MultiSum: A Multi-Facet Approach for Extractive Social Summarization Utilizing Semantic and Sociological Relationships

Tanglong Zhao[1,2], Ruifang He[1,2*], Jing Xu[1,2], Bo Wang[1,2]

[1] College of Intelligence and Computing, Tianjin University, Tianjin, China
[2] Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin, China
{zhaotl,rfhe,jingxu,bo_wang} @tju.edu.cn

Abstract

Social summarization aims to provide summaries for a large number of social texts (called posts) about a single topic. To extract a summary, both the representation of post and summary selection method are crucial. Previous methods introduce social relation to enhance post embedding to mitigate the sparse representation due to its brief and informal expression. However, they ignore that there are multiple relations between posts. Besides, existing graph-based centrality calculation approaches tend to select posts from one aspect. This leads to facet bias especially when there are multiple viewpoints. In this paper, we propose a model named MultiSum to improve social summarization. Specifically, 1) We use graph convolutional networks to fuse text content with social and semantic relations to improve post representation; 2) The similarity between the summary and all aspects is incorporated into the centrality score during the selection phase, encouraging the model to pay attention to different facets. Experimental results on English and Chinese corpora support the effectiveness of this model. Furthermore, external evaluations by human experts and large language models demonstrate the validity of MultiSum in facet coverage and redundancy reduction.

Introduction

Social summarization aims to produce a short and compact summary for a collection of posts on a specific topic. It can efectively alleviate information overloading problem and help people acquire key information on social media quickly. A notable technique in the field of text summarization is the 'Extractive' approach, in which representative sentences are directly extracted from the source document. An alternative method, called the 'Abstractive' approach, generates summaries based on the core idea within the text. Both of them have made great progress within conventional summarization domains, such as news summarization and dialogue summarization (Liu et al. 2022; Liu, Jia, and Zhu 2022; Dixit, Wang, and Chen 2023; Li et al. 2023; Gao et al. 2023).

Similarly, there have been strides in social summarization. Traditional approaches typically use different types of ranking algorithms to get summaries based on the textual content of the posts, such as graph-based ranking algorithms (Erkan and Radev 2004; Dutta et al. 2018; Sharifi, Inouye, and Kalita 2013) and clustering-based methods (Andy, Wijaya, and Callison-Burch 2019; Wang et al. 2019; Gillani et al. 2017). These methods mainly consider each post as a separate unit and extract features independently based on its content. However, because of post's brief content and informal expression, its representation tends to be sparse. Work in recent years has focused on improving the summarization by enriching post representation. Ali et al. (2020) propose a technique for summarizing microblogs on Twitter that simultaneously considers topic sentiments. But this approach still focuses on the content of the posts and ignores the relationships between them. Liu et al. (2021) use graph convolutional network (GCN) on social networks to integrate text content and social relation features into a universal representation. Despite the validity of it, existing studies usually involve only one type of edges when constructing graphs. However, posts often have multiple types of relations, such as social and semantic relationships (Jing, Park, and Tong 2021). Different relationships provide relational information from distinct aspects, and jointly modeling various relations of them will improve the performance of the model (Yan et al. 2021; Jing et al. 2021a).

In addition to the representation of posts, the approach for summary selection is also important for extractive summarization. Graph-based centrality calculation methods are commonly used in this task (Radev, Jing, and Budzikowska 2000; Erkan and Radev 2004). They suggest the more similar a sentence is to other sentences, the more important it is. Zheng and Lapata (2019) introduce a directed centrality method called PacSum, premise on the assumption that the contribution of any two nodes to their respective centralities is influenced by their relative positions in the document. For documents focusing on a single aspect, this approach typically works well. However, in longer documents or social media posts, there are often multiple facets to consider. Taking into account the property that posts have various opinions can enhance the diversity of summary. Existing methods, such as clustering (Andy, Wijaya, and Callison-Burch 2019), can enhance the variety of aspects, but they need manual feature extraction from posts and the number of human-defined facets, which require expert experiences

*Corresponding author.

and have poor flexibility and generalization ability.

In this paper, we propose an unsupervised extractive social summarization model, named MultiSum. We improve summaries by modeling multiple relationships between posts and extracting posts that contain various viewpoints as summaries. Specifically, We first model the social and semantic relationships between posts and then fuse the relations and textual content using a multi-stream encoder consisting of two GCNs. This encoder aggregates information from neighbors with two perspectives. To make the summaries cover as many facets as possible, we use an improved graph-based ranking method to select posts. It encodes all tweets into a vector space, which is used to capture all aspects. For each candidate summary, we compute the similarity score between the summary and all posts, called summary-document similarity. During the ranking phase, we combine summary-document similarity with post's centrality score, ensuring the selected posts are representative while covering numerous aspects. Our contributions are as follows:

- For the task of social summarization, we are the first to model multiple relationships between posts simultaneously, considering both semantic and social relationships.
- Utilizing the inherent 'multi-facet' nature of social media text, we employ an improved graph-based ranking method to obtain summaries that cover multiple facets.
- The validity of the model is evaluated through automatic assessments of the English and Chinese datasets, as well as through evaluations by humans and large language models.

Related Work

Social Summarization

Current methods can be divided into two categories. The first method is a content-based approach that select salient posts based on textual content features, such as n-grams (Ganesan, Zhai, and Viegas 2012), TF-IDF weights (Inouye and Kalita 2011), and phrase frequency (Sharifi, Hutton, and Kalita 2010). Mere consideration of textual content proves insufficient, as social posts are often short, leading to highly sparse content (Chang et al. 2013). Integrating social network features can alleviate the problem of insufficient information (Liu et al. 2012). Consequently, the second method considers social signals, such as the number of reposts (Alsaedi, Burnap, and Rana 2016), topic tags (Dutta et al. 2018), the quantity of followers (Liu et al. 2012), Reddit's "karma" scores (Kano et al. 2018), and user influence (Li et al. 2015), along with social relationships (Shalizi and Thomas 2010). These social signals assist in identifying the importance and popularity of posts. In this paper, we utilize both semantic and sociological relationships to mitigate the problem of insufficient information.

