International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Exploiting Recommendation on Social Media Networks

Swati A. Adhav [1], Sheetal A. Takale [2]

[1] ME-II Computer Engineering, Savitribai Phule Pune University, Vidya Pratishthan's College of Engineering, Baramati,

[2] Professor, Department of Information Technology, Savitribai Phule Pune University, Vidya Pratishthan's College of Engineering, Baramati,

Abstract: ***Social media where the user interacts with each other to form social networks is becoming very popular these days. In social media networks, users are influenced by others for various reasons. For example, the colleagues have a strong impact on one's work, while the friends have a strong impact on one's daily life. Social influence mining in social networks is important in real world applications such as photo recommendation. Since the Social network is multi-modal and heterogeneous, it is insignificant to effectively exploit the social media information to learn the topic distribution for users and images accurately. To tackle this Problem, Exploiting Recommendation on Social Media Networks is used to find topical influential for users and images with the help of Hypergraph Learning approach. The Hypergraph learning method is used to model user, images and social link relationship. It combines the content of images and social links to determine the topic-specific influence for users and images in the social media network. Exploiting Recommendation on Social Media Networks consists of three learning stages Hypergraph Construction, topic distribution learning, and topic sensitive influence ranking.***

Keywords: Hypergraph Construction, topic distribution learning, topic sensitive influence ranking.

1. Introduction

Social media has reform the way people share and access information. In communities there are many social sharing websites, which allows user to share photos, web links, songs, pictures etc. with social media networks, users necessarily interact with each other in communities. The social communications include two-way links like connect in LinkedIn, add friend in Facebook, and one-way links like contact in Flickr network. For example, the colleagues on LinkedIn will largely impact one's choice in work, while the friends from Facebook have strong influence on one's preferences in daily life. With the increase of social sharing networks such as Facebook, and Flickr, Social Influence becomes more important. In large social media networks, users are influenced by others for various reasons. Social influence [2] is a most important topic in Social Networks and looks at how individual thoughts and feelings are influenced by social groups. Social influence mining in social media networks is important in real-world applications such as photo recommendation. Fig. 1 shows the contact network of user Sudhir in Flickr which includes three influencers: Sunil, Geeta and Ram. On the right there are expertise of each influencer regarding topics on fashion, travel and technology. Assuming Sudhir is searching photos of Goa for his trip, obviously Sunils preference will influence him most. when Sudhir searches photos of DG fashion show, the advice from Geeta will be most appreciated. This demonstrates that some influencers in certain topics might be more trustworthy than others and the influence value is topic-sensitive.

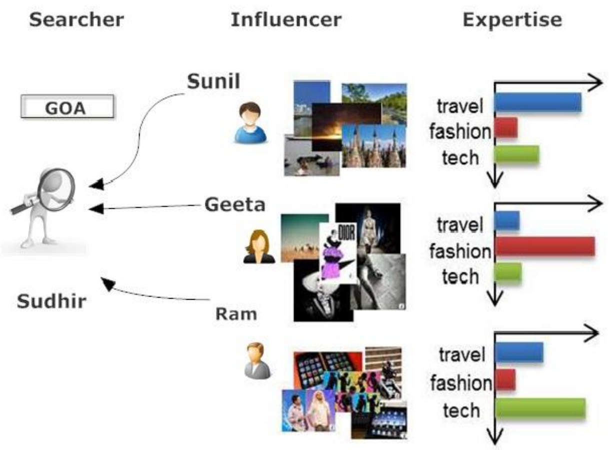


Figure 1: Example of Topic Sensitive Influencer

Social network is multi-modal and heterogeneous, it is insignificant to effectively exploit the social media information to learn the topic distribution for users and images accurately. Real social media networks like Facebook, Flickr are getting bigger with thousands or millions of user's, to efficiently find out the topic based influential strength for each social network is really a challenging problem. To tackle this Problem, Exploiting Recommendation on Social Media Networks is used to find topical influential for users and images with the help of Hypergraph Learning approach. In this paper, we investigate the problem of topic sensitive influencer mining in social media networks.

