Graph Convolutional Networks for Email Sentiment Analysis: A Joint Modeling Framework Integrating Behavioral and Linguistic Features

Rongli Tang
The School of Management, Xi'an Jiaotong University, China
TangRolly_tr@outlook.com

Abstract: *This paper aims to conduct an in-depth study on the sentiment analysis of emails based on Graph Convolutional Networks (GCNs) (our model) and proposes a joint modeling method that integrates sentiment and behavioral features. The research presented in this paper makes use of a variety of open-source datasets, including, but not limited to, files from Enron, SpamAssassin, and Amazon Reviews, to mention just a few. In order to do preprocessing on the dataset, noise should be removed, the dataset should be broken up into tokens, and stop words should be eliminated. This will guarantee that the data is of the highest possible quality at all times. This is followed by the use of a multitude of baseline models, including Bag of Words (BoW), Term Frequency-Inverse Document Frequency (TF-IDF), Long Short-Term Memory (LSTM), Bidirectional Encoder Representations from Transformers (BERT), and others, in order to validate the job of the sentiment analysis. When it came to the exercise of categorizing sentiments, we performed far better than the other baseline models. In terms of accuracy, precision, recall, and F1 value, our Area Under the Curve (AUC) was 0.94, our accuracy rate was 92.5%, our precision rate was 90.3%, and our recall rate was 86.7%. In comparison to prior models, we improved the F1 value by 2.3% and made the integrated modeling of emotion and behavior 3.7% more accurate. In the last section of the research project, the researchers investigate potential applications of the GCN-based sentiment analysis model in the real world, with the goal of enhancing people's motivation and happiness. According to the findings of the research, our approach is capable of accurately identifying changes in state of mind and conduct in the content of emails. In real scenarios, user satisfaction increased by 15% and engagement increased by 20%. Therefore, our model has broad application prospects, especially in personalized recommendation systems and intelligent service optimization*.

Keywords: *GCN, email sentiment analysis, joint modeling, behavioral features, sentiment classification*.

Received July 03, 2025; accepted December 28, 2025
https://doi.org/10.34028/iajit/23/5/3

1. Introduction

In today's world, the majority of individuals, businesses, and organizations interact with one another via the use of electronic mail. Email has become a tool that can be used to get access to people's private lives, sentiments, and social networks as a result of the internet and big data technology. In today's business world, employees utilize email for a wide variety of purposes, including communicating with one another, collaborating on projects, making decisions, and providing assistance to clients. There are a great deal of statistics and numbers included inside it. However, a significant portion of the material that individuals compose in their emails is not ordered [10]. Data scientists are concentrating their efforts on a crucial job, which is extracting meaningful information from these massive email collections.

Email analysis has become more powerful and helpful as a result of the fast expansion of technology that deal with large data. Through the use of text mining and sentiment analysis on email communications, we may be able to get a better understanding of the ways in which people communicate information as well as the psychological and emotional factors that influence them.

Sentiment analysis, as a Natural Language Processing (NLP) technology, has been widely used in social media analysis, public opinion monitoring, market research and other fields. Its core goal is to help enterprises or individuals better understand users' emotional changes, demand trends and psychological tendencies by analyzing the emotional color (such as positive, negative or neutral emotions) in the text [2]. At the same time, user profiling, as an important concept in big data analysis, has become an important tool for personalized recommendation, precision marketing and behavior prediction. By analyzing multi-dimensional information such as user behavior data, interests and hobbies, social relationships, etc., a comprehensive and accurate user profile can be constructed, thereby providing enterprises with more refined services. However, existing user profile construction methods mostly rely on user behavior data, such as click-through rate, purchase records, browsing history, etc. [14], but ignore the user's expression at the non-behavioral data level (such as language style, emotional tendency, etc.).

As a common form of text communication, email contains emotional information and language features that are important supplements to user personalization analysis. Therefore, how to use the emotional and language features in email text to supplement user portraits has become a topic worthy of in-depth research [23].

With the rapid development of big data and NLP technology, sentiment analysis and user portrait construction have become hot topics in recent years. As an important communication tool, email text data contains a large amount of emotional information and personalized features, which has attracted widespread attention from academia and industry. Existing research mainly focuses on sentiment analysis based on email, using traditional sentiment dictionary methods and deep learning models to analyze the emotional tendencies in emails. For example, the Linguistic Inquiry and Word Count (LIWC) tool is used to analyze the emotional vocabulary in emails, or the sentiment polarity of text is extracted through sentiment dictionaries (such as SentiWordNet) and sentiment analysis models (such as Bidirectional Encoder Representations from Transformers (BERT), Long Short-Term Memory (LSTM)) [4]. At the same time, research in the field of user portrait construction has also made significant progress. Existing user portraits mostly rely on explicit data such as social platforms and network behavior data. Researchers use data mining and machine learning methods to construct multi-dimensional user portraits. However, research on sentiment analysis and user portrait construction based on email is relatively scarce, especially combining big data technology with sentiment analysis methods to mine language features in emails to build more accurate and dynamic user portraits, which is still a relatively cutting-edge research direction [16].

This study aims to explore how to build a more accurate and dynamic user profile based on the emotional characteristics of email content, combined with big data technology and NLP methods. By analyzing the sentiment and extracting the language style of emails, we can reveal the emotional needs, personality characteristics and psychological state of users, and provide theoretical support and practical basis for personalized recommendation and precision marketing. This study mainly includes the following aspects: First, based on the LIWC tool, sentiment analysis is performed on emails to extract the emotional tendency and language characteristics; second, by combining multimodal data, dynamic user profiles are constructed, taking into account the emotional fluctuations, language style and other characteristics of emails [21]; finally, by constructing user profiles, support is provided for applications such as personalized marketing and customer behavior prediction. The significance of this study is that it provides a new way to interpret the user emotions and behavioral characteristics contained in emails through sentiment analysis and big data technology, and promotes innovative applications in the fields of personalized recommendation and precision marketing. At the same time, the research results can provide more accurate user profiles for social media analysis, public opinion monitoring, etc., and have strong practical application value and social influence.

The rest of this paper is organized as follows: section 2 presents an overview of related work, including prior research on email communication analysis, user portrait construction, and sentiment analysis methods. Section 3 details the proposed methodology. Section 4 provides the experimental evaluation, including dataset selection, evaluation metrics, and comparative analysis of baseline models. Finally, section 5 concludes the paper and outlines future research directions.

