DOI 10.29042/2018-3846-3848

Helix Vol. 8(5): 3846- 3848

Data Extraction by Using Natural Language Processing Tool

[1]Sujata D. More, [2]Mangala S. Madankar, Dr. M.B.Chandak
[1]Research Scholar, [2]Assistant Professor, Computer Science and Engineering, G. H. Raisoni College of Engineering, Nagpur.
Email: sujatamore123@gmail.com, msmadankar@gmail.com

Received: 09th July 2018, Accepted: 14th August 2018, Published: 31st August 2018

Abstract
In case of smart cities, main concern lies in analysing the huge amount of data coming from various sources in real-time. In todays world, usage of social networking sites has expanded exponentially. So we may say that people acts as a detector by using these social networking sites and detects the information in various forms. So every individual here acts as a detector who gets the information from real world. Anyone can report such information made by others as their own. The information could either be true or false depending on individual perception. This paper presents the role of data detection from social networking sites. People are sources and represented as detector. The information they provide is represented as data. The data is identified by detector to find out which data is correct or not and it split the data from Interference by using Natural Language Processing.

Keywords: People as Detector, Social Detecting, Natural Language Processing, Data Reliability.

Introduction

People obtains the information about physical world from social networking sites such as Facebook, Twitter, Whatsapp, etc. These social networking sites plays vital role in our society and are very helpful in getting the information about the world. In March 2011 Japan suffers horrible tragedy as Tsunami. In that so many people destroy there life's. Twitter is the first social Network which describe the damage of Japan Tsunami . The Japan Government later suggest to encouraging the use of social network in the study for calamity recovery[6].

It is important to know that social networking apps like twitter contains various kinds of information about the world. Many people posts updates about their day to day activities or we can say post the data on twitter. Out of which some data are correct, some are irrelevant or some data are personal. We can use such data which is in the form of tweets posted by people, for our analysis purpose and find so many real world problems in the form of calamities or disasters.

This paper presented role of social networks as sensor networks. People sources represent as Detector. The information they make represent data. The data is identified by detector to find which data is correct or not. And it split the data from interference by using natural language processing.[3]

Literature Review:

A Data mining and machine learning technique used for Natural calamity and Crisis[1]:

This paper use the machine learning and data mining technique to support for decision making problems. Situational awareness and real-time threat assessment problem solving by these two technique.[1]

A framework for detecting unfolding emergencies using humans as sensors [2]: This paper implements the architectural framework for detecting the emergencies with the help of human as a sensor paradigm. [2]

Human-Agent Collectives act as a Calamity Response system. [3]: This paper implements a calamity response system called HAC -ER. Human and agents plays important role to collect the information. Software and robotics which individuals find the social relationships. [3]

Statistical based analysis of Disaster management[10]: In this paper, The statistical analysis which find word association. With help of twitter post a real earthquake disasters identified.[10]

System for proctor Natural Tragedy by using Natural language processing in social network. Twitter[9]: This paper shows the innovation and Effectuation of an machine controlled system and use the twitter for proctorial the data And API is use for filtering purposes. The data are store into the database for analyse which is obtained from twitter.[9]

Proposed Methodology

Data Collection module:- We compile the data from twitter. And collect many disaster related words or their synonym and store it into the database. And also store the places or area name to find the disaster related sites.

Sentence Detector Module:- The sentence detector module will detect the twits available on twitter. Which is posted by human. which act as an Human as a Sensor.

Tokens Detector module:-By using token detector module we can break the text into small

word, Symbol, Phrases, or many meaningful sentences**.**

POS Tag Detector Module:- The part of speech tags are used to find the sentence whether it is Noun, Pronoun, Adverb, Verb, Adjective, Preposition, Interjection, Conjunction.

Stop Word Removing Module: Removing the stop word which will help to show the sentence.

Frequent Term to Detect Social Alert Module: After applying all module we get the frequent term to show the social alert message.

General Scheme Design

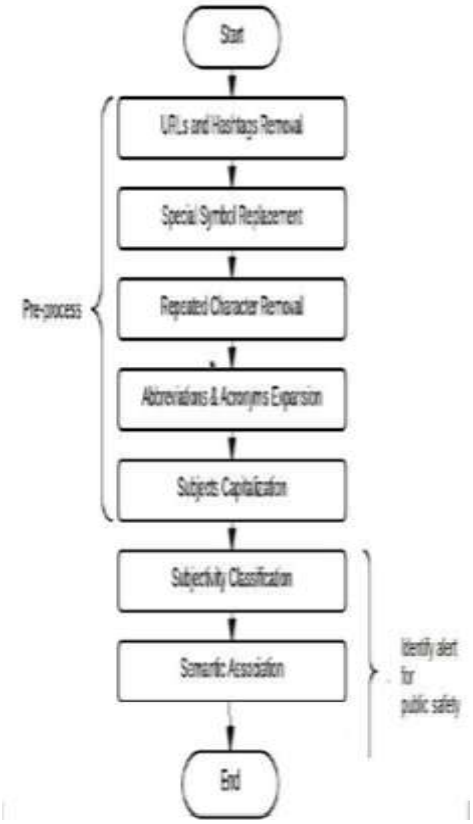

Figure 1. Flow Graph of Work

Proposed Algorithm

Consider the words which is input W1,W2 there are two paragraphs p1i which is used for word W1 and p2j paragraph for word W2, we have to finding the word from twitter and compare with dictionary(database) N1i and N2j.word will be match with each other then we have to find unique list and that unique list compare with initial W1 word. then we get final unique list will be compare with relation which is defined in algorithm. if word is strong and that will store in our database. then we have to show the result.

