

International Journal of Innovative Research in Science, Engineering and Technology*An ISO 3297: 2007 Certified Organization**Volume 5, Special Issue 14, December 2016***3rd National Conference on Recent Trends in Computer Science and Engineering and Sustainability in Civil Engineering (TECHSYNOD'16)****19th, 20th and 21st December 2016****Organized by****Department of CSE, MCA &CE, Mohandas College of Engineering and Technology, Thiruvananthapuram, Kerala, India**

Survey on Connection Discovery Using Non-user Generated Tagging on User Shared Images

Chippy Jacob^{#1}, Rakhi Ramachandran Nair^{#2}, Tessy Vincent^{#3}, Rekha V R^{*4}M.Tech Student, Dept. of Computer Science, College of Engineering, Kidangoor, Kerala, India^{#1#2#3}HOD & Assistant Professor, Dept. of Information Technology, College of Engineering, Kidangoor, Kerala, India^{*4}

ABSTRACT: The interests of users are always useful for customized or personalized content recommendations on friendships, events, news, products and videos from the social big data. However, those interests may not be specified by the user make the recommendations challenging. Social graphs, representing the online friendships among users, are one of the most fundamental types of data for numerous online networking applications, such as recommendation, virality prediction and marketing. As some users also hide or restrain the information of their connections from the public in social media platforms due to security concerns. Another possible solution to analyze the user's interests from the shared content, especially images with manually annotated tags, but it is unreliable. The user-shared images are proved to be an easier and effective alternative to discover user connections. It is observed that the shared images from users with a follower/followee relationship show relatively higher similarities. This survey focuses the most important techniques for connection discovery among users without the usage of social graphs and manually annotated tags.

KEYWORDS: Big data; connection discovery; recommendation; social network analysis; user-shared images.

I. INTRODUCTION

Social media is gaining a great amount of popularity in recent years, with millions of users visiting sites like Facebook, Google Plus for social networking; Twitter for micro-blogging; Drupal for blogging; Instagram and YouTube for photo and video sharing, respectively. These social networking sites depend primarily on their users to create and contribute content, to comment on others' content with tags, ratings, and comments, to form online connections and to join online communities and so on.

As the web is over flooded with data, the need of providing users with personalized recommendations is much essential. The objective of a personalized recommender system is to adapt the content based on characteristics of the individual users. Social media and personalized recommender systems can mutually benefit from one another. On the one hand, social media introduces new types of data and such as tags, ratings, comments, and explicit people relationships, which can be utilized to enhance recommendations. On the other hand, recommender systems can play a vital role in the success of social media applications and the social web as a whole for expanding their relations and connections. More recently, some of the leading social media sites have also added personalized recommendation features: video-sharing site YouTube has launched a personalized homepage that includes recommendations based on past views and favorites. This feature is reported to have led to an increase in the popularity of such online social media sites.

Recommending friends or related people, groups or pages on social networking sites is worth studying because it is not quite the same as conventional recommendations of books, movies, restaurants, etc. due to the social implications of "friending".

International Journal of Innovative Research in Science, Engineering and Technology*An ISO 3297: 2007 Certified Organization**Volume 5, Special Issue 14, December 2016***3rd National Conference on Recent Trends in Computer Science and Engineering and Sustainability in Civil Engineering (TECHSYNOD'16)****19th, 20th and 21st December 2016****Organized by****Department of CSE, MCA &CE, Mohandas College of Engineering and Technology, Thiruvananthapuram, Kerala, India****II. LITERATURE SURVEY****A. RECOMMENDING PEOPLE ON SOCIAL NETWORKING SITES USING FOUR DIFFERENT ALGORITHMS**

This paper [1] proposed people recommendation systems designed to help users find known, offline contacts and discover new friends on social networking sites.

