A Real Time Stream Data Processing and Analysis Model and Catchments over Twitter Stream Data

Ankit Sarawagi
Department of Computer
Science & Engineering,
UIT RGPV, Bhopal; India

Rajeev Pandey
Department of Computer
Science & Engineering,
UIT RGPV, Bhopal; India

Raju Barskar
Department of Computer
Science & Engineering,
UIT RGPV, Bhopal; India

S. P. Pandey
RBS Engg Technical
Campus Bichpuri, Agra;
India

ABSTRACT

Big data processing is an important aspect in todays world. Twitter produce a large number of tweets and different segment of data according user usage and post. Understanding the proper sentiments, extracting the proper meaning from it is an objective task which is required different processing tools and methodology. Real time data gathering, storing them and analyzing efficiently to produce effective and fast accessible result approach is always a required work today. For this purpose in this research work a technique PSWNSWAP is proposed, which use Twitter stream data gathering in real time as well as Fast indexing, processing and performed sentiment analysis of gathered data. Distance computation, finding the right place to perform some operation is the tedious task for business operation or any brand to get established in new areas. Here's an algorithm which is St-QAP algorithm, is investigated and processed with the Apache Storm tool and NLP library. The Objective is to produce an efficient path mapping and catchments for new brands to establish in a new area and solving investigation behind it. Our proposed algorithm computed efficient result, while comparing with existing traditional solution with it.

Keywords

Big Data processing, Real Time streaming, Twitter, NLP computation, Storm processing, PSWNSWAP, St-QAP, Distance computation, Catchments.

1. INTRODUCTION

Stream data analysis in a Real time [1], emerging as the quickest and most proficient way to get useful information about what is going on now such as tweets on Twitter [2], enabling associations to respond immediately when issues show up or to identify new patterns enhancing their performance. Large number of data processing, finding an efficient pattern and solution for them is an important task. This must be important for SNA (social network analysis) [3]. Massive or huge amount of complex data generated rapidly per unit time [4] from various social sites such as Twitter, Facebook, YouTube, Instagram and other Big Data application domains [5]. Micro blogging and social media Twitter gather millions of data in a day for any specific post or product. This processing and analysis of massive amount of stream data must be needed to perform in Real Time. This research paper offers a framework for processing and analyze the Real Time stream data in an efficient manner.

Natural language processing [6] is an important library and approach to understanding the data's significance. Data mining [7] and processing large data keeps a track of usable entity . Sentiment Analysis [8] is an overall attitude of a speaker, writer, reader, or any other entities, with respect to some topic written in a piece of text. It is an effective technique for discovering public opinions. So a technique PSWNSWAP is proposed, which use Twitter stream data gathering in real time as well as Fast indexing, processing and performed sentiment analysis on gathered data.

Various offline market research and directory investigation is required for any company or brand to get established in new areas [9]. In this paper an approach to investigate a brand occupancy over an area and performing the visiting area by their defined rule is performed. St-QAP is an approach which used for Catchments in business, distance mapping and solve the travelling salesman problem, short distance and path optimization issue.

2. PROBLEM DEFINITION

In the previous research work, There are different techniques with the data mining and twitter data analysis [10] with the data storage, it's applicability over the data center, server and accessing is performed by different user. Previous techniques worked on data analysis provided by their static dataset which is not real time stream data and thus a proper analysis cannot be performed. As the study is taken and performed with various strategies & techniques and distinct outcomes from the algorithms were monitored such as PSWAP [11] and various other approaches to solve the tweet analysis and further be finding efficient locality data over it. Spatial data distribution [12], data location description, bandwidth determination and other relevant research performed are limited to particular area and moreover limited to statically investigation or research. Upon verifying distinct scenario and the available strategies, techniques various short comes with the Existing algorithm for Geo-tagging [13] and relevant data, finding with twitter file based sentiment detection, which is taken as a base for our analysis work.

The following are the issue which can be monitored and identify as a problem. These problems can be analyzed and performed further with upgrades and enhancements-

i. Previous technique such as twitter analysis over the formulation to find relevant location is limited to the statically defined dataset (not in real time stream dataset). Thus the approach can't be able to work beyond the given data. [13]

ii. The existing algorithm takes advantage over previous traditional techniques, but still more refinements are required as per today's standard. And the existing algorithm is also limited for static datasets. Thus a proper sentiment analysis, hashing mechanism is required in real time, which can make it more reliable and executable to tackle with current cloud scenario in the world. [13]

iii. Multiple user from anywhere, tweets over the different product, data availability, data review and comment on them. These reviews or comments become more important for any company or brand. Existing approach not specify this, in real time.

iv. A combination of twitter and data optimization is taken in the consideration which is neither more reliable while talking about accuracy, again an extra procedure is required to do the real time data exchange. Thus, it exhibits extra computational time as well as computation cost for cloud server. [11]

v. FP growth algorithm makes a repeated computation and accuracy over the repeated value.

vi. Sarcastic keyword analysis outperforms low accuracy, precision and other parameter analysis over the data. [11]

vii. A bulk number of Spam tweets give poor data combination, data verification over large noise tweets are observed.

viii. The existing approach for KDE [13], it allows the fixed bandwidth over the data availability.

ix. Previous shortest path derivation algorithm such as ANT, Dijkstra's and other approach works on the fixed pattern and no dynamic decision is described.

x. Spatial key distribution is performed over statical map. [13]

3. PROPOSED METHODOLOGY

In this proposed Methodology, we modify the existing technique [13] by new and more efficient technique of data, finding and collection as well as trend finding. We replace some previous concept which is necessary for retail market searching, an aspect that will help to increase accuracy and reduce the computation cost, computational time as well as total execution time.

i. In this proposed architecture Processing and analysis of twitter data can be performed in Real time. So the Real time stream data is used for this research work. This can reduce the issues and problems identified in the existing framework.

ii. Perform proper sentiment analysis of comments, reviews, Tweets, Opinion of users in Real Time. PSWNSWAP is used in this proposed architecture for the purpose of sentiment analysis. PSWNSWAP is enhancement of PSWAP. The algorithm performed using NLP library. This can reduce limitation of existing framework.

iii. Storm [14] framework is used for calculating Tweet value.

iv. Use an improved form of assignment problem, St-QAP technique in place of existing distance finding technique.

