PERD: Personalized Emoji Recommendation with Dynamic User Preference

Xuanzhi Zheng
zhengxuanzhi@stu.xjtu.edu.cn
Xi'an Jiaotong University
Xi'an, China

Guoshuai Zhao*
guoshuai.zhao@xjtu.edu.cn
Xi'an Jiaotong University
Xi'an, China

Li Zhu
zhuli@mail.xjtu.edu.cn
Xi'an Jiaotong University
Xi'an, China

Xueming Qian*
qianxm@mail.xjtu.edu.cn
Xi'an Jiaotong University
Xi'an, China

ABSTRACT

Emoji recommendation is an important task to help users find appropriate emojis from thousands of candidates based on a short tweet text. Traditional emoji recommendation methods lack personalized recommendation and ignore user historical information in selecting emojis. In this paper, we propose a personalized emoji recommendation with dynamic user preference (PERD) which contains a text encoder and a personalized attention mechanism. In text encoder, a BERT model is contained to learn dense and low-dimensional representations of tweets. In personalized attention, user dynamic preferences are learned according to semantic and sentimental similarity between historical tweets and the tweet which is waiting for emoji recommendation. Informative historical tweets are selected and highlighted. Experiments are carried out on two real-world datasets from Sina Weibo and Twitter. Experimental results validate the superiority of our approach on personalized emoji recommendation.

CCS CONCEPTS

• **Information systems** → *Information systems applications*; *Retrieval tasks and goals*; **Recommender systems**.

KEYWORDS

recommendation system, emoji recommendation, deep learning

ACM Reference Format:
Xuanzhi Zheng, Guoshuai Zhao, Li Zhu, and Xueming Qian. 2022. PERD: Personalized Emoji Recommendation with Dynamic User Preference. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '22), July 11–15, 2022, Madrid, Spain.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3477495.3531779

*Corresponding authors

SIGIR '22, July 11–15, 2022, Madrid, Spain

ACM ISBN 978-1-4503-8732-3/22/07…$15.00
https://doi.org/10.1145/3477495.3531779

1 INTRODUCTION

Emoji is a new type of symbolic language expressing diversified sentiments. They are popular on social platforms, with about 90% of users using them in tweets frequently. And they transmit 6 billion emojis every day[10]. Compared with plain text, emojis are artistic imitations of human facial expressions and body languages, which help users express feelings vividly and enhance the sentiment expressions of short tweets[3]. The number of emojis has expanded considerably in recent years. In Unicode 10.0, 1144 emojis are defined and 112 emojis added in Unicode 14.0. Thus, users may feel difficult to quickly select emojis from numerous candidate emojis in social media platforms. Therefore, emoji prediction task is proposed which aims to predict the accompanying emojis which might be used in plain text. It gradually become an important task in recommendation system.

Several traditional matrix factorization based methods[20] and deep learning based methods[2][11][15] are proposed for emoji prediction task. However, all these existing methods either do not consider user preferences in emoji choice or learn fixed user preferences when facing different tweet text. In addition, user historical tweets are ignored, which may be benefit for modeling user preferences and implementing personalized and diverse emoji recommendation.

To address these issues, we propose a personalized emoji recommendation with dynamic user preference (PERD) model. The core of our approach is text encoder and personalized attention. In text encoder module, a pre-trained BERT model is contained to learn text representations of tweets. Since historical tweets with different semantics and sentiments imply different aspects of the user preference, a personalized attention is designed to highlight informative historical tweets and generate dynamic user preferences.

The main contributions of this paper are as follow.

- We propose a personalized emoji recommendation with dynamic user preference (PERD) model based on plain text and user historical tweets. Experiment results show that our model achieves better performance than baselines.
- We propose personalized attention module to learn dynamic user preference representations from text and emojis in historical tweets. Personalized attention module recognize and highlight informative historical tweets. Emoji labels in historical tweets weighted sum to generate dynamic preferences.