1) Content Matching

Content matching algorithm is based on the idea that “if two users both post content or image on similar topics, they might be interested in getting to know each other”. In other words, the algorithm helps to find users associated with similar content on social networking sites. It creates a bag-of-words representation of each user, using textual content profile entries and status messages of users, as well as the title, description, tags, and any textual content associated with their photos and shared lists, job title and the city of the user’s work location. After doing Porter-Stemming, the remaining word stems associated with a user u are used to create a word vector V_u . The similarity of two users a and b is then measured by the cosine similarity of their word vectors V_a and V_b . Users similar to the recipient user u are recommended in decreasing order of similarity. The limitation of content matching algorithm recommends mostly the unknown people with no common friends.

2) Content-plus-Link (*CplusL*)

The content-plus-link algorithm improves the content matching algorithm by combining it with social link information derived from social graphs. This algorithm computes similarity in the same way as the content matching algorithm. However, instead of recommending users with top similarity scores, we boost the similarity of a candidate user c and u by 50% if a valid social link from u to c exists.

3) Friend-of-Friend Approach (*FoF*)

This algorithm utilizes just social network information of friending based on the idea that “if many of my friends consider Bob a friend, perhaps Bob could be my friend as well”. This intuition is the primary algorithmic foundation of the “People You May Know” feature on Facebook, which is one of the few known people recommenders deployed on a social networking site. For a user u being the receiver of the recommendation, its recommendation candidate set $RC(u)$ is created first. For each candidate c of $RC(u)$, its mutual friends set $MF(u,c)$ is calculated. The score of each candidate c for recipient u is defined as the size of $MF(u,c)$ and the high score indicates their relationship or connection. The demerit of this algorithm is that it requires existing friends because it cannot generate recommendations for people with no or a limited number of friends.

4) SONAR

SONAR system helps to aggregate relationship information from the different other data sources like publication database, patent database, social networking sites, blogging system, people tagging system and project wiki. A relationship is discovered if within that data source two people have somehow interacted with each other, such as co-authoring a paper or leaving comments on each others’ blog. Given a user u , SONAR returns a list of users related to u and their aggregated relationship score from different systems with u .

This method is a good approach which recommends people who are mostly related or known.

B. CONTENT-BOOSTED COLLABORATIVE FILTERING (CBCF) ALGORITHM FOR PERSONALIZED RECOMMENDATION

There are two prevalent approaches for building recommender systems: content-based (CB) and collaborative filtering (CF). The CB approach is based on recommending items that are similar to those in which the user has shown interest in the past. The CF approach recommends items to the user based on other individuals who are found to have similar

International Journal of Innovative Research in Science, Engineering and Technology*An ISO 3297: 2007 Certified Organization**Volume 5, Special Issue 14, December 2016***3rd National Conference on Recent Trends in Computer Science and Engineering and Sustainability in Civil Engineering (TECHSYNOD'16)****19th, 20th and 21st December 2016****Organized by****Department of CSE, MCA &CE, Mohandas College of Engineering and Technology, Thiruvananthapuram, Kerala, India**

preferences or tastes. Traditionally, both CB and CF systems have been based on explicit input from the user, usually provided by rating or commenting a set of items. To avoid this extra burden on the user, implicit interest indicators, such as purchase history, views, clicks, or queries, has recently become more popular in recommender systems.

This paper [2] presents a hybrid recommender approach - Content-Boosted Collaborative Filtering (CBCF), which uses both CB and CF methods. An item cannot be recommended unless a user has rated it before. This problem is called First-rater Problem and it affects the new items badly. This problem can be avoided by combining a CB predictor to enhance existing user data, and then provides personalized suggestions through CF. This method combines CB and CF techniques.

C. PERSONALIZED RECOMMENDATION OF SOCIAL SOFTWARE ITEMS BASED ON SOCIAL RELATIONS

This paper [3] explores the effects of familiarity and similarity between the user and the people who have rated the items. Similarity is defined in two forms – profile similarity (age, hobbies, film genre preferences, etc.) and rating overlap. Familiarity is simulated through repeated exposure to profiles and thus does not reflect real acquaintance.

