# Multidimensional Private Information Portrait in Social Network Users

Fangfang Shan[1]*, Mengyi Wang[2], Huifang Sun[3]
School of Computer Science, Zhongyuan University of Technology, Zhengzhou 450007, Henan, China[1, 2, 3]
Henan Key Laboratory of Cyberspace Situation Awareness, Zhengzhou 450001, Henan, China[1]

***Abstract*—In order to tackle the challenges of users' weak privacy awareness and frequent disclosure of private information in social network, this paper proposes a multidimensional privacy information portrait model of users in Chinese social networks. Because the TF-IDF (Term Frequency-Inverse Document Frequency, TF-IDF) algorithm does not consider the distribution of feature terms among and within classes, uses the TF-IDF algorithm based on the bag-of-words model to calculate the sensitivity of user privacy information. Considering the diversity of user privacy information, this paper proposes the PROLM (Positive reverse order lookaround matching ) algorithm, which is combined with the Flashtext+ (improved Flashtext) algorithm and SMA (string matching algorithm, SMA), the PROLM_FlashText+_SMA to extract user personal privacy information and location where the privacy information is located, and return the sensitivity. Using the BERT (Bidirectional Encoder Representation from Transformers, BERT)-Softmax privacy information classification model, the privacy information is classified into high, moderate and mild privacy information, and a multidimensional privacy information portrait of the user is constructed based on the privacy information and sensitivity. The experiments show that the accuracy of PROLM_FlashText+_SMA algorithm for privacy information extraction reaches 93.63%, and the overall F1 index of privacy information classification using the BERT-Softmax model reaches 0.9798 on the test set, better than baseline comparison model, has better privacy information classification effect.**

***Keywords*—Social network; personal privacy information; privacy information portrait; sensitivity; privacy protection; BERT**

## I. Introduction

In the current stage, social networking platforms like QQ, WeChat, Weibo, and Facebook have spread at an unprecedented speed, becoming indispensable parts of people's lives. According to the 50th Statistical Report on China's Internet Development, As of June 2022, 38% of internet users said they had encountered any online security problems in the past six months. In addition, the proportion of internet users who experienced personal information disclosure was the highest, at 22.1%. The emergence of social networks has changed the way people communicate and interact with each other [1], but people's use of social networks to share their lives inevitably brings the risk of privacy leakage. A particular piece of data posted by a user on a social networking platform may contain one privacy item of the user, but if many pieces of information contain privacy items of the user, these privacy items may be associated to expose a whole privacy chain of the user. For the published information, users cannot control its dissemination path, and once the information is accessed by unlawful elements, there is a risk of causing economic and property losses, and even threatening personal safety [2]. For example, in August 2022, an individual posted on a hacker forum claiming to auction the Shanghai Health Code database for $4000. The post stated that the database contained personal information of 48.5 million users of the Shanghai Health Code, including ID numbers, names, and phone numbers of individuals who have either resided in or visited Shanghai since the implementation of the health code system. The post also included the release of 47 sets of sample data. Such incidents have raised concerns about personal privacy and have led to widespread interest in the study of privacy protection technologies in social networks.

At present, many scholars both domestically and internationally are conducting research on privacy protection schemes for social networks. AL-Asbahi R [3] introduced the concept of structural anonymity as a means to reduce data anonymization. Lian *et al.* [4] achieved privacy protection for users by calculating sensitive attribute levels and using different anonymization methods for different levels, but this algorithm has high time expenditure compared to other algorithms. To meet the differences in privacy protection needs of different users, Yin X et al. [5] proposed a social network attributes graphs algorithm under personalized differential privacy, which is for the independent attribute information between users. Ning et al. [6] integrated noisy weights into the generated graph to solve the problem of edge weights and frequent structure privacy and realized the privacy protection of graph structure in the process of frequent mining. Huang et al. [7] proposed the PBCN method based on the joint clustering and randomization algorithm to resist node attacks and degree attacks in social networks and lost adjacency information. However, the error generated by the method includes noise error and graph reconstruction error.

However, these privacy protection studies have not yet visualized the personal privacy information that users expose on social networking platforms, making it difficult for users to intuitively understand the potential harm caused by such privacy information. Social networking platforms store a large amount of personal privacy information from users, and due to their massive user base and strong communication interactivity, users' privacy information is more prone to leakage. In order to identify potential privacy leakage risks and

This paper is supported by the National Natural Science Foundation of China (No. 62302540), Open Foundation of Henan Key Laboratory of Cyberspace Situation Awareness (No. HNTS2022020); Natural Science Foundation of Henan Province Youth Science Fund Project (No. 232300420422) and the Natural Science Foundation of Zhongyuan University of Technology (No. K2023QN018).

attack threats, enhance users' privacy awareness and behavior, and strengthen personal privacy protection, reduce the probability and impact of privacy breaches, the author extracts users' exposed privacy information from their historical data posted on social networking platforms and uses it to construct a multidimensional privacy information portrait of users.

The main contributions of this paper are as follows:

- Use the BOW-TF-IDF algorithm to measure the sensitivity of privacy information. The BOW-TF-IDF algorithm can better consider the distribution of privacy information within and between classes, more accurately reflect the differences between different categories, and has certain advantages in measuring privacy information sensitivity.
- In response to the limitation of the FlashText algorithm that can only extract English text, this paper improves it to make it suitable for Chinese scenarios. Propose PROLM_FlashText+_SMA algorithm extracts user privacy information. Experiments have shown that this algorithm has advantages in terms of precision in extracting privacy information.
- Build a privacy information classification model. Using the BERT-Softmax model to classify privacy information. This paper uses one hot to encode text, converting unstructured text sequences into structured feature vectors, and inputting text features into the fully connected layer to calculate the probability of each category label through the Softmax function. Experimental analysis shows that this classification model is significantly superior to traditional classification models.
- This paper proposes the concept of user portrait for privacy information, which determines the risk of user privacy leakage by calculating the average sensitivity of privacy information.

This article is mainly divided into four chapters. In Section I, introduction is mentioned which mainly introduces the background and significance of this study. Section II is related work, which introduces the research of domestic and foreign scholars on user portrait. Section III is algorithms and model, which introduces the algorithm and model of this article. Section IV is experimental results and analysis and Section V concludes the paper.

## II. Related Work

At present, research on user portrait can be broadly classified into three main categories: user behavior-based user portrait model, interest-based user portrait model, and topic-based user portrait model.

Alan Cooper was the first to suggest the idea of a user portrait. A user portrait is a designated user model based on some actual data points (such as social traits and consumption variables) [8].

