International Journal of Computational Intelligence Systems, Vol. 7, No. 4 (August 2014), 650-659

An Approach to Enhance the Quality of Recommendation Using Collaborative Tagging

Latha Banda
Associate Professor, Department of Computer Science
Lingayas University
latha.banda@gmail.com

K.K. Bharadwaj
Professor, School of Computers and System Sciences, JNU
kbharadwaj@gmail.com

Received 1 October 2012; Accepted 26 September 2013

Abstract

Collaborative labeling portrays the process by which numerous users put in metadata in the form of keywords to shared data. Nowadays, collaborative labeling has grown in reputation on the web, on sites that permit users to label bookmarks, photographs and other details. It has been recently become useful and well known as one effective way of classifying items for future search, sharing information, and filtering. So, as to predict the future search of users, we propose a novel collaborative tagging-based page recommendation algorithm using fuzzy classifier. The method consists of three phases: Grouping, Rule Generation Phase and Page Recommendation Phase. In the proposed method, we calculate the resemblance of users in selecting tags and thereby, calculate the nearest neighbors of each user and cluster them. Then, the priority of tags and items for each user is calculated for constructing a Nominal Label Matrix and Nominal Page Matrix. Finally, the fuzzy rules are generated for page recommendation. The experimentation is carried out on delicious datasets and the experimental results ensured that the proposed algorithm has achieved the maximum hit ratio of 6.6% for neighborhood size of 20, which is higher than the existing technique which obtained only 5.5%.

Keywords: Page recommendation, fuzzy classifier, Collaborative labeling, Rule generation.

1. Introduction

The exponential development of web information has brought us into an information overload that leads to discover the most appropriate and fascinating information from it. User-contributed metadata, also known as labeling, offers users to correlate individually the significant keywords or labels with content entities [4, 5], facilitating them to hit upon the content afterward via information predisposed to distinguish or recollect. Labeling helps users "package" information for future information seeking and reuse [6]. Labels are free-form text strings that are attached to an information object, such as a webpage or an image [3]. The process of collaborative labeling produces a set of labels that can be used to describe a resource. This labeling process, made popular by sites like del.icio.us, can be made use of to produce grass root ontology, generally called Folksonomies. Folksonomies insert semantics to resources through a social markup process, facilitating the use of label clouds to illustrate entities, and decide relevancy [1]. Collaborative labeling systems such as del.icio.us and citeulike.org publicly depicts individual users relations between content entities and labels, thereby providing visibility into words others have used to label alike entities [7]. Collaborative labeling is a resource-efficient mechanism to pursue human-driven labeling of web pages, to empower the users of the system [26].

Labeling systems permit a great deal of flexibility and adaptability in arranging information than do proper categorization systems. Labeling is known as a classification process, in comparison to a pre-optimized classification procedure as demonstrated by Semantic Web ontologies [2]. Collaborative labeling is perfectly useful in the absence of a 'librarian' or when there is a large content for a solitary authority to classify; both of these behaviors are correct in the context of the web, where collaborative labeling has become well-liked [8, 9]. Collaborative labeling of this type recommends a motivating substitute to current efforts at semantic web ontologies which have been a center of attention of research by a variety of groups. The collaborative labeling schemes are supposed to gather information from the public and the quality of information users can acquire will augment as the magnitude of data people supplied expands [19]. At present, the Internet's numerous collaborative labeling sites are in existence, but there is the requirement for a service to incorporate the data from the manifold sites, to build a huge and compact set of collaborative data, from which users can have more precise and richer updates than from a particular site. There are a number of websites that are based on collaborative labeling, like del.icio.us, Technorati, and Flickr which permit their users to interpret possessions, like a web page, a blog post, or an image, in general using free sets of labels and facilitating their distribution and recycle [10, 11, 12]. Label suggesting systems suggest applicable labels for an unlabeled user resource [13-18].

Generally, Collaborative tagging is the practice of allowing any user to freely annotate the content with any kind of tags. In some circumstances, such as, when there is no ''librarian" to classify items or there are too many items to classify by a single authority, collaborative tagging is one of the most useful ways of categorizing or indexing content. Moreover, tags are directly published and discussed on the Web and may be applied to any kinds of items, even people. Collaborative tagging can play a key role in sharing content in social networks and also, it is an effective technique to solve the two problems such as, the sparsity problem and the cold start problem that happen in collaborative filtering despite, its success and popularity [25, 26]. In this paper, we have developed a new method for collaborative tagging based on web page recommendation using fuzzy classifier. The proposed algorithm consists of two main steps such as, rule generation phase and page recommendation phase. Initially the resemblance between users is calculated and nearest neighbors is found out and clustered. Then the priority of users for pages and labels are computed. Using these data, a Nominal Label Model and Nominal Page Model are created. These models are used for creating the fuzzy rules required for page recommendation. Based on the fuzzy rules generated, pages are recommended for new users. At last, the experimentation is carried out using delicious datasets to prove the efficiency of the proposed approach. From the comparative analysis, the proposed approach obtained the maximum hit ratio of 6.6% for neighborhood size of 20. This value is higher than the existing technique which obtained the maximum hit ratios is 5.5% neighborhood size of 20. The recall ratio of the proposed approach is 4.78 % for the all the different values of N.