Personalized recommendations within an enterprise social software application suite which consist of different social software applications including social bookmarking, file sharing, blogging, communities, and wikis.

3) Social Network Aggregation

SONAR, a system for collecting and aggregating social network information across different sources within the application suite, is used to generate the users' familiarity and similarity networks. Familiarity relationships include all direct people-people relations, as well as two types of indirect relations: co-authorship of a paper and persons having the same manager. Similarity relationships include indirect relations only, such as co-usage of the same tag, co-tagging of the same item, co-commenting on the same blog entry, or co-membership in the same community. The familiarity relationships are more effective in obtaining interesting recommended items, yet similarity relationships are also productive and may diversify the recommended items.

4) Item Recommendation Algorithm

The recommendation score of item i to user u is determined by considering the relevance of the item i , SONAR relationship score between users u and v based on the familiarity and similarity networks, weightage score for the user-item relationship between user v and item i . So the recommendation score of an item i reflects its likelihood to be recommended to the user. This score may increase due to the following factors: more people within the user's network are related to the item, stronger relationships of these people to the user, stronger relationships of these people to the item, and freshness of the item.

D. RECOMMENDATION BASED ON PEOPLE AND TAGS

Tags applied on the user by other people are found to be highly effective in representing that user's topics of interest. This paper [4] uses the combination of related people and related tags to recommend social media items.

Social networks and discovery in the enterprise (SaND) is an aggregation system like SONAR that models relationships among people, items, and tags, through data collected across different applications. The aggregation system aggregates any kind of relationships between its three core entities—people, items, and tags. A user is directly related to: (1) another person: as a friend, as a tagger of or tagged by that person; (2) an item (e.g., a shared file or a community): as an author, a commenter, a tagger, or a member; or (3) a tag: used by the user or applied on the user by others.

The user profile, $P(u)$, is given as an input to the recommender engine and it is used to personalize the recommended items for u . It consists of 30 related people, $N(u)$, and 30 related tags, $T(u)$, retrieved through the aggregation system.

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization

Volume 5, Special Issue 14, December 2016

3rd National Conference on Recent Trends in Computer Science and Engineering and Sustainability in Civil Engineering (TECHSYNOD'16)19th, 20th and 21st December 2016**Organized by****Department of CSE, MCA &CE, Mohandas College of Engineering and Technology, Thiruvananthapuram, Kerala, India**

The set of people related to the user is extracted by considering both direct and indirect people-people relations. The user tag relations such as - used tags (tags the user has used), incoming tags (tags applied on the user by others), and indirect tags (tags applied on items related to the user) - are used to extract the user's related tags.

5) Recommendation Algorithm

Given the user profile, $P(u) = (N(u), T(u))$, we suggest items to the user u that are related to people and/or tags in his profile. The recommendation score of item i for user u is determined by using a recommendation algorithm. It uses relationship strengths of user u to user v and tag t i.e $w(u,v)$ and $w(u,t)$ obtained from the user profile and the relationship strengths between user v and tag t respectively to item i i.e $w(v,i)$ and $w(t,i)$ which are as determined by SaND based on the direct relations graph.

This method is recommending social media items based on both related people and related tags. A combination of directly used tags and tags applied by others is most effective in representing the user's topics of interest. A recommender based on this tag profile yields items that are significantly more interesting to the user. A hybrid people-tag-based recommender has other advantages, such as low proportion of expected items, high diversity of item types, richer explanations, and the simple fact that for some users, recommendations based on people work better, while for others, recommendations based on tags are more effective.

