ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

A Trust based Approach to reducing Noise and Spam wrt Tagging in online social networks (OSNs)

Prasada Rao Mandala. M.Tech, (Ph.D), ACCP
Associate Professor, Department of CSE
Avanthi's St.Theressa College of Engineering, Technology & Management, Vijayanagaram, AP.

D.R.S.Swetha
M Tech, Department of CSE
Avanthi's St.Theressa College of Engineering, Technology & Management, Vijayanagaram, AP.

Abstract: ***A social networking service is a platform to build social networks or social relations among people who share interests, activities, backgrounds or real-life connections. Labeling and tagging are carried out to perform functions such as aiding in classification, marking ownership, noting boundaries, and indicating online identity. They may take the form of words, images, or other identifying marks. Noisy and spam annotations often make it difficult to perform an efficient search. Users may make mistakes in tagging and irrelevant tags and content may be maliciously added for advertisement or self-promotion. This article examine recent advances in techniques for combating such noise and spam in social tagging. The trust relationship among users has a direct impact on the sharing and transmission mode of digital contents. To effectively assess direct or recommended trust between users, this paper proposed a multimedia social networks trust model based on small world theory. Online and Internet databases and early websites deployed them as a way for publishers to help users find content. In this paper we implement A Trust based Approach to reducing Noise and Spam wrt Tagging in online social networks (OSNs). The proposed system has been designed with an objective to minimize spam tagging and posting in social networking scenario with the adaptation of classification algorithms.***

Keywords: ***Tagging, OSNs, Trust Model, Spam, User Profile, Trust Modelling.***

Introduction: A social network service consists of a representation of each user (often a profile), his or her social links, and a variety of additional services. Social networks are web-based services that allow individuals to create a public profile, to create a list of users with whom to share connections, and view and cross the connections within the system. Most social network services are web-based and provide means for users to interact over the Internet, such as e-mail and instant messaging. Social network sites are varied and they incorporate new information and communication tools such as mobile connectivity, photo/video/sharing and blogging.

Tagging was popularized by websites associated with Web 2.0 and is an important feature of many Web 2.0 services. It is now also part of some desktop software. An analogous example of tags in the physical world is museum object tagging. In the organization of information and objects, the use of textual keywords as part of identification and classification long predates computers. However, computer based searching made the use of keywords a rapid way of exploring records.

In 2003, the social bookmarking website Delicious provided a way for its users to add "tags" to their bookmarks (as a way to help find them later); Delicious also provided browseable aggregated views of the bookmarks of all users featuring a particular tag.[1] Flickr allowed its users to add their own text tags to each of their pictures, constructing flexible and easy metadata that made the pictures highly searchable. The success of Flickr and the influence of Delicious popularized the concept, and other social software websites – such as YouTube, Technorati, and Last.fm – also implemented tagging. Other traditional

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

and web applications have incorporated the concept such as "Labels" in Gmail and the ability to add and edit tags in iTunes or Winamp.

Tagging has gained wide popularity due to the growth of social networking, photography sharing and bookmarking sites. These sites allow users to create and manage labels (or "tags") that categorize content using simple keywords. The use of keywords as part of an identification and classification system long predates computers. In the early days of the web keywords meta tags were used by web page designers to tell search engines what the web page was about. Today's tagging takes the meta keywords concept and re-uses it. The users add the tags. The tags are clearly visible, and are themselves links to other items that share that keyword tag.

Knowledge tags are an extension of keyword tags. They were first used by Jumper 2.0, an open source Web 2.0 software platform released by Jumper Networks on 29 September 2008. Jumper 2.0 was the first collaborative search engine platform to use a method of expanded tagging for knowledge capture.

Websites that include tags often display collections of tags as tag clouds. A user's tags are useful both to them and to the larger community of the website's users.

Tags may be a "bottom-up" type of classification, compared to hierarchies, which are "top-down". In a traditional hierarchical system (taxonomy), the designer sets out a limited number of terms to use for classification, and there is one correct way to classify each item. In a tagging system, there are an unlimited number of ways to classify an item, and there is no "wrong" choice. Instead of belonging to one category, an item may have several different tags. Some researchers and applications have experimented with combining structured hierarchy and "flat" tagging to aid in information retrieval.