The rest of the paper is organized as follows; Section 2 describes the related work in the field of collaborative labeling and section 3 depicts the problem statement and the terms used in the proposed method are defined in section 4. The new proposed method is described in the section 5. The results and experimentation is explained in section 6. Finally, the conclusion is given in section 7.

2. Review of Related Works

Lots of research works in the field of page recommendation system [30-35] have been carried out earlier. In this section, some of the earlier works related to collaborative labeling of web pages for easy searching and recommendation to new users are discussed briefly. Panagiotis Symeonidis *et al*. [19] developed an integrated framework to model the three categories of entities that existed in a social tagging system: users, items, and tags. A lot of users insert metadata in the form of keywords, to interpret and classify items (songs, pictures, web links, products, etc). Their method could offer three different types of suggestions: 1) tags to users, based on tags, what other users have used for the similar items, 2) items to users, based on tags which they had in common, with other similar users, and 3) users with common social interest, based on common tags on similar items. But, users might have different interests for an item, and items might have numerous facets. So those data were modeled by a 3-order tensor, on which multi-way latent semantic analysis and dimensionality reduction, was carried out using both the Higher Order Singular Value Decomposition (HOSVD) method and the Kernel-SVD smoothing technique.

Iyad Abu Doush *et al*. [20] presented a framework for adding semantics into e-learning system. Their approach depended on two principles. The first principle was the automatic addition of semantic information, while generating the mathematical contents. The second principle was the collaborative tagging, annotation of the e-learning contents, and the use of ontology to classify the e-learning contents. Their scheme encoded the mathematical contents using presentation MathML with RDFa interpretation. The classification permitted students to emphasize and annotate explicit parts of the e-learning contents. The objective was to add sense into the e-learning contents, to append relationships between contents, and to create a framework to ease searching of the contents. The semantic information could be used to respond semantic queries (e.g., SPARQL) to recover information request of a user. Their work was executed as an embedded code into Moodle e-learning system.

Cheng-Lung Huang *et al*. [21] proposed a recommender system that considered the users current tag priorities as social bookmarking, which was an environment in what the user gradually changes interests over time so that the tag data linked with the present temporal period was frequently, more significant than tag data sequentially far from the existing period, which implied that in the social tagging system, the newly tagged items by the user are more appropriate than the older items. Their system included the subsequent stages: grouping alike users into clusters by using an E-M clustering algorithm, finding related resources based on the user's bookmarks, and suggesting the top-N items to the target user. Their study investigated the system's information retrieval performance using a dataset from del.icio.us, a famous social bookmarking web site.

Cheng-Lung Huang and Cheng-Wei Lin [22] proposed a recommender system based on the collaborative folksonomy. The purpose of their system was to recommend Internet resources (such as books, articles, documents, pictures, audio and video) to users. The proposed method was integrated with four steps: creating the user profile based on the tags, combining the identical users into clusters with the help of an agglomerative hierarchical clustering, finding similar resources based on the user's past collections by means of content-based filtering, and suggesting alike items to the target user. Their study examined the system's performance for the dataset collected from "del.icio.us", a famous social bookmarking website.

Yibo Chen *et al.* [23] made use of tag grouping method to cluster the tag according to the resemblance of co-occurrence distributions. Last.FM and MovieLens permitted user to share items which they like with their family, friends, or the online society at bulk. A significant facet of those services was that users themselves annotate the items using so called tags, which illustrated the contents of the items or offer additional contextual and semantical information. There was a lot of researches concern about the use of a tag to develop the excellence of suggestion. Based on it, their paper suggested an approach to group synonymy tags and fusing the association among users-tag with the collaborative filtering algorithms. The results of the empirical evaluation showed that the approach was effectiveness in augmenting recommendation. Yu-Kyung Kang *et al.* [24] have proposed a folksonomy data mining approach based on FCA for discovering hidden knowledge easily from folksonomy. The tagging data of users, tags and resources constitutes a folksonomy, which was the user-driven and bottom-up approach for coordinating and categorizing information on the Web. They illustrated a suitable approach for evaluating tagging data, and determining concealed knowledge from them. Also they have demonstrated how their approach could be applied in the collaborative tagging system through experiment. Their approach could be applied to a number of motivating areas such as social network analysis, semantic web mining and so on.

