Location Impact Identification Over Unanimous Users

J.S. Kanchana
Department of Information Technology,
K.L.N. College of Engineering,
Madurai, India
kanchjs@gmail.com

V.J. Prasanna
Department of Information Technology,
K.L.N. College of Engineering.
Madurai, India
prasanna0794@gmail.com

S. Prabavathy
Department of Information Technology,
K.L.N. College of Engineering,
Madurai, India
s.praba.pranauv@gmail.com

Dr. S.Sujatha,
Department of Computer Applications
Anna University Regional Centre
BIT Campus, Tiruchirappalli, India
sujathaaut@gmail.com

Abstract— **Social networks provide a framework for connecting users, often by allowing one to find people with common interest in their location. However, users would ideally like the option of finding others who share many of their interests, thus allowing one to find highly-similar, like-minded individuals. This paper describes a way on analysing the impact of location over the like-minded users with the rating and location factor. A user-sign is constructed by establishing set representing resource tag and its rating with the Location based Rating algorithm. Location factor is generated as a text string specifying the category id, location id and its rating average. Further the impact of location of a product in a particular location can be justified by applying a parameter based location influence algorithm**.

Keywords— ***Unanimous users, Location Influence, clustering like-minded individuals.***

I. INTRODUCTION

The web is a space for people to find many things people not only search for information in web but also for people with common interest to share knowledge. Few of the facilities in web are tagging and rating. A people powered metadata is referred as Tag which may be descriptive or subjective. Tags are used in categorising the items. Using the tag, and the frequency of the usage of the resource, the resource is valued and its web presence is increased. The other way of finding the efficient resource in web is with its user rating.

Rating is usually given on a scale of 5 or 10, which is how the user rates the item present in the web. Based on how much users rated and how the users rated, the item's presence in the web is increased. Search engines which are available these days, use the rating as a major keyword to display the item in

the search results. Whenever people search for particular categories of items, they search for the high rated item and also people rate it with their own view on the item. For example, Windows apps store from Microsoft which is an application store for windows powered devices, sort its applications based on the rating provided by users.

This rating based searching idea paved a way to find like-minded using the categories and rating. Many websites use rating as a key factor to promote the resources. Every user has their own view and rate for the item present on the web. The items in the web are categorised with the help of tags. The tag/category is used to identify what type of resource. The rating depicts how the user rate/view the resource. This rating projects the user's point of view on the resource. .So with the help of these tags and rating a user can be uniquely identified by generating a unique location factor for him/her. The location factor is a unique text string which uniquely identifies the interestingness and the mindedness of people over the resource available. The location factor is generated using the Location Based Rating Algorithm.

With these location factor the unanimity (i.e) the like mindedness of people can be found since people with unanimous mind, rate the resources more or less on a similar level. So the location factor which is created using location based rating algorithm is used in finding the unanimity. The

Location influence algorithm is used to check the location impact among the different users/people on the particular location.

The web is a place with huge number of people. So comparing two individuals is only good for few users. So a clustering mechanism is introduced which help in grouping like -minded individuals. A signature parameter is set and people who fall in those category are clustered. A Parameter Based Clustering algorithm is used to create cluster. Those parameters are used as a signature for the group.

In the existing system, the signature creation mechanism was used and the signature was created using tag pop and tag frequency. A fuzzy logic is implemented to check the user's likely mindedness. Also the cluster creation is done with the OWA operator which aggregates the signature of the users.

II. RELATED WORK

A. Adamic and E. Adar on their paper[1] discussed about how to identify similarity by analysing text, links, and mailing list. Considering these items, the usage of the items for two different users (user A and User B) may vary. Also they may have few usage resources in common. With the commonly used resources among the users the similarity is identified. Items/Resources that are unique to a few users are weighted more than commonly occurring items.

G. Smith, in his paper [2] elaborated that Tags are people powered metadata which is used to represent a particular resource available on the web. A tag is a text string which may be subjective or descriptive. A tag represents the type of the resource. People tag the resource on their own interest, which helps in identifying the people's point of view on the resource.

C. Tanner, I. Litvin, and A. Joshi, on their paper[3] proposed that Similarity can also be found using a matrix representation. W\ith the help of the min hashing technique, in which user is represented in rows and the user interests are

represented in column and the hashing algorithms are implemented to check the similarity between the users. This method is suitable if the users and the available resources are limited.

A. Nisgav and B. Patt-Shamir, on their paper [4] discussed that, similar users can be identified using the resource they see and how they see the resource in the web and how many times they see the resource. for example if a user tags a resource as "funny" and the same tag "funny" is done by other person for the same resource, which proves that they both have unanimous view on that resource. On calculating for many resources with many tags the unanimity can be identified.

