IJASRM

International Journal of Advanced Scientific Research and Management, Volume 4 Issue 6, June 2019

www.ijasrm.com

ISSN 2455-6378

Recommendation of Product in Social Network using STD

D.Gandhimathi[1] and A.JohnSanjeevKumar[2]

[1] Ph.D Scholar, Research and Development Center,Bharathiar University, Coimbatore, Tamilnadu, India

[2]Department of MCA,Thiagarajar College of Engineering, Madurai, Tamilnadu, India

Abstract

Social Networks and their role played in daily life increased considerably over the last few years. The advances in location-acquisition techniques and the proliferation of mobile devices have generated an enormous amount of Spatial and Temporal data of user's activities. People can upload a geo-tagged video, photo or text to social networks like Facebook and Twitter. Similar friend can be found using hierarchal clustering. Based on this information, we discover the relationship and cluster the groups. The recommend their interest to the user and suggest the friend in the friend list based on their interest rather than friend's circle. Using Spatio Temporal Data(STD), recommend the product to particular classes of user with respect to time. Nearest point algorithm is used in this work which has KNearest Neighbor with Euclidean distance, that identify the product or points to the particular region of user.

Keywords: *Spatial Temporal Data,Social Network, Twitter,KNearestNeighbor,Euclideandistance.*

1.Introduction

The high quality of Geographical data usage is increased by varies application of GIS & GPS technology. Through these feature incorporate spatial information and analysis by data collection, Integration, Sharing and so on. Social media is a collection of internet based application that allows the users to create, share or exchange information [1]. Most common examples for social media are Social Network Sites (SNS), blogs, micro blogs etc. Social networks are getting closer to our real physical world, social media have advantages that the volume of available data that are time-stamped and exactly placed frequently [2]. Due to this reason, social media data are increasingly recognized as different sources to public approach examination [3]. Social media analysis is a social media mining technique that involves constructing systems to collect and analyze opinions about different products that appear in blog posts, comments, reviews or tweets. People share the exact location and time of their check-ins and are influenced by their friends. The social media data are increasingly recognized as different sources to public approach examination. People share the exact location and time of their check-ins and are influenced by their friends. Twitter could potentially provide a valuable source of social data. The mined processes of these data are can be done to produce planners, marketers and researchers with beneficial instruction and also it is about enterprises and assessment beyond time and space. Reflection of "Social Neighborhood "is Twitter.Twitter is a free social networking micro blogging service that allows registered members to broadcast short posts called *tweets* [4]. Twitter members can broadcast tweets and follow other users' tweets by using multiple platforms and devices[5]. Tweets and replies to tweets can be sent by cell phone text message, desktop client or by posting at the Twitter.com website. The use of twitter social media data as indicators of human activity has received considerable attention in recent years as researchers and businesses seek out alternative data from which to derive insights into population dynamics [6]. It connects to individual in distant space through Geographically. It offers a public application programming interface (API) that enables anyone to request a sample of Tweets according to particular search criteria [7]. Overall work is described in Flowchart. The focus of this paper is identification of STD to suggest product .From the users point identify the friendspoint using nearestpoint group and suggest a list of product using predict point extract,point frequency process.

2.Related works

Taylor (2008) acknowledges that different social networks give us access to our "friends" in different ways, which differentiate our relationships. Whether it is checking your friend's Facebook page to find out what they did this weekend, getting business referrals on LinkedIn, getting the latest news updates in your Twitter feed, or checking out the pictures on Flicker that your cousin took of her trip to the Grand Canyon, you are connecting with people in various ways[8].

Longley et al. (2015) proposed the Tweets that are geolocated facilitate profiling of usage across space as well as time – although we are unaware of any attempt to use any geographically extensive area. Kwak et al. (2010) proposed the research into the content of geotagged Tweets has ranged from identifying new trends across time and space[10].

