Data Science and Engineering
https://doi.org/10.1007/s41019-023-00225-8

RESEARCH PAPER

A Neural Inference of User Social Interest for Item Recommendation

Junyang Chen[1,2] · Ziyi Chen[3] · Mengzhu Wang[1] · Ge Fan[3] · Guo Zhong[4] · Ou Liu[5] · Wenfeng Du[1] · Zhenghua Xu[6] · Zhiguo Gong[7]

Received: 13 May 2023 / Revised: 18 July 2023 / Accepted: 8 August 2023

Abstract
User-generated content is daily produced in social media, as such user interest summarization is critical to distill salient information from massive information for recommendation tasks. 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 neural inference method (*NIGraph-Net*) by mining user social interest for item recommendation. It can unearth user latent topics combined with user relation learning. Specifically, we exploit a neural variational inference approach to learn the distributions between user interests and hidden topics. (We denote it as interest-topic distributions in the following.) Then, we adopt a unified graph-based training loss that jointly learns the hidden topics and user relations for item recommendation. 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 is conducive to the user's automatic language understanding.

Keywords Neural variational inference · User interest summarization · Item recommendation

1 Introduction

Social networking platforms have become immensely popular among individuals for sharing opinions and exchanging information [1, 2]. However, the sheer volume of user-generated posts produced daily on these platforms far surpasses human beings' reading and comprehension abilities. Therefore, the ability to extract essential information from a large volume of posts has become a critical capability for current applications.

Two main 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 in which their word co-occurrences are usually sparse. To extract the key points of these sentences, the keyphrases generation [3] can be adopted to summarize the whole posts of a POI. These generated keyphrases can be further used in the downstream tasks, such as similar POI search [4, 5], user sentiment analysis [6, 7], and POI recommendation [8, 9].

However, most previous work focuses on extracting existing phrases from target posts. For example, [7, 10] employ topic models to generate topical words as the keyphrases of a group of posts. These methods, ascribed to the limitation of most topic models, are incapable of generating non-existed keyphrases for each targeted post. To deal with this problem, [3] recently introduces a sequence generation framework that can generate keyphrases beyond the target post. It presents a neural seq2seq model based on integrating more tweets related to the target post to generate keyphrases in a

✉ Junyang Chen
junyangchen@szu.edu.cn

1 College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China

2 Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ), Shenzhen, China

3 Tencent Inc., Shenzhen, China

4 School of Information Science and Technology, Guangdong University of Foreign Studies, Guangzhou, China

5 Wenzhou Institute, University of Chinese Academy of Sciences, Zhejiang, China

6 State Key Laboratory of Reliability and Intelligence of Electrical Equipment, Hebei University of Technology, Tianjin, China

7 State Key Laboratory of Internet of Things for Smart City, Department of Computer Information Science, University of Macau, Zhuhai, China

Published online: 29 August 2023

Table 1 POI: "The Vortex Bar And Grill" on Yelp.

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

Posts are usually the short comments from different users for the POI. User interest summarization denotes a succinct description from these posts

word-by-word training manner. But still, the above methods encounter a common challenge: only a 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 is 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. One way is to combine more relevant posts [3] 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 first propose a graph-based neural interest summarization model (UGraphNet) [11] that includes three complementary innovations. The first one is *user collaboration* that leverages neighboring information by constructing 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. UGraphNet has achieved improvement compared with the baselines.

Moreover, we further study UGraphNet to improve its performance by considering the optimization in the second part. The previous method refers to a matrix factorization [12] to obtain the hidden topical representations of user interests, which is a lack of nonlinear changes in the learning process and will have a negative impact on the final outcomes. More concretely, user-interested posts may only be related to several topics, and each topic may only involve several important words in reality. But the matrix factorization with linear transform is unable to follow such conditions, which prevents obtaining better user interest representations. To this end, we explore a neural variational inference method (*NIGraphNet*) to endow with the ability of nonlinear transform in the topical representation learning. Finally, we adopt a unified graph-based training loss that jointly learns the hidden topics and user relations for item recommendation.

