

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSI*net*

Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Research on Adaptive Multi-filtering Model of Network Sensitive Information

Xiao-Feng Cao, Wei Kang, Quan Shi and Fang-fang Shi
School of Computer Science and technology, Nantong University, Nantong, 226019, China

Abstract: Sensitive information (i.e., negative hot topics, negative incident, bad information etc.) filter is the information filtering that is important and very difficult task. Aim at the problems of detection time lag, low accuracy and Poor adaptability etc in internet sensitive information detecting, This project, with the Chinese text media (web page, blog, BBS *et al*) in internet as the research object,using technologies of opinion mining,machine learning, High performance computing and natural language processing etc, research on adaptive multi-filtering model of network sensitive information, to adaptive identify sensitive information from the overall and semantic view. The research results will provide technical support for public opinion monitoring, business intelligence and aid making decision application system development.

Key words: Sensitive information, sensitive word, information filtering, self-adaption

INTRODUCTION

Sensitive information's structure is complex, diverse, fast time-varying, has obviously large data's characteristics. It can come through many carriers, such as sensitive images, sensitive sound, video and text. The sensitive information in this article only refers to sensitive text.

From the content point of view, the sensitive information means the misuse of sensitive information or unauthorized use of personal information, if without modification, would be harmful to the national interest or the implementation of government programs or violate the law of personal privacy (Sensitive). From the expression it includes negative hot topics or unexpected negative events or negative information.

Sensitive information is different from a hot topic, because hot topic is divided into positive, neutral, negative hot topic; sensitive information differs from harmful information, because harmful information is with aggressive means of hackers, viruses, etc. sensitive information differs from the information of interest, because interest information is the user's online behavior ; sensitive information differs from spam, spam is useless because the information is not necessarily harmful interference and bad information ; sensitive information is different from the sensitive word, sensitive word is the word prone to sensitive information, such as " Falun Gong " is a sensitive word but not necessarily sensitive information, because you can be on "Falungong" criticism, on the other hand, sensitive words are pre-constructed static dictionary and sensitive information is dynamic.

Filter sensitive information has become a very important component in security system. So, the development of sensitive information filtering is the basis to build a network public opinion monitoring system,which has very important academic significance and practical value.

Due to the sensitive information filtering has very big potential value,many scholars have started the related researches. Back to the late 1950s, Luhn proposed the selective release of information-concept (Greevy and Smeaton, 2004). designed text classification method using general preliminary study of the detection and discovery of racist propaganda on the Internet-related information. But the similarity of text extraction and text classification model is highly dependent on the characteristics of the calculation (Zhou *et al.*, 2005). studied the structure and content of the information extremist propaganda sites, the use of statistical methods Multidimensional Scaling (MDS) conducted a sensitivity analysis of the content and then digging through intersite topology and then analyzes the sensitive site distribution but the study relies to a large extent the properties of some manual calibration data, limiting the range of applications. Company BBN developed probabilistic model applied to their topic tracking and identification systems, the system combines several classifiers,put the results of the individual classifiers together to make the judgment. Stanford University professor TAKW (Bermingham and Smeaton, 2010). AYN created SIFT filtration system. Using the matching rule, comparing the daily news ranking list to source template and the demand information of the users and then give the desired information to the users according to user features and then give feedback (SIFTsystem).

ADAPTIVE MULTISENSITIVE INFORMATION FILTERING FRAMEWORK

Adaptive filter model underlayer: The underlayer of the filter is to identify sensitive information (including: Sensitive word, hot topic, emergencies, sensitive information feature), laying the foundation for the next level filter.

Under normal circumstances, the user wants to achieve zero mistake rate which would give up some of filtering accuracy, rather than filter out the normal error page. Therefore, the underlayer filter, the filter should have enough but not strong too much to avoid filter out normal page, to achieve closing zero mistaken result.

The main task of the underlayer of the filter:

- Model in conjunction with the user heat hot topic find sensitive words

- Model to identify sensitive words from emergencies events
- Model of identify meaning of sensitive words

Model of finding sensitive words through the heat topic of

Therese: The model of current hot topic finder is based on the hot level of media and press which according to the number of articles from media reports to determine the heat of the event. Since the report is based on the reaction of the masses and the state to carry out, this led to sensitive information filtering lag behind. From a sociological perspective, public behavior choices and common concerns of the hot topics have an important role in promoting the sensitive words. Therefore, the combination of media and user attention to hot issue is the key to find the hot point. Users inner psychological state to an certain extent are determined by the heat level of the user.