Greg Linden et al (2003) discussed that Recommendation algorithms provide an effective form of targeted marketing by creating a personalized shopping experience for each customer. For large retailers like Amazon.com, a good recommendation algorithm is scalable over very large customer bases and product catalogs, requires only sub second processing time to generate online recommendations, is able to react immediately to changes in a user's data, and makes compelling recommendations for all users regardless of the number of purchases and ratings. Unlike other algorithms, item-to-item collaborative filtering is able to meet this challenge[11].

Marius Kaminskas et al (2015) adopt an item-centric approach, by establishing a degree of relatedness between any two products in a retailer's product catalog. We identify two techniques for computing item relatedness – one based on textual descriptions of products, and the other based on product co-occurrence in shoppers' browsing histories. The proposed recommender is based on a combination of the two techniques. Being able to compute a relatedness score for any pair of products allows us to implement a service which provides product recommendations when a user is viewing a product web page. The viewed product acts as a 'seed' or 'query' for recommending the top-N most related products from the catalog, which can be displayed in a recommendation panel on the product page[12].

Farman Ullah and Sungchang Lee (2016)proposed a social content recommendation method based on spatial-temporal aware controlled information diffusion modeling in OSNs. Users interact more frequently when they are close to each other geographically, have similar behaviors, and fall into similar demographic categories. Multicriteria-based social ties relationship and temporal-aware probabilistic information diffusion are proposed modeling for controlled information spread maximization in OSNs. The proposed social ties relationship modeling takes into account user spatial information, content trust, opinion similarity, and demographics. A ranking algorithm is used that considers the user ties strength with friends and friends-of-friends to rank users in OSNs and select highly influential injection nodes. These nodes are able to improve social content recommendations, minimize information diffusion time, and maximize information spread. Furthermore, the proposed temporal-aware probabilistic diffusion process categorizes the nodes and diffuses the recommended content to only those users who are highly influential and can enhance information dissemination. The experimental results show the effectiveness of the proposed scheme[13].Ye et al (2010) proposed Geo- measured friend-based collaborative filtering that produces recommendations by using only ratings that are from a uerying user's social-network friends that live in the same city. This technique only addresses user location embedded in ratings. LARS, on the other hand, addresses three possible types of location-based ratings. More importantly, LARS is a complete system (not just a recommendation technique) that employs efficiency and scalability techniques (e.g., merging, splitting, early query termination) necessary for deployment in actual large-scale applications[14].

P.Venetis and H Gonzalez proposed problem of hyper-local place ranking. Given a user location and query string, hyper-local ranking provides a list of top-k points of interest influenced by previously logged directional queries. Hyper-local ranking does not personalize answers to the querying user, i.e., two users issuing the same search term from the same location will receive exactly the same ranked answer set[15].

3.Proposed methodology

Product suggestion to users on social sites is based on their social activities like location tag and time of the status update. Products suggestion means spatial items in that location