In general, the contributions of this work are as follows:

- We propose a novel method to leverage user relations and latent topics by social media interest summarization for item recommendation. Our model enables an end-to-end training process through a unified graph-based training loss.
- We propose three main components: a contrastive learning loss, a topic modeling loss, and a graph-based learning loss to achieve the above purposes through their joint learning. We further explore a neural variational inference method to endow with the ability of nonlinear transform in topical representation 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 Related Work

This work is mainly in the line of three domains: user interest summarization, item recommendation, and topic modeling.

2.1 User Interest Summarization

Most of the previous works employ supervised or unsupervised methods to extract words selected from target documents to form the summarization. For supervised learning, [10] use deep recurrent neural networks with sequence tagging to conduct keyphrase extraction. Further,

[13] incorporate expert knowledge into the extraction. For unsupervised learning, various algorithms are also proposed, such as graph ranking [14] and document clustering [15]. However, these works only select keyphrases from source documents to make a summary, which would encounter a sparsity problem in short posts on social media. After that, [16] proposes to predict keyphrases in a sequence generation manner that allows the creation of absent keyphrases. Besides, some previous works [17, 18] that are based on topic modeling can also effectively alleviate the data sparsity with the corpus-level latent topics. Different from them, we propose to leverage user relations and latent topics on social media for user interest summarization that has been ignored in previous research and will be extensively studied here. In this way, our model can generate keyphrases beyond the limited number of relevant posts for the target user.

2.2 Item Recommendation

This refers to the social recommendation that adopts social relations to improve the content recommendation performance. Earlier work has typically used the directly linked neighbors to constrain and learn the representation of target users via matrix factorization approaches [19]. Recently, with the rise of the graph neural networks (GNN) such as GCN [20], GraphSAGE [21], and GAT [22], lots of effort have been devoted into the social recommendation. [23] use the directly linked relations of target users/items with GNN to learn their representations. Then, [24, 25] further exploit behavior patterns into user–item graphs to learn more powerful representations for both users and items. And [26] adopts a hypergraph neural network to explore high-order information in recommendation scenarios. Different from the above models, the proposed user interest summarization method is learned together with language generation, which has not been explored before in the existing work.

2.3 Topic Modeling

To enhance the topic modeling aspect of UGraphNet, we propose incorporating relevant work on hidden topic analysis. One prominent research approach for exploring the relationships between documents and their hidden topics is the matrix factorization method [12]. This method takes the document–word matrix as input and produces the document–topic matrix and the topic–word matrix as output. It has been widely employed as a key component in previous works on hidden topic analysis [11, 27]. However, there are two primary limitations associated with the matrix factorization method. Firstly, the previous use of this method lacked the ability to perform nonlinear transformations in topical representation learning. Consequently, previous models were unable to engage in nonlinear parameter learning. Secondly, we adopt a variational inference method [28] for topical learning. By utilizing the reparameterization trick in the inference process, this method addresses the issue of overfitting and enables models to capture more generalized user interests during the generative learning process.

3 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.

3.1 Contrastive Learning Loss

As shown in the left part of Fig. 1, 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 a unique embedding in large-scale scenarios, which will inevitably make the number of parameters tremendous. As such, inspired by the work [29] that they represent the users by the terms of queries, we represent the users with a smaller number of tag embeddings. In other words, 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 target user and its neighbors. First, we perform the message

