Journal of Computer and Knowledge Engineering, Vol. , No.. 2019.
DOI:

Description-based Post-hoc Explanation for Twitter List Recommendations*

Havva Alizadeh Noughabi[1], Behshid Behkamal[2], and Mohsen Kahani[3]

***Abstract*-- Twitter List recommender systems can generate highly accurate recommendations, but since they employ heterogeneous information of users and Lists and apply sophisticated prediction models, they may not provide easy understandable intrinsic explanations. To address this limitation, Twitter List descriptions can play a critical role in providing post-hoc explanations that help users make informed decisions. In this paper, we present a model to provide relevant and informative explanations for recommended Twitter Lists by automatically generating descriptions for them. The model selects the most informative tweets from a List as its description to inform users more with the recommended List that positively contributes to the user experience. More specifically, the explanation model incorporates three categories of features: *content relevance* features, *tweet-specific* features and *publisher's authority* features that are used in a learning to rank model to rank the List's tweets in terms of their informativeness. Experimental results on a Twitter dataset validate the effectiveness of our proposed model in generating useful explanations for recommended Twitter Lists.**

***Index Terms*-- Explainable recommender systems, Post-hoc explanation, Description generation, Twitter Lists.**

I. Introduction

With the substantial increase in user-generated content on social media, several platforms assist users in organizing related information into a single bin. For instance, Twitter introduced *Lists* as a solution to tackle the issue of information overload [1]. A Twitter List is a group of accounts that anyone with an interest in the topics covered by the List can subscribe to for free. While List recommender systems have been highly effective in using of different user and List features as well as advanced hybrid models to improve their performance [2], [3], their lack of explainability remains a significant challenge.

Nowadays, the user experience with social media platforms, which includes factors such as trust, understandability, and satisfaction, is increasingly influenced by the availability of explainability in social recommender systems [4], [5]. This has motivated us to provide informative explanations for the recommended Twitter Lists. Users who were provided with explanations for recommended Lists were found to be more inclined to engage with them and to perceive the recommendations as relevant and useful.

On the other hand, the development of methods to provide post-hoc explanations, which are generated after recommendations have been made, has recently attracted a lot of interest in the field of recommender systems [6]–[8]. Post-hoc explanations can make it easier for users to make informed decisions about the recommendations by providing them with detailed information about the recommended item. This is especially useful for users who may not be familiar with the technical aspects of the recommender system and may find it difficult to understand intrinsic explanations, which can be too complex or technical for them. Our main goal is to provide a post-hoc explanation for the recommended Twitter List, which not only helps to maintain the predictive accuracy of the current complex recommender systems but also ensures that users can gain more information and insights into the recommended Lists.

Description of Twitter Lists can plays an important role to provide post-hoc explanations for the recommended Lists. However, only a few popular Lists have a description written by the List owner on Twitter. For example, the List named *NTD* does not have any informative description on Twitter and it is quite hard for users to understand the main topics of the List or guess the content of it. For such Lists, the user needs to manually check the tweets of the List and read some of its recent tweets to figure out the topics of it which is tedious and highly time-consuming. Automatically generating an accurate and informative description about Lists on Twitter not only helps users make an informed decision but also improves the likelihood of subscribing to Lists on Twitter.

Recently, there has been increasing interest in automatic item description generation [9]–[11]. Some studies have

* Manuscript received , accepted .
[1] PhD Candidate, Department of Computer Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
[2] **Corresponding author.** Associate Professor, Department of Computer Engineering, Ferdowsi University of Mashhad, Mashhad, Iran **Email**: behkamal@um.ac.ir
[3] Professor, Department of Computer Engineering, Ferdowsi University of Mashhad, Mashhad, Iran

utilized statistical frameworks, such as [12], that incorporate statistical methods with templates for generating product descriptions. Other research, like [13], have proposed to generate summarization of item reviews by applying a template based Natural Language Generation (NLG) framework. To overcome the need for handcrafted templates, researchers have turned to deep learning-based models and presented various conditions to the generation model [14], thereby addressing the limitations of traditional approaches. In our paper, to generate a Twitter List's description, our main idea is to select the most informative tweets of a List as its description to inform users more with the recommended List that improves their overall experience.

Specifically, compared to existing explainable models with various goals [14], [15], our work focuses on the recommended Twitter List to present an informative explanation with the aim of helping users to know more about the content of the recommendation. Therefore, providing a set of K tweets from the recommended List can give pertinent information to users, although despite noisy tweets, considering the quality of selected tweets is so important to obtain a relevant and informative tweet-level explanation. In this paper, to catch the high quality description for Twitter Lists, we use three feature categories to consider various perspectives: *Content Relevance Features*, *Tweet Specific Features* and *Publisher's Authority Features*. Then highly-useful tweets are obtained which provide explanations to help users know more about the List.