3. Problem Statement

The dynamics of social labeling has been an active research area in recent years. However the related literature primarily focuses on the problems of label prediction, including cold-start recommendation to facilitate web-based activities. The widespread use and popularity of online collaborative labeling of sites have created new challenges and opportunities for designers of "web entities" such as electronics products, travel itineraries, popular blogs, etc. Various websites today (e.g., Flickr for photos, YouTube for videos, Amazon for different products) encourage users to actively participate by assigning labels to online resources with a purpose to promote their contents and allow users to share, discover and organize them. An increasing number of people are turning to online reviews and user-specified labels to choose among the competing entities. For example, a cell phone that has been labeled lightweight by several users is likely to influence a prospective customer decision in its favor. This creates an opportunity for designers to build entities that are likely, to attract desirable labels when published. In addition to traditional marketplaces like electronics, autos or apparel, label desirability also expands to other various domains.

Suppose we are given a set of entities, each having a set of attributes and a set of user submitted labels (e.g., cell phones on Amazon's website, each described by a set of attributes, and associated user labels). From this training data, for each individual label, we consider a classifier for predicting the label specifying the entities. Label forecast and page recommendation is a recent area of research, and the existence of such classifiers is a key assumption in our work. One of the classifier used in existing work given in [26] is k-NN classifier. K-NN classifier is a data mining classification technique; working based on majority voting and distance matching. But, incorporating the experts' knowledge and fuzzy information can lead to effective classified result than k-NN classifier. So that, enriched information through fuzzy classifier can lead a path of better recommendation to discover hidden information varied in the data set.

4. Definition of Terms

Definition 1 *User - Page Binary Matrix (A)*: If there exists a mapping between 'x' a set of users, $S = \{s_1, s_2, \ldots s_x\}$ and a set of y pages, $P = \{p_1, p_2, \ldots p_y\}$ then the corresponding mapping information can be represented as x by y binary matrix, called as User - Page Matrix(A). The rows and columns of the matrix represent the users and pages respectively. The value $A_{x,y}$ $G = \{g_1, g_2 \ldots g_z\}$ shows a binary 1 value if a user labels a particular page and binary 0 otherwise.

Definition 2 *User - Label Frequency Matrix (B)*: If $G = \{g_1, g_2 \ldots g_z\}$ is a set of z labels, the preference of labels by a set of x users can be shown as x by z matrix which is termed as User-Label Frequency Matrix (B). The rows and columns of the matrix represent the users and labels respectively and each value $B_{x,z}$ in the matrix represents the number of items that have been labeled by the corresponding user with the corresponding label.

Definition 3 *Label - Page Frequency Matrix (C)*: If $G = \{g_1, g_2 \ldots g_z\}$ is a set of z labels and $P = \{p_1, p_2, \ldots p_y\}$ is a set of y items, then the Label-Item Frequency Matrix (C) is z by y matrix in which the rows represent labels and columns represent items and each member of the matrix $C_{z,y}$ shows the number of users, who labeled the corresponding item with the corresponding label.

Definition 4 *Reverse User Frequency (ruf_t)*: Let l be the total number of users in the system and n_t the number of users tagging with a tag t. Then, the inverse user frequency for a tag t, ruf_t, is computed: *Reverse User Frequency,* $ruf_t = \log_{10}(m) * \log_m(l / n_t)$. If all users have tagged using tag t, then the value of ruf_t is zero, $ruf_t = 0$.

Definition 5 *User-User Resemblance Matrix (D)*: When the inverse user frequency is applied to the cosine resemblance technique, the resemblance between two users, u and v, is measured by the following equation: where T is a set of tags and $A_{u,t}$ and $A_{v,t}$ denote the tag frequencies of users u and v each in the user–tag matrix, A. In addition, ruf_t refers to the inverse user frequency of tag t. The resemblance score between two users is in the range of [0, 1]. The higher score a user has, the more similar he/she is to a target user. Finally, for l users, we compute x by x user–user resemblance matrix D.

Definition 6 *User-Label Priority Matrix (E)*: The label preference of users can be represented as x by z matrix, *E,* where rows represent users and columns represent labels.

Definition 7 *User-Page Priority Matrix (F)*: For estimating the priority value of the target user u for page, the sum of the page frequency given by neighbor users on the page are used. Each frequency is weighted by the equivalent resemblance between the target user and each neighbor. The weighted sum leads to a higher prediction value for a more similar user to a target user.

Definition 8 *Nominal Label Matrix (NLM)*: Nominal Label Matrix refers to the set of preferred labels for a particular cluster which is a subset of the total label set, based on the user label priority matrix.

Definition 9 *Nominal Page Matrix* (NPM): Nominal Page Matrix refers to a set of preferred pages for a particular cluster which is a subset of the total page set, based on the user page priority matrix.

5. Proposed Approach to Enhance the Quality of Recommendation Using Collaborative Tagging

The proposed approach to enhance the quality of recommendation using collaborative tagging consists of three phases: (1) Grouping of users, (2) a rule generation phase and (3) the page recommendation phase. In the grouping phase, we input three matrices, user-page matrix, user-label matrix and label-page matrix. Then, we compute the resemblance of users thereby constructing a two dimensional matrix which signifies the user resemblance. This user resemblance matrix will facilitate to infer the K nearest neighbors, of each user by calculating the k highest values in the matrix. These K nearest neighbors, are then clustered to reduce computation time. So, the user resemblance matrix will be converted to a cluster resemblance matrix. Based on the corresponding cluster resemblance matrix, we can infer the user-label priority matrix which denotes the priority of a label by the users in a cluster. Subsequently, the user page priority matrix is also calculated to indicate the priority of a particular user.