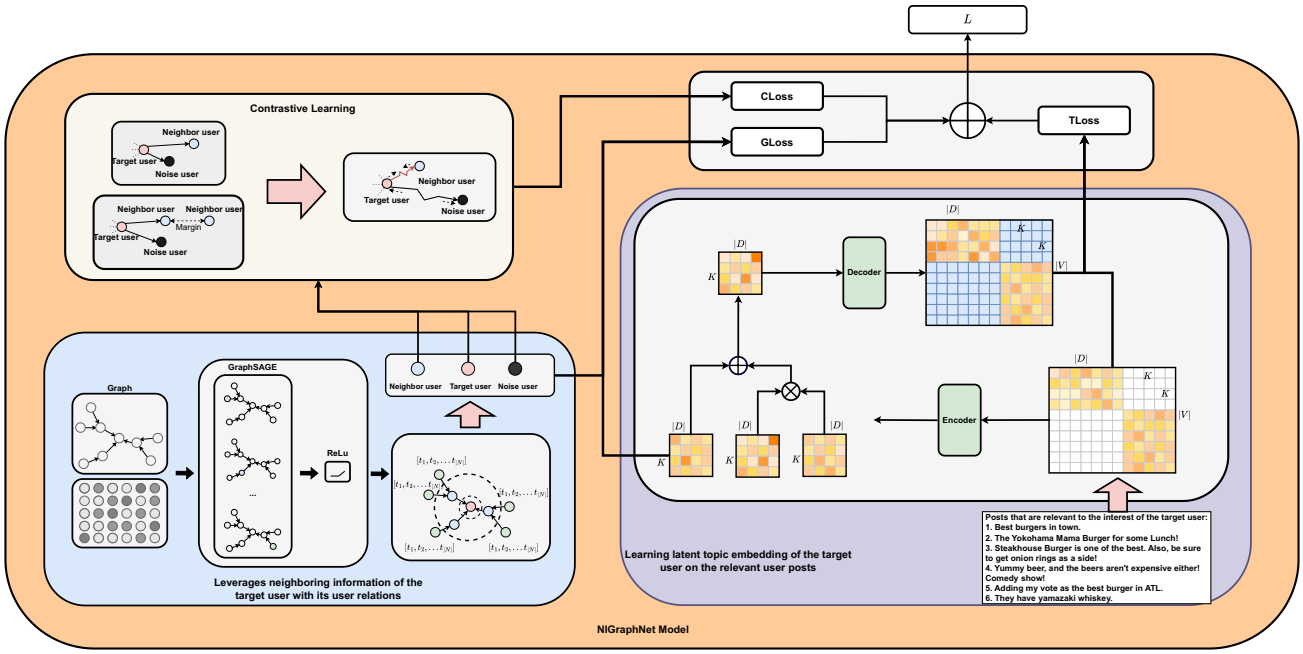

Fig. 1 Overview of the proposed NIGraphNet model

propagation step for dealing with the messages passing from neighboring nodes, which is given by:

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

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)), $\text{MLP}(\cdot) \in \mathbb{R}^{d \times d}$ 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 attentive way. The weight coefficient α_{v_i, v_j} between two nodes can be formulated by:

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

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 [30].

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:

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

where ReLU is an activation function [31] and L denotes the last layer of the network.

Finally, inspired by the recent advances in the contrastive learning work [32, 33], 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 [34] that only takes $O(1)$ time, ∇ is a margin hyperparameter 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.

3.2 Topic Modeling Loss

In this part, the previous method UGrahpNet [11] refers to a matrix factorization [12] 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 interesting to the target user. Besides, the obtained document–topic embedding $\mathbf{\Theta}$ will be used in generative learning.

Nevertheless, Eq. (6) lacks nonlinear transform in the topical feature learning, which may prevent obtaining better user interests. To this end, in our proposed NIGraphNet, we further explore a variational inference method [28] in topical learning, where the detail is the following.

As shown in the right part of Fig. 1, given a collection of posts, i.e., the document–word matrix $\mathbf{D}$, we adopt an encoder to obtain the document–topic embedding $\mathbf{\Theta}$ by estimating the parameters: mean μ and variance δ, where the formula is given by:

$$\mu = \text{ReLU}(\mathbf{D}\mathbf{M}_\mu), \quad \delta = \text{ReLU}(\mathbf{D}\mathbf{M}_\delta), \tag{7}$$

where $\mathbf{M}_\mu \in \mathcal{R}^{V \times k}$ and $\mathbf{M}_\delta \in \mathcal{R}^{V \times k}$ are trainable weights, and ReLU is an activation function [17]. In our approach, we utilize a Gaussian distribution with mean μ and variance δ to estimate the document–topic embedding $\mathbf{\Theta}$. This enables us to incorporate model uncertainty in parameter inference, thereby mitigating the overfitting issue commonly associated with the original matrix factorization method. By considering the distribution of the document–topic embedding, we introduce a level of flexibility that helps to alleviate the limitations of the previous approach. Then, we follow the generative process:

1. Generate hidden topical features $\mathbf{\Theta} \sim \mathcal{N}(\mu, \delta^2)$;
2. For each document $i \in D$:
   (a) Draw a document–topic distribution: $\Phi_i = \text{Softmax}(\text{Sigmoid}(W_\Theta \mathbf{\Theta}_i))$;
   (b) For each word w in the vocabulary, draw a word–topic distribution: $\phi_w = \text{Softmax}(\text{Sigmoid}(W_\Phi \Phi_i))$.

where $\mathbf{\Theta} \in \mathbb{R}^{|D| \times k}$, M_Θ and M_Φ are trainable linear transformation, $\Phi \in \mathcal{R}^{|D| \times k}$ denotes the document–topic distributions, k denotes the dimension of topic embeddings, and $\phi \in \mathcal{R}^{V \times k}$ denotes the word–topic distributions. In an end-to-end training, $\mathbf{\Theta} \sim \mathcal{N}(\mu, \delta^2)$ can be re-parameterized [28] as:

$$\mathbf{\Theta} = \mu + \epsilon \cdot \delta, \tag{8}$$

where $\epsilon \sim \mathcal{N}(0, 1)$. After that, a decoder is adopted for $\mathbf{\Theta}$ to reconstruct the input $\hat{\mathbf{D}}$. The detail is given as follows:

$$\hat{\mathbf{D}} = \text{Softmax}(\mathbf{\Theta}\mathbf{W}_d), \tag{9}$$

where $\mathbf{W}_d \in \mathcal{R}^{k \times V}$. In general, we need to minimize the original input $\mathbf{D}$ and the reconstructed output $\hat{\mathbf{D}}$. The final topic modeling loss is given by:

$$\hat{\mathcal{L}}_t = ||\mathbf{D} - \hat{\mathbf{D}}||^2 + \eta(||\mu||^2 + ||\delta||^2), \tag{10}$$

where η is set to 0.001 in the experiments. In summary, we improve the exploration of Eq. (6) with Eq. (10). The obtained document–topic embedding $\mathbf{\Theta}$ of Eq. (8) will be used in the following section.

By incorporating the aforementioned neural inference method, we enhance our ability to perform nonlinear transformation in topical representation learning. Moreover, the utilization of the reparameterization trick during the inference process helps to mitigate the problem of overfitting and allows our models to capture more generalized user interests during the generative learning process.

3.3 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. (8) that represents the interests of the target user, we can construct the generative learning loss $\mathcal{L}_g$ as follows:

$$\mathcal{L}_g = -\sum_{v_t \in \mathcal{T}} \log(\sigma([\mathbf{h}_{v_t}^L ; \mathbf{\Theta}_{v_t}]\mathbf{W}_v)), \tag{11}$$

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. (11), 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.

3.4 Learning and Inference

In the training stage, we adopt stochastic gradient descent [35] to minimize the loss function of the total loss, which is given by:

$$\mathcal{L}_{total} = \mathcal{L}_c + \hat{\mathcal{L}}_t + \mathcal{L}_g. \tag{12}$$

With the above learning objective as shown in Eq. (12), 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. 10), and (3) fuse the above information (Eq. 11) to simultaneously learn them in an end-to-end way.

3.4.1 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.

3.5 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.

4 Experiments

In the experiments, we first evaluate the performance on user interest summarization tasks. Then, we conduct an ablation study for estimating the effect of the proposed components, including contrastive learning, topic modeling, and generative learning. At last, we evaluate whether jointly learning user interests can be conducive to the item recommendation task.

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

4.1 Datasets

We adopt two real-world datasets to estimate the performance: *Delicious*[1] and *Yelp*[2] which are widely used in social recommendation [36, 37]. 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.

4.2 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 the selected methods:

GSDMM [38] is a traditional and widely used probabilistic graph model which is designed for short text modeling. The word and document representations are learned by combining Dirichlet and multinomial distributions.