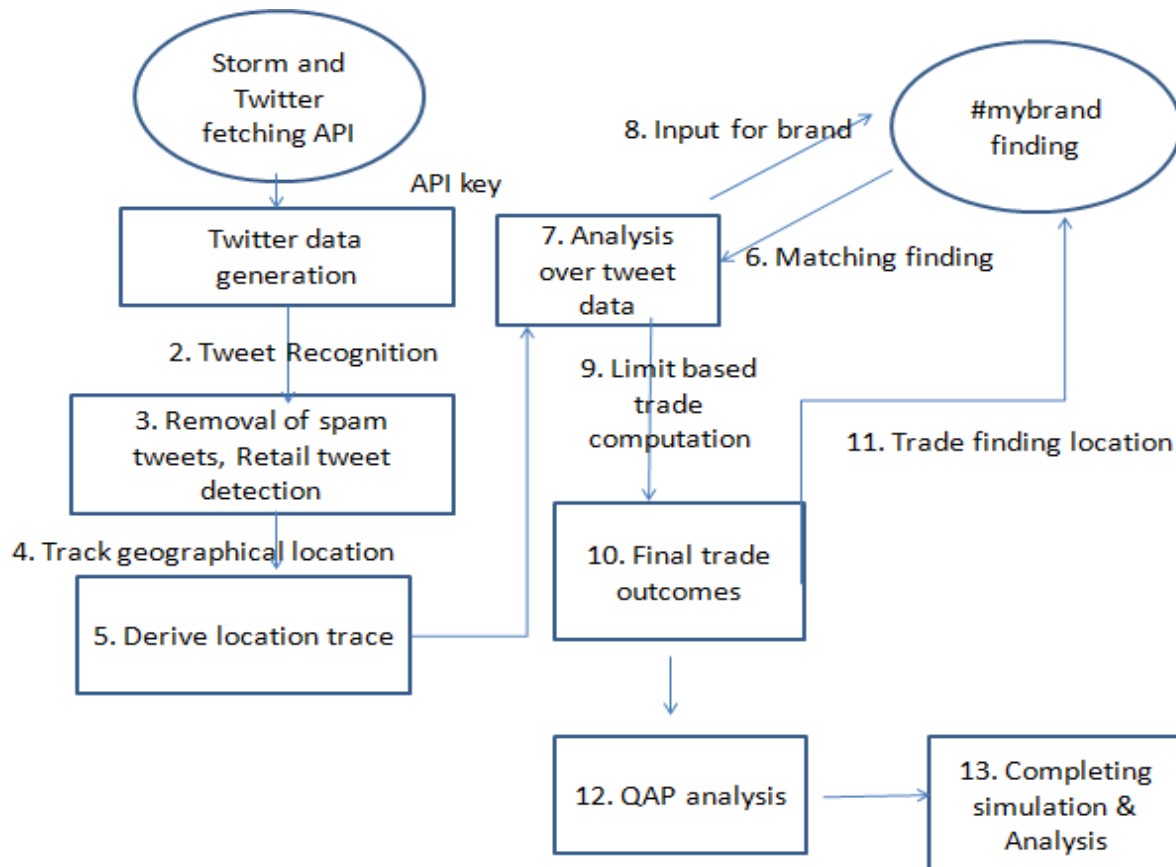v. Perform large file integrity processing in the proposed system.

Fig 1: Architecture of Proposed System

In proposed architecture there are three components called Storm & Twitter API, Proper sentiment analysis of real time stream data, Location Matching &St-QAP technique, which are used to provide a location finding in cloud [15]. Detailed descriptions of these components are shown below.

i. Tweet Collection API & Storm

A collection by providing input as the keyword and keys for the tweet searching API is given by the input values. Storm is the framework available with Zookeeper and other programming tool, a platform which help in working with the authentication with the high dimensional spatial platform. It also gives the input to the availability of data. Data tweets in real time collection is performed by the Storm, which is useful for a proposed work analysis.

