

# Hashtags, Emotions, and Comments: A Large-Scale Dataset to Understand Fine-Grained Social Emotions to Online Topics

Keyang Ding   Jing Li\*   Yuji Zhang

Department of Computing, The Hong Kong Polytechnic University, HKSAR, China

{keyang.ding, yu-ji.zhang}@connect.polyu.hk

jing-amelia.li@polyu.edu.hk

## Abstract

This paper studies social emotions to online discussion topics. While most prior work focus on emotions from writers, we investigate readers’ responses and explore the public feelings to an online topic. A large-scale dataset is collected from Chinese microblog Sina Weibo with over 13 thousand trending topics, emotion votes in 24 fine-grained types from massive participants, and user comments to allow context understanding.<sup>1</sup> In experiments, we examine baseline performance to predict a topic’s possible social emotions in a multi-label classification setting. The results show that a seq2seq model with user comment modeling performs the best, even surpassing human prediction. More analyses shed light on the effects of emotion types, topic description lengths, contexts from user comments, and the limited capacity of the existing models.

## 1 Introduction

Social media have become a popular outlet for people to voice opinions, share viewpoints, and exchange ideas. It provides us with rich resource to research public opinions of the trending topics and hear people’s voice over important social events, such as the global COVID-19 crisis. However, our ever-changing physical world leads to the rapid evolution of discussion topics in online world; it is far beyond humans’ capability to catch them in real time. It consequently presents a pressing need for automatic sentiment (Wang et al., 2011; Yang and Eisenstein, 2017) and emotion (Abdul-Mageed and Ungar, 2017; Aragón et al., 2019) analysis.

Nevertheless, most of the related work focus on the feelings from writers (Tang et al., 2014; Huang and Carley, 2019; Singh et al., 2019) and

[H]:#张艺兴整蛊GAI#
[T]: Lay played tricks on GAI.
[E]: 😂: lol; 🙄: facepalm; 🐶: doge (tease).

Figure 1: A Weibo hashtag and its resulting social emotions.  $H$  is the original hashtag in Chinese and  $T$  is its English translation.  $E$  shows the top three emojis voted by online users and their meanings (seperated by “:”).

the existing studies concerning reader emotions mostly tackle well-written texts, such as news reports (Li et al., 2016; Zhou et al., 2018; Yang et al., 2019). Limited work has been done to characterize collective feelings from the public (henceforth **social emotions**) to an online topic described with fragmented and colloquial social media language. Where some previous efforts gather viewpoints from limited readers through user replies (Alhothali and Hoey, 2015; Li et al., 2019) or manual annotations (Buechel and Hahn, 2017; Bostan et al., 2019), we focus on social emotions reflecting aggregated feelings from large amount of people.

In light of these concerns, we present a novel task to infer *social emotions to online topics*. Its goal is to predict the possible emotions from the majority of readers given a few words description of a trending social media topic. The task will benefit various applications, such as topic analysis (Wang et al., 2011) and event detection (Yang et al., 2019). It would help people understand and foresee how the general public thinks of an event even before it becomes a trending topic on social media. More importantly, it is to induce a society’s collective emotions to a discussion topic, which is helpful to gain machines’ affective analysis ability for more appropriate response behaviors and will potentially advance human-computer interactions (Strapparava and Mihalcea, 2007).

To better illustrate our task, Figure 1 displays a topic trended on a popular Chinese microblog

\*Jing Li is the corresponding author.

<sup>1</sup>The dataset can be found at: <https://github.com/polyusmart/HEC-Dataset>

Sina Weibo (henceforth **Weibo**). Here *hashtags* (phrases between two “#”s) are considered as user-annotated discussion topics following the common practice (Wang et al., 2011, 2019). Also shown are the top three reader emotions voted by the online users.<sup>2</sup> As can be seen, the social media style writing — short and informal — makes it challenging to capture what feelings a topic is likely to evoke. The example concerns the story of two celebrities (Lay and GAI) in a variety show, where one needs to make sense of “整蛊” (an internet slang means “play tricks on”) and access some background to get why most voters leaned on hilarious emotions.