DP-BMM [2] is another often used probabilistic graph model 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 [10] is a state-of-the-art deep recurrent neural network model that can combine keywords and context information to automatically extract keyphrases from short texts.

[1] https://grouplens.org/datasets/hetrec-2011/.

[2] https://www.yelp.com/dataset.

Table 3 Main comparison results displayed with scores in %.

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 art								
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	34.56	35.51	33.07	34.63	55.99	35.09	25.30	48.20
NIGraphNet (ours)	**35.70**	**36.39**	**34.08**	**36.33**	**70.52**	**63.05**	**49.86**	**65.00**
Improv	3.30%	2.48%	3.05%	4.91%	25.95%	79.68%	97.08%	34.85%

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

SEQ2SEQ-CORR [39] exploits a sequence-to-sequence (seq2seq) architecture for keyphrase generation which captures correlation among multiple keyphrases in an end-to-end fashion.

TAKG [16] introduces a seq2seq-based neural keyphrase generation framework that takes advantage of the recent advance of neural topic models [28] 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. Our proposed models include:

UGraphNet [11] propose a graph-based neural interest summarization model that includes contrastive learning, topic modeling, and generative learning.

NIGraphNet is our proposed improved version that considers the optimization in the second part of UGraphNet, which endows with the ability of nonlinear transform in the topical representation learning.

4.3 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:

(1) Our model *UGraphNet* and *NIGraphNet* consistently outperform other comparisons on all datasets under various metrics. This shows the usefulness of leveraging user neighboring information for their interest summarization. Moreover, *NIGraphNet* increases 3.30%, 2.48%, 3.05%, and 4.91% over *UGraphNet* in terms of HR@1, HR@5, HR@10 and MAP on Delicious, respectively. And *NIGraphNet* increases 25.95%, 79.68%, 97.08%, and 34.85% on Yelp. One interesting observation is that *NIGraphNet* can gain larger improvements on Yelp compared with Delicious. We explain that Yelp contains more text information than Delicious as shown in Table 2. These improvements demonstrate our improved version can better explore the document–topic distributions.

(2) Besides, the second-best method, *UGraphNet*, achieves up to 24.86%, 14.70%, 14.67%, and 24.75% improvements over the third-best method *TAKG* in terms of HR@1, HR@5, HR@10 and MAP on Delicious. *UGraphNet* 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 methods by jointly modeling user relations and user interests.

(3) 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,

Table 4 Ablation analysis

Method	Delicious				Yelp			
	HR@1	HR@5	HR@10	MAP	HR@1	HR@5	HR@10	MAP
NIGraphNet	**35.70**	**36.39**	**34.08**	**36.33**	**70.52**	**63.05**	**49.86**	**65.00**
w/o CLoss	28.02	20.24	13.42	17.21	61.15	46.37	35.19	58.15
w/o TLoss	4.92	5.63	4.44	5.28	52.07	25.79	15.81	14.29
w/o GLoss	31.51	29.32	28.08	26.98	20.86	17.47	14.99	11.03
Improv	13.30%	24.11%	21.37%	34.66%	15.32%	35.97%	41.69%	11.78%

The highest scores are marked in boldface, and '_' denotes the second-best results

which suggests the help of exploiting latent topics in short texts. Interestingly, our model achieves larger improvements with a step further by exploring the user relations and their latent topics.

4.4 Ablation Analysis

To analyze the effectiveness of the proposed components on user interest summarization (introduced in Sect. 3) 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 contrastive 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. 10) aims to exploit the latent topics in short texts which can especially effectively alleviate the data sparsity in the user interest summarization.
- III. w/o GLoss (without generative learning loss): The GLoss (Eq. 11) utilizes the assumption that relevant users share similar interests. We adopt it to generate keyphrases that are relevant to users' latent topics.