E. BAG-OF-FEATURES TAGGING APPROACH USING USER'S SHARED IMAGES

This paper [5] proposes a novel methodology utilizing Bag-of-Features (BoF) based tagging that improves recommendations through user's interests obtained from the images they have uploaded or shared. Sometimes user's interests may not be specified in their profile, so it can be analyzed from the images that are shared by them in the social networking sites. Tag is a text to annotate a social content or image and it can describe what the image is about. As the users with similar interests are more likely to upload images with similar tags, it is assumed that connections can be discovered by find out the persons who are sharing the images with the same tag. However, in most of the social networks, the tags are not always reliable or not suitable for analysis or sometimes even available.

BoF is an image based approach that detects low-level features, and encodes an image into a feature vector. Images are grouped based on the similarity of their features vectors and hence the similarity of two users can be calculated.

The different stages of the BoF tagging are:

6) Feature Extraction

Feature extraction is a procedure to obtain the local features of an image. These local features can be identified by Harris Affine detector, or Maximally Stable Extremal Regions detector. They are relatively consistent with viewing angles and lighting conditions and are represented in a way that is independent of the size and orientation, such as scale invariant feature transform (SIFT).

7) Codebook Generation

It is a process to acquire the visual words that can represent the features obtained in the feature extraction. It is a clustering process, which is a k-mean clustering that groups similar features. The mean vectors of each group can be defined as the visual word, which can be utilized to represent the features in these images.

8) Feature Coding and Pooling

Feature coding is to encode features with the visual words. Each feature in every image is represented by a visual word in feature coding. The images are then represented by a feature vector in the feature pooling.

International Journal of Innovative Research in Science, Engineering and Technology*An ISO 3297: 2007 Certified Organization**Volume 5, Special Issue 14, December 2016***3rd National Conference on Recent Trends in Computer Science and Engineering and Sustainability in Civil Engineering (TECHSYNOD'16)****19th, 20th and 21st December 2016****Organized by****Department of CSE, MCA &CE, Mohandas College of Engineering and Technology, Thiruvananthapuram, Kerala, India****9) Clustering and Tagging**

Group images with similar feature vectors and clusters are formed. After obtaining the cluster, the images in any cluster are assigned with the same auxiliary tag to reflect that they are visually similar and belong to the same group.

10) Profile Learning

It is not required to input the user profile by the user as it can be produced from the images uploaded by the user. The histogram of the BoFT labels of their shared images is used as the user profile in this approach.

11) User Profile and Recommendation

After getting the user profile as a histogram of tags, recommendations can be found based on the tie, the strength of the relationship between two people. An item favored by a user may also be liked by friends of the user with strong ties. Content recommendation is then possible by using similarity cosine of set of tags in the images of two users as people with similar interests have a higher similarity cosine value.

As user friendship/interests are not always available, this paper proposes an unsupervised approach to classify image according to their visual features. Those visual features represent the user's interest and hence can make a recommendation.

F. AUTOMATIC TAGGING OF IMAGE

Tag based Image Retrieval and Connection Discovery has become popular and practical in extensive applications. However, manually tagging images is extremely expensive and time-consuming. So a lot of high-quality image tagging algorithms are proposed to automatically tag the images.

1) By merging multiple features

This method [6] suggested a new approach to fusing semantic, visual features and their correlation simultaneously for image tagging. The whole image tagging process mainly contains two stages which are training and tagging. In the training phase, a Feature Correlation Graph (FCG) is computed to capture the correlations between different features in an integrated manner, which represent features as nodes and their correlations as edges. Every tag or query image can be represented by a local FCG. The FCG of a query image comprises of its visual features and they are connected if they have a correlation by using an intra-type correlation method. The FCG of a tag comprises of all the tags named semantic features which are correlated with it and the visual features which have a co-relation with those tags obtained by using an inter-type correlation method. In the tagging stage, a probabilistic model, i.e. Markov Random Field is used to evaluate the relevance of assigning a candidate tag t_1 to image I_q by calculating the joint distribution probability of such a graph. A large value of joint distribution probability implies high relevance of the tag to the query image. Finally, all the tags are ranked by their relevance value and the top-ranked tags are returned as the result tags.