From the computed user – label priority matrix and user page priority matrix, the corresponding Nominal Label Matrix (NLM) and Nominal Page Matrix (NPM) are created. These are based on the previous user history. We can infer the fuzzy rules required for the page recommendation from, this Nominal Label Matrix and Nominal Page Matrix. In the page recommendation phase, the items which will be of interest to a user are recommended using the deduced fuzzy rules. The proposed approach is discussed in detail in the following diagram shown in figure 1.

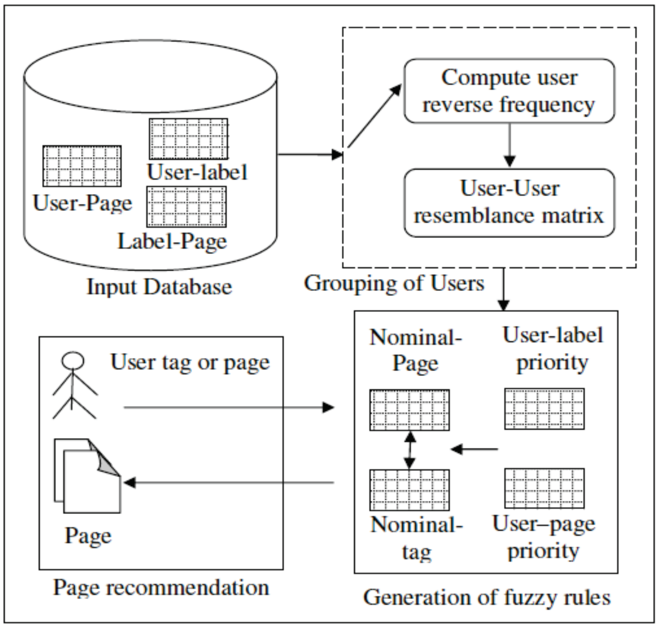

Fig 1. Block diagram of the proposed collaborative tag-based page recommendation

5.1 Grouping of Similar Users

1) Calculation of Reverse User Frequency: As explained in the above section, the input required for proposed approach is in the form of three matrices namely, user-page matrix, user-label matrix and label-page matrix. Once the input is obtained, reverse user frequency is calculated using the following equation (1).

Reverse User Frequency, $ruf_t = \log_m(x/n) * \log_{10}(m)$ (1)

If there are x number of users and n_l be the number of users labeling with a label L, then reverse user frequency can be calculated, as the ratio of logarithm to the base m of users labeling with a particular label to the total number of users. The proposed method can be easily understood with the help of an example. Figure 2(a), (b) and (c) shows the user-page, user-label and label-page matrices respectively which are the inputs for the proposed approach. Based on these data the reverse user frequency, ruf_l , for each label is found out using equation (1) and the user-user resemblance matrix D is calculated.

	P1	P2	P3	P4
U1	1	1		
U2	1	1	1	
U3			1	1
U4				1

(a)

	U1	U2	U3	U4
L1	2	1	1	
L2		3	1	
L3		1	1	
L4	2		1	
L5	1	1		
L6	1	1	1	1
L7		1	2	

(b)

	P1	P2	P3	P4
L1	1	2		1
L2	1	1	2	
L3			2	
L4	1	1		1
L5	2			
L6	2			2
L7			2	1

(c)

Fig.2. (a) user-page, (b) user-label and, (c) label-page

2) Construction of User-User Resemblance Matrix: From the user-user resemblance matrix D, similar users can be clustered together. Clustering is done to reduce the size of the database so as to enable faster calculations as well as to reduce the memory consumption. After clustering, the similarity between the clusters is calculated and k nearest neighbors (clusters) are found out. We can now use this reverse user frequency to find the resemblance between two users by applying this value in cosine resemblance technique. In this method, the resemblance of two users can be calculated as the cosine value of two vectors. The formula used for calculating the resemblance is as follows.

User-User Resemblance Matrix

$$res(u,v) = \cos ine(\vec{u},\vec{v}) = \frac{\sum_{l\varepsilon L}(B_{u,l}\cdot ruf_l)(B_{v,l}\cdot ruf_l)}{\sqrt{\sum_{l\varepsilon L}(B_{u,l}\cdot ruf_l)^2}\sqrt{\sum_{l\varepsilon L}(B_{v,l}\cdot ruf_l)^2}} \tag{2}$$

Where, L is a set of labels and $B_{u,l}$ and $B_{v,l}$ denote the label frequencies of users u and v each in the user–label matrix, A. In addition, ruf_l refers to the reverse user frequency of label l. The resemblance score between two users will be in the range of [0, 1]. The higher score a user has, the more related he/she is to a target user. Finally, for x users, we compute a x by x user–user resemblance matrix D. From the resulting matrix, we could find the K Nearest Neighbors of each user based on the values identified. Figure 3 shows an example for a user-user similarity matrix *D*.