ii. Sentiment analysis of Twitter Real Time Stream data

At Step 2 and 3 on fig 1: there is more architecture and process is performed which is processed using NLP (Natural language processing). The algorithm PSWNSWAP performed using NLP library, which is able to process tweets extracted. Thus, in order to understand the tweet semantics NLP is applied. PSWNSWAP (positive sentiment with negative sentiment with antonym pair) algorithm is used in this research work for sentiment analysis on real time stream Twitter data. This algorithm calculates the positive or negative tweets, comments, reviews, from Twitter data. This must be important for any company for marketing, business purpose in analyzing the product rating or review. PSWNSWAP performed using NLP library, which is able to process tweets extracted. Thus, in order to understand the tweet semantics NLP is applied. First of all segmentation is performed over the tweet, further recognizing nouns, pronouns, verb, adjective are determined from the input sentence. A further steps pruning and thus sentence understanding are performed at step 3, which process the input tweets.
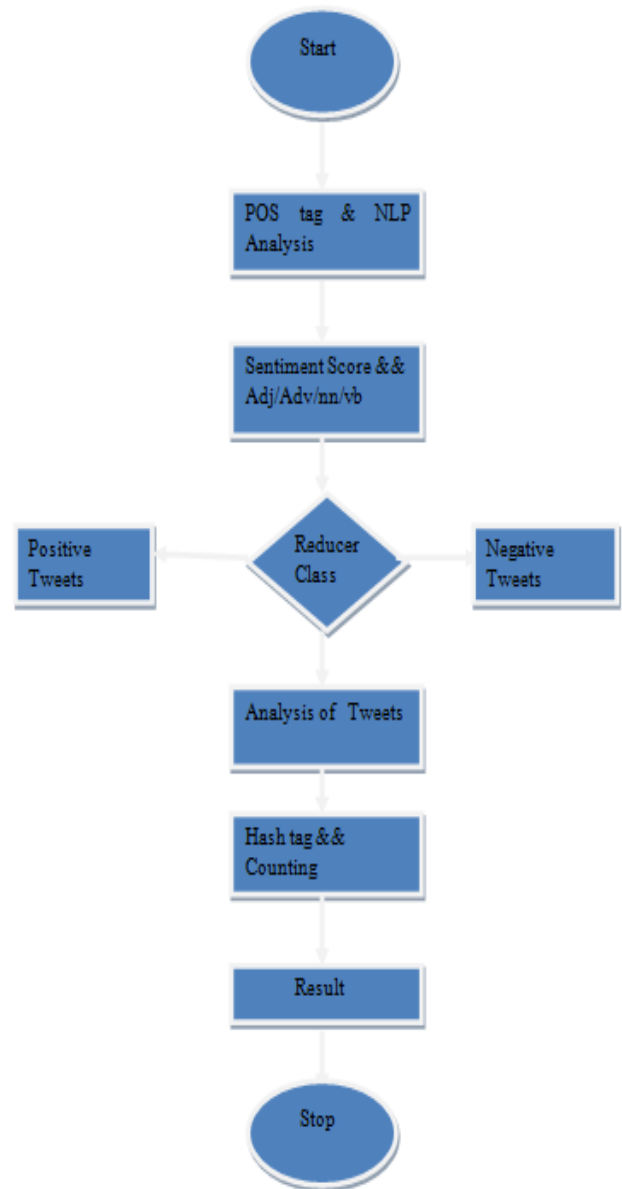

Fig 2: Flowchart of PSWNSWAP

- **Pre-processing Training Data**

Cleaning the data: Since tweets contain several syntactic features that may not be useful for machine learning, the data needs to be cleaned. The module provides these functions:

- Remove quotes - provides the user to choose to remove the quotes () from the text.
- Remove @ - provides a choice of removing the @ symbol, removing the @ along with the user name, or replace the @ and the user name with a word 'USERNAME'.
- Remove # - removes the HashTag.

Some of the ways that data can be represented are feature-based or bag-of-words representation. By features, it is meant that some attributes that are thought to capture the pattern of the data are selected and the entire dataset must be represented in terms of them before it is fed to a machine learning algorithm. Different features such as n-gram presence or n-gram frequency, POS tags, syntactic features, or semantic features can be used. For example, one can use the keyword

lexicons that here saw above as features. Then the dataset can be represented by these features using either their presence or frequency.

There are following steps are used for that purpose:

- Firstly input of that dataset is provided, then a parsing technique is used to provide a part of speech tagging for that data.
- Sentiwordnet   is used to define polarity of the reviews.
- Then frequency of the keywords is also calculated.
- Pruning is performed to refine these reviews.

Then FP (frequent Pattern)   algorithm is used to generate different patterns from the data, for that purpose Pattern technique is used, in that for any topic frequency of the word is calculated if any word having frequency more than 3 will be considered for review and then reviews are categorized as positive or negative. Following are the benefits to apply proposed PSWNSWAP-

1:-   Find positive and negative tweets, comments antonym pair. And to generate score, noun, verb and adjective on using tweets.

2:-   Then find out top # HashTag and bottom # HashTag count and location.(see fig 3,4,5,6,7,8,9)

- **Pseudo code of PSWNSWAP**

Input :-Real time Data Set Twitter

Result :- Algorithm process, parameter computation and find sarcastic twitt &&hash tag

Steps :-

Active Proposed Algo.

Twitter processing ();

If (Scorefunc())

While (Sentence in corpus) do

If (word = "_NN"){

Current _ tag = NLP tag of current word

Add func (Current _ tag);

}

End if

Else if (Word = "_Abj"){

Current_ tag = NLP tag of current word.

Add func (current_ tag);

}

End if

Else if (word = "_VB"){

Current_ tag =NLP tag of current word.

Add func (current_ tag);

}

End if

End

Count = 0

Sarcasmflag = false;

While (word in tweet){

If (word == positive sentiment){

Count = 1;

Continue;

}

Else if (word = negative sentiment){

Sarcasm flag = true

break ;

}

End if

Else

Give tweet is not sarcastic

End

End

While (sentence = #){

Hashtag = find _ has_ tag();

Addfunc(Hashtag);

}   If (sentence ! = #){

Hashtag = find_ hash_ tag ();

Hashtag = "# no hash tag";

Add (Hashtag);

}

End if

Result computation;

Set status = finish and exit;