Chien-Liang Liu, Wen-Hoar Hsaio, Chia-Hoang Lee, Gen-Chi Lu, and Emery Jou ,on their paper[5], discussed that rating of a particular resource, for instance considering a movie, its rating can be evaluated using a sentiment classification. The rating and the comments are semantically reviewed. The positive sentiment and the negative sentiment are compared for a particular resource and the rating is evaluated and summarized.

Yunseok Noh, Yong-Hwan Oh and Seong-Bae [6], discussed that Personalized news article recommendation provides interesting articles for a user based on the users' preference. With increasing use of hand-held devices the interests of users are reflected by their location.

J.S. Kanchana and Dr.S. Sujatha[7],proposed a way on how to identify unanimous people with the tags and rating using the user rating and user signature.

III. PROPOSED WORK

To overcome the drawbacks like high processing overhead and complex processing in the existing system, in this paper we discuss a few mechanisms to find impact of location with the "scale of location influence" factor and a new parameter based algorithm to create cluster. This frame work consists of a location factor generator, a cluster formation and impact analysis. A new factor called "scale of

location influence" is introduced to estimate the location impact. The location factor is generated by means of the Location based Rating algorithm and the scale of influence is found using the location influence algorithm.
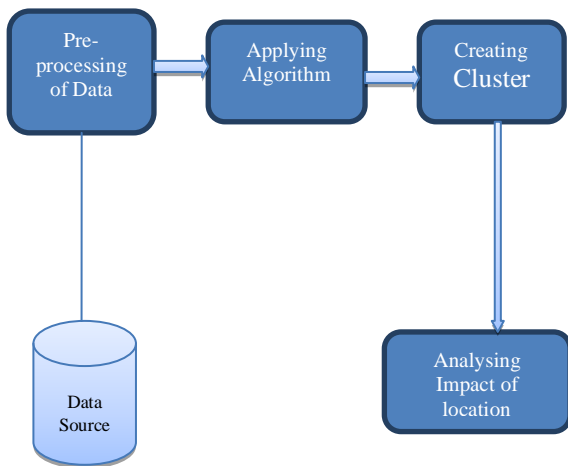
Data Pre-processing:

Data pre-processing is a technique of removing unwanted and noisy data and transforming the required data into the desired format. The data source consist of information about books, users, and the user rating for each book and the desired location of the user . Users who didn't rate any book are eliminated for the next level processing as the signature for the user cannot be created. Also Data is separated into many chunks of data, as it enables easy processing.

Location factor Generation:

A location factor is a text string which uniquely identifies the user characteristics. In this module, A location factor is created for valid users with the help of the category ID, Location information and the Rating given by the user. The Location Based Rating algorithm is used to generate the sign.

$$loc(k)= \pi_0^i (Cat\_ID\ C(i) * Avg\_Rating\ r(i,k,l))$$

$$Avg.\ Rating\ r(i,k,l) = \Sigma_0^i \underline{Rate}_{(i)}$$

Where n is the no of ratings done by user 'k' and 'l' the location factor

Location Based Rating Algorithm

begin

initialise user,catID,loc;

for each user U

    for each category i

        find average rating(U) for category(i);

        ifavgrating(U,i) <= 0

    remove the category for user U;

        else

        generate location factor by catID,AvgRating (u,l);

        end If

    end for {category}

    end for{user}

end

Cluster Formation:

Since web is a place of many people we cluster many people together with a particular parameter which is the signature of that cluster. A cluster consists of unanimous individuals. A cluster is created by setting a particular parameter for individual categories and people who are satisfying the criteria are formed under a common cluster. Those parameters may be either set to high or low as to find liked cluster or disliked cluster.This is done with Parameter based clustering algorithm.

Parameter Based Algorithm:

begin

    for each category n

        set parameter for p(n)

    for each user A

        for each category n

            if (A(n)>p(n)

                in=1;

            else

                in=0;

                break;

        end for {category}

        if(in=1)

            add to cluster;

        else

go to next user;

      end for {user}

end

Impact analysis:

For any kind of product or products there will be an influence of location. People or users in a location should possess same mind by considering some parameters and thus it justifies the influence of location over a product. The impact of location is identified by the location influence algorithm. It is evaluated using the location factor using the location based rating algorithm. The clustered users by every location are taken into consideration. For example, impact of a location ID/State code is found by number of users for each category books and their average rating. The threshold value is set for poor, average and good rating. When all the category books come under same threshold then the maximum of threshold will be taken as the result.