	U1	U2	U3	U4
U1	1	0.129	0.338	0.278
U2	0.129	1	0.654	0.781
U3	0.338	0.654	1	0.549
U4	0.278	0.781	0.549	1

Fig 3. User-User resemblance matrix

5.2 Rule Generation for Tagging and Page Recommendation

1) Construction of User Label Priority Matrix: For estimating the priority value of the target user u for label, the sum of the label frequency given by neighbor users on the label is used. Each frequency is weighted by the equivalent resemblance between the target user and each neighbor. The weighted sum leads to a higher prediction value for a more related user to a target user. The measurement of how much the target user u prefers label is given by:

User Label Priority Matrix $$N_{u,l} = \sum_{i\varepsilon KNN(u)} (B_{i,l})\cdot res(u,i) \tag{3}$$

where KNN(u) is a set of k nearest neighbors of user u, and $B_{i,l}$ is the label frequency of neighbor user i on label l. The $res(u,i)$ represents the resemblance between users, u and i, which is calculated as mentioned in Eq. (2). The label priority of users can be represented as a matrix, N, where rows represent users and columns represent labels. An entire x by z user–label priority matrix, N, can be filled in using Eq. (3). Fig.4. shows the user label priority matrix that contains label priority of group users obtained after finding nearest neighbors. Finding of nearest neighbors of users are explained in the previous section 5.1. Since the users are clustered, the size of the user label priority matrix is reduced here as compared with User-User resemblance matrix. Figure 4 shows the label priority of every group of user (here, two groups) based on the equation 3. The computation of user–label priority matrix is an intermediate step of generating fuzzy rule matching.

	L_1	L_2	L_3	L_4	L_5	L_6	L_7
C_1	0.46	0.72	0.46	0.33	0.12	0.46	0.82
C_2	0.91	0.65	0.65	0.91	0.12	0.72	1.3

Fig 4. user label priority matrix

2) Construction of User Page Priority Matrix: For estimating the priority value of the target user u for page, the sum of the page frequency is given by the, neighbor users on the page that are used. Each frequency is weighted by the equivalent resemblance between the target user and each neighbor. The weighted sum leads to a higher prediction value for a more related user to a target user. The measurement of how much the target user u prefers page is given by:

$$\text{User Page Priority Matrix } F_{u,p} = \sum_{i\varepsilon KNN(u)} (A_{i,p}) \cdot res(u,i) \quad (4)$$

Where, KNN (u) is a set of k nearest neighbors of user u , and $A_{i,p}$ is the page frequency of neighbor user i on page p . The $res(u,i)$ represents the resemblance between users, u and i , which is calculated as mentioned in Eq. (2). The Page priority of users can be represented as a matrix, N, where rows represent users and columns represent pages. An entire x by y user–page priority matrix, F, can be filled in using Eq. (4). Fig. 5 shows the user page priority matrix that is obtained for two groups of users.

	P_1	P_2	P_3	P_4
C_1	0.13	0.13	0.47	0.34
C_2	0.13	0.13	0.65	0.65

Fig 5. user page priority matrix

3) Generation of Nominal Label Matrix: The basic idea of generating nominal label starts from, assuming that a target user is likely to prefer labels that have been used by similar users, or by the target user before. After calculating the priority of clusters on labels, a Nominal Label Matrix can be generated. We are giving the user label priority matrix, N, and a threshold value 'h' which determines the nominal labels. The threshold 'h' represents the model size. The output we get from this step is a two dimensional user-label matrix, R. The two dimensional matrix R, is converted into a binary valued matrix $R_{u,l}$ which, implies the uth user for the l^{th} label is denoted as 1 if the corresponding priority value $N_{u,l}$ is greater than the 'h' highest priority value in the uth row of N, and denoted as zero otherwise. The labels which are marked as 1 for each user in the nominal label matrix are taken as the nominal label model of each user. These are the preferred label set for a particular user. Like this, the label set which is of higher priority for each user is calculated.

C_1	L_1,L_2,L_6,L_5
C_2	L_3,L_7,L_1,L_4

Fig 6. Nominal Label Matrix

4) Generation of Nominal Page Matrix: The basic idea of generating nominal page matrix starts from assuming, that a target user is likely to prefer pages that have been used by, similar users or by the target user before. After calculating the priority of clusters on pages, a Nominal Page Matrix can be generated. We are giving the user page priority matrix, *F*, described in 6.1.2 and a threshold value 'h' which determines the nominal pages. The threshold 'h' represents the model size. The output we get from this step is a two dimensional user-page matrix, *Q*. The two dimensional matrix *Q*, is converted into a binary valued matrix $Q_{u,p}$ which implies the uth user for the l^{th} label is denoted as 1 if the corresponding priority value $F_{u,p}$ is greater than the 'h' highest priority value in the uth row of *Q* and denoted as zero otherwise. The pages which are marked as 1 for each user in the nominal page matrix is taken as the nominal page model of each user. These are the preferred page set for a particular user. Like this, the page set which is of higher priority for each user is calculated. Figure 7 shows the Nominal Page Model.

