International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 6 | ISSN : 2456-3307

Rule Based Tag Recommendation for Images

Harshada A. Karande[1], Prof. A. G. Phakatkar[2]
[1]Student, PICT, Pune, Maharashtra, India
[2]Assistant Professor, PICT, Pune, Maharashtra, India

ABSTRACT

Tag recommendation is focused on recommending useful tags to a user who is searching for images. The recommendation of additional tags to partially annotated resources, which may be based on either personalized or collective knowledge. Analyzed tag collection can be stored in different abstraction level by applying GenIO (Generalized input-output) algorithm in generalized association rule mining for it. An association between two levels can be found in Wordnet lexical database. Tag selection and Ranking algorithm assign the desirable tags to the image. In rule-based tag recommendations use of generalization that improves performance significantly.
Keywords: Generalized Association Rule Mining, Rule-Based Systems, Tag Recommendation.

I. INTRODUCTION

Tagging systems have become major infrastructures on the web. They allow users to create tags that annotate and categorize content and share them with other users, very helpful in particular for searching multimedia content.

However, as tagging is not constrained by a controlled vocabulary and annotation guidelines, tags tend to be noisy and sparse. Especially new resources annotated by only a few users have often rather idiosyncratic tags that do not show a common perspective useful for search.

Tagging is very useful for users to figure out other users with similar interests within a given category. Users with similar interests might post similar tags and similar resources might have similar tags posted to them. Tagging refers to the behaviour of bookmarking resources with keywords (tags). In recent years, social tagging is becoming more and more popular in many web 2.0 applications where users can freely annotate various resources, such as web pages, academic publications, and multimedia objects. Tag recommendation is concerned with suggesting relevant tags to the users, which they could potentially use to bookmark the web resources they visited. The motivation of tag recommendation is twofold. From the systems perspective, it aims at expanding the set of tags annotating a resource, thus enriches the index of resources. For the users perspective, like all other recommendation systems, the target is to improve the recommendation in her tagging process.

II. LITERATURE SURVEY

Cagliero L, et al. [1] presents a novel personalized tag recommendation system that discovers and exploits generalized association rules, that is, tag correlations holding in different abstraction levels, to find additional tags to suggest. The use of generalized rules relevantly improves the effectiveness of traditional rule-based systems in coping with sparse tag collections, because: (i) correlations hidden at the level of individual tags may be anyhow figured out at higher abstraction levels and (ii) low-level tag associations discovered from collective data may be exploited to specialize high-level associations discovered in the user-specific context.

R. Jaschke, et al. [3] present two tag recommendation algorithms: an adaptation of user-based collaborative filtering and a graph-based recommender built on top of FolkRank, an adaptation of the well-known PageRank algorithm that can cope with undirected triadic hyperedges. They check and compare both algorithms on large-scale real life datasets and show that both provide better results than non-personalized baseline methods.

Y. Toshihiko, et al. [2] address two emerging yet challenges, problems in social media: (1) scoring the text tags in terms of the influence of the numbers of views, comments, and favorable ratings of images and videos on content sharing services, and (2) recommending additional tags to increase such popularity-related numbers. For these purposes, they present the FolkPopularityRank algorithm, which can score text tags based on their ability to influence the popularity-related numbers. The FolkPopularityRank algorithm is inspired by the PageRank and FolkRank algorithms, but the scores of the tags are calculated not only by the co-occurrence of the tags, but also by considering the popularity-related numbers of the content. This is the first attempt to recommending tags that can enhance popularity attributes of social media.

T. H. Hanh, et al. [4] present a personalized content-aware image tag recommendation approach that combines both historical tagging information and image based features in a factorization model. They apply state of the art deep learning image classification and object detection techniques to extract powerful features from the images.

Chitra, et al. [9] propose P-TAG, a method which automatically generates personalized tags for web pages. Upon browsing a web page, P-TAG produces keywords relevant both to its textual content, but also to the data residing on the surfers Desktop, thus expressing a personalized viewpoint. Empirical evaluations with several algorithms pursuing this approach showed very promising results.