In a typical tagging system, there is no explicit information about the meaning or semantics of each tag, and a user can apply new tags to an item as easily as applying older tags. Hierarchical classification systems can be slow to change, and are rooted in the culture and era that created them. The flexibility of tagging allows users to classify their collections of items in the ways that they find useful, but the personalized variety of terms can present challenges when searching and browsing.

Existing System

When information is exchanged on the Internet, malicious individuals are everywhere, trying to take advantage of the information exchange structure for their own benefit, while bothering and spamming others. Before social tagging became popular, spam content was observed in various domains: first in e-mail, and then in Web search networks have been also influenced by malicious peers, and thus various solutions based on trust and reputation have been proposed, which dealt with collecting information on peer behavior, scoring and ranking peers, and responding based on the scores . Today, even blogs are spammed. Ratings in online reputation systems, such as eBay, Amazon, and Epinions, are very similar to tagging systems and they may face the problem of unfair ratings by artificially inflating or deflating reputations. Several filtering techniques for excluding unfair ratings are proposed in the literature. Unfortunately, the countermeasures developed for e-mail and Web spam do not directly apply to social networks.

Proposed System

In a social tagging system, spam or noise can be injected at three different levels: spam content, spam tag-content association, and spammer. Trust modeling can be performed at each level separately or different levels can be considered jointly to produce trust models, for example, to assess a user's reliability, one can consider not only the user profile, but also the content that the user uploaded to a social system. In this article, we categorize trust modeling approaches into two classes according to the target of trust, i.e., user and content trust modeling. Table 1 summarizes

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

representative recent approaches for trust modeling in social tagging. Presented approaches are sorted based on their complexity from simple to advanced, separately for both content and user trust models.

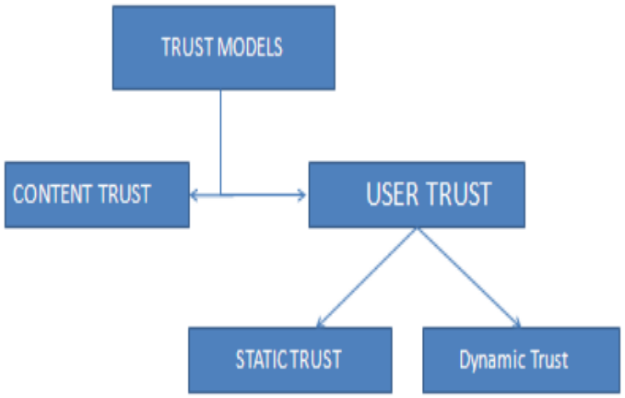
Module Description:

1. Content Trust Modeling
2. User Trust Modeling(Static)
3. User Trust Modeling(Dynamic)
4. Data Set

Modules Description

1. Content Trust Modeling

Content trust modeling is used to classify content (e.g., Web pages, images, and videos) as spam or legitimate. In this case, the target of trust is content (resource), and thus a trust score is given to each content based on its content and/or associated tags. Content trust models reduce the prominence of content likely to be spam, usually in query-based retrieval results. They try to provide better ordering of the results to reduce the exposure of the spam to users.



Koutrika et al. [20] proposed that each incorrect content found in a system could be simply removed by an administrator. The administrator can go a step further and remove all content contributed by the user who posted the incorrect content, on the assumption that this user is a spammer (polluter).

2. User Trust Modeling (Static)

The aforementioned studies consider users' reliability as static at a specific moment. However, a user's trust in a social tagging system is dynamic, i.e., it changes over time. The tagging history of a user is better to consider, because a consistent good behavior of a user in the past can suddenly change by a few mistakes, which consequently ruins his/her trust in tagging.

3. User Trust Modeling (Dynamic)

A dynamic trust score, called SocialTrust, is derived for each user. It depends on the quality of the relationship with his/her neighbors in a social graph and personalized feedback ratings received from neighbors so that trust scores are updated as the social network evolves.

The dynamics of the system is modeled by including the evolution of the user's trust score to incent long-term good behavior and to penalize users who build up a good trust rating and suddenly "defect." It was shown that SocialTrust is resilient to the increase in number of malicious users, since the highly trusted users manage to keep them under control thanks to the trustaware feedback scheme introduced in this approach. It was also shown that SocialTrust outperforms TrustRank-based models, because SocialTrust model incorporates relationship quality and feedback ratings into the trust assessment so that bad behavior is punished.

