Computers, Materials & Continua CMC, vol.62, no.1, pp.137-155, 2020

Research on Privacy Disclosure Detection Method in Social Networks Based on Multi-Dimensional Deep Learning

Yabin Xu [1, 2, *], Xuyang Meng[1], Yangyang Li[3] and Xiaowei Xu[4, *]

Abstract: In order to effectively detect the privacy that may be leaked through social networks and avoid unnecessary harm to users, this paper takes microblog as the research object to study the detection of privacy disclosure in social networks. First, we perform fast privacy leak detection on the currently published text based on the fastText model. In the case that the text to be published contains certain private information, we fully consider the aggregation effect of the private information leaked by different channels, and establish a convolution neural network model based on multi-dimensional features (MF-CNN) to detect privacy disclosure comprehensively and accurately. The experimental results show that the proposed method has a higher accuracy of privacy disclosure detection and can meet the real-time requirements of detection.

Keywords: Social networks, privacy disclosure detection, multi-dimensional features, text classification, convolutional neural network.

1 Introduction

Social networks are popular and actively pursued by netizens for their convenient, flexible information dissemination and fast and efficient network communication, which fully meets the needs of people to express personal appeals and share various information. Internet users are also increasingly keen to communicate and share information with each other through social networks. Social networks have become an important way for people to communicate, share and to express their opinions.

However, the form of information sharing in social networks represented by microblogs is limited by the number of characters in the early days, which determines that most of the contents published are limited to information closely related to themselves, such as life dynamics, mood state, learning work, daily routine, friends gathering, and so on. By looking closely at or tracking a person's social network sharing, you can capture the footprints of his life and even sum up the bits and pieces of his growth. The mass disclosure of this information has led to the disclosure of many personal private

[1] School of Computer, Beijing Information Science &Technology University, Beijing, 100101, China.

[2] Beijing Key Laboratory of Internet Culture and Digital Dissemination Research, Beijing, 100101, China.

[3] Innovation Center, China Academy of Electronics and Information Technology, Beijing, 100041, China.

[4] Department of Information Science, University of Arkansas at Little Rock, Little Rock, 72204, USA.

[*] Corresponding Authors: Yabin Xu. Email: xyb@bistu.edu.cn;

Xiaowei Xu. Email: xwxu@ualr.edu.

information intentionally or unintentionally, and spread quickly through social networks. Incidents of privacy breaches are common, and have a serious impact on personal life and family peace.

With the frequent occurrence of privacy leakage events and the increasing awareness of people's privacy and security, how to accurately and efficiently detect the privacy disclosure of information to be published on social networks, and promote users to avoid the impact and harm to themselves and others caused by privacy leaks. This is a hot concern and urgent problem for the social network platform, network regulatory agencies and the whole society. Therefore, realizing the privacy disclosure detection and warning to the social network can not only greatly reduce the leakage of the privacy information of the social network user, but also increase user's trust in the social network platform and improve the activity on the social network. That will contribute to the healthy development of social networks, reduce unnecessary legal disputes, and contribute to social stability.

2 Related work and innovation

In 1890, Warren and Brandeis published the article "The Right to Privacy" in the Harvard Law Review, and put forward the concept of the right to privacy for the first time [Warren and Brandeis (1890)]. They see privacy as a right, defined as the right to be left alone.

With the development of society and the vigorous development of social network, the transition from traditional space to virtual space makes it more difficult to define the concept of privacy completely and accurately in the Internet environment. Zhao [Zhao (2002)] holds that privacy is a kind of personality right that citizens enjoy private life peace on the Internet, and private information can be protected according to law, not being illegally invaded, known, collected, utilized and disclosed by others. The right also refers to the prohibition on the Internet disclosure of sensitive information related to individuals and so on. In the personal information protection policy of Sina Weibo, personal sensitive information is defined as a kind of personal information that may not only endanger personal and property safety, but also easily lead to personal reputation, physical and mental health damage or discriminatory treatment once it is disclosed, illegally provided or abused.

There are many kinds of private information in social networks. Machida et al. [Machida, Shimada and Ecizen (2013)] divides the privacy information in social networks into three categories: 1) personal intrinsic characteristics; 2) personal physical condition and mental state; 3) life status. Islam et al. [Islam, Walsh and Greenstadt (2014)] classifies the private information in Twitter into 10 types, such as geographical location, medical care, drunkenness, personal emotion, etc.

Clarke [Clarke (1999)] explicitly divided privacy into four dimensions: personal privacy, behavioral privacy, interpersonal privacy, and data privacy. On this basis, Belanger et al. [Belanger and Crossler (2011)] merged the latter two categories into information privacy. Domestic scholar [Qiu (2012)] summarized the types of private information in social networks into four aspects: personal information of users, information shared by users, interpersonal information of users and information obtained through data mining.

Private information in social networks is not only diverse, but also various ways of privacy disclosure. Kim et al. [Kim, Jung and Park (2013)] pointed out that some privacy attributes that are not exposed in the social network user profile can be inferred through social relationships. Hou et al. [Hou, Wei, Wang et al. (2018)] proposed a Privacy Preserving Medical Recommendation (PPMR) algorithm to protect patients' treatment information and demographic information during online recommendation process. Yao [Yao (2014)] used the same public information in the community to infer the unpublished private information of the target user. It can be seen that privacy information may also be mined through social relationships and so on.

Many scholars at home and abroad have studied privacy disclosure detection methods in social networks such as Weibo, Facebook, Twitter, and so on, and proposed some models and methods of privacy disclosure detection.