To make our idea more clear, we give an example. Consider a recommended Twitter List named *'Web Dev'* that has many tweets within it that are posted by its members. Some of which are below and let's assume that one of these tweets is going to be presented as an explanation with the aim of helping users to know more about the content of the List.

A) "Roadmap for web development: 1-HTML 2-CSS 3-CSS frameworks 4-JS 5-DOM 6-Git and GitHub 7-React 8-Node.js 9-API 10-Database 11-Web3.js 12-Solidity."

B) "JavaScript is Awesome!"

C) "It's 5:30 and the sun is still up."

D) "Keep wearing masks if you want to survive. Dropping all the mitigation measures will bring another wave."

It is evident that tweets C and D are not associated with the primary topic of the '*Web Dev'* List. For example, members posted tweet D due to trending topics of it (Covid19). Tweets A and B are related and can be considered as potential explanations to provide information about the List. Given that tweet A contains more detailed information that helps to clarify the recommended List, it is considered to be more informative than tweet B.

The main contributions of the paper are summarized as follows.

- We propose a method to give a post-hoc explanation for the recommended List on Twitter to help users be more successful in decision making.
- We propose an explanation ranking model to catch a high quality description for a recommended List by utilizing three feature categories: content relevance features, tweet-specific features, and publisher's authority features.
- By conducting experiments on a Twitter dataset, we have shown that our tweet-level explanation can provide helpful information about the recommended List for users.

The structure of this paper is as follows. In the subsequent section, we provide an overview of related research on explainable social recommendation. We outlines the proposed model that is designed to generate an informative post-hoc explanations for the recommended Twitter List in Section 3. Our evaluation of the model is presented in Section 4, and we conclude with Section 5.

II. Related Works

The explainability of recommendation models is considered so important in order to increase users' trust and encourage them to adopt recommender systems. These explanations may serve various purposes, such as transparency, effectiveness, trust, persuasiveness, satisfaction, scrutability, and efficiency [14], [16]. To achieve a specific purpose, it is essential to carefully consider what information should be conveyed to the user through the explanation facility [15].

Explainable recommendation models can be classified into two categories: *model-intrinsic* and *model-agnostic* [17]. Model-agnostic approaches generate explanations (often called the post-hoc explanation) using separate models or techniques that are independent of the recommendation algorithm. Post-hoc explanations are often more flexible and can be applied to a wider range of recommendation algorithms, but they may not be as accurate or specific compared to intrinsic explanations. Frequently employed techniques for post-hoc explanation generation include surrogate models [18],[19] and data mining methods such as association rule mining [20] or subgraph discovery [21]. Despite their approximate nature, post-hoc approaches are effective in maintaining the accuracy of the underlying model [22]. Intrinsic explanations are generated using features that are built into the recommendation algorithm itself. These features aim to provide a more detailed and specific explanation of why a particular item was recommended. Previous research has employed various explainable models to generate intrinsic explanations including factorization models [23],[24], knowledge graph based models [25]–[27], deep learning models [28]–[30], and rule-based models [31], [32].

The explainability of *social recommender systems* is vital in establishing users' trust in the recommendations, which is fundamental to maintain the sustainability of social networks [33]. In some previous studies on explainable social recommender systems, models provide explanations based on the user's social network [34], [35]. For example, Wang et al. [35] have developed social explanations that follow the structure of "A and B also like the item". To generate the pertinent social explanation, they developed an algorithm for identifying the optimal set of users to include in the explanation. In some other works, a heterogeneous information network is created according to user and item information and then explanations are presented as paths connecting users and the recommended items [36], [37]. For instance, Zhang et al. [37] have introduced a knowledge distillation approach to explain black-box models for recommendation. Given an embedding-based model that generates black-box recommendations, their proposed approach explained recommended items based on differentiable paths on the knowledge graph. Recently, researchers in the field of social recommender systems have started to consider user-generated content such as reviews and posts as a form of explanation for the recommendations made by these systems [38]–[40]. For instance, Ren et al. [38] have presented the social collaborative viewpoint regression model, which utilize viewpoints as explanations. These viewpoints are represented by a combination of concept, topic, and sentiment label that is extracted from both user reviews and social connections.

Compared with explainable social recommendation, our paper focuses on providing an informative post-hoc explanation for the recommended Twitter List by automatically creating descriptions based on the List's content. We believe that such explanations can assist users in making well-informed decisions.

III. Methodology

This section is devoted to the formulation of our model to generate an informative explanation for the recommended List. Formally, let L be the set of Twitter Lists and $l \in L$ is a recommended List, given $M_l = \{m_1, m_2, \ldots, m_N\}$ as a tweet collection of l, we aim to identify top-K most informative and relevant tweets of l, as a post-hoc explanation. To address this problem, the proposed approach, comprising several components, is illustrated in Fig. 1.