The results of the ablation tests are shown in Table 4. Our method *NIGraphNet* outperforms the other variants. Specifically, *NIGraphNet* achieves 13.30%, 24.11%, 21.37%, and 34.66% improvements over the second-best variant in terms of HR@1, HR@5, HR@10, and MAP on Delicious, and obtains 15.32%, 35.97%, 41.69%, and 11.78% gains on Yelp, respectively. These results validate that the user attention update gate is more appropriate to explore user interests. These results demonstrate the effectiveness of jointly learning different components. We observe that the performance order is presented as w/o GLoss > w/o CLoss > w/o TLoss on Delicious. These results demonstrate that the topic modeling loss contributes the most to the learning. By contrast, the performances of components on Yelp are as: w/o CLoss > w/o TLoss > w/o GLoss, which shows that the generative loss contributes the most while the contrastive learning loss contributes the least. In general, all parts contribute to the final performance, which evidently demonstrates their effectiveness.

4.5 Item Prediction

In this part, we evaluate if unearthing potential user relations and jointly learning the latent topic representations can facilitate item prediction. Concretely, we adopt a standard evaluation metric area under the curve (AUC) [40] to predict a link between users and items. This metric represents the probability that users and items in a random unobserved link are more similar than those in a random non-existed link. The AUC metric has been widely used in recommendation tasks [9, 25]. When the prediction results perfectly match the ground truth, the AUC value will be one, otherwise, it will be zero. The baselines include AMOUNT [41], IMP-GCN [42], and IRLM [43], where are the stat-of-the-art methods used for user–item prediction. We report the comparison results as shown in Fig. 2. Observations derived from this figure are as follows:

Our methods achieve the best performance over the other baselines in terms of AUC. Specifically, our NIGraphNet obtains 0.9574 and 0.8896 on Delicious and Yelp, respectively. Besides, UGraphNet obtains 0.9423 and 0.8775 on Delicious and Yelp, respectively. Among the baselines, IRLM gets the second-best performance, 0.9276, on Delicious, and IMP-GCN obtains better performance, 0.8633, on Yelp. All the comparison methods either ignore the semantic features or regard them as the static values associated with nodes. By contrast, our models enable an end-to-end training process that jointly learning the latent topics and the user–item relations.

Table 5 Parameter analysis for η in Eq. (10), where the highest scores are marked in boldface

NIGraphNet	Delicious				Yelp			
	HR@1	HR@5	HR@10	MAP	HR@1	HR@5	HR@10	MAP
$\eta = 0.1$	0.22	0.86	0.61	0.58	1.55	1.03	0.85	0.76
$\eta = 0.01$	0.38	0.24	0.28	0.34	45.02	28.25	20.63	21.92
$\eta = 0.001$	**35.70**	**36.39**	**34.08**	**36.33**	**70.52**	**63.05**	**49.86**	**65.00**
$\eta = 0.0001$	10.60	20.67	18.58	21.65	0.48	0.60	0.62	0.58

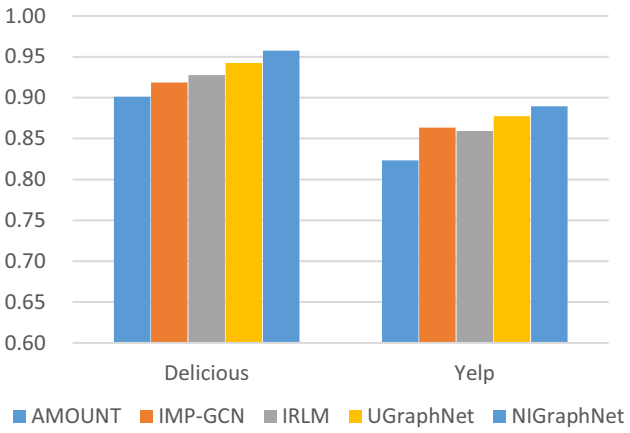

Fig. 2 Results of AUC comparison on the Delicious and Yelp datasets

4.6 Parameter Analysis

This section aims to perform experiments with different values of η in Eq. (10) and analyze their impact on the model's performance. The results of these experiments are presented in Table 5. From the table, we can observe that NIGraphNet achieves the best performance on both datasets when $\eta = 0.001$. Additionally, NIGraphNet obtains the second-best performance when $\eta = 0.0001$ on the Delicious dataset and $\eta = 0.01$ on the Yelp dataset. It is important to note that we introduce additional constraints parameterized by η in Eq. (10). The inclusion of a penalty term serves the purpose of preventing the parameters from reaching excessively large values. By imposing these restrictions, we aim to regulate the parameter values and facilitate a more balanced and stable optimization process.

