A Topic-Aware Graph-Based Neural Network for User Interest Summarization and Item Recommendation in Social Media

Conference Paper · April 2023

DOI: 10.1007/978-3-031-30672-3_36

A Topic-Aware Graph-Based Neural Network for User Interest Summarization and Item Recommendation in Social Media

Junyang Chen[1,2(✉)], Ge Fan[3], Zhiguo Gong[4], Xueliang Li[1,5(✉)], Victor C. M. Leung[1], Mengzhu Wang[1], and Ming Yang[4]

[1] College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China
{junyangchen,lixueliang}@szu.edu.cn
[2] Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ), Shenzhen, China
[3] Tencent Inc., Shenzhen, China
[4] State Key Laboratory of Internet of Things for Smart City, Department of Computer Information Science, University of Macau, Macau, China
[5] National Engineering Laboratory for Big Data System Computing Technology, Shenzhen University, Shenzhen, China

Abstract. User-generated content is daily produced in social media, as such user interest summarization is critical to distill salient information from massive information. While the interested messages (e.g., tags or posts) from a single user are usually sparse becoming a bottleneck for existing methods, we propose a topic-aware graph-based neural interest summarization method (UGraphNet), enhancing user semantic mining by unearthing potential user relations and jointly learning the latent topic representations of posts that facilitates user interest learning. Experiments on two datasets collected from well-known social media platforms demonstrate the superior performance of our model in the tasks of user interest summarization and item recommendation. Further discussions also show that exploiting the latent topic representations and user relations are conductive to the user automatic language understanding.

1 Introduction

Social networking platforms enjoy a great popularity for individuals to share opinions and exchanging information [4,5]. Then, millions of user-generated posts are produced

This work was supported by National Natural Science Foundation of China under Grant No. 62102265, by the Open Research Fund from Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ) under Grant No. GML-KF-22-29, by the Natural Science Foundation of Guangdong Province of China under Grant No. 2022A1515011474, by the Science and Technology Development Fund, Macau SAR, China, under Grant No. (0068/2020/AGJ, SKL-IOTSC(UM)-2021-2023), and by Shenzhen Talents Special Project - Guangdong Provincial Innovation and Entrepreneurship Team Supporting Project under Grant No. 2021344612).
X. Li—unique corresponding author

X. Wang et al. (Eds.): DASFAA 2023, LNCS 13944, pp. 537–546, 2023.
https://doi.org/10.1007/978-3-031-30672-3_36

Table 1. POI: "The Vortex Bar And Grill" on Yelp. Posts are the short comments from different users for the POI. User interest summarization is the succinct description for these posts.

POI: The Vortex Bar And Grill - Midtown
Source posts from social media:
User 1: Best burgers in town.
User 2: The Yokohama Mama Burger for some Lunch!
User 3: Steakhouse Burger is one of the best. Also, be sure to get onion rings as a side!
User 4: Yummy beer, and the beers aren't expensive either! Comedy show!
User 5: Adding my vote as the best burger in ATL.
User 6: They have yamazaki whiskey.
User interest summarization: American (Traditional), Burgers, Restaurants, Bars, Nightlife

daily, far outpacing the human being's reading and understanding capacity. As such, discovering gist information from large volume of posts becomes vital capability for current applications. Two characteristics can be summarized for current posts: short texts and word sparsity. As shown in Table 1, the source posts of one point-of-interest (POI) from social media (e.g., Yelp) usually only contain a few short sentences and their word co-occurrences are usually sparse. The user interest summarization can be considered as the keyphrases [22] to summarize the whole posts of a POI. These keyphrases can be further used for the downstream tasks, such as similar POI search [2,7], user sentiment analysis [6,20], POI recommendation [18,19], and so forth. Despite the widespread use of keyphrases, millions of posts are generated daily without summarization. Therefore, there exists a pressing need for automating user interest summarization for daily posts.

Most previous work focuses on extracting existed phrases from target posts. [20, 24] employ topic models to generate topical words as the keyphrases for a group of posts. These methods, ascribed to the limitation of most topic models, are incapable of generating non-existed keyphrases for each target post. More recently, [22] introduces a sequence generation framework that can generate keyphrases beyond the target post. It present a neural seq2seq model based on integrating more tweets related to the target post to generate keyphrases in a word-by-word training manner.