2. Literature Survey

A. Hypergraph Learning

Zhou.et.al. [6] proposed hypergraph is nothing but a generalization of the simple graph in which the edges, called hyperedges. Hypergraph can be used to model both various types of entities and complex relations, which is extremely appropriate in social media modeling. Hypergraph has been used in many information retrieval and data mining tasks, such as image retrieval and object recognition. Zhou et al used a general hypergraph framework and apply it to clustering, classification and embedding tasks. Liu et al [5] intended a transductive learning framework for image retrieval. In this method, each image is represented as a vertex in the constructed hypergraph, and the visual clustering results are used to construct the hyperedge. Liu et.al [8] used a transductive learning framework for image retrieval. It constructs a hypergraph by generating a hyperedge from each sample and its neighbors, and hypergraph-based ranking is then performed. Based on the visual similarity matrix computed from various feature descriptors, it take each image as a centroid vertex and create a hyperedge by a centroid and its k-nearest neighbors. To more exploit the correlation information among images, it propose a probabilistic hypergraph, which assigns each vertex to a hyperedge in a probabilistic way. This work demonstrates the effectiveness of hypergraph model in capturing higher-order relationship. Yue et.al [4] intended an approach that simultaneously utilizes both visual and textual information for social image search. In the proposed method, both visual content and tags are used to generate the hyperedges of a hypergraph, and a relevance learning method is performed on the hypergraph structure where a set of pseudo-relevant samples are used. Different from the conventional hypergraph learning algorithms, our method learns not only the relevance scores among images but also the weights of hyperedges. By using the learning of hyperedge weights, the effects of uninformative tags and visual words can be minimized.

B. Topic Distribution Learning

Topic models, such as PLSA [5], LapPLSI, LDA, Corr-LDA, have shown impressive success in many fields. Corr-LDA is used, to jointly model textual and visual features for latent topic representation. Latent Dirichlet Allocation (LDA) is the simplest and most popular statistical topic modeling (Blei et el. 2003). Many other models came up as an extension of LDA is the simplest topic model [8]. The intuition behind LDA is that documents exhibit multiple topics. In LDA [9], the topics are distributions over words and this discrete distribution generates observations (words in documents). LDA cannot model the correlations among topics. For example the topic “genetics” is more likely to be similar to “disease” than to “astronaut”. LDA fails to depict this correlation of topics. Correlated topic model (CTM) is introduced by Blei and Lafferty (2007) which is an extension of of LDA that can model the correlations among topics. Hoffman et.Al has proposed PLSA is a statistical model for analyzing two modes and finding the co-occurrence of data. PLSA is based on a generative probabilistic model. We can apply the Probabilistic Latent Semantic Analysis (PLSA) to model the generation of each image and its word co-occurrences. Probabilistic Latent Semantic Analysis is a novel statistical technique for the analysis of two mode and co-occurrence information, which has applications in information retrieval, machine learning from text and natural language processing. It can be used to extract topics from a collection of documents. The starting point for Probabilistic Latent Semantic Analysis is a statistical model which has been called aspect model. This model is a latent variable model for co-occurrence data which associates an unobserved class variable.

3. Implementation Details

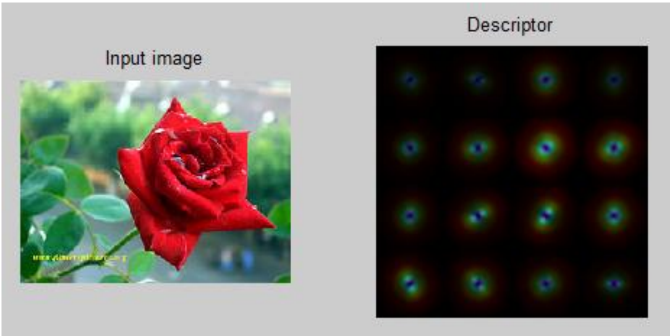
The proposed framework is organized into three stages Hypergraph Construction, Topic distribution learning, and Topic sensitive influence ranking respectively. The system architecture is defined in Fig.2.