C_1	**P_1,P_3,P_4**
C_2	**P_2,P_3,P_4**

Fig 7. Normal Page matrix

5.3 Page Recommendation Using Fuzzy Rule Matching

The final step in our algorithm is to generate a recommendation such as, a list of top N pages that the target user will prefer the most. In our research, a page recommendation by collaborative labeling is regarded as a categorization dilemma. Based on a Nominal Label Matrix and Nominal Page Matrix for each user, top-N pages are recommended with the help of *Fuzzy Classifier*. Fuzzy classification is the process of classifying the elements into a fuzzy set [28] in which the membership function are derived by the truth value of a fuzzy propositional function. Then, some authors [29] have utilized this for classification due to the features of Soft labeling, Interpretability, Limited data, available expertise.

In fuzzy classifier, Rule generation and rule weighting are significant steps for developing fuzzy-based decision support scheme. The deviation vectors, Nominal Label Model and Nominal Page Model acquired from the preceding step are used here to generate the decision rules that specified the priority of labels and pages of users. The rules are generated automatically from, the two deviation vectors that enclose the deviation of each characteristic contrasting two classes. Decision rules are created from each element by associating the corresponding elements, of both vectors from the equivalent size deviation vector. Once the fuzzy rules are derived, the recommendation of pages or label is carried out using the fuzzy classifier. Here, the rules are matched with the initial user behavior of the test input so that, the recommendation is identified from the class label using the fuzzy inference procedure.

6. Results & Discussion

The evaluation of the page recommendation algorithm via collaborative tagging is presented in this section. The proposed algorithm is implemented using JDK tool and experiments were performed on core2duo 3.0 GHz, 4 RAM computers, running a MS-Window 2007 server.

6.1. Dataset and Evaluation Metrics

Dataset: This dataset [27] contains social networking, bookmarking, and tagging information from a set of 2K users from Delicious social bookmarking system. The dataset contains 1867 users, 69226 URLs and 53388 tags. To evaluate the quality of the recommendation, the dataset was divided into two sets; 80% of the data was used as a training set and 20% of the data was used as a test set.

Evaluation metrics: The evaluation of the page recommendation algorithm will be done using hit ratio and recall measures. Hit ratio identifies the similar number of pages in test data with the pages recommendation provided by the algorithm. Recall identifies effectiveness of the algorithm in providing the different number of recommendations.

$$HitRatio(u) = \frac{|Test_u \cap TopN_U|}{Test_u}$$

$$\mathrm{Re}call = \frac{\sum_{u=1}^{k} Hitratio(u)}{k_u} * 100$$

$Test_u$ → A set of the item list of a target user u in the test data

$TopN_U$ → Top-N recommended item list for the user u.

6.2 Performance Evaluation

The performance of the proposed algorithm of web pages recommendation is analyzed with the two different evaluation metrics and computation time. These evaluations are carried out for various neighborhood size (k) and the Top-N recommendation. ***Effectiveness:*** The effectiveness of the collaborative tagging-based algorithm is analyzed using three tables. From the table 1, the hit ratio of the proposed approach is high for neighborhood size of 20, signifies that the whenever the k is increased, the hit ratios of the proposed approach is also varied. The results given in table is taken for N=20. Table 2 presents the hit ratio of different number of recommendation pages. Similarly, table 3 presents the recall of the proposed approach for various N values. Here, the recall ratios of the proposed approach is 4.78 % for the all the different values of N.

Table 1. Neighborhood size vs Hit ratio

Neighborhood size (k)	Hit ratio (%)
4	1.20
8	1.34
12	3.97
16	5.78
20	6.62

Table 2. N-recommendation vs Hit ratio

N-recommendation	Hit ratio (%)
15	3.03
20	3.03
25	6.62
30	6.62
35	6.62

Table 3. N-recommendation vs Recall

N-recommendation	Recall (%)
15	4.780771263
20	4.780771263
25	4.780771263
30	4.780771263
35	4.780771263

Efficiency: The efficiency of the proposed approach is extensively analyzed with the help of table given in 4 and 5. Here, the computation time is computed for various neighborhood size and the Top N-recommendation. From table 4, the proposed approach took 0.86 sec for the neighborhood size of 20 and 2.22 sec has been taken for N=35

Table 4. Neighborhood size vs time

Neighborhood size (k)	Time (sec)
4	0.861
8	0.86
12	0.85
16	0.875
20	0.864

Table 5. N-recommendation vs Time

N-recommendation	Time (sec)
15	0.547
20	0.85
25	1.168
30	1.486
35	2.275

Comparison: In figure 8, recall is computed for the various neighborhood size (k) of proposed algorithm and the recent existing technique [26]. From the figure 8, we can identify that, the proposed approach obtained the maximum hit ratio of 6.6% for neighborhood size of 20. This value is higher than the existing technique which, obtained the maximum hit ratios is 5.5% for k=20. Here, the performance is improved for the proposed approach for the high value of neighborhood size. If we take the neighbors of the test user as high, then the performance is automatically improved.