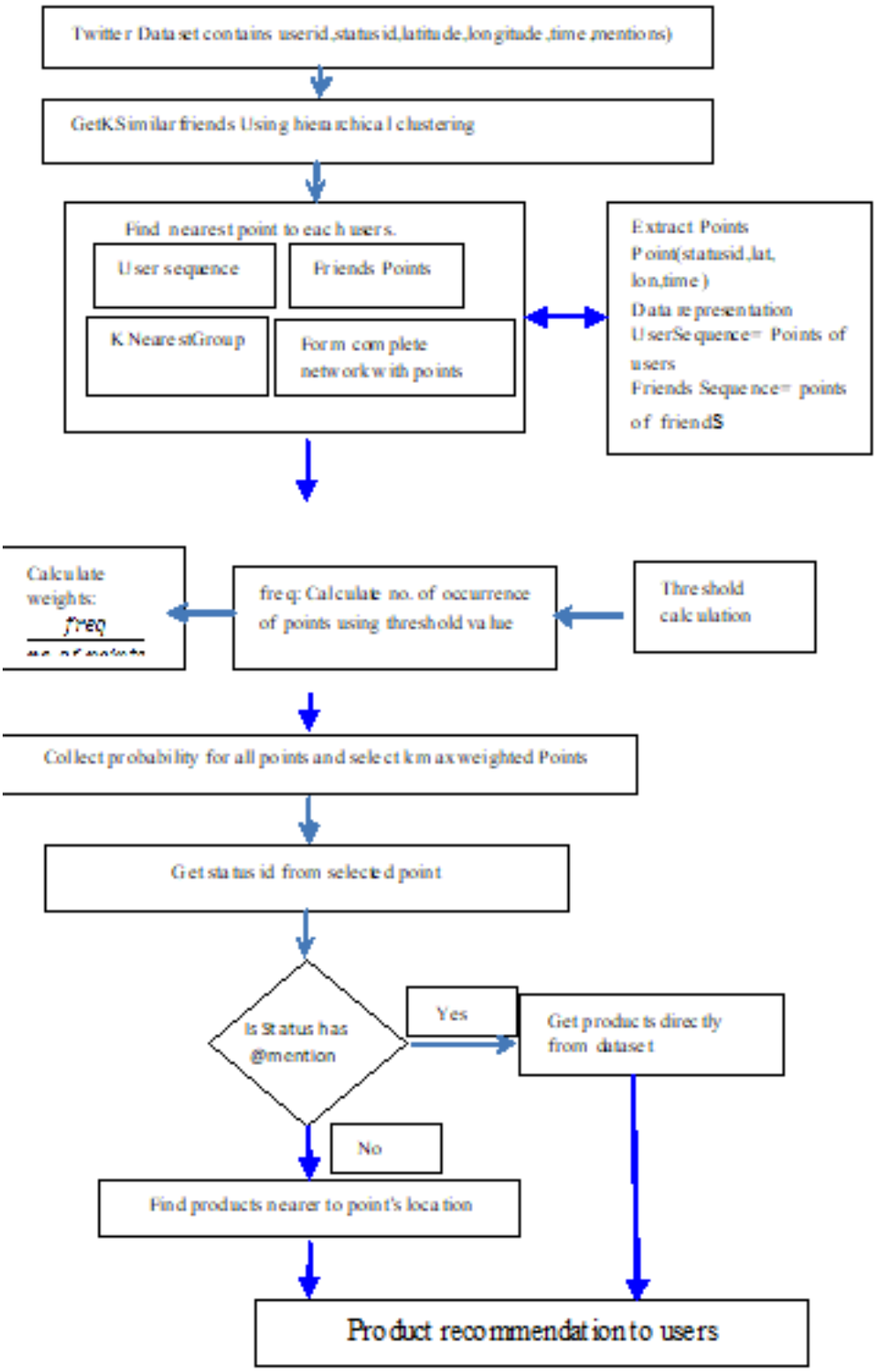

Fig 1. Proposed Flowchart

3.1 Datasets:

Dataset is created by extracting from social network twitter resource URL link

https://api.twitter.com/1.1/statuses/home_timeline.json using twitter4j library. Twitter4J is a one of the Java library for the Twitter API. It is Twitter API binding library for the Java language licensed under Apache License 2.0. With Twitter4J, integrate Java application with the Collect user data from twitter and find similar users using spatial data mining. Dataset contains every users tweets/status userid, statusid, latitude, longitude, time, mentions from status. This represents real tweets and user data. Mention is term used to get the product directly from the database.

Resultset 1

userid	statusid	place	lat	lon	mentions	
947729065	400	Dindukal			@hayat_lens	1538
16120265	1000	SaravanaPatti,Coimbatore	11.00559	73.73658	@lex_watz	1537

Fig 2. Dataset

Prediction of Nth friends in social networks is identified using friend ID to extract their status. If their locations are enabled extract their status. Through similar like group extract the similar followers group. Cluster is formed based on above detail using Hierarchical cluster algorithm . Twitter data are harvested into a simple clustering which can be useful to suggest product for similar group of clustered friend. Hierarchical clustering is used to identify similar users, using location, favorites, mutual friends data. To suggest a product for similar group of friends through STD additional data sets are prepared such as Product data.

3.2 Spatial Temporal Data:

It can be represented as Point. Point ϕ, contains <statusid,latitude,longitude,time>.