A. Hypergraph Construction

For hypergraph construction, there are two types of vertices corresponding to the users and images, which constitute the vertex set denoted as V= {u, o}. The construction of the hyperedges is illustrated as follows:

1. Homogeneous hyperedges:

It is used to represent the visual-textual content relations among image vertices. There are two types of homogeneous hyperedges including visual content relation hyperedge and textual content relation hyperedge. In our experiment, from each image 512-dimensional feature vector is extracted as the content representation, including 512-dimensional GIST feature.



Feature extraction using GIST

The visual content similarity hyperedge’s weight is set based on the visual similarity matrix [8]. A is calculated according to

$$A_{ij} = \begin{cases} \exp\left(\dfrac{-\| x_i - x_j \|}{2\sigma^2}\right) \text{if } j \in N_t(i) \text{ or } i \in N_t(j) \\ 0 \text{ otherwise} \end{cases}$$

Where N_t denotes the index set for t nearest neighbors, x_i and x_j are feature vectors associated with images respectively, σ is a scaling parameter.

To construct $\in_{textual}$, we build a tag vocabulary. Each tag is used to build a hyperedge, i.e., the images containing the same tag are connected by a hyperedge. The homogeneous hyperedges are used to learn the topic distribution.

Visual Similarity Matrix is

0.0	0.52	0.51	0.55	0.56	0.58	0.45	0.48	0.53	0.55
0.52	0.0	0.52	0.76	0.65	0.66	0.6	0.51	0.74	0.78
0.51	0.52	0.0	0.83	0.51	0.76	0.75	0.5	0.46	0.55
0.55	0.76	0.83	0.0	0.64	0.75	0.65	0.56	0.56	0.54
0.56	0.65	0.51	0.64	0.0	0.8	0.76	0.66	0.72	0.68
0.58	0.66	0.76	0.75	0.8	0.0	0.66	0.58	0.51	0.5
0.45	0.6	0.75	0.65	0.76	0.66	0.0	0.8	0.71	0.62
0.48	0.51	0.5	0.56	0.66	0.58	0.8	0.0	0.61	0.65
0.53	0.74	0.46	0.56	0.72	0.51	0.71	0.61	0.0	0.6
0.55	0.78	0.55	0.54	0.68	0.5	0.62	0.65	0.6	0.0

Textual Similarity Matrix is:

1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 1 1 1 1 1 1 0 0 1 0 1 0 1 1 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 1 0 0 1 0 0 0
0 0 0 0 0 1 0 0 0 1 0 1 1 1 1 1 1 0 1 0 0 1 0 0 0 0 0 1 1 0 0 0 1 0 0 0 1 1 1 1 1 1 1 0
0 0 0 0 0 1 1 1 0 0 0 1 0 0 0 1 0 1 1 1 1 1 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 1 0 0 0 1 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 1 0 0 1 0 0 0 0 0 0 0 1 1 1 1 0 1 0 0 0 1 1 1 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 1 1 0 0 1 0 1 0 0 0 0 0
0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 1 1 1 0 1 0 0 0 0 0
0 0 0 0 0 1 0 0 0 0 0 0 1 0 1 1 0 0 1 0 0 0 0 0 0 0 0 1 1 0 0 0 1 0 0 0 1 1 1 1 0 0 1 0
0 0 0 0 0 0 0 0 0 1 0 1 1 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1

