ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 3, Issue 12, December 2015**

# A Survey on Data Mining Based POI Recommendation System Using Geo Tagged Images

Rajebhosale Supriya[1] , Prof. Shilpa Gite[2]

[1]Student, Dept. of M-Tech Computer science, Symbiosis Institute of Technology, Symbiosis International University, Pune, Maharashtra, India.

[2]Professor, Dept. of M-tech. Computer science Symbiosis Institute of Technology, Symbiosis International University, Pune, Maharashtra, India.

**ABSTRACT:** Online Social Network (OSN) has been extensive over the time, however it faces problems such as recommendation of similar locations to users, significantly it is difficult to determine the travel preferences and point-of-interest (POI) of the similar locations based on the tagging. Travel preferences and locations based on the accurate information lacks. The point-of-interest based on flicker images and recommendation is inaccurate and inefficient. In this paper, ID3 is used to overcome data sparsity problem. By this algorithm, system can show exact location and user opinions about the geo location are significantly predicted. It is based on user activity and point-of-interest with geo tag location accuracy and user opinions with comments.

**KEYWORDS :** Text mining, image mining, prediction algorithm, Apriori algorithm, collaborative filtering(CF),data sparsity.

## I. INTRODUCTION

Tagging is an important application of Social websites these days. Whatever object be, the keywords assigned to it gives a good description to the object. It is really easy these days, with the new applications to keep people updated of one's life. Photo sharing applications relatively make task easy by allowing people not just to upload their pictures but also tag their geo location, the moment experience etc. The universal fact of almost each and every person carrying a Smart phone makes this quite challenging when millions of pictures are getting uploaded and shared every single day. Tagging here really needs to be done in a smart manner. [1]

There are many tag recommendation methods already introduced to facilitate users to do the tagging in a proper manner which could help community get a rich data in return. Also the varying sizes of the smart devices makes it difficult to tag the content which can be further resolved with the help of these methods. This personalized assistance of these methods on behalf of users filter and suggest important keywords for tagging. Generally, the user profile is one of the important parameters in designing any tag recommendation method. The geo tags are made in reference to the place where the picture was taken. However, it becomes very difficult for these assisting methods to identify there are multiple places very similar to each other. For example Forts in Maharashtra. The pinnacles of these forts are so similar if seen in pictures, it becomes highly impossible to distinguish between them [1,3].

## II. RELATED WORK

### *A.generic tag recommendation*

Using the classification techniques, and with the help of cosine similarity the method [3] suggests the possible automatic tag from the models learnt with the help of Training data sets. Image Content is considered as the most important for making the tag recommendations. A multi task support vector machine SVM algorithm [4] was designed to check for the associations between the object and the already tagged images. The algorithm works like KNN

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 3, Issue 12, December 2015**

algorithm where it finds the very closely resembled images from the set and then checks for the tags given to them and in this way the tags are recommended.

The textual content cannot be declined or underrated when it comes to make the tag recommendations. On the basis of the relation built between the textual documents the tags can be suggested. This concept was further made strong by doing indexing on such relevant documents [5].

The tags too can be categorized when there is a huge database of the pictures or text with the tagging done. Here, the tags can be categorized into geo locations, context, places, regions, landmarks etc. and further can be used while making the tag recommendations [6].

***B. Tag Recommendation In Product Websites***

The method basically focuses on three major aspects. The tag recommendation can be seen as multi label learning problem where the problem firstly needs to be transformed. For this, Binary Relevance and Label Power set methods [8,9] are preferred. Later to which multi label data needs to be controlled through some algorithms. The method also focuses on recommending tags for the similar objects for which the component is designed by evaluating the cosine similarity with the help of term frequency and inverse document frequency. The method also checks for the terms and after the analysis suggests the tags for that. This is done by establishing the relationship between the tags and terms. This relationship is measured by finding the affinity score between the two. It can be evaluated as follows:

$$\mathbf{Aff\ (tag,t) = n_{t,tag} / n_{tag}}$$

$n_{t,tag}$ : Number of software objects where 't' terms and tag 'tag' both appear.
$n_{tag}$ : Number of objects in which the tag 'tag' appears.

***C. Collaborative Tag Recommendation [1]***

Every user has some interest when it comes to the profile on a particular website. In this tag recommendation method [1] the tags recommended are on the basis of user's interests or preferences. The recommended tags can be multiple and a user can make a choice for the most suitable keyword. A strategy is applied to create a user profile. Firstly the items are selected with the higher values. Then items beyond user's interests are eliminated. Following to which preferences are established between the selected items. Finally the ranking is set between all the preferred tags and further the tags are recommended to the user.

***D. Personalized Tag Recommendation [2]***

This tag recommendation technique works online as well as offline. The offline technique involves the operations like data collection which is done from Flicker that has a huge number of pictures that are tagged with some information regarding picture, the location where it has been captured etc. Along with this the information about the preferences of a user is also stored. The third operation is providing the geo location preferences. The unified space for the visual and the textual stats is now evaluated which eventually is the comparison between visual features and tagging information.