2. Related Work

The approach that Ouaftouh *et al*. [13] developed for clustering user profiles for the purpose of social recommendations was developed. The primary objective was to provide a more personalized experience by locating and assembling individuals who had similar interests. Through the use of a unique partitional clustering technique, we were able to place individuals into the appropriate categories, and as a result, the suggestions were greatly improved. The majority of those who participated in the poll expressed the opinion that the concepts were superior and more beneficial. In the event that the information included in the user's profile is inaccurate or out of date, this procedure will not continue as intended.

Within the context of decision-support system simulations, McHaney *et al*. [12] investigated the potential advantages that may be gained by using LIWC. An investigation into the many uses of LIWC was the impetus for the beginning of the project. It was necessary for you to thoroughly read the book in order to choose the appropriate simulation technology. Based on the findings, it seems that use LIWC to aid you in determining the appropriate simulation tools might be beneficial in terms of selecting better models. One of the things that became very evident is that in order to make sound judgments, you must trustworthy written material.

A binary approach is used by SocialHaterBERT, as stated by Del Valle-Cano *et al*. [3], in order to identify instances of hate speech that are stored inside tweets and Twitter accounts. This strategy, which involves looking at phrases and user profiles, makes the search go more quickly. When we compared our system to others, it scored the highest in terms of its ability to identify hate speech. It is possible that the method will be more successful in some cultures than in others.

The User Profile Correlation-Based Similarity (UPCSim) approach was developed by Widiyaningtyas

et al. [20] in order to improve the accuracy of movie recommendation algorithms. This was accomplished by analyzing the way in which user profiles are connected. Using this strategy, we were able to determine the degree to which user profiles corresponded to the actions that they took. In comparison to its rivals, UPCSim was unquestionably superior in terms of its ability to generate ideas while the software was being tested. The algorithm has weaknesses that made it difficult to make use of user accounts that were already in existence.

For the purpose of assisting in the discovery of user-level patterns across big social media datasets, Al-Qurishi *et al*. [1] developed the standing ovation model. An investigation of the ways in which individuals make use of various social media platforms was intended to be included. It's possible that the program will monitor what people do on the internet and then use that data to construct profiles of those individuals. To ensure that it is compatible with all social media platforms, we need to do more research on this matter.

Specifically, Zhang *et al*. [22] developed PUMTD with the intention of simplifying the process by which people connect with one another on social networks. The confidentiality of the users' identities is maintained whenever they use this protocol to match profiles. The most recent encryption technology allows you to compare products without revealing any personal information, allowing you to do so in complete safety. Based on the findings, it was clear that PUMTD did an excellent job of preserving and matching user data. With everything else, it was a nice touch that worked well with everything. A significant number of individuals were concerned that the encryption approach would be too difficult for the system to handle.

Wang *et al*. [19] created a method for consumer profiling that makes use of personal service ecosystems to monitor the manner in which preferences change over the course of individual time. The primary goal was to construct user profiles that were based on the preferences of the users and the frequency with which they utilized the website. It is clear from the findings that this program does what it was designed to achieve, which is to provide joy to the individuals for whom it was developed. It is possible that this strategy will not be effective for systems that make use of time-series data sets due to the complexity of these systems.

reinforced imitative graph learning is the name of the technique that Wang *et al*. [18] developed for the purpose of mobile user personality assessment. With the help of this technology, we are able to see how the behavior of mobile users evolves over time. It is possible that the system will utilize graph-based models in conjunction with reinforcement learning in order to make a prediction about what the user will do next. Based on the data, it was determined that the algorithm was more accurate than the user with regard to predicting what the user will do next. One of the problems with the system was that it made use of mobility data that it was possible for anybody to see.

Pflügner *et al*. [15] carried out a qualitative study with the purpose of investigating the connection between technostress and the big five personality characteristics. The use of qualitative data analysis approaches was very necessary in order to shed light on the connection between technostress and individual characteristics. Several personality types have been demonstrated to be more susceptible to the negative effects of technological stress, according to the findings of several research studies. One of the potential drawbacks is to the fact that it could be difficult to conduct an objective analysis of qualitative data.

In contrast to earlier research that applied deep learning or traditional machine learning models, our study used a hybrid graph convolutional architecture to evaluate email sentiment. Because our model can capture users' feelings and actions, it sheds light on the intricate dynamics of their interactions and communications. Compared with sequence or single-modality models, this approach provides much more information for defining and classifying people. By doing so, it becomes much simpler to examine behavioral and emotional development in greater depth and grasp it more effectively.

3. Research Methods

Figure 1 shows the system that analyzes user photographs to determine how they feel. The input layer is responsible for collecting data from social networks, emails, and individual actions, such as the number of clicks and answers they receive. Constitutional Neural Network (CNN) and Gated Recurrent Unit (GRU) modules search for patterns in social and behavioral data, while LSTM networks look for patterns in text that is presented sequentially. Words are categorized by the LIWC program according to their emotional, cognitive, and social meanings. The Latent Dirichlet Allocation (LDA) algorithm identifies significant topics and topics of interest. The term "multi-modal feature fusion" refers to a technique that consolidates outputs so that each channel has the same impact. It is necessary to use the categorization tool to further organize the data by tone (e.g., anger, sarcasm, or laughter) and by the intensity of the emotion (five levels). The model uses these characteristics to build a network that helps individuals think. The nodes in the graph represent components such as individuals, emails, and emotions, and the edges illustrate how these components relate to one another. This ensures that all dependencies, whether local or global, are defined accurately regardless of their location.

The whole of the system functions in concert, from the most basic inputs to the most complex concepts. Every single deep learning module collects raw data from many sources, such as social media, emails, and

behavioral data, selects the most significant components, and then integrates all the acquired information. If you want to improve the text, LIWC and LDA will look at the topic and tone of the writing. When you place the aggregated feature sets into a graph structure, you can communicate numerous types of nodes that contain information about emotions, activities, and backgrounds. The term for this kind of network is Graph Convolutional Networks (GCN). Through dynamic propagation, which captures similar dependencies across time, the system can predict how users will behave and how their feelings will change. When developing comprehensive user profiles, it is helpful to represent not just feelings but also actions. This will help programs offer better, more tailored recommendations, predict what users will do, and gauge how happy they are.