Initially, a recommender system provides a recommended List, denoted as l. For the set of tweets in l (i.e., M_l), three categories of features are extracted within the *Feature Extraction* component. These tweets are subsequently ranked based on their informativeness and relevance to the main topic of the List in *Tweet Ranking* component. To train the ranking model, a learning-to-rank technique is applied, ensuring that the most relevant and informative tweets are prioritized. The following subsections provide a detailed introduction to these components.

A. Feature Extraction

Our emphasis when defining features is on *Informativeness* and *Relevance* of List's tweets that are utilized to provide an explanation. Adopted from [41], in our work, *Relevance* means the degree of content closeness to the main subject of recommended Twitter List and *Informativeness* means degree of information acceptability which can explain the recommended List understandable to maximum people. The explanation ranking model employs three distinct categories of features, which are: (1) *content relevance* features, (2) *tweet-specific* features and (3) *publisher's authority* features. We provide further details about these categories below.

1) Content Relevance Features

The features of this category are used to measure the relatedness of a tweet from the recommended List $l \in \mathrm{L}$, i.e., $m \in M_l$, to the main subject of l. According to our intuition, an explanation will be more beneficial to the user if it is more semantically comparable to the recommended List. Below is a description of two features belonging to this category.

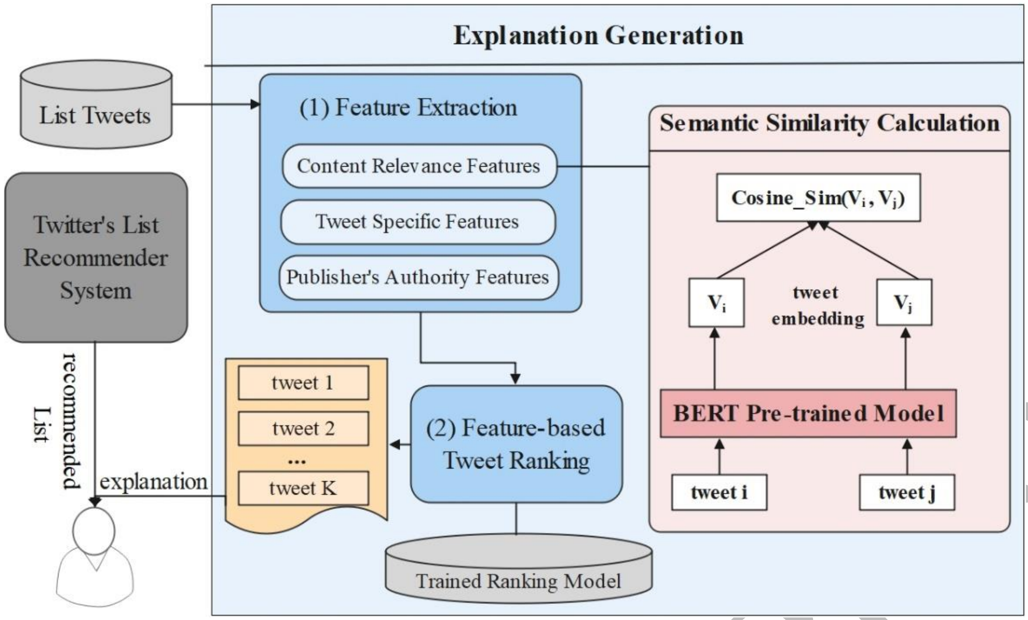

Fig. 1. Overview of the proposed explanation generation model

1) Semantic Relatedness: we compute the semantic similarity between the tweet $m \in M_l$ and the l's tweets (up to N:100 tweets) as a relatedness score. We use the sentence-BERT model [42] to get the embedding vectors and then compute the cosine similarity of the two embedding vectors. We calculate relatedness score as follows:

$$relatedness\_score(m,l) = \frac{1}{N}\sum_{i=1}^{N} cosine\_similarity(\vec{m}, \vec{m}_i)$$

Where m is a tweet of l which is compared with other tweets of l (i.e. m_i). If m is semantically related to the main subject of l, the relatedness score will be higher than if m is a noise tweet.

2) Relevance to Hashtags: A number of hashtags may be used in the $m \in M_l$ to highlight its main keywords. This feature determines the count of hashtags that are present in tweet m and are also included in the set of l's top-10 hashtags. (top-10 hashtags of l is specified through tweet history of it).

2) Tweet Specific Features

The quality of the tweet $m \in M_l$ regardless of its relatedness to the main subject of l, is measured by this category of features. Our hypothesis is that a tweet will be more helpful to serve as an explanation if it is more popular and informative than a tweet. The following features are included in this category, which were inspired by those mentioned in the literature on tweet ranking [43]–[46]:

1) Length: It is determined by how many words a tweet contains. Intuitively, a longer tweet is more likely to contain a greater amount of information than a shorter one.

