International Journal of Engineering & Technology, 7 (2.20) (2018) 49-51

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

An efficient approach to predict emergency calls and locations

G.V.S. Narayana[1], T. Venkatesh[2*], P. Mourya[3], B. Ramya[4]

[1]*Professor, Department of CSE,Koneru Lakshmiah Educational Foundation, India.*
[2]*Student, Department of CSE,Koneru Lakshmiah Educational Foundation, India.*
[3]*Student, Department of CSE,Koneru Lakshmiah Educational Foundation, India.*
[4]*Student, Department of CSE,Koneru Lakshmiah Educational Foundation, India.*
**Corresponding author E-mail:venkateshtammisetty98@gmail.com*

Abstract

Big data stands for huge set or collecting information which can't be prepared by modern techniques, for example, data processing. Examining enormous information has the quality in the arena of interpersonal organizations, spot business patterns, web, drug, science, fund, commerce informatics and indeed in government. Dissecting information would assistance in extraordinary basic leadership, which may achieve change happening productivity, diminishment in cost and disappointment dangers. Enormous information examination turns into an awesome hunger for the creating associations since it winds up plainly troublesome for those associations to process a excess of tera bytes of information. Huge information investigation even discovers its request in considerate the purpose behind regular or man influenced failures by gathering enormous information keeping in observance conclusion to recuperate from the catastrophe and to develop the correspondence since correspondence is the primary test that the common people face while confronting sudden failures..

Keywords: *Big data, Hadoop, K-Nearest neighbour (KNN), map reduce, support vector model (SVM).*

1. Introduction

The present situation with behavioral examination might want of separating enthusiastic and wistful information from the social field which is rich with suppositions, regardless of the way that generally exists in internet , was countered some path with big- data arrangements utilizing machine learning. Be that as it may hadoop upset the entire setup Big data hadoop has been basically instrumental in adding an additional valency to the social data accumulated from social destinations like Facebook, Twitter, Pinterest, Instagram, though meanwhile amalgamating innovation with business interests for common benefit and participation. With machine learning and option apparatuses like listening instruments and conclusion examination, bigdata hadoop has been profession with progress to the professional world, whereby it uncovers unstructured information from innumerable Facebook Posts, Twitter Tweets, and Pinterest Pins.
Ventures locales utilize bigdata hadoop for putting away, revealing, and process information like "what number of people checked-in region all through twelvemonth festivals?" Not exclusively successful measure the business homes, inns, and furthermore the avionics business making best utilization of this mined data, however conjointly the social destinations like Facebook. One will gage the value of big data in online networking investigation if one experiences the "inclination list" of Facebook - the rundown with assortment of feelings: positive or negative, made calculative scientific for nostalgic examination, since the figuring dialect neglects to require into thought disordered up human feelings. despite the way that the etymology wide acknowledges parallel articulations like "the flight was cozy however didn't quite recently like the sustenance served on-load up," Facebook has crisped its rundown of feelings felt with the intention of downsize quality to least, consequently expanding the effectuality of data collected, that has been reportable as efficient, authentic and agile for certain effective measures.

2. Related work

Aditya Bhardwaj and Ankit kumar(2015)[1] have talked about on enormous information examination. As showed by them, Hadoop defines to the bulk of information past customary database innovation ability to store, get to, oversee then register proficiently. They said by breaking down this huge measure of information organizations can foresee the client conduct, enhanced promoting methodology, and get upper hands in the market.
As per them. hadoop is an adaptable and open source usage for breaking down expansive datasets utilizing Map Reduce They centered different developing innovations, for example, Apache Pig, Hive, Sqoop, HBase, Zookeeper, and Flume that can be utilized to enhance the execution of essential Hadoop Map Reduce structure. They said Apache Pig is a scripting dialect is utilized to diminish advancement time of Map Reduce program because it needs only less number of code lines of code it gives settled information sorts that are missing from Map Reduce.
Hive gives simple to utilize stage to the engineers who are agreeable in SQL dialect for Map Reduce programming, if HDFS fails then the arbitrary read/keep in touch with Big Data is given by HBase. They transferred information amongst Hadoop and RDBS framework utilizing Sqoop, Zookeeper is utilized for integrate of Hadoop bunch lastly Flume can be utilized for rapid spilling of web log information to Hadoop.