As the pilot study for online topics’ social emotions, *we also present the very first dataset for this task, which contains large-scale popular topics, their corresponding social emotions, and user comments for context understanding.* We first gather over 13 thousand Weibo topics (in hashtag forms) associated with fine-grained emotion types represented as 24 emoji labels and contributed by more than 3 thousand online users on average. Compared with the related resource from news websites, our dataset prepared via social media crowd sourcing exhibits larger scale, more annotators, and finer-grained emotion types. Moreover, we collect abundant user comments for each topic, which enables context modeling to access public thoughts. It will later benefit future work to examine how social emotions are shaped in online discussions.

Extensive experiments are carried out on our dataset. We first discuss how the baselines and popular multi-label classification models work to predict the top three reader emotions. The results show that seq2seq-based models obtain the best overall results and user comments can further boost the performance via providing richer contexts. We also find that machines exhibit a superiority on our task compared with humans and point out the possible bias from individuals to sense public emotions. Afterwards, model performances are quantified over varying emotions and hashtag length, where user comments consistently result in the performance gain. Finally, we probe into our output to analyze how user comments help and the limitation of the existing models.

<sup>2</sup>The information comes from a vote on Weibo. It encourages online users to select an emoji from a total of 24 to describe their responsive emotions to a trending hashtag. More details will be discussed later in Section 3.

## 2 Related Work

Our work is related with emoji studies, which mostly focus on how writers will tag emojis to represent the emotions reflected in the texts, e.g., posts, tweets, news (Abdul-Mageed and Ungar, 2017; Barbieri et al., 2018; Demszyk et al., 2020). Different from them, we concern readers’ collective responses to an online topic, which has never been studied before. We are also inspired by topic-oriented sentiment analysis (Wang et al., 2011) concerning writer’s sentiment polarity in positive and negative. Compared with them, we investigate fine-grained public emotions from readers, which is ignored in the existing research and extensively studied here.

As for the analysis of readers’ emotions (a.k.a., affective analysis) (Strapparava and Mihalcea, 2007; Tang and Chen, 2012), some of them collected emotion signals (or emojis) from user replies (Alhothali and Hoey, 2015; Li et al., 2019) or manual annotation (Buechel and Hahn, 2017; Bostan et al., 2019). Responsive emotions are collected from limited readers and hence cannot reflect social emotions from the public. Other studies adopt emotion votes on news websites to gather public feelings on news (Li et al., 2016; Zhou et al., 2018; Yang et al., 2019). However, news reports usually exhibit more formal style than online topics in social media language. None of the above work examine public emotions to online discussion topics, which is the gap filled in this study.

## 3 Data Collection and Analysis

**Data Collection.** Our dataset is built based on a Weibo emotion vote, where it provides users to vote for an emoji from a total of 24 emojis in the form of a questionnaire to represent their feelings to a trending hashtag.

Here comes the steps to collect the data. First, we tracked the trending hashtags following the everyday Weibo topic summary list<sup>3</sup> from Apr to May 2020. For topics trended before that, we resort to a webpage listing the historical hot topics since Nov 2019.<sup>4</sup> Then, we searched and parsed their emotion vote webpage via querying the hashtag in HTTP requests<sup>5</sup> with the `selenium` package.

<sup>3</sup><https://s.weibo.com/top/summary>

<sup>4</sup>[https://github.com/Writeup001/weibo\\_Hot\\_Search](https://github.com/Writeup001/weibo_Hot_Search)

<sup>5</sup><https://m.s.weibo.com/hot/attitude?query=xxx>, where xxx refers to a hashtag name.

Next, the crawled pages were parsed and analyzed using `lxml` package to gather the topics’ emotion voting results. At last, hashtags with less than 100 voters were removed to filter out biased results.