Graph Convolutional Networks

GCN can be regarded as a message-passing framework. The hidden state of nodes is updated through a weighted aggregation of their neighbors' states. In the past few years, with the development of GCN, it has been used in a wide variety of fields (Liu, Fu, and Strube 2023; Yang, Bao, and Qiu 2023; Sharma et al. 2022). In the field of recommender systems, Chen et al. (2023) use GCN to capture the substitutable and complementary relationships between items to enhance the recommendation effect. For social summarization, Liu et al. (2021) utilize GCN to fuse social relation and BERT-encoded textual representations. Consequently, the nodes' information can be propagated along edges, thereby influencing the nodes' neighbors. Jing et al. (2021b) employ GCNs to separately capture the shared keyword information among sentences and the semantic relationships between them when engaged in the task of summarizing news.

Graph-Based Summarization

A document can be represented as a graph, where the nodes correspond to sentences. PageRank determines the importance scores of each node through textual similarity, thereby producing a summary (Brin and Page 1998). In the work by Zheng and Lapata (2019), a directed centrality method named PacSum is introduced. Dong, Romascanu, and Cheung (2021) further improve PacSum by amalgamating hierarchical and positional information into the directed centrality method. Liang et al. (2021) suggest a pioneering approach to augment unsupervised extractive summarization by learning facet-aware modeling, comprising a graph-based ranking model that computes sentence centrality with a sentence-document weight. This weight mirrors the importance of different facets in the document.

Proposed Method

Problem Definition

Our study focus on topic-oriented social summarization, which implies that under a single topic, there exists a set of posts $S = \{s_1, s_2, ..., s_N\}$. Our objective is to select L posts from these N posts ($L \ll N$), and these L posts can retain the core information of all the posts. Let $U = \{u_1, u_2, ..., u_M\}$ represent the M users in the social network. These M users and N posts form a heterogeneous graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{V} = \{\mathcal{V}^s \cup \mathcal{V}^u\}$, in which $\mathcal{V}^s$ stands for the post nodes, and $\mathcal{V}^u$ represents the user nodes. $\mathcal{E} = \{\mathcal{E}^{uu} \cup \mathcal{E}^{us}\}$ where $\mathcal{E}^{uu} \subseteq M \times M$ indicates the relationships between users, and $\mathcal{E}^{us} \subseteq M \times N$ signifies the user-post relationships. More specifically, if users u_i and u_j have a social relationship, they are connected in the graph. Meanwhile, if post s_j is published by user u_i, then user u_i and post s_j are linked.

Multi-Relation Graph Construction

Social Relation Graph Based on sociology theories, the social relationships between posts have been proven to be effective for many downstream tasks (Kipf and Welling 2017). The most prominent of these are social consistency and social contagion in social networks. Social consistency refers to the tendency for a single user's opinions on a specific topic to remain consistent over a short period. Social contagion refers to the ability of the emotions contained in a user's post, such as happiness (Fowler and Christakis 2008)

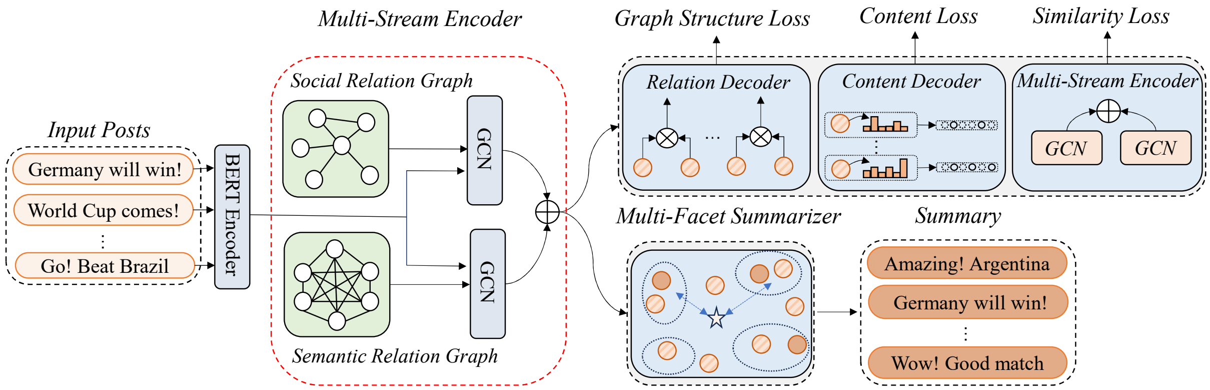

Figure 1: Model architecture of MultiSum. $\oplus$ means concatenated tensors. Post embeddings are represented by circles. Dark cell indicates that it has been selected as a summary. In the Multi-Facet Summarizer module, the star is the representation of all posts, the dotted ellipse means a facet, and the line between it and the star illustrates the distance, which measures the similarity of summary and all facets

and loneliness (Cacioppo, Fowler, and Christakis 2009), to spread within the social network (He and Duan 2018; He, Zhao, and Liu 2020). Driven by these two social theories, we propose a simple strategy to simulate the social relationships of posts, considering both social consistency and social contagion.

To construct the social graph, we make the following definitions. We build the social relationship graph $\mathcal{G}_{soc} = (\mathcal{V}^s, \mathcal{E}_{soc})$, which includes post nodes $\mathcal{V}^s$ and social relationships between posts $\mathcal{E}_{soc}$. For a user u_i, let $\mathcal{N}(u_i) = \{u_k \mid \mathcal{E}^{uu}_{ik} = 1\}$ represent the set of neighbors of user u_i, and $\mathcal{P}(u_i) = \{s_k \mid \mathcal{E}^{us}_{ik} = 1\}$ represent the set of posts published by user u_i. According to these definitions, the construction of the social graph follows two principles:

- **Social consistency** If posts $s_i \in \mathcal{P}(u_k)$ and $s_j \in \mathcal{P}(u_k)$, s_i and s_j are connected. It means that a link is established between posts published by the same user. This relationship reflects the relationship of social consistency (He and Duan 2018; Abelson and Prentice 1983).
- **Social contagion** If $s_i \in \mathcal{P}(u_i)$, $s_j \in \mathcal{P}(u_j)$, and $u_i \in \mathcal{N}(u_j)$, we allow the interactions between posts published by users who have social connections. This kind of relationship reflects the relationship of social contagion (Centola et al. 2013; Centola 2010).