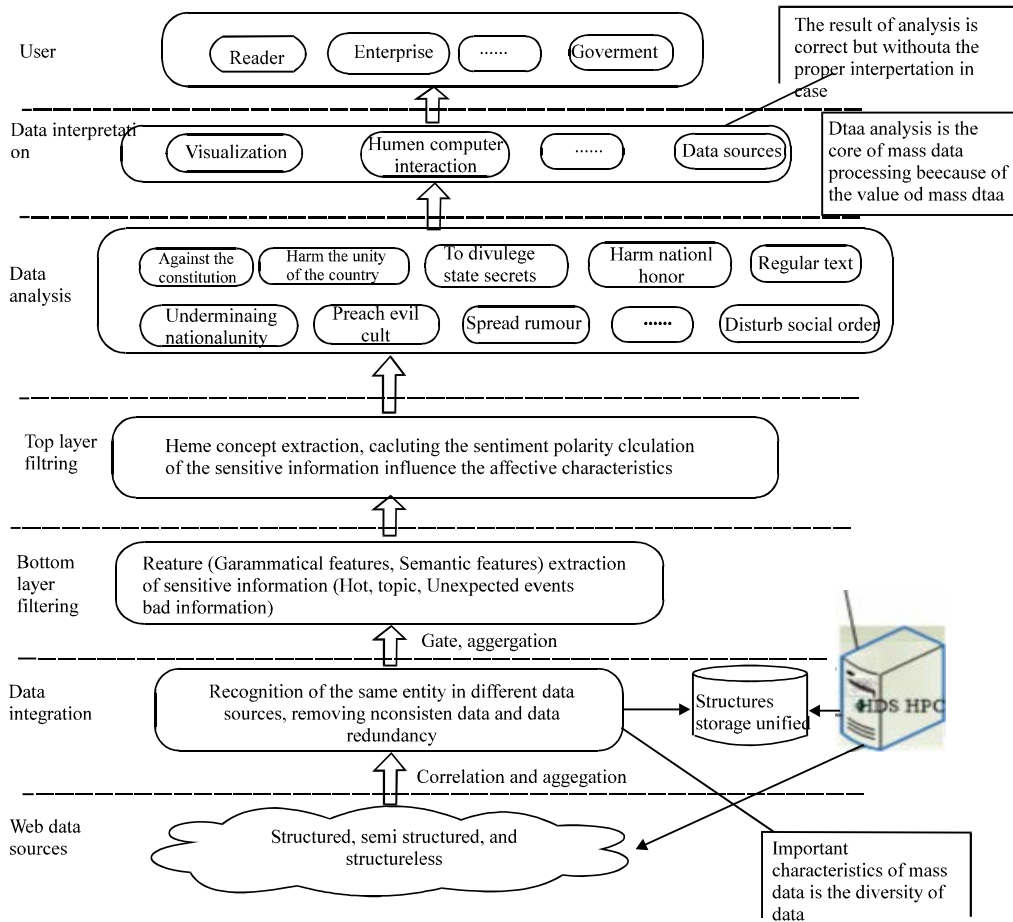


Fig. 1 Adaptive filter model of multiple sensitive information

Users heat level is calculated as:

$$Hot_v(j, t) = (\text{occurrence}(t, d) \times \text{postCount}(t, d) \times \text{place}(t, d) \times \text{length}(t) + I_r / (I_r + I_c)) / AR_s$$

where, ARs is related texts average reading speed (words/sec), I_r as hits, I_c as Replies, occurrence (t, d) is characterized by the number of occurrences in the post entry t d in ; postCount (t, d) d t is the number of articles contained; place (t, d) t appear in the position corresponding to d, the title corresponding to the word appeared value>appeared in the main post words corresponding value>appears only replies word corresponding value over; length (t, d) corresponds to the word length, word length longer the greater the value .

So the model of hot topic sensitive assessment is:

$$\text{Topic}(j, t) = w_1 \text{Hot}_M(j, t) + w_2 \text{Hot}_v(j, t)$$

w_1, w_2 are influencing factors.

Unexpected Incident and emergencies sensitive temporal model of word recognition: Media heat is not necessarily to be high to incident, such as in a school " exam cheating expelled," " dormitory turn off lights ", " cafeteria prices increase", " campus network charges " events with the following differences: The information of this kind of incidents is more accurate and comprehensive than the general information in time, place, person, reason, events, etc. In most time the time, place and person appeal in title, this we call " The title heat " Emergencies has the limit of its location and time and the overall heat may not be particularly high but special location and time it can be very high, in this article, we call this instantaneous frequency of occurrence of such an IP address" time and space heat."