A hypergraph G = {V, E, W} consists of the vertex set V, the hyperedge set E, and the hyperedge weight vector w. each edge is assigned a weight $w(e_i)$ the hypergraph G can be denoted by incident matrix H

$$h(v,e)=\begin{cases}1 \text{ if } v\in e\\ 0 \text{ if } v\notin e\end{cases}$$

h (v,e) is

1 0 0 0 0 1 1 0 0 0
1 1 1 0 0 0 0 0 0 0
0 0 1 0 0 0 0 1 1 0
1 0 0 1 0 0 0 0 0 1
0 1 1 0 1 0 0 0 0 0
1 0 0 0 0 1 0 0 0 1
1 1 0 0 0 0 1 0 0 0
1 0 1 0 0 0 0 1 0 0
0 0 1 0 0 1 0 0 1 0
1 0 0 0 0 1 0 0 0 1
0 0 0 0 0 0 0 0 0 0
0 0 1 1 0 0 0 0 0 1
0 0 1 0 0 0 0 0 1 1
0 0 1 0 0 0 0 0 0 1
0 0 1 0 0 1 1 1 1 0
0 0 1 1 0 1 0 0 1 0
0 0 1 0 0 0 0 0 0 0
0 0 0 1 0 0 0 1 0 0
0 0 1 1 0 1 0 0 1 0
0 0 0 1 0 0 0 0 0 0

For a hyperedge e ∈ E, hyperedge degree can be estimated by:

$$d(e)=\sum_{v\in V} h(v,e)$$

Vertex degree of each vertex v ∈ V is:

$$d(v)=\sum_{e\in E} w(e)h(v,e)$$

B. Topic Distribution Learning

Input: Image-Term matrix. Where, M: No of images. $\{o_1, o_2, \ldots\ldots o_m\}$ W: No of Words in Tag Vocabulary. $\{w_1, w_2, \ldots\ldots w_m\}$

Output: the Topic Distribution θ_o, θ_u.

We can apply the Probabilistic Latent Semantic Analysis (PLSA) [5] to model the generation of each image and its word co-occurrences by the following steps:

1. Select an image o_i with probability $P(o_i)$,
2. Select a potential topic of interest z_k with probability $P(z_k \mid o_i)$,
3. Create a word w_j with probability $P(w_j \mid z_k)$.

Following the maximum likelihood principle, we can determine the model parameters

$$l=\sum_{i=1}^{N}\sum_{j=1}^{M} n(o_i, w_j)\log\sum_{k=1}^{K} P(w_j,z_k)P(z_k,o_i)$$

By exploring the homogeneous relations among images in the hypergraph and utilizing hypergraph regularizer, represented by

$$\mathcal{R}=\frac{1}{2}\sum_{k=1}^{K}\sum_{e\,\varepsilon\, e0}\sum_{o_i,o_j\,\varepsilon\, v_o}\frac{w(e)h(o_i,e)h(o_j,e)}{d(e)} * \left(\frac{P(z_k,o_i)}{\sqrt{d(o_i)}}-\frac{P(z_k,o_j)}{\sqrt{d(o_j)}}\right) \quad (1)$$

to maximize the regularized log-likelihood from equation 1as follows:

$L = l - \lambda\mathcal{R}$

$$=\sum_{i=1}^{N}\sum_{j=1}^{M} n(o_i, w_j)\log\sum_{k=1}^{K} P(w_j,z_k)P(z_k,o_i) - \lambda\mathcal{R}$$

$P(w_j \mid z_k)$ can be calculated by using M step re-estimation:

$$P(w_j \mid z_k)=\frac{\sum_{i=1}^{N} n(o_i,w_j)P(z_k \mid o_i,w_j)}{\sum_{j'=1}^{M}\sum_{i=1}^{N} n(o_i,w'_j)P(z_k \mid o_i,w'_j)} \quad (2)$$

$P(z_k \mid o_i)$ Calculated by using:

$$P(z_k \mid o_i)=\frac{\sum_{i=1}^{N} n(o_i,w_j)P(z_k \mid o_i,w_j)}{\sum_{j=1}^{M} n(o_i,w_j)} \quad (3)$$