2) Tag Recommendation System based on Collective Knowledge

Given a photo with user-defined tags, an ordered list of m candidate tags is derived for each of the user-defined tags, based on tag co-occurrence. The lists of candidate tags are then used as input for tag aggregation and ranking, which ultimately produces the ranked list of n recommended tags. This method [7] consists of two stages.

i) Tag co-occurrence - It is the key to the tag recommendation approach. The user-generated tags provide the collective knowledge base upon which this method works on. In this step, the co-occurrence coefficients between two tags are calculated. The raw co-occurrence coefficients between two tags is just the number of photos which share these two tags. As this measure is not very meaningful, make it normalized by using two different normalization methods: Symmetric measures and Asymmetric measures. It is more likely that asymmetric tag co-occurrence will provide a

International Journal of Innovative Research in Science, Engineering and Technology*An ISO 3297: 2007 Certified Organization**Volume 5, Special Issue 14, December 2016***3rd National Conference on Recent Trends in Computer Science and Engineering and Sustainability in Civil Engineering (TECHSYNOD'16)****19th, 20th and 21st December 2016****Organized by****Department of CSE, MCA &CE, Mohandas College of Engineering and Technology, Thiruvananthapuram, Kerala, India**

more suitable diversity of candidate tags than its symmetric measures. Finally, this method creates a candidate set of tags for each user tag.

ii) Tag Aggregation and Ranking - A recommended set of tags are to be find out from the large candidate set of tags. For this two aggregation strategies are used: Voting and Summing. Voting does not take the co-occurrence values of the candidate tags into account, while the summing strategy uses the co-occurrence values to produce the final ranking. Promotion function is derived using stability promotion function, descriptiveness- promotion function, rank promotion function. This function is applied to a tag pair of user-tag and candidate tag. Then applying the promotion function in combination with either the voting or summing aggregation function, the score function of a candidate tag is updated. It is found that voting method with promotion outperforms the summing method with promotion.

Even though this technique can be used to improve the user-generated tags, some uploaded images may not be tagged or tagging with incorrect tags.

3) Image Tag Refinement via Ternary Semantic Analysis

This method [8] can reduce the effect of noise due to the missing or incorrect tag of images in the automated tagging system . Three types of interrelated entities are involved, i.e., *image*, *tag*, and *user* in the tagging system. The incorporation of *user* information contributes to a better understanding and description of the tagging data. It can reduce the semantic gap issue which is prevalent in such systems. A ranking optimization scheme is proposed, which intuitively considers the user tagging behaviors. In addition to the ternary interrelations, we also collect multiple intra-relations among users, images, and tags. These intra-relations constitute the user, image, tag affinity graphs respectively A tag affinity graph based on tag context and semantic relevance is build and tags with the k-highest affinity values are considered semantic-relevant or context-relevant. This method succeeds to mine the semantic correlations among tags and images from the observed tagging data.

G. CONNECTION DISCOVERY USING BIG DATA OF USER-SHARED IMAGES IN SOCIAL MEDIA

This paper [9] is the modification of the methods used in Bag-of-Features Tagging Approach in [5]. It is observed that the shared images from users with a follower/followee relationship show relatively higher similarities. Moreover, [9] discovered that distribution of the number of shared images and follower/followee relationships on both social networks follows the power law distribution. So this method can discover user or relationships using users' shared images directly without the usage of any user annotated tags. Through discovering user connections from their shared images, the connections can be used to identify user gender or recommend followers/followee to users. A multimedia big data system that utilizes this observed phenomenon is proposed as an alternative to user-generated tags and social graphs for follower/followee recommendation and gender identification.

