Prompt-Based Generative Multi-label Emotion Prediction with Label Contrastive Learning

Yuyang Chai[1], Chong Teng[1], Hao Fei[2], Shengqiong Wu[1], Jingye Li[1], Ming Cheng[3], Donghong Ji[1], and Fei Li[1](✉)

[1] Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, School of Cyber Science and Engineering, Wuhan University, Wuhan, China
{yychai,tengchong,whuwsq,theodorelee,dhji,lifei_csnlp}@whu.edu.cn

[2] School of Computing, National University of Singapore, Singapore, Singapore
haofei37@nus.edu.sg

[3] The first Affiliated Hospital of Zhengzhou University, Zhengzhou, China
fccchengm@zzu.edu.cn

Abstract. Multi-label emotion prediction, which aims to predict emotion labels from text, attracts increasing attention recently. It is ubiquitous that emotion labels are highly correlated in this task. Existing state-of-the-art models solve multi-label emotion prediction in sequence-to-sequence (Seq2Seq) manner, while such label correlations are merely leveraged in decoding side. In this work, we propose an emotion prediction framework to jointly generate emotion labels and template sentences via Seq2Seq language model. On the one hand, our template-based natural language generation method makes better use of generative language model compared with generating label sequences in the prior Seq2Seq-based generative classification model. On the other hand, we introduce the Correlation-based Label Prompts (CLP) through soft prompt learning and contrastive learning, which enables our model to further consider emotion label correlations in encoding side. To demonstrate the effectiveness of our prompt-based generative multi-label emotion prediction model, we perform experiments on the GoEmotions and SemEval 2018 datasets, achieving competitive results, outperforming 7 baselines w.r.t. 3 evaluation metrics. In-depth analyses show the generation manner is much more impressive compared with generating label sequences and our model is particularly effective in label correlation modeling.

Keywords: Emotion prediction · Text generation · Prompt learning · Contrastive learning

1 Introduction

Emotion prediction refers to automatically identify all the possible emotions expressed by individual in a piece of text [1,2], which has been applied to real-world applications, such as stock prediction [3], AI chat robots [4,5], etc. Since

W. Lu et al. (Eds.): NLPCC 2022, LNAI 13551, pp. 551–563, 2022.
https://doi.org/10.1007/978-3-031-17120-8_43

Sentences		*Emotion labels*
(S1)	I hate the message that's being pushed, and it's even worse that it's being pushed by this guy.	anger, disgust, embarrassment
(S2)	Worst advice ever. This would piss me off.	anger, disgust
(S3)	Exactly, I remember loving those videos but I was also like 12	love, realization
(S4)	Love situations like this...excited to see if he'll make it by tomorrow	love, excitement
(S5)	Gods don't die	neutral

Fig. 1. Illustration of the label correlation issue in multi-label emotion prediction.

an instance may often simultaneously involve multiple emotion labels, most relevant works solve this task as multi-label classification problem, a.k.a., multi-label emotion prediction [6,7]. A widely adopted approach in tackling multi-label emotion prediction is the binary relevance (BR) [8], which merely considers each emotion label as an independent binary problem.

Nevertheless, correlations do exist in emotion expressions [9]. Intuitively, emotions in the same polarity own strong correlations. As can be exemplified in Fig. 1, the negative emotion *"disgust"* is more likely to co-exist with *"anger"* rather than the positive emotion *"love"*. However, the correlations between predicted emotions are ignored in BR, which significantly benefits the overall predictions. Thus, the classifier chain (CC) [10] is proposed, which integrates label correlation information along a chain of classifiers. Very recent state-of-the-art works solve multi-label emotion classification in a sequence-to-sequence (Seq2Seq) manner (i.e., one decoding position for one emotion label), achieving considerable improvements [9,11–13]. Nevertheless, clear limitations can still be witnessed in those methods. First, the correlations between emotion labels are merely involved in decoding side and not considered in encoding side. Besides, the Seq2Seq architecture's inherent advantage, i.e., sequential language generation, is not respected in existing label sequence generation work. These hamper Seq2Seq from giving its most for multi-label emotion prediction.