According to the above rules, we can construct a social relation graph. It can be represented by the adjacency matrix $\mathbf{A}_{soc} \in \mathbb{R}^{N\times N}$. The relationships allow information to be transmitted between posts, thereby capturing the social context information of the posts.

Semantic Relation Graph The semantic relation graph $\mathcal{G}_{sem} = (\mathcal{V}^s, \mathcal{E}_{sem})$ explores the semantic relationships between posts. To construct $\mathcal{G}_{sem}$, we first need to obtain representations of the posts. Specifically, for each post $s_i \in S$, we input s_i into a pre-trained BERT encoder. The output of the first token ([CLS]) of the last layer is then taken as the post embedding.

Given the differences between posts and traditional documents, we adopt the BERTWEET model (Nguyen, Vu, and Tuan Nguyen 2020). It shares a similar architecture with BERT, but is trained on a large corpus of tweets, making it more suitable for handling posts:

$$\mathbf{x_i} = BERTWEET(s_i) \tag{1}$$

After obtaining the posts representations, we use the cosine similarity between posts as the weight of the edges between them. We build $\mathcal{E}_{sem}$ and represent it using an adjacency matrix $\mathbf{A}_{sem} \in \mathbb{R}^{N\times N}$.

Multi-Stream Encoder

We propose a multi-stream encoder that leverages both sociological and semantic relations between posts. This consists of two distinct GCN encoders : a social relation encoder and a semantic relation encoder. Together, they alleviate information shortages in individual posts by integrating social signals and textual relationships.

In formal terms, the social relation encoder adopts the representation of the post, $\mathbf{X} \in \mathbb{R}^{N\times D}$, and the adjacency matrix representation of the social relation graph, $\mathbf{A}_{soc} \in \mathbb{R}^{N\times N}$, as inputs. It aggregates information through the following rules:

$$\mathbf{H}^{(l+1)}_{soc} = \sigma\left(\tilde{\mathbf{A}}_{soc}\mathbf{H}^{(l)}_{soc}\mathbf{W}^{(l)}_{soc} + \mathbf{b}^{(l)}_{soc}\right) \tag{2}$$

$$\tilde{\mathbf{A}}_{soc} = \tilde{\mathbf{D}}^{-1/2}_{soc}(\mathbf{A}_{soc} + \mathbf{I})\tilde{\mathbf{D}}^{-1/2}_{soc} \tag{3}$$

Where $\mathbf{H}^{(0)}_{soc} = \mathbf{X}$, $\tilde{\mathbf{A}}_{soc}$ is the normalized adjacency matrix after adding self-connections. $\mathbf{I}$ is the identity matrix and $\tilde{\mathbf{D}}_{soc}$ is the degree matrix of $\mathbf{A} + \mathbf{I}$. $\mathbf{W}^{(l)}_{soc}, \mathbf{b}^{(l)}_{soc}$ are the trainable weights for the l-th layer. In each layer of the GCN, the node uses the normalized adjacency matrix $\tilde{\mathbf{D}}_{soc}$ to aggregate information from its neighbors, applies a linear transformation using $\mathbf{W}^{(l)}_{soc}, \mathbf{b}^{(l)}_{soc}$, and then employs an activation function. Likewise, the rule for updating node representations in the semantic relation encoder follows the same methodology.

In the multi-stream encoder, we adopt residual connections, enabling the model to directly inherit information from the input, while simultaneously avoiding the over-smoothing problem caused by deep GCN.

$$\mathbf{H}_i = f(\mathbf{X}) + \sigma\left(\tilde{\mathbf{A}}_i \mathbf{X} \mathbf{W}_i^{(0)} + \mathbf{b}_i^{(0)}\right) \tag{4}$$

Where $i \in \{soc, sem\}$, $\mathbf{H} \in \mathbb{R}^{N \times Z}$ is the hidden state, where Z is the dimension of the representation. $\sigma(\cdot)$ is the activation function, and $f(\cdot) \in \mathbb{R}^{D \times Z}$ is the mapping function that transforms node features to the latent hidden space $\mathbb{R}^Z$. The outputs of sentence features from the two encoders are then concatenated for the final representation, which is formulated as equation 5. Obviously, by connecting the outputs of the two encoders, the final representation includes both sociological and semantic information.

$$\mathbf{H} = \mathbf{H}_{soc} \oplus \mathbf{H}_{sem} \tag{5}$$

Multi-Stream Reconstruction

Having obtained the representation of the post that integrates social information and content relevance, we allow the model to reconstruct the representation of the post through the multi-stream reconstruction module. Specifically, we separately reconstruct the original text of the post, social and semantic relation graph, and the representation of the summary.

For reconstructing the original text of the post, the approach is to predict the relationship between the post and the words, that is, whether the post contains this word. Hence, we transform this task into a classification problem:

$$\hat{\mathbf{s}}_i = \sigma(\mathbf{W}_d \mathbf{h}_i + \mathbf{b}_d) \tag{6}$$

Where, $\mathbf{W}_d \in \mathbb{R}^{V \times 2Z}$ and $\mathbf{b}_d \in \mathbb{R}^V$ are learnable parameters. V is the size of the vocabulary. $\hat{\mathbf{s}}_i \in \mathbb{R}^V$ is the predicted result, where $\hat{\mathbf{s}}_{ij}$ denotes the probability of post $\hat{\mathbf{s}}_i$ containing the word $\mathbf{w}_j$. Given the similarity of content within posts on the same topic and our processing of the noisy representations in the posts, we maintain a relatively small vocabulary size.

