

Application for Counter Terrorism using Word Substitution

Londhe Shital S.¹, Devare Vishakha R.², Shinde Asmita B.³, Gujarathi Madhuri R.⁴

^{1,2,3,4} Student of Computer Engineering, SVIT Chincholi

Abstract - Terrorist attacks are the major problem for the society. To know their intentions early and avoid the attacks is very necessary. In our technique we are going to allow automatic detection of suspicious messages. We are developing a desktop application where the admin who is monitoring all chats can crack the chat between two person whose chat seems to be suspicious. By performing Natural Language Processing Algorithm on the chat admin will able to know the exact meaning of the conversation by substituting the irrelevant words by the point of view of attacker. For this we have used Hadoop, Map-reduce, Word substitution like techniques.

Key Words: Counter terrorism ,Data Mining, Document Analysis ,Map Reduce , Sentence Substitution , Natural Language Processing.

1. INTRODUCTION

In our system admin and user roles, Admin is able to view the live chat and monitor the same. User can chat with any other user which he cannot monitor. He is only able to send and receive the messages to and from the other user. While chatting if there is any occurrence of obfuscated word then the algorithm is being able to detect the same and is able to convert the obfuscated word with a normal one. Database is being used for analysis of the same where all the chat will be stored in a text file, the same text file will be stored into database as a storage purpose. Regarding the various propose models this system is motivated to find the obfuscated chat in terrorism, this techniques focuses on collection, observation and preservation of suspected words and information occurs during conversation, so as to analyze the suspected words. Our paper explains the tools in the digital media (storage), digital forensic methods and cyber crime. This system proposes a new tool which uses the word Substitution technique. The proposed system is designed for finding the suspected words occurs in conversation and replaces these words with normal ones which are stored in our hadoop database and the message is received in the form of normal text.

2. LITERATURE SURVEY

Factored language models and generalized parallel back: The problem of detecting a word that is somehow out of context occurs in a number of settings. For example, speech recognition algorithms model the expected next word, and back up to a direct interpretation when the next word becomes sufficiently unlikely. These problem differs from the problem addressed here because of the strong left context that is used to decide on how unlikely the next word is, and

the limited amount of resources that can be applied to detection because of the near real time performance requirement Application for counter terrorism using word substitution.

Winnow-based approach to context-sensitive spelling correction: Detecting words out of context can also be used to detect (and correct) misspellings. These problem differs from the problem addressed here because the misspelled words are nonsense, and often nonsense predictably transformed from the correctly spelled word, for example by letter reversal.

Using common sense knowledge base for detection word obfuscate in adversarial communication future information security (Swati Agrawal , Ashish Sureka) : use of concept net to compute conceptual similarity between two term solving problems of word conduct empirical analysis on large and real word data set Conflict occur in substitution of words due to the complex design.

Detecting threats of violence in online discussion using bi-diagrams of important words (Hugo Lewi Hammer): Use of Classification method within text mining classification performed using LASSO logistic regression .But the drawback of this system is rate of wrongly classification of word is about 20 percent so cannot used for real time substitution.

Detecting word substitution in text. IEEE Transaction on knowledge and data (SW. Fong, D.Roussinovei and D.B.Skillicorn): design the system of measures that applied to sentence, positive detection rate is 90 percent and false rate is around 10 percent. Difficult to understand which design is used and conflict data structure.

Surveillance issue for security over computer communication and legal implication. proceeding (Mr Shilpa Mehta, Dr.U.Eranna, Dr.k.Soundatatajan)

3. PROPOSED SYSTEM

In our system both user as well as the admin logs into the system. The user can chat with the other users. The admin can monitor the chat between the users. The admin can crack the chat between them. In our proposed system we are going to use natural language processing, word substitution, map-reduce like techniques. The input to our system will be in the form of simple text message, or posts from the social Medias or emails. The admin will provide the input for word substitution. Our system uses hadoop database to store data. To extract meaningful word or substitute the relevant word

for the suspicious word from the hadoop database we are using map-reduce function. Our proposed system also inculcates some algorithms for word substitution. After performing all the operations the final result is provided to the admin in the form of normal text message again.