However, the aforementioned models encounter a common challenge: only limited number of relevant posts existed for one POI are encoded when processing posts from social media. To illustrate this challenge, we display Table 1 where a batch of posts are from Yelp. Such posts are the comments related to the POI "The Vortex Bar And Grill - Midtown". We can observe that each post only contains a few words, as such, this will inevitably encounter the sparsity problem. On way is to combine more relevant posts [22] to enrich the contents. However, even though all posts are focused on the same topic, it is difficult to summarize the keyphrases "American (Traditional), Burgers, Restaurants, Bars, Nightlife" due to the limited number and colloquial nature of social media language by looking at posts from 1 to 6 in Table 1.

To address the above challenges, we propose a novel graph-based neural interest summarization model (UGraphNet) that includes three complementary innovations. The first one is *user collaboration* that leverages neighboring information by construct the bipartite graph of user-post-user to enrich sparse contents. The second one is corpus-level *user latent topic modeling* with the constructed graph and the users interested posts. The last one is joint modeling the *latent topic embedding* of all users and the interest prediction of the target users. These approaches can effectively improve the accuracy and alleviate data sparsity in the tasks of user interest summarization and item recommendation. In general, the contributions of this work are as follows:

– To the best of our knowledge, our work is the first to study the benefit of leveraging user relations and latent topics on social media interest summarization and item recommendation. Also, our model enables an end-to-end training process.
– We propose there main components: a contrastive learning loss, a topic modeling loss, and a graph-based learning loss to achieve the above purposes by their jointly learning.
– We experiment on two newly constructed social media datasets. Our model can significantly outperform all the comparison methods. Ablation analysis also demonstrates the effectiveness of exploiting the latent topic representations and user relations in user automatic language understanding.

2 Proposed Model

In this section, we describe the proposed framework that how to leverage user collaboration and latent topics for the user interest summarization. Figure 1 shows the overall architecture consisting of three modules - a contrastive learning loss, a topic modeling loss, and a graph-based generative learning loss. Formally, given a collection D of social media posts, we process each post into bags-of-words word vector $[t_1, t_2, ..., t_{|V|}]$, which is a V-dim vector over the vocabulary and V denotes its size. Besides, each post consists of latent topics and we denote the topic size as $|K|$. Below we first introduce our three modules and then describe how they are jointly trained.

2.1 Contrastive Learning Loss

We exploit user collaboration by constructing the adjacent graph of users. Specifically, when two users are interested in the same posts, we make a connection between these users. Besides, it is difficult for every user has an unique embedding in the large-scale scenarios, which will inevitably make the number of parameters tremendous. As such, inspired by the work [8] that they represent the users by the terms of queries, we represent the users with a smaller number of tag embeddings. In other word, each user can be represented with a limited number of tags. Here we also use one-hot encoding to represent the tag (or call word) lexicon $(t_1, ..., t_{|V|})$. Then, we map the tags to d-dimensional vectors with a mapping function f to represent users as follows:

$$\mathbf{h}_{v_t} = f((t_1, ..., t_{|V|}), \mathbf{M}), \tag{1}$$

where $\mathbf{h}_{v_t} \in \mathbb{R}^d$ denotes the embedding of a user v_t, and $\mathbf{M} \in \mathbb{R}^{|V| \times d}$ is the transformation matrix. After that, we adopt an attention method to fuse the information of a