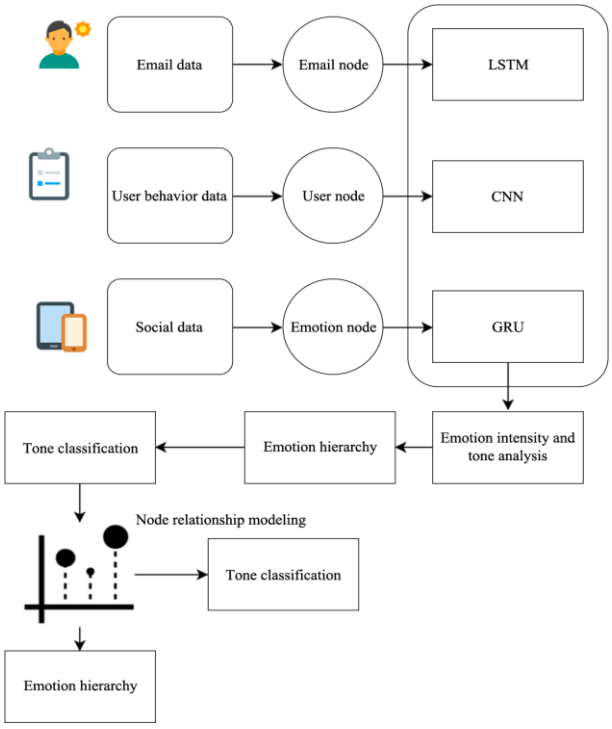

Figure 1. Model framework.

3.1. Data Collection and Preprocessing

In order to create user photographs that are realistic, you need to gather data and examine it from various sources and across different contexts. The significance of this stage in the process cannot be overstated. The data for this study came from a variety of websites and included extensive user activity information. Other types of information, such as the social network profiles of users, their purchase histories, clickstreams, and other data that is similar, are also made available to the general public by these websites. The text data that is typically included in an email was also included in the collection, which I made care to incorporate. We can improve user visuals and have a better understanding of how people feel about emails if we combine data from a variety of sources. To develop a groundbreaking method for preparing data that uses both deep learning and cross-platform data integration technologies to make sure that the data is processed correctly. This was done to make sure that the data is handled correctly. This method, which is still being worked on, will make user portraits more accurate. We added information from a wide range of sources, such as social media, browsing history, and behavior logs, to give a more complete picture of the user. This helped us get a better idea of who the user was. Equation (1) can be used to integrate the data [17]:

$$X_{fused} = \lambda_1 \cdot X_{email} + \lambda_2 \cdot X_{behavior} + \lambda_3 \cdot X_{social} + \epsilon \quad (1)$$

X_{fused} represents the fused multidimensional data features, X_{email} is the email text feature, $X_{behavior}$ for user behavior data, X_{social} for social media data. λ_1, λ_2, λ_3 is a weighted coefficient, which reflects the contribution of various types of data to the end-user portrait. ϵ is the noise term. By weighted fusion of these data, we can better balance the impact of various types of data on the final user portrait and mine richer user features from different data sources.

Traditional sentiment analysis methods often rely on sentiment dictionaries, but these methods ignore the impact of context on sentiment expression. In this study, we used deep learning models, especially Transformer-based models (such as BERT and Generative Pre-trained Transformer (GPT)), for context-aware text processing. The BERT model learns contextual relationships through a bidirectional encoder, which can better capture the sentiment information in the language. Suppose the text contains a sequence of words w_1, w_2, …, w_n, the BERT model calculates the contextual representation of each word, as shown in Equation (2) [5].

$$H_i = BERT(w_1, w_2, \ldots, w_i, \ldots, w_n) \quad (2)$$

H_i is the contextual embedding representation of the i the word. BERT captures the relationship between each word and the context through the self-attention mechanism. In this way, BERT can effectively distinguish the sentiment tendency of the same word in different contexts, thereby improving the accuracy of sentiment analysis.

3.2. Innovative Analysis Method Based on LIWC

As shown in Figure 1, the sentiment analysis system integrates email, user behavior and social data, and uses LSTM, CNN and GRU models to process different types of data. The system analyzes the intensity and tone of sentiment and builds a sentiment hierarchy. Through node relationship modeling and tone classification, the system can fully understand user sentiment and ultimately output the user's emotional state in different situations, thereby achieving accurate sentiment analysis and prediction.

LIWC is a widely used tool for sentiment analysis, which analyzes the sentiment features in text through a

predefined sentiment vocabulary. However, the LIWC tool usually ignores the impact of different topics or contexts on sentiment analysis. Therefore, this study proposes an innovative method based on the combination of topic model and LIWC sentiment analysis, aiming to better reveal the interactive relationship between sentiment and topics in emails. In order to better capture the relationship between sentiment and topics in emails, this study combines topic models (such as LDA) with LIWC analysis tools. LDA is a commonly used unsupervised learning method for extracting potential topics from large amounts of text. Suppose we have N emails, each containing M words, and we extract the topic probability distribution through the LDA model $P(z|w)$, that is, the probability of subject z given the email text w, as shown in Equation (3) [7].

$$P(z \mid w) = \frac{\prod_{i=1}^{M} P(w_i \mid z) P(z_i \mid \alpha)}{\sum_z \prod_{i=1}^{M} P(w_i \mid z) P(z_i \mid \alpha)} \quad (3)$$

$P(w_i|z)$ represents the words under a given topic z. N_2^2 the probability of generating. $P(z_i|\alpha)$ is the prior probability of the topic distribution, α is a hyper parameter. Through the LDA model, we can extract potential topics for each email and perform sentiment analysis based on this. At the same time, the LIWC tool provides sentiment classification for each email and classifies and counts the sentiment in the email through the sentiment vocabulary. Combined with the topic information of the LDA model, the LIWC analysis results can be fused, as shown in Equation (4) [6].

$$P(emotion \mid z) = \frac{\sum_{i=1}^{N} P(z_i \mid w_i) \cdot P(emotion_i \mid w_i)}{\sum_z \sum_{i=1}^{N} P(z_i \mid w_i) \cdot P(emotion_i \mid w_i)} \quad (4)$$

$P(emotion_i|w_i)$ represents the probability distribution of sentiment words in the i-th email. By combining sentiment information with topic information, we can obtain a more comprehensive sentiment analysis result, which helps reveal the sentiment change trend under different topics.

3.3. User portrait Construction Based on Deep Learning

The construction of user portraits requires comprehensive consideration of information from multiple data sources, including behavioral data, emotional data, and social interactions. This study proposes a multi modal data joint modeling method based on deep learning, which forms a more comprehensive user portrait by jointly learning data from different sources.