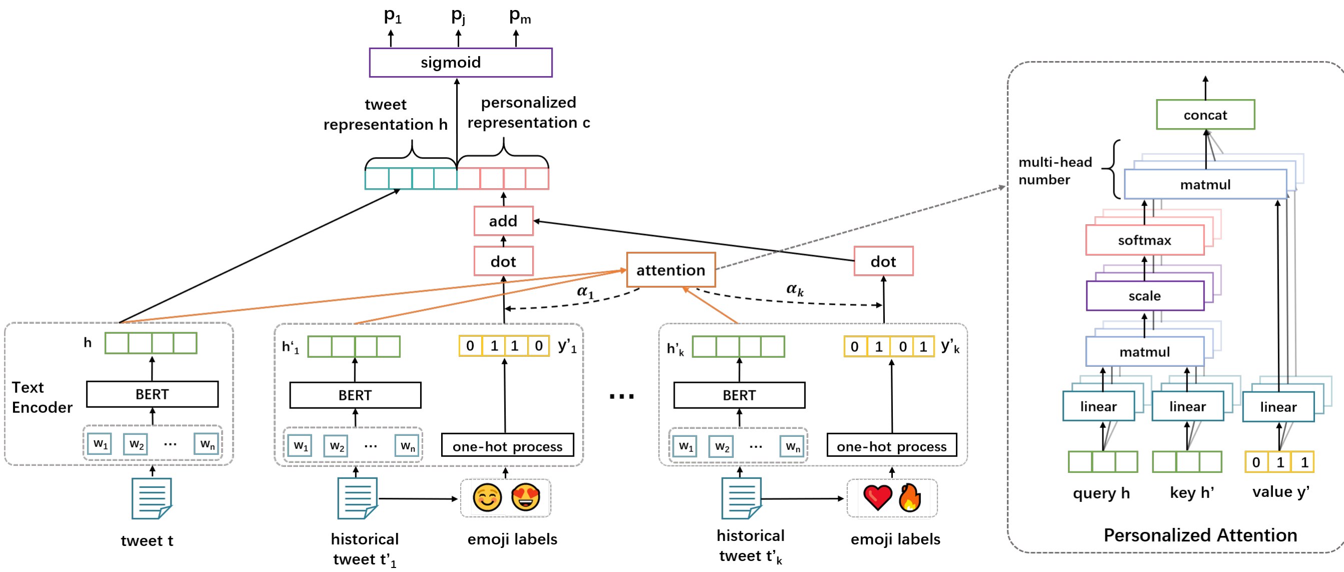

Figure 1: The architecture of our PERD model for emoji recommendation.

2 RELATED WORK

A series of studies have been carried out on analysing and recommending emojis. At first, scholars focused on analyzing the usage of emojis. With further research on emojis, emoji prediction task has begun to attract more attention. Barbieri et al.[4] defined the emoji prediction task and designed a multi-modal method through integrating textual and visual information. This research confirms that emojis can be predictable based on text information. Felbo et al.[8] proposed a sentiment and sarcasm detection method based on neural networks, which pre-trained by emoji prediction task on a dataset with 1246 million tweets and 64 common emojis. Further, tens of studies on emoji prediction in multi-language were published at SemEval 2018 as shared task[1]. The above studies focus on single emoji prediction. It is observed that multiple emojis are often contained in a piece of text in real life. This emoji usage habit indicates that the multi-emoji prediction task is more practical for application. Peng et al.[16] proposed a multi-emoji prediction model, Seq2Emoji. They considered the correlation between emojis and introduced diverse beam search method to enhance the diversity of prediction. Wu et al.[18] proposed a neural method for multi-emoji prediction through combining CNN and LSTM. An attention mechanism is used to choose the words with more information in text.

The above works predict emojis only through considering the plain text. However, the contextual and personal information are important factors about emoji choice. Zhao et al.[19] proposed a matrix factorization method which makes full use of the context sensitive information for emoji prediction. To take sentiment factors and emoji position prediction into account, researchers proposed the sentiment-aware emoji prediction task and emoji position prediction task. Zhao et al.[20] proposed a mmGRU model to predict categories and positions of emojis, which leveraged the strong correlations between emoji categories and their positions. Lin et al.[14] proposed a BERT-BiLSTM-CRF model, which predicted emojis based on tweet text and sentiment polarities. Lee et al.[13] presented a multi-task method for emoji recommendation. An emotion detection task is introduced as an auxiliary task in order to improve the performance of emoji recommendation.

3 OUR APPROACH: PERD

We first define the problem and then describe our personalized emoji recommendation with dynamic user preference (PERD) model in detail. PERD method learns tweet representations through text encoder. Then a user personalized attention is proposed to learn user dynamic preferences from their historical tweets. Tweet representations and user preference representations are integrated to rank candidate emojis by calculating their scores. The architecture of our approach is shown in Figure 1.