For reconstructing the social and semantic relation graph, we follow Kipf and Welling (2016). For each pair of nodes in the graph, the decoder predicts the weight of the edge between their posts. The prediction factor is implemented using the inner product function between the representations of the posts:

$$P\left(\mathbf{A}_{i,j} \mid \mathbf{h}_i, \mathbf{h}_j\right) = \sigma\left(\mathbf{h_i} \cdot \mathbf{h_j}^{\mathrm{T}}\right) \tag{7}$$

Regarding the reconstruction of the representation of the summary, we follow the ideas of (Chu and Liu 2019; Ma et al. 2018). Since there is no golden summary as a supervisory signal, we compress the summary obtained by the model into the post space $s \in V$, which can be regarded as the golden summary. We then re-encode the summary and calculate the similarity loss between the average representation of the summary and the post, further ensuring semantic similarity between the summary and the original post.

Based on the above explanation, the loss of our model is divided into four parts. The first part is the text reconstruction loss. The objective function utilized for content reconstruction is the binary cross entropy, which is computed between the predicted outcomes, denoted as $\hat{s}_i$, and the corresponding ground-truth values, represented by s_i.

$$L_c = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{V} \left(\mathbf{s}_{ij} \log\left(\hat{\mathbf{s}}_{ij}\right) + \left(1 - \mathbf{s}_{ij}\right) \log\left(1 - \hat{\mathbf{s}}_{ij}\right)\right) \tag{8}$$

The second and third parts are the reconstruction losses of the social and semantic relation graph, respectively. $\mathbf{A}$ is the adjacency matrix of the actual graph, while $\hat{\mathbf{A}}$ is the reconstructed graph predicted by the decoder:

$$\begin{aligned} L_{soc} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} (\mathbf{A}_{ij} \log(\hat{\mathbf{A}}_{ij}) \\ + (1 - \mathbf{A}_{ij}) \log(1 - \hat{\mathbf{A}}_{ij})) \end{aligned} \tag{9}$$

$$L_{sem} = \sum_{i}^{N} \sum_{j}^{N} |\mathbf{A}_{ij} - \hat{\mathbf{A}}_{ij}| \tag{10}$$

For the final part, which is the reconstruction loss of the summary. We use the mean cosine distance (denoted as d_{cos}). It measures the difference between hidden states of each encoded post ($\mathbf{h}_j$) and the encoded summary ($\mathbf{h}_s$).

$$L_{sim} = 1 - \frac{1}{N} \sum_{j=1}^{N} d_{cos}(\mathbf{h}_s, \mathbf{h}_j) \tag{11}$$

The sum of these four losses constitutes the final overall training objective:

$$L = L_c + L_{soc} + L_{sem} + L_{sim} \tag{12}$$

Multi-Facet Summarizer

The key idea of graph-based ranking is to compute the centrality score for each post. Traditionally, this score is measured through degree or ranking algorithms, based on PageRank. In this section, we introduce a variant method based on a directed graph, allowing the summary to cover a broader range of facets (Liang et al. 2021). Firstly, when calculating the centrality score for s_1, we regard posts with a large semantic gap from it as noise, filtering them by setting a threshold ϵ:

$$\begin{aligned} \mathrm{DC}\left(s_i\right) = \lambda_1 \sum_{i>j} \mathrm{Max}\left(\left(e_{ij} - \epsilon\right), 0\right) \\ + \lambda_2 \sum_{i<j} \mathrm{Max}\left(\left(e_{ij} - \epsilon\right), 0\right) \end{aligned} \tag{13}$$

Where each e_{ij} is calculated by the dot product of two posts $\mathbf{h}_i^T \mathbf{h}_j$, $\epsilon = \beta \cdot (\max(e_{ij}) - \min(e_{ij}))$, β is a hyperparameter. We use the following formula to rank posts and select the top k posts to be extracted as summary.

$$\text{summary } = \operatorname{topK}\left(\{\mathrm{DC}\left(s_i\right)\}_{i=1,..,N}\right) \quad (14)$$

We also present a method to model different aspects within posts. It adds a summary-document similarity that calculates the similarity between the candidate summary $\mathcal{C}$ and all posts d to measure the relevance. $\mathcal{C}$ is pre-selected from the posts with the highest k scores of $\mathrm{DC}\left(s_i\right)$ to reduce the search range. We combine summary-document similarity with sentence centrality to obtain the best candidate summary:

$$\text{summary} = \arg\max_{\mathcal{C}}\left(\operatorname{sim}(d, \hat{v}) \cdot \sum_{s_i \in \mathcal{C}} \mathrm{DC}(s_i)^{\alpha}\right) \quad (15)$$

Where α is a hyperparameter controlling the influence of directed centrality. $\operatorname{sim}(d, \hat{v})$ refers to the summary-document similarity, where d is the representation of whole posts and $\hat{v}$ is the candidate summary representation. $\hat{v}$ is obtained by $\frac{\sum_{i \in \mathcal{C}} v_i}{|\mathcal{C}|}$, which is the average representation of the summary sentences. We choose the cosine similarity for $\operatorname{sim}(\cdot)$. For d, we first collect all of post representations $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_N$. To compress all valuable information in the posts, we apply the max-pooling function to the sentence representations. The document representation d is calculated as:

$$d = \text{Maxpooling}\{\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_N\} \quad (16)$$

Experimental Setup

Dataset

We test our method using two real-world social media datasets: TWEETSUM (He, Zhao, and Liu 2020) and Weibo (Li et al. 2015), which are in English and Chinese respectively. TWEETSUM is a collection of tweets, focused on 12 specific topics. The tweets, totaling 44,034 from 11,240 users, are posted shortly before or after an event, reflecting the immediacy of the platform. Each set of tweets has four gold-standard summaries extracted by annotators. In our experiments, urls, mentions and other special characters are removed. The Weibo dataset is derived from Sina Weibo, a major social media platform in China. It consists of 130,000 posts from 126,000 users, centered 10 trending topics. Posts are organized as a tree-structure according to their interaction relations such as reposts and replies. Three editors write gold-standard abstractive summaries independently for each event.