Motivated by the recent progress on language generation together with large pre-trained language models (PLM) [14,15], in this work, we consider modeling the multi-label emotion prediction as a natural sentence generation problem. As shown in Fig. 2, based on a pre-trained Seq2Seq architecture, we generate a template-based natural sentence from which all the possible emotion labels will be yielded. Compared with existing multi-label emotion prediction works, we advance in two aspects. First of all, instead of generating a label sequence in prior Seq2Seq-based method, the text-to-text scheme naturally helps take full advantage of the power of the current generative PLM, e.g., BART [16], T5 [15]. Secondly, inspired by the success of soft prompt tuning [17], we introduce Correlation-based Label Prompts (CLP) to incorporate label correlation in the encoding side of our model.

We firstly introduce label prompts that aim to learn "label-like" representations similar to soft prompts [17], which helps our model better predict emotion labels by taking label information in prompts as references. Then, to involve label correlations in label prompts, we propose an extended contrastive loss to incorporate label correlations during label prompts learning stage and thus obtain CLP. During the contrastive learning, two label representations are forced to be pulled *closer* according to the two label co-exist frequency in corpus-level scope, and two *never* co-existed label's prompts would be pushed *further*. It is noteworthy that the polarity of emotions is not involved in contrastive learning. In this way, we extend the advantages of generative Seq2Seq model and meanwhile enhance the label correlation learning at the encoding side.

We conduct experiments on GoEmotions [1] and SemEval 2018 [2] datasets. Results show that our method achieves competitive performances on both two datasets, in terms of both the settings with and without using PLM. Compared with the current best-performing Seq2Seq models, our model consistently outperforms them on all metrics. Besides, our proposed contrastive loss is proved to be effective in label correlations learning. Our contributions can be summarized as follows:

- We propose transforming the multi-label emotion prediction into a natural language generation paradigm that considers label correlations and make full use of the existing generative PLM.
- We utilize label prompts to capture label information that helps our model better predict emotion labels.
- We propose an extended contrastive loss to obtain Correlation-based Label Prompts, which makes our model be aware of correlations modeling in encoding side.
- Experiments show that our proposed methods are effective and achieve competitive results on two widely-used datasets.[1]

2 Related Work

Emotion Prediction. Emotion prediction is an important branch of sentiment analysis and opinion mining in natural language processing (NLP) community [18–22]. Early works employ emotion lexicons for emotion multi-label classification [7,23]. This method constructs emotion lexicons to tackle emotion classification tasks, while most lexicons cover few domains, which restricts its performances on the short informal text like tweets. Besides, words may convey different emotions in different context [24], which results in mismatch problems with this approach. With the presence of the Transformer-based pre-trained language models, e.g., BERT [25], have been successfully employed in multi-label emotion classification task [1,26].

Similar to our work, Fei et al. (2020) [9] and Huang et al. (2021) [11] tackle multi-label classification problem in Seq2Seq manner. Fei et al. (2020) propose a Latent Memory Network based on the encoder-decoder framework that views multi-label classification as a label sequence generation problem. Huang et al. (2021) propose an LSTM-based Seq2Seq framework Seq2Emo to handle emotion classification and consider emotion correlations implicitly in decoding stage.

[1] Code is available at https://github.com/yychai74/Generative-MultiEmo.

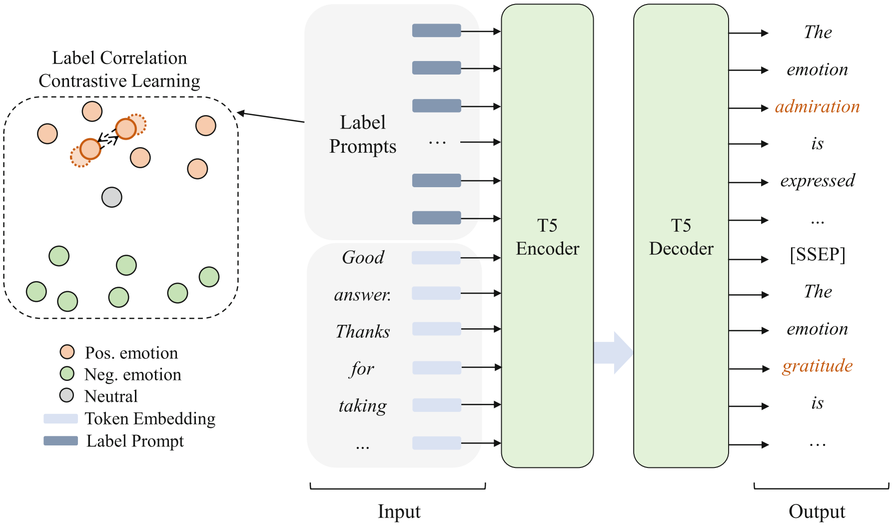


Fig. 2. The overall framework. The hidden representations of CLP are not used in decoder. Particularly, the Label Correlation Contrastive Learning only depends on the co-existence of labels and the distribution of emotions learned by CLP correctly reflects the correlations among emotions with different polarities (cf. Sect. 5.3).