Yao Kai gave a keyword extraction method and then used the keyword matching to detect the privacy in Weibo.

Foreign scholar [Mao, Shuai and Kapadia (2011)] analyzed three types of privacy leakage in Twitter: holiday plan, drunkenness, medical treatment(disease), classified content by keyword matching, and then used naive Bayes, pattern matching and rule matching method respectively to detect privacy leakage of the above three types of tweets. On this basis, the domestic scholar [Jiang (2013)] applied the relevant research conclusions to Sina Weibo in China, and used the two-level naive Bayesian model to detect the above three types of privacy content.

Tesfay et al. [Tesfay and SernaOlvera (2016)] adopted the support vector machine (SVM) model to discriminate private content through a two-classification method. Islam et al. [Islam, Walsh and Greenstadt (2014)] combined topic modeling, named entity recognition, privacy ontology, sentiment analysis, and text normalization to represent privacy features to discover whether the text contains private content. Machida et al. [Machida, Shimada and Ecizen (2013)] divided the privacy content into three categories: personal intrinsic characteristics, personal physical condition and mental state, and life status, and detected the first two types of privacy content through keyword analysis and semantic analysis.

Liu et al. [Liu and Terzi (2009)] proposed a model and algorithm for calculating the privacy scores of online social network users, and used mathematical models to estimate the sensitivity of information. Srivastava et al. [Srivastava and Geethakumari (2013)] used a method similar to Liu to calculate privacy disclosure of personal data attributes in social networks.

Ji et al. [Ji and Xu (2015)] detected the privacy content of the referee's documents. According to litigant's personal information, he combined the method of named entity identification and privacy template matching to detect. For the body part, he extracted features and constructed the SVM decision tree to detect the privacy content. Jiao [Jiao (2017)] used the Naive Bayesian model to classify the text content, and then used the decision tree method to judge the privacy leakage. Zhang [Zhang (2016)] used logistic regression method to conduct privacy detection on Weibo content.

From the above, it can be found that domestic and foreign scholars mainly use two

methods to detect privacy leakage, namely, the method based on keyword matching and the method of machine learning. The former is relatively simple and efficient, but the detection accuracy is low. The latter classifies the text content first, selects its privacy features for each given privacy category, and constructs a corresponding model for detection. The efficiency of this two-step detection process is not high, and it is difficult to classify accurately due to the ambiguity and uncertainty of the privacy categories. In addition, existing methods only considered the current published text, but ignore the privacy leakage of different ways in social networks and the aggregation effect of scattered private information.

The innovations of this paper are as follows:

1) Faced with massive social network data, we divide privacy disclosure detection process into two steps. Firstly, we detect privacy disclosure on the currently published text efficiently based on the fastText model. When it determines that the current text contains private information, we carry out a comprehensive privacy disclosure detection based on multi-dimensional convolutional neural network model (MF-CNN) proposed in this paper. Thereby, it not only improves the efficiency of privacy disclosure detection, but also ensures the accuracy of privacy disclosure detection.

2) There is no need to classify the privacy of the text content, which not only avoids the difficulty of classification caused by the ambiguity and uncertainty of the privacy category, but also enhances the applicability of the detection method because it is not limited to the specified privacy categories.

3) We detect possible privacy leakage from different ways, and fully consider the aggregation effect of private information. It also adopts information fusion technology and deep learning method to conduct comprehensive privacy disclosure detection so that it can improve the accuracy of privacy disclosure detection.

3 The way and framework of privacy disclosure

3.1 Analysis of privacy disclosure

It is generally believed that private information is leaked through the published text, but it is not the case. Social networks not only display content through texts such as Weibo, but also display user profiles, complex social relationships, and information that has been published. Therefore, we not only need to regard the text to be published by the user as one of the contents of privacy leakage detection, but also should identify the private information which is directly leaked from the user profile, inferred through complex social relationships, or exposed from published content.

The key is that just one aspect of the information may not be enough to reveal the privacy of the user, or the degree of privacy leaked is not serious enough. But if the private information leaked from the above aspects is brought together, it may reveal users' privacy very serious. Therefore, it is not enough to establish a detection model for a single aspect. It is necessary to combine the above several aspects, considering various possible combinations, and establish a comprehensive detection model for comprehensively mining privacy leakage. And the model can ensure the efficiency and accuracy.

3.2 The way of privacy disclosure and its information expression paradigm

3.2.1 User profile and its expression paradigm

Taking the Sina Weibo as an example, the Sina Weibo platform provides the personal profile section for each user, so that users can express themselves by labeling, which makes it easy to make friends. However, in order to meet the satisfaction of these social needs, there is a certain risk of privacy leakage. The more items of personal profile information are disclosed, the greater the probability of revealing personal privacy. The information items in the Sina Weibo user profile section are shown in Tab. 1.

Table 1: Information items in user profile

Number	Information items	Number	Information items
1	Real Name	11	QQ
2	Location	12	Types of School
3	Gender	13	School Name
4	Sexual Orientation	14	Admission time
5	Relationship Status	15	Department
6	Birthday	16	Company Name
7	Blood Type	17	Department/Position
8	Slogan	18	Work Time Period
9	Blog	19	Work Location
10	MSN	20	Tags

It can be seen from Tab. 1 that the user profile section has the above 20 optional information. But through statistical analysis of a large number of microblog user information obtained, it is found that only these five items have more actual values: "Location", "Gender", "Educational information", "Occupation" and "Tags", while other information items have little actual value and need not be considered. Therefore, we only consider the five information items with more actual values.