*1) User behavior-based user portrait model:* Li et al. [9] developed a model that utilizes big data mining and analysis technology to construct a student portrait based on extensive data from campuses. By extracting the characteristics and attributes from a vast amount of behavioral data, they are able to create a comprehensive portrait of a student. Minghui You et al. [10] proposed a behavior-aware user profiling technique that utilizes data mining of user attributes to construct an initial user portrait by identifying user behavior patterns through perception. Zhang et al. [11] introduced an enhanced fuzzy MLKNN multi-label learning algorithm based on MLKNN, aiming to address the challenges of subjective augmentation caused by credit data discretization and the absence of multi-dimensional *credit user portrait in current credit data research.*

*2) Interest-based* user portrait model: Shufang Wu et al. [12] put forward an interest transfer-based user portrait building approach, which serves as a remedy for the inadequacy and inaccuracy of current microblog user portrait creation methods in capturing user characteristics. Ding Z et al. [13] proposed the LDA-RCC model to analyze the interests of forum users and create user portrait.

*3) Topic-based user portrait model:* Jianyun Wu et al. [14] proposed by analyzing the crawled videos, users, and their viewing data through text mining, this paper build a single user portrait, cluster the users, and extract themes through K-Means and LDA models to explore the characteristics of group users. Deng et al. [15] proposed a user portrait fraud warning scheme based on publicly available data on Weibo. They conducted preliminary screening and cleaning based on the keyword "being scammed" to obtain effective fraudulent user identity documents. Through feature engineering techniques such as avatar recognition, artificial intelligence (AI) sentiment analysis, data filtering, and fan blogger type analysis, these images and texts were abstracted into user preferences and personality characteristics, and multi-dimensional information was integrated to construct user portrait.

With the increasing awareness of privacy among individuals, personal privacy issues are increasingly receiving attention. To address these issues, this paper applies user profiling technology to social networks and proposes the concept of user privacy information profiling for social networks. By extracting five dimensions of privacy information from the historical information shared by users on social networks, a user portrait of privacy information is constructed to visually display potential privacy threats and risks to users, and to intuitively understand potential privacy issues and leakage risks.

To benefit the readers with a quick reference, the major notations of this paper are listed in Table I.

TABLE I. MAJOR NOTATIONS

| Notation | Description |
|---|---|
| *TF* | Term Frequency |
| *IDF* | Inverse document frequency |
| *TF-IDF* | sensitivity $w_i$ |
| $w_i$ | Sensitivity of privacy information |
| *Current* ($P_i$) | Location of the node where the current keyword is located |
| *next(Current,*$P_i^j$*)* | Location of the node where the next keyword is located |
| *F* | Transfer Functions |
| *Query(Q)* | The relationship between the current word to be queried and other words |
| *Key(K)* | Vector used to calculate attention weights |
| *Value(V)* | Weighted results |
| $W^Q, W^K, W^V$ | Weight matrix |
| $d_k$ | Dimension of input information |
| $W^O$ | Mapping Vector for Multi head |
| *Concat(·)* | Concatenation function |
| $h_\theta(x^i)$ | Softmax regression model discriminant function |
| $J_\theta$ | The Cost Function of Softmax Regression Model |
| $x^i$ | Input samples |

## III. ALGORITHMS AND MODEL

The multidimensional privacy information portrait of social network users is mainly divided into three modules: the privacy information sensitivity calculation model, the social network user privacy information extraction model, and the BERT-based privacy information classification model. The overall structure of the Chinese social network user multidimensional privacy portrait model is shown in Fig. 1.

The user portrait of social network privacy information depicts the personal privacy data exposed by users while using social networks from various perspectives and dimensions. Firstly, the word bag model is utilized to extract features related to privacy information. It involves transforming the sparse word frequency matrix regarding privacy information into a word bag vector. Subsequently, the TF-IDF algorithm is employed to determine the sensitivity of the privacy data. Next, the PROLM_FlashText+_SMA algorithm is utilized to extract the user's privacy information and its corresponding sensitivity. The average sensitivity of the user's privacy data is calculated to classify the risk of privacy leakage into high, medium, and low levels. Lastly, the BERT Softmax model is employed to categorize privacy information as high privacy, moderate privacy, or mild privacy, enabling the construction of a user portrait for social network privacy information.

### A. A Model for Calculating the Sensitivity of Privacy Information of Social Network Users

This paper employs the TF-IDF algorithm, utilizing the word bag model, to assess the sensitivity of private information. The concept of TF-IDF was introduced by Karen Spärck Jones [16] and entails a fusion of term frequency (TF) and inverse document frequency (IDF).

Bag-of-words model, is a simple mathematical model for describing text and a common way to extract text features [17]. This model disregards the grammatical and sequential aspects of the text, treating it as a collection of words. Each word occurrence in the document is considered as an independent entity.

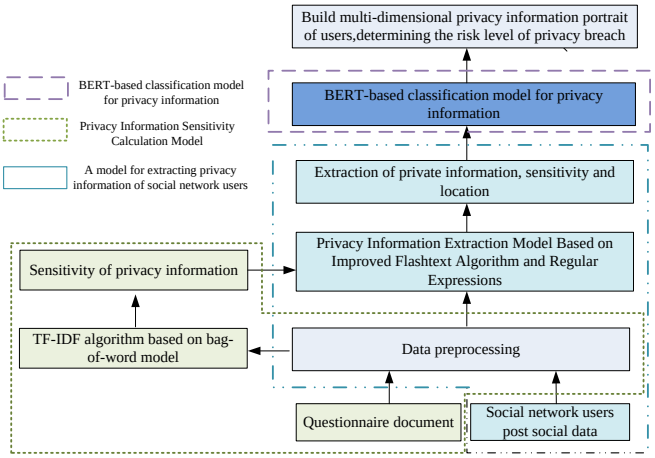


Fig. 1. Overall structure of the multidimensional privacy portrait model of social network users.

The TF-IDF algorithm takes into account both the frequency of a term in a specific document and the importance of the term in the entire document collection, thereby reflecting the significance of the term more accurately in the text. By computing the inverse document frequency, the TF-IDF algorithm can consider the distribution of privacy information across the entire corpus and capture the global characteristics of terms. This enables it to calculate privacy sensitivity more accurately and better represent the importance of privacy information to users. Therefore, this paper utilizes the TF-IDF algorithm to calculate the sensitivity of privacy information.

The model for calculating privacy information sensitivity consists of two main modules: Privacy information feature extraction, which involves extracting features from pre-processed text using the bag-of-words model. Privacy information sensitivity calculation, which includes transforming the text into a bag-of-words vector and ultimately calculating the sensitivity of privacy information using the TF-IDF algorithm.