In our experiments, we use ROUGE (Lin 2004) as our evaluation standard. This measures the quality of automatic summaries based on the overlap of n-grams between the golden references and the summaries produced by the model. The F-measures of ROUGE-1, 2, L and ROUGE-SU* are reported in our experiment results, where they are referred to as R-1, 2, L and R-SU*.

Implementation

For English tweet data, we use the pre-trained BERTWEET model as the content encoder. For Chinese Weibo posts, we use the pre-trained Chinese BERT model to obtain representations. During the training process, we apply the Adam optimizer, with an initial learning rate $lr = 0.01$, gradually reducing the learning rate. We also apply edge dropout (Rong et al. 2020) on the two relation graphs to avoid over-smoothing, and the edge dropout rate is also set to 0.3. The multi-facet summarizer has 4 hyperparameters, the best set of which is chosen from the following settings: $\alpha \in \{1, 2\}, \beta \in \{0.0, 0.1, \cdots 0.9\}, \lambda_1 \in \{0.0, 0.1, \cdots 1.0\}, \lambda_2 = 1 - \lambda_1$. In our experiments, we use grid search to find the combination with the highest R-1 score, and the best parameter combination is $\alpha = 1, \beta = 0.1, \lambda_1 = 0.4, \lambda_2 = 0.6$.

Comparison Methods

Our methodology is contrasted with two distinct categories of methods. The first category exclusively employs text content information and encompasses the subsequent methods. The **Centroid** (Radev, Blair-Goldensohn, and Zhang 2001) identifies sentences of high relevance to a cluster of related documents using centroid-based features. The **LSA** (Gong and Liu 2001) applies singular value decomposition to the feature matrix and selects posts with superior singular values. The **Lexrank** (Erkan and Radev 2004) employs a PageRank-like algorithm on the constructed similarity graph to select prominent posts. The **DSDR** (He et al. 2012) perceives summarization as a reconstruction task and extracts summaries by minimizing reconstruction error. The **MDS-Sparse** (Wang et al. 2015) conducts multi-document summarization using a sparse coding technique and extracts summaries by reconstructing the source document. The **PacSum** (Zheng and Lapata 2019) is a graph-based extractive model that uses BERT as sentence features and considers the relative position of posts and model documents as a directed graph. The **Spectral** (Liu et al. 2015) proposes a spectral-based hypothesis and identifies the importance of a sentence according to the spectral impact of posts. The **MT-GNN** (Doan, Nguyen, and Bui 2022) uses finer-grained semantic units in text posts to capture the complex relationship between words and posts, and constructs a heterogeneous graph neural network for extracting summaries.

The second category capitalizes on social relation and includes the subsequent models. The **SNSR** (He and Duan 2018) introduces a social regularization term into the sparse reconstruction framework to find salient tweets as summaries. The **SCMGR** (Liu et al. 2021) uses GCN on original social networks to extract socialized post representations and adopts a sparse reconstruction framework to select summary posts.

Results and Analysis

Main Results

Table 1 illustrates the performance of the MultiSum model and other comparative models on TWEETSUM and Weibo datasets. The first part encompasses three baseline methods: Expert, Oracle, and Random, representing the consistency between different standard reference summaries, the upper limit of performance, and the lower limit of performance.

	TWEETSUM				Weibo			
	R-1	R-2	R-L	R-SU*	R-1	R-2	R-L	R-SU*
Oracle	58.229	34.902	56.688	29.940	42.024	18.577	18.409	16.511
Expert	47.384	15.972	45.111	21.047	47.139	25.289	28.535	20.556
Random	41.480	9.6710	39.149	16.408	32.880	7.8740	12.504	10.483
Centroid	38.172	12.442	36.430	15.409	29.712	7.9580	13.718	9.7410
LSA	43.524	13.077	41.347	18.197	29.181	8.3750	12.717	8.9020
LexRank	42.132	13.302	39.965	18.192	34.802	8.1000	12.762	11.593
DSDR	43.335	13.106	41.055	17.264	19.771	5.3620	8.6790	4.5880
MDS-Sparse	42.119	10.059	40.101	16.686	33.019	7.5620	12.599	10.621
PacSum	42.603	13.021	40.375	17.268	32.664	8.7600	13.554	10.811
Spectral	43.488	11.980	41.229	17.794	33.862	8.6570	12.842	11.260
MTGNN	44.852	12.481	42.102	20.013	33.762	8.2190	13.558	11.519
SNSR	44.886	13.891	42.800	19.990	34.009	7.6220	12.566	10.925
SCMGR	45.829	14.081	43.433	20.141	36.405	10.495	14.469	12.722
MultiSum	**45.953**	**14.236**	**43.802**	**20.618**	**36.501**	**10.529**	**14.942**	**12.935**

Table 1: Results of MultiSum and comparison methods on TWEETSUM and Weibo

The second part displays the results of summarization methods that solely consider textual content, while the final part reveals the performance of summarization methods that integrate relationship between posts. Based on the experimental results, we have the following observations:

- The MultiSum model surpasses other comparative models on both datasets, validating the efficacy. These results indicate that our model is capable of integrating social network signals between posts and capturing semantic relationships, which alleviating the problem of shortage information in posts. Among all baseline models, SCMGR achieves the highest performance, suggesting that social relationships between posts can provide additional clues for content analysis, thereby enhancing model performance.
- Our model performs better on the TWEETSUM dataset than Weibo dataset. This is because the reference summaries in the TWEETSUM dataset are extractive, which is consistent with MultiSum, while the reference in Weibo is abstractive.
- Among all comparative methods, approaches that integrate relationships bewteen posts perform better than that only consider text content. Between models that utilize social relations, SNSR uses a rule-based method to model social relationships as simple regular terms. The SCMGR model uses GCN to integrate social relationship information and text content information of posts, which can more flexibly capture the social relationship structure between posts. However, this method overlooks the semantic relationships, which can also provide information for posts.