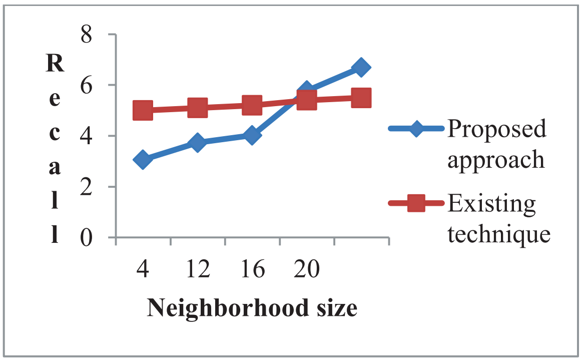

Fig 8. Neighborhood size vs Recall

7. Conclusion

We have developed a new method, for web page recommendation using collaborative tagging, which is a technique to solve the two problems such as, sparsity problem and cold start problem. The proposed algorithm contains three major steps such as, grouping of users, rule generation and page recommendation. Initially, the grouping has been done using the similarity measure by considering user, tag and page related matrices. Then, the priority of users for pages and labels are computed. Using these data, a Nominal Label Model and Nominal Page Model are created for creating the fuzzy rules required for page recommendation. Based on the fuzzy rules generated, pages and label are recommended for new users. At last, the experimentation was performed using delicious dataset to validate the proposed approach and the result signifies that, the proposed approach obtained the performance improvement in recall compared with existing technique. The future work can be done in a way that, i) utilizing soft computing techniques to choose a best set of neighbors, 2) Practical constraints like, time spent, quality can be incorporated into the page recommendation.

References

1. I. A. Doush, F. Alkhateeb, E. Al. Maghayreh, I. Alsmadi and S. Samarah, Collaborative Tagging, and Searching Mathematics in E-Learning, *International Journal of Advanced Computer Science and Applications*, **2** (4) (2011) 30-39.
2. H. Halpin, V. Robu and H. Shepherd, The Complex Dynamics of Collaborative Tagging, *International World Wide Web Conference Committee*, (2007) pp. 211-220,.
3. Samarth Swarup and Les Gasser, "*Collaborative Tagging as Information Foraging*", Technical Report, pp. 1-7, 2008.
4. S. Golder and B. A. Huberman, Usage patterns of collaborative tagging systems, *Journal of Information Science,* **32**(2) (2006) 198–208.
5. J. Vig, M. Soukup, S. Sen and John Reidl, Tag expression: tagging with feeling, *In Proceedings of the 23nd annual ACM symposium on User interface software and technology*, (2010), pp. 323-332.
6. L. M. Markus, Toward a theory of knowledge reuse: Types of knowledge reuse situations and factors in reuse success, *Journal of MIS*, **18**(1) (2001) 57 – 93.
7. E. Rader and R. Wash, Influences on Tag Choices in del.icio.us, *In Proceedings of CSCW*, (2008), pp. 239-248.
8. L. B. Marinho and L. S. Thieme, Collaborative Tag Recommendations, *Data Analysis Machine Learning and Applications*, **37**(5) (2005) 533-540.

