**Related Work**

This section mainly describes the work related to sensitive information detection.

In 2011, Mao [1] proposed three types of tweets that could leak privacy and are worthy of attention. Vacation tweets, drunk tweets and illness tweets. Mao's team points out that holiday tweets can lead to burglaries, drunk tweets can send law enforcement a drink-driving alert, and sick tweets may alert insurance companies. The detection methods of the three types of tweets in this paper are very similar, so this paper only introduces the author's analysis method for the holiday tweets. Mao randomly selected 600 pieces of data from available tweets and labelled them either "sensitive" or "non-sensitive." Use the NER tool to automatically detect the person, location, and time information in tweets. In addition, some keywords related to classification were selected to judge the sensitivity. Add keywords like "beach," "hotel," and "flight" when checking vacation tweets. And some words denoting virtual and negative as indicators of non-sensitive information. After classifying, obtain 300 sensitive data and 300 non-sensitive data. 80% of the data were used as training sets, and naive Bayes and SVM classifiers were used for training respectively. In naive Bayes, the accuracy of real holiday tweets was 76%. Through the above methods, the author accurately analyzed the sensitive information of holidays, drunkenness and illness. But the classification scope involved was relatively small. In 2014, Islam [2] divided tweets into 200 topics. The detection probability of privacy information is greatly increased by pre-defining the theme of the content published on social networks and detecting it according to the related features of the theme. The author also found that naive Bayes had a better performance in classification.

Both authors' studies, however, were limited to identifying leaks of sensitive information from published tweets. In the context of increasingly serious privacy issues and people's increasing attention to privacy, it reflects the importance of real-time monitoring of sensitive information. In 2017, Cappellari [3] and the team built a privacy decision tool to alert users of potential privacy disclosure risks before they posted messages to social platforms. Cappellari collected 500 samples and labelled them, and used Wordnet to preprocess labelled tweets. They used 80% of the processed tweets as a training set and 20% as a test set. After using the nearest neighbour, rule induction, random forest, naive Bayes and SVM methods, the SVM obtained the highest accuracy. In the same year, Neerbeky [4] developed a real-time privacy detection desktop application.

Today's social networks are all based on the Web or mobile end. If it is a privacy detection system for social platforms, it should be embedded in the Web end to achieve better results. But few people are integrating real-time monitoring of private information into the Web. This paper will focus on the implementation of sensitive information detection in the Web site of workplace social platforms. The data of choice were tweets containing workplace keywords, including work, job, colleague, workmate, boss, salary, wage, overtime, and a host of other privacy terms. As the research direction of this paper is highly targeted and only one topic is studied, the selection of keywords is small and accurate. For the same number of tweets, using keywords yields more tweets available. After manual annotation of the acquired data , preprocess these data. The method of selecting sensitive information for machine learning is to use the Bayes model with high accuracy.

References

1. Mao H, Shuai X, Kapadia A. Loose tweets: an analysis of privacy leaks on twitter. Proceedings of the 10th annual ACM workshop on Privacy in the electronic society; Chicago, Illinois, USA: Association for Computing Machinery; 2011. p. 1–12.

2. Islam AC, Walsh J, Greenstadt R. Privacy Detective: Detecting Private Information and Collective Privacy Behavior in a Large Social Network. Proceedings of the 13th Workshop on Privacy in the Electronic Society; Scottsdale, Arizona, USA: Association for Computing Machinery; 2014. p. 35–46.

3. Cappellari P, Chun SA, Perelman M. A Tool for Automatic Assessment and Awareness of Privacy Disclosure. Proceedings of the 18th Annual International Conference on Digital Government Research; Staten Island, NY, USA: Association for Computing Machinery; 2017. p. 586–7.

4. Neerbeky J, Assentz I, Dolog P, editors. TABOO: Detecting Unstructured Sensitive Information Using Recursive Neural Networks. 2017 IEEE 33rd International Conference on Data Engineering (ICDE); 2017 19-22 April 2017.