Different from these works, our goal is to generate a natural language sentence rather than a label sequence, which makes full use of generative PLM.

Prompt Learning. Recently, prompt learning has become a new way to make use of the masked language model (MLM) [27,28]. Through designing a discrete prompt for specific downstream work, the MLMs can be fine-tuned to solve various tasks by simply changing prompts, which does not require model reconstructions. Besides, soft prompt (i.e., continuous prompts) [17] is utilized in prompt-tuning where only the prompt parameters need to be trained. Inspired by these works, we introduce label prompts into our framework. The prompts are the same length as the label category size and trained with our model jointly. The prompts would learn label representations during training step, and then our model can take label prompts as references to better predict emotion labels.

Contrastive Learning. Contrastive learning aims to pull "positive" examples together and push "negative" examples apart, which has been widely adopted in the computer vision area [29]. Recently, supervised contrastive learning has been employed to tackle many NLP tasks. Suresh and Ong (2021) [30] introduce a weighted contrastive loss to help model better classify indistinguishable labels in fine-grained classification tasks. In this paper, we propose an extended contrastive loss that aims to incorporate emotion label correlations into the label prompts training stage.

3 Framework

3.1 Task Formulation

In our work, we solve multi-label emotion prediction task in a generative manner. Given a sentence X, we aim to generate a template-based natural sentence Y that contains all predicted emotion labels with our model. If the model generates a sentence with wrong templates or emotions that are beyond the label set, the predictions would be abandoned.

We utilized a template sentence $\mathbf{T}$, *"The emotion a_i is expressed in this sentence."* to construct the target sentences, where a_i is the label the instance X possesses. For multi-labeled instances, we first transfer labels into multiple sentences with template $\mathbf{T}$, then we concatenate them with a pre-defined separation token `[SSEP]` in the order of labels to obtain the target sequence Y.

3.2 Encoder

Given an input sentence X, we obtain its token embedding matrix $\mathbf{E}_n$ through the internal embedding lookup matrix of PLM, where $\mathbf{E}_n \in \mathbb{R}^{n\times d}$, d denotes the hidden dimension of the encoder and n is the tokenized length of sentence X. Then we concatenate **label prompts** $\mathbf{P}_l$ with $\mathbf{E}_n$:

$$\mathbf{E} = \mathrm{Concat}(\mathbf{P}_l, \mathbf{E}_n) \tag{1}$$

where $\mathbf{P}_l \in \mathbb{R}^{l\times d}$ and l denotes the size of emotion categories[2] of the dataset.

The label prompts are trained jointly to learn emotion label representations and the label correlations would be incorporated into label prompts with the usage of our proposed contrastive loss. Then the vectorized word representations are fed into an encoder to obtain the context-sensitive representation:

$$\mathbf{H}^{enc} = \mathrm{Encoder}(\mathbf{E}) \tag{2}$$

where $\mathbf{H}^{enc} \in \mathbb{R}^{(l+n)\times d}$. With the help of self-attention mechanism [31], the correlation-based label information is naturally integrated into the hidden representations of the sentence X.

3.3 Decoder

At the inference stage, the hidden representations obtained from the encoder are fed to the decoder via cross-attention layers. Noteworthily, we do not feed the label prompts to the decoder as in Fig. 2, which will be discussed in Sect. 5.5.