3.3 Product data:

Products db contains productsid,name,lat,lon,type.Spatial data is treated as products in most of the social spatial data research. name=unique name of products . lat, long are latitude and longitude of place. type= indicate type of product whether it is place or not. 0 =spatial product 1=non spatial product.

Resultset 1

name	pid	lat	long	fullname
horlicks_nut	1368812			Horlicks nutrie
parle_mom	4578966			moms magic
Beets	4789662			beets gee
GreatDealer	5478964	14.7586	80.01425	Dress GreatD
Karpgam_complex	5682145	9.39138	81.97239	karpagam cor
BlackThunder	12543369	7.59284	79.58185	BlackThunder
sf_treat	12545275			sunfeast trea l

Fig 3.Product data

Twitter service. Twitter4j library is used to get twitter users data like userId, Favourites tweet ids. Location Details Extract from user tweets if location share is enabled. If location is not enabled by user, then it has been taken from user profile.

3.3.1 Implementation Process

1. *SF*=getSimilarFriends();
2. For *Uid* from *SF*
   //SF=similar friends ,Uid= user id ,//UserSeq={$\phi_1,\phi_2,\ldots \phi_c$*} all points of Uid*
   //friendSeq={ $\phi_1,\phi_2,\ldots \phi_f$ *} all points of users from SF except Uid*
   *//*Point ϕ *=<Latitude,Longitude,Time,Statusid>*
3. For ϕ from UserSeq// ϕ *=point*
4. NrG_ϕ =nearestPointsGroup(ϕ,friendSeq) *//*NrG_ϕ *=set of nearest pointsGroup*
   //nearestPointsGroup() function returns points nearest to ϕ
5. End For
6. AllPoints={*UserSeq*∪NrG} *//AllPoints contains list of all points from UserSeq and NrG*
7. For each < ϕ > from UniquePoints
8. Probability Score< ϕ >= $\frac{freq<\phi>}{\text{no of UniquePoints}}$ //Probability score is list contain all points score
9. End For
10. PredictedPoints ={select the k points which has maximum weightage score}
11. For ϕ from PredictedPoints
12. Status= ϕ.statusid;
13. If(Status has @mentions)
14. Get Product from Products db where name=mention
15. Else
16. Get Products nearest to ϕ.latitude and ϕ.longitude from products db
17. End For
18. Suggest products to user Uid.
19. End For

The above algorithm is used to identify similar users, using location, favorites, mutual friends data using KNearest Neighbor with Euclidean distance. The unknown instance represented by some feature vectors as a spatial data can be classify using, the *k*-NN classifier calculates the distances between the point and points in the training data set. The Euclidean distance is used as the distance metric.

Subsequently, it assigns the point to the class among its *k*nearest neighbours (where *k* is an integer).SF represents similar friends list contains user ids. Similar users got as like previous phase using clustering. Select user id from SF. To find products suitable to particular userid from list . *SF* is list contains userids from step 1. *Uid* is user id selected from *SF* for an instance.

3.3.2 Data representation:

UserSeq is a list contains points. *UserSeq* represents sequence of points related to user *Uid*. Each point in the UserSeq contains status details updated by user Uid. friendSeq is a list contains list of points from users in *SF* other than *Uid*. To Find the group of points related to UserSeq ,Each pointϕ from UserSeq is selected and find the nearest points using nearest neighbour algorithm with eculidean distance. Group of points nearest to ϕ from friendSeq is found and stored inNrG_{ϕ}. NrG contains all list of points nearest to all ϕ. This helps to remove outliers which is far away from UserSeq points and predict points related to user.Form a complete network of points All Points is a list contains UserSeq and NrG points. Calculate probability of occurrence of each point from Unique points. Number of occurrence of point divide by total number of points. *UniquePoints is* set of points from All Points where duplicate points are removed. Predict points which have more weightage score. Sort points as per score and select k number of points which has higher score.