Ablation Study

For this work we focus on two issues. Specifically, 1) Does using GCN to fuse the two relations between posts help to improve the effect of social summarization; 2) Whether the multi-facet summarizer can improve the summary by considering the various facets when selecting posts. In response to these two questions, we conduct the following experiments on the English data set. To better understand the contribution of different modules to performance, we conduct an ablation study on the TWEETSUM dataset using our proposed MultiSum model. Initially, we remove the social relation encoder from the multi-stream encoder to explore the impact of social relationships on the model. We later drop the semantic relation encoder. To ascertain that our multi-stream encoder aids in producing improved post representations, we directly input the post representation after BERTWEET encoding into our summarizer.

TWEETSUM	R-1	R-2	R-L	R-SU*
MultiSum	45.953	14.236	43.802	20.618
-social	44.981	13.126	43.079	19.893
-semantic	45.122	13.841	43.213	20.182
-context	44.916	13.017	42.915	19.792

Table 2: Ablation study on TWEETSUM. We remove various modules and explore their influence on our model

The above three methods correspond to -social, -semantic, -context in Table 2. We observe that the performance scores of the model decreases after the removal of any module, validating the effectiveness of these modules. Specifically, the R-2 score decreases by more than 1% after removing the social relation encoder, indicating that the inclusion of social relationships allows the model to capture more keyword combinations. This phenomenon might suggest a high correlation between text content and its social relationships. Removing the semantic relation encoder leads to the most significant drop in R-1 and R-L scores. This confirms that integrating semantic relationships enables the model to learn more accurate word representations and the order relationships between words. Further, ROUGE scores are lower when we don't use the multi-stream encoder. This shows that by combining social and semantic relationships, the multi-stream encoder improves post representation.

Analysis of Multi-Facet Summarizer

In order to explore the improvement of multi-facet summarizer for model performance, we test it on the English dataset with another summarization methods. Subsequently, summaries are produced using the multi-facet summarizer (MFS) and the sparse reconstruction-based extractor (SRE) (Liu et al. 2021). It conducts a sparse reconstruction process by selecting tweets that can best reconstruct the original tweets in a specific topic. The ROUGE scores of both are shown in Table 3. According to the results, they have similar effects on the TWEETSUM dataset. The reason for this result may be that SRE uses diversity regularization to avoid duplicate posts, whereas MFS considers multiple aspects and selects representative posts among the candidate summaries.

TWEETSUM	R-1	R-2	R-L	R-SU*
MFS	44.916	13.017	42.915	19.792
SRE	44.886	13.891	42.800	19.990

Table 3: Results of MFS and SRE on TWEETSUM

In order to assess the capability of MFS in reducing facet bias and redundancy, we selected a sub-topic from TWEETSUM. We ask three human annotators to read all the tweets under this topic, a total of 787 tweets. Then read the summaries produced by the two types of summarizers. The three annotators are required to provide the number of facets and rate the redundancy of the summaries on a scale of 0-5, with lower scores indicating less redundant content. To ensure a fair evaluation of the summaries and avoid the influence of individual human differences, we also involve the most advanced large language models (LLMs), namely GPT-3.5, GPT-4, and Claude 2, in the evaluation process. The evaluation results are presented in Tables 4 and 5.

Model	**Human**	**GPT-3.5**	**GPT-4**	**Claude 2**
MFS	8	7	9	7
SRE	7	6	7	6

Table 4: Evaluations of covering multiple facets

Model	**Human**	**GPT-3.5**	**GPT-4**	**Claude 2**
MFS	2	2	4	2
SRE	3	4	4	3

Table 5: Evaluation of the redundancy level

In the assessment of the number of facets incorporated, we initially established that the tweets under the topic comprised a total of 10 facets. It is evident that the MFS includes more facets than the SRE across all evaluation systems, indicating that our model not only enhances the performance scores of the summaries but also yields summaries with a greater number of facets. GPT-4 aligns most closely with human evaluation results and is capable of identifying more facets than other LLMs. We infer that its stronger model capabilities and understanding of post text lead to this result. The evaluation outcomes of GPT-3.5 and Claude 2 are closely aligned, which is in line with our expectations. We further visualized GPT-4's evaluation by calculating the number of tweets corresponding to each facet in the summary and plotting Figure 2. As can be seen from it, the MFS not only covers more facets, but the distribution of the number of tweets among each facet is also more balanced.

In terms of redundancy level assessment, MFS achieved lower redundancy than SRE in all evaluations, with the exception of GPT-4. The evaluation results of GPT-4 for the two models are relatively proximate, but there is a substantial gap with other LLMs and human evaluation results. We surmise that this outcome may be attributed to GPT-4's more strict evaluation standards, resulting in elevated redundancy

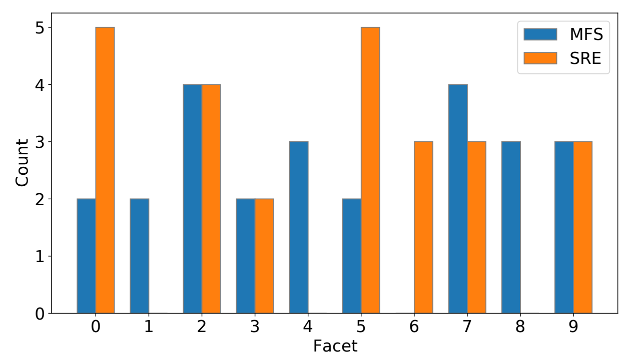

Figure 2: Number of Tweets in Each Facet

Conclusion

In this paper, we propose an unsupervised extractive method, which improves social summarization by considering multiple relationships between posts and various facets in summary simultaneously. We first employ GCNs to collect information about the neighbors of post from both social and semantic relations, which mitigates the information shortage about a single post. Next, in selection phase, we introduce an improved ranking algorithm to alleviate the facet bias caused by centrality calculation approaches and improve the diversity of aspects in the summary. Experiments on two real-world datasets prove the effectiveness of our approach, and evaluations by both humans and three large language models reveal that the summaries cover more viewpoints with less redundancy.