4. SYSTEM ARCHITECTURE

The above architecture of the proposed system shows, there are two suspicious persons i.e., user A and user B. Both the users communicate with each other. The admin monitors the chat between the two suspicious persons, cracks the chat between them. The admin then provides the chat in the form of normal text message for natural language processing. After performing natural language processing on the data the relevant word for the suspicious word is substituted from the Hadoop database. Then the output is provided back to the admin in the form of normal text file.

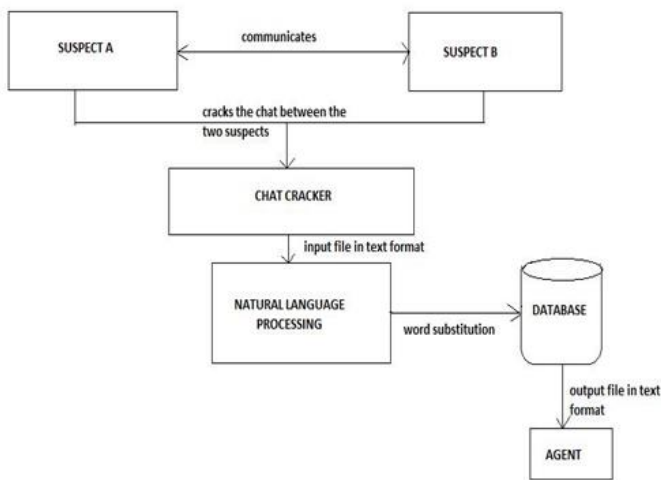


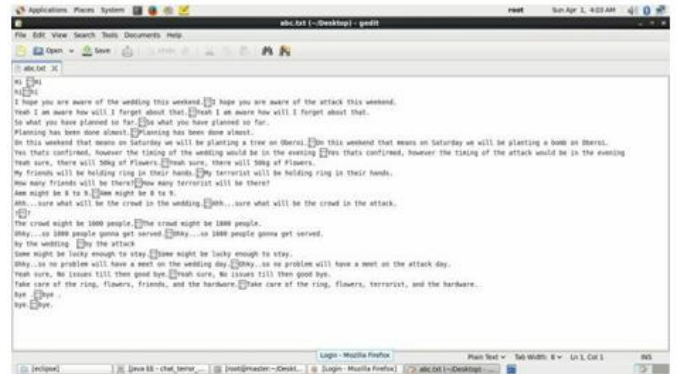
Fig 1. System Architecture



In the above screenshot the user 1 and user2 are chatting with each other.



And admin will monitor all the chat and substitute the suspicious words.



After doing all the substitution the final result is provided to the admin in the form of normal text message file again.

5. Mathematical module

Mathematical Formulation:

$$S=I, F, O$$

Where,

I=Set of inputs

i1=Normal text messages

i2=Images

i3=Audio conversation

F=Set of functions

f1=Hadoop

f2=Map reduce

f3=word substitution

O=Set of outputs

o1=Messages in normal form

o2=Identity of suspected person

o3=Location of suspected person

Success Condition=Gets the relevant information from the communication of suspected person.

Failure condition=Connection is loss system does not work, internet connection is compulsory.

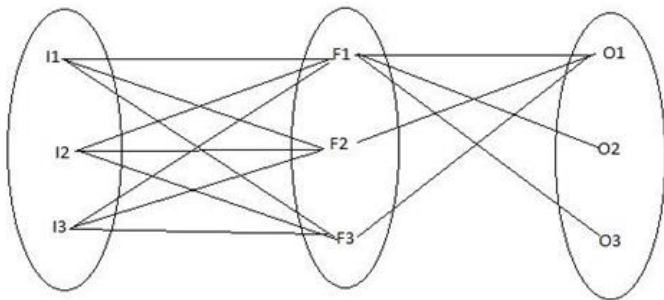


Fig 3.Mathematical module

6. CONCLUSIONS

We have tested how word substitutions within textual communication can be detected. Our technique allows us to automatically detect suspicious messages, so that they can be further investigated either by a more sophisticated data mining techniques or manually. The task of detecting substitutions is being important because terrorists, criminals, attackers and other adversarial parties may use substitution in order to avoid the leakage of information because of the use of certain words.