Location Influence Algorithm

begin

      forlocation l

          for each category n

             find r(l,n)

             influence(n)=10-|r(l,n)|

             repeat for each category n

          end for

             SOI(l)=influence(n)/n;

      end for

end

IV. EXPERIMENTAL SETUP

The data set consist of 3 tables, a user information table(2,00,000+) users , a books information table(2,00,000+) books , a rating table with location information (11,00,000 entries) . For efficient processing, the books information other that category and ISBN were eliminated since they are not used for current processing. The same way, after generating the location factor, users who didn't have any location factor are also eliminated since they cannot be used in comparing with other users. To verify the effectiveness of the algorithm, initially location factor was with only 10,000 rating entries which could able to create signature for few 1000's of users and also few users remain unrated. Also the unanimity range was under 3 to 4. When 2,00,000 entries were considered, more user signature were created and also the unanimity level also increased gradually to the scale of 5 to 6. When all the 11,00,000 entries were considered, signature was created for more than 1,82,092 users and the unanimity among the users were increased. The number of people who fall under each cluster was also increased with the increase in the number of ratings considered.

	Category A	Category B	Category C	Catego D
CA	2.4461	2.233	6.4236	2.463
NY	6.363	2.3716	6.3913	6.470
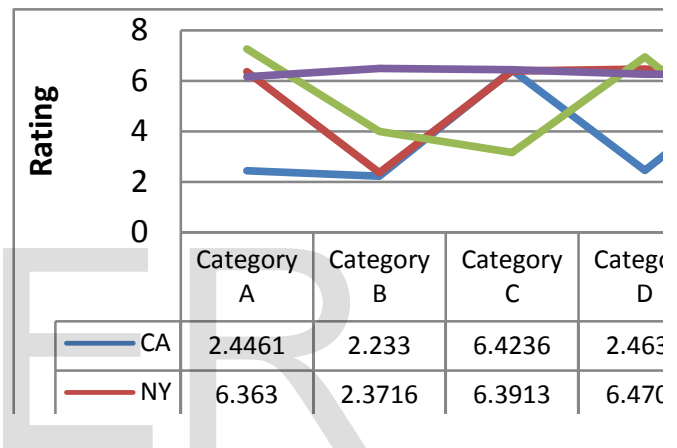
V. CONCLUSION AND FUTURE ENHANCEMENT

The perception of the web has increased due to the introduction of new social platform which are in need of methods and tools to support users' and search for other user groups which communicates their own interests. The advent of social networks, web user groups and other user groups changed the ways of sharing information between the users. In this paper, rating with the location information plays a major role in web. Rating is a process that pulls the users for estimating the content in the web. Users who involved in rating will rate the resources available in the web, based on their perception. Users will represent their interests and likes using rating. This creates an opportunity for using location factor based on rating to represent the users, since web is a place where people search for people who have similar interest with them. To find the location impact, location influence

algorithm is used. Also a new way of clustering the users is proposed as the parameter based clustering mechanism. This paper proposed a way of identifying the impact of location only with five categories of books and four statecodes. In future more categories and more state codes   may be involved which increases the scale of impact. An effective methodology may be designed to manage more no. of categories. It deals with the books domain where the user rating for books is considered. It may be extended to various other domains like food, movies, television shows, Electronic gadgets etc. An effective change in the generation of location factor and the location influence algorithm which increases the scale of influence to an optimistic level. Further the unanimity between the states i.e. the unanimity between the location can be identified effectively.

REFERENCES

[1] L. A. Adamic and E. Adar, "Friends and neighbours on the web," Social Netw., vol. 25, pp. 211–230, 2003.

[2] G. Smith, Tagging: People-Powered Metadata for the Social Web: New Riders, 2008.

[3] C. Tanner, I. Litvin, and A. Joshi, "Social networks: Finding highly similar users and their inherent patterns," Social Netw., 2008.

[4] A. Nisgav and B. Patt-Shamir, "Finding similar users in social networks," Theory Comput. Syst., vol. 49, pp. 720–737, 2011.

[5] Chien-Liang Liu, Wen-Hoar Hsaio, Chia-Hoang Lee, Gen-Chi Lu, and Emery Jou , "Movie Rating and Review Summarization in Mobile Environment", Systems, Man and Cybernetics part C- application and reviews, vol 42, no 3, may 2012

[6] Yunseok Noh, Yong-Hwan Oh and Seong-Bae, "A Location-based Personalized News Recommendation", June 2012

[7] J.S.Kanchana and Dr.S.Sujatha "Clustering Unanimous Web Users Based on Rating and User-signature",VOL. 6, NO. 3, AUGUST 2014