C. Chen, et al. [12] suggest the problem of personalized, interactive tag recommendation for Flickr: While a user enters/selects new tags for a particular picture, the system suggests related tags to her, based on the tags that she or 2 other people have used in the past along with (some of) the tags already entered. The suggested tags are dynamically updated with every additional tag entered/selected.

III. EXISTING SYSTEM

In an existing system, a collaborative filtering method is used and integrated in a scalable architecture. The subject of an interactive recommendation of the Flickr brand is addressed. The proposed marks are first selected from a set of previously assigned tags based on coexistence measures.

According to the recommendation, the user’s file is reduced so that the proposal is more specific. However, coexistence methods are questioned by the lack of data because computational complexity may increase exponentially depending on the number of marks, or the score associated with each tag may not be directly comparable. In other approaches there is

IV. PROPOSED SYSTEM

We propose a new personalized photo tag recommendation system. given With a series of images and user-defined tags, the system assigns new related tags the user-specific (i.e. tags that are already commented To each photo of the same user), the rest of the collective knowledge (i.e. Note from other users). Figure 1. Architectural blocks are shown.

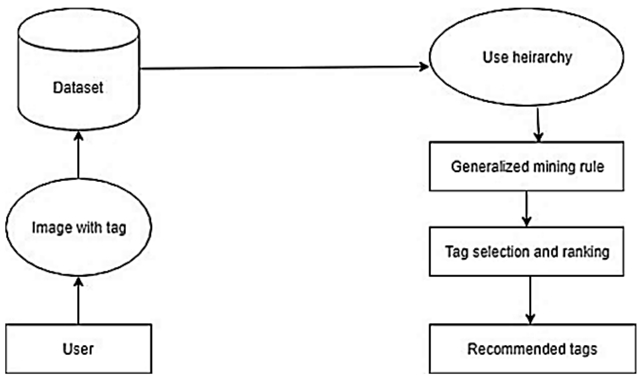

Figure 1. Architecture Diagram

Pre-processing: This block is explain to create a collection of previous tag annotations It is suitable for generalized rule mining process. Tag set Transaction data format. It is associated with a given picture by the user and includes the appropriate set of tags assigned. The history tag collection is also used to derive a series of generalization hierarchies established Wordnet lexical database [Wordnet 2012].

Generalized association rule mining: This block focuses on finding high-level tags in the form of generalized related rules and correlation of transaction. Discover the tag correlation with different levels of abstraction. Two different rule sets: (i) User-defined rule set, previous annotations from users targeted by recommendations, (ii) Collection Rule Set, This includes generalized rules derived from previous annotations it was created by another user. GenIO is a generalized itemset mining Apriori algorithm that addresses the discovery of a smart subset of all the possible frequent (generalized) itemsets.

Tag selection and ranking: A set of photos and tags has already been given, This block is intended to generate a ranked list with additional tags that you can suggest. To accomplish this goal is to create generalized rules for user-specific and collective rules. An already assigned tag was selected. The ranking of the proposed tag is the set of selected rules is based on highest strength of tags and tags from Genio algorithm.

V. SYSTEM ANALYSIS

The GENIO Algorithm: GENIO is a generalized itemset mining algorithm that addresses the discovery of a smart subset of all the possible frequent (generalized) itemsets.

Given a source dataset, a set of generalization hierarchies, and minimum support threshold *minsup*, it discovers all frequent not generalized itemsets and all frequent generalized itemsets having at least an infrequent descendant, that is, a descendant that does not satisfy *minsup*. To achieve this goal, the generalization process is support driven, that is, it generalizes an itemset only if it is infrequent with respect to the minimum support threshold.

Algorithm 1. GenIO Generalized Itemset Discoverer
Input: minimum support *minsup*, dataset D
Output: set of generalized frequent itemsets L

k=1, L= ϕ
repeat
Gen = ϕ // generalized itemset container
if k = 1 then
C_1= scan D and count support for each item
else
C_k= candidate-generation(L_{k-1})
scan D and count support for each c ∈ C_k
end if
for all c ∈ C_k **do**
new-itemset = apply RR ∩ c
update Gen with *new-itemset*
if support of c < *minsup* **then**
mark *new-itemset*
end if
end for

17: L_k = (items in C_k whose support = *minsup*) U (mark itemset of Gen whose support *minsup*)
18: k=k+1
19: **until** $L_2 \neq \phi$
20 : **return** L

Tag selection: Given a photo p_i, a set of user-defined tags $T(p_i,u_j)$ assigned by user u_j to p_i and the sets of generalized rules $R_T(u_j)$ and $R_T(-u_j)$ mined, respectively, from $T(u_j)$ and $T(-u_j)$, this block entails the selection and the ranking of the additional tags to recommend to u_j for p_i.