3.1 Problem Definition

In this paper, uppercase letters represent sets and corresponding lowercase letters with superscripts indicate elements in sets. We define user set $U = [u_1, u_2, \ldots, u_g]$ with size g, tweets data $T = [t_1, t_2, \ldots, t_n]$ with size n and emoji set $E = [e_1, e_2, \ldots, e_m]$ with size m. The i-th tweet $t_i = [w_1, w_2, \ldots, w_x]$ is composed of a sequence of words. According to above definitions, our emoji recommendation task can be formulated as a multi-label task. Given a tweet set and a user set with their recent posted tweets, our goal is to recommend emojis which might be used in each tweet from candidate emojis E. The ground-truth of predictions can be denoted as $Y = [y_1, y_2, \ldots, y_n]$. $y_i = \{0, 1\}^m$ indicates whether emojis are used in the tweet t_i. $y_{ij} = 1$ means that the j-th emoji e_j is included in tweet t_i.

3.2 Text Encoder

In our approach, the text encoder module is responsible for learning tweet representations. The input of the text encoder is a sequence of words $t = [w_1, w_2, \ldots, w_x]$, and it will be converted to low-dimensional dense vectors named tweet representations. Since the superior performance of the BERT model for NLP tasks, a pre-trained BERT model is contained in the text encoder to generate tweet representations. In BERT architecture, the first token of each sequence is a unique token([CLS]) added in pre-processing stage. The embedding of this unique token aggregates the semantics of whole sequence. It is denoted as tweet representation h.

$$h = BERT\ (t) = BERT\ (w_1, w_2, \ldots, w_x) \tag{1}$$

3.3 Personalized Attention

We propose personalized attention module to generate the user dynamic personalized representations based on text and emojis in their historical tweets. Considering the limitation of the training time and run-time memory, we cannot fully utilize the whole user historical tweets. For each user, we select their s recent historical tweets to form historical tweet list $T' = [t'_1, t'_2, \ldots, t'_s]$. If the whole number of historical tweets is less than s, the empty part of T' will be filled with empty strings. The emoji labels Y' corresponding to T' can be written as:

$$Y' = [y'_1, y'_2, \ldots, y'_s] \tag{2}$$

The tweet representations of T' can be calculated by text encoder, denoted as:

$$H' = [BERT(t_1), BERT(t_2), \ldots, BERT(t_s)] = [h'_1, h'_2, \ldots, h'_s] \tag{3}$$

Due to the high correlation between emojis and semantics of text, historical tweets with different semantics and sentiments usually have different informativeness. When model predicts emojis for the tweet t_i, the higher similarity in semantics and sentiments between t_i and historical tweets, the more user preference clues included. For example, when we recommend emojis for a tweet "I got a good grade in yesterday's exam", the emojis used in historical tweets with positive sentiment like "What a nice day!" are more likely to be recommended, compared with emojis used in tweets with negative sentiment. Therefore, recognizing important historical tweets is useful for characterizing user preferences.

In the vanilla attention mechanism, attention weights are usually calculated based on the same input of key and value vectors. In this way, the information in historical emoji labels will lose. Hence, we propose a personalized attention mechanism, which calculates the weight parameters of s historical tweets by defining the function of the tweet representation h, representations of historical tweets H' and corresponding emoji labels Y'. The query vector q in the personalized attention is calculated by the projection computation of h:

$$q = W_q * h + bias_q \tag{4}$$

where W_q and $bias_q$ are parameters. The inputs of the key vectors K and the value vectors V in personalized attention are representations of historical tweets H' and corresponding emoji labels Y'. They are formulated as:

$$K = W_k * H' + bias_k \tag{5}$$

$$V = W_v * Y' + bias_v \tag{6}$$

The attention weight of each historical tweet label is computed through inner product and softmax operation between tweet query q and historical tweet keys K. The attention weight of j-th emoji label can be formulated as:

$$a_j = k_j^T q \tag{7}$$

$$\alpha_j = \frac{exp(a_j)}{\sum_{j=1}^{s} exp(a_j)} \tag{8}$$

The final dynamic personalized representation c_i of user u and tweet t_i is the summation of historical emoji labels weighted by attention weights:

$$c_i = \sum_{j=1}^{s} \alpha_j v'_j \tag{9}$$

With the help of user dynamic personalized representations, our model implement personalized recommendation. For same text edited by different users, our model can predict different emojis which are tailored to user preferences.