*1) Privacy information feature extraction:* The structure diagram of privacy information feature extraction is shown in Fig. 2.

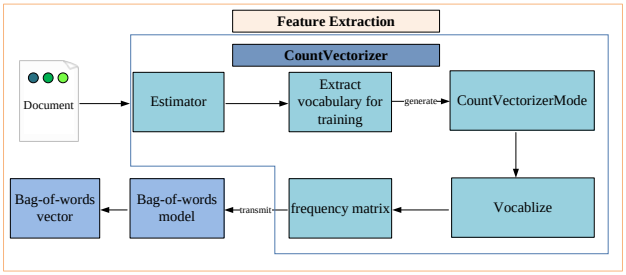


Fig. 2. Privacy information feature extraction structure diagram.

The specific steps are.

Step 1: Construct a corpus Document (pre-processed survey questionnaire document).

Step 2: Training a CountVectorizer model and creating CountVectorizerMode, extracting privacy information words in the document for training and obtaining the vocabulary i.e. all privacy information words present in the corpus.

Step 3: Transforming the text into a word frequency matrix of documents regarding privacy information words.

Step 4: Build a word bag model and convert the word frequency matrix into a word bag vector.

*2) Privacy information sensitivity calculation:* In this stage, the TF-IDF (sensitivity $w_i$) value of privacy information is calculated based on the bag-of-words vector. The specific steps are.

Step 1: Calculate the TF value. The TF value depends only on the number of times the privacy information words are in the corpus and is calculated as:

$$TF = \frac{D_{w,\mathrm{d}}}{\sum_k D_{k,d}} \tag{1}$$

where, D denotes the corpus, the symbol $D_{w,d}$ denotes the number of occurrences of privacy word w in document *d*, and the denominator denotes the total number of occurrences of all privacy words w in D.

Step 2: Calculate the IDF value, which is calculated as:

$$IDF = \log \frac{|D|}{DF_{w,D} + 1} \tag{2}$$

where, $|D|$ denotes the total amount in the corpus, $DF_{w,D}$ denotes the number of documents containing the privacy word w. To avoid the denominator and the occurrence of arithmetic errors, the IDF needs to be optimized, i.e., denominator + 1.

Step 3: Calculate TF-IDF value.TF-IDF value of a privacy word is the product of TF and IDF, which is calculated as:

$$TF - IDF = TF * IDF \tag{3}$$

### B. A Model for Extracting Privacy Information of Social Network Users

The FlashText algorithm, created by Singh V, is designed to specifically match complete English words. It serves as a novel and efficient algorithm for keyword search and replacement [18]. However, considering the disparities between English and Chinese texts, this paper enhances the FlashText algorithm to cater to Chinese texts. The enhanced algorithm, known as FlashText+, incorporates jieba word segmentation and constructs a Chinese Trie dictionary. By doing so, the FlashText+ algorithm not only retains the advantages of the original FlashText algorithm but also becomes applicable in Chinese scenarios. Fig. 3 illustrates the structural diagram of the model for extracting user privacy information in social networks.

In the privacy information extraction stage, in order to facilitate the extraction of user privacy information, this paper have constructed two Chinese privacy information dictionaries, dic1 and dic2, which store two parameters, respectively, where dic1 = (privacy word, sensitivity) and dic2 = (keyword, sensitivity). The privacy word indicates user privacy information that can be accurately collected and defined, while the keyword indicates user privacy information that contains a specific keyword or word.

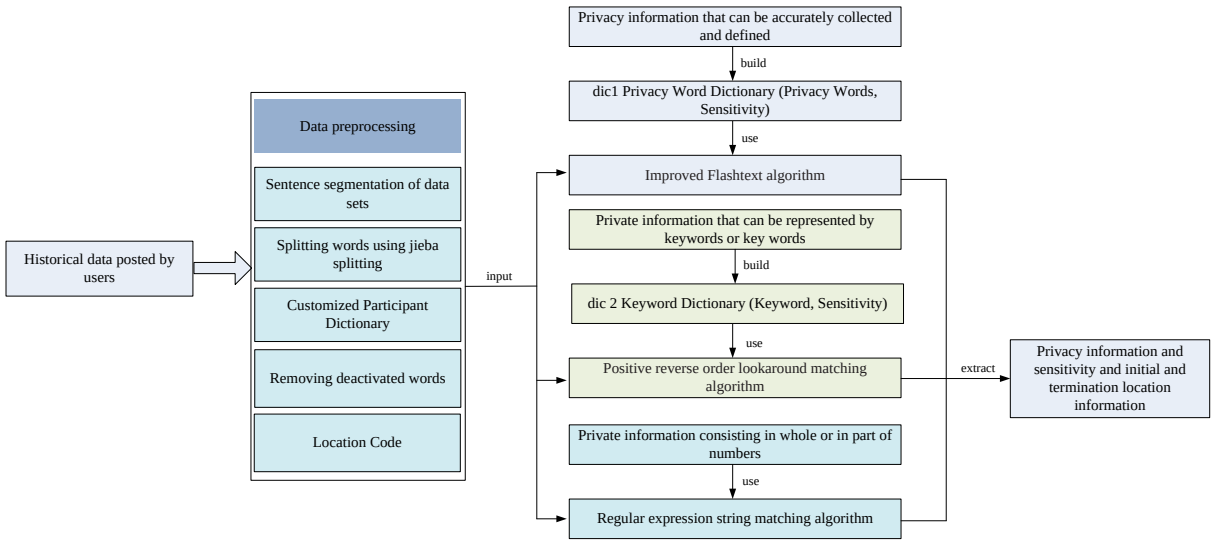


Fig. 3. Structure of social network user privacy information extraction model.

In this paper, this study classify users' privacy information into three types, i.e., privacy information that can be accurately collected and defined, privacy information that cannot be accurately collected and defined but can be represented by a keyword or word, and privacy information that consists entirely or partially of numbers. The first type of privacy information extracted using the improved FlashText algorithm; for the second type of privacy information, the keywords defined in dic2 are used to extract the user's privacy information and its sensitivity and its location using the PROLM algorithm; for the third type of privacy information, the corresponding rules are defined to extract the privacy information, sensitivity and its location using the string matching algorithm.

Example 1. A message posted by a user T = {Punch up Henan University, near Longting District.}

*1) Extracting* user privacy information using an improved FlashText algorithm.

The improved FlashText algorithm extracts user privacy, sensitivity, and location information in the following steps.

Step 1: Data pre-processing. Data preprocessing is performed on the user-posted information T. The word order and location information of T after preprocessing are shown in Fig. 4.