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization

Volume 5, Special Issue 14, December 2016

3rd National Conference on Recent Trends in Computer Science and Engineering and Sustainability in Civil Engineering (TECHSYNOD'16)

19th, 20th and 21st December 2016

Organized by

Department of CSE, MCA &CE, Mohandas College of Engineering and Technology, Thiruvananthapuram, Kerala, India

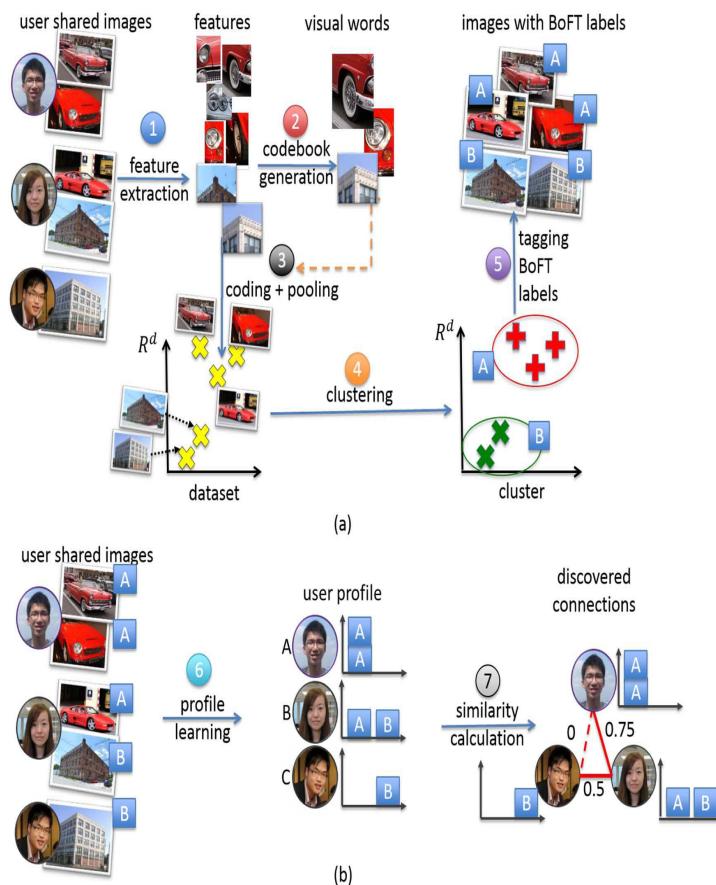


Figure 1 BoFT: (a) annotation with BoFT labels, and (b) user similarity calculation based on BoFT labels.

The follower/followee relationship between users can be possibly detected from the higher similarity of visual features in their shared images. Bag-of-Features Tagging (BoFT) is an image-based approach that detects low-level features and encodes an image into a feature vector. Images are analyzed using BoFT, which annotates each image with a BoFT label. In order to handle the big data, the feature vectors are first split into multiple blocks in the Hadoop Distributed File System (HDFS) and distributed to virtual machines (VMs) for the k-means clustering process. Each VM is in charge of computing the distribution of different labels for several users, and the BoFT similarity is also calculated in a distributed way.

A user profile, which reflects the content of a user's shared images, is the key in connection discovery. A user is represented by his/her user profile and it contains the distribution of the BoFT labels that the user's image has obtained through BoFT labeling. After getting the user profile of each user, BoFT similarity between a pair of users is identified which will be higher if they share highly similar images. So a higher BoFT similarity between a user pair (i,j) implies a higher probability that the two users i,j are a related pair. The list of recommended users are those with higher BoFT similarity to a user, and a recommendation system is built accordingly. If the BoFT similarities of all users with a particular user are given, then a list of recommended users those with the higher BoFT similarity can be obtained. Those user list should be the most likely related pairs with user.

H. NON-USER GENERATED ANNOTATION ON USER SHARED IMAGES FOR CONNECTION DISCOVERY

This paper [10] suggests the idea that Non-user generated labels annotated on shared images can be used for connection discovery along with different color-based and feature-based methods.