}

Twitter	place
#TripleTalaq पर सुप्रीम कोर्ट के फैसले का पालन करने लिए नया कानून बनाने की जरूरत नहीं: @Swamy39 #HallaBol लाईव... https://t.co/jU2ppVxo9g	India
RT @Oo7Ranjan: @ZeeTV Really #Bhootu is nice program for everyone...Thanks @ZeeTV	Mumbai
Who's watching with us? #Bhootu https://t.co/gVBbtrIRlQ	Mumbai
RT @Daffodilssssss: Arshiya, the cutest ghost ever♥☺ #Bhootu @ZeeTV	Mumbai
RT @ZeeTVME: #Bhootu and #Kaanha are coming to meet you! Don't forget to watch #Bhootu starts tonight at 7:30 PM only on #ZeeTVME https://t...	Mumbai
RT @OzeeApp: Head to @OzeeApp to watch the first episodes of #Bhootu here: https://t.co/3VVDkrZNUq & #JGTPM here: https://t.co/s4i7WP7LEN h...	Mumbai
They met at Mumbai Clinic to change gender. Marriage and baby next https://t.co/iGufG6kXLY https://t.co/x1bytVen9L	India
Kya aap bhi apni maa se utna hi pyaar karte hai, jitna bhootu karti hai? Dekhiye inki anokhi kahani, Mon-Fri, sha... https://t.co/kxTsnU6hD8	Mumbai
RT @DSYWMP: #Bhopal Raja Bhoj Multiclass Sailing Championship (Ranking) Day 3 https://t.co/8WVrY99LsS https://t.co/nPU0Tn5sAZ	Madhya Pradesh
As #blockchain hits the mainstream, IBM wants to make it easier to use: https://t.co/nEouGUZ4dk @FastCoDesign https://t.co/1gYwEzwR0x	Armonk, New York
RT @Gadgets360: Vivo, Oppo See Strongest Growth in Global Smartphone Sales: Gartner https://t.co/UQ4aJrdW1l	India
RT @Gadgets360: iPhone 8's New Gestures Without Touch ID Spotted in iOS 11 Beta: Report https://t.co/MxIAz0eKeH	India

Fig 3: Identify Tweets with their location

Twitter	HashTag
Will Alastair Cook overtake @imVkohli in the @MRFWorldwide ICC Test batting rankings? https://t.co/5M6e4uX6cr https://t.co/tR1auBYkU0	MyBrand
RT @FriendsofMP: #MilBaancheMP countdown has started, only 4 days to go. More than 2 lakh volunteers will share their valuable knowledge wi...	MilBaancheMP countdown has started, only 4 days to go. More than 2 lakh volunteers will share their valuable knowledge wi...
RT @MPTourism: Enchanting Bamniya kund waterfall near Indore is one of the best waterfalls to visit. Pic courtesy @prithvi2 #MPTourism #Hea...	MPTourism
Which @WWE #RAW Superstar do you want to see FREE AGENT @JohnCena face? VOTE NOW! https://t.co/vHqnet9SGS	RAW Superstar do you want to see FREE AGENT @JohnCena face? VOTE NOW! https://t.co/vHqnet9SGS
RT @Gadgets360: EA, NFL Partner on Madden NFL 18 Fan Competition for Players and Teams https://t.co/ASqMBNamBy	MyBrand
Will Alastair Cook overtake imVkohli in the MRFWorldwide ICC Test batting rankings? https://t.co/GEPYH2BNzM https://t.co/dsKEJZclyD	MyBrand
RT @Gadgets360: Skype for iOS Gets Mood Messages, Improved Sign-In Process, and More https://t.co/iRYSpEGKnS	MyBrand
RT @NDTVProfit: Infosys Crisis: Co-Chairman Ravi Venkatesan Meets Arun Jaitley https://t.co/b2LqzeeKLY	MyBrand

Fig 4: Identify HashTags on Tweets

Top5 #HashTag	Count	Bottom5 #HashTag	Count
MyBrand	1700	proudteamself…	1
#NoHashTag	399	WWCC17	1
Sufyan	16	SLvIND https://t.co/ZASOhnHhl4	1
SLvIND	11	UPDATE	1
JUSTIN	10	Rakhigifts	1

Fig 5: HashTag Counts in Real time

Top HashTag	
HashTag	Location
MyBrand	Dubai
#NoHashTag	Dubai
Sufyan	Kolkata, India
SLvIND	Kolkata, India
JUSTIN	New Delhi

Fig 6: Identify Trend HashTags with location

Collected NN,VB & JJ

Noun	Verb	Adj
#accessoriess	#breaking	15mp
#alertp	's	15mp
#alertp	2dxztztqss	15mp
#ayurvedap	2dxztztqss	1st
#factsmatter	2dxztztqss	1st
#giftss	40d	2day
#jayalalithaap	40d	2day
#jayalalithaap	4v88bw	2hr
#justin	4v88bw	2hr
#justin	4v88bw	2hr
#justinp	4v88bw	2nd

Applied Frequency Count

Noun	Verb	Adj

Fig 7: FP Growth & Pruning

Search reviews

Enter search term | Search

Search Term: car

Positive Reviews

1	son has me looking at cars online i hate car shopping would rather go to the dentist anyone with a good car at a good price to sell
2	nice segment on the google car interesting how ideas like this could disrupt the insurance industry 571xaf3w5r
3	driverless cars now street legal in california a good reason to leave ca no one is behind the wheel http t co 1keemoph
4	saw the google car again asked if i could drive the driver showed me how the car can see awesome http t co htuamykmsy
6	self driving cars powered by uber this google investment is awesome
7	google has a self driving car awesome