The decoder generates template-based natural sentences in an auto-regressive manner. At the c-th time step, the hidden representations and previous outputs $y_{<c}$ are applied to compute the decoder outputs:

$$\mathbf{H}^{dec}_c = \mathrm{Decoder}(\mathbf{H}^{enc}_{l:}, y_{<c}) \tag{3}$$

[2] We do not use the tokenized label token size in this work, so as to ensure that the model focuses label prompts on label level rather than token level.

where l is the size of emotion categories and $\mathbf{H}_{l:}^{enc}$ denotes the hidden representations with label prompts peeled. To obtain the probability for the next token in the vocabulary set, the softmax function is utilized:

$$P(y_c|y_{<c}, \mathbf{H}_{l:}^{enc}) = \text{Softmax}(\boldsymbol{W}^T \mathbf{H}_c^{dec} + b) \quad (4)$$

$\boldsymbol{W} \in \mathbb{R}^{d \times |\mathcal{V}|}$ maps decoder outputs $\mathbf{H}_c^{dec}$ to a logit vector which can be used to compute probability distribution over vocabulary set, and $|\mathcal{V}|$ represents the vocabulary set size of PLM. $\boldsymbol{W}$ and b are all learnable parameters.

3.4 Training

Label Correlation Contrastive Loss. The Supervised Contrastive Loss (SCL) makes the representations of samples belonging to the same class stay closer [32,33]. For an instance x_i, the positive set among batch B is given by $\mathcal{P} = \{p|y_p = y_i, p \neq i\}$, where y_i is the label of x_i. Let I denotes the indexes of examples in training batch B, then the Supervised Contrastive Loss over batch B is defined as:

$$\mathcal{L}_B^{SCL} = \sum_{i=1}^{k} \frac{-1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} \log \frac{\exp(h_i \cdot h_p/\tau)}{\sum_{b \in I, b \neq i} \exp(h_i \cdot h_b/\tau)} \quad (5)$$

where k is batch size, h_i is the representation vector of x_i obtained from an encoder and τ is temperature hyper-parameter.

However, the SCL brings representations of samples in the same class closer together, which is inconsistent with our goal of label correlations modeling. In this case, we extend the aforementioned contrastive loss to enhance label prompts learning with the awareness of label correlations. We aim to train two label representations closer if the two label co-exists with a higher probability. Our proposed supervised Label Correlation Contrastive Loss (LCCL) is defined as:

$$\mathcal{L}_B^{LCCL} = \sum_{i=1}^{l} \frac{-1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} \log \frac{\exp(e_i \cdot e_p/\tau)}{\sum_{p \in \mathcal{P}} \exp(e_i \cdot e_p/\tau) + \sum_{n \in \mathcal{N}} \exp(e_i \cdot e_n/\tau) + \varepsilon} \quad (6)$$

In Eq. 6, l is the label category size of the dataset, e_i denotes the normalised label prompt of label a_i[3]. The positive set $\mathcal{P}$ contains the indexes of co-existed labels with label a_i in training batch B, and the negative set $\mathcal{N}$ contains the indexes of labels that *never* co-existed with a_i in the overall train set. The hyper-parameter ε is to ensure label prompts can be correctly trained when $\mathcal{N} = \emptyset$.

Here, unlike Suresh and Ong's work [30] that introduces extra weights, we model different label correlations in LCCL implicitly since two label embeddings are pushed closer based on their co-exist frequency. In this case, label correlations are incorporated into label prompts learning so that our model can better handle multi-label emotion prediction task with the help of the CLP.

[3] The label prompt index are aligned with the emotion label order in corresponding dataset.

Generative Loss. The cross-entropy loss between the output sentence and the target template sentence is used to optimize the PLM:

$$\mathcal{L}^{CE} = -\sum_{c=1}^{m} \log P(y_c|y_{<c}, \mathbf{H}_c^{enc}) \tag{7}$$

where m denotes the length of target sentence.

Finally, the overall generative framework is optimized using the combination of cross-entropy loss and our proposed LCCL:

$$\mathcal{L} = (1-\alpha)\mathcal{L}^{CE} + \alpha\mathcal{L}^{LCCL} \tag{8}$$

where α is a hyper-parameter to balance the two parts of the loss function.

Table 1. Data statistics. *Emo.* is the numbers of emotion categories. *Multi.* denotes the percentage of multi-labeled sentences. *Avg.Len.* is the average length of sentences.