3.4 Prediction and Model Training

In this section, we aim to predict the probability p_{ij} of each emoji e_j for tweet text t_i and define loss function to optimize our model. We concatenate the tweet embedding h_i from the text encoder and the user personalized representation c_i from personalized attention. We compute probability through sigmoid functions. Considering that a tweet may contain several different emojis, we transform the multi-label emoji prediction problem into m binary classification tasks. m sigmoid activation functions are used to predict the probability of m categories of emojis, which can be calculated as:

$$p_{ij} = sigmoid(W * concat(h_i, c_i) + bias) \tag{10}$$

where $p_{ij} \in [0, 1]$. W, $bias$ are parameters. We use the cross-entropy function for model optimization, which is formulated as:

$$\zeta = -\sum_{i}^{n}\sum_{j}^{m}\left(y_{ij} log\left(p_{ij}\right) + \left(1 - y_{ij}\right) log\left(1 - p_{ij}\right)\right) \tag{11}$$

4 EXPERIMENTS

4.1 Experimental Setting

In this paper, our experiments are performed on two real-world datasets named Weibo and Twitter, which are collected by Zhao et al.[19]. There are 1.53 Million tweets posted by 89,600 users and 50 unique emojis in Weibo dataset. Twitter dataset contains 1.63 Million tweets, 6,500 users and 50 unique emojis.

In our experiments on Weibo dataset, the Chinese BERT model pre-trained by Cui et al.[6] is used. The depth of the model is set to 12, the dimension of word vectors is set to 768, and the multi-head number in attention is set to 12. In experiments on Twitter dataset, the BERT-Medium model pre-trained by Devlin et al.[7] is used, with 8 layers, 512 hidden states and 8 heads. The dimension of user preference representations is same as that of word vectors. Due to the limitation of run-time memory, the length of user historical tweet list s is set to 4. Adam[12] is used for gradient descend. The initial learning rate is set to $5e{-}5$. Model training is completed until the training iteration reaches the maximum epoch number or the training loss stops decreasing. The model with the best performance on the validation set is used for test. The metrics in our experiments are Precision, Recall, F1-Score, and NDCG scores. Our code is released on GitHub[1].

Table 1: Experimental results on Weibo Dataset

Methods	Kim-CNN	libFM	B-LSTM	DeepFM	mmGRU	CAPER	PERD
P@5	0.0849	0.0929	0.1054	0.1011	0.1302	0.1151	**0.1242**
R@5	0.4238	0.3687	0.3962	0.3765	0.5191	0.4472	**0.5337**
F@5	0.1415	0.1484	0.1665	0.1594	0.2082	0.1831	**0.2016**
P@10	0.0604	0.0632	0.0814	0.0741	0.0789	0.0817	**0.0822**
R@10	0.6043	0.5013	0.3136	0.5522	0.6291	0.6349	**0.7061**
F@10	0.1098	0.1122	0.1318	0.1307	0.1402	0.1448	**0.1472**
N@5	0.2903	0.3105	0.3187	0.2688	0.5024	0.3399	**0.3886**
N@10	0.3321	0.3593	0.3663	0.3287	0.5408	0.3932	**0.4307**

Table 2: Experimental results on Twitter Dataset

Methods	Kim-CNN	libFM	B-LSTM	DeepFM	mmGRU	CAPER	PERD
P@5	0.0763	0.0798	0.0837	0.1098	0.0916	0.1357	**0.1691**
R@5	0.3816	0.3225	0.1896	0.4201	0.3473	0.5242	**0.6407**
F@5	0.1272	0.1279	0.1161	0.1741	0.1450	0.2148	**0.2676**
P@10	0.0521	0.0590	0.0585	0.0748	0.0601	0.0909	**0.1026**
R@10	0.5211	0.4769	0.2652	0.5725	0.4558	0.6884	**0.7778**
F@10	0.0948	0.1050	0.0959	0.1324	0.1063	0.1606	**0.1814**
N@5	0.2301	0.2566	0.0872	0.3332	0.2835	0.4352	**0.5826**
N@10	0.3022	0.3435	0.1359	0.3833	0.3230	0.4831	**0.6168**

4.2 Performance Evaluation

The performance of PERD is compared with the following baselines: libFM[17], B-LSTM[5], DeepFM[9], mmGRU[20], CAPER[19]. The performance of different methods on Weibo and Twitter datasets are shown in Table 1 and 2. We observe that our approach performs better than all baselines. Compared with CAPER, our approach achieves 33.86% and 27.67% improvement on NDCG@5 and NDCG@10 on Twitter dataset. Different from baseline methods, PERD can recognize important historical tweets and characterize user dynamic preferences with the help of personalized attention. Therefore, our method PRED outperforms these methods. Furthermore, the performance of matrix factorization based methods such as libFM, DeepFM and CAPER is close to the performance of deep learning based methods such as mmGRU and Kim-CNN.