Acknowledgments

Our work is supported by the National Natural Science Foundation of China (No.61976154), the National Natural Science Foundation of China (No.62376192) and the National Natural Science Foundation of China (No.62376188).

References

Abelson, R. P.; and Prentice, D. A. 1983. Whatever became of consistency theory? *Personality and Social Psychology Bulletin*, 37–54.

Ali, S. M.; Noorian, Z.; Bagheri, E.; Ding, C.; and Al-Obeidat, F. 2020. Topic and Sentiment Aware Microblog Summarization for Twitter. *Journal of Intelligent Information Systems*, 129–156.

Alsaedi, N.; Burnap, P.; and Rana, O. F. 2016. Automatic Summarization of Real World Events Using Twitter. In *Proceedings of the 2016 ICWSM*, 511–514.

Andy, A.; Wijaya, D. T.; and Callison-Burch, C. 2019. Winter is here: Summarizing twitter streams related to pre-scheduled events. *In Proceedings of the Second Workshop on Storytelling*, 112–116.

Brin, S.; and Page, L. 1998. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 107–117.

Cacioppo, J. T.; Fowler, J. H.; and Christakis, N. A. 2009. Alone in the crowd: The structure and spread of loneliness in a large social network. *Journal of personality and social psychology*, 1–31.

Centola, D. 2010. The spread of behavior in an online social network experiment. *science*, 1194–1197.

Centola, D.; Becker, J.; Brackbill, D.; and Baronchelli, A. 2013. The simple rules of social contagion. *Scientific reports*, 1–7.

Chang, Y.; Wang, X.; Mei, Q.; and Liu, Y. 2013. Towards twitter context summarization with user influence models. *In Proceedings of the 2013 WSDM*, 527–536.

Chen, H.; He, J.; Xu, W.; Feng, T.; Liu, M.; Song, T.; Yao, R.; and Qiao, Y. 2023. Enhanced Multi-Relationships Integration Graph Convolutional Network for Inferring Substitutable and Complementary Items. In *Proceedings of the 2023 AAAI*, 4157–4165.

Chu, E.; and Liu, P. J. 2019. MeanSum: A Neural Model for Unsupervised Multi-document Abstractive Summarization. In *Proceedings of the 2019 ICML*, 1223–1232.

Dixit, T.; Wang, F.; and Chen, M. 2023. Improving Factuality of Abstractive Summarization without Sacrificing Summary Quality. In *Proceedings of the 2023 ACL*, 902–913.

Doan, X.-D.; Nguyen, L.-M.; and Bui, K.-H. N. 2022. Multi Graph Neural Network for Extractive Long Document Summarization. In *Proceedings of the 2022 COLING*, 5870–5875.

Dong, Y.; Romascanu, A.; and Cheung, J. C. K. 2021. HipoRank: Incorporating Hierarchical and Positional Information into Graph-based Unsupervised Long Document Extractive Summarization. In *Proceedings of the 2021 ACL*, 1629–1640.

Dutta, S.; Das, A. K.; Bhattacharya, A.; Dutta, G.; Parikh, K. K.; Das, A.; and Ganguly, D. 2018. Community detection based tweet summarization. *Advances in Intelligent Systems and Computing*, 797–808.

Erkan, G.; and Radev, D. R. 2004. Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of artificial intelligence research*, 457–479.

Fowler, J. H.; and Christakis, N. A. 2008. Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study. *BMJ*, 1–9.

Ganesan, K. A.; Zhai, C.; and Viegas, E. 2012. Micropinion generation: an unsupervised approach to generating ultra-concise summaries of opinions. *In Proceedings of the 2012 WWW*, 869–878.

Gao, S.; Cheng, X.; Li, M.; Chen, X.; Li, J.; Zhao, D.; and Yan, R. 2023. Dialogue Summarization with Static-Dynamic Structure Fusion Graph. In *Proceedings of the 2023 ACL*, 13858–13873.

Gillani, M.; Ilyas, M. U.; Saleh, S.; Alowibdi, J. S.; Aljohani, N.; and Alotaibi, F. S. 2017. Post summarization of microblogs of sporting events. *In Proceedings of the 2017 WWW*, 59–68.

Gong, Y.; and Liu, X. 2001. Generic Text Summarization Using Relevance Measure and Latent Semantic Analysis. In *Proceedings of the 2001 SIGIR*, 19–25.

He, R.; and Duan, X. 2018. Twitter summarization based on social network and sparse reconstruction. In *Proceedings of the 2018 AAAI*, 5787–5794.

He, R.; Zhao, L.; and Liu, H. 2020. TWEETSUM: Event oriented Social Summarization Dataset. In *Proceedings of the 2020 COLING*, 5731–5736.

He, Z.; Chen, C.; Bu, J.; Wang, C.; Zhang, L.; Cai, D.; and He, X. 2012. Document summarization based on data reconstruction. In *Proceedings of the 2012 AAAI*, 620–626.

Inouye, D.; and Kalita, J. K. 2011. Comparing twitter summarization algorithms for multiple post summaries. *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, 298–306.

Jing, B.; Park, C.; and Tong, H. 2021. HDMI: High-order Deep Multiplex Infomax. *In Proceedings of the 2021 WWW*, 2414–2424.

Jing, B.; Xiang, Y.; Chen, X.; Chen, Y.; and Tong, H. 2021a. Graph-MVP: Multi-View Prototypical Contrastive Learning for Multiplex Graphs. *arXiv*.