Step 2: Build a Chinese Tree dictionary.

Construct a Chinese tree dictionary using the privacy words in dic1 as input. First, create an empty node root which is the starting point of all private words, then insert all private words into the Trie tree one by one, if there is already a node pointing to the current character in the path during the insertion process, then visit the node, if there is no corresponding node, then create a new node pointing to the current character, if a private word completes the insertion, mark the last node in its path. The process is as in Algorithm 1.

**Algorithm 1:** Build a Chinese Tree dictionary

Input: P= {$p_1$, $p_2$, ..., $p_r$}
Output: Chinese Tree dictionary
Create an initial state empty root 0
For i∈1, 2 ,....., r Do
  Current ← root
  j ← 1
  Pi ← Current ($P_i$)
  While j ⩽ $m_i$ AND next (Current, $P_i^j$) ≠ end Do
    Current ← next (Current, $P_i^j$)
    j←j+1
  End of While
  If j ⩾ $m_i$ Do
    Create a new non empty node Stat
      next (Current, $P_i^j$ ) ←State
      Current ← State
      j← j+1
      break
  End If
  If Current is already terminal Then F(Current)←F(Current) ∪ {i}
  Else mark Current as end, F(Current)←{i}
End of For

This study defines some of the notations used in Algorithm 1. *P* is the privacy word in dic1, root is the root node, *i* is the position of the privacy word, *j* is the position of the node in the tree dictionary, *Current* ($P_i$) is the position of the node where the current keyword is located, *next (Current,* $P_i^j$*)* is the position of the node where the next keyword is located, and *F* is the transfer function.

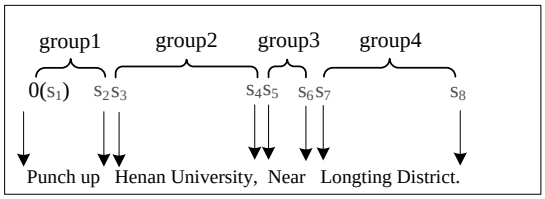


Fig. 4. Phrase order and position information of L after pre-processing.

Step 3: Iterate through the Chinese tree dictionary and g output the results. The pre-processed T is used as input, and each node in the Chinese tree dictionary is traversed one by one in terms of phrases, starting from the root node root (0) for matching. If the characters match, the current state jumps to the corresponding state. When the output state (the state with shaded background) is reached, the corresponding private word for the sequence of matched nodes is output, and the sensitivity is returned. Additionally, the location of the privacy information is saved. If the characters do not match, the next matching step is performed. The traversal process is shown in Algorithm 2.

**Algorithm 2:** Privacy Information Extraction

Input: P={$p_1$,$p_2$,...,$p_r$}, T=($t_1$,$t_2$,...,$t_n$)
Output: Privacy Words pi, Sensitivity, Start and end positions
Preprocessing
  FT← Build_FT(P)
Searching
  Current ← Initial state of the Flashtext FT
  For pos ∈1,2,...,n Do
    While $next_{FT}$ (Current, $t_{pos}$) = end AND $S_{FT}$(Current) ≠ end Do
      Current ← $S_{FT}$(Current)
    End of While
    If $next_{FT}$ (Current, $t_{pos}$) ≠ end Then
      Current ← $next_{FT}$ (Current, $t_{pos}$)
    Else Current ← initial state of FT
    End of If
    If Current is terminal Then
      Mark all the occurrences (F(Current), pos)
    End of If
  End of For
F

This study define some symbols used in Algorithm 2, P is the privacy word in dic1, T is the input text, $next_{FT}$ is the next matching node position, $S_{FT}$ is the current matching node position, and *F* is the transfer function.

*2) Extracting* user privacy information using the PROLM algorithm.

PROLM algorithm focuses on matching results and omits duplicated content, enhancing matching speed by searching from right to left. However, it cannot match privacy information consisting of digits alone. A combination with a string matching algorithm is proposed to extract such information.

PROLM algorithm：Matching text from right to left can be abstractly represented as: ( ?<=SubExp1)|(?<=SubExp2).

According to the characteristics of SubExp1 can be summarized into three categories.

- The subexpression SubExp1 is of fixed length and is in general mode.
- The subexpression SubExp1 has a variable length, for non-greedy mode.
- The subexpression SubExp1 is not fixed in length, but it contains match-first quantifiers, which is a greedy mode.

The matching principle of the PROLM algorithm is shown in Fig. 5, which can be divided into main matching and sub-matching.

Its main idea is to extract phrases containing keywords in dic2 from data published by users.

Taking the text in Fig. 5 as an example, the main matching process is shown Algorithm 3.

**Algorithm 3:** Main matching

Input: $T=\{t_1,t_2,...,t_n\}$, $dic2=\{k_1, k_2,....,k_n\}$
Output: Privacy Words, Sensitivity, Start and end positions
for i∈1,2,...,n Do
  i← 1
  while there is no keyword ki in dic2 in T Do
  end of while
  if T contains the keyword ki in dic2 Do
    find the keyword position Si from the initial position Sk to the right
    $S_k$←$S_{i\text{-}len}$(SubExp1)
    Enter sub matching
  End of If
  if next round of sub matching is required
    give control to subsequent word expressions
  else location of report Si, matching failed
  end of if
  if Subsequent sub expression matching succeeded
    location of report $S_i$ , match succeeded
  else location of report $S_i$, matching failed
  end of if
End of for

This study define some of the symbols used in Algorithm 3, T is the input text, $k_n$ is the keyword in dic2, $S_k$ and $S_i$ are the location of the text, and *len(SubExp1)* is the length of "SubExp1".

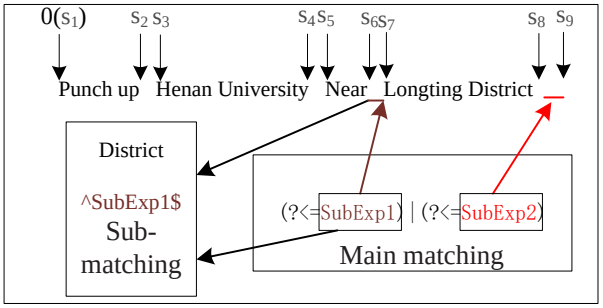


Fig. 5. PROLM algorithm matching principle.

The main matching process is divided into the following steps.

Step 1: Matching is attempted from position 0 to the right, before finding the position that satisfies (? <=SubExp1) and contains the keyword in dic2, the matching must fail until position $S_9$ is found and its requirements are met.

Step 2: Locate the position $S_6$, which satisfies the minimum length requirement for "SubExp1", by moving leftwards from $S_9$.