4.3 Ablation Study

To validate the advantage of personalized attention, we compare the performance of PERD and its variants PERD-avg and PRED-. PERD-avg averages the recent historical emoji labels to compute fixed user preferences. PERD- model predict emojis only based on plain tweet text without considering user preferences. According to Table 3, we can observe that PERD and PERD-avg consistently outperform PRED-. Firstly, compared with PRED-, the performance of PERD-avg is greatly improved by using user historical information through a simply average operation. This is probably because user preference is an important factor that affects the choice of emojis. And it can be well captured from user historical tweets. Secondly, our model with personalized attention performs better than its variant method, PERD-avg with average operation. This is probably because emojis are closely related to the semantics and sentiments of text. It leads to that historical tweets with different semantics and sentiments usually contain different informativeness for emoji recommendation. Therefore, it is important to recognize and highlight important historical tweets through attention mechanism in order to learn dynamic user preferences toward different text. The experimental results confirm that dynamic user preferences can achieve more accurate and personalized recommendation compared with fixed user preferences.

[1]https://github.com/Zhengbaibai/PRED

Table 3: Ablation studies on Weibo and Twitter Dataset

	Weibo			Twitter		
Methods	PERD-	PRED-avg	**PERD**	PERD-	PRED-avg	**PERD**
P@5	0.1152	0.1213	**0.1242**	0.1562	0.1591	**0.1691**
R@5	0.4950	0.5211	**0.5337**	0.5916	0.6028	**0.6407**
F@5	0.1870	0.1968	**0.2016**	0.2471	0.2518	**0.2676**
P@10	0.0792	0.0808	**0.0822**	0.0970	0.0984	**0.1026**
R@10	0.6804	0.6945	**0.7061**	0.7355	0.7460	**0.7778**
F@10	0.1419	0.1448	**0.1472**	0.1715	0.1740	**0.1814**
N@5	0.3578	0.3795	**0.3886**	0.5386	0.5481	**0.5826**
N@10	0.4047	0.4227	**0.4307**	0.5767	0.5856	**0.6168**

Table 4: Discussion on the parameter s on Weibo and Twitter Datasets

Dataset	s	P@5	R@5	F@5	P@10	R@10	F@10	N@5	N@10
weibo	1	0.1199	0.5149	0.1945	0.0806	0.6923	0.1444	0.3715	0.4152
	2	0.1203	0.5168	0.1952	0.0804	0.6907	0.1441	0.3734	0.4167
	4	**0.1242**	**0.5337**	**0.2016**	**0.0822**	**0.7061**	**0.1472**	**0.3886**	**0.4307**
Twitter	1	0.1585	0.6006	0.2509	0.0984	0.7455	0.1738	0.5497	0.5846
	2	0.1636	0.6196	0.2588	0.1002	0.7591	0.1770	0.5637	0.5995
	4	**0.1691**	**0.6407**	**0.2676**	**0.1026**	**0.7778**	**0.1814**	**0.5826**	**0.6168**

4.4 Parameter Discussion

We explore the influence of hyperparameter s on the performance of our approach, which is the length of user historical tweet list. The experimental results are shown in Table 4. It can be observed that the performance of our approach consistently rises with the increase of s. And the ascending trend of the model performance does not weaken with the increase of s. This is probably because more user historical tweets can help the model characterize better and comprehensive user preferences. However, as s increasing, the training time and the running memory also rise. Due to these limitations, we only conduct experiments with a maximum s of 4.

5 CONCLUSION

In this paper, we propose a personalized emoji recommendation with dynamic user preference (PERD) model. We design a text encoder based on BERT to learn tweet representations from text. And we propose a personalized attention to recognize the important user historical tweets and characterize user dynamic preferences. The experiments on real-world datasets validate the effectiveness of our approach. In the future work, we try to combine the emoji category prediction task with emoji position prediction task. Due to an observation that the position of emojis is strongly related with the category of emojis. The combination of these two related tasks may be benefit for the recommendation performance.

ACKNOWLEDGMENTS

This work was supported in part by the National Key Research and Development Program, China under Grant 2019YFB2102500, in part by the National Natural Science Foundation of China under Grants 61902309, in part by China Postdoctoral Science Foundation under Grant 2020M683496 and BX20190273, in part by the Humanities and Social Sciences Foundation of Ministry of Education, China under Grant 16XJAZH003, and in part by the Science and Technology Program of Xi'an, China under Grant 21RGZN0017.