9. S. A. Golder and B. A. Huberman, Usage patterns of collaborative tagging systems, *Journal of Information Science*, **32** (2) (2006) 198–208.
10. D. A. Cernea, E. D. Moral, and J. E. L. Gayo, SOAF: Semantic Indexing System Based on Collaborative Tagging, *Interdisciplinary Journal of E-Learning and Learning Objects*, **4** (2008).
11. G. W. Furnas, C. Fake, L. V. Ahn, J. Schachter, S Golder, K. Fox, M. Davis, Marlow, and M. Naaman, 2006. Why do tagging systems work?, *In Proceedings of CHI Extended Abstracts*, (2006), pp. 36-39.
12. C.Musto, F. Narducci, M. De. Gemmis, P. Lops, G. Semeraro, A Tag Recommender System Exploiting User and Community Behavior, *In Proceedings of the Workshops on Recommender Systems the Social Web*, (2009).
13. M. Tatu, M. Srikanth, and T. D'Silva, Tag recommendations using book- mark content,*In Proceedings of ECML PKDD Discovery Challenge (RSDC08)*, (2008), pp.96-107.
14. A. Nanopoulos, Item Recommendation in Collaborative Tagging Systems, *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, **41** (4) (2011) 760 - 771.
15. R. Jaschke, L. Marinho, A. Hotho, L. Schmidt-Thieme, and G. Stumme, Tag recommendations in Folksonomies, *In Proceedings of Workshop of Lernen - Wissensentdeckung – Adaptivit*, pp. (2007), pp.13–20.
16. G. Mishne, Autotag: a collaborative approach to automated tag assignment for weblog posts, *In Proceedings of the 15th international conference on World Wide Web*, (2006), pp. 953–954.
17. S.Ming-Sheng and Z. Zi-Ke, Diffusion-Based Recommendation in Collaborative Tagging Systems, *Chineese Physics Letters*, **26**(11) (2009).
18. C.Musto, F. Narducci, M. de Gemmis, P. Lops, and G. Semeraro, STaR: a Social Tag Recommender System , *In Proceedings of ECML PKDD Discovery Challenge*, (2009), Vol. 497, pp. 215-227.
19. P. Symeonidis, A. Nanopoulos, and Y. Manolopoulos, A Unified Framework for Providing Recommendations in Social Tagging Systems Based on Ternary Semantic Analysis, *IEEE Transactions On Knowledge And Data Engineering*, (2010), Vol. 22, pp. 1-14.
20. I. A. Doush, F. Alkhateeb, E. Al. Maghayreh, I. Alsmadi and S. Samarah, Annotations, Collaborative Tagging, and Searching Mathematics in E-Learning, *(IJACSA) International Journal of Advanced Computer Science and Applications*, Vol. 2, No. 4, pp. 30-39, 2011.
21. C. Huang, H. Chien, M. Conyette, Folksonomy-based Recommender Systems with User's Recent Preferences, *World Academy of Science, Engineering and Technology*, (2011), Vol. 78, pp. 183-187.
22. C. Huang and C. Lin, Collaborative and Content-based Recommender System for Social Bookmarking Website , *World Academy of Science, Engineering and Technology*, (2010), Vol. 68, pp. 748-753.
23. Y. Chen, C. Wu, X. Guo, M. Xie and F. Long, Augmenting Collaborative Recommendation by Grouping Synonymy Tags, *Journal of Computational Information Systems*, **7**(4) (2011) 1350-1357.
24. Y. Kang, S. Hwang, and K.Yang, FCA-based conceptual knowledge discovery in Folksonomy, *World Academy of Science, Engineering and Technology*, (2009), Vol. 53, pp. 842-846.
25. M. Das, G. Das, V. Hristidis, Leveraging Collaborative Tagging for Web Item Design, *In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, (2011), pp. 538-546.
26. H. Kim, A. Ji, I. Ha, G. Jo, Collaborative filtering based on collaborative tagging for enhancing the quality of recommendation, *Electronic Commerce Research and Applications*. **9(**2010) 73–83.
27. Dataset from the framework of the 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec 2011), "http://ir.ii.uam.es/hetrec2011/datasets.html" and http://www.grouplens.org/node/462.
28. L. A. Zadeh, (1965). Fuzzy sets. *Information and Control* (8), pp. 338–353.
29. D. Amo, A. Montero, J. & Cutello, V., On the principles of fuzzy classification. *Proc. 18th North American Fuzzy Information Processing Society Annual Conf,* (New York, 1999), pp. 675 – 679.
30. Adomavicius G, Tuzhilin A (2005) Personalization toward the next generation of recommender systems: a survey of the state-of-theart and possible extensions. IEEE Trans Knowl Data Eng 17(6):734–749.
31. Latha Banda and Kamal K. Bharadwaj "Improving Scalability Issues in Collaborative Filtering based on Collaborative Tagging using Genre Interestingness Measure", ARTCom 2012, LNEE pp. 403–409, 2012.(Published by Springer Digital Library).
32. Heung-Nam Kim, Ae-Ttie Ji, Inay Ha, Geun-Sik Jo. collaborative filtering based on collaborative tagging for enhancing the quality of recommendation. Vol 9, issue 1, jan-feb 2010, page 73-83.
33. Shweta Tyagi, Kamal K. Bharadwaj: Enhancing collaborative filtering recommendations by utilizing multi-objective particle swarm optimization embedded association rule mining. Swarm and Evolutionary Computation 13: 1-12 (2013)
34. Vibhor Kant, Kamal K. Bharadwaj: Integrating Collaborative and Reclusive Methods for Effective Recommendations: A Fuzzy Bayesian Approach. Int. J. Intell. Syst. 28(11): 1099-1123 (2013)
35. Vibhor Kant, Kamal K. Bharadwaj: Fuzzy Computational Models of Trust and Distrust for Enhanced Recommendations. Int. J. Intell. Syst. 28(4): 332-365 (2013)
36. Latha Banda, Kamal K. Bharadwaj, Evaluation of Collaborative Filtering Based on Tagging with Diffusion Similarity Using Gradual Decay Approach, *Smart Innovation, Systems and Technologies*, Vol. 27, (2014), 421–428.