Dataset	Train	Dev	Test	Emo.	Multi.	Avg.Len.
GoEmotions	43,410	5,426	5,427	28	16.2%	12.82
SemEval 2018	6,838	886	3,259	11	86.1%	16.04

4 Experiments

Dataset. We conduct our experiments on GoEmotions [1] and Semeval 2018 Task 1 [2] datasets. The details of the two datasets are shown in Table 1. Particularly, we utilize *neutral* to generate template-based sentences for those non-labeled instances in Semeval 2018 dataset, and all predictions that include neutral are restored to original label in the evaluation stage.

Baselines and Metrics. Since GoEmotions is a newly released dataset, we include **Seq2Emo** results reported by Huang et al. (2021) [11] and **BERT** results reported by Demszky et al. (2020) [1]. Seq2Emo is a Seq2Seq model to tackle multi-label classification problem, and label correlations are leveraged in decoding stage. Additionally, Huang et al. report results of **BR** and **CC** with their implementations, which are also included for comparison. For SemEval 2018 dataset, we compare our model with following baselines: **NTUA-SLP** [7] trains a two-layer LSTM network where a large amount of external emotion lexicons are used. **DATN** [6] employs dual attention mechanism to encode tweets into features. Zhou et al. (2020) [26] introduce an emotion network (**EmNet**) that can alleviate the domain mismatch and emotion ambiguity problems of using external lexicons, and we obtain **BERT** and **EmNet+BERT** results from this paper. We include **SGM** results from [11] additionally. Besides, **Seq2Emo** also reports results on SemEval 2018 dataset. We follow former works [2,11,26], using Micro F1 (**MiF**), Macro F1 (**MaF**) scores, and Jaccard Index (**Jacc**) to measure the model performances. The results are averaged over 5 runs.

Experiment Setting. We adopt `T5-base` [15] from Huggingface Transformers[4] as the backbone framework. Besides, we also employ `T5-large` for further comparison and all the BERT mentioned in baselines are *base* version. The learning

[4] https://huggingface.co/.

rate is set to 4e−5, the batch size is set to 16 and our model is trained for 20 epochs where early stopping strategy is utilized. The hyper-parameter α and ε are set to 0.1 and 0.01 respectively. The temperature parameter of LCCL is found to be 0.3 for GoEmotions and 0.07 for SemEval 2018. Label prompts are randomly initialized and the dimensions are the same as the hidden size of T5. The beam sizes are set to 3 for GoEmotions and 2 for SemEval 2018.

5 Results and Analysis

5.1 Main Results

The results on GoEmotions are shown in Table 2. The first observation we can find is that our proposed model surpasses all baselines on three metrics, and $T5_{large}$ makes further progress. Compared with the state-of-the-art model Seq2Emo with BERT encoder, our model gains 1.03 more scores on Micro F1 and much more consistent improvement on Macro F1 (+4.48 scores), Jaccard Index (+5.11 scores). This proves the significance of our proposed generative multi-label prediction model.

Table 2. Results on GoEmotions dataset. Best results except $T5_{large}$ are shown in bold and the results of $T5_{large}$ are underlined. † and ‡ means the result is significant with $p < 0.01$ and $p < 0.05$ compared with $T5_{base}$ respectively. ∗ denotes results obtained by our implementations.

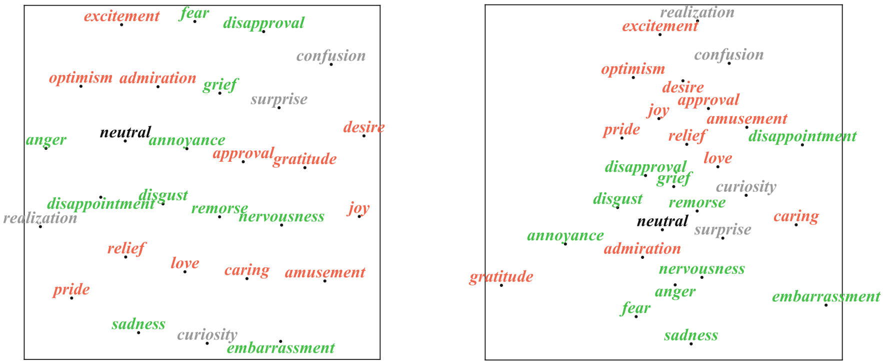
Model	MiF	MaF	Jacc
BERT	–	46.00	–
BR	58.21	45.38	52.76
CC	58.38	43.92	55.61
Seq2Emo	59.57	47.28	53.79
Seq2Emo+BERT*	59.96	49.04	54.45
$T5_{base}$	60.52	51.43	59.08
$T5_{base}$+CLP	**60.99**‡	**53.52**†	**59.56**‡
$T5_{large}$+CLP	<u>61.32</u>†	<u>53.78</u>†	<u>59.73</u>†

Table 3. Results on SemEval 2018 dataset. The notations are the same with the left table.