2) Retweet Count: It is described as the quantity of retweets a tweet receives. The fundamental idea is that a tweet is more informative and useful if it is retweeted frequently.

3) Favorite Count: A tweet's quality and amount of appreciation may be suggested by the frequency with which users have expressed positive feelings about it.

4) URL Count: Publishers frequently augment their tweets with URLs that direct readers to more information on different web pages. As a result, the number of URLs in a tweet may affect how informative it is.

5) Hashtag Count: A tweet becomes more informative and useful the more hashtags it has.

3) Publisher's Authority Features

This particular group of features quantifies the level of authority of the individual who posted the tweet. According to our hypothesis, tweets that are shared on social media by more authoritative users, are perceived to be of higher quality and more compelling, and thus may be more effective in providing a description for the recommended Twitter List to the user. Adopted from [43]–[45], the following features are considered as potential markers to determine a user's authority:

1) Follower Count: The amount of followers a user has, is recorded by this feature.

2) Status Count: This feature counts the total amount of tweets a user has ever posted.

3) Mention Count: This feature is used to estimate the number of times a user is mentioned in tweets.

B. Feature-based Tweet Ranking

Given M_l, the set of tweets within the recommended list l, our objective is to rank these tweets, identifying top-K most relevant and informative ones as the descriptive explanation. To achieve this, each $m \in M_l$ is represented by a feature vector, where each dimension of the vector quantifies the relevance and informativeness of m with respect to l. Specifically, utilizing the features extracted in the previous component (i.e., Feature Extraction), we apply a learning-to-rank model to efficiently rank the tweets. In the following, we describe the process of training the ranking model illustrated in Fig. 1, titled *Trained Ranking Model*, which is utilized by the *Feature-based Tweet Ranking* component.

First, to collect the training data, we randomly selected 100 Twitter Lists from our dataset and randomly chose 30 tweets from each List as potential explanations. These tweets were then manually annotated according to the specified annotation guidelines, resulting in the creation of an explanation pool. Each explanation was assessed by human annotators and assigned one of three points according to the following criteria:

0: Explanation is unrelated to the main subject of the List.

1: Explanation is related to the List but lacks informativeness.

2: Explanation is both related to the List and informative.

Table 1 presents examples of explanations for each annotation category. With this annotated data, we proceeded to train a feature-based learning-to-rank (LTR) model, employing LambdaMART specifically. The experiments aimed at identifying the optimal LTR method are elaborated in Section IV.B.

This trained ranking model enables the systematic ranking of tweets based on their feature vectors, ensuring that the most relevant and informative tweets are prioritized effectively. This capability represents a critical step towards enhancing the quality and relevance of descriptive explanations for recommended Twitter Lists. Indeed, the *Feature-based Tweet Ranking* component, leveraging the trained ranking model, has the capacity to organize input tweets according to their quality, with a focus on relatedness and informativeness.

Table 1. Explanation annotation examples for each category.

Human annotation	**List name**	**Explanation**
unrelated	virus scientists	*"What'd you eat for breakfast?"*
	crypto currency	*"The weather is perfect for along run."*
related and non-informative	virus scientists	*"Wearing a mask isn't fun."*
	crypto currency	*"Don't trip, buy the dip[4]!"*
related and informative	virus scientists	*"New on 3rd shot (booster) effectiveness vs Omicron infections from Spain in 7 million people: 51%, across all age groups, Moderna 13% better than Pfizer."*
	crypto currency	*"At the very least, one should be expecting a #Bitcoin bounce right around now. STH-SOPR has fallen below 1, which means top-buyers are spending their $BTC at a realized loss. When top buyers capitulate, it is historically a local bottom."*

IV. Experiments

A. Dataset

We collected a dataset from Twitter using Tweepy[5]. Similar to [47], the crawling process began with the Lists of 'Ashton Kutcher', a well-known user on Twitter. Given his Lists, we initially gathered all users who have subscribed to these Lists in order to expand the set of users. Then, we added more Lists to the collection by gathering every List that these users had subscribed to. After four iterations, our final dataset includes roughly 17,000 Lists covering a diverse range of topics. By investigating the random subset of 3000 Lists, it is realized that 61% of them do not have a description written by the List's owner. For the remaining 39%, we depict the Twitter Lists distribution by the number of description tokens in Fig. 2 which states that the length of List's description is short in most cases. Therefore, providing a description with the amount of related information is important to help users be more successful in decision making.