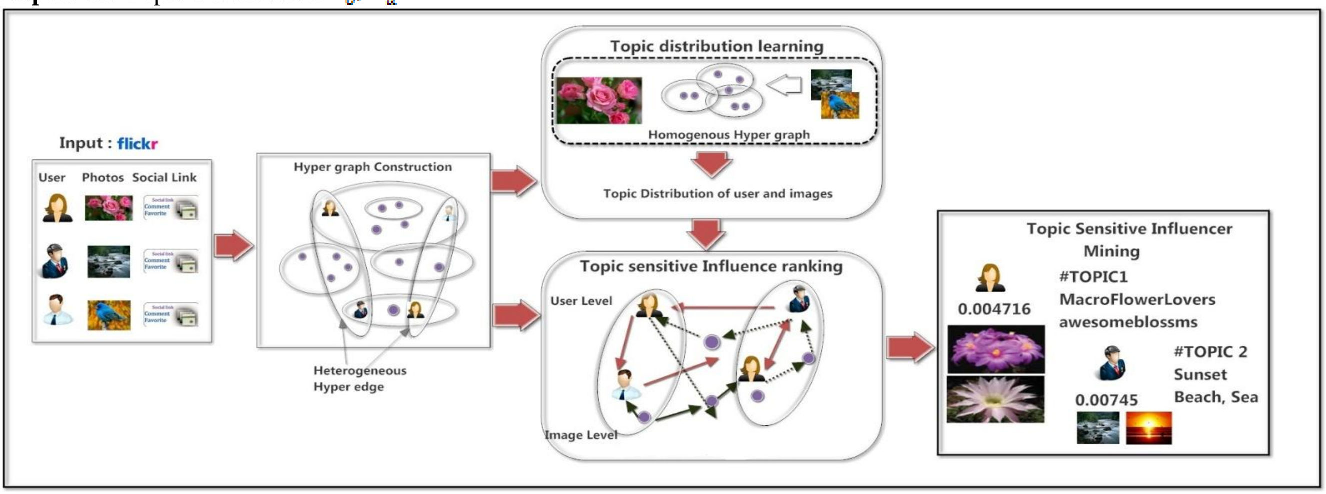


Figure 2: System Architecture

the topics extracted are as follows:

topic1	topic2	topic3
technology	cooolflowers	coolsunset
street	lotus	beautyofsunse
coolcity	awesome	beach
architecture	lovely	blueishsky
city	pink	sky
buildings	flowers	skysunset
beautifulenvironment	orangeflowers	sunriise
lightningbuildings	flower	naturalsunset
lightsss	gorgeousflowe	skyblue
cities	pinkflowers	beautifulsky
	rose	

C. Topic-sensitive Influence Ranking-

1. Learning Influence score for Users:

To calculate influence score of user we require user's contact network and their social link relationship like comments. First we calculate user's individual influence on each topic based on images that he/she has uploaded on particular topic and after that we calculate commenter influence based on which user has comment on which topic of image, after combining user's individual influence and commenter influence we get Final influence score of users.

2. Learning Influence score for Images:

To calculate influence score of images we require each user's uploaded images including tags. First we consider the tags of each image and calculating the probability of tags that comes under each topic.