Let the feature matrix of the user profile information be A, then

$$A = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{bmatrix} a_{1,1} & \cdots & a_{1,k} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,k} \end{bmatrix}$$

Among them, n is the number of information items, vector $x_i=(a_{i,1}, a_{i,2}\cdots a_{i,k}\}$， $a_{i,1}=a_{i,2}=\cdots=a_{i,k}=0$ or 1, 0 represents the information item does not exist, 1 means the information item exists.

3.2.2 Social relationship and its expression paradigm

Although the text content published through social networks is the main way of privacy leakage, some studies have shown that some privacy of the user can be inferred from their friendships [Yao (2014); Mao, Shuai and Kapadia (2011)], and attackers can infer some

personal information that users do not want to expose from their social relations. For example, the target user does not want to expose his educational information, but since most of his friends who are concerned about each other are students of a certain school, it can be inferred that the target user is highly likely to be a student of the school.

Through research and analysis, it is found that the user information items that can be inferred through social relationships are limited to some information items shown in Tab. 2.

Table 2: Information items in user profile

Number	Information items
1	Location
2	Education information
3	Occupation

In social networks, social relationship is mainly reflected in the relationship of "follow" and "followed". Generally, users who follow each other are acquaintances relationship based on identity. So the inferred information generally has certain credibility. Based on this, through the social network, we can obtain the user profile item with the mutual concern relationship for the target user, and construct a friend information matrix named Friends_Info with the size of m×3 for the target user, as shown in Tab. 3. Among them, M is the number of friends who are concerned about each other.

Table 3: Friend information matrix with mutual concern

	info 1	info 2	info 3
friend 1			
…			
friend j		V(i,j)	
…			
friend m			

As shown in Tab. 3, the matrix has three columns (three information items in Tab. 2), and each column represents an item of information. The number of rows in the matrix is determined by the number of friends of mutual concern that target users have, and each row represents the information item disclosed by a friend. Each cell V(i, j) represents the value of the information item j disclosed by the friend i, that is, the content item is filled in and set to be visible to everyone.

The information item to be inferred can be represented by a four-tuple, That is, Infer_info=<info_i, target_user, infer_value, confidence_degree>. The info_i represents the unknown information item of the target user and is also the inferred information item; the target_user represents the target user; the infer_value represents the inferential value of the information item; and the confidence_degree represents the credibility of the inferred value.

The inference principle and process of information items are as follows:

First, we look at the three information items of the target user (in Tab. 2), and get the unknown item set named unknown_info, that is, the information item that the target user does not disclose. Secondly, in Friends_Info, the unknown item info_i column in the unknown item set unknown_info is sequentially queried, and the existing attribute value v_j and the number of friends with the same attribute value num_v_j are recorded. We can get the attribute value v_j corresponding to the maximum num_v_j and use it to infer the attribute value infer_value. Finally, we calculate the confidence_degree of this inferred attribute value.

The attribute inference algorithm is described as follows:

Input: user_id of target user, confidence threshold R.

Output: the target user's inferred value infer_value for each unknown information item info_i.

1) Get a friend list F{f_1, f_2...f_m} of mutual concern relationship with the target user.

2) Obtain the profile information item of the friend in F and construct the information matrix Friends_Info.

3) Query each info_i and get the attribute value v_j with the most common attribute and num_v_j which is the number of friends who have the value v_j .

4) Calculate the confidence confidence_degree of the attribute value v_j by the confidence formula.

5) Determine whether v_j is used as the infer value infer_value of info_i, according to the confidence threshold R.

3.2.3 Confidence degree calculation

The correct rate of the inferred information item is affected by the number of friends concerned with each other and the consistency of the exposed information items. Therefore, it is necessary to calculate the confidence degree of the inferred information item. When the confidence degree is greater than a certain threshold, we consider this value is credible.

The confidence formula of confidence_degree to infer attribute value infer_value by the user unknown information items info_i is as shown in (1):

$$confidence\_\deg ree = \max\left(num\_v_j\right) \Big/ m \tag{1}$$

In the formula (1), num_v_j represents the number of friends having the attribute value v_j, and m represents the number of friends of mutual concern owned by the target user.

Different confidence thresholds can be set according to the actual application. After obtaining the inferred value and confidence degree of the target user's unknown information item, it is possible to determine whether the inferred attribute value is available according to a preset confidence threshold.

Let the information feature matrix inferred from social relations be B, then

$$B = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{bmatrix} b_{1,1} & \cdots & b_{1,k} \\ \vdots & \ddots & \vdots \\ b_{m,1} & \cdots & b_{m,k} \end{bmatrix}$$

Among them, m is the number of information items that can be inferred, vector $y_i = (b_{i,1}, b_{i,2} \cdots b_{i,k}\}$, $b_{i,1}=b_{i,2}=\cdots=b_{i,k}=0$ or 1, 0 represents the information item cannot be inferred through the social relationship, 1 means the information item can be inferred through the social relationship.

3.2.4 Information released in history

The emerging privacy breaches indicate that the aggregate effect of historical text and current text content will lead to more privacy leaks. Although the short current text to be published does not contain enough complete information to reveal privacy, the current text content combined with some historical postings is sufficient to reveal personal privacy information. This is the reason why the privacy leakages continue to climb. The existing methods of detecting privacy leaks generally lack the analysis and effective use of historical text information.

Therefore, this paper makes full use of historical text information to expand the scope of detection and conduct more comprehensive and in-depth privacy leak detection.

In the face of massive historical data, not all historical texts are available. Therefore, it is necessary to screen historical data. Personal Weibo data has strong timeliness and content is closely related to the release time. Obviously, the historical data closer to the current time point is easier to be seen and remembered by other users. What's more, the closer to the user's own state, the greater the impact on privacy leakage.

Through statistics and analysis of Weibo data, it is found that the average number of Weibo users sending Weibo is 0.5 per day. If the first 10 messages are selected, it can represent the user's situation for nearly 20 days. Therefore, in this paper, we select the first 10 historical information released by the user for detection. If the number of historical information released by the user is less than 10, all historical data of the user is taken for detection.

3.3 Privacy disclosure detection architecture

The overall architecture of privacy leak detection in this paper is shown in Fig. 1:

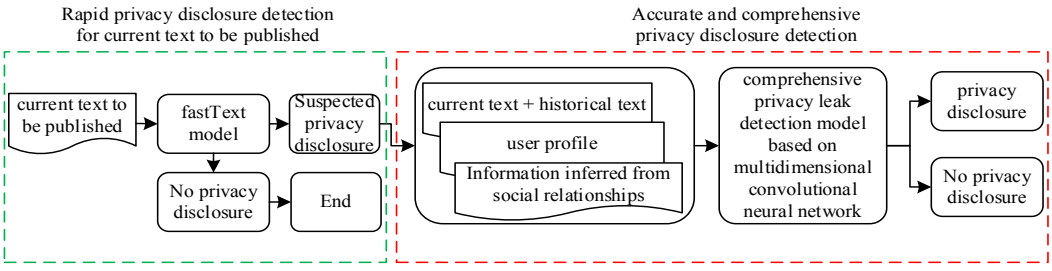

Figure 1: The privacy disclosure detection architecture

Firstly, we perform rapid privacy disclosure detection on the currently published text by building a text classification model based on fastText. Here, the issue of whether the current text contains a privacy disclosure can be regarded as a binary classification problem. If the classification result is no privacy disclosure, the detection is over. If the classification result contains private information, we will further make an accurate and comprehensive privacy disclosure detection synthesizing all aspects of information which includes current text, historical text, personal data, social relationships, and so on. In this part, we build a comprehensive privacy leak detection model of convolution neural network based on multidimensional features by combining all kinds of information. Among them, it is necessary to transform all aspects of information into the form of vector matrix as the input of the model. Finally, the classification model after training is used for privacy leakage detection to obtain the final result.

4 Privacy disclosure detection in social networks

4.1 Privacy disclosure detection for text to be published

Due to the wide range of social networking applications, social networks represented by Sina Weibo can publish thousands of messages per second on average. In the face of massive information release, it is important to improve the efficiency of privacy disclosure detection. This requires us to find a fast and effective privacy leakage detection method to detect the current text to be published on the premise of ensuring high accuracy.

fastText [Joulin, Grave, Bojanowski et al. (2016); Bojanowski, Grave, Joulin et al. (2016)], an open source rapid text classification model released by Facebook in 2016, provides a simple and efficient method for supervised text categorization and characterization learning, which has good applicability in Chinese text categorization. It uses n-grams to narrow the difference between the linear model and the deep learning model in accuracy, and can obtain the classification accuracy similar to deep learning. And it is much faster in training and evaluation than the deep learning classifier [Dai and Jiang (2018)], especially in large-scale, short text data [Wang (2018)]. Therefore, we implement a fast privacy classification of the current social text to be published based on the fastText model.

The architecture of the fastText privacy leakage detection model based on social network text content is shown in Fig. 2.

In Fig. 2, for the text information to be published on the social network, the preprocessing and word segmentation are first performed. The input layer of the model is the initial word vector of the word. Among them, fastText not only initializes the words after the word segmentation, but also adds the n-gram feature, taking the local word order into account and obtaining the word order information. The order of the words is different, and the semantics of the sentences may be completely different. For example, “I miss him”. If you do not consider the word order, the features are “I”, “miss”, “him”, and there is no way to distinguish it from “He misses me”. By adding 2-gram text features (i.e., adding “I miss” and “miss him”), the word order is considered, which makes the semantic expression more complete.

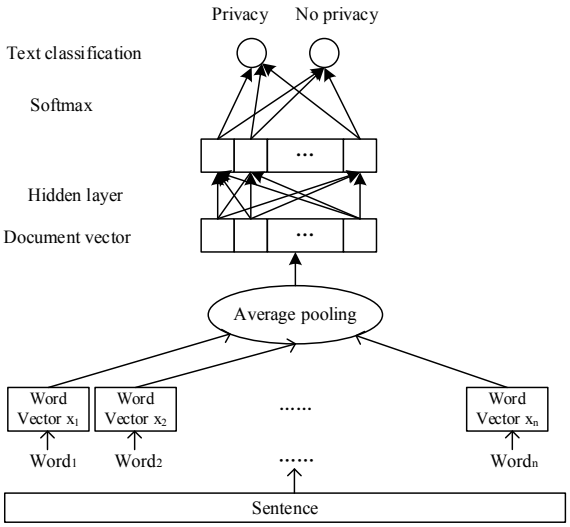

Figure 2: Current text privacy leakage detection model based on fastText

After vectorising the words in the sentence, the model calculates the document vector y of each sentence through the hidden layer. The calculation of y is obtained by averaging each word vector x_i in the sentence, as shown in Eq. (2):

$$y = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{2}$$

N is the number of words in the sentence and x_i is the word vector corresponding to each word.

The document vector y will be used as the input of the hidden layer, which is multiplied by the weight matrix B of the hidden linear layer to obtain the classification vector z, which is calculated as shown in Eq. (3):

$$z = \begin{pmatrix} z_1 \\ \vdots \\ z_m \end{pmatrix} = \begin{bmatrix} b_{1,1} & \cdots & b_{1,n} \\ \vdots & \ddots & \vdots \\ b_{m,1} & \cdots & b_{m,n} \end{bmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = B \cdot y \tag{3}$$

where y is an n-dimensional vector, z is an m-dimensional vector, B is a matrix of size m*n, and m is the number of categories of text classification.

Finally, the predicted label is obtained by using the softmax function to calculate the probability of the category to which the text belongs. The formula for calculating the class probability of a category is shown in Eq. (4):

$$p_j = \frac{e^{z_j}}{\sum_{k=1}^{m} e^{z_k}} \tag{4}$$

Among them, p_j is the class probability of text belonging to category j; z_j and z_k are the components of the classification vector z.

4.2 Comprehensive privacy disclosure detection

In order to effectively mine and express private information, this paper constructs a privacy leak detection model (MF-CNN) of convolutional neural network based on multi-dimensional feature. This method processes the user's current text to be published, recent historical texts, profile information, and social relationship information and transform them into vector form. What's more, we act them together in the input layer of the convolutional neural network. In this way, we can study the characteristic information of the privacy leakage more comprehensively, so as to detect user's privacy leakage in all direction and effectively.

Compared with the LSTM network, this method can receive multi-dimensional information of parallel input, which greatly reduces the training time of the network model. At the same time, the method effectively compensates the deficiency of only relying on the content of the text to be published by combining the multi-dimensional information, so that the model can obtain more comprehensive privacy feature information and effectively detect privacy leakage of users.

A multi-dimensional feature-based convolutional neural network privacy leak detection model (MF-CNN) is shown in Fig. 3.

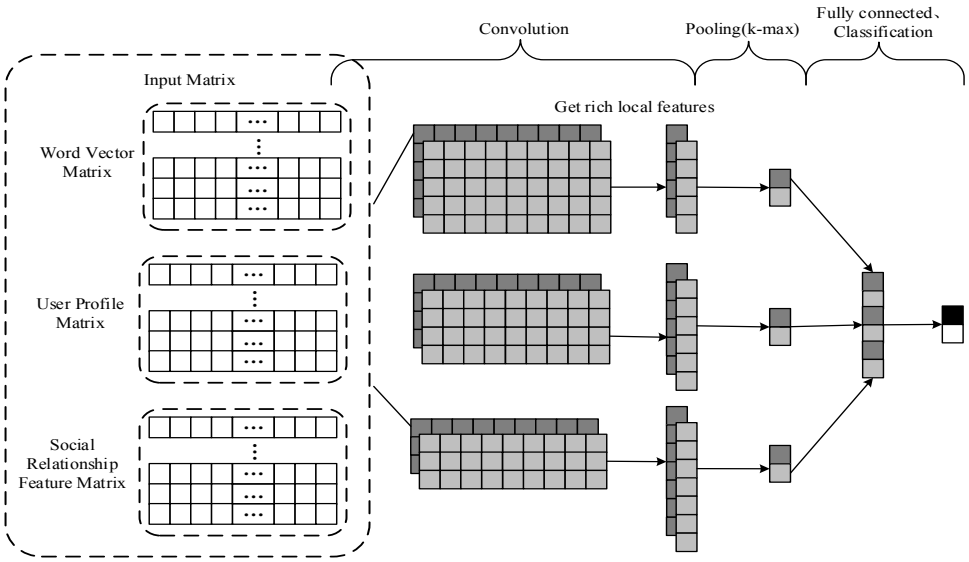

Figure 3: Privacy disclosure detection model (MF-CNN) based on multidimensional features

As can be seen from Fig. 3, the input matrix of the model MF-CNN is composed of three parts of information matrices.

(1) Word Vector Matrix. This part corresponds to the user's text to be published and the user history Weibo text information obtained in Section 3.2.4. Firstly, the preprocessing and word segmentation are performed, and the order of word is mapped to word vectors by word vector model, so the text content is converted into a corresponding word vector matrix. For example, if the text contains n words and k is the dimension of the word

vector, the input data of the model corresponds to a two-dimensional data of n×k.

(2) User Profile Matrix. That is, the profile information obtained in Section 3.2.1.

(3) Social Relationship Feature Matrix. That is, the information feature matrix inferred from social relations obtained in Section 3.2.2.

Then, the MF-CNN model performs a convolution operation on the vector matrix of the input layer. In the convolution operation, the model can automatically obtain abundant local feature vectors, including not only the user's text content, but also the user's profile characteristics and social relationship features of the user. In the process of the model training, we can integrate three aspects of information and learn a series of rules of privacy leakage.

In this paper, the convolution operation of the input feature matrix is performed by using a convolution kernel of size h×d, that is:

$$C_i = f(w \cdot X_{i:i+h-1} + b) \tag{5}$$

Among them, C_i represents the i-th feature value in the feature graph, f() is the activation function, $w \in R^{hd}$ is the filter, h is the size of the sliding window, and b is the bias term. $X_{(i:i+h-1)}$ represents the local feature matrix composed of the rows from i to i+h-1. Therefore, feature map C is:

$$C = [C_1, C_2, C_3, \ldots C_{n-h+1}] \tag{6}$$

After the convolution, it is the pooling operation, that is, dimension reduction. The pooling operation is to select the most important features that contribute to the final classification.

With the output vectors from each convolution kernel, we merge them into a long feature vector. In order to make the network more anti-interference and prevent overfitting, Dropout mechanism is used in this paper to improve the generalization ability of the model by randomly discarding some intermediate calculation results during the model training process.

Finally, the probability distributions of different categories are obtained through the softmax layer. In the process of model training, we use cross entropy loss as the supervision information, and introduce L2 regularization on this basis. It can not only avoid overfitting effectively, but also make our optimization solution stable and fast.

$$L_{loss} = L_{cross\_entropy} + \lambda_{12} * L_{l2} \tag{7}$$

Among them, $L_{cross\_entropy}$ represents the cross-entropy loss, L_{l2} represents the L2 regular loss, λ_{l2} is a custom parameter. The calculation formula is shown in (8) and (9).

$$L_{cross\_entropy}(x, y; \theta) = -\sum_{i=1}^{K} y_i \ln P(C_j|x) \tag{8}$$

$$L_{l2} = \frac{\lambda}{2n} \sum_{W} W^2 \tag{9}$$

5 Experiment

5.1 Experimental data and experimental environment

We used the Sina Weibo dataset purchased on the Datatang and the dataset includes the user table and Weibo table, which contains 2,649,567 pieces of Weibo data. In order to carry out related experiments and analysis on the proposed method, 33,920 pieces of microblogs data with privacy leakage from these microblog datasets are manually annotated, including 25,951 users, and also labeled 53880 microblog data that without privacy disclosure. In the experiment, the problem of data imbalance is solved by slightly up-sampling the privacy microblog data.

For Weibo content, due to its complexity and variety, it may include emoji, links, pictures, and @symbols. Therefore, these special representations must be replaced before the text is segmented, otherwise it will have a greater impact on the final classification results. Since the emoji can clearly reflect a user's emotional state, it must be identified and processed; similarly, the @symbol will reveal a user's social information to a certain extent, so it also needs to be identified and converted. We do not consider image in this article because the semantics of the image are difficult to identify and understand. At the same time, we do not consider the links and regard them as useless information.

The pretreatment results are shown in Tab. 4:

Table 4: Microblogs preprocessing

Before processing	After processing
emoticons	[expression name]
@username	[friends]
pictures	*img*
links	*URL*

The experimental environment of this paper is shown in Tab. 5.

Table 5: Experimental environment

Server	Sugon Xmachine W780-G20
CPU	2* Intel E5-2650v4 2.2G 9.6QPI 30M 12C 105W
RAM	8* 32G DDR4 2133/2400 ECC REG
GPU	8* NV TESLA P100 16GB -E3x16 250W
Operating System	linux-ubuntu16.04
Development language	Python3
Development environment	PyCharm

5.2 Experiment of privacy disclosure detection of text to be published

5.2.1 Experimental scheme

The current text privacy leakage detection based on the fastText model not only considers

the semantics of the text, but also adds n-gram information to the network model, and finally obtains the category probability through supervised training. The specific experimental steps are as follows:

(1) Preprocessing of special symbols: Loading the original training set text train.txt to perform the replacement processing of the special symbols in microblog text;

(2) Word segmentation and removal of stop words: Using jieba word segmentation module to segment words, loading the stop word file stopwords.txt into the stop words dictionary, and removing the stop words in the text according to the stop words dictionary;

(3) Format processing of training data: After the previous two steps, the data is processed according to the training format of the fastText model, so that each piece of data is one line, and each line starts with the label label_XXX corresponding to the text, and end with the text content;

(4) Training and optimization: Getting training text data set from Step (3), and the supervised function is used to train and optimize the model;

(5) Optimizing the parameters of the model during the training process, and using the 10-fold cross-validation to obtain the accuracy, recall and F1 values of the model. Finally, selecting the best combination of parameters by comparing the effects of the model.

5.2.2 Experimental results and analysis

In order to select the best word vector dimension, we selected 6 dimensions (50, 100, 150, 200, 250, 300) to perform a 10-fold crossover experiment (that is, divide the data into 10 parts, For each experiment, take 1 of them as a test set and the remaining as the training set) . The changes of the F1 value of the fastText model in each cross-validation is observed under different word vector dimensions. The experimental results are shown in Fig. 4.

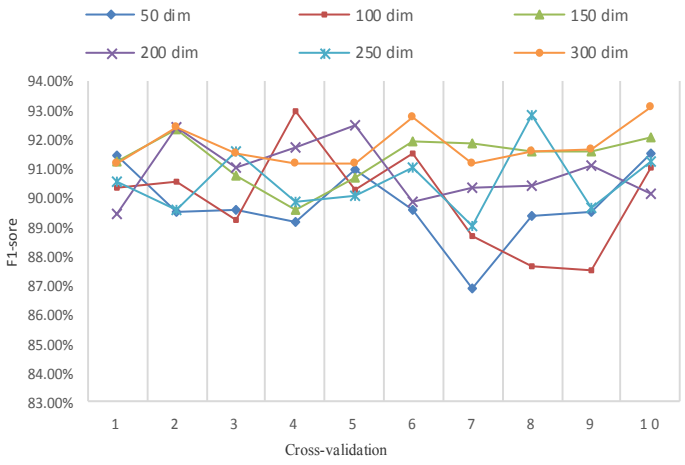

Figure 4: Cross-validation of word vector dimensions of fastText model

As can be seen from Fig. 4, when the word vector dimension is 300, the fastText model has an average optimal F1 value. Therefore, the word vector dimension we choose is 300.

Because fastText adds n-gram to get local word order information, it improve the accuracy of classification. In order to select the best n-gram value, we experimented on different n-gram values under the same data set, and obtained the privacy text content detection effect (accuracy, recall rate, F1 value) under different n-grams. The result is shown in Fig. 5.

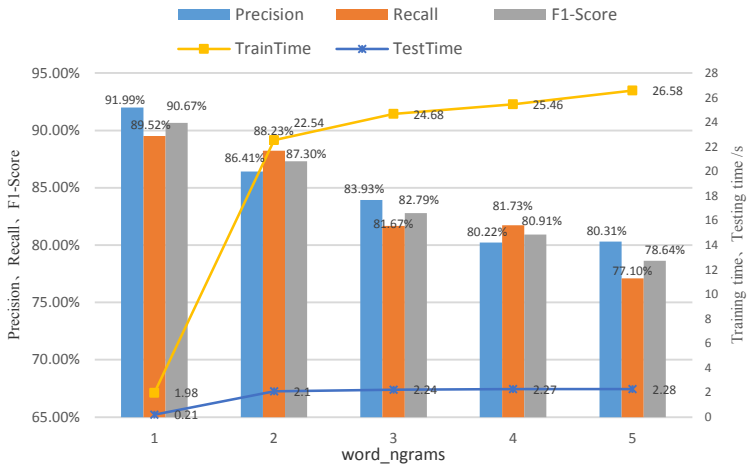

Figure 5: Cross-validation of word vector dimensions of fastText model

It can be seen from Fig. 5 that the model has the best effect when the value of n-gram is 1, and the accuracy, recall rate and F1 value decrease as the n-gram value increases. In addition, the training time and test time of the model are also increasing. Therefore, when the n-gram value is 1, it works best in the privacy leakage detection in this paper.

In order to verify the effect of fastText on text privacy leakage detection, this paper compares it with traditional machine learning methods and deep learning methods on the same data set. We used 10-fold cross-validation method, and used the average precision, recall, F1-Score as the evaluation criteria. The experimental results are shown in Tab. 6, and the time used in the experiment is shown in Tab. 7.

As can be seen from Tab. 6, fastText is significantly better than the traditional machine learning methods in the performance of text privacy leakage detection, and slightly higher than CNN, but slightly inferior to RNN, LSTM.

As can be seen from Tab. 7, fastText has obvious advantages in training time and test time, so it can be concluded that in the context of massive social network data, using fastText for text privacy leakage detection can greatly shorten the detection time and can meet the needs of real-time detection very well.

Table 6: Model comparison experiment

	P/%	R/%	F1/%
Naive Bayes	64.39	73.82	68.78
SVM	95.21	35.98	52.22
Logistic regression	85.23	83.27	84.24
C4.5	91.42	76.37	83.22
fastText	91.99	89.52	90.67
CNN	84.96	96.64	90.42
RNN	88.99	93.85	91.36
LSTM	89.57	95.48	92.43

Table 7: Model training time and test time

	Train/s	Test/s
fastText	1.98	0.21
CNN	6454.43	2.99
RNN	37596.51	4.90
LSTM	57373.17	5.81

5.3 Comprehensive privacy disclosure detection (MF-CNN) experiment

5.3.1 Experiment and model parameter settings

On the Sina Weibo dataset as described above, we obtain the user's personal data, social relationships, and historical data for each text, and organize them into new training sets and test sets.

The programming language used in the comprehensive privacy leakage detection experiment is Python3.6, the toolkit is TensorFlow, an open source deep learning framework of Google. Other parameter settings in the network model are shown in Tab. 8.

Table 8: Parameter Settings in MF-CNN

Adjustable parameters	**value**
Convolution kernel function	ReLu funtion
Filter sliding window size h	3,4,5
Number of filters	128
Optimizer	AdagradOptimizer
learning rate	Exponential attenuation method
batch size	100
dropout probability	0.1
training iterations	1000

5.3.2 Experimental results and analysis

In the first group of experiments, in order to verify the effectiveness of MF-CNN model based on multi-dimensional features in privacy leakage detection, we compare the performance of MF-CNN model with that existing CNN model based on single text content. Under the same experimental conditions, the experimental results are shown in Tab. 9.

Table 9: Model comparison experiment

	P/%	R/%	F1/%
CNN	84.96	96.64	90.42
MF-CNN	96.44	91.07	93.68

As can be seen from Tab. 9, the MF-CNN model combines the user's personal data, social relationships, and historical data, which enables a more comprehensive privacy leakage detection, effectively improving the privacy leakage detection effect in social networks.

In the second group of experiments, based on the multi-dimensional feature method, different models are used for training. The experimental results are shown in Fig. 6.

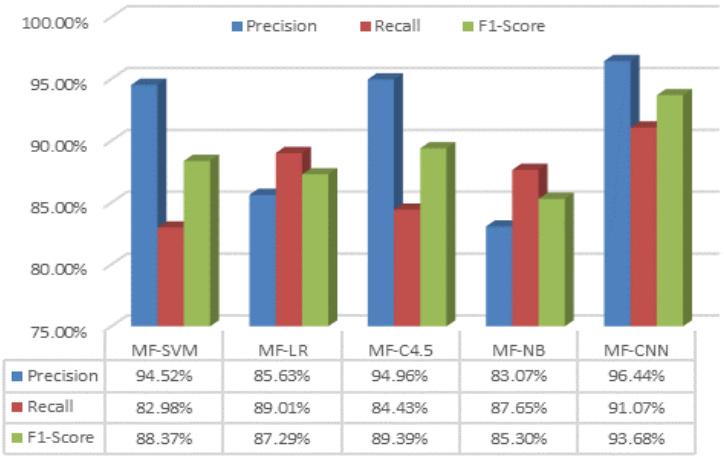

Figure 6: Model comparison based on multidimensional features

It can be seen from Fig. 6 that in the case of multidimensional feature information, the method MF-CNN has better effect and effectively improves the privacy leakage detection effect in the social network. It can also be seen from the comparison between Fig. 6 and Tab. 6 that the multi-dimensional information-based detection method is superior to the method of relying only on the current text for detection in each model, and the effectiveness of the method based on multi-dimensional features is illustrated again.

6 Conclusion

In view of the characteristics of social network with huge user groups and massive information publishing, this paper firstly implements fast privacy leakage detection for the text to be published based on fastText model, avoiding comprehensive and accurate

detection in the case of text without privacy disclosure, in order to improve the pertinence and rapidity of detection.

In the case that the text to be published contains certain privacy information, this paper comprehensively and deeply analyzes the various possible ways of privacy leakage for the characteristics of Weibo. By considering the privacy information that may be leaked in the current texts to be published, historical texts, user profiles and social relationships, a multi-dimensional feature-based convolutional neural network model is established, and a comprehensive privacy leak detection is performed based on this model.

The comparison experiment results show that the multi-dimensional comprehensive privacy leakage detection method has higher accuracy than the existing research that only rely on the text content to be published for privacy leakage detection.

At the same time, needn't to first classify the privacy of the content, but integrate the private information of different ways, and then directly uses the convolutional neural network to receive the parallelized input multi-dimensional information, to obtains deep privacy leakage information, and to conduct comprehensive privacy leak detection, so it effectively improves the efficiency and accuracy of privacy leakage detection.

Acknowledgement: This work was supported by the National Natural Science Foundation of China (No. 61672101), the Beijing Key Laboratory of Internet Culture and Digital Dissemination Research (ICDDXN004)* and Key Lab of Information Network Security, Ministry of Public Security, China (No. C18601).

References

Belanger, F.; Crossler, R. E. (2011): Privacy in the digital age: a review of information privacy research in information systems. *Society for Information Management and the Management Information Systems Research Center*, vol. 35, no. 4, pp. 1017-1042.

Bojanowski, P.; Grave, E.; Joulin, A.; Mikolov, T. (2016): Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135-146.

Clarke, R. (1999): Internet privacy concerns confirm the case for intervention. *Communications of the ACM*, vol. 42, no. 2, pp. 60-67.

Dai, L. L.; Jiang, K. (2018): Chinese text classification based on fastText. *Computer and Modernization*, no. 5, pp. 35-40.

Hou, M. W.; Wei, R.; Wang, T. G.; Cheng, Y.; Qian, B. Y. (2018): Reliable medical recommendation based on privacy-preserving collaborative filtering. *Computers, Materials & Continua*, vol. 56, no. 1, pp. 137-149.

Islam, A. C.; Walsh, J.; Greenstadt, R. (2014): Privacy detective: detecting private information and collective privacy behavior in a large social network. *Workshop on Privacy in the Electronic Society*, pp. 35-46.

Ji, X.; Xu, Y. B. (2015): Detection method of privacy content for judgment documents. *Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition)*, vol. 27, no. 5, pp. 639-646.

Jiang, Z. S. (2013): *Research on Microblogging Privacy Detection Based on Bayesian (Ph.D. Thesis)*. Harbin Engineering University, Heilongjiang, China.

Jiao, Y. Q. (2017): *Detection of Bad Information and Privacy Disclosure in Social Network (Ph.D. Thesis)*. Beijing Information Science and Technology University, Beijing, China.

Joulin, A.; Grave, E.; Bojanowski, P.; Mikolov, T. (2016): Bag of tricks for efficient text classification. arXiv: 1607.01759.

Kim, C.; Jung, K.; Park, S. (2013): A method of predetecting privacy leak in social network service using collaborative filtering. *International Conference on Database Systems for Advanced Applications. Springer*, pp. 153-163.

Liu, K.; Terzi, E. (2009): A framework for computing the privacy scores of users in online social networks. *IEEE International Conference on Data Mining*, pp. 1-30.

Machida, S.; Shimada, S.; Ecizen, I. (2013): Settings of access control by detecting privacy leaks in SNS. *International Conference on Signal-Image Technology & Internet-Based Systems*, pp. 660-666.

Mao, H.; Shuai, X.; Kapadia, A. (2011): Loose tweets: an analysis of privacy leaks on twitter. *ACM Workshop on Privacy in the Electronic Society*, pp. 1-12.

Qiu, J. P. (2012): A study of privacy security issues of social network users. *Information & Documentation Services*, vol. 33, no. 6, pp. 34-38.

Srivastava, A.; Geethakumari, G. (2013): Measuring privacy leaks in online social networks. *International Conference on Advances in Computing*, pp. 2095-2100.

Tesfay, W. B.; Serna-Olvera, J. (2016): Towards user-centered privacy risk detection and quantification framework. *IFIP International Conference on New Technologies, Mobility and Security*, pp. 1-5.

Wang, Y. J. (2018): Implementation of classification of prevention and control targets based on fasttext. *China Public Security (Academy Edition)*, no. 1, pp. 29-32.

Warren, S.; Brandeis, L. (1890): The right to privacy. *Harvard Law Review*, vol. 4, no. 5, pp. 193-220.

Yao, K. (2014): *Design and Implementation of Chinese Weibo Privacy Mining System (Ph.D. Thesis)*. Xidian University, Shanxi, China.

Zhang, L. (2016): *Research on Privacy Disclosure Detection Method for Microblog Content (Ph.D. Thesis)*. Beijing Information Science and Technology University, Beijing, China.

Zhao, H. M. (2002): On the legal protection of online privacy rights. *Journal of Peking University*, pp. 165-171.