[4] dip in the world of cryptocurrency stands for 'Drop In Price'

[5] https://www.tweepy.org/

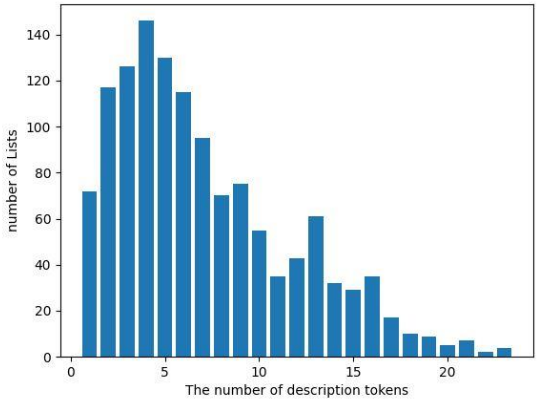

Fig. 2. Twitter Lists distribution by count of description tokens

B. Experimental Settings

1) BERT-Sentence. As explained in Section 3, we utilize BERT-Sentence to measure the relatedness between a tweet and the main subject of the recommended List. Before applying BERT-Sentence, we slightly preprocessed the tweets by removing special characters, URLs and mentions. We apply the pre-trained model (i.e., "all-MiniLM-L6-v2") to transform the tweets into dense vector embeddings in our experiments.

2) Learn to Rank. In our experiment, we used the RankLib[6] library for learning to rank (LTR). We used three machine learning techniques to train the explanation ranking model, consisting of one linear method (Coordinate Ascent [48]) and two non-linear methods (MART [49] and LambdaMART [50]). To select the best-performing LTR method, given the train dataset, a 5-fold cross validation approach is applied to evaluate various ranking models in terms of Expected Reciprocal Rank (ERR) and Normalized Discounted Cumulative Gain (NDCG). Considering the results in Table 2, LamdaMART is selected for the rest of our experiments.

Table 2. NDCG and ERR reported by different LTR models on test data

Model	**NDCG@10**	**ERR@10**
Coordinate Ascent	0.921	0.832
MART	0.884	0.798
LamdaMART	0.928	0.840

C. Evaluation

Similar to [51], we evaluated two aspects of our proposed model: (1) the generated explanations' quality with regards to relatedness and informativeness to the recommended List through a user study. (2) the significance of the ranking component by a pairwise analysis. As explained in Section 3, we identify top-K most informative and relevant tweets of the recommended List, as a post-hoc explanation. We set K to one in our experiments.

1) Explanations' quality evaluation by user study

In order to assess the quality of the generated explanations, we initially selected 100 Lists at random and then the proposed model generated an explanation for each one. These explanations were then annotated by human annotators according to the guidelines introduced in Section III.B.

The results of user study are reported in Fig. 3 which shows the percentage of explanations annotated by each label. Based on our findings, we observed that the number of explanations that were both informative and relevant was higher than the number of explanations belonging to other categories.

2) Analysis on the Ranking Component

As discussed in Section 3, the tweets of recommended List *l* as potential explanations are ranked by the ranking component based on their relatedness to the main subject of the recommended List and their informativeness. To investigate the impact of ranking component on the quality of the final explanation, we design a pairwise evaluation. In detail, for randomly selected 100 Lists, the explanation generated by our model named *A* and the randomly selected potential explanation named *B*. Human annotators conducted pairwise evaluations between two explanations, using one of the following points:

1: A is more related and informative than B.

2: B is more related and informative than A.

3: A and B are almost the same, both related and equally informative.

4: A and B are almost the same, both unrelated.

Fig. 4 illustrates the results of the pairwise evaluation. For 52% of recommended Lists, top-1 ranked explanations are more related and informative than the randomly selected potential ones. In addition, for 39% of recommended List, it is challenging for annotators to determine which one is better because both explanations are useful. We conclude that in the majority of cases, top-1 ranked explanations perform as well as, or even better than, other potential explanations, which demonstrates the explanation ranking component's beneficial impact to improving explanation quality. potential ones. In addition, for 39% of recommended List, it is challenging for annotators to determine which one is better because both explanations are useful. We conclude that in the majority of cases, top-1 ranked explanations perform as well as, or even better than, other potential explanations, which demonstrates the explanation ranking component's beneficial impact to improving explanation quality.

3) Feature Analysis

To specifically evaluate the effectiveness of each feature group (i.e., content-relevance, tweet-specific and publisher's authority), we performed an ablation study in which we removed the features of every category separately. The performance of LamdaMART as affected by the ablation study is presented in Table 3. In this table, symbols * shows statistical

[6] https://sourceforge.net/p/lemur/wiki/RankLib/

significance on a paired t-test with p-value less than 0.05. According to the findings, all three feature categories are effective on how well the explanation ranking model performs, and performance is decreased in terms of NDCG and ERR by removing them. The performance of the ranking component is demonstrated to be more significantly impacted by content relevance features and tweet-specific features than by publisher's authority features.