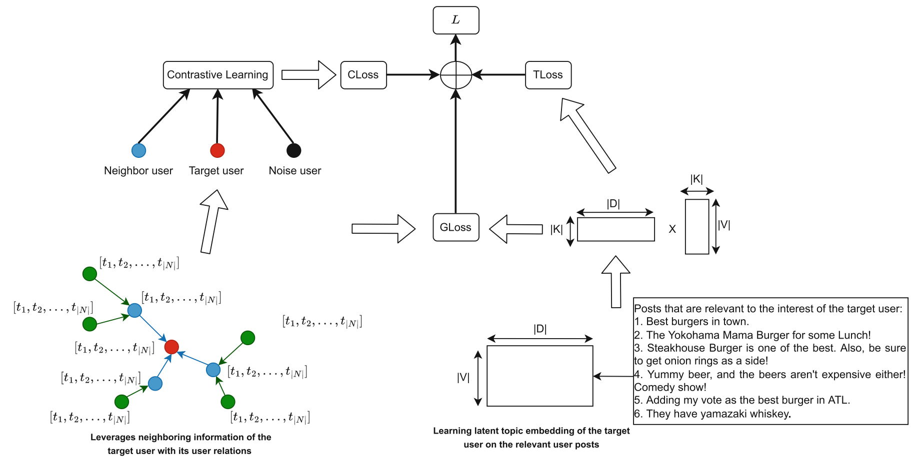

Fig. 1. Overview of the proposed UGraphNet model.

target user and its neighbors. First we perform the message propagation step for dealing with the messages passing from neighboring nodes, which is given by:

$$\boldsymbol{m}_{v_i \leftarrow v_j} = \mathrm{MLP}(n_{v_j v_i} \oplus \boldsymbol{h}_{v_j}) \cdot \boldsymbol{h}_{v_j}, \tag{2}$$

where $\boldsymbol{m}_{v_i \leftarrow v_j} \in \mathbb{R}^d$ denotes the information passing from node v_j to v_i, $n_{v_j v_i}$ is one-hot encoded of the neighbor type (e.g., one-hop (0, 1) or multi-hop neighbors (1, 0)), $\mathrm{MLP}(\cdot) \in \mathbb{R}^{d \times d}$ denotes a Multi-Layer Perception that takes as inputs both the neighbor type $n_{v_j v_i}$ and the representations of the user $\boldsymbol{h}_{v_j}$, and $\oplus$ represents the concatenation.

Then, we aggregate the information of the target node and the messages passing from its neighbors in an attention way. The weight coefficient α_{v_i,v_j} between two nodes can be formulated by:

$$\alpha_{v_i,v_j} = \frac{\exp\Big(\sigma(\mathbf{a}^T \cdot [\boldsymbol{W}\boldsymbol{h}_{v_i} || \boldsymbol{W}\boldsymbol{m}_{v_i \leftarrow v_j}])\Big)}{\sum_{v_k \in \mathcal{N}_{v_i}} \exp\Big(\sigma(\mathbf{a}^T \cdot [\boldsymbol{W}\boldsymbol{h}_{v_i} || \boldsymbol{W}\boldsymbol{m}_{v_i \leftarrow v_k}])\Big)}, \tag{3}$$

where $\boldsymbol{W} \in \mathbb{R}^{d \times d}$ is a shared weight matrix for mapping nodes into the same embedding space, $\mathbf{a} \in \mathbb{R}^{2d}$ denotes a weight vector for learning the relations of the target node and its neighbors, and $\mathcal{N}_{v_i}$ is the set of neighbors of node v_i, and σ denotes the sigmoid function [11].

After that, with the learned weight coefficients α_{v_i,v_j} and the neighboring message information $\boldsymbol{m}_{v_i \leftarrow v_j}$, the final representations of node v_i can be formulated by:

$$\mathbf{h}_{v_t}^L = \mathrm{ReLU}\Bigg(\sum_{v_j \in \mathcal{N}_{v_i}} \alpha_{v_i,v_j} \mathbf{W}\mathbf{m}_{v_i \leftarrow v_j}\Bigg), \tag{4}$$

where ReLU is an activation function [9], and L denotes the last layer of the network.

Finally, inspired by the recent advances in the contrastive learning work [12,13], we introduce a contrastive learning loss $\mathcal{L}_c$ formulated by:

$$\mathcal{L}_c = \sum_{(v_t, v_p, v_n) \in \mathcal{T}} [\sigma(v_t, v_p; \boldsymbol{h}) - \sigma(v_t, v_n; \boldsymbol{h}) + \nabla]_+, \tag{5}$$

where $\boldsymbol{h}$ denotes the hidden embeddings of users, v_t is the target user, v_p denotes its neighbor users, v_n is the negative users drawn from the whole set by using the alias table method [15] that only takes $O(1)$ time, ∇ is a margin hyper-parameter separating the positive pair and the corresponding negative one (we set it as 0.5 in the experiments), $\mathcal{T}$ denotes a training batch, and $[\cdot]_+$ denotes the positive part of the calculation. The above contrastive learning loss (Eq. (5)) explicitly encodes similarity ranking among node pairs into the embedding vectors.