Step 3: Perform a sub-matching process by applying "SubExp1" from position $S_6$ and moving rightwards.

Step 4: Successfully match "(? <=SubExp1)" and proceed to the subsequent sub-expression "(? <=SubExp2)". Keep attempting to match until the entire expression either matches or fails. Report whether the entire expression at position S9 matches successfully or fails.

Step 5: If necessary, continue to find the next position $S_5$ and start a new round of attempted matching.

The sub-matching process is divided into the following main steps.

Step 1: Upon entering sub-matching, the source string has been determined as the string between S9 and S6, and the regular expression at this point becomes "^SubExp1$". In this round of sub-matching, once a match is successful, it must start at S9 and end at S6.

Step 2: Once the sub-expression is fixed, whether the match succeeds or fails, the match result is returned, and there is no need for further rounds of matching.

Step 3: When the length of the subexpression is not fixed, in the case of a greedy mode, if the match fails, it is reported as a failure and the next round of matching is requested. If the match is successful, all backtracking states are discarded and the success is reported, eliminating the need to try the next round of matching.

Step 4: In the case of a greedy mode, if the match fails, it is reported as a failure and the next round of matching is requested. If the match succeeds, all backtracking states are discarded, the success is reported, and the successful content of this match is recorded. The next round of matching is requested until the longest match is obtained.

The sub-matching process is shown in Algorithm 4.

**Algorithm 4:** Sub matching

```
Input: S_k, S_i
Output: return to main matching
Sub Match Start
    if fixed length of sub-expressions do
        report matching results, no need for next round of sub matching
    else
        next
    en of if
    if subexpression is not greedy mode and matching succeeded
        report matching results, no need for next round of sub matching
    else
        matching failure, request to enter the next round of sub matching and return to main matching
    end of if
    if the sub expression is greedy and matched successfully
        matching success, record this success, request to enter the next round of sub matching
    end of if
    if the sub expression is greedy and matched failure
        return to main matching
    end of if
```

*3) Extracting* user privacy information using regular expression string matching algorithm.

Use regular expression string matching algorithm to define corresponding matching rules to extract privacy information and sensitivity from the preprocessed text for privacy information that is entirely or partially composed of numbers, such as a mobile phone number, license plate number, mailbox, etc.

### C. BERT-Softmax Classification Model for Privacy Information

*1) BERT model:* BERT is a pre-trained language model based on a Transformer encoder, which enables reading the whole text at once for bi-directional linguistic representation to achieve the prediction task, and the input Embedding is encoded and transformed through layers of Encoder. In this paper, a pre-trained BERT-based model is utilized to classify the extracted user-privacy information.

Compared to traditional text classification methods, text classification based on BERT pre-trained models has the following advantages:

- Better semantic understanding: The BERT model can learn contextual information from the text, leading to a better understanding of the text's meaning and improving the accuracy of text classification.
- Improved robustness: The BERT model can handle texts of different lengths, enhancing the model's robustness and enabling it to process various types of text data.
- Good generalization capability: Text classification models based on BERT can be fine-tuned to adapt to different text classification tasks, thereby exhibiting good generalization capability.

The structure of the BERT model [19] is shown in Fig. 6. It consists of a multilayer bidirectional Transformer encoder. Where, the middle Trm denotes the Transformer encoder, $E_1$, $E_2$,...$E_N$ denotes the text input vector of the word, the vectorized representation of the text is obtained after the Trm module, and $T_1$, $T_2$,...$T_N$ denotes the final text representation.

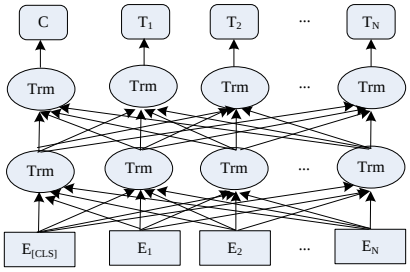


Fig. 6. Structure of the BERT model.

*2) BERT-Softmax* classification model. The structure of the BERT- Softmax classification model is shown in Fig. 7.

The classification process is as follows: the extracted user-privacy information is used as input text data, the text features are extracted using the BERT pre-training model, i.e., the text is encoded using one-hot, the unstructured text sequence is converted into a structured feature vector, the text features are input into the fully connected layer and the probability of each type of label is calculated by the Softmax function, and the label corresponding to the maximum probability is the result of the model classification.

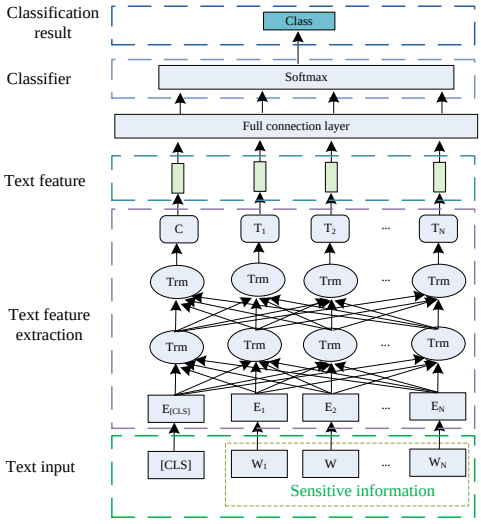


Fig. 7. Structure of BERT-based pre-trained classification model.

*3) Attention mechanism:* whose core idea is to calculate the interrelationship between each word in the input text and all the words in that text, and to measure the relevance and importance of different words in the input text, and to adjust the weight of each word by these interrelationships to obtain a

new expression for each word, which contains not only the semantics of the word itself but also its relationship with other words [20].

The calculation process of Self-attention for a single word is shown in Fig. 8, where Query denotes the current word to go to a query concerning other words, Key denotes waiting to be checked, and Value denotes the actual feature information.

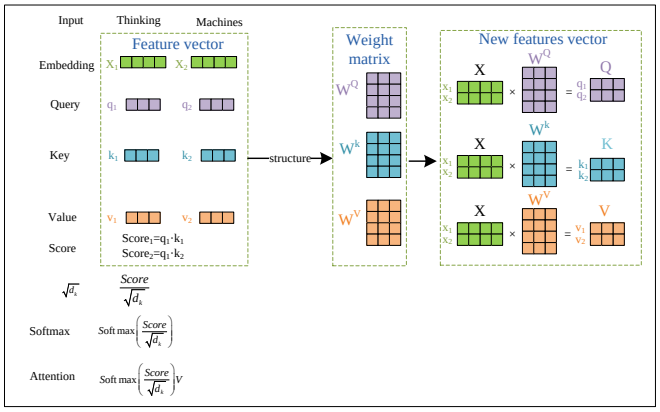


Fig. 8. Self-attention calculation process for a single word.