In the online technique, for every picture with a description and a geo location associated is searched for its neighboring tags in user specified space and also the geo specified space. These are now stored in the community set where the semantically relevant picture can be specified with the similar tags.

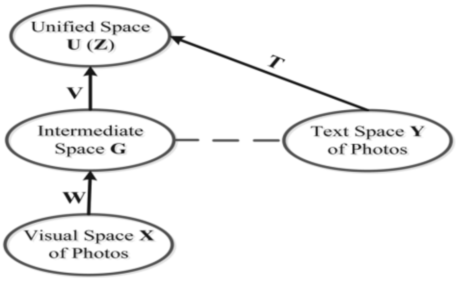


**Fig No 1: The subspace learning [2]**

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

# International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 3, Issue 12, December 2015**

The Fig 1 shows that the structure of the text space and the visual space are different. To bring out the correlation between the Visual Space and the Text Space the mapping of the visual space is done in the Intermediate space G. It is then when the unified space is obtained. There are three transformation matrices applied, two of which are for the visual features and one for the textual feature. An untagged photo is now mapped with the unified space for the user information and the geographic information. After this the relevant tags are generated and recommended.

## III. PROPOSED ALGORITHM

### A .Proposed Technology

ID3 stands for Iterative Dichotomiser3.Algorithm used to generate a decision tree. By using ID3 classifies data. Steps of algorithm.

Split (node,(example))

1. A←The best Attribute for splitting the{example}.
2. Decision attribute for this node←A.
3. For each value of A, Create new child node.
4. Split training{example}to child nodes.
5. For each child /subset:

   If subset is pure: Stop.

   Else: Split (child-node,{subset}).

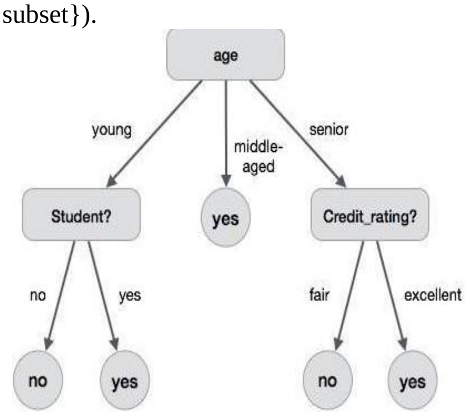


Fig No 2 ID3

## IV. COLLABORATIVE FILTERING MECHANISM

The most important element in making a decision for buying any product is a 'review' written by the customer on product forum. But the problem here is the highly populated customer base due to which the number of reviews is so huge. Finding the most useful reviews from these is always a challenge. With Collaborative Filtering, the best few can be thus filtered and shortlisted which can be further recommended to the customers. Apart from this, on the basis of like mindedness the customer can also be suggested what else he can buy along with this particular product as many customers went for the other products after opting this. This can be now seen in many social media websites like Facebook, where a person is suggested which page to like. Or it can be also seen on the online shopping websites like Amazon, Flipkart where a person is suggested what to buy along with what on the basis of the search history and the past database of other customers. But again, Collaborative Filtering faces so many challenges too. In the race of well known product recommendations the new products which are newly added would not be recommended even after having a good quality. There are so many products which might not have been labeled or rated yet and would thus be out of the recommendations [13, 14]. Many a times, the customer might have a different taste and just because of the majority customer base, the user won't get uncommon recommendations as per his need, but rather would get the regular choices being made. With such rapid search results and search expectations there are many technical challenges

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 3, Issue 12, December 2015**

as well which need to be considered in the process of Collaborative Filtering. There are thousands of people making a query at the same time and it becomes so difficult for the servers to make a result of the clustering and pipelining of the requests which eventually results in the slow response time. This big number of requests affects the computational power and speed of these servers [14].

The user item matrix is used to make the recommendations to a customer and it is really challenging to do it for a new user who has made very few searches or purchases before. This is made through a vast history of searching and buying and since the data is too less of the new users, it becomes a big challenge. Also the items listed recently do not have huge information and the rating is also very less, the reviews posted are also too less, as a result the item recommendation too is a challenge here.

In the sparse databases, the it is also important that the user makes the rating and provides a review if possible which further becomes handy to provide a recommendation to the neigbour customer or the customer with similar interests. However, this problem can be solved to some extent using Singular Value Decomposition Techniques [18]. Even in the reduced space the relation has to be established in these cases for which Latent Semantic Indexing can be used.

### *A.Data Sparsity*

In practice, many commercial recommender systems are used to evaluate very large product sets. The user-item matrix used for collaborative filtering will thus be extremely sparse and the performances of the predictions or recommendations of the CF systems are challenged. The data sparsity challenge appears in several situations, specifically, the cold start problem occurs when a new user or item has just entered the system, it is difficult to find similar ones because there is not enough information (in some literature, the cold start problem is also called the new user problem or new item problem[19]). New items cannot be recommended until some users rate it, and new users are unlikely given good recommendations because of the lack of their rating or purchase history. Coverage can be defined as the percentage of items that the algorithm could provide recommendations for. The reduced coverage problem occurs when the number of users' ratings may be very small compared with the large number of items in the system, and the recommender system may be unable to generate recommendations for them. Neighbours transitivity refers to a problem with sparse databases, in which users with similar tastes may not be identified as such if they have not both rated any of the same items. This could reduce the effectiveness of a recommendation system which relies on comparing users in pairs and therefore generating predictions.