Model	MiF	MaF	Jacc
SGM	55.11	–	45.14
NTUA-SLP	70.10	52.80	58.80
DATN	–	54.40	58.30
BERT	70.10	53.00	58.00
Seq2Emo	70.02	51.92	58.67
Seq2Emo+BERT*	70.52	53.48	59.03
EmNet+BERT	**71.60**	**56.50**	59.60
$T5_{base}$	70.95	55.62	60.24
$T5_{base}$+CLP	71.34‡	55.84	**60.80**‡
$T5_{large}$+CLP	<u>71.86</u>†	<u>57.39</u>†	<u>61.41</u>†

Table 4. Generation using different templates on GoEmotions.

Templates	MiF	MaF	Jacc
The emotion a_i is expressed in this sentence	**60.99**	**53.52**	**59.56**
It expressed emotion a_i	60.61	52.75	59.25
It is clear that the emotion a_i is expressed in this sentence.	60.66	52.49	59.16
No template, directly output labels	60.40	52.05	58.81



(a) Label prompts. (b) Correlation-based label prompts.

Fig. 3. Visualization of the label prompts on GoEmotions, where orange, green and gray colors correspond with positive, negative and ambiguous emotions respectively. (Color figure online)

As for SemEval 2018 dataset, we show our results in Table 3. Our model achieves state-of-the-art performance on Jaccard Index score and slightly lower but competitive results compared with EmNet. One possible reason is that the SemEval 2018 dataset has fewer training data so that label prompts cannot learn heuristic label representations during training. In this case, CLP does not bring that great improvements as the GoEmotions dataset. Particularly, T5$_{large}$ achieves the state-of-the-art results, verifying the potential of generative multi-label emotion prediction.

We conduct paired t-tests and ablation studies to validate the contributions of CLP to our framework, and the results are shown in Table 2 and Table 3. We can notice that the CLP brings significant performance increases on both datasets, which verifies the effectiveness of CLP in multi-label emotion prediction. Besides, even the T5$_{base}$ outperforms the BERT-based Seq2Emo on two datasets. Further, CLP brings 4.1% Macro F1 score improvement on the more fine-grained GoEmotions dataset, indicating our model can better predict emotion labels that less occur in the dataset with the help of label information contained in the CLP.

5.2 Effect of Natural Language Template

To further analyze the effect of templates to our model, we train our model using different templates, and the results are shown in Table 4. We can observe that different templates make differences in the performance of our model, which is consistent with findings of Liu et al. (2021) [14] and Gao et al. (2021) [28]. Additionally, we also train our model to generate label sequences as SGM and achieve the worst performances on GoEmotions dataset, which verifies the advantage of natural language generation compared with generating label sequences.

5.3 Effectiveness of Contrastive Learning

We perform visualization of the CLP and non-LCCL-used pure label prompts learned on GoEmotions using t-SNE to disclose the effectiveness of contrastive learning. As seen in Fig. 3, we scatter each learned label representation with corresponding color based on the sentiment polarity provided in [1] officially. It is obvious that label representations trained with LCCL own stronger label correlations. The emotions in Fig. 3(a) scatter randomly, whereas the positive and negative emotions cluster with emotions in the same polarity in Fig. 3(b), which is consistent with the intuitive conclusion mentioned in Sect. 1. Besides, ambiguous emotions and neutral fall in reasonable places in Fig. 3(b), e.g., *realization* is closer to positive emotions whereas *curiosity* stays closer to negative emotions.

5.4 Predictions on Different Emotion Label Numbers

We further investigate the predictions of our model for sentences with different numbers of labels, as shown in Table 5. First, the performance on single-labeled sentences is improved, which verifies the helpfulness of the CLP in providing emotion label information during prediction. Then, the CLP enhances the ability of multi-label emotion prediction of our model, indicating that label correlations are highly considered in our model.