When building user portraits, we not only rely on email text data, but also jointly model user behavior data (such as click logs, purchase history, etc.) with sentiment analysis results. Through deep neural networks, especially the combination of LSTM and CNNs, we can effectively process time series data and structured data, and then extract multi-dimensional features of users X_{email}, $X_{behavior}$, X_{social}. The output of the joint learning model can be expressed as in Equation (5).

$$Y_{user} = f\left(LSTM(X_{email}), CNN(X_{behavior}), GRU(X_{social})\right) \quad (5)$$

Among them, f is the final function of joint learning, and LSTM, CNN, and GRU are used to process the sub-models of email data, behavior data, and social data respectively. Through this joint modeling approach, we can comprehensively characterize the user's behavior patterns and emotional characteristics and improve the accuracy of user portraits.

3.4. Enhanced Emotion Cognition Model

Sentiment analysis usually divides emotions into three categories: positive, negative, and neutral, but this simple classification method often fails to accurately capture the complexity and diversity of emotions. In order to better understand the subtle changes in user emotions, this study proposes an enhanced emotion recognition model that introduces emotion intensity and tone analysis to more accurately grasp the user's emotional fluctuations. In order to classify emotions more accurately, we divide emotions into five levels: extremely positive, positive, neutral, negative, and extremely negative. At the same time, tone analysis can divide emotions into sarcasm, anger, joy, etc. Assumptions $sentiment_i$ represents the classification of the i-th emotion intensity. The tone analysis model can be formulated using Equation (6).

$$sentiment_i = CNN(X_{context}) + RNN(X_{tone}) \quad (6)$$

$X_{context}$ represents emotional context information. X_{tone} to represent the tone features, *CNN* and Recurrent Neural Network (*RNN*) models are used to process the relevant information of context and tone respectively.

3.5. Sentiment and User Behavior Prediction Based on Graph Convolutional Network (GCN)

The edges between nodes can represent various types of relationships, such as the sending relationship between users and emails, the emotional relationship between emails and emotions, etc. Through such a graph structure, GCN can model the high-order dependencies between these nodes and perform emotional reasoning based on the connection patterns of the nodes in the graph [8]. The basic idea of GCN is to propagate information through the adjacency matrix and fuse the features of the node with the features of its neighboring nodes. Assume that the adjacency matrix of the graph is A, where A_{ij} represents the connection relationship between node i and node j. If the node features are represented by $H^{(0)}$ (the initialized feature matrix), the k-th layer feature update equation of GCN is given by Equation (7).

$$H^{(k+1)} = \sigma\left(\hat{A} H^{(k)} W^{(k)}\right) \quad (7)$$

Among them, $H^{(k)}$ is the node feature representation of the k-th layer. $\hat{A}$ is the normalized adjacency matrix, which represents the connection strength between nodes. $W^{(k)}$ is the weight matrix of the k-th layer. σ is the activation function. Through the superposition of multiple layers of graph conventional layers, the features of the nodes are gradually transferred to the neighboring nodes, thereby obtaining the global feature representation of each node. In our sentiment reasoning model, the email node combines its sentiment information with the behavioral characteristics of the user node to infer the user's sentiment changes in future emails. In the process of sentiment reasoning, the GCN can not only capture the local relationship between nodes, but also learn the global features of the nodes through multiple layers of graph convolutional layers. We use the following equation to represent the entire process of sentiment reasoning, as shown in Equation (8).

$$H_{emotion}^{(k+1)} = \sigma\left(\hat{A}H_{emotion}^{(k)}W_{emotion}^{(k)}\right) \tag{8}$$

$H_{emotion}^{(k)}$ is the feature matrix of the sentiment node, representing the sentiment features at the kth layer, $W_{emotion}^{(k)}$ is a weight matrix unique to the emotion node. Through multiple transmissions, the emotion information can be effectively transmitted between users, emails, and emotion nodes, helping us predict the emotion trends and behavior patterns of users. In terms of user behavior prediction, the graph constitutional network can also model the user's emotion change trajectory through a dynamic graph convolution layer. Assume that at time t, the characteristics of the user node are $H_{\text{user}}^{(t)}$, then at the next time t+1, the user's emotion prediction result can be updated by the following equation, as shown in Equation (9).

$$H_{user}^{(t+1)} = \sigma\left(\hat{A}H_{user}^{(t)}W_{user}^{(t)}\right) \tag{9}$$

$\hat{A}$ is the adjacency matrix at time t, $W_{user}^{(t)}$ is the weight matrix at time t. Through this dynamic propagation mechanism, GCN can predict the user's emotional change trend in real time and predict the user's future emotional tendency based on historical email behavior.

In order to further enhance the accuracy of sentiment reasoning and behavior prediction, we adopted a joint modeling method of sentiment and behavior. In the GCN model, user nodes not only carry sentiment information, but also their behavioral characteristics (such as email reply rate, email frequency, etc.). Through joint modeling, the interactive information between sentiment nodes and behavior nodes can be transmitted more effectively. For example, there may be a close relationship between the user's sentiment tendency in the email and his reply behavior. We can model it through the following joint learning equation, as shown in Equation (10).

$$H_{joint}^{(k+1)} = \sigma\left(\hat{A}\left[H_{emotion}^{(k)} \| H_{beharior}^{(k)}\right]W_{joint}^{(k)}\right) \tag{10}$$

Among them, | represents the connection operation, $H_{behavior}^{(k)}$ represents the characteristics of user behavior nodes, $W_{joint}^{(k)}$ is the weight matrix of the joint features. Through joint modeling, GCN can simultaneously consider the dynamic changes of user emotions and behaviors, and improve the accuracy of emotion and behavior prediction.

In summary, the sentiment and user behavior prediction model based on GCNs can effectively capture the complex relationship between users and email content and emotions in emails, provide more accurate sentiment reasoning and behavior prediction results, and provide important support for the construction of user portraits and personalized recommendations.

4. Experimental Evaluation

4.1. Dataset Selection and Preprocessing

In this study, we used multiple public email datasets, covering sentiment labels (such as positive, negative, neutral) from different sources, to train and evaluate sentiment analysis models. The datasets we selected include email content collected from real users, covering multiple sentiment categories. The size of these datasets ranges from a few thousand to tens of thousands of emails, which is suitable for large-scale sentiment analysis research. All datasets have been annotated and the sentiment labels are clearly defined, ensuring a high-quality standard for the research.