4.7 Summary for Experimental Study

In general, our proposed extension method, *NIGraphNet*, achieves improvements over the original *UGraphNet* in various evaluation metrics. Specifically, on the Delicious dataset, *NIGraphNet* outperforms UGraphNet by 3.30%, 2.48%, 3.05%, and 4.91% in terms of HR@1, HR@5, HR@10, and MAP, respectively. On the Yelp dataset, *NIGraphNet* achieves improvements of 25.95%, 79.68%, 97.08%, and 34.85% over *UGraphNet* in the same metrics. Additionally, *NIGraphNet* demonstrates superior performance in terms of AUC, surpassing *UGraphNet* by 1.60% and 1.38% on the Delicious and Yelp datasets, respectively. The results of the ablation study further highlight the effectiveness of jointly learning different components within *NIGraphNet*.

5 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. Moreover, we further propose an improved version, NIGraphNet, that can explore hidden topics with a variational inference approach. The main innovations of our work include a contrastive learning loss, a topic modeling loss, and a graph-based learning loss that can leverage user relations and latent topics on social media through joint 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.

Acknowledgements The preliminary version of this article has been published in DASFAA 2023 [A Topic-Aware Graph-Based Neural Network for User Interest Summarization and Item Recommendation in Social Media]. 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 and 2023A1515012534, by the National Statistical Science Research Project of China Grant No. 2022LY096, by the Science and Technology Development Fund, Macau SAR, China, under Grant No. (0068/2020/AGJ, SKL-IOTSC(UM)-2021-2023), and by the Natural Science Foundation of Hebei Province, China, under the Grant No. F2021202064. (Corresponding authors: Wenfeng Du,

Zhenghua Xu, Email: junyangchen@szu.edu.cn, duwf@szu.edu.cn, zhenghua.xu@hebut.edu.cn).

Author Contributions JC, MW, ZC, and GF are responsible for writing the first draft of the paper, conducting code experiments, organizing experimental results, analyzing experimental results, and drawing diagrams. WD, and GZ are responsible for writing and proofreading paper formulas, deriving algorithms, and debugging experimental code. OL, ZX, and ZG are responsible for conceptualizing paper ideas, summarizing related work, and refining innovative points

Funding Not applicable

Availability of Data and Materials Not applicable.

Declarations

Conflict of interest Not applicable.