International Journal of Innovative Research in Science, Engineering and Technology*An ISO 3297: 2007 Certified Organization**Volume 5, Special Issue 14, December 2016***3rd National Conference on Recent Trends in Computer Science and Engineering and Sustainability in Civil Engineering (TECHSYNOD'16)****19th, 20th and 21st December 2016****Organized by****Department of CSE, MCA &CE, Mohandas College of Engineering and Technology, Thiruvananthapuram, Kerala, India**

The connection discovery using SIFT descriptor is already done in [9]. It is found that GIST descriptor has shown good results for scene categorization and image search. The idea is to build a holistic and low dimensional representation of a scene, which does not require explicit segmentation of image regions and objects. The method filters each pixel with a sequence of (Gabor) filters, producing feature maps for the image. Each feature map is divided into blocks and the GIST descriptor is obtained by concatenating the averaged value of each block in all feature maps. The GIST descriptor summarizes the gradient information (scales and orientations) for different parts of an image, which provides a rough description of the scene.

The images are encoded into vectors with different approaches for comparison. The first two approaches are color-based methods, Color Histogram Descriptors, where both the HSV color space and RGB color space are used. The HSV color space is uniformly discretized as 16x4x4 and the RGB color space is discretized as 16x16x16. The other two methods are feature-based, one based on SIFT and the other based on GIST. SIFT is based on BoF and in this method the SIFT descriptor is obtained for each image and the vocabulary is learned by clustering similar features. The last method is the GIST descriptor. It is configured with a common setting, namely the number of blocks is 4x4 and the number of orientations for all four scales is 8. After the vectors for all the images are obtained, k -means clustering is used to group images with respect to each method. Each cluster obtained is assigned with the same non-user generated label to reflect that the images in the cluster are visually similar and belong to the same group. The label distribution of each method is then calculated by counting the occurrences of each label.

GIST has a higher precision rate and recall rate than the other methods. In terms of precision rate, GIST is 20% better than SIFT. And both GIST and SIFT are much better than RGB and HSV, with a higher rate of more than 100% and 70% respectively.

III. DISCUSSIONS

According to the survey conducted in [1], it is found that users are interested to connect with unknown people, if they have any friends or interests in common. That's where the importance of content-plus link algorithm lies. SONAR recommends mostly known people and those people may have common friends or common interests. Relationship based algorithms (FoF and SONAR) outperform content similarity ones (Content and CplusL) in terms of user response. If relationship information is less available, then content similarity method can be used. The relationship-based algorithms are better at finding known contacts while content similarity algorithms were stronger at discovering new friends.

The CBCF algorithm introduced in [3] incorporates the advantages of both CBF and CF, while not inheriting the disadvantages of either. As it is a hybrid approach, improvement of the performance of either CB or CF certainly improves the performance of the whole system. A hybrid people-tag-based recommender system can recommend people or items that are significantly more interesting to the user and it can make use of both the merits and demerits of people and tag based approach. In [4], a combination of directly used tags and tags applied by others(incoming tags) is used as the user's profiles and it is found to be the most effective way to represent the user's topics of interest.

Through this survey we found that tags are very useful in finding the relationships between users. So a lot of algorithms were proposed for generating automatic tags and to refine such tags. As user-generated tags are elaborated to generate recommended tags in [7], it is effective in generating tags automatically. The performance of tag generation using promotion function along with voting aggregation is more evenly distributed for any number of user-generated tags. It is also particularly good at recommending locations, artifacts, and objects. In [9], it is observed that related pairs of users have a higher BoFT similarity. We found that BoFT method has two problems. As SIFT-based BoF tagging method extracts the local features of images, same feature vector could be created even though those images are different. The value of k , which is the number of labels, used in k -means clustering algorithm has to be predefined and the result of clustering the images depends on its value.