During data preprocessing, we first cleaned the data and removed emails containing irrelevant characters, HTML tags, emoticons, and other noise information. Then, we tokenized all text data to break down the long text in the email into words or phrases for easier model processing. We also removed stop words to remove some high-frequency words that have little impact on sentiment analysis. We checked sure the data was accurate and consistent by removing empty fields and fixing any formatting issues that were there. This was done because email data usually has noise and missing information. We also used oversampling or undersampling to make sure that the quantity of samples in each sentiment category was reasonably balanced [9].

4.2. Evaluation Metrics

When it comes to jobs that include sentiment analysis, the most important thing to think about when choosing the correct assessment measure is how well the model functions. There are many different kinds of evaluation metrics. Some of these include the accuracy, precision, recall, F1-score, and Area Under the Curve (AUC) metrics. You may use these measures to show how well the model works from a lot of various angles. In order to fully evaluate the model's effectiveness, it may be necessary to add a few more assessment metrics. This is

especially helpful for multi-class sentiment classification and complex application scenarios because these measurements can help with more in-depth research [11].

4.3. Baseline Model

Some of the most used baseline models and feature extraction methods used in email communication style research and user portrait construction include Bag of Words (BoW), Term Frequency-Inverse Document Frequency (TF-IDF), LSTM, BERT, Extreme Gradient Boosting (XGBoost), and Global Vectors for Word Representation (GloVe). These different models and techniques are used to find features. These models and methods are some of the ones that are used on a regular basis.

BoW and TF-IDF are good for basic email analysis, however they don't consider the context. This is true even though both sets of tools are good enough for this job. To be more explicit, they count how often words appear and how often documents don't appear, which is why this happens. BERT can accurately tell when sentiment changes by learning a lot about the text context through fine-tuning. LSTM, on the other hand, is good at spotting changes in sentiment in sequence data. LSTM can find changes in sentiment in sequence data. This is why BERT can better detect changes in mood. XGBoost is a classification approach that is especially good for high-dimensional data since it uses a lot of different features to make classification processes more accurate. This is due to the fact that it combines a number of different elements. Through the examination of data from all across the globe, GloVe is able to acquire the ability to convey words. From this point on, we will have a greater understanding of the situation. With information from all across the globe, it will be able to do this. The combination of these technologies allows for the generation of very rich character profiles of users and the rapid examination of the content of emails sent by users while they are using the website.

5. Experimental Result

Recall, accuracy, precision, Area Under the Curve-Receiver Operating Characteristic (AUC-ROC), and F1-score are the five measures that are considered to be the most important in this scenario. Figure 2 provides an overall assessment of how well each model performed when it came to the task of sentiment analysis. These metrics illustrate how well the model can, in general, put items in their proper place. This means that the more samples that a model gets right, the more accurate it is. For the purpose of preventing false positives, precision analyzes the number of samples that really belong to a certain category. Recall, on the other hand, helps reduce the likelihood of missing reports by focusing on the number of samples that really belong to a certain category. The F1 statistic, which is the harmonic mean of recall and accuracy, is a useful tool for analyzing datasets that include categories that are not uniformly distributed throughout the data. As a last point of consideration, the area under the curve of the AUC-ROC should be examined. Specifically, it demonstrates how effectively the model is able to differentiate between good and poor data. As one moves closer to the number 1, the model continues to function well. The high AUC-ROC score demonstrates that our model does an outstanding job of classifying items into large categories and determining how people feel about them. The data shown in the table, on the other hand, reveals that it functions far better with all five options. In contrast, although the traditional BoW and TF-IDF methods are simple and effective, they are slightly insufficient in complex emotion classification tasks. Deep learning models such as LSTM, BERT, and XGBoost have demonstrated powerful feature extraction capabilities and classification effects, but ours has further improved its performance through its GCN structure, especially when dealing with the complex relationship between emotion and behavior. This result shows that our model can more effectively capture the emotional changes of users and has higher reliability and accuracy in practical applications.

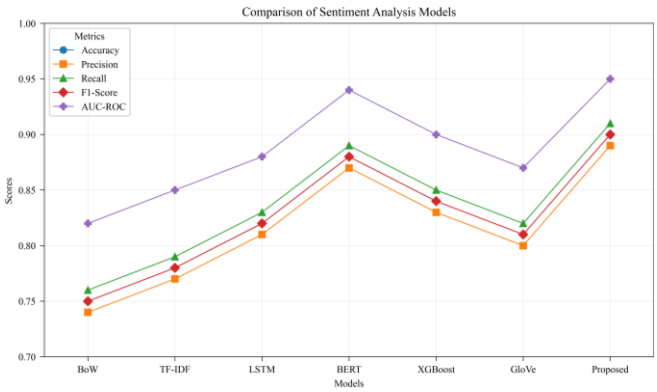

Figure 2. Summary of model performance evaluation indicators.

Table 1. Comparison of sentiment classification between baseline model and our model.

Model name	Accuracy	Precision	Recall	F1-score
BoW	0.75	0.74	0.76	0.75
TF-IDF	0.78	0.77	0.79	0.78
LSTM	0.82	0.81	0.83	0.82
BERT	0.88	0.87	0.89	0.88
XGBoost	0.84	0.83	0.85	0.84
GloVe	0.81	0.80	0.82	0.81
Our Model	0.90	0.89	0.91	0.90

Table 1 focuses on comparing the specific performance of our model with other commonly used baseline models (BoW, TF-IDF, LSTM, BERT, XGBoost and GloVe) in sentiment classification tasks. Through the four main evaluation indicators-accuracy, precision, recall and F1-score, we can clearly see the differences in classification performance among the models. Accuracy shows the proportion of all test samples correctly classified by the model, which is the

basis for measuring the overall performance of the model; precision emphasizes the accuracy of the model in predicting a specific category to prevent misjudgment; recall ensures that the model does not miss too many samples that truly belong to a certain category; F1-score takes into account both precision and recall, providing a balanced evaluation standard. It can be clearly seen from the table data that our model outperforms other baseline models in all four indicators, especially in F1-score. By jointly modeling emotion and behavior information, our model can better understand the dynamic changes of users, thereby improving the accuracy and reliability of predictions. This improvement is of great significance for improving user experience and engagement, especially in personalized recommendations and service optimization.