2.2 Topic Modeling Loss

In this part, we refer to a matrix factorization [1] method to obtain the topic modeling loss $\mathcal{L}_t$. More concretely, given the document-word matrix $\mathbf{D}$, we decompose it into the product of the document-topic embedding matrix $\mathbf{\Theta}$ and the topic-word embedding matrix $\mathbf{T}$ with regularization as follows:

$$\mathcal{L}_t = \sum_{i \in \mathcal{T}} (\mathbf{D}_i - \mathbf{\Theta}_i \mathbf{T})^2 + \lambda(||\mathbf{\Theta}_i||_2^2 + ||\mathbf{T}||_2^2), \tag{6}$$

where $\mathbf{D} \in \mathbb{R}^{|D| \times V}$, D denotes the set of documents, V is the vocabulary size, $\mathbf{\Theta} \in \mathbb{R}^{|D| \times k}$, $\mathbf{T} \in \mathbb{R}^{k \times V}$, k is the dimension of the topic embedding, $|| \cdot ||_2^2$ is the l_2 norm regularization of the parameters, and λ is a harmonic factor for regularization. In Eq. (6), we explore the latent topics of the posts that are interested by the target user. Besides, the obtained document-topic embedding $\mathbf{\Theta}$ will be used in the generative learning in the following section.

2.3 Generative Learning Loss

With the target user embedding $\mathbf{h}_{v_t}^L$ from Eq. (4) that represents the user collaboration information, and the document-topic embedding $\mathbf{\Theta}_{v_t}$ from Eq. (6) that represents the interests of the target user, we can construct the generative learning loss $\mathcal{L}_g$ as follows:

$$\mathcal{L}_g = -\sum_{v_t \in \mathcal{T}} \log(\sigma([\mathbf{h}_{v_t}^L; \mathbf{\Theta}_{v_t}]\mathbf{W}_v)), \tag{7}$$

where $\mathbf{h}_{v_t}^L \in \mathbb{R}^{1 \times d}$, $\mathbf{\Theta}_{v_t} \in \mathbb{R}^{1 \times k}$, $\mathbf{W}_v \in \mathbb{R}^{(d+k) \times 1}$ are trainable weights, and $[;]$ denotes the concatenation operation. In Eq. (7), we aim to fuse the information of the two domains (i.e., the user relations and the interested latent topics) which exploits the assumption that relevant users may share similar interests.

2.4 Learning and Inference

In the training stage, we adopt stochastic gradient descent [14] to minimize the loss function of the total loss, which is given by:

$$\mathcal{L}_{total} = \mathcal{L}_c + \mathcal{L}_t + \mathcal{L}_g. \tag{8}$$

Table 2. The statistics of datasets.

Datasets	Delicious	Yelp
#Users	1847	7913
#Items	68755	12462
Avg. items interacted by per user	195.72	41.52
Avg. length of user summarization per item	3.67	6.36
Avg. length of description per item	7.09	12.15

With the above learning objective as shown in the Eq. (8), we can: (1) exploit the user collaboration information with the contrastive learning loss (Eq. (5)), (2) explore the latent topics of the semantic information to summarize user interests (Eq. (6)), (3) fuse the above information (Eq. (7)) to simultaneously learn them in an end-to-end way.

User Interest Inference: Based on the concatenated embedding of user collaborative information $\mathbf{h}_{v_t}^{L}$ and user historical interest information $\mathbf{\Theta}_{v_t}$, we can conduct dot product with the topic-word embedding $\mathbf{T}$ to generate a ranking list of output words, where the top K ones serve as the user interest summarization in the evaluation.

Post Recommendation Inference: Similarly, based on the $\mathbf{h}_{v_t}$ and $\mathbf{\Theta}_{v_t}^{L}$ of the target user, we generate a ranking list with the document-topic embedding $\mathbf{\Theta}$ of the output posts, where the top N ones serve as the post recommendation.