As Weibo only keeps emotions gaining the top three votes, we will hence focus on the top three emotions in the following discussions. These emotions were selected by over 83% voters on average and can still reflect feelings from the majority.

Furthermore, to access the contexts of hashtags, we collected some user comments involved in a hashtag’s discussion. Concretely, we first visited the hashtag page<sup>6</sup> through HTTP requests and obtained the popular Weibo messages that carry the hashtag. Then from the HTML codes, we extracted the IDs of the top four messages. Their comments were later gathered from the messages’ comment pages<sup>7</sup> using an open source toolkit.<sup>8</sup>

Dataset	Size	Len	Voters	Emos
Zhou et al. (2018)	5,586	702.4	157	6
Bostan et al. (2019)	5,000	11.3	331	8
Our dataset	13,766	5.4	3,250	24

Table 1: Statistics: our data vs. prior resource. Size and emos are the # of instances and emotion types. Len and voters are the average # of words (after Chinese word segmentation) and the involved voters per instance.

**Data Analysis.** The statistics of our dataset in comparison with the related resources are shown in Table 1. Both Zhou et al. (2018) and Bostan et al. (2019) present social emotions over news: the former contains Chinese news articles while the latter English news headlines. Our data contains more instances, each with less words, more voters (for annotations), and emotions with finer-grained types. In addition, our dataset presents 408.7 user comments on average for each instance (hashtag), whose average length is 12 Chinese words.

In our dataset, most of the topics are hot events in real life. For example, a few topics collected in the late 2019 concern the social unrest in Hong Kong, while many topics trended in 2020 are about the COVID-19 outbreak. We characterize the topics by keywords and find that 15% topics contain the word “新冠” (COVID-19) while the number rises

<sup>6</sup><https://s.weibo.com/weibo?q=xxx>, where xxx refers to a hashtag name.

<sup>7</sup><https://weibo.cn/comment/hot/MID?&page=PID>, where MID refers the message ID and PID can be changed to turn pages and crawl more comments.

<sup>8</sup>[https://github.com/keyucui/weibo\\_topic\\_analyze](https://github.com/keyucui/weibo_topic_analyze)

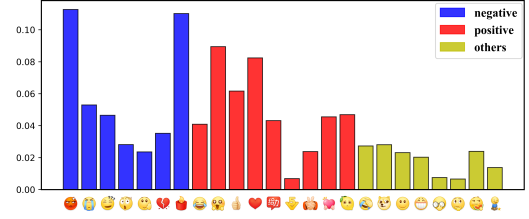


Figure 2: User preferences over varying emotions. X-axis: 24 emotions in emojis; y-axis: proportions of voters who selected the emotion. Emotions are grouped into positive, negative, and others by our interpretations and shown in blue, red, and yellow bars.

to 32% if more relevant keywords are included, such as “医生” (doctor), “口罩” (face mask), and “武汉” (Wuhan). We also examine the relations of the keywords in topics and the responsive emojis voted by users. It is seen that hashtags with “新冠” (COVID-19) are most likely to result in 😭 (cry) while users tend to choose 😏 (tease) to respond to hashtags containing “特朗普” (Trump).

To further analyze the 24 fine-grained emotions, Figure 2 shows how the voter number distribute over varying emotions. This implies the diverse preferences of user voters to pick up varying emotion labels and the challenging label imbalance issue of our task. We also observe that positive and negative emotions are approximately equally distributed, both used more often than emotions in “others” group, probably due to their complicated and vague meanings.

## 4 Experiments and Discussions

### 4.1 Experimental Setup

**Data Preprocessing and Model Setup.** We employed an open-source toolkit `pkuseg` for Chinese word segmentation.<sup>9</sup> For experimental setup, we follow a multi-label classification setting to predict the top three emotions and split the dataset into 85% for training, 5% for validation, and 10% for test. All the non-neural models have hyperparameters tuned on the validation set via grid search and employ features from the pre-trained publicized Chinese embeddings (Li et al., 2018). Neural models are set up following the original papers.

