoders for domain adaptation," in *ICML*, 2012.

[37] A. Conneau and G. Lample, "Cross-lingual language model pretraining," in *NeurIPS*, 2019.

[38] Z. Chen, S. Shen, Z. Hu, X. Lu, Q. Mei, and X. Liu, "Emoji-powered representation learning for cross-lingual sentiment classification," in *IJCAI*, 2020.