This paper is mainly used for executing the Twitter tweets by utilizing Hive question on HDFS Insight group and results demonstrates that as we increment number of hubs in the bunch, at that point Map Reduce space time increment however general aggregate time utilized for and RituTiwari(2016) [2] have concentrated on person to person communication sites which is a wellspring of different sort of data.

They said this is an immediate consequence of the idea of these sites on which people groups remarks and post their suppositions on various sorts of subjects i.e. they express pros and cons slants about any item that they use in day by day life, gripes and executing Hive inquiry expire. Raj Kumar Verma current issues etc etera. They said the assumptions help in getting data about different current patterns besides utilized further in choosing convenience of a few undertakings, items and subjects.

Likewise social web information like twitter has more information that individuals post so it's ended up being vital to chip away at effective shrewd frameworks that can do information refinement, investigation of undertakings astutely and proficiently. Dhiraj Gurkhe and NirajPal(2014)[3] have examined the viable Assessment Analysis of Social Media Datasets Using Naive Bayesian Classification.

The procedure includes outcome of subject data from literary information. A typical human can without much of a stretch comprehend the slant of a record written in normal in perspective of its data of understanding the limit of words (unigram, bigram and n-grams) and once in a while the general semantics used to portray the subject.

The paper plans to influence the machine to separate the extremity (positive, negative or nonpartisan) of online networking dataset as for the questioned watchword. The paper presented an approach for consequently ordering the conclusion of online networking information by utilizing the accompanying strategy: First the preparation information is nourished to the Sentiment Analysis Engine for learning by usage of machine learning calculation.

After the learning is finished with qualified precision, the machine begins tolerating singular social information regarding watchword that it dissect and translates, and after that arranges it as positive, negative or unbiased as for the question term. Laurie Butgereit(2015) [4] has focused on the occasion hung on 1 November in South Africa, 2014 in which a coal storehouse crumbled at Eskom's most up to date control station, Majuba. The paper concentrated on the harm constrained Eskom to execute moving piece outs (called stack shedding) all through the nation.

The paper explored on the off chance that it was conceivable to measure the relative outrage against Eskom as communicated in sets of posts on Twitter (called tweets). The paper proposed a calculation was created that deliberate certain qualities of the tweets, for example, swear words, emoji's, emoticons, capitalized letters, and certain accentuation marks. The outcomes were evaluated against comes about gave by two free individuals going about as coders.

These two individuals likewise assessed similar tweets. The outcomes demonstrate that as the polarity (or contrast) in outrage in two tweets builds, the calculation is almost as exact as two human coders. A.K.Santra and S. Jayasudha(2012) [5] have concentrated on conduct of the intrigued clients against investing energy in general conduct.

The current model utilized upgraded variant of choice tree calculation C4.5.

In the paper, they utilize the Naive Bayesian Classification calculation for characterizing the intrigued clients and furthermore they introduced a correlation investigation of utilizing upgraded variant of choice tree calculation C4.5 and Naïve Bayesian Classification calculation for recognizing intrigued clients. The execution of this calculation is measured for web log information along session planning, rehashed client profiling, and page profundity to the webpage length.

3. Proposed algorithm

A. Design considerations

- Establishing Connection Twitter Authorization using FLUME or Twitter4J
- Storing and Preserving Data (Tweets) which is in JSON in HDFS along-with HBASE
- Creating Meta Structures and Tables in HIVE
- Integrating and Mapping JSON data with HIVE Meta Structures
- Extracting data using HIVEQL and Map Reduce
- Pre-Processing on extracted Data
- Forming KNN/SVM classifiers for results.

