International Engineering Journal For Research & Development
Vol.1
Issue 2

E-ISSN NO:-2349-0721

Impact factor : 3.56

Trust modeling in social tagging of multimedia content

Prdeep Mane, likhesh kolhe

Asst.Professor, Dept. of Information Technology, Alamuri Ratnmala Engineering College, Shahapur., India.

ABSTRACT*: Social networks are very popular now days, as it facilitates search and retrieval of multimedia features. Anyway, 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.*

Keywords: *Spam, Multimedia Social Networks, Websites.*

INTRODUCTION

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.

BACKGROUND

Social networks and multimedia content sharing Web sites have become increasingly popular in recent years. Their service typically focuses on building online communities of people who share interests and activities, or are interested in exploring the interests and activities of others. At the same time, they have become a popular way to share and disseminate information. For

Example, users upload their personal photos and share them through online communities, letting other people comment or rate them.

One important challenge in tagging is to identify the most appropriate tags for given content, and at the same time, to eliminate noisy or spam tags. The shared content is sometimes assigned with inappropriate tags for several reasons. First of all, users are human beings and may commit mistakes. Moreover, it is possible to provide wrong tags on purpose for advertisement, self-promotion, or to increase the rank of a particular tag in automatic search engines. Consequently, assigning free-form keywords (tags) to multimedia content has a risk that wrong or irrelevant tags eventually prevent users from the benefits of annotated content.

TRUST MODELING

The social network approach to design large-scale systems has significant benefits including scalability, low cost of ownership, robustness, and ability to provide site autonomy. However, this approach has several drawbacks as well including trust issues and lack of coordination and control among the peers. We present a trust model for a social network structured large-scale network computing system and completely define the trust model and describe the schemes used in it. Central to the model is the idea of maintaining a recommender network that can be used to obtain references about a target domain. Simulation results indicate that the trust model is capable of building and maintaining trust and also identifying the bad domains. 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. Presented approaches are sorted based on their complexity from simple to advanced, separately for both content and user trust models.

CONTENT TRUST MODELING

Approaches for content trust modeling utilize features extracted from content information, users profiles and/or associated tags to detect specific spam content. Trust Rank relies on an important empirical observation called approximate isolation of the good set: good pages seldom point to bad ones. It starts from a set of seeds selected as highly qualified, credible and popular Web pages in the Web graph, and then iteratively propagates trust scores to all nodes in the graph by splitting the trust score of a node among its neighbors according to a weighting scheme. Trust Rank effectively removes most of the spam from the top-scored Web pages however it is unable to effectively separate low-scored good sites from bad ones, due to the lack of distinguishing features. 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. The administrator can go a step further and remove all content contributed by the user who posted the incorrect content

USER TRUST MODELING

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.

In user trust modeling, trust is given to each user based on the information extracted from a user's account, his/her