Furthermore, to assess the individual impact of each feature, we conducted an ablation study by systematically removing one feature at a time. The results of this study, presented in Table 4, provide valuable insights into the relative importance of features within each category. The findings highlight *semantic relatedness* as the pivotal feature within the content relevance category, with its removal resulting in a significant decrease in NDCG, from 0.928 to 0.889. *Tweet length* emerges as the primary feature within the tweet-specific category. Conversely, *follower count* within the publisher's authority category exhibits a negative impact on NDCG. Its removal leads to a slight increase in NDCG, from 0.928 to 0.931. Overall, our analysis demonstrates that the majority of features significantly contribute to the effectiveness of ranking explanations.

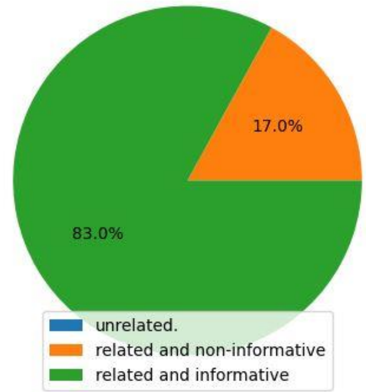

Fig. 3. The results of user study

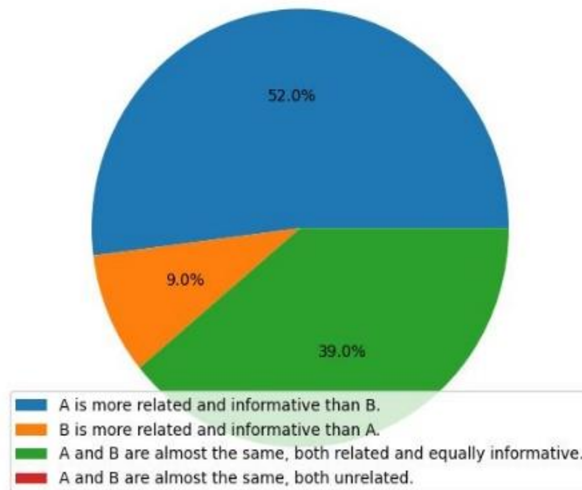

Fig. 4. Pairwise annotation results

Table 3. Ablation study results (feature category removal).

	NDCG@10	▼	ERR@10	▼
All Features	0.928		0.840	
- Content Relevance Features	0.879	5.28 % *	0.812	3.33 % *
- Tweet Specific Features	0.868	6.47 % *	0.693	17.5 % *
- Publisher's Authority Features	0.926	0.22 %	0.833	0.83 %

Table 4. Ablation study results (feature removal).

	NDCG@10	ERR@10
All Features	0.928	0.840
- Semantic Relatedness	0.889	0.828
- Relevance to Hashtags	0.923	0.824
- Length	0.894	0.807
- Retweet Count	0.926	0.839
- Favorite Count	0.924	0.831
- URL Count	0.921	0.832
- Hashtag Count	0.916	0.829
- Follower Count	0.931	0.838
- Status Count	0.925	0.839
- Mention Count	0.924	0.830

V. CONCLUSION

In this paper, in order to help users to make an informed decision on social media, we proposed a post-hoc explanation model for List recommendations on Twitter. The proposed model provides a high quality description as explanation using the content of the recommended List. More specifically, by using three feature categories from different aspects, our model ranks explanations according to their relatedness and informativeness. In our experiments, the quality of final explanations are evaluated by user study. Also, the importance of the explanation ranking component is investigated.

In the current work, we only use one related and informative tweet of the recommended List as a final explanation. In future studies, our aim is to determine the minimum number of tweets required to provide reliable indicators of the usefulness of an explanation.

References.

[1] S. de la Rouviere and K. Ehlers, “Lists as coping strategy for information overload on Twitter,” presented at the Proceedings of the 22nd International Conference on World Wide Web, 2013, pp. 199–200.

[2] V. Rakesh, D. Singh, B. Vinzamuri, and C. K. Reddy, “Personalized recommendation of twitter lists using content and network information,” presented at the Proceedings of the International AAAI Conference on Web and Social Media, 2014, pp. 416–425.

[3] L. Chen, Y. Zhao, S. Chen, H. Fang, C. Li, and M. Wang, “iplug: Personalized list recommendation in twitter,” presented at the Web Information Systems Engineering–WISE 2013: 14th International Conference, Nanjing, China, October 13-15, 2013, Proceedings, Part II 14, Springer, 2013, pp. 88–103.

[4] C.-H. Tsai and P. Brusilovsky, “The effects of controllability and explainability in a social recommender system,” *User Modeling and User-Adapted Interaction*, vol. 31, pp. 591–627, 2021.

[5] A. Papadimitriou, P. Symeonidis, and Y. Manolopoulos, “A generalized taxonomy of explanations styles for traditional and social recommender systems,” *Data Mining and Knowledge Discovery*, vol. 24, pp. 555–583, 2012.