B. Description of the Proposed Algorithm vide KNN and SVM classifiers

Step 1: KNN linear correlation

It is used to measure the ability of the relationship among 2 variables. If there is no change among the two valuables, Then there is no certain way then the values of the first capacity may raise or reduce the with values of the second capacity.

The strength of the linear relationship lies between -1 and 1 only.

-1 which means it is pure negative and +1 which means pure positive.

Step 2: SVM Categorization:

Consideration of both ability and Efficacy are important. This selection feature is applied for appliances and to learn different methods for text grouping. To decrease the count of the features, the steps are first, by considering the frequency counts we remove the features. Later, adopt a less number of features that fit into the categories.

4. Pseudo code

Step 1: Extraction data via Flume from twitter or Twitter4J
Step 2: Hive Script
Step 3: KNN model KNN (dataset, sample) {
1. Calculate distance for each item.
2. Classify the pattern as the majority class between K samples in the dataset having minimum length to the sample.
3. Compute dataset for K and calculate Distances.
4. return0 }
Step 4: SVM 4 Model.

5. Conclusion

For analyzing the user opinion, first of all twitter data is extracted using flume. The data extracted is available is in unstructured (JSON) format.

The data is integrated with Hadoop. Using hive it is given a tabular form, i.e. a structured form of data is obtained. Maven framework is used to get the executable jar to integrate eclipse and Hadoop. Data needs to be filtered before analyzing. Data is cleaned by removing stop words. For classification, KNN/SVM has been used.

For using naïve KNN/SVM technique, we have used a dictionary which stores a list of words that are positive, negative and neutral. Lastly, data is imported to excel to give a graphical form and to get the results. In the scheme, we can identify the user opinion with the help of user id whether the user is positive, negative or in drifting mode.

Also, the system tells the general behavior of users country-wise as well as city-wise for a particular topic. The system is 7080% accurate. In future, the data can be from multiple sources at the same time. Also, various different tools like R, a table can be integrated, also we can continue with ontology in it. Finally,

multiple topics also can be taken into consideration. Further works can be done to improve the efficiency and accuracy.

References

[1] Bhardwaj A, Kumar A, Narayan Y & Kumar P, "Big data emerging technologies: A CaseStudy with analyzing twitter data using apache hive", *2nd International Conference on Recent Advances in Engineering & Computational Sciences (RAECS)*, (2015), pp.1-6.

[2] Butgereit L, "An algorithm for measuring relative anger at Eskom during load-shedding using Twitter", *AFRICON*, (2015), pp.1-5.

[3] Santra AK & Jayasudha S, "Classification of Web Log Data to Identify Interested Users Using Naïve Bayesian Classification", *IJCSI International Journal of Computer Science Issues*, Vol.9, No 2, (2012).

[4] Sagiroglu S & Sinanc, D, "Big data: A review", *IEEE International Conference on Collaboration Technologies and Systems (CTS)*, (2013), pp. 42-47.

[5] Pal A & Agrawal S, "An experimental approach towards big data for analyzing memory utilization on a Hadoop cluster using HDFS and MapReduce", *IEEE, First International Conference on Networks & Soft Computing (ICNSC)*, (2014), pp.442-447.

[6] Bedi P, Jindal V & Gautam A, "Beginning with Big Data Simplified", *IEEE International Conference on Data Mining and Intelligent Computing (ICDMIC)*, (2014), pp.442-447.

[7] Hassan S, Yulan H & Alani H, "Semantic sentiment analysis of Twitter", *The Semantic Web– ISWC*, (2012), pp. 508-524.

[8] Abdul-Mageed M, Diab M & Korayem M, "Subjectivity and sentiment analysis of modern standard Arabic", *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Vol. 2, (2011).

[9] Abdul-Mageed M & Diab M, "AWATIF: A multi-genre corpus for Modern Standard Arabic subjectivity and sentiment analysis", *Proceedings of LREC, Istanbul, Turkey*, (2012).