Negative Reviews

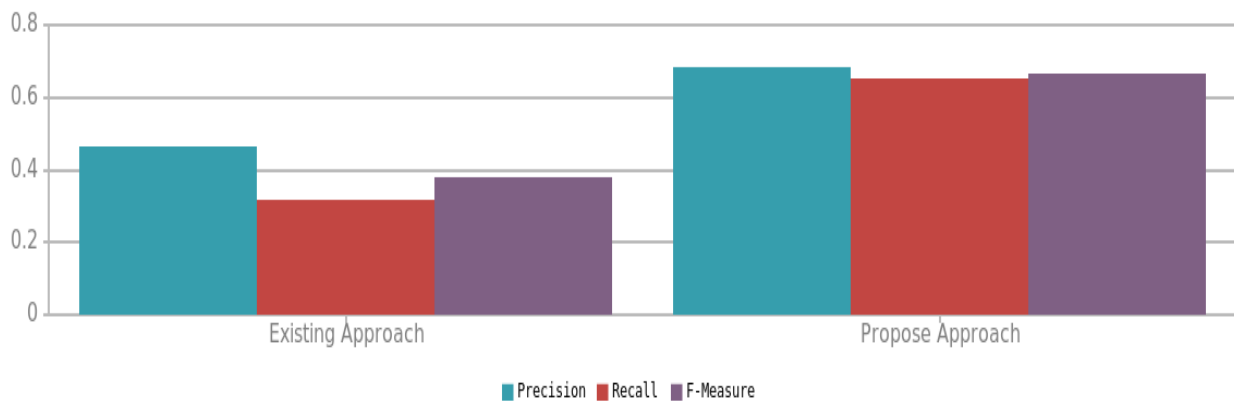|	
2	google driver less car is not viable insurance companies would have no part of it
11	oh hell no i know enough about computers to not trust a driverless car
14	drgitlin yes i just meant in the sense that people on it aren t driving we don t need f n driverless passenger cars
43	driver less cars fuck that i don t want machines ruling the world ftw
45	#BREAKING	Government will not allow driverless car to come to India, says Union minister @nitin_gadkari
48	#BREAKING	Government will not allow driverless car to come to India, says Union minister @nitin_gadkari

Fig 8: Identify Positive and Negative Reviews

Twitter upload for using check Precision , Recall , F-measure			
	Precision	Recall	F-Measure
Existing Approach	0.4666666666666667	0.3181818181818182	0.3783783783783784
Propose Approach	0.6818181818181818	0.6521739130434783	0.6666666666666666

Fig 9: Precision, Recall and F-Measure

iii   St-QAP for Catchments

A distance finding and finding the least measure in between the given scenario is driven in this approach. This research work proposes a new more fast algorithm St-QAP is performed along with Map functionality and Similarity measure score as a more stable value approach. St-QAP is the catchment approach used in order to find the least distance over a MAP. It helps in lowest distance to cover maximum point on a given location. As per observation about the existing technique and their limitations in different terms and scenario's. This work represents a new approach which consumes low travel time and therefore travel costs over the number of available locations. Our algorithm also checks for proper access control using more secure and reliable parameters.
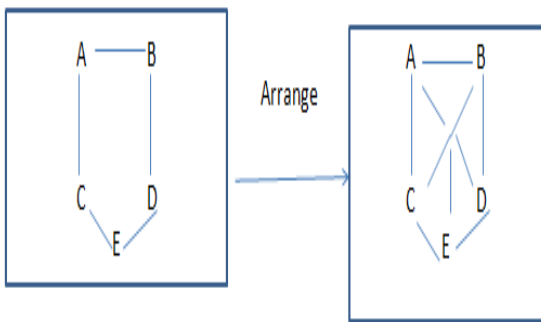

Fig 10: St-QAP arrangement flow diagram

- *Algorithm Pseudo Code :*

Enhance StQAP approach:

Input : Input Tweets, Input brand, City .

Output : Communication process, data matching result MS, Computation time.
Steps :

While(true) do{
Tweet file listing{t1,t2....tN};
DataUploadRequest();
Authentication Storm();
Performing Tweet collection();
FetchTweet();
session Verificatoin :
If(session()==true)
{
Tweet processed();
Input mybrand;
Set status=Active; generate statistics ();
generateRelevantCity();
Apply StQAP();
StQAP function();
Plotting over Map();
}else
{
Status=exit;
generating data for request;
}
Return Computation time;
}
End.

4. IMPLEMENTATION ENVIRONMENT & RESULT ANALYSIS

Java language over NETBEANS IDE simulator with the Twitter API, in Storm framework   is used to implement the proposed methods and a comparison of results with the existing technique is presented. Zookeeper framework installation and further starting its shell, which is going to help in authentication and initialization.   Here we have demonstrated our work in various respects and observed the result and measure the results based on the experiment performance. Both The algorithm are developed in Java language with storm framework, Java net-beans tool setup using Intel i3 processor, 750 GB hard disk, 8 GB RAM. The comparison analysis and execution result shows that our proposed approach outperform best while comparing with existing algorithm

4.1 Computation Time

A training time of a dataset in Java is computed with the help of start and end time class variables defined in the tool and here as we load the dataset and verifies the eligibility and taking their features for consideration or not is the time taking process to identify and to load the data and selection of twitter data and retail location comes under the training time of a dataset, extracting the properties and making them in process format is training time.

$$CT = final\ time\ completion - initial\ time$$

In the figure 11, the comparison between both the technique computation is presented through line chart graphically. The proposed and existing technique is performed with the different real time stream data sets, where the data is processed and following output results were monitored:

Table1: Statically analysis of computation time

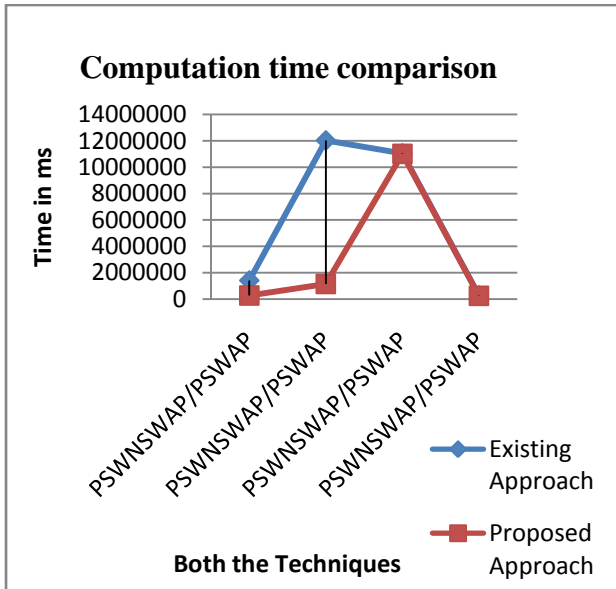| Technique Approach Tweets / Real time stream data sets	Existing Technique PSWAP (Computation time in ms)	Proposed Technique PSWNSWAP (Computation time in ms)
Real time stream data set in 1 Iteration	1412827	279351
Real time stream data set in ii Iteration	12023506	1156053
Real time stream data set in iii Iteration	11045234	11012345
Real time stream data set in iv Iteration	282798	253941