The calculation steps are as follows:

Step 1: For each word of the input after one-hot encoding get the initial feature vector noted as $X=[x_1,x_2]$, and initialize three weight vectors q, k, v;

Step 2: With training the entire weight matrix $W^Q$, $W^K$, $W^V$ is constructed;

Step 3: Use Eq. (4) to obtain three new feature expression matrices Q(Query), K(Key), and V(Value) for each word;

$$\begin{cases} Q=XW^Q \\ K=XW^K \\ V=XW^V \end{cases} \tag{4}$$

Step 4: Calculate the inner product Score using Eq. (5), which indicates the relationship between the current word and other words.

$$Score = QK^T \tag{5}$$

Step 5: To prevent the Score from increasing with vector dimension, the inner product is calculated by dividing by the factor $\sqrt{d_k}$, $d_k$ being the vector dimension of the input information and normalized to a probability distribution by the softmax function.

Step 6: The distribution of scoring by inner product is the weighted average of Value, and Attention is calculated using Eq. (6).

$$Attention(Q,K,V) = \text{Soft max}\left(\frac{Score}{\sqrt{d_k}}\right)V \tag{6}$$

In practice, the Transformer encoder uses Multi-head attention mechanism, Multi-head mechanism that is Multi-head attention is a network optimization technique used in BERT network structure, using different heads to focus on different context dependency patterns, similar to the model integration effect, which can achieve parallel operations, the input of the network into multiple branches, respectively, do attention mechanism, and finally, the results of each branch will be spliced to get, its calculation as in Eq. (7) and Eq. (8).

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \tag{7}$$

$$Multi-head(Q,K,V) = Concat(head_i)W^O \tag{8}$$

where, $W_i^Q$, $W_i^K$, and $W_i^V$ represent the $W^Q$, $W^K$, and $W^V$ weight matrices of the ith head, $W^O$ denotes the mapping vector of Multi-head, and Concat(·)denotes the splicing function.

*4) Classification:* In this paper, introduce softmax regression model for privacy information classification. The softmax regression, like linear regression, does a linear superposition of the input features and ourights. One major difference from linear regression is that the number of output values in softmax regression is equal to the number of categories in the label. Assuming there is a training sample set $\{(x^1, y^1), (x^2, y^2),..., (x^m, y^m)\}$, where $x^i$ represents the privacy information vector corresponding to the i-th training sample, with a dimension of n for a total of m training samples, $y^i \in (1,2,..., n)$ represents the category corresponding to the i-th training sample, and n is the number of categories. For a test input sample x, the distribution function of the Softmax regression model is the conditional probability p(y=j|x), indicating the probability that x belongs to category j, where the category with the highest probability of occurrence is the category to which the current sample belongs, and the hypothesis function of belonging to each category is as in Eq. (9).

$$h_\theta(x^i) = \begin{bmatrix} p(y^i=1|x^i;\theta) \\ p(y^i=2|x^i;\theta) \\ \cdots \\ p(y^i=n|x^i;\theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^{n} e^{\theta_j^T x^i}} \begin{bmatrix} e^{\theta_1^T x^i} \\ e^{\theta_2^T x^i} \\ \cdots \\ e^{\theta_n^T x^i} \end{bmatrix} \tag{9}$$

Among them, the probability that any element $p(y^i=n|x^i; \theta)$ in $h_\theta(x^i)$ is the current input sample $x^i$ belonging to the current category *n*.

For a data set with *m* training samples, the cost function of the Softmax regression model is as in Eq. (10).

$$J_\theta = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{n} I\left\{y^{(i)}=j\right\}\log\frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^{n} e^{\theta_j^T x^{(i)}}}\right] \tag{10}$$

where, *I{.}* is the indicative function, *m* is the number of samples, *n* is the number of categories, *i* denotes a certain sample, $x^i$ denotes the vector representation of the *ith* sample *x*, and *j* denotes a certain category.

After obtaining the model parameters $\theta$, the probability of belonging to category *j* for the sample *x* to be measured is:

$$p\left(y^i = j \mid x^i;\theta\right) = \frac{e_j^T x^i}{\sum_{l=1}^{n} e_l^T x^i} \quad (11)$$

The probability of x belonging to all categories is calculated using Eq. (11), and the category with the highest probability is chosen as the classification category for x.

### D. Building a Portrait of User Privacy Information

The construction of user privacy portrait first extracts user privacy information through the PROLM_ FlashText+_ SMA algorithm, and then uses BERT-Softmax to classify the privacy information. Then, use Eq. (12) to calculate the average sensitivity of privacy information to determine the level of user privacy leakage risk. Finally, use word cloud technology to construct user portrait of privacy information.

$$W = \frac{\sum_{i=0}^{n} w_i}{n}, (n \geq 0, i \geq 0) \quad (12)$$

where, $W$ denotes the average sensitivity of privacy information, $n$ denotes the number of privacy information, $i$ denotes the i-th privacy word, and $w_i$ denotes the sensitivity of the i-th privacy word.

Set a threshold to divide the risk of user privacy leakage into three levels.

Heavy leakage risk: average sensitivity $W > 0.11$.

Moderate leakage risk: average sensitivity $0.07 < W < 0.11$.

Mild leakage risk: average sensitivity $W < 0.07$.

## IV. Experimental Results and Analysis

Due to the lack of public datasets related to the precise expression of user privacy, the experimental datasets for user privacy information in this paper are collected by crawlers and collected manually. The dataset in this paper is divided into four parts: 1071 valid questionnaire data for user privacy information sensitivity calculation, 1000 social network users' microblogging history data, a dictionary containing 7673 user privacy words and a dictionary containing 199 user privacy keywords for privacy information extraction, and 60,000 user privacy data for privacy information classification. Experiment in this paper is divided into three parts: user privacy information sensitivity calculation, user privacy information extraction, and user privacy information classification.

### A. User Privacy Information Sensitivity Calculation

The dataset used in this experiment is a 1071 valid survey questionnaire, which defines a total of 62 types of privacy information.

This paper uses the BOW-TF-TDF algorithm to calculate privacy sensitivity, and arranges the sensitivity calculation results in descending order, as shown in Table II.