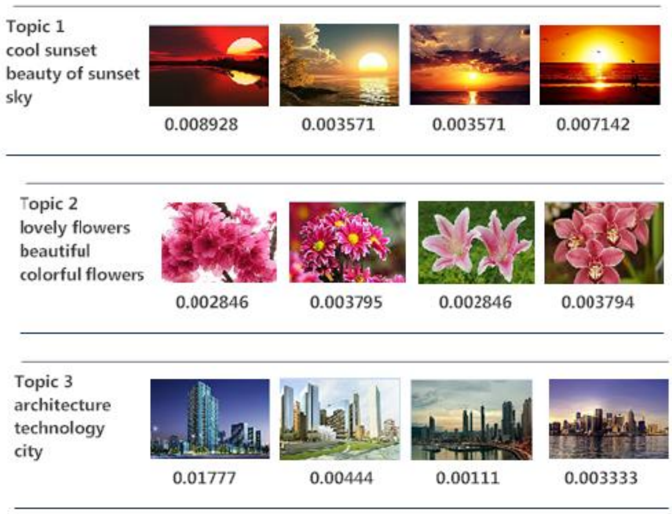
4. Results and Dataset

We collect data from Flickr API to evaluate our approach. We downloaded each user's contact information and each user's uploaded images. The social links including contacts, views and comments are recorded. In this dataset, each user has different metadata information including textual title and tags, comments. We also classified that each photo has at least one type of social links, containing comments.

Top 3 representative users and their photos on Topic1 from collected Flickr dataset, ranked by topic-specific social influence scores of users.

Top 3 representative users and their photos on Topic2 from collected Flickr dataset, ranked by topic-specific social influence scores of users.

Top 3 representative users and their photos on Topic3 from collected Flickr dataset, ranked by topic-specific social influence scores of users.



Representative photos on three topics on the collected Flickr dataset, ranked by topic-specific social influence scores of images.

5. Conclusion

Exploiting recommendation on social media network is proposed to deal with the problem of topic-sensitive influencer mining in social sharing websites. It aims to find the influential user and images in the social media networks. Social links between users and images indicate the social influence of users and images in the social media networks. We also demonstrate that this system can improve the performance significantly in the applications of photo recommendation.

6. Acknowledgment

I avail this opportunity to express my deep sense of gratitude and whole hearted thanks to my guide Prof. S. A. Takale madam for giving her valuable guidance, inspiration and encouragement to embark this paper. Without their Coordination, guidance and reviewing, this task could not be completed alone. Prof. S. A. Takale madam gave me all the freedom I needed for this project. I also take great pleasure in thanking Prof. S. S. Nandgaonkar for providing appropriate guidance. I would personally like to thank Prof. G. J. Chhajed, Head of Computer Dept. at Vidya Vidya Pratishthan's Collage of Engineering (VPCOE), and our Honble principal Dr. S. B. Deosarkar sir who creates a healthy environment for all of us to learn in best possible way.

References

[1] J. Sang, Q. Fang, "Topic-Sensitive Influencer Mining in Interest-Based Social Media Networks via Hypergraph Learning", *IEEE Trans. Multimedia*, vol. 16, no. 3, April. 2014.

[2] J.Tang, J. Sun, C. Wang, and Z. Yang, "Social influence analysis in large-scale networks," in *Proc. KDD*, 2009, pp. 807–816..

[3] J. Sang, C. Xu, "Right Buddy Makes the Difference: an Early Exploration of Social Relation Analysis in Multimedia Applications", Proc. *ACM Multimedia*, 2012.

[4] Y. Gao, M. Wang, "Visual-Textual Joint Relevance Learning for Tag-Based Social Image Search", in Proc. *ACM Multimedia*, 2011, pp. 15171520.

[5] T. Hofmann, "Probabilistic latent semantic indexing," in *Proc. SIGIR*, 1999, pp. 50–57.

[6] D. Zhou, J. Huang, and B. Schokopf, "Learning with hypergraphs: Clustering, classification, and embedding," in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 1–8.

[7] S. Agarwal, K. Branson, and S. Belongie, "Higher order learning with graphs," in *Proc. ICML*, 2006, pp. 17–24.

[8] Y. Huang, Q. Liu, S. Zhang, and D. N. Metaxas, "Image retrieval via probabilistic hypergraph ranking," in *Proc. CVPR*, 2010, pp. 3376–3383.

[9] David M.Blei, "Introduction to Probabilistic Topic Models"