4. Data Set

Data sets used for development and evaluation of trust modeling techniques have a wide range of diversity in terms of content, numbers of resources, tags and users, and type of spam. Social bookmarking is the most popularly explored domain for trust modeling, especially user trust modeling.

Algorithm

Trust modeling can be formulated as either a classification problem or a ranking problem, depending on the way of treatment.

In the classification problem, the results of an algorithm can be summarized by a confusion matrix from ground-truth data and predicted labels, which

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

contains the number of true positives, true negatives, false positives, and false negatives. From these values, classical measures such as a receiver operating characteristic (ROC), the area under the ROC curve (AUC), precision-recall (PR) curves, and F-measure can be derived.

Conclusion

In this article, we dealt with one of the key issues in social tagging systems: combatting noise and spam. We classified existing studies in the literature into two categories, i.e., content and user trust modeling. Representative techniques in each category were analyzed and compared. In addition, existing databases and evaluation protocols were re viewed. An example system was presented to demonstrate how trust modeling can be particularly employed in a popular application of image sharing and geotagging. Finally, open issues and future research trends were prospected. As online social networks and content sharing services evolve rapidly, we believe that the research on enhancing reliability and trustworthiness of such services will become increasingly important.

References

[1] Ivanov, I. , Lausanne, Vajda, P. , Jong-Seok Lee , Ebrahimi, T, In Tags We Trust: Trust modeling in social tagging of multimedia content, Signal Processing Magazine, IEEE (Volume:29 , Issue: 2)

[2] Malipeddi Kishore & MVB Chandra Sekhar, Privacy Protection & Surveillance over Online and Mobile Social Networking Sites, IJMETMR, http://www.ijmetmr.com/olnovember2014/MalipeddiKishore-MVBChandraSekhar-37.pdf, Volume No: 1(2014), Issue No: 11 (November)

[3] C. Marlow, M. Naaman, D. Boyd, and M. Davis, ―HT06, tagging paper, taxonomy, Flickr, academic article, to read,‖ in Proc. ACM HT, Aug. 2006, pp. 31–40.

[4] K. Liu, B. Fang, and Y. Zhang, ―Detecting tag spam in social tagging systems with collaborative knowledge,‖ in Proc. IEEE FSKD, Aug. 2009, pp. 427–431.

[5] L. S. Kennedy, S.-F. Chang, and I. V. Kozintsev, ―To search or to label?: Predicting the performance of searchbased automatic image classifiers,‖ in Proc. ACM MIR, Oct. 2006, pp. 249–258.

[6] P. Heymann, G. Koutrika, and H. Garcia-Molina, ―Fighting spam on social Web sites: A survey of approaches and future challenges,‖ IEEE Internet Comput., vol. 11, no. 6, pp. 36–45, Nov. 2007.

[7] L. von Ahn, M. Blum, N. J. Hopper, and J. Langford, ―CAPTCHA: Using hard AI problems for security,‖ in Proc. Eurocrypt, May 2003, pp. 294–311.

[8] L. von Ahn, B. Maurer, C. Mcmillen, D. Abraham, and M. Blum, ―reCAPTCHA: Human-based character recognition via Web security measures,‖ Science, vol. 321,no. 5895, pp. 1465–1468, Aug. 2008.

[9] Yahoo, Inc. (2011, Dec.). Flickr—Tags. [Online].Available: http://www. flickr.com/help/tags.

[10] G. Mori and J. Malik, ―Recognizing objects in adversarial clutter: Breaking a visual CAPTCHA,‖ in Proc.IEEE CVPR, June 2003, pp. I-134–I-141.

[11] M. Sahami, S. Dumais, D. Heckerman, and E. Horvitz,―A Bayesian approach to filtering junk e-mail,‖ AAAI Workshop on Learning for Text Categorization, Madison: WI, Tech. Rep. WS-98-05, July 1998.

Author Details:

Prasada Rao Mandala. M.Tech, (Ph.D), ACCP

D.R.S.Swetha