References

1. Chen J, Gong Z, Liu W (2019) A nonparametric model for online topic discovery with word embeddings. Inf Sci 504:32–47
2. Chen J, Gong Z, Liu W (2020) A Dirichlet process biterm-based mixture model for short text stream clustering. Appl Intell 50:1609–1619
3. Wang Y, Li J, King I, Lyu MR, Shi S (2019) Microblog hashtag generation via encoding conversation contexts. arXiv preprint arXiv:1905.07584
4. Efron M (2010) Hashtag retrieval in a microblogging environment, pp 787–788
5. Bansal P, Jain S, Varma V (2015) Towards semantic retrieval of hashtags in microblogs, pp 7–8
6. Davidov D, Tsur O, Rappoport A (2010) Enhanced sentiment learning using twitter hashtags and smileys, pp 241–249
7. Wang X, Wei F, Liu X, Zhou M, Zhang M (2011) Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach, pp 1031–1040
8. Wang W et al (2019) Trust-enhanced collaborative filtering for personalized point of interests recommendation. IEEE Trans Ind Inf 16:6124–6132
9. Wang W, Chen J, Wang J, Chen J, Gong Z (2019) Geography-aware inductive matrix completion for personalized point-of-interest recommendation in smart cities. IEEE Internet Things J 7:4361–4370
10. Zhang Q, Wang Y, Gong Y, Huang X-J (2016) Keyphrase extraction using deep recurrent neural networks on twitter, pp 836–845
11. Chen J et al. (2023) A topic-aware graph-based neural network for user interest summarization and item recommendation in social media, pp 537–546
12. Ahmed NK et al. (2018) Learning role-based graph embeddings. arXiv preprint arXiv:1802.02896
13. Gollapalli SD, Li X-L, Yang P (2017) Incorporating expert knowledge into keyphrase extraction, Vol. 31
14. Mihalcea R, Tarau P (2004) Textrank: bringing order into text, pp 404–411
15. Liu Z, Huang W, Zheng Y, Sun M (2010) Automatic keyphrase extraction via topic decomposition, pp 366–376
16. Wang Y et al. (2019) Topic-aware neural keyphrase generation for social media language. arXiv preprint arXiv:1906.03889
17. Zeng J et al. (2018) Topic memory networks for short text classification. arXiv preprint arXiv:1809.03664
18. Li J, Song Y, Wei Z, Wong K-F (2018) A joint model of conversational discourse and latent topics on microblogs. Comput Linguist 44:719–754
19. Tang J, Hu X, Liu H (2013) Social recommendation: a review. Soc Netw Anal Min 3:1113–1133
20. Kipf TN, Welling M (2016) Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907
21. Hamilton W, Ying Z, Leskovec J (2017) Inductive representation learning on large graphs, pp 1024–1034
22. Veličković P et al. (2017) Graph attention networks. arXiv preprint arXiv:1710.10903
23. Fan W et al. (2019) Graph neural networks for social recommendation, pp 417–426
24. Wu L et al. (2020) Diffnet++: a neural influence and interest diffusion network for social recommendation. IEEE Trans Knowl Data Eng
25. Chen J et al. (2022) Meta-path based neighbors for behavioral target generalization in sequential recommendation. IEEE Trans Netw Sci Eng
26. Yu J et al. (2021) Self-supervised multi-channel hypergraph convolutional network for social recommendation, pp 413–424
27. Chen J, Gong Z, Wang W, Liu W, Dong X (2021) Crl: collaborative representation learning by coordinating topic modeling and network embeddings. IEEE Trans Neural Netw Learn Syst 33:3765–3777
28. Miao Y, Grefenstette E, Blunsom P (2017) Discovering discrete latent topics with neural variational inference, pp 2410–2419 (PMLR)
29. Fan S et al. (2019) Metapath-guided heterogeneous graph neural network for intent recommendation, pp 2478–2486
30. Han J, Moraga C (1995) The influence of the sigmoid function parameters on the speed of backpropagation learning, Springer, pp 195–201
31. Glorot X, Bordes A, Bengio Y (2011) Deep sparse rectifier neural networks, pp 315–323
32. Jaiswal A, Babu AR, Zadeh MZ, Banerjee D, Makedon F (2020) A survey on contrastive self-supervised learning. Technologies 9:2
33. Khan A, AlBarri S, Manzoor MA (2022) Contrastive self-supervised learning: a survey on different architectures, pp 1–6 (IEEE)
34. Li AQ, Ahmed A, Ravi S, Smola AJ (2014) Reducing the sampling complexity of topic models, pp 891–900
35. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980
36. Wang H, Wu Q, Wang H (2017) Factorization bandits for interactive recommendation

37. Guo Z, Wang H (2020) A deep graph neural network-based mechanism for social recommendations. IEEE Trans Ind Inf 17:2776–2783
38. Yin J, Wang J (2014) A dirichlet multinomial mixture model-based approach for short text clustering, pp 233–242 (ACM)
39. Chen J, Zhang X, Wu Y, Yan Z, Li Z (2018) Keyphrase generation with correlation constraints. arXiv preprint arXiv:1808.07185
40. Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143:29–36
41. Wang W et al (2021) A multi-graph convolutional network framework for tourist flow prediction. ACM Trans Internet Technol (TOIT) 21:1–13
42. Liu F, Cheng Z, Zhu L, Gao Z, Nie L (2021) Interest-aware message-passing GCN for recommendation, pp 1296–1305
43. Chen J et al. (2022) Irlm: inductive representation learning model for personalized poi recommendation. IEEE Trans Comput Soc Syst