3.3.3 To Predict Products:

From predicted point extract statusid(ϕ.statusid) from point and check it has mention. If it has mention can fetch product directly from products database or select product which is nearer to (ϕ.latitude, ϕ.longitude) point. Proposed model predict not only spatial items but also non spatial items by utilizing mentions in the tweets. Mention are part of Tweet which is used to include any Twitter id(e.g-@jionetwork) Mention is unique name for the user. Selected products add to suggested list for the user id. Repeat for other userid from SF and suggest products to each similar user.

3.4 Point Frequency <ϕ>

Frequency of points in calculating number of points occurred in same distance .It is done by calculating location distance which other points in Unique Points(UP) list and count the points which has distance less than threshold value. Frequency of point is calculated using $freq = |di = Loc\_dist(pt, UP_i) \forall di < \lambda|$ calculate location distance between point<pt> and points in UP and then count di less than . Location distance is calculated using : $Loc\_dist(pt_1, pt_2) = \sqrt{(pt_1.lat - pt_2.lat)^2 + (pt_1.lon - pt_2.lon)^2}$

λ Calculation: $\lambda = \frac{\sum_{i=1}^{s} D_i}{s}$ where D is list contains sorted distance. S is arbitrary number less than size of list(D). $ds = \{(UP_i - UP_j) \forall\, i,j\}$ where ds is distance between two points. Distance between all points in UP and store in list $D = \{ds_1, ds_2, \ldots, ds_m\}; where\ ds_1 < ds_2 < \ldots < ds_m$

Freq(<pt>):

$$freq = |di = Loc\_dist(pt, UP_i) \forall di < \lambda|$$
$$\lambda = \frac{\sum_{i=1}^{s} D_i}{s}$$
$$ds = \{(UP_i - UP_j) \forall\, i,j\}$$
$$D = \{ds_1, ds_2, \ldots, ds_m\}; where\ ds_1 < ds_2 < \ldots < ds_m$$

4.Nearest point algorithm :

Nearest points against ϕ is identified using KNearestNeighbor with Euclidean distance.

NearestPointsGroup(ϕ, FriendSeq):
X1=<ϕ.lat> X2=<ϕ.lon> X3=<ϕ.time>
For each point y from FriendSeq
Y1=y.lat; y2=y.lon; y3=y.time;

$$distance = \sqrt{(x1 - y1)^2 + (x2 - y2)^2 + (x3 - y3)^2}$$

Return k minimum distance points{<y>}

Φ is assigned to variable and point from FriendSeq , which is assigned to Y variable.STD such as time and location is used to calculate related points with distance formula.Sort the list and select K minimum distance and corresponding points from FriendSeq. Return the required minimum FriendSeq point.

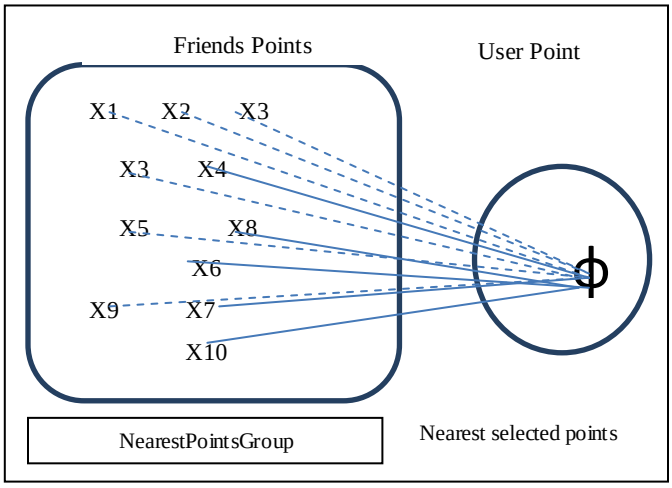

Fig 4. NearestPointsGroup

5.Experimental analysis

In experimental analysis, performance of proposed method is analyzed using java and MYSQL database. The performance and accuracy of the proposed method is evaluated in terms of precision, recall, F-measure and accuracy values.