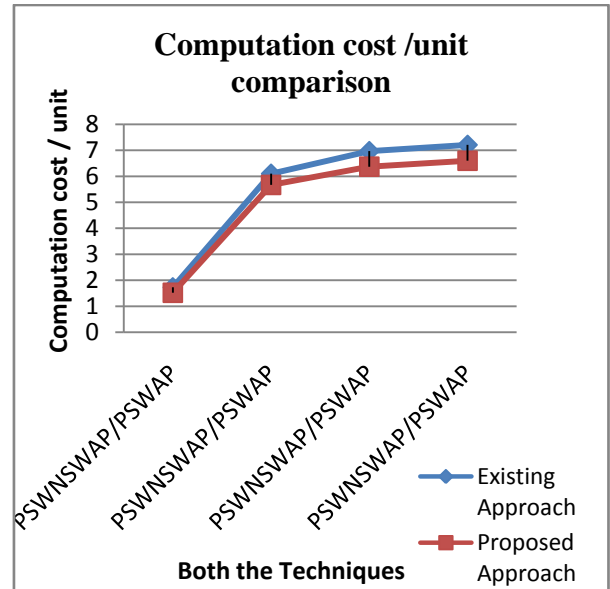**Fig 11: Comparison Line graph for Technique Analysis**

Fig 12: Comparison Line graph for Technique   Analysis

4.2 Computation Cost Comparison

The graph representation shows the efficiency of our proposed algorithm work and it outperform the low computational time, thus the low computational cost with the number of different query and data processing. The proposed and existing technique is performed with the different real time stream data sets, where the data is processed and following output results were monitored:

Table2: Statically analysis of computation cost

| Technique Approach Tweets / Real time stream data sets	Existing Technique PSWAP (Computation cost/Unit)	Proposed Technique PSWNSWAP (Computation cost/ Unit)
Real time stream data set in 1 Iteration	1.73 Cost/Unit	1.52 Cost/Unit
Real time stream data set in ii Iteration	6.10 Cost/Unit	5.67 Cost/Unit
Real time stream data set in iii Iteration	6.97 Cost/Unit	6.37 Cost/Unit
Real time stream data set in iv Iteration	7.21 Cost/Unit	6.60 Cost/Unit
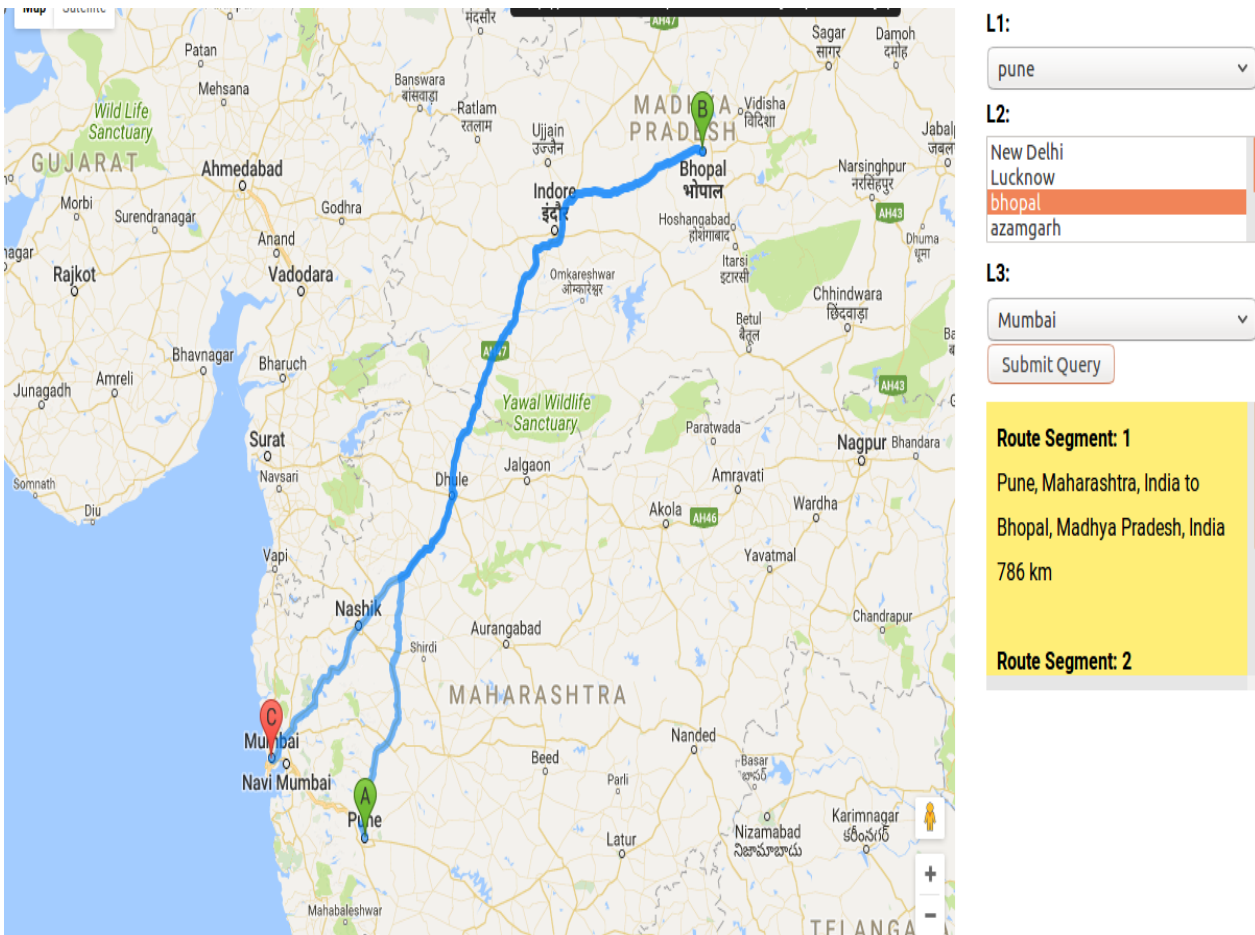
Fig 13: Map Optimization