**Comparison Models.** We first consider two weak baselines, one yields random prediction (henceforth `RANDOM`) and the other ranks the emotions by frequency (henceforth `FREQ`). Then,

<sup>9</sup><https://github.com/lancopku/PKUSeg-python>

four non-neural baselines are selected: binary relevance (Boutell et al., 2004) (BR), classifier chain (Read et al., 2011) (CC), multi-label KNN (Zhang and Zhou, 2007) (ML-KNN), and label powerset (Tsoumakas and Katakis, 2007) (LP).

In addition, a popular neural sequence generation model is involved, which is based on seq2seq with a weighted decoder to generate label sequence (henceforth SGM) (Yang et al., 2018). To further exploit contexts from user comments, we consider an extension of SGM that is able to leverage user comments. Our intuition is that readers may voice opinions there with words possibly reflect public emotions. To that end, we first use TextRank (Mihalcea and Tarau, 2004) to extract the 50 keywords from comments. Then we concatenate keywords with the hashtag, feed them both into SGM, and name the new model as SGM+UC.

**Evaluation Metrics.** Here, we adopt three popular metrics from multi-label classification (Qin et al., 2019): label-F1 (average F1 over labels), instance-F1 (average F1 over instances), and hamming loss measured on the predicted and ground-truth label sequences (Koyejo et al., 2015).

## 4.2 Experimental Results

Table 2 shows the main comparison results, where we draw the following observations. First, SGM generally perform better than the non-neural baselines. Its gain on label-F1 is whereas small (even outperformed by LP) indicating its incapability to well handle label imbalance compared with non-neural models. Second, external features from user comment results in the better results from SGM+UC than SGM. We further examine words in hashtags and find only 11.6% on average appear in the sentiment lexicon built based on NTUSD (Ku and Chen, 2007) and HowNet (Yan et al., 2008). It shows that most words in hashtags are not explicit sentiment indicators, rendering the challenge to tackle our task. If additionally considering comments, the number will benefit 49% relative increase (to 17.3%), because readers’ viewpoints exhibited there may narrow the gap between topic description and social emotions.

In addition, we investigate whether an individual’s feelings are always consistent with the public’s. Here 96 test hashtags were sampled and two native Chinese speakers were invited to select the first three emotions occurred to them. The results are also displayed in Table 2 (bottom) together with

	L-F1	I-F1	HL
<b>Weak Baselines</b>			
RANDOM	0.100	0.125	0.219
FREQ	0.064	0.343	0.164
<b>Non-neural Models</b>			
BR	0.254	0.336	0.300
CC	0.227	0.304	0.234
ML-KNN	0.245	0.366	0.158
LP	0.279	0.423	0.144
<b>Neural Models</b>			
SGM	0.260	0.450	0.137
SGM+UC	<b>0.308</b>	<b>0.532</b>	<b>0.117</b>
<b>Human vs. Machine</b>			
HUMAN 1	0.244	0.330	0.168
HUMAN 2	0.192	0.267	0.183
SGM	0.249	0.392	0.152
SGM+UC	<b>0.289</b>	<b>0.465</b>	<b>0.133</b>

Table 2: Comparison results of multi-label classification models to predict the top three emotions. L-F1 (label-F1) and ins-F1 (instance-F1): the higher the better; HL (Hamming Loss): the lower the better.

the performance of SGM and SGM+UC on the same test samples. We find model performance is better than human results, suggesting the possible bias of individual thoughts from public viewpoints. Then, we asked the same annotators to review the hashtags and their emotions from votes, where they thought 75% and 76% voted emotions make sense to them. It means that most social emotions can still be understood by individuals though they may sometimes disagree with the majority.