Jing, B.; You, Z.; Yang, T.; Fan, W.; and Tong, H. 2021b. Multiplex Graph Neural Network for extractive text summarization. *In Proceedings of the 2021 EMNLP*, 133–139.

Kano, R.; Miura, Y.; Taniguchi, M.; Chen, Y.-Y.; Chen, F.; and Ohkuma, T. 2018. Harnessing Popularity in Social Media for Extractive Summarization of Online Conversations. In *Proceedings of the 2018 EMNLP*, 1139–1145.

Kipf, T. N.; and Welling, M. 2016. Variational graph auto-encoders. *arXiv*.

Kipf, T. N.; and Welling, M. 2017. Semi-supervised classification with graph convolutional networks. In *Proceedings of the 2017 ICLR*, 1–14.

Li, J.; Gao, W.; Wei, Z.; Peng, B.; and Wong, K.-F. 2015. Using Content-level Structures for Summarizing Microblog Repost Trees. In *Proceedings of the 2015 EMNLP*, 2168–2178.

Li, Y.; Peng, B.; He, P.; Galley, M.; Yu, Z.; and Gao, J. 2023. DIONYSUS: A Pre-trained Model for Low-Resource Dialogue Summarization. In *Proceedings of the 2023 ACL*, 1368–1386.

Liang, X.; Wu, S.; Li, M.; and Li, Z. 2021. Improving unsupervised extractive summarization with facet-aware modeling. *In Proceedings of the 2021 ACL*, 1685–1697.

Lin, C.-Y. 2004. ROUGE: A Package for Automatic Evaluation of Summaries. In *Text summarization branches out*, 74–81.

Liu, F.; Liu, Y.; Liu, F.; and Zhao, D. 2015. A spectral method for unsupervised multi-document summarization. In *Proceedings of the 2015 IJCAI*, 2279–2285.

Liu, H.; He, R.; Zhao, L.; Wang, H.; and Wang, R. 2021. SCMGR: Using Social Context and Multi-Granularity Relations for Unsupervised Social Summarization. In *Proceedings of the 2021 CIKM*, 1058–1068.

Liu, W.; Fu, X.; and Strube, M. 2023. Modeling Structural Similarities between Documents for Coherence Assessment with Graph Convolutional Networks. In *Proceedings of the 2023 ACL*, 7792–7808.

Liu, X.; Li, Y.; Wei, F.; and Zhou, M. 2012. Graph-Based Multi-Tweet Summarization using Social Signals. In *Proceedings of the 2012 COLING*, 1699–1714.

Liu, Y.; Jia, Q.; and Zhu, K. 2022. Length control in abstractive summarization by Pretraining Information Selection. *In Proceedings of the 2022 ACL*, 6885–6895.

Liu, Y.; Liu, P.; Radev, D.; and Neubig, G. 2022. Brio: Bringing order to abstractive summarization. *In Proceedings of the 2022 ACL*, 2890–2903.

Ma, S.; Sun, X.; Lin, J.; and Wang, H. 2018. Autoencoder as Assistant Supervisor: Improving Text Representation for Chinese Social Media Text Summarization. In *Proceedings of the 2018 ACL*, 725–731.

Nguyen, D. Q.; Vu, T.; and Tuan Nguyen, A. 2020. BERTweet: A pre-trained language model for English Tweets. In *Proceedings of the 2020 EMNLP*, 9–14.

Radev, D. R.; Blair-Goldensohn, S.; and Zhang, Z. 2001. Experiments in single and multidocument summarization using MEAD. In *First document understanding conference*, 1–7.

Radev, D. R.; Jing, H.; and Budzikowska, M. 2000. Centroid-based summarization of multiple documents: sentence extraction, utility-based evaluation, and user studies. In *Proceedings of the 2000 NAACL*, 21–29.

Rong, Y.; Huang, W.; Xu, T.; and Huang, J. 2020. DropEdge: Towards Deep Graph Convolutional Networks on Node Classification. In *Proceedings of the 2020 ICLR*, 1–17.

Shalizi, C. R.; and Thomas, A. C. 2010. Homophily and Contagion Are Generically Confounded in Observational Social Network Studies. *Sociological Methods & Research*, 211–239.

Sharifi, B.; Hutton, M.-A.; and Kalita, J. K. 2010. Summarizing Microblogs Automatically. In *Proceedings of the 2010 NAACL*, 685–688.

Sharifi, B. P.; Inouye, D. I.; and Kalita, J. K. 2013. Summarization of twitter microblogs. *The Computer Journal*, 378–402.

Sharma, A.; Saxena, A.; Gupta, C.; Kazemi, S. M.; Talukdar, P. P.; and Chakrabarti, S. 2022. TwiRGCN: Temporally Weighted Graph Convolution for Question Answering over Temporal Knowledge Graphs. In *Proceedings of the 2022 EACL*, 2049–2060.

Wang, R.; Luo, S.; Pan, L.; Wu, Z.; Yuan, Y.; and Chen, Q. 2019. Microblog summarization using paragraph vector and semantic structure. *Computer Speech and Language*, 1–19.

Wang, Z.; Liang, X.; Xu, B.; Liang, J.; and Duan, H. 2015. Multi-document summarization based on two-level sparse representation model. In *Proceedings of the 2013 AAAI*, 196–202.

Yan, Y.; Liu, L.; Ban, Y.; Jing, B.; and Tong, H. 2021. Dynamic Knowledge Graph Alignment. In *Proceedings of the 2021 AAAI*, 4564–4572.

Yang, K.; Bao, Q.; and Qiu, H. 2023. Identifying Multiple Propagation Sources With Motif-Based Graph Convolutional Networks for Social Networks. *IEEE Access*, 61630–61645.

Zheng, H.; and Lapata, M. 2019. Sentence Centrality Revisited for Unsupervised Summarization. In *Proceedings of the 2019 ACL*, 6236–6247.