4.3 Map Optimization

In figure 13, a map page for the user is presented. This page helps companies or brands to show locations for catchments. Which maintains user's action over analysis and details.

4.4 St-QAP Catchments Results

The Objective is to produce an efficient path mapping for new brands to establish in a new area. Here investigate a brand occupancy over an area and performing the visiting area by their defined rule is performed. Minimum travel time and travel cost between location are the measure parameters (fig14, fig 15) for catchments to establish a brand or company in new areas.
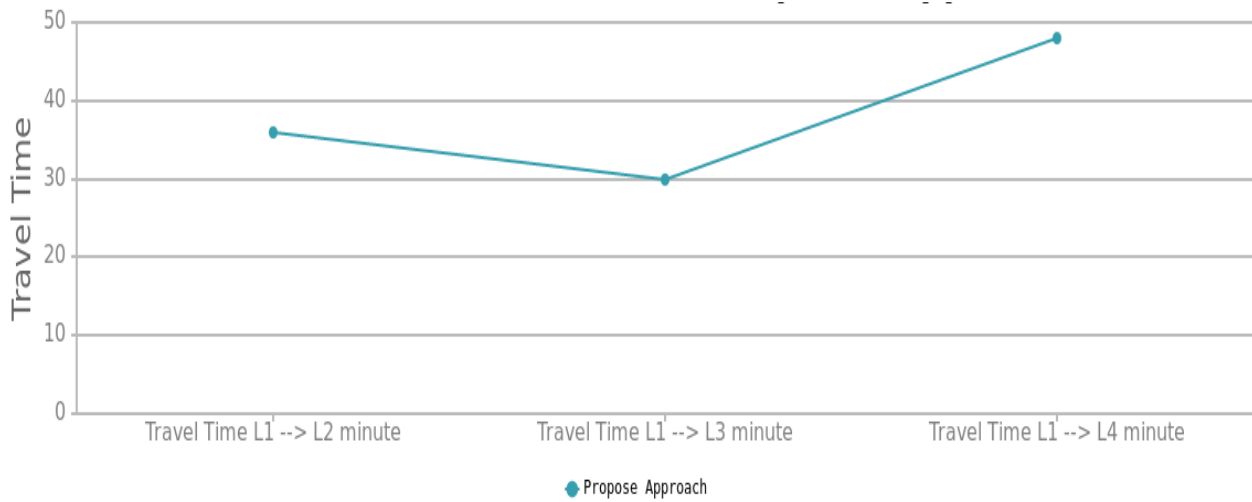
Check Travel Time L1 to			
	L2	L3	L4
Travel Time	36 Minute	30 Minute	48 Minute

Fig 14: Travel Time Between Locations
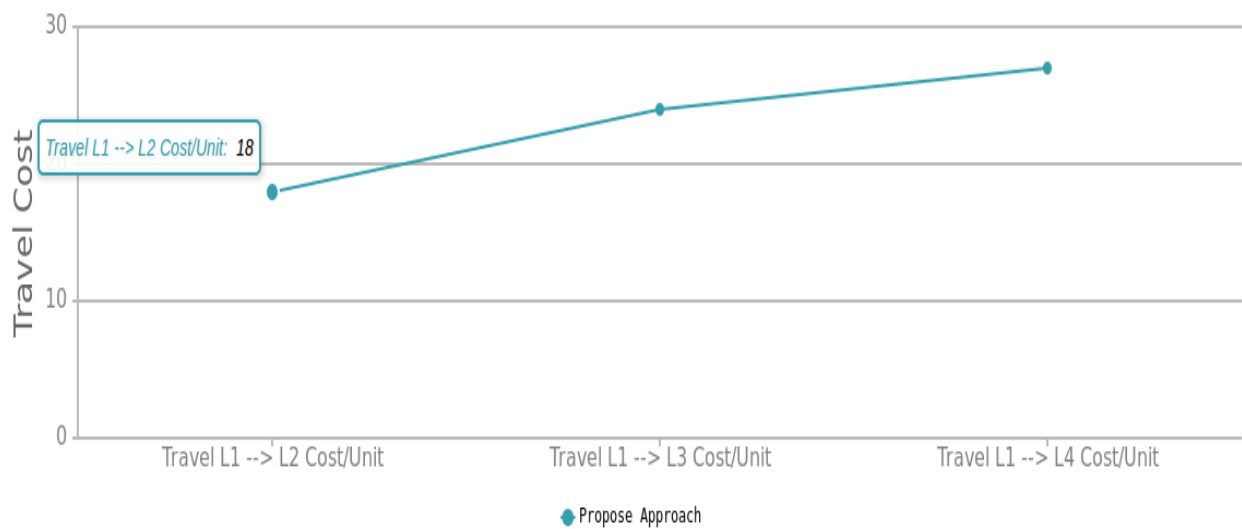
Check Travel Cost L1 to			
	L2	L3	L4
Travel Cost	18 Cost/Unit	24 Cost/Unit	27 Cost/Unit