Another point leading to the disagreement is the prominence of Janus emojis (with two opposite meanings) in social media; for instance, 😊, originally positive has turned out to mean “I don’t care”.

**Results over Varying Emotions.** We then examine how models perform to predict varying emotions and Table 3 shows the results of SGM and SGM+UC.<sup>10</sup> It is first observed that models exhibit diverse F1 over varying emotions, which indicates the difficulty levels vary to understand different emotions. We also notice that user comments boost precision, recall, and F1 for most emotions, which again indicates the usefulness of comments to indicate various social emotions.

**Results over Varying Topic Length.** Next, we explore how hashtag length affects model prediction and show the results from the neural models and LP over varying hashtag length in Figure 3(a). SGM may heavily rely on training data scale and exhibits a performance drop for long hashtags (with

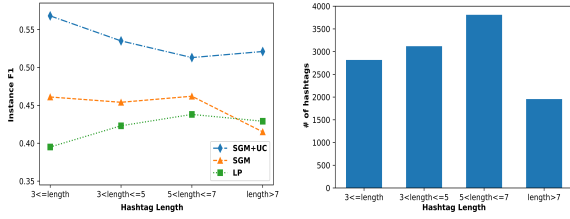
<sup>10</sup>We only consider emotions appearing in the top three of at least 100 test hashtags to avoid bias.



Emotion	Precision	Recall	F1
😏 (onlooker)	.230→.339	.346→.268	.277→.299
😱 (shocked)	.428→.484	<b>.675</b> →.667	.524→.561
😂 (lol)	.347→.398	.339→.471	.343→.432
😡 (angry)	.368→.460	.576→.665	.449→.544
😮 (wow)	.448→.588	.432→.541	.440→.564
🕯️ (candle/RIP)	.506→.578	.336→.448	.404→.505
😭 (cry)	.440→.533	.416→.587	.428→.559
🤦 (facepalm)	.491→.543	.516→.615	.503→.577
👍 (thumbs up)	.565→.611	.490→.552	.525→.580
❤️ (love)	<b>.576</b> → <b>.690</b>	.514→ <b>.731</b>	.543→ <b>.710</b>
🐶 (doge/tease)	.583→.650	.609→.631	<b>.596</b> →.641

Table 3: The prediction results of the SGM model for varying emotions before and after user comments modeling (separated with →).

over 7 words). Because such hashtags have less instances available for training (as shown in Figure 3(b)). Nevertheless, SGM+UC consistently perform better, suggesting that the context from user comments can helpfully alleviate data sparsity.



(a) Length vs. Instance F1 (b) Length vs. Hashtag Count

Figure 3: Instance F1 (left y-axis) in prediction and training hashtag number (right y-axis) over hashtag length (Chinese word count shown in x-axis).

### 4.3 Further Discussions

**Case Study.** As discussed above, it is sometimes difficult to induce readers’ responses from a short and informal hashtag, such as the example in Figure 1. Recall that 整蛊 (*play tricks on*) is essential to predict emotions, while as an uncommon slang, its semantics may not be easy to capture in the limited context. However, keywords in comments, such as 可爱 (*cute*), 搞笑 (*funny*), 调皮 (*naughty*), etc., will contribute to signal the amusements from readers. Without them, SGM only produces common emotions: 😮 (wow), 😱 (shocked), and ❤️ (love).

**Error Analysis.** Here we probe into the model outputs and analyze the errors occur. One major error type comes from the heavy reliance on trending words. For example, most models predict negative feelings when observing 新冠肺炎 (*COVID-19*),

even for hashtags raising good points and touching stories. The other is the incapability to correctly predict some uncommon emotions, such as 😏 (*not easy*). Future work should concern how to handle imbalanced labels in fine-grained social emotions.