[6] Y. Zhang and X. Chen, “Explainable recommendation: A survey and new perspectives,” *Foundations and Trends® in Information Retrieval*, vol. 14, no. 1, pp. 1–101, 2020.

[7] M. T. Ribeiro, S. Singh, and C. Guestrin, “Model-agnostic interpretability of machine learning,” *arXiv preprint arXiv:1606.05386*, 2016.

[8] D. Shmaryahu, G. Shani, and B. Shapira, “Post-hoc Explanations for Complex Model Recommendations using Simple Methods.,” presented at the IntRS@ RecSys, 2020, pp. 26–36.

[9] Q. Chen, J. Lin, Y. Zhang, H. Yang, J. Zhou, and J. Tang, “Towards knowledge-based personalized product description generation in e-commerce,” presented at the Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, pp. 3040–3050.

[10] A. Nadamoto, K. Fukumoto, R. Takeuchi, H. Terada, and M. Bato, “Automatic generation of product description using deep learning methods,” *Journal of Data Intelligence*, vol. 4, no. 1 & 2, pp. 134–148, 2023.

[11] Q. Zhang, B. Guo, S. Liu, J. Liu, and Z. Yu, “CrowdDesigner: information-rich and personalized product description generation,” *Frontiers of Computer Science*, vol. 16, no. 6, p. 166339, 2022.

[12] J. Wang, Y. Hou, J. Liu, Y. Cao, and C.-Y. Lin, “A statistical framework for product description generation,” presented at the Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers), 2017, pp. 187–192.

[13] S. Gerani, Y. Mehdad, G. Carenini, R. Ng, and B. Nejat, “Abstractive summarization of product reviews using discourse structure,” presented at the Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 1602–1613.

[14] K. Balog and F. Radlinski, “Measuring recommendation explanation quality: The conflicting goals of explanations,” presented at the Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval, 2020, pp. 329–338.

[15] I. Nunes and D. Jannach, “A systematic review and taxonomy of explanations in decision support and recommender systems,” *User Modeling and User-Adapted Interaction*, vol. 27, pp. 393–444, 2017.

[16] N. Tintarev and J. Masthoff, “A survey of explanations in recommender systems,” presented at the 2007 IEEE 23rd international conference on data engineering workshop, IEEE, 2007, pp. 801–810.

[17] M. Caro-Martínez, G. Jiménez-Díaz, and J. A. Recio-García, “Conceptual modeling of explainable recommender systems: an ontological formalization to guide their design and development,” *Journal of Artificial Intelligence Research*, vol. 71, pp. 557–589, 2021.

[18] C. Nóbrega and L. Marinho, “Towards explaining recommendations through local surrogate models,” presented at the Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing, 2019, pp. 1671–1678.

[19] B. C. G. Lee, K. Lo, D. Downey, and D. S. Weld, “Explanation-based tuning of opaque machine learners with application to paper recommendation,” *arXiv preprint arXiv:2003.04315*, 2020.

[20] G. Peake and J. Wang, “Explanation mining: Post hoc interpretability of latent factor models for recommendation systems,” presented at the Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 2060–2069.

[21] C. Lonjarret, C. Robardet, M. Plantevit, R. Auburtin, and M. Atzmueller, “Why should i trust this item? explaining the recommendations of any model,” presented at the 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), IEEE, 2020, pp. 526–535.

[22] M. Du, N. Liu, and X. Hu, “Techniques for interpretable machine learning,” *Communications of the ACM*, vol. 63, no. 1, pp. 68–77, 2019.

[23] Y. Zhang, G. Lai, M. Zhang, Y. Zhang, Y. Liu, and S. Ma, “Explicit factor models for explainable recommendation based on phrase-level sentiment analysis,” presented at the Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval, 2014, pp. 83–92.

[24] N. Wang, H. Wang, Y. Jia, and Y. Yin, “Explainable recommendation via multi-task learning in opinionated text data,” presented at the The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, 2018, pp. 165–174.

[25] Q. Ai, V. Azizi, X. Chen, and Y. Zhang, “Learning heterogeneous knowledge base embeddings for explainable recommendation,” *Algorithms*, vol. 11, no. 9, p. 137, 2018.

[26] H. Wang *et al.*, “Ripplenet: Propagating user preferences on the knowledge graph for recommender systems,” presented at the Proceedings of the 27th ACM international conference on information and knowledge management, 2018, pp. 417–426.

[27] Y. Xian, Z. Fu, S. Muthukrishnan, G. De Melo, and Y. Zhang, “Reinforcement knowledge graph reasoning for explainable recommendation,” presented at the Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval, 2019, pp. 285–294.

[28] F. Fusco, M. Vlachos, V. Vasileiadis, K. Wardatzky, and J. Schneider, “RecoNet: An Interpretable Neural Architecture for Recommender Systems.,” presented at the IJCAI, 2019, pp. 2343–2349.