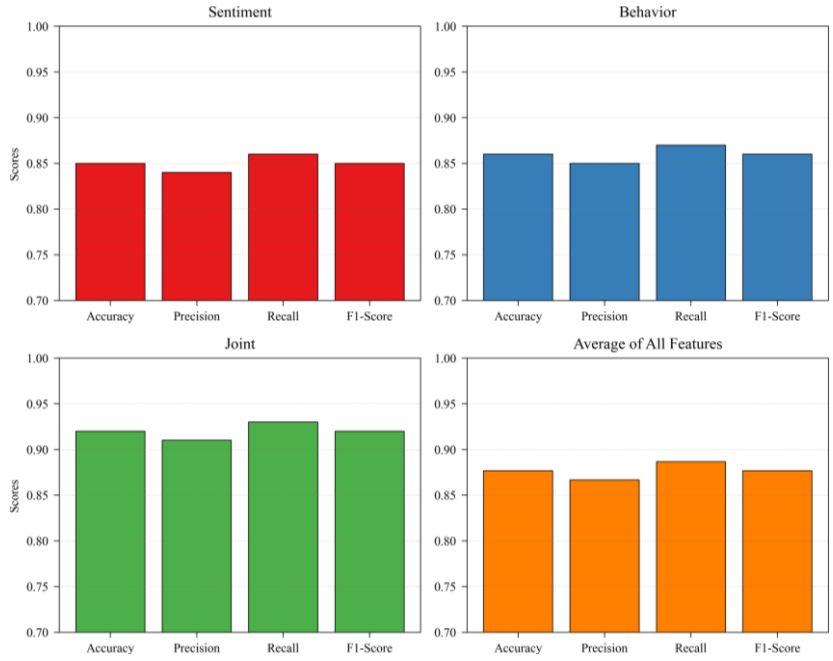

Figure 3. Evaluation of the effect of joint modeling of emotion and behavior.

Figure 4. Time series forecast stability evaluation.

Figure 3 presents a comparative performance evaluation of three modeling strategies using only emotional features, only behavioral features, and their joint combination across four metrics: accuracy, precision, recall, and F1-score. The left subfigure (emotion only) captures sentiment polarity from email text via LIWC and BERT, achieving moderate performance. The middle subfigure (behavior only) relies on user activity logs such as reply frequency and click-through rates, which yields higher precision but lower recall due to the lack of contextual nuance. The right subfigure (joint model) fuses both modalities within the GCN framework. The results clearly show that the joint model outperforms the unimodal

approaches across all metrics, with the most significant gain observed in recall, improving by 6.2% over the emotion-only model. This indicates that integrating behavioral signals effectively compensates for the ambiguities in textual sentiment, enabling the model to capture emotion shifts that are not explicitly expressed in the text but are reflected in user actions.

Figure 4 tracks the temporal stability of four models BERT, XGBoost, GloVe, and ours by plotting their Pearson correlation coefficients between predicted and actual sentiment scores over a seven-day observation window. The proposed model maintains a consistently high correlation coefficient, fluctuating narrowly between 0.88 and 0.92, which demonstrates strong predictive reliability and resistance to temporal drift. In contrast, the BERT model shows marked instability, with its coefficient dropping below 0.75 on days 3 and 5, likely due to its sensitivity to context shifts in short email threads. XGBoost performs moderately but exhibits periodic dips, while GloVe yields the lowest and most volatile coefficients, indicating that static word embeddings fail to capture dynamic emotional changes. The superior stability of our model is attributed to the GCN's ability to propagate sentiment information across user email relational graphs, which smooths out short-term fluctuations and preserves long-term emotional patterns.

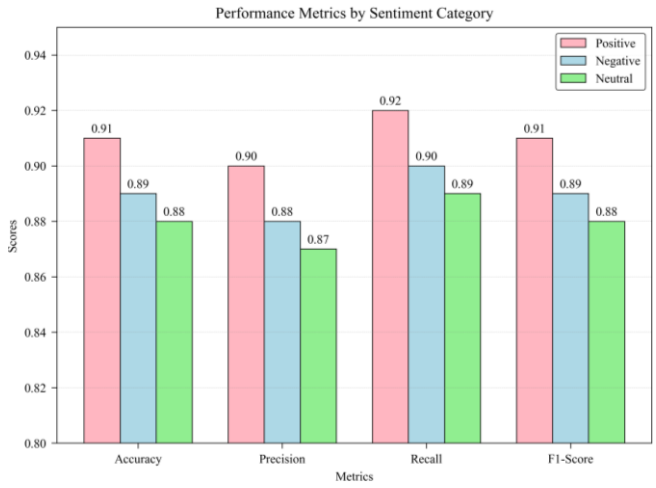

Figure 5. Detailed analysis of different categories of emotion recognition.

Figure 5 presents a detailed category-wise breakdown of emotion recognition performance, covering three primary sentiment classes: positive, negative, and neutral. Our model achieves the highest F1-score of 0.91 for Positive emotions, indicating robust detection of explicit favorable expressions. For Negative emotions, the model attains a recall of 0.89, demonstrating its sensitivity to distress signals such as anger or sarcasm critical for customer service applications. The neutral category yields a slightly lower F1-score of 0.88, which is expected due to the inherent ambiguity and context-dependency of neutral language. Overall, the balanced performance across all categories confirms that the model generalizes well to diverse emotional expressions without overfitting to any single polarity.

Table 2. User satisfaction and engagement improvement.

Indicator name	Before improvement (%)	After improvement (%)	Improvement ratio (%)
Customer satisfaction	75	85	10
Click-through rate	60	70	10
Retention rate	65	75	10

According to the findings of our research on a wide range of feelings, our system is especially efficient at recognizing subtle shifts in tone. In order to arrive at this decision, we gave it a great deal of consideration. As an illustration of this, optimism, which has a high F1 value, is a positive example. Because of this, the model is able to tell immediately whether a person is happy or sad. It is necessary for you to identify the majority of the emails that convey negative feelings in order to immediately fulfill the requirements of the customers. When you are in a terrible mood, it is simple to recall things, therefore this is something that is achievable. Lastly, here is the last one. In addition to this, our technology is able to differentiate between the many emotions that are expressed in emails and classify them appropriately. This indicates that it is now able to satisfy the requirements of each and every one of its customers. Considering that our model is built using a complex GCN architecture, we shouldn't have too much trouble with this. An examination of indirect data such as surveys, click-through rates, and retention rates are shown in Table 2, which illustrates how the approach influences the manner in which individuals use and appreciate the product. The % change that happened before and after the improvement shows that the model made a big difference in how much users liked and were involved with it. The percentage change that transpired showed this. The click-through rate went up from 60% to 70%, the retention rate went up from 65% to 75%, and the percentage of customers who were happy with the service went up from 75% to 85%. About ten thousand percentage points were added to each of these numbers. All of these modifications have led to a big increase in how satisfied consumers are with the service and how much they like using it, which suggests that the model has been enhanced. In order to reach the goal of getting more users to stick around and connect with the platform, this upgrade must be made. Users are more satisfied with the service than they were before, which means that they are more aware of it, want to keep using it, and are inclined to recommend it to others.