TABLE II. Sensitivity of User Privacy Information

| Privacy words | Sensitivity | Privacy words | Sensitivity |
|---|---|---|---|
| ID number | 0.155 | Purchase preferences | 0.125 |
| Email | 0.154 | Private financial list | 0.124 |
| health condition | 0.154 | Payment records | 0.124 |
| phone number | 0.153 | Income situation | 0.123 |
| Home Address | 0.153 | Purchase Record | 0.123 |
| Property status | 0.152 | major | 0.122 |
| Bank card number | 0.152 | Purchase frequency | 0.121 |
| Card Number | 0.152 | Consumer credit | 0.121 |
| academy | 0.152 | Shopping habits | 0.121 |
| Political landscape | 0.151 | Consumption amount | 0.12 |
| height | 0.15 | Dressing hobbies | 0.118 |
| Current address | 0.15 | Literary Hobbies | 0.118 |
| height | 0.15 | Sports hobbies | 0.116 |
| Gender | 0.149 | Consumption level | 0.116 |
| marital status | 0.149 | Travel Hobbies | 0.115 |
| Age | 0.143 | Game Hobbies | 0.1118 |
| weight | 0.14 | Weibo account | 0.106 |
| position | 0.137 | Chat records | 0.104 |
| educational | 0.137 | QQ number | 0.102 |
| Online behavior | 0.134 | Tiktok | 0.102 |
| Mode of travel | 0.134 | Kuaishou | 0.102 |
| Life experience | 0.134 | Station B | 0.101 |
| Social relationship | 0.133 | WeChat signal | 0.099 |
| Family member | 0.133 | Alipay account | 0.096 |
| Device Information | 0.133 | The Little Red Book | 0.095 |
| Work unit | 0.132 | Investment hobbies | 0.013 |
| Personal itinerary | 0.132 | Art Hobbies | 0.011 |
| IP address | 0.132 | Pets | 0.003 |
| name | 0.13 | constellation | 0.0001 |
| occupation | 0.13 | religious belief | 0.00002 |
| Training experience | 0.13 | blood group | 0.00002 |

According to Table II this paper classifies privacy information into three levels.

High privacy: when $wi$ >0.13, the privacy information includes the user's ID number, home address, name, IP address, and other information, which can accurately locate the user's identity and location once exposed.

Moderate privacy: when 0.13 >$wi$ >0.1, the privacy information includes information about the user's hobbies, social behavior habits, etc., which can only be roughly understood about the user even if exposed.

Mild privacy: when $w_i < 0.1$, this type of information, even if exposed, will not have a major impact on the user.

### B. User Privacy Information Extraction

The dataset used in this experiment is the basic Weibo information of 1000 social network users and 65536 historical and dynamic information. This paper evaluates the superiority of this scheme based on the precision of privacy information extraction.

*1) Data preprocessing：* This paper mainly collects data from Weibo users on social networking platforms.

Data preprocessing mainly includes filtering the empty data in the dataset, using the Jieba word segmentation library to segment the dataset based on the stop list proposed by the Harbin Institute of Technology laboratory, and removing stop words, special symbols, and meaningless words.

*2) Evaluation indicators:* This paper evaluates the performance of different models by extracting privacy information precision.

The goal of privacy information extraction is to correctly extract all the privacy information from the historical data posted by users, and Precision is an important indicator of how good the information extraction model is. Precision is calculated by the following formula.

$$\Pr ecision = \frac{|\text{privacy}_u|}{|extreact_u|} * 100\% \tag{13}$$

where, $|privacy_u|$ denotes the number of extracted privacy information, $|extreact_u|$ denotes the total number of extracted information.

This paper selects classic text information extraction models BILSTM, BILSTM-CRF, and LSTM-CRF as comparative experiments. We divided the dataset into training, validation, and testing sets in a ratio of 8:1:1. The distribution of the dataset is shown in Table III.

TABLE III. PRECISION COMPARISON OF DIFFERENT PRIVACY INFORMATION EXTRACTION MODELS

| Model | Precision |
|---|---|
| BILSTM | 87.92 |
| BILSTM-CRF | 90.32 |
| LSTM-CRF | 89.64 |
| **PROLM_FlashText+_SMA** | **93.63** |

According to the data in Table III, it can be seen that the PROLM_FlashText+_SMA algorithm proposed in this paper has the highest precision rate in privacy information extraction tasks (93.63), followed by BILSTM-CRF (90.32), and the model BILSTM has the lowest precision rate (87.92). Experiments have shown that the PROLM_FlashText+_SMA algorithm outperforms the comparative model in terms of precision in extracting privacy information.

### C. User Privacy Information Classification

*1) Experimental dataset：* The classification dataset used in this paper is the THUCNews text classification dataset provided by the Tsinghua NLP group and the author collection of 60,000 pieces of data about users' privacy information, containing three types of data: highly privacy, moderately privacy and mildly privacy information, and this paper divide the dataset into the training set, validation set, and test set according to 8:1:1, and the distribution of privacy information classification dataset is shown in Table IV.

TABLE IV. CLASSIFICATION OF PRIVACY INFORMATION CLASSIFICATION EXPERIMENTAL DATASET

| Category | Label | Train Set | Test Set | Validation Set |
|---|---|---|---|---|
| Highly privacy | 1 | 8000 | 1000 | 1000 |
| Moderately privacy | 2 | 8000 | 1000 | 1000 |
| Mildly privacy | 3 | 8000 | 1000 | 1000 |

*2) Model and parameter setting：* The experimental environment is Windows 10, based on the PyTorch framework, CUDA version 11.7.1, CUDANN 8.5, and the GPU used to accelerate the training is RTX3050, and the model parameters are set as shown in Table V.

TABLE V. PARAMETERS OF THE MODEL

| Parameters | Numerical value |
|---|---|
| hidden_size | 768 |
| epoch | 3 |
| batch_size | 32 |
| pad_size | 32 |
| learning_rate | 5e-5 |
| Attention Dimension | 64 |

*3) Evaluation indicators：* The common evaluation metrics for classification problems are the Precision, Recall, and F1 (F1-Score) index. Precision and recall are important metrics to measure how good an information extraction model is. The F1-score is the summed average of Precision and Recall, which is used to evaluate the precision and recall together.

The precision calculation formula equation is as follows.

$$Precision = \frac{Sum_{true}}{Sum_{forecast}} \tag{14}$$

where, $Sum_{true}$ indicates the number of privacy information correctly classified, and $Sum_{forecast}$ indicates the number of privacy information predicted to be in that category.

The recall is calculated as follows.

$$Recall = \frac{Sum_{true}}{Sum_{actual}} \tag{15}$$

where, $Sum_{actural}$ indicates the actual amount of privacy information in that category.

The formula for calculating the F1-Score is as follows.

$$F1 = \frac{\Pr ecision * \text{Re} call}{\Pr ecision + \text{Re} call} * 2 \tag{16}$$

In this paper, we choose the classical text classification models BERT_CNN, BERT_RNN, BERT_RCNN, and ERNIE as the comparison experiments. The training set, validation set, and test set data are kept consistent with the BERT model during the training process, and the Precision, Recall, and F1 index of different models are compared as shown in Table VI.