- For this we have to use natural language processing, hadoop, map-reduce, word substitution, etc like techniques. When users are doing their conversation, if admin monitors some suspected words in their communication admin will browse the chat of this person and send it to hadoop map reduce model for analysis. Hadoop and mapreduce then analyzed the chat and calculate the occurrence or frequency number of chat, mapped it and create text file of chat. This text file is deliver to Hadoop database for searching most related words to that suspected words and word substitution is done in this process. Finally the system will generate the message in normal form which can understood by the admin. This system is useful for Government and Private Security Agencies to find the most important information.

REFERENCES

- 1)Vaibhav Shinde,Komal Salunke ,Punam Teli Deducing and conclusive analysis of declassified obfuscate chat in terrorism using word substitution ,IERJ, March 2017.
- 2)Hugo Lewi Hammer, detecting treats of violence in online discussion using bigram of important words.2014 IEEE Joint Intelligence and Security Informatics Conference.
- 3) Sonal N. Deshmukh, Ratnadeep R. Deshmukh and Sachin N. Deshmukh Finding Real Semantic of Replaced Words Using K-gram and NGD. World Congress on Engineering 2013 Vol III, WCE 2013, July 3-5, 2013,London, U.K.
- 4) Mrs. Shilpa Mehta, Dr. Eranna, Dr.K.Soundararajan, Surveillance Issues for Security over Compute Communications and Legal Implications, Proceedings of the World Congress on Engineering 2010 Vol I WCE2010, June 30 - July 2, 2010, London, U.K.
- 5)SW. Fong, D. Roussinov, and D.B. Skillicorn. Detecting word substitutions in text. IEEE Transactions on Knowledge and Data Engineering, 20(8):10671076, 2008.
- 6) J. Bartlett, J. Birdwell, and M. Littler, The rise of populism in Europe can be traced through online behaviour..., 21-January-2014.
- 7) I. M. Sunde, Preventing radicalization and violent extremism on the Internet (Norwegian), The Norwegian Police University College 2013:1,2013.
- 8)S. N.Deshmukh ,R. R.Deshmukh ,performance analysis of different sentence oddity measure applied on google news repository for detection of substitution,International referred journal of Engg and science,2014.
- 9)H. Zhao and X. Li, S3PAS: A Scalable Shoulder Surfing Resistant Textual Graphical Password Authentication Scheme, in Proc. of 21st international Conference on Advanced Information Networking and Applications Wprk shops, 2007.
- 10) D Mitri r Oussinov , S Zewang F Ong , D Avid S Killcorn ,Detecting Word Substitution : PMI VS HMM, Sigr 2007,A Msterdam Proceedings.
- 11) Transcript: Senator Dick Durbin on Surveillance, Terrorism and the Constitution Newstin Oct 5 2009.
- 12) Arlen Specter, The Need to Roll Back Presidential Power Grabs The New York Review of Books, Volume 56,Number 8 May 14, 2009.
- 13) S. Fong, D. Skillicorn, and D. Roussinov, Measures to Detect Word Substitution in Intercepted Communication, Proc. IEEE Intl Conf. Intelligence and Security Informatics (ISI 06), May 2006.

BIOGRAPHIES

Ms. Shital Suresh
Londhe, Student, at Department of
Computer Engg, PRES's Sir
Visvesvraya Institute of Technology,
Chincholi, Nashik, 422102, Savitribai
Phule, Pune, University, Maharashtra,
India



Ms. Vishakha
Rajendra Devare Student, at
Department of Computer Engg PRES's
Sir Visvesvraya Institute of
Technology, Chincholi, Nashik,
422102, Savitribai Phule, Pune
University, Maharashtra, India



Ms. Asmita Bapu Shinde
Student, at Department of Computer
Engg PRES's Sir Visvesvraya Institute
of Technology, Chincholi, Nashik,
422102, Savitribai Phule, Pune
University, Maharashtra, India



Ms. Madhuri Rajesh Gujrathi Student,
at Department of Computer Engg
PRES's Sir Visvesvraya Institute of
Technology, Chincholi, Nashik,
422102, Savitribai Phule, Pune,
University, Maharashtra, India