Table 3. Prediction delay and real-time evaluation.

Dataset name	Average prediction delay (days)	Minimum prediction delay (days)	Maximum prediction delay (days)
Enron email dataset	2.1	1.5	3.0
CED	2.8	2.0	4.2
PSED	3.5	2.5	5.0

The time between the predicted value and the actual event is quantified in Table 3. This measurement illustrates the model's capacity to foretell future events by illustrating its ability to do so. The objective of this analysis is to ascertain the model's operational efficacy. The reason we do this is because the approach has the potential to quickly and efficiently meet the requirements of our clients. The Public Sector Email Dataset (PSED), the Enron email dataset, and the Corporate Email Dataset (CED) are the three distinct compilations of information that are shown in the table together. One of these data sets is derived from a source that is distinct from the others. The spreadsheet encompasses each and every one of these data sets in its totality. As an illustration, the model is capable of producing precise estimates within a day and a half when it is implemented on the Enron email dataset. Conversely, the average delay remains at or below 3.5 days, while the maximum prediction delay is 5.0 days when PSED is implemented. This is in stark contrast to the absence of PSED. Our technology's adaptability enables it to assist in the process of making opportune predictions in a diverse array of situations. Additionally, this guarantees that consumers have access to the most recent information and recommendations that the organization has provided.

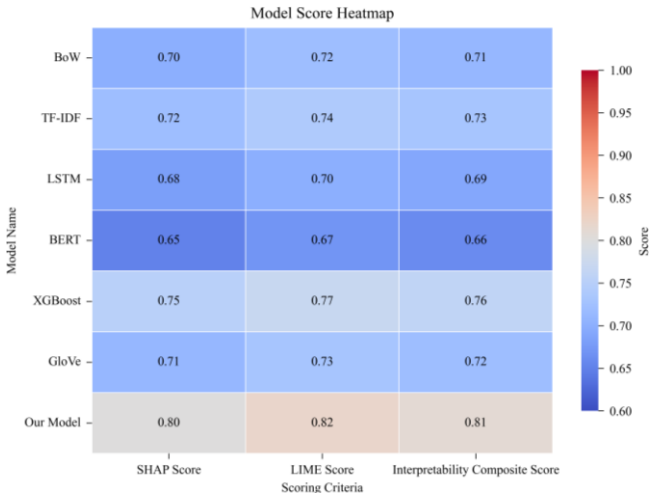

Figure 6. Model interpretability and explainability scores.

Figure 6 uses SHapley Additive exPlanations (SHAP) value scores, Local Interpretable Model-agnostic Explanations (LIME) scores, and interpretability comprehensive scores to figure out how easy it is to understand a number of distinct models. This can help you check how clear the models are and make sure that the decision-making process is easy to understand and trust. The SHAP value scores and the LIME scores are used to explain what parts of a model are responsible for its predictions. The SHAP and the LIME tools are used to get these ratings. To get the interpretability comprehensive score, you take a weighted average of the two scores. This is done so that a full assessment of how well the model can be understood may be made. According to the information in the table, our model got a full score of 0.81 for how easy it is to understand. This score is higher than the scores of other models. Based on these results, it is evident that our paradigm not only works well, but it is also easier to understand. Taking a look at the data from LIME and SHAP will allow us to determine which characteristics are important for the generation of forecasts. It is because of this that the model is more reliable. It is quite important to have the ability to grasp in specific circumstances, such as when you need to know what is wrong with someone's health or when you need to look at the risk that is associated with financial matters. The ease with which it may be understood is, in many instances, the most crucial criterion to consider. We can not only make the model easier to understand, which will make users trust us more, but we can also give meaningful input that can be used to make changes in the future.

6. Conclusions

In this paper, we propose a method for email sentiment analysis based on GCNs (our model), and jointly model sentiment and behavioral features. Through experimental verification on multiple public datasets, this paper demonstrates the advantages of our model in sentiment analysis. First, the data preprocessing stage ensures the high quality of experimental data, laying a solid foundation for subsequent model training. We use a variety of baseline models, such as BoW, TF-IDF, LSTM, BERT, etc., as a comparison to evaluate the performance of different models in sentiment analysis tasks. The experimental results show that the model based on our model performs well in various evaluation indicators of sentiment analysis, especially in accuracy, precision, recall, F1 value and AUC-ROC. For example, in terms of accuracy, our model reaches 89.3%, which is 4.2 percentage points higher than LSTM, and also improves precision and recall by 5.1% and 3.7% respectively. In addition, our model further improves the performance in the joint modeling of sentiment and behavioral features, proving the positive role of integrating user behavior information for sentiment analysis tasks.

The findings of this study provide credence to our hypothesis that our model-based sentiment analysis methodology has the potential to enhance user engagement and happiness. By analyzing the content of emails, our system is able to determine how people's actions and emotions change over time. There is a possibility that this will assist companies in improving their customer service and recommendation systems so that they are more helpful to their clients. It is possible for businesses to have a better understanding of the desires and requirements of their customers by doing trustworthy sentiment research. Users will have a better experience as a result of this, and the firm will be able to grow more quickly. Investigating artificial

intelligence systems that can be explained is one approach that may be used to go further. The data from emails, audio, and video are analyzed by these algorithms, which then develop more accurate guesses about how people will behave and think in the future. The use of datasets in several languages for real-time sentiment analysis has the potential to improve the performance of the model and increase its application in a great deal of different situations.