To alleviate the data sparsity problem, many approaches have been proposed. Dimensionality reduction techniques, such as Singular Value Decomposition (SVD), remove unrepresentative or insignificant users or items to reduce the dimensionalities of the user-item matrix directly. The patented Latent Semantic Indexing (LSI) used in information retrieval is based on SVD in which similarity between users is determined by the representation of the users in the reduced space.

## IV.PROPOSED MECHANISM

In the existing paper collaborative filtering algorithm was used but collaborative filtering face the data scarcity problem. We aim to capture how different parts such as user preference, geographical influence and user mobility affect user POI check-in decisions. The key idea is that overall user preferences are the result of the interaction between geographical preferences and interest preferences. Our system aim to effectively capture that interaction.

(a) (b) (c)

An example of a typical user check-in pattern: (a) all the POIs; (b) the user's check-ins over different regions: Maharashtra, Andhra Pradesh , Arunachal Pradesh , Assam, Bihar (c) the user's check-ins in Bihar area.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

# International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 3, Issue 12, December 2015**


**Fig No 2. Basic Architecture of POI**

## V.CONCLUSION

By using the ID3 algorithm the geo-tag and location can be easily optimized for tags effectively. The purpose based metadata gives information about image. The tokenizing gives the relevant user comments that are related to the geo-location and thus optimizes the point of interest of user. By using text mining algorithm, the system can download the comments from OSN and recommend the point of interest according to feedback based on users comments.

In future work we are performing Android Based Implementation can be done.

## REFERENCES

[1] Dihua Xu, Zhijian Wang, Yanli Zhang, Ping Zong. A Collaborative Tag Recommendation based on User Profile. 2012 4th International Conference on Intelligent Human-Machine Systems and Cybernetics.

[2] Jing Liu, Zechao Li, Jinhui Tang, Yu Jiang, and Hanqing Lu. Personalized Geo - specific tag recommendation for photos on social websites. IEEE transactions on multimedia, Vol.16, No.3, April 2014

[3] H. Chen, M. Chang, P. Chang, M. Tien,W. Hsu, and J.Wu. Sheepdog: Group and tag recommendation for flikr photos by automatic search based learning. *Proc. ACM Multimedia*, 2008.

[4] Y. Shen and J. Fan. Leveraging loosely-tagged images and inter-object correlations for tag recommendation. *ACM Multimedia*, 2010

[5] A. Sun, S. S. Bhowmick, and J.-A. Chong, "Social image tag recommendation by concept matching," in *Proc. ACM Multimedia*, 2011.

[6] E. Moxley, J. Kleban, and B. S. Manjunath, "Not all tags are created equal: Learning Flickr tag semantics for global annotation," in *Proc. ICME*, 2009.

[7] Xin Xia, David Lo, Xinyu Wang, and Bo Zhou. Tag Recommendation in Software information sites. MSR 2013, San Francisco, CA, USA

[8] G. Tsoumakas and I. Katakis, "Multi-label classification: An overview," International Journal of Data Warehousing and Mining (IJDWM), vol. 3, no. 3, pp. 1–13, 2007.

[9] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Mining multi-label data," Data mining and knowledge discovery handbook, pp. 667–685, 2010.

[10] Rakesh Agrawal and Ramakrishnan Srikant Fast algorithms for mining association rules in large databases. Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, pages 487-499, Santiago,Chile,September1994.

[11] Saeed Abu-Nimeh, Dario Nappa, Xinlei Wang, Suku Nair, A comparison of machine learning techniques for phishing detection, in: Proceedings of the Anti-Phishing Working Group eCrime Researchers Summit,2007.

[12] G. Tsoumakas and I. Katakis, "Multi-label classification: An overview," International Journal of Data Warehousing and Mining (IJDWM), vol. 3, no. 3, pp. 1–13, 2007.

[13] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Mining multi-label data," Data mining and knowledge discovery handbook, pp. 667–685, 2010.

[14] Rakesh Agrawal and Ramakrishnan Srikant Fast algorithms for mining association rules in large databases. Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, pages 487-499, Santiago,Chile,September1994.

[15] Saeed Abu-Nimeh, Dario Nappa, Xinlei Wang, Suku Nair, A comparison of machine learning techniques for phishing detection, in: Proceedings of the Anti-Phishing Working Group eCrime Researchers Summit,2007.

[16] D. Antoniades, I. Polakis, G. Kontaxis, E. Athanasopoulos, S. Ioannidis, E. P. Markatos, and T. Karagiannis. we.b: The web of short URLs. In Int. World Wide Web Conf. (WWW),2011.

[17] P. Jaccard. The distribution of flora in the alpine zone. The New Phytologist, 11(2):37–50, 1912.

G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6,pp. 734–749, Jun. 2005