## 5 Conclusion

We have investigated social emotions to online discussion topics. A large-scale Chinese Weibo dataset is built containing trending hashtags, emotion votes, and user comments (for context modeling). In experiments, we have shown that the prediction of social emotions is challenging and the modeling of user comments may usefully bridge topic descriptions and public emotions.

## Acknowledges

This work is supported by the Hong Kong Polytechnic University internal funds (1-BE2W and 1-ZVRH) and NSFC Young Scientists Fund 62006203. The authors would like to thank Ruofei Wang (North East University, China) and the anonymous reviewers for the insightful comments.

## References

- Muhammad Abdul-Mageed and Lyle H. Ungar. 2017. Emonet: Fine-grained emotion detection with gated recurrent neural networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, pages 718–728. Association for Computational Linguistics.
- Areej Alhothali and Jesse Hoey. 2015. Good news or bad news: Using affect control theory to analyze readers’ reaction towards news articles. In *NAACL HLT 2015, The 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Denver, Colorado, USA, May 31 - June 5, 2015*, pages 1548–1558. The Association for Computational Linguistics.
- Mario Ezra Aragón, Adrián Pastor López-Monroy, Luis Carlos González-Gurrola, and Manuel Montesy-Gómez. 2019. Detecting depression in social media using fine-grained emotions. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 1481–1486. Association for Computational Linguistics.
- Francesco Barbieri, Luis Espinosa Anke, José Camacho-Collados, Steven Schockaert, and Horacio

- Saggion. 2018. Interpretable emoji prediction via label-wise attention lstms. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pages 4766–4771. Association for Computational Linguistics.
- Laura Ana Maria Bostan, Evgeny Kim, and Roman Klinger. 2019. Goodnewseveryone: A corpus of news headlines annotated with emotions, semantic roles, and reader perception. *arXiv Preprint*, abs/1912.03184.
- Matthew R. Boutell, Jiebo Luo, Xipeng Shen, and Christopher M. Brown. 2004. Learning multi-label scene classification. *Pattern Recognit.*, 37(9):1757–1771.
- Sven Buechel and Udo Hahn. 2017. Readers vs. writers vs. texts: Coping with different perspectives of text understanding in emotion annotation. In *Proceedings of the 11th Linguistic Annotation Workshop, LAW@EACL 2017, Valencia, Spain, April 3, 2017*, pages 1–12. Association for Computational Linguistics.
- Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan S. Cowen, Gaurav Nemade, and Sujith Ravi. 2020. Goemotions: A dataset of fine-grained emotions. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 4040–4054. Association for Computational Linguistics.
- Binxuan Huang and Kathleen M. Carley. 2019. Syntax-aware aspect level sentiment classification with graph attention networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 5468–5476. Association for Computational Linguistics.
- Oluwasanmi Koyejo, Nagarajan Natarajan, Pradeep Ravikumar, and Inderjit S. Dhillon. 2015. Consistent multilabel classification. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, pages 3321–3329.
- Lun-Wei Ku and Hsin-Hsi Chen. 2007. Mining opinions from the web: Beyond relevance retrieval. *J. Assoc. Inf. Sci. Technol.*, 58(12):1838–1850.
- Shen Li, Zhe Zhao, Renfen Hu, Wensi Li, Tao Liu, and Xiaoyong Du. 2018. Analogical reasoning on chinese morphological and semantic relations. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 2: Short Papers*, pages 138–143. Association for Computational Linguistics.
- Shoushan Li, Jian Xu, Dong Zhang, and Guodong Zhou. 2016. Two-view label propagation to semi-supervised reader emotion classification. In *COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11-16, 2016, Osaka, Japan*, pages 2647–2655. ACL.
- Xintong Li, Qinke Peng, Zhi Sun, Ling Chai, and Ying Wang. 2019. Predicting social emotions from readers’ perspective. *IEEE Trans. Affect. Comput.*, 10(2):255–264.
- Rada Mihalcea and Paul Tarau. 2004. Textrank: Bringing order into text. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, EMNLP 2004, A meeting of SIGDAT, a Special Interest Group of the ACL, held in conjunction with ACL 2004, 25-26 July 2004, Barcelona, Spain*, pages 404–411. ACL.
- Kechen Qin, Cheng Li, Virgil Pavlu, and Javed A. Aslam. 2019. Adapting RNN sequence prediction model to multi-label set prediction. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 3181–3190. Association for Computational Linguistics.
- Jesse Read, Bernhard Pfahringer, Geoff Holmes, and Eibe Frank. 2011. Classifier chains for multi-label classification. *Mach. Learn.*, 85(3):333–359.
- Abhishek Singh, Eduardo Blanco, and Wei Jin. 2019. Incorporating emoji descriptions improves tweet classification. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 2096–2101. Association for Computational Linguistics.
- Carlo Strapparava and Rada Mihalcea. 2007. Semeval-2007 task 14: Affective text. In *Proceedings of the 4th International Workshop on Semantic Evaluations, SemEval@ACL 2007, Prague, Czech Republic, June 23-24, 2007*, pages 70–74. The Association for Computer Linguistics.
- Duyu Tang, Furu Wei, Nan Yang, Ming Zhou, Ting Liu, and Bing Qin. 2014. Learning sentiment-specific word embedding for twitter sentiment classification. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014, June 22-27, 2014, Baltimore, MD, USA, Volume 1: Long Papers*, pages 1555–1565. The Association for Computer Linguistics.
- Yi-jie Tang and Hsin-Hsi Chen. 2012. Mining sentiment words from microblogs for predicting writer-reader emotion transition. In *Proceedings of the Eighth International Conference on Language*