Table 5. Jaccard Index score of T5$_{base}$ and T5$_{base}$+CLP on sentences with different numbers of labels.

	GoEmotions			SemEval 2018			
Number of Labels:	1	2	≥ 3	1	2	3	≥ 4
T5$_{base}$	61.84	42.72	34.28	34.70	66.55	66.21	52.33
T5$_{base}$+CLP	62.54	43.68	34.43	35.41	67.26	66.58	53.52

5.5 Should CLP Be Used in Decoder?

After obtaining the hidden representations from encoder, only the encoded token representations are used in decoder via cross-attention layer. One may naturally ask the follow question: *If the CLP are used in decoder, how does the performance become?* We evaluate such hypothesis using the experiment shown in Table 6 and find that the results get hurt for both datasets. Despite the CLP are learning

Table 6. Results with using CLP in decoder.

	MiF	MaF	Jacc
GoEmotions	60.99	53.52	59.56
w. CLP	60.69	51.93	59.11
SemEval 2018	71.34	55.84	60.80
w. CLP	70.75	55.69	60.14

"label-like" representations [17], they are not any true token embeddings in T5 (i.e., not actual tokens in the vocabulary set of T5) and thus do not contain any text meaning. In this case, using CLP that do not own practical text meaning has an adverse effect on text-to-text based generation in T5. Therefore, the CLP should not be used in the decoder.

6 Conclusion

In this work, we propose a generative framework to tackle multi-label emotion prediction, where we introduce Correlation-based Label Prompts (CLP) via prompt learning and contrastive learning. Experimental results on two datasets show that our model outperforms 7 baselines on two datasets significantly and further experiments prove the advantage of template-based language generation compared with generating label sequence. Besides, our model is identified that label correlations are effectively incorporated in the encoding side, which helps our model predict multi-labeled sentences more efficaciously.

Acknowledgment. This work is supported by the National Natural Science Foundation of China (No. 62176187), the National Key Research and Development Program of China (No. 2017YFC1200500), the Research Foundation of Ministry of Education of China (No. 18JZD015), the Youth Fund for Humanities and Social Science Research of Ministry of Education of China (No. 22YJCZH064), the General Project of Natural Science Foundation of Hubei Province (No. 2021CFB385), the Science and Technology Project of Henan Province (No. 222102210112).

References

1. Demszky, D., Movshovitz-Attias, D., Ko, J., Cowen, A., Nemade, G., Ravi, S.: GoEmotions: a dataset of fine-grained emotions. In: Proceedings of the ACL, pp. 4040–4054 (2020)
2. Mohammad, S., Bravo-Marquez, F., Salameh, M., Kiritchenko, S.: Semeval-2018 task 1: affect in tweets. In: Proceedings of the SemEval, pp. 1–17 (2018)
3. Nguyen, T.H., Shirai, K., Velcin, J.: Sentiment analysis on social media for stock movement prediction. Expert Syst. Appl. **42**(24), 9603–9611 (2015)
4. Zhou, H., Huang, M., Zhang, T., Zhu, X., Liu, B.: Emotional chatting machine: Emotional conversation generation with internal and external memory. In: Proceedings of the AAAI, pp. 730–739 (2018)
5. Fei, H., Ren, Y., Zhang, Y., Ji, D.: Nonautoregressive encoder-decoder neural framework for end-to-end aspect-based sentiment triplet extraction. IEEE Trans. Neural Netw. Learn. Syst. 1–13 (2021). https://ieeexplore.ieee.org/abstract/document/9634849
6. Yu, J., Marujo, L., Jiang, J., Karuturi, P., Brendel, W.: Improving multi-label emotion classification via sentiment classification with dual attention transfer network. In: Proceedings of the EMNLP, pp. 1097–1102 (2018)
7. Baziotis, C., et al.: NTUA-SLP at SemEval-2018 task 1: Predicting affective content in tweets with deep attentive RNNs and transfer learning. In: Proceedings of the SemEval, pp. 245–255 (2018)