Feature-based methods are also proven to be 95% better than color-based methods, and 65% better than tag-based methods in terms of precision rate and recall rate. Although the performance of the color-based methods is worse than

International Journal of Innovative Research in Science, Engineering and Technology*An ISO 3297: 2007 Certified Organization**Volume 5, Special Issue 14, December 2016***3rd National Conference on Recent Trends in Computer Science and Engineering and Sustainability in Civil Engineering (TECHSYNOD'16)****19th, 20th and 21st December 2016****Organized by****Department of CSE, MCA &CE, Mohandas College of Engineering and Technology, Thiruvananthapuram, Kerala, India**

UserT, nonuser generated labels from the color-based methods are more accessible because they do not have any language barriers and are easy to compute. But the problem is that color-based methods do not include any spatial information so different objects with the same color could have the same histogram, such as a blue car and the ocean. In contrast, feature-based methods can capture spatial information. In [10], GIST gives a rough glimpse of the whole image by capturing the holistic information, while SIFT-based BoF focuses on local features. Therefore the GIST descriptor should be more suitable for scene categorization, especially in image-based social networks, such as Flickr. In other social networks where users share images of their daily lives, such as WeChat, SIFT could be more suitable to capture the nature in their shared images.

IV. CONCLUSION

From the overview of the “connection discovery among users on social media sites or applications” presented in this survey, it is clear that a lot of researches have been done in this area. In this paper, various techniques for connection discovery and personalized recommendations on friendships or media contents are discussed. The automation of tags or non-user generated tags is much effective for finding the connections or similarity of their interests among the users on social networking sites. A SIFT based Bag-of-Feature Tagging approach (BoFT) makes use of computer vision techniques in object recognition tasks and is proven to be better than approaches using user annotated tags. But SIFT might not be accurate in some cases as it creates feature vector by concentrating on the local features. So it is recommended to use other feature extraction techniques such as GIFT, instead of using SIFT. The researches for finding the correct value of k in k-means clustering in advance is still going on.

REFERENCES

- [1] J. Chen, W. Geyer, C. Dugan, M. Muller, and I. Guy, “Make new friends, but keep the old: Recommending people on social networking sites,” in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2009, pp. 201–210.
- [2] Melville, P., Mooney, R. J., & Nagarajan, R. 2002. Content boosted collaborative filtering for improved recommendations. *Proc. AAAI '02*, 187-192.
- [3] I. Guy, N. Zwerdling, I. Ronen, D. Carmel, and E. Uziel, “Social media recommendation based on people and tags,” in *Proc. 33rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2010, pp. 194–201.
- [4] I. Guy *et al.*, “Personalized recommendation of social software items based on social relations,” in *Proc. 3rd ACM Conf. Recommender Syst.*, 2009, pp. 53–60.
- [5] M. Cheung and J. She, “Bag-of-features tagging approach for a better recommendation with social big data,” in *Proc. 5th Int. Conf. Adv. Inf. Mining Manag.*, 2014, pp. 83–88.
- [6] X. Zhang, Z. Li, and W. Chao, “Tagging image by merging multiple features in a integrated manner,” *J. Intell. Inf. Syst.*, vol. 39, pp. 87–107, 2012.
- [7] B. Sigurbjrnsson and R. Van Zwol, “Flickr tag recommendation based on collective knowledge,” in *Proc. the 17th Int. Conf. World WideWeb*, 2008, pp. 327–336.
- [8] J. Sang, C. Xu, and J. Liu, “User-aware image tag refinement via ternary semantic analysis,” *IEEE Trans. Multimedia*, vol. 14, no. 3, pp. 883–895, Jun. 2012.
- [9] M. Cheung, J. She., and Z. Jie., ”Connection discovery using Big Data of user-shared images in social media,” in *IEEE Transactions on Multimedia*, vol.17, no.9, pp.1417-1428, Sept. 2015
- [10] Ming Cheung, James She, Xiaopeng Li., “ Non-user Generated Annotation on User Shared Images for Connection Discovery,” in *IEEE International Conference on Data Science and Data Intensive System*, , December 2015, pp. 1486-1491