Algorithm 2: Tag Selection
Input: the user-specific rule set $R_T(u_j)$, the collective rule set $R_T(-u_j)$ and the user-specified tags $T(p_i, u_j)$.
Output: the tag selection C
1) Covered rules (u_j) = select pertinent user-specific rules ($R_T(-u_j)$,$T(p_i, u_j)$).
2) Covered rules($-u_j$) = select pertinent collective rules ($R_T(-u_j)$,$T(p_i, u_j)$).
3) **for all** user-specific rules R in covered rules(u_j) do
4) insert tags in R.*consequent* into C
5) **for all** generalized tags g in C do
6) **for all** collective rules R_2 in covered rules($-u_j$) do
7) **if** R_2.*consequent* includes any tag t^* in g.*leafdescendant* **then**
8) insert t^* in C
9) **end if**
10) **end for**
11) **end for**
12) **end for**
13) **remove** generalized tags from C
14) **return** C

An analysis of a real-life use-case for our system by the discovered generalized tag associations. Figure 2 shows Real life example of hierarchy.

VI. RESULT AND ANALYSIS

To evaluate the performance of both our recommendation system and its competitors, we exploited three standard information retrieval metrics, let Q be the set of relevant tags, namely the tags really assigned by the user to the test photo and C the tag set recommended by the system under evaluation. The adopted evaluation measures are defined as follows.

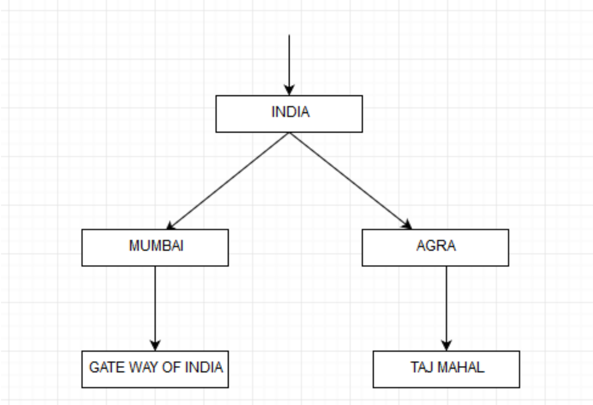

Figure 2. Real life example

Mean Reciprocal Rank (MRR): This measure captures the ability of the system to return a relevant tag (i.e., a held-out tag) at the top of the ranking. The measure is averaged over all the photos in the testing collection and is computed by

$$MRR = max_{q \in Q} \frac{1}{c_q}$$

Success at Rank k (S@k): This measure evaluates the probability of finding a relevant tag among the top-k recommended tags. It is averaged over all the test photos and is defined as follows.

$$S@k = \begin{cases} 1 & \text{if } Q \cap C_k \neq \phi \\ 0 & \text{otherwise} \end{cases} \quad where$$

$q \in Q$ is a relevant tag and C_k is the set of the top-k recommended tags.

Precision at Rank k (P@k): This metric evaluates the Percentage of relevant tags over the set of retrieved ones. The measure, averaged over all test photos, is defined as follows.

$$P@k = \frac{|Q \cap C_k|}{|Q|}$$

Our system on the real-life collection by varying the minimum support and confidence threshold enforced during the generalized rule mining process. The confidence threshold may slightly affect the recommendation system performance. By enforcing very low confidence threshold values (30%), a large amount of low-confidence rules are selected.

Differently, when increasing the confidence threshold a more selective pruning of the low-quality rules may allow enhancing the recommender system performance.