REFERENCES

[1] Larisa Alexa, Alina Beatrice Lorent, Daniela Gîfu, and Diana Trandabat. 2018. The Dabblers at SemEval-2018 Task 2: Multilingual Emoji Prediction. In *SemEval@NAACL-HLT*. Association for Computational Linguistics, 405–409.

[2] Francesco Barbieri, Luis Espinosa Anke, José Camacho-Collados, Steven Schockaert, and Horacio Saggion. 2018. Interpretable Emoji Prediction via Label-Wise Attention LSTMs. In *EMNLP*. Association for Computational Linguistics, 4766–4771.

[3] Francesco Barbieri, Luis Espinosa Anke, and Horacio Saggion. 2016. Revealing Patterns of Twitter Emoji Usage in Barcelona and Madrid. In *CCIA (Frontiers in Artificial Intelligence and Applications, Vol. 288)*. IOS Press, 239–244.

[4] Francesco Barbieri, Miguel Ballesteros, Francesco Ronzano, and Horacio Saggion. 2018. Multimodal Emoji Prediction. In *NAACL-HLT (2)*. Association for Computational Linguistics, 679–686.

[5] Francesco Barbieri, Miguel Ballesteros, and Horacio Saggion. 2017. Are Emojis Predictable?. In *EACL (2)*. Association for Computational Linguistics, 105–111.

[6] Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, and Ziqing Yang. 2021. Pre-Training With Whole Word Masking for Chinese BERT. *IEEE ACM Trans. Audio Speech Lang. Process.* 29 (2021), 3504–3514.

[7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL-HLT (1)*. Association for Computational Linguistics, 4171–4186.

[8] Bjarke Felbo, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann. 2017. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In *EMNLP*. Association for Computational Linguistics, 1615–1625.

[9] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. 2017. DeepFM: A Factorization-Machine based Neural Network for CTR Prediction. In *IJCAI*. ijcai.org, 1725–1731.

[10] Douglas Heaven. 2017. Say it with feeling. *New Scientist* 234, 3131 (2017), 43. https://doi.org/10.1016/S0262-4079(17)31225-3

[11] Joongyum Kim, Taesik Gong, Bogoan Kim, Jaeyeon Park, Woojeong Kim, Evey Huang, Kyungsik Han, Juho Kim, JeongGil Ko, and Sung-Ju Lee. 2020. No More One Liners: Bringing Context into Emoji Recommendations. *ACM Trans. Soc. Comput.* 3, 2 (2020), 9:1–9:25.

[12] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *ICLR (Poster)*.

[13] SangEun Lee, Dahye Jeong, and Eunil Park. 2022. MultiEmo: Multi-task framework for emoji prediction. *Knowl. Based Syst.* 242 (2022), 108437.

[14] Fuqiang Lin, Yiping Song, Xingkong Ma, Erxue Min, and Bo Liu. 2021. Sentiment-Aware Emoji Insertion Via Sequence Tagging. *IEEE Multim.* 28, 2 (2021), 40–48.

[15] Weicheng Ma, Ruibo Liu, Lili Wang, and Soroush Vosoughi. 2020. Multi-resolution Annotations for Emoji Prediction. In *EMNLP (1)*. Association for Computational Linguistics, 6684–6694.

[16] Dunlu Peng and Huimin Zhao. 2021. Seq2Emoji: A hybrid sequence generation model for short text emoji prediction. *Knowl. Based Syst.* 214 (2021), 106727.

[17] Steffen Rendle. 2012. Factorization Machines with libFM. *ACM Trans. Intell. Syst. Technol.* 3, 3 (2012), 57:1–57:22.

[18] Chuhan Wu, Fangzhao Wu, Sixing Wu, Yongfeng Huang, and Xing Xie. 2018. Tweet Emoji Prediction Using Hierarchical Model with Attention. In *UbiComp/ISWC Adjunct*. ACM, 1337–1344.

[19] Guoshuai Zhao, Zhidan Liu, Yulu Chao, and Xueming Qian. 2021. CAPER: Context-Aware Personalized Emoji Recommendation. *IEEE Trans. Knowl. Data Eng.* 33, 9 (2021), 3160–3172.

[20] Peijun Zhao, Jia Jia, Yongsheng An, Jie Liang, Lexing Xie, and Jiebo Luo. 2018. Analyzing and Predicting Emoji Usages in Social Media. In *WWW (Companion Volume)*. ACM, 327–334.