Fig 15: Travel Cost Between Locations

5. CONCLUSION & FUTURE WORK

5.1 Conclusion

Data processing is a platform which use for different type of analysis, it works with the input data processing and extracting proper knowledge from it. Twitter data generation having its diversity in various fields and tweets over multiple concept help in utilizing for various decisions . Here the problem associate with the previous knowledge extraction approach and twitter analysis is discussed. In various research work, processing and analysis can be performed on static data set. The existing base paper discussed about the static distribution and They also used statical graph analysis for distance computation. The existing data matching algorithm also not much effective . This research work proposed an efficient framework for processing and analysis the massive amount of complex stream data in Real Time. This framework covers the real time data fetching using storm framework, data processing through NLP, use PSWNSWAP algorithm for proper sentiment analysis with comparison parameter as computation time as well as computation cost to compute the comparative analysis and use St-QAP distance measure and finding distance optimization. The proposed algorithm St-QAP takes an input brand name and find proposition for it, with efficient results having parameters travel time and travel cost. The data processing technique produces efficient parameter computation with real time fast and effective process over Zookeeper server.

5.2 Future work

1. In future the real time implementation can be done to determine the largest number of tweets, which can apply over the industry level cloud infrastructure and to find it more secure, reliable than the other alternate available over the web.
2. Categorize implementation with the largest real time stream dataset.
3. This research work will be use in future for various types of analysis such as-
   - Mobility pattern analysis
   - Business Planning & Marketing
   - Flow of business analysis
   - Catchments for Business
   - Social Network Analysis

6. REFERENCES

[1]. Saeed Shahrivari, "Beyond Batch Processing: Towards Real-Time and Streaming Big Data", Computers, Vol. 3, pp. 117.129, 2014.

[2]. Intel IT center. Big Data in the Cloud: Converging Technologies [online]. Available: http://www.intel.com/content/www/us/en/big-data/big-data-cloud-technologies-brief.html. Date accessed: (April 2015).

[3]. Jos´e Antonio Iglesias, Aaron Garc´ıa-Cuerva, Agapito Ledezma and Araceli Sanchis, "Social Network Analysis: Evolving Twitter Mining", 2016 IEEE International Conference on Systems, Man, and Cybernetics • SMC 2016 |October 9-12, 2016 • Budapest, Hungary

[4]. Susan Gunelius, The Data Explosion in 2014 [Online]. Available: http://aci.info/2014/07/12/the-data-explosion-in-2014-minute-by-minute-infographic/ Date accessed: July 12, 2014.

[5]. Ankit Sarawagi, Rajeev Pandey, Raju Barskar, "Big data Applications: A Technical Review" International Journal of Computer Technology & Applications, Vol 8(4) ,431-440 July-August 2017 Available online@www.ijcta.com.

[6]. Xabier Artola, Zuhaitz Beloki, German Rigau, Aitorsoroa "Big data for natural language processing: A Streaming Approach" in knowledge-based systems volume 79, May 2015, pages 36–42.

[7]. Bharati Punjabi, Sonal Honale, "A Research on Big Data Analysis & Processing With Data Mining Techniques", International Journal of Scientific Engineering and Applied Science (IJSEAS) - Volume-1, Issue-9, December 2015.

[8]. Walaa Medhat, Ahmed Hassan, Hoda Korashy,"Sentiment analysis algorithms and applications:A survey"Production and hosting by Elsevier B.V. on behalf of Ain Shams University.accepted 19 April 2014Available online 27 May 2014.

[9]. "Big Data Analytics: managing business data costs and data quality in the capital markets, available online-https://www.emc.com/collateral/software/white-papers/h10834-business-data-cost-optimization-wp.pdf.

[10]. L.Jaba Sheela, "A Review of Sentiment Analysis in Twitter Data Using Hadoop", International Journal of Database Theory and Application, Vol.9, No.1(2016), pp.77-86, http://dx.doi.org/10.14257/ijdta.2016.9.1.07.

[11]. S.K. Bharti n, B.Vachha, R.K.Pradhan, K.S.Babu, S.K.Jena, "Sarcastic sentiment detection in tweets streamed in real time: a big data approach" Contents lists available at Science Direct Digital Communications and Networks, Accepted 15 June 2016 , Available online 12 July 2016

[12]. M. R. Evans, D. Oliver, K. Yang, X. Zhou, S. Shekhar, "Enabling Spatial Big Data via CyberGIS: Opportunities and Challenges," M. F. Goodchild, Ed. S. Wang, CyberGIS: Fostering a New Wave of Geospatial Innovation and Discovery. Springer, 2014.

[13]. Alyson Lloyd, James Cheshire, "Deriving retail centre locations and catchments from geo-tagged Twitter data" , Contents lists available at Science Direct Computers, Environment and Urban Systems, Article Accepted 28 September 2016 , Available online 20 October 2016.

[14]. J.P. Nivash, Ebin Deni Raj , L.D. Dhinesh Babu, M. Nirmala, V. Manoj Kumar, "Analysis on enhancing storm to efficiently process big data in real time', Published in: Computing, Communication and Networking Technologies (ICCCNT), 2014 International Conference, Date of Conference: 11-13 July 2014.Publisher: IEEE, Date Added to IEEE Xplore: 23 April 2015.

[15]. Thanh Trung Nguyen,Tin Khac Vu, Minh Hieu Nguyen,"BFC: High-Performance Distributed Big-File CloudStorage Based On Key-Value Store",978-1-4799-8676-7/15copyright 2015 IEEE,SNPD 2015, June 1-3 2015, Takamatsu, Japan.