8. Godbole, S., Sarawagi, S.: Discriminative methods for multi-labeled classification. In: Proceedings of the PAKDD, pp. 22–30 (2004)
9. Fei, H., Zhang, Y., Ren, Y., Ji, D.: Latent emotion memory for multi-label emotion classification. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 7692–7699 (2020)
10. Read, J., Pfahringer, B., Holmes, G., Frank, E.: Classifier chains for multi-label classification. Mach. Learn. **85**(3), 333–359 (2011)
11. Huang, C., Trabelsi, A., Qin, X., Farruque, N., Mou, L., Zaiane, O.R.: Seq2Emo: a sequence to multi-label emotion classification model. In: Proceedings of the NAACL, pp. 4717–4724 (2021)
12. Fei, H., Ren, Y., Wu, S., Li, B., Ji, D.: Latent target-opinion as prior for document-level sentiment classification: a variational approach from fine-grained perspective. In: Proceedings of the WWW: the Web Conference 2021, pp. 553–564 (2021)
13. Fei, H., Li, J., Ren, Y., Zhang, M., Ji, D.: Making decision like human: joint aspect category sentiment analysis and rating prediction with fine-to-coarse reasoning. In: Proceedings of the WWW: the Web Conference, pp. 3042–3051 (2022)
14. Liu, J., Teng, Z., Cui, L., Liu, H., Zhang, Y.: Solving aspect category sentiment analysis as a text generation task. In: Proceedings of the EMNLP, pp. 4406–4416 (2021)
15. Raffel, C., et al.: Exploring the limits of transfer learning with a unified text-to-text transformer. J. Mach. Learn. Res. **21**(140), 1–67 (2020)
16. Lewis, M., et al.: BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In: Proceedings of the ACL, pp. 7871–7880 (2020)
17. Lester, B., Al-Rfou, R., Constant, N.: The power of scale for parameter-efficient prompt tuning. In: Proceedings of the EMNLP, pp. 3045–3059 (2021)
18. Fei, H., Zhang, M., Ji, D.: Cross-lingual semantic role labeling with high-quality translated training corpus. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 7014–7026 (2020)
19. Fei, H., Wu, S., Ren, Y., Li, F., Ji, D.: Better combine them together! integrating syntactic constituency and dependency representations for semantic role labeling. In: Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021, pp. 549–559 (2021)
20. Fei, H., Li, F., Li, B., Ji, D.: Encoder-decoder based unified semantic role labeling with label-aware syntax. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 12794–12802 (2021)
21. Fei, H., Zhang, M., Li, B., Ji, D.: End-to-end semantic role labeling with neural transition-based model. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 12803–12811 (2021)
22. Shi, W., Li, F., Li, J., Fei, H., Ji, D.: Effective token graph modeling using a novel labeling strategy for structured sentiment analysis. In: Proceedings of the ACL, pp. 4232–4241 (2022)
23. Wu, S., Fei, H., Ren, Y., Ji, D., Li, J.: Learn from syntax: Improving pair-wise aspect and opinion terms extraction with rich syntactic knowledge. In: Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, pp. 3957–3963 (2021)
24. Fei, H., Ji, D., Zhang, Y., Ren, Y.: Topic-enhanced capsule network for multi-label emotion classification. IEEE/ACM Trans. Audio Speech Lang. Process. **28**, 1839–1848 (2020)

25. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the NAACL, pp. 4171–4186 (2019)
26. Zhou, D., Wu, S., Wang, Q., Xie, J., Tu, Z., Li, M.: Emotion classification by jointly learning to lexiconize and classify. In: Proceedings of the COLING, pp. 3235–3245 (2020)
27. Fei, H., Ren, Y., Ji, D.: Retrofitting structure-aware transformer language model for end tasks. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, pp. 2151–2161 (2020)
28. Gao, T., Fisch, A., Chen, D.: Making pre-trained language models better few-shot learners. In: Proceedings of the ACL, pp. 3816–3830 (2021)
29. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: Proceedings of the ICML, pp. 1597–1607 (2020)
30. Suresh, V., Ong, D.: Not all negatives are equal: Label-aware contrastive loss for fine-grained text classification. In: Proceedings of the EMNLP, pp. 4381–4394 (2021)
31. Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems, pp. 5998–6008 (2017)
32. Khosla, P., et al.: Supervised contrastive learning. In: Advances in Neural Information Processing Systems, pp. 18661–18673 (2020)
33. Fei, H., Wu, S., Ren, Y., Zhang, M.: Matching structure for dual learning. In: Proceedings of the 39th International Conference on Machine Learning, ICML, pp. 6373–6391 (2022)