- Resources and Evaluation, LREC 2012, Istanbul, Turkey, May 23-25, 2012*, pages 1226–1229. European Language Resources Association (ELRA).
- Grigorios Tsoumakas and Ioannis Katakis. 2007. Multi-label classification: An overview. *IJDWM*, 3(3):1–13.
- Xiaolong Wang, Furu Wei, Xiaohua Liu, Ming Zhou, and Ming Zhang. 2011. Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach. In *Proceedings of the 20th ACM Conference on Information and Knowledge Management, CIKM 2011, Glasgow, United Kingdom, October 24-28, 2011*, pages 1031–1040. ACM.
- Yue Wang, Jing Li, Hou Pong Chan, Irwin King, Michael R. Lyu, and Shuming Shi. 2019. Topic-aware neural keyphrase generation for social media language. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 2516–2526. Association for Computational Linguistics.
- Jiajun Yan, David B. Bracewell, Fuji Ren, and Shingo Kuroiwa. 2008. The creation of a chinese emotion ontology based on hownet. *Engineering Letters*, 16(1):166–171.
- Pengcheng Yang, Xu Sun, Wei Li, Shuming Ma, Wei Wu, and Houfeng Wang. 2018. SGM: sequence generation model for multi-label classification. In *Proceedings of the 27th International Conference on Computational Linguistics, COLING 2018, Santa Fe, New Mexico, USA, August 20-26, 2018*, pages 3915–3926. Association for Computational Linguistics.
- Yang Yang, Deyu Zhou, Yulan He, and Meng Zhang. 2019. Interpretable relevant emotion ranking with event-driven attention. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 177–187. Association for Computational Linguistics.
- Yi Yang and Jacob Eisenstein. 2017. Overcoming language variation in sentiment analysis with social attention. *Trans. Assoc. Comput. Linguistics*, 5:295–307.
- Min-Ling Zhang and Zhi-Hua Zhou. 2007. ML-KNN: A lazy learning approach to multi-label learning. *Pattern Recognit.*, 40(7):2038–2048.
- Deyu Zhou, Yang Yang, and Yulan He. 2018. Relevant emotion ranking from text constrained with emotion relationships. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 1 (Long Papers)*, pages 561–571. Association for Computational Linguistics.