Optimized Word Similarity Algorithm.

Step1: for the two words W1, W2, W1 have paraphrases {p1i} ,1 is less than i less than I, W2 have paraphrases { p2j}, 1 is less than j less than J..

Step2: take paraphrases p1i, p2j, find some similar code in database : N1i, N2j.

Step3: If N1i = N2j, then they are synonyms, Sim (N1i, N2j)
=Sim(p1i, p2j)= 1, Exit

Step4: If N1i ≠ N2j, find there node i, factor θi, θi-1 by the twist.

Step5: If N1i, N2j compactness according to Database.

Step6: If N1i, N2j find X strong relation correlation between them, then the similarity is :
Sim_N1i,N2j_ = (θ + δ1) + β · · ε 3
=0.1go to Step 8

Step7: If N1i, N2j fulfil Z relation as a weak correlation
between them is said, then the similarity is:
Sim_N1i,N2j_ = (θ + δ2) + β · · ε 4
(β = 0.1), go to Step 8.

Step8: If W1 And W2 identify by all words(paraphrases)
then exit.
else
go to Step2

Result and Discussion

Figure 2. Home page

Figure 3. Data Base Entries

This paper searching the utilization of social network as sensor network to find the correct perception that something has occurred or some problem exists. In this problem, everyone is act as Detector. Who time to time find the information about the Real world. People can detect disaster related sentence through twitter and with the help of natural language processing tool. we can analyse the sentence with the help of sentence detector module then chopping the sentence into pieces with the help of token detector module. part of speech apply to identify the word whether it is noun, pronoun, verb, adverb and adjective.
then stop word removing module remove the unwanted word ,it will help to exact the meaningful word. After applying all module we get the frequent term to identify the social alert message. In database so many disaster related words which I store and also the area and city name to identify in which area of city the disaster happen.

Conclusion & Future Work

This paper searching the use of social network as sensor network to find the correct perception that something has occurred or some problem exists. The information they get from twitter represent as a data . The sensing problem Which shows that which data is correct or which data is wrong . Which is to say, Split the data from noise using natural language processing tool. The solution was implemented in our application and find the best accuracy .This paper can be Extend in broad level by creating mobile application. This application will be publicly available and people will download all notification, This can be as a real-time application. We can create our own new Social networking Site, So the people will easily create the account, & get the real-time updates related to Disaster. So this will help for society.

References

1. A. Caragliu, C. Del Bo, and P. Nijkamp,2013 "Smart cities in europe," Journal of urban technology, vol.18,no. 2, pp.65–82
2. N. Komninos, M. Pallot, and H. Schaffers, 2015. "Special issue on smart cities and the future internet in europe," Journal of the Knowledge Economy, vol. 4, no. 2, pp.119–134
3. D. Doran, S. Gokhale, and A. Dagnino, 2016 "Human sensing for smart cities," in Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. ACM,2013 ,pp.1323–1330.
4. H.N. Teodorescu, 2015 "Using Analytics and Social Media for Monitoring and Mitigation of Social Disasters", Procedia Engineering, DOI 10.1016j.proeng.2015.06.088, 107C:325-334,
5. P. Anantharam, P. Barnaghi, K. Thirunarayan, and A. Sheth, Jul. 2015 "Extracting city traffic events from social streams", ACM Trans. Intell Syst. Technol. Vol 6, No 4, pp 43: 1-43:27. [Online] Available: http://doi.acm.org/10.1145/2717317
6. Yuki Takeichi, Kazutoshi Sasahara, Reiji Suzuki and Takaya Arita,2014"Twitter as Social Sensor: Dynamics and Structure in Major Sporting Events"http://dx.doi.org/10.7551
7. J. M. L´opez-Higuera, L. Rodriguez Cobo, A Q. Incera, and L. RCobo, , 2011 "Fiber optic sensors in structural health monitoring," Lightwave Technology, Journal of, vol. 29, no. 4, pp. 587–608.
8. A. C. Perchet. (2013) Reduce traffic congestion And information services in urban places Available at http://bit.ly/XtEeqm/
9. Miguel Maldonado, Darwin Alulema, 2016 "System for Monitoring Natural Disasters using Natural Language Processing in the Social NetworkTwitter"978-1-5090-1072-1/16/$31.00
10. Mironela Pirnau,2017"Word Associations in Media Posts Related to Disasters - A Statistical Analysis978-1-5090-6497-7/17/$31.00
11. S. F. Apache, 2015. "Apache zookeeper," Available at https://zookeeper.apacheorg/,
12. Geohash.org, 2015."Geohash," Available at https:// en.wikipedia .org/wiki/ Geohash.
13. D. Jurafsky and J. H. Martin, Speech & language processing. 2ed. Pearson Education India, 2009.
14. NILC, 2015. Inter institutional Centre for Computational Linguistics, USP, "Lemmatize for Portuguese,"Available at http:// www.nilc.icmc.usp.br/ nilc/index.php/tools-and-resources
15. I. H. Witten, E. Frank, and M. A. Hall, 2011. Data mining: practical machine learning tools and techniques. Morgan Kaufmann,T. M. Mitchell, Machine Learning, 1st ed.1997. New York, NY, US McGraw-Hill, Inc.
16. S. Haykin, Neural Networks: 2007. A Comprehensive Foundation (3rd Edition). Upper Saddle River, NJ, USA: Prentice-Hall, Inc.