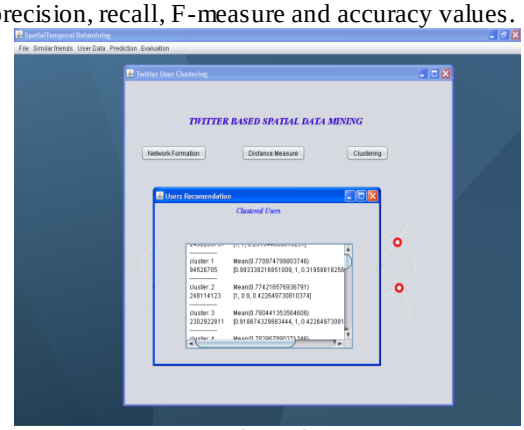

Fig 5: Clustered Users

The Fig 5 illustrates the clustering of users in a group based on the mean value. The cluster 1 has mean value of 0.770, cluster 2 has mean value of 0.774, cluster 3 has mean value of 0.780 ,etc.

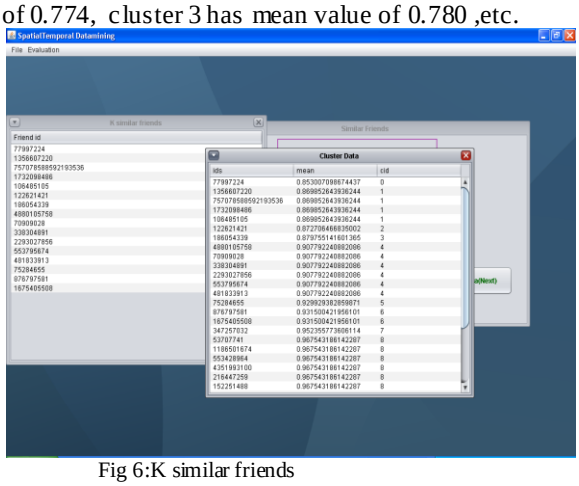

Fig 6:K similar friends

The Fig 6 illustrates the k similar friends, using hierarchical clustering algorithm.

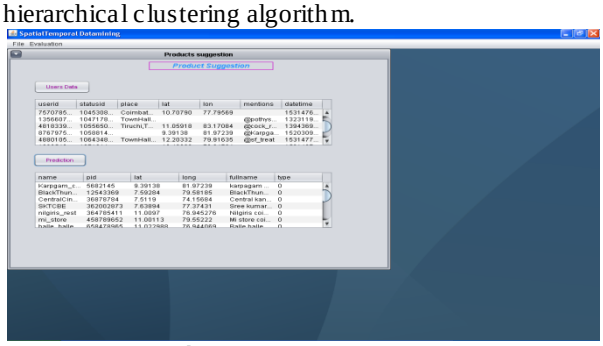

Fig7: Product suggestion

The exchange of product related information can be found in all forms of social media. From different perspective of multiple users in social media-Twitter , ongoing experiment provide us selected list of user from cluster identify the suggested products ,are listed based on the latitude and longitude values, which are represented as STD. Analysis of these relevant data can help a business people to predict the most wanted product in various location.

5.1 Performance metrics

1.Precision value : Precision value specified to the retrieved dataset. This is calculated by the total amount of relevant datasets separated by the total number of resultant datasets.