As an extreme case, when enforcing very high confidence thresholds (e.g., 90%), rule pruning selectivity becomes too high to generate a considerable amount of interesting patterns.

The system analyses the user-specific and collective knowledge bases to suggest additional tags to recommend. By setting the standard configuration (minimum support threshold *minsup*=40%, minimum confidence threshold *minconf*=30%) the following strong rule is discovered by our system from the collective transactional tag set.

Following figure 3 is the example of tag recommendation for "Mi" suggested tags with minimum support(30%) and confidence(40%).

Figure 3. Tags with highest confidence

Best values of support and confidence threshold actually depend on the analysed data distribution. For instance, when coping with the benchmark dataset the best minimum support threshold values are around 40%, because the analysed dataset is relatively sparse.

VII. CONCLUSION AND FUTURE ENHANCEMENT

Tag recommendation is one of the challenging problem in data mining. In this project, we proposed a system which can recommend the additional tags from the images which provided by the user with a tag. User-defined and collected tags can be used in the hierarchy. Genio algorithm is used in extraction of strong association rules. Tags can be selected from the user and collective data hierarchy. Top ranked tags consider for recommendation. We demonstrated the efficiency, effectiveness and usability of solution experimental results. From the results, we see that our system could recommend over 90% of related tags.

Our system has so far not been concerned with the analysis of the textual content related to the annotated Web resources (e.g., photo descriptions, related blogs, or articles). We plan to extend it by also considering the user-generated textual content coming from social networks and online communities. Furthermore, to take the evolution of photo annotations over time we will investigate the integration of incremental rule mining approaches as well.

VIII. REFERENCES

[1]. L. Cagliero, A. Fiori and Grimaud, "Personalized tag recommendation based on generalized rules," ACM Transactions on Intelligent Systems and Technology, vol. 5, pp. 15-31, 2013.

[2]. Y. Toshihiko, Hu Jiani, S. Shumpei, "FolkPopularityRank: Tag Recommendation for Enhancing Social Popularity," Twenty-Sixth International Joint Conference on Artificial Intelligence, pp. 231-237, 2017.

[3]. R. Jaschke, L. Marinho, A. Hotho, "Tag Recommendations in Folksonomies," European Conference on Principles of Data Mining and Knowledge Discovery, pp. 506-514, 2006.

[4]. T. H. Hanh, W. Martin, "Personalized Tag Recommendation for Images Using Deep Transfer Learning," University of Hildesheim, Universit atsplatz, Hildesheim, Germany, 2014

[5]. R. Agarwal, T. Imielinski, "Mining association rules between sets of items in large databases," ACM SIGMOD international, conference on Management of data, pp. 207-216 , 1993.

[6]. S. Bao, G. Xue, "Optimizing web search using social annotations," 16th International Conference on World Wide Web (WWW07), New York, pp. 501-510, 2007.

[7]. E. Barali, L. Cagliero, T. Cerquitell, "Support driven opportunistic aggregation for generalized item set extraction," IEEE Conference of Intelligent Systems IEEE, Los Alamitos, pp. 102-107, 2010.

[8]. S. Brin and L. Page, "The anatomy of a large-scale hyper textual web search engine," 7th International Conference on World Wide Web (WWW98), pp. 107-117, 1998.

[9]. P. A. Chitra, S. Costache, W. Nejdl, "P-tag: Large scale automatic generation of personalized annotation tags for the web," 16th International Conference on World Wide Web (WWW07), New York, pp. 845-854, 2007.

[10]. R. Datta, J. Wang, "Toward bridging the annotation-retrieval gap in image search," IEEE Multimedia vol. 14, pp. 24-35, 2007.

[11]. P. A. Dmitriev, N. Eiron, M. Fontoura and E. Shekita, "Using annotations in enterprise search," 15th International Conference on World Wide Web (WWW06), pp. 811-817, 2006.

[12]. A. Wesley, N. Garg and I. Weber," Personalized, interactive tag recommendation for Flickr," ACM Conference on Recommender Systems, pp. 67-74, 2006.

[13]. C. Chen and C. Wang, "Personalized tag recommendation using graph based ranking on multi-type interrelated objects," 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval , New York, pp. 540-547, 2009.