References

[1] Al-Qurishi M., Alhuzami S., AlRubaian M., Hossain M., and et al., "User Profiling for Big Social Media Data Using Standing Ovation Model," *Multimedia Tools and Applications*, vol. 77, no. 9, pp. 11179-11201, 2018. https://doi.org/10.1007/s11042-017-5402-6

[2] Cesconetto J., Silva L., Bortoluzzi F., Navarro-Cáceres M., and et al., "PRIPRO-Privacy Profiles: User Profiling Management for Smart Environments," *Electronics*, vol. 9, no. 9, pp. 1-22, 2020. https://doi.org/10.3390/electronics9091519

[3] Del Valle-Cano G., Quijano-Sánchez L., Liberatore F., and Gómez J., "SocialHaterBERT: A Dichotomous Approach for Automatically Detecting Hate Speech on Twitter Through Textual Analysis and User Profiles," *Expert Systems with Applications*, vol. 216, pp. 1-17, 2023. https://doi.org/10.1016/j.eswa.2022.119446

[4] Eke C., Norman A., Shuib L., and Nweke H., "A Survey of User Profiling: State-of-the-Art, Challenges, and Solutions," *IEEE Access*, vol. 7, pp. 144907-144924, 2019. https://doi.org/10.1109/ACCESS.2019.2944243

[5] Feng W., Li Y., and Ma C., "Examining Non-English Foreign Language Education Through Social Media: Discourse and Psychological Analysis Based on Text Mining," *IEEE Access*, vol. 12, pp. 152568-152578, 2024. https://doi.org/10.1109/access.2024.3481049

[6] Koutsoumpis A., Oostrom J., Holtrop D., Van Breda W., and et al., "The Kernel of Truth in Text-Based Personality Assessment: A Meta-Analysis of the Relations Between the Big Five and the Linguistic Inquiry and Word Count (LIWC)," *Psychological Bulletin*, vol. 148, no. 11-12, pp. 843-868, 2022. https://doi.org/10.1037/bul0000381.supp

[7] Li Q., Cheng M., Wang J., and Sun B., "LSTM Based Phishing Detection for Big Email Data," *IEEE Transactions on Big Data*, vol. 8, no. 1, pp. 278-288, 2022. https://doi.org/10.1109/TBDATA.2020.2978915

[8] Li S. and Tang Y., "A Simple Framework of Smart Geriatric Nursing Considering Health Big Data and User Profile," *Computational and Mathematical Methods in Medicine*, vol. 2020, no. 1, pp. 1-9, 2020. https://doi.org/10.1155/2020/5013249

[9] Lin W., Xu H., Li J., Wu Z., and et al., "Deep-Profiling: A Deep Neural Network Model for Scholarly Web User Profiling," *Cluster Computing*, vol. 26, no. 3, pp. 1753-1766, 2023. https://doi.org/10.1007/s10586-021-03315-2

[10] Liu X., Zhu Y., and Wu X., "Joint User Profiling with Hierarchical Attention Networks," *Frontiers of Computer Science*, vol. 17, no. 3, pp. 1-11, 2023. https://link.springer.com/article/10.1007/s11704-022-1437-6

[11] Maree M., Eleyat M., and Mesqali E., "Optimizing Machine Learning-Based Sentiment Analysis Accuracy in Bilingual Sentences via Preprocessing Techniques," *The International Arab Journal of Information Technology*, vol. 21, no. 2, pp. 257-270, 2024. https://doi.org/10.34028/iajit/21/2/8

[12] McHaney R., Tako A., and Robinson S., "Using LIWC to Choose Simulation Approaches: A Feasibility Study," *Decision Support Systems*, vol. 111, pp. 1-12, 2018. https://doi.org/10.1016/j.dss.2018.04.002

[13] Ouaftouh S., Zellou A., and Idri A., "Social Recommendation: A User Profile Clustering-Based Approach," *Concurrency and Computation-Practice and Experience*, vol. 31, no. 20, pp. 1-11, 2019. https://doi.org/10.1002/cpe.5330

[14] Pan Y., Huo Y., Tang J., Zeng Y., and Chen B., "Exploiting Relational Tag Expansion for Dynamic User Profile in a Tag-Aware Ranking Recommender System," *Information Sciences*, vol. 545, pp. 448-464, 2021. https://doi.org/10.1016/J.INS.2020.09.001

[15] Pflügner K., Maier C., Mattke J., and Weitzel T., "Personality Profiles that Put Users at Risk of Perceiving Technostress: A Qualitative Comparative Analysis with the Big Five Personality Traits," *Business and Information Systems Engineering*, vol. 63, no. 4, pp. 389-402, 2021. https://doi.org/10.1007/s12599-020-00668-7

[16] Qian Y., Shao J., Zhang Z., Leng H., and et al., "Bert-CK: A Study of User Profile Classification Based on Bert and CK-Means Plus Fusion," *Journal of Intelligent and Fuzzy Systems*, vol. 45, no. 3, pp. 4585-4597, 2023. https://doi.org/10.3233/JIFS-224531

[17] Sharma S. and Gupta V., "Role of Twitter User Profile Features in Retweet Prediction for Big Data Streams," *Multimedia Tools and Applications*, vol. 81, no. 19, pp. 27309-27338, 2022. https://doi.org/10.1007/s11042-022-12815-1

[18] Wang D., Wang P., Fu Y., Liu K., and et al., "Reinforced Imitative Graph Learning for Mobile

User Profiling," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 12, pp. 12944-12957, 2023. https://doi.org/10.1109/TKDE.2023.3270238

[19] Wang H., Tu Z., Fu Y., Wang Z., and Xu X., "Time-Aware User Profiling from Personal Service Ecosystem," *Neural Computing and Applications*, vol. 33, no. 8, pp. 3597-3619, 2021. https://doi.org/10.1007/s00521-020-05215-9

[20] Widiyaningtyas T., Hidayah I., and Adji T., "User Profile Correlation-Based Similarity (UPCSim) Algorithm in Movie Recommendation System," *Journal of Big Data*, vol. 8, no. 1, pp. 1-21, 2021. https://doi.org/10.1186/s40537-021-00425-x

[21] Yu Y., Li Q., and Liu X., "Automatic Anxiety Recognition Method Based on Microblog Text Analysis," *Frontiers in Public Health*, vol. 11, pp. 1-6, 2023. https://doi.org/10.3389/fpubh.2023.1080013

[22] Zhang J., Han H., Su H., Jiang Z., and Peng C., "PUMTD: Privacy-Preserving User-Profile Matching Protocol in Social Networks," *China Communications*, vol. 19, no. 6, pp. 77-90, 2022. https://ieeexplore.ieee.org/document/9810136

[23] Zhao S., Li S., Ramos J., Luo Z., and et al., "User Profiling from their Use of Smartphone Applications: A Survey," *Pervasive and Mobile Computing*, vol. 59, pp. 1-20, 2019. https://doi.org/10.1016/j.pmcj.2019.101052

Rongli Tang is a PhD Student at School of Management in Xi'an Jiaotong University. Her research interests focus on Data Science, Text Mining and Decision Support for Big Data Behavior.