TABLE VI. COMPARISON OF PRECISION, RECALL, AND F1 INDEXES OF DIFFERENT CLASSIFICATION MODELS

| Model Name | Precision | Recall | F1-Score(F1) |
|---|---|---|---|
| BERT_CNN | 0.9231 | 0.9530 | 0.9378 |
| BERT_RNN | 0.9382 | 0.9550 | 0.9419 |
| BERT_RCNN | 0.9384 | 0.9640 | 0.9510 |
| ERNIE | 0.9412 | 0.9730 | 0.9568 |
| **BERT-Softmax** | **0.9846** | **0.9750** | **0.9798** |

From Table VI, it can be seen that the privacy information classification model proposed in this paper based on BERT Softmax has an accuracy and recall rate of over 0.97 in privacy information classification experiments, and the F1 index reaches 0.9798. From the overall experimental results, it can be seen that the F1 value of the BERT model is higher than that of the BERT_CNN model is four percentage points higher than the ERNIE model by two percentage points, indicating that this model is compared to BERT_CNN, BERT_RNN, BERT_RCNN and ERNIE classification models can better represent the semantic information of short text privacy information and have better privacy information classification performance.

## V. CONCLUSION

In order to address the issue of user privacy and social relationships being prone to privacy leakage during data publishing and information sharing, this paper proposes a user privacy information portrait model for social networks. To address the issue of difficulty in measuring user privacy information sensitivity, the BOW-TF-IDF algorithm is used to calculate user privacy information sensitivity; Due to the diversity of user privacy information, this paper proposes the PROLM_FlashText+_SMA privacy information extraction algorithm, which extracts the privacy information exposed by users from historical data published on the social Internet of Things. The privacy information is classified into high, moderate, and mild using the BERT-Softmax based privacy information classification model. Finally, Use the extracted privacy information to construct a multidimensional user portrait of privacy information and determine the level of privacy leakage risk by calculating the average sensitivity of privacy information. The experiment shows that the privacy information extraction model proposed in this paper exhibits its superiority in terms of accuracy; the privacy information classification model proposed in this paper outperforms traditional classification models in terms of accuracy, recall, and F1 index. It can better represent the semantic information of short text privacy information and has better privacy information classification performance.

In the future, this study will be combined with trust between users to generate access controls to protect the privacy of users in social networks.

## REFERENCES

[1] T. L. Gu, F. R. Hao, L. Li, J. J. Li and L. Chang, "Behavior Accountability of Agents Responsible for Privacy Negotiation in Social Networks," Ruan Jian Xue Bao/Journal of Software, vol. 33, no. 9, pp. 3453-3469, 2022.

[2] R. N. Xie, X. N. Fan, Y. Lin et al., "Research on extended access control mechanism in online social network," Chinese Journal of Network and Information Security, vol. 7, no. 5, pp. 123-131, 2020.

[3] Al-Asbahi R, "Structural Anonymity For Privacy Protection In Social Network," International Journal of Scientific and Research Publications (IJSRP), vol. 11, no. 6, pp.102-107, 2021.

[4] C. Lian and Z. Chen, "Anonymous privacy protection algorithm based on sensitive attribute classification," 2020 2nd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI), pp. 222-226, 2020.

[5] X. Yin, S. Zhang and H. Xu, "Node Attributed Query Access Algorithm Based on Improved Personalized Differential Privacy Protection in Social Network," International journal of wireless information networks, vol. 26, no. 3, pp.165-173, 2019.

[6] B. Ning, Y. Sun, X. Sun, and G. Li, "Differential privacy protection on weighted graph in wireless networks," Ad Hoc Networks, vol. 110, no.102303 pp. 1-10, 2021.

[7] H. Huang, D. Zhang, F. Xiao et al., "Privacy-Preserving Approach PBCN in Social Network With Differential Privacy," IEEE Transactions on Network and Service Management, vol. 17, no. 2, pp. 931-945, 2020.

[8] Z. Y. He, Q. H. Zhu and M. Bai, "The Construction of Urban Elderly User Portrait from the Perspective of Pension Service," Journal of Information, vol. 40, no. 9, pp. 154-160, 2021.

[9] X. Li and S. He, "Research and Analysis of Student Portrait Based on Campus Big Data," 2021 IEEE 6th International Conference on Big Data Analytics (ICBDA), pp. 23-27, 2021.

[10] M. H. You, Y. F. Yin, L. Xie and S. L. Lu, "User profiling based on activity sensing," Journal of Zhejiang University(Engineering Science ), vol. 55, no. 4, pp. 608-614, 2021.

[11] Z. Zhang, L. Han, M. Chen, "Fuzzy MLKNN in Credit User Portrait," Appl. Sci, vol, 12, no.22, pp. 11342, 2022.

[12] S. F. Wu, C. C. Wu and J. Zhu, "Microblog User Dynamic Portrait Generation Based on Interest Transfer," Information Science, vol. 39, no. 8, pp. 103-111, 2021.

[13] Z. Ding, C. Yan, C. Liu, et al., "Short Text Processing for Analyzing User profiles: A Dynamic Combination," International Conference on Artificial Neural Networks, vol. 12397, pp. 733-745, 2020.

[14] J. Y. Wu and M. Z. Xu, "Video Personalized Recommendation Based on User Portrait and Video Interest Tags," Information Scence, vol. 39, no.1 , pp. 128-134, 2021.

[15] L. An, J. Y. Hu and G. Li, "Research on profiles of High-impact Users on Social Media in the Context of Emergencies," Information and Documentation Services, vol. 41, no. 6, pp. 6-16, 2020.

[16] K. S. Jones, "A statistical interpretation of term specificity and its application in retrieval," Journal of Documentation, vol. 28, no. 1, pp. 11-21, 1972.

[17] X. G. Hu, X. H. Li, F. Xie and X. D. Wu, "Keyword Extraction Based on Lexical Chains for Chinese News Web Pages," Pattern Recognition and Artificial Intelligence, vol. 23, no. 1, pp. 45-51, 2010.

[18] V. Singh, "Replace or Retrieve Keywords in Documents at Scale," https://arxiv.org/pdf/1711.00046v2.pdf, 2017.

[19] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," North American Association for Computational Linguistics (NAACL), 2019.

[20] P. Yang and W. Y. Dong, "Chinese named entity recognition method based on BERT embedding," Computer Engineering, vol. 46, no. 4, pp. 40-45, 2020.