3 Experiments

3.1 Datasets

We adopt two real-world datasets to estimate the performance: *Delicious*[1] and *Yelp*[2] which are widely used in social recommendation [10,17]. The statistics of the datasets are shown in Table 2. Each dataset contains of users, items, the interactions including browse or access between users and items, user summarization of items, and item description. The "Avg. items interacted by per user" represents the average number of items that have been browsed or visited by users before. The "Avg. length of user summarization per item" denotes the average length of words that users summarize items. The "Avg. length of description per item" denotes the average length of words that are used to comprehensively describe the characters of items.

3.2 Comparison Methods

We include several traditional and state-of-the-art approaches that can be applied to user interest summarization, including probabilistic graph models and sequential learning models. Here are descriptions of selected methods: **GSDMM** [23] is a traditional and widely used probabilistic graph model which is designed for the short text modeling. The word and document representations are learned by combining Dirichlet and multinomial distributions. **DP-BMM** [5] is an another often used probabilistic graph model

[1] https://grouplens.org/datasets/hetrec-2011/.
[2] https://www.yelp.com/dataset.

Table 3. Main comparison results displayed with scores in %. Boldface scores in each column indicate the best results. The underlined scores denote the second best performance. Our model outperforms the strongest baselines with p-value < 0.05.

Model	Delicious				Yelp			
	HR@1	HR@5	HR@10	MAP	HR@1	HR@5	HR@10	MAP
Traditional models								
GSDMM	14.83	12.02	10.05	13.83	1.67	14.34	7.40	7.93
DP-BMM	16.56	14.75	14.38	17.15	4.16	17.68	11.12	12.41
State of the arts								
SEQ-TAG	20.16	22.03	21.79	21.03	12.21	27.49	21.62	20.56
SEQ2SEQ-CORR	23.21	29.59	24.27	23.10	15.17	32.70	22.31	22.96
TAKG	27.68	30.96	28.84	27.76	42.27	33.22	23.93	38.37
UGraphNet (Ours)	**34.56**	**35.51**	**33.07**	**34.63**	**55.99**	**35.09**	**25.30**	**48.20**
Improv.	24.86%	14.70%	14.67%	24.75%	32.46%	5.63%	5.73%	25.62%

which explicitly exploits the word-pairs constructed from each document to enhance the word co-occurrence pattern in short texts. It can deal with the topic drift problem of short text streams naturally. **SEQ-TAG** [24] is a state-of-the-art deep recurrent neural network model that can combines keywords and context information to automatically extract keyphrases from short texts. **SEQ2SEQ-CORR** [3] exploits a sequence-to-sequence (seq2seq) architecture for keyphrase generation which captures correlation among multiple keyphrases in an end-to-end fashion. **TAKG** [21] introduces a seq2seq based neural keyphrase generation framework that takes advantage of the recent advance of neural topic models [16] to enable end-to-end training of latent topic modeling and keyphrase generation.

Different from the above methods, we exploit the potential usefulness of user collaboration and the latent topics exhibited in the user interest and the item contents, which have been ignored in previous research and will be extensively studied here. We also present an ablation study to show the effectiveness of our proposed components.

3.3 User Interest Summarization Results

In this section, we examine our performance in user interest summarization for social media. The performance of the user summarization is accessed by calculating how many "hits" in an n-sized list of ranked words. To this end, we use popular information retrieval metrics *Hit Ratio* (HR) and *Mean Average Precision* (MAP) for evaluation. For the datasets Delicious and Yelp, most items are summarized by users with 3 to 6 on average (Table 2), thus HR@1, HR@5, HR@10 are reported. Besides, MAP is measured over the top 10 prediction for all datasets.

The main comparison results are shown in Table 3, where the highest scores are highlighted in boldface and the underlined ones denote the second best. The last row is the improvements of our method compared with the best baseline. In general, we can observe that:

Table 4. Ablation analysis. The highest scores are marked in boldface, and '_' denotes the second-best results.