[29] Y. Lu, R. Dong, and B. Smyth, “Why I like it: multi-task learning for recommendation and explanation,” presented at the Proceedings of the 12th ACM Conference on Recommender Systems, 2018, pp. 4–12.

[30] Z. Chen *et al.*, “Co-attentive multi-task learning for explainable recommendation.,” presented at the IJCAI, 2019, pp. 2137–2143.

[31] X. Wang, X. He, F. Feng, L. Nie, and T.-S. Chua, “Tem: Tree-enhanced embedding model for explainable recommendation,” presented at the Proceedings of the 2018 world wide web conference, 2018, pp. 1543–1552.

[32] W. Ma *et al.*, “Jointly learning explainable rules for recommendation with knowledge graph,” presented at the The world wide web conference, 2019, pp. 1210–1221.

[33] W. Sherchan, S. Nepal, and C. Paris, “A survey of trust in social networks,” *ACM Computing Surveys (CSUR)*, vol. 45, no. 4, pp. 1–33, 2013.

[34] A. Sharma and D. Cosley, “Do social explanations work? Studying and modeling the effects of social explanations in recommender systems,” presented at the Proceedings of the 22nd international conference on World Wide Web, 2013, pp. 1133–1144.

[35] B. Wang, M. Ester, J. Bu, and D. Cai, “Who also likes it? generating the most persuasive social explanations in recommender systems,” presented

at the Proceedings of the AAAI Conference on Artificial Intelligence, 2014.

[36] C. Shi, Z. Zhang, P. Luo, P. S. Yu, Y. Yue, and B. Wu, "Semantic path based personalized recommendation on weighted heterogeneous information networks," presented at the Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, 2015, pp. 453–462.

[37] Y. Zhang, X. Xu, H. Zhou, and Y. Zhang, "Distilling structured knowledge into embeddings for explainable and accurate recommendation," presented at the Proceedings of the 13th international conference on web search and data mining, 2020, pp. 735–743.

[38] Z. Ren, S. Liang, P. Li, S. Wang, and M. de Rijke, "Social collaborative viewpoint regression with explainable recommendations," presented at the Proceedings of the tenth ACM international conference on web search and data mining, 2017, pp. 485–494.

[39] J. Zheng, Q. Li, J. Liao, and S. Wang, "Explainable link prediction based on multi-granularity relation-embedded representation," *Knowledge-Based Systems*, vol. 230, p. 107402, 2021.

[40] J. Zheng, Z. Qin, S. Wang, and D. Li, "Attention-based explainable friend link prediction with heterogeneous context information," *Information Sciences*, vol. 597, pp. 211–229, 2022.

[41] D. Rudrapal, A. Das, and B. Bhattacharya, "Ranking of event-focused english tweets based on relevance and informativeness," *Computación y Sistemas*, vol. 23, no. 2, pp. 491–500, 2019.

[42] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi, "Bertscore: Evaluating text generation with bert," *arXiv preprint arXiv:1904.09675*, 2019.

[43] K. Chen, T. Chen, G. Zheng, O. Jin, E. Yao, and Y. Yu, "Collaborative personalized tweet recommendation," presented at the Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval, 2012, pp. 661–670.

[44] C. De Maio, G. Fenza, M. Gallo, V. Loia, and M. Parente, "Time-aware adaptive tweets ranking through deep learning," *Future Generation Computer Systems*, vol. 93, pp. 924–932, 2019.

[45] Y. Duan, L. Jiang, T. Qin, M. Zhou, and H. Y. Shum, "An empirical study on learning to rank of tweets," presented at the Proceedings of the 23rd international conference on computational linguistics (Coling 2010), 2010, pp. 295–303.

[46] K. Sailunaz, J. Kawash, and R. Alhajj, "Tweet and user validation with supervised feature ranking and rumor classification," *Multimedia Tools and Applications*, vol. 81, no. 22, pp. 31907–31927, 2022.

[47] D. Kim, Y. Jo, I.-C. Moon, and A. Oh, "Analysis of twitter lists as a potential source for discovering latent characteristics of users," presented at the ACM CHI workshop on microblogging, Citeseer, 2010.

[48] D. Metzler and W. Bruce Croft, "Linear feature-based models for information retrieval," *Information Retrieval*, vol. 10, pp. 257–274, 2007.

[49] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

[50] Q. Wu, C. J. Burges, K. M. Svore, and J. Gao, "Adapting boosting for information retrieval measures," *Information Retrieval*, vol. 13, pp. 254–270, 2010.

[51] C. Chen, M. Zhang, Y. Liu, and S. Ma, "Neural attentional rating regression with review-level explanations," presented at the Proceedings of the 2018 world wide web conference, 2018, pp. 1583–1592.