$$Precision\ value = \frac{True\ Positive}{True\ positive + False\ Positive}$$

2. Recall value : Recall value is specified to as the relevant datasets that are related to the other request search.

$$Recall\ value = \frac{True\ Positive}{False\ \ positive + False\ Negative}$$

3. F-measure : The F-measure is the harmonic mean of precision and recall.

$$F\ measure = 2\frac{Precision \times Recall}{Precision + Recall}$$

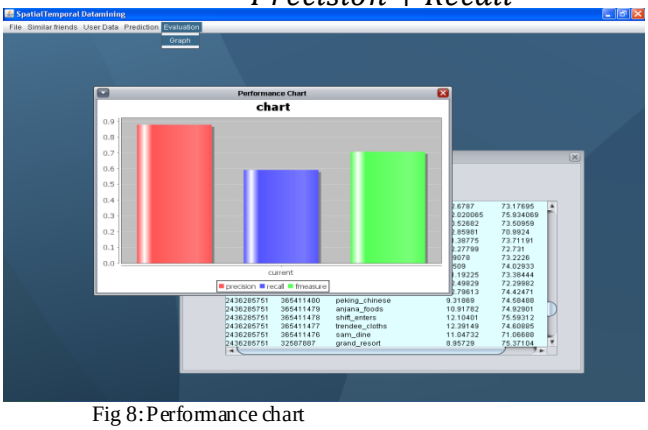

Fig 8: Performance chart

Figure 8 illustrate

6.Conclusion

Valuable information concerning human behavior is available in huge volume of data in Social media platform . Most research has focused on the content of Tweet messages and the characteristics of Twitter users. The analysis of Twitter data demonstrate that aggregative huge amount of message offer valuable approach. Twitter has become a popular data source for opinion mining and trend tracking. Similar product usage is obtained from publicly available data from Twitter. The findings revealed unique traits of interest and likeness across location, and also between Twitters themselves. Product suggestion to users on social sites is based on their social activities like location tag and time of the status update.

Future Enhancement

A frame work of product suggestion through Twitter data can enhance in various dimensions of CRM process. The information available from usage of various customers can progress the marketing specifically for any new customers, through life time value of customers available in Social media.

Reference

[1] TanjumKamboj, ManojDayal "Social Media: Collaborating Web2.0and User-Generated Content (Ugc)" International Multidisciplinary Journal of Applied Research, Volume no. 1, Issue 3, Year-2, October-2014.

[2] YU ZHENG "Trajectory Data Mining: An Overview" ACM Transactions on Intelligent Systems and Technology, Volume no.6, Issue No. 3, May 2015.

[3] M. Nick Hajli "A study of the impact of social media on consumers" International journal of market research volume no. 56 issue 3 January 2013.

[4] Tina P. Singh, RatnaSinha "The Impact of Social Media on Business Growth and Performance in India" International Journal of Research in Management & Business Studies, Volume no. 4, Special Issue -1, January - March 2017.

[5] Mathieu Perreault and Derek Ruths "The Effect of Mobile Platforms on Twitter Content Generation"Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 2011.

[6] Francisco Jose Grajales III, Samuel Sheps, etal., "Social Media: A Review and Tutorial of Applications in Medicine and Health Care" Journal of medical internet research, February 2014.

[7] Jonathan Wheeler "Mining the First 100 Days: Human and Data Ethics in Twitter Research" ***Journal of Librarianship and Scholarly Communication, 2018***

[8] Taylor, D. (2008, Winter2008). Connect you to humanity?. Phi Kappa Phi Forum, 88(4), 18-21.

[9] Longley, P.A., Adnan, M. and Lansley, G. (2015) 'The geotemporal demographics of Twitter usage', *Environ. Plan. A*, Vol. 47, No. 2, pp.465–484.

[10] Kwak, H., Lee, C., Park, H. and Moon, S. (2010) 'What is Twitter, a social network or a news media?', *Proceedings of the 19th International Conference on World Wide Web*, pp.591–600, ACM Press, New York, NY.

[11] Greg Linden, Brent Smith and Jeremy York (2003, Feb.). Item-to-Item Collaborative Filtering" IEEE Internet ComputingPublished by the IEEE Computer Society.pp.71-83

[12] Marius Kaminskas et al.," Product-Seeded and Basket-Seeded Recommendations for Small-Scale Retailers" Journal on Data Semantics, March 2017, Volume 6, Issue 1, pp 3–14

[13] Farman Ullah and Sungchang Lee *"Social Content Recommendation Based on Spatial-Temporal Aware Diffusion Modeling in Social Networks"

[14] ,Symmetry, 2016, volume no.8, Issue no. 9. Page no 1-19.

[15] Ye, M.; Yin, P.; and Lee, W.-C. 2010. Location recommendation for location-based social networks. In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS '10, 458– 461.

[16] P.Venetis and H Gonzalez "Hyper-Local, Directions-Based Ranking of Places", Proceedings of the VLDB Endowment, Vol. 4, No. 5, 2011.