Method	Delicious				Yelp			
	HR@1	HR@5	HR@10	MAP	HR@1	HR@5	HR@10	MAP
UGraphNet	**34.56**	**35.51**	**33.07**	**34.63**	**55.99**	**35.09**	**25.30**	**48.20**
w/o CLoss	26.36	12.30	6.88	10.80	<u>53.02</u>	<u>33.56</u>	<u>22.23</u>	<u>46.99</u>
w/o TLoss	4.92	5.63	4.44	5.28	52.07	25.79	15.81	14.29
w/o GLoss	<u>28.64</u>	<u>27.22</u>	<u>26.44</u>	<u>24.76</u>	12.54	11.15	9.71	3.25
Improv.	20.67%	30.46%	25.08%	39.86%	5.60%	4.56%	13.81%	2.58%

(1) Our model *UGraphNet* consistently outperforms other comparisons on all datasets under various metrics. This shows the usefulness of leveraging user neighboring information for their interest summarization. More concretely, *UGraphNet* achieves up to 24.86%, 14.70%, 14.67%, and 24.75% improvements over the second-best method *TAKG* in terms of HR@1, HR@5, HR@10 and MAP on Delicious. Besides, *TAKG* gains 32.46%, 5.63%, 5.73%, and 25.62% improvements on average against the second ones on Yelp. In general, the above improvements demonstrate the effectiveness of our method by jointly modeling user relations and user interests.

(2) Among the results of the baselines, the traditional methods including *GSDMM* and *DP-BMM* give poor performance. This indicates that user interest summarization is a challenging task. It is hard to rely on probabilistic graphical models to yield acceptable performance. On the contrary, seq2seq-based models consisting of *SEQ-TAG*, *SEQ2SEQ-CORR*, and *TAKG* yield better results than the traditional ones. Particularly, *TAKG* outperforms the other baselines, which suggests the helpful of exploiting latent topics in short texts. Interestingly, our model achieve larger improvements with a step further by exploring the user relations and their latent topics.

3.4 Ablation Analysis

To analyze the effectiveness of the proposed components on user interest summarization (introduced in Sect. 2) in our method, we conduct an ablation analysis as follows. In general, we have three ablated variants of our model:

I. w/o CLoss (without constrastive learning loss): The CLoss (Eq. (5)) is used to exploit user relations that help to distinguish the target user from its neighboring users and negative users. We remove the CLoss and keep the TLoss and GLoss for comparison.

II. w/o TLoss (without topic modeling loss): The TLoss (Eq. (6)) aims to exploit the latent topics in short texts which can alleviate the data sparsity in the user interest summarization.

III. w/o GLoss (without generative learning loss): The TLoss (Eq. (7)) utilizes the assumption that relevant users share similar interests. We adopt it to generate keyphrases that is relevant to users' latent topics.

The results of the ablation tests are shown in Table 4. Our method *UGraphNet* outperforms the other variants. Specifically, *UGraphNet* achieves 20.67%, 30.46%, 25.08%, and 39.86% improvements over the second-best variant in terms of HR@1, HR@5, HR@10, and MAP on Delicious, and obtains 5.60%, 4.56%, 13.81%, and 2.58% gains on Yelp, respectively. These results validate that the user attention update gate is more appropriate to explore user interests. These results demonstrates the effectiveness of jointly learning different components. We observe that the influence levels of each component are presented as GLoss > CLoss > TLoss on Delicious. These results are identic with expectation that the generative loss contributes more than the topic modeling loss. This is because there are more dense interactions between users and items (i.e., Avg. items interacted by per user: 195.72) as shown in the statistics of datasets (Table 2), comparing with the sparsity of text information (i.e., Avg. length of user summarization per item (3.67)). Interestingly, the influence levels of components on Yelp are as: CLoss > TLoss > GLoss, which shows that the contrastive learning loss contributes the most while the generative loss contributes the lest. These results are also in accord with the characteristics in Table 2. There are more users information (#User: 7913) while less interactions between users and items (Avg. items interacted by per user: 41.52) in Yelp. Note that the above interesting points are observed by comparing results across the statistics of the datasets.

4 Conclusion

In general, we propose a topic-aware graph-based neural interest summarization method, called UGraphNet, that can enhance user semantic mining for user interest summarization and item recommendation in social media. The main innovations of our work include a contrastive learning loss, a topic modeling loss, and a graph-based learning loss that leverage user relations and latent topics on social media by jointly training. Experiments on two newly constructed social media datasets demonstrate that our model can significantly outperform all the comparison methods. Ablation analysis is also conducted to show the superiority of our proposed components.

References

1. Ahmed, N.K., et al.: Learning role-based graph embeddings. arXiv preprint arXiv:1802.02896 (2018)
2. Bansal, P., Jain, S., Varma, V.: Towards semantic retrieval of hashtags in microblogs. In: Proceedings of the 24th International Conference on World Wide Web, pp. 7–8 (2015)
3. Chen, J., Zhang, X., Wu, Y., Yan, Z., Li, Z.: Keyphrase generation with correlation constraints. arXiv preprint arXiv:1808.07185 (2018)
4. Chen, J., Gong, Z., Liu, W.: A nonparametric model for online topic discovery with word embeddings. Inf. Sci. **504**, 32–47 (2019)
5. Chen, J., Gong, Z., Liu, W.: A dirichlet process biterm-based mixture model for short text stream clustering. Appl. Intell. **50**(5), 1609–1619 (2020)
6. Davidov, D., Tsur, O., Rappoport, A.: Enhanced sentiment learning using twitter hashtags and smileys. In: Coling 2010: Posters, pp. 241–249 (2010)

7. Efron, M.: Hashtag retrieval in a microblogging environment. In: Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 787–788 (2010)
8. Fan, S., et al.: Metapath-guided heterogeneous graph neural network for intent recommendation. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 2478–2486 (2019)
9. Glorot, X., Bordes, A., Bengio, Y.: Deep sparse rectifier neural networks. In: Proceedings of the 14th International Conference on Artificial Intelligence and Statistics, pp. 315–323 (2011)
10. Guo, Z., Wang, H.: A deep graph neural network-based mechanism for social recommendations. IEEE Trans. Industr. Inf. **17**(4), 2776–2783 (2020)
11. Han, J., Moraga, C.: The influence of the sigmoid function parameters on the speed of backpropagation learning. In: Mira, J., Sandoval, F. (eds.) IWANN 1995. LNCS, vol. 930, pp. 195–201. Springer, Heidelberg (1995). https://doi.org/10.1007/3-540-59497-3_175
12. Jaiswal, A., Babu, A.R., Zadeh, M.Z., Banerjee, D., Makedon, F.: A survey on contrastive self-supervised learning. Technologies **9**(1), 2 (2020)
13. Khan, A., AlBarri, S., Manzoor, M.A.: Contrastive self-supervised learning: a survey on different architectures. In: 2022 2nd International Conference on Artificial Intelligence (ICAI), pp. 1–6. IEEE (2022)
14. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
15. Li, A.Q., Ahmed, A., Ravi, S., Smola, A.J.: Reducing the sampling complexity of topic models. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 891–900 (2014)
16. Miao, Y., Grefenstette, E., Blunsom, P.: Discovering discrete latent topics with neural variational inference. In: International Conference on Machine Learning, pp. 2410–2419. PMLR (2017)
17. Wang, H., Wu, Q., Wang, H.: Factorization bandits for interactive recommendation. In: 31st AAAI Conference on Artificial Intelligence (2017)
18. Wang, W., Chen, J., Wang, J., Chen, J., Gong, Z.: Geography-aware inductive matrix completion for personalized point-of-interest recommendation in smart cities. IEEE Internet Things J. **7**(5), 4361–4370 (2019)
19. Wang, W., Chen, J., Wang, J., Chen, J., Liu, J., Gong, Z.: Trust-enhanced collaborative filtering for personalized point of interests recommendation. IEEE Trans. Industr. Inf. **16**(9), 6124–6132 (2019)
20. Wang, X., Wei, F., Liu, X., Zhou, M., Zhang, M.: Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, pp. 1031–1040 (2011)
21. Wang, Y., Li, J., Chan, H.P., King, I., Lyu, M.R., Shi, S.: Topic-aware neural keyphrase generation for social media language. arXiv preprint arXiv:1906.03889 (2019)
22. Wang, Y., Li, J., King, I., Lyu, M.R., Shi, S.: Microblog hashtag generation via encoding conversation contexts. arXiv preprint arXiv:1905.07584 (2019)
23. Yin, J., Wang, J.: A dirichlet multinomial mixture model-based approach for short text clustering. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 233–242. ACM (2014)
24. Zhang, Q., Wang, Y., Gong, Y., Huang, X.J.: Keyphrase extraction using deep recurrent neural networks on Twitter. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 836–845 (2016)