By event extraction model, we can get all the events that may occur within a certain period of time selected by the user but for the convenience of users to quickly and accurately grasp the most critical and important information, we need to get in some way to sort this events, therefor obtaining emergencies.

Definition of time and space heat is:

Heat of title = matching degree of characters, location, time with sensitive information

Time and space heat = Number of occurrence nearby IPaddress

Therefore, sensitive word of emergency assessment model is:

Event = wk × title heat × time and space heat.

Model of semantic recognition of sensitive word: The sensitive words have a very large amount and it changing very quickly, the traditional dictionary is not capable of matching this job! With time going, the new sensitive words will continue to appear, in order to cope with this dynamic trend, machine learning is the most effective method to solve this problem. No matter how sensitive words changes, no matter how innovative, semantic templates are relatively stable, with high-performance computers, the use of semi-supervised learning algorithm weakening combine semantic ontology reasoning to identify the word change the template for the sensitive part of the word, as is it is sensitive information at the bottom of the filter does not need to consider .

If the definition of sensitive words is in the dictionary, the word heat of sensitive words is calculated by matching degree.

After calculating the heat of sensitive words, we usually want to select some similar words to this sensitive word in the text to be analyzed, these similar words can be in anywhere in the text we are going to analyze and the same word, meaning not necessarily the same but different words can also mean the same. In order to be accurate, comprehensive, high-quality filtration to achieve results, sensitive information must be able to model the underlayer filtering synonymous relationship effectively determine which filtration system can make the necessary inference, based on keyword semantic relationship to reasoning based, combined with a certain database information, conduct discourse analysis and reasoning, in order to identify similar words and sensitive words .

Model of top lever of adaptive filtering: After the filtration from underlayer, sensitive words are marked, due to the sensitive word is only one feature of sensitive information it could be not sensitive information, so the top-level study of adaptive filtering sensitive information model is designed as a whole recognition and semantic perspective polarity sensitive words . It mainly using opinion mining, semi-supervised learning and ontology reasoning techniques.

Opinion mining technology is a technology branch combined with the practical Web data mining technology, the point of the study is to determine emotional tendencies of documents, sentences, words, the opinion digging is also called sentiment classification. Opinion mining "opinions" consists of four elements: The theme, the holder, comments, emotions (Zeitsoff, 2011) .

Since each comment includes the information of the opinion holder, so this study omitted the identification of these persons. For statement selection, we selected a

brief way in the process to resolve the problem. Sensitive information to build an adaptive filter model of top-level main task is to excavate rule : < sensitive word, polarity >. Extracting program of the theme is the same as hot topic finder.

Theme concept extraction: Top-level model based on adaptive filter choose candidates sensitive words and text, without further processing, then the selected range becomes too broad, although it contains a lot of text candidate filtered sensitive word but its appear sensitive words may be sensitive to criticism, such as "spreading pornography ban," first appeared in the sentence "pornography" Obviously, this is a great concern of sensitive words, the text message after sensitive information from adapt the underlayer filter it will be temporarily retained as a candidate filter text, however and it will soon be shielded if it so. So judged by sensitive words then conclude its sensitive information is not correct.

How to summarize the central meaning of the filtered candidate text, is the basis to determine the emotional polarity of the text, each candidate text has its own peculiarities, according to different criteria, can be divided into several types, such as single theme and multiple themes, explicit and implicit theme and so on. In practice, a candidate filtering text is usually not restricted to a division of a type but you can put it into several different types of division . In view of the above, during the theme concept extraction, conventional methods often appear the following conditions:

- Too little of refining the concept theme
- Too much refining the concept theme
- Theme concept refining error

Since each of the review included the information of the opinion holder, so this study omitted the identification of the opinion holder. For statement of selective question, after confirm the sensitive word, take all sentences which include sensitive word as candidate emotional sentences, each word of the sentence S contains classified to the corresponding theme concept, declarative vector to correspond to S (T1, W1; T2, W2; ... Tn, Wn), wherein Ti is contained in the thematic sentences concept, Wi Ti corresponding to the frequency. In order to determine the final statement from the candidate's emotional declarative text, you need to assess the importance of the sentence by the calculation.

Importance sentence S is calculated:

$$I(S) = P_s \frac{\sum_{i=1}^n W_i}{n} \quad (1)$$

Ps is the weighting coefficient of sentence S. Different sentence has different weighting coefficient in different locations. Generally, sentence in first paragraph is more important than the sentence in last paragraph.

According to the above equation, we can calculate the degree of importance of the one concept, the maximum value candidate of important emotional declarative set Q.

To extract the emotional topic sentence from the topic sentence in the text set Q, we use CRFs model . In the training model, based on the characteristics of Chinese emotional sentences, select two types of characteristics: Emotional tendencies characteristic features and transfer word. In the CRF model training feature selection, the main features of the model selected two types of training. One is emotional tendencies characteristic features of the observation point y on the choice of words with emotional tendencies as a feature . The words expressed emotional tendencies are mainly adjectives, adverbs and some verbs.

Polarity word extraction: Polarity word contains words expressed by emotional tendencies. There are three types of words are usually polar emotional tendencies which include compliment, neutral and derogatory. Opinion mining mostly concern which containing positive emotions (commendatory) and contain negative emotions (derogatory) polarity of words. For neutral words, you cannot consider.

October 2007, HowNet issued a " sentiment analysis using word set (beta version) ", contains English words with six sets of sentiment analysis which are positive emotion words, negative emotion words, positive evaluation words, negative evaluation words advocates words and degree level terms. This paper selects Chinese dictionary words to build polarity. Polar dictionary includes three parts:

- Evaluation polarity dictionary word, usually an adjective
- Negative word dictionary
- degree adverb dictionary

In this study, in this article we only extract only polar context, constitute < left polarity word, bad information, the right polarity word >, polar word is empty also taking one polarity placeholder .

THE EXPERIMENTAL RESULTS

Data set: The actual test data classified as sensitive information corpus in Table 1 (using cloud spider web information gathering cloud services platform acquisition). When the actual tests, arbitrary selection of which 2/3 as the training set, 1 / 3 as the test set.

Table 1: Classification corpus of sensitive information

Classification	Candidate sensitive text								Normal Text
	Falun gong		Fishing Island		Independence of taiwan		South China Sea		
	n	s	n	s	n	s	n	s	
No. of doc	58	30	120	60	51	42	166	28	100

Table 2: Experimental results based on the characteristics of the sensitive information

	Falun gong		Fishing Island		Taiwan Independent		South China Sea		n
	n	s	n	s	n	s	n	s	
	P	61.2	78.6	76.2	77.5	63.0	83.3	54.6	
R	82.2	88.4	78.3	77.9	87.7	90.2	76.6	79.1	15.3
E	3.4	0.8	2.1	1.8	1.3	0.1	2.7	0.7	0.3
F1	70.2	84.2	77.2	77.7	73.3	86.6	63.7	70.4	19.7

Table 3: Experimental results based on sensitive word (Where n is Normal Text, s is sensitive Text)

	Falun gong		Fishing Island		Taiwan Independent		South China Sea		n
	n	s	n	s	n	s	n	s	
	P	98.3	79.9	77.4	87.5	83.5	93.3	84.4	
R	92.4	81.1	88.1	85.7	82.2	91.4	90.6	89.7	100
E	21.2	45.2	68.1	54.6	33.3	44.2	31.3	38.4	0.0
F1	95.2	80.5	82.4	86.6	82.8	92.3	87.4	87.0	99.9

Where n is Normal Text, s is sensitive Text.

Evaluation criteria: Evaluation criteria to sensitive information filtering system include two aspects, one is the rate of accuracy, another is timeliness of the filter.

Accuracy is fundamental indicator of the evaluation of a filtration system. The main function of a filter system is to filter the sensitive information and shield it. One of the most important parts of the system is truly filter bad links, other functions which including the establishment of sensitive words, founding emergencies and to identify hot topics are for this function to work normally. Therefore, any model or a filter filtration system, should improve the rate of accuracy as the ultimate goal, to improve the accuracy is often said that the error rate is reduced, a perfect filtration system should achieve zero error sentenced rate.

Timeliness is also an important indicator of the filtration system, even it is not as important as accuracy but it can't be ignored, timeliness determines if the accuracy is meaningful. We all know, whether it is sensitive word, or unexpected events it keeps on changing, if sensitive words can grasp the latest formation at the earliest time, then it will be able to intercept sensitive information in a time. If the emergency is over, the hot topic is no longer a hot topic, a new word has spread sensitive networks and filtration systems for the information began to filter and block, even though the filter achieve 100% accuracy rate it will still be useless. This also show the importance of the timeliness of a filtration system, more sensitive words can grasp the latest information in the first time make the system more competitive.

Through high-performance computers dealing with massive statistical data, we can find the relationship between users of heat topics, the media heat topics, heat titles, semantic heat word, such as heat and timeliness heat and context, over the incident, the weight will be more realistic, enhance the timeliness of the system. Results of the assessment will affect word in the sensitive word dictionary.

Analysis of experimental results: Figure 2 shows the experimental results of the heat medium based on the model. Fig. 3 shows the combination of the heat sensitive information based on the user to find the model results. Fig. 2 shows the degree of media attention (in million articles reported) synchronized with the social hot topics, Fig.3 shows that sensitive information is not necessarily hot, due to the limitations of the dictionary structure, if the threshold value of 0.5, then the first five months of sensitive information is " Fishing Islands " event. Experimental results are basically in line with reality.

Use classifier provided by weka3.5.4. From each type of sensitive text, extract 18 grammatical features, 16 semantic features, eight emotional features. Classification using only the sensitive word to classify sensitive information, were carried out 30 times, the average results of the experiments are shown in Table 2 and Table 3.

From Table 2 and 3 we can see, the classification precision and recall rate based on the characteristics of sensitive information is significantly less than sensitive word it is because we do not give weight to the features, without considering the overall semantics (is our next step research to be carried out); results based on the characters of negative sensitive information(Falun Gong,

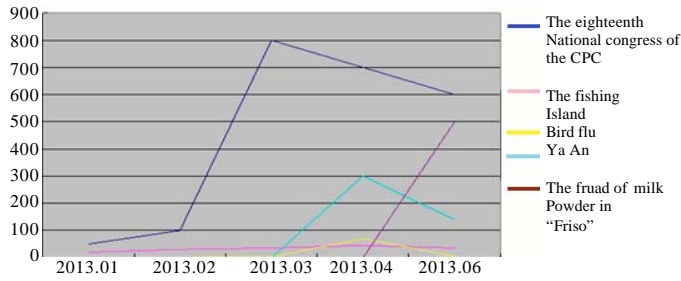


Fig. 2: Experimental results base on the media heat model.

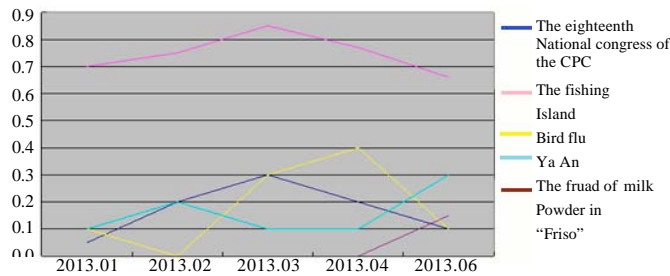


Fig. 3: Experimental results based on combination of user heat and sensitive information

Taiwan independence) is better than positive (Diaoyu Island, South China Sea) which is comply to reality ; Result based on characteristics of the normal text is less accuracy than results based on sensitive words which is also realistic. The error rate based on the characteristics of sensitive information is significantly less than the one based on sensitive words it is because the classification of sensitive word cannot distinguish normal and sensitive.

CONCLUSION

The current status of the information filtering model is with the aid of a sensitive dictionary, without distinguishing sensitive word and sensitive information. It brings the problem that filter the information which should not be filtered and does not filter information which should be filtered. In this article it gives a hierarchical filtering sensitive information model. At first it filter the text which does not contain sensitive word, the second step based on the analysis of the tendency of sensitive words, then filter sensitive information. This layered model has two effects: reduce the complexity of large data calculation; give overall evaluation from sensitive information and semantic perspective. This model is also scalable, if required, can be divided into finer granularity filtering model to go deeper .

ACKNOWLEDGEMENT

The National Natural Science Foundation of China (61171132) and Jiangsu Natural Science Foundation (BK2010280) funded

REFERENCES

Bermingham, A. and A.F. Smeaton, 2010. Classifying sentiment in microblogs: Is brevity an advantage. Proceedings of the 19th ACM International Conference on Information and Knowledge Management, October 26-30, 2010, Toronto, Canada, pp: 1833-1836.

Greevy, E. and A.F. Smeaton, 2004. Classifying racist texts using a support vector machine. Proceedings of the 27th Annual International ACM SIGIR Conference, July 25-29, 2004, Sheffield, UK., pp: 468-469.

Zeitsoff, T., 2011. Using social media to measure conflict dynamics: An application to the 2008-2009 gaza conflict. J. Conflict Resolut., 55: 938-969.

Zhou, Y.L., E. Reid, J.L. Qin, H.C. Chen and G. Lai, 2005. US domestic extremist groups on the Web: Link and content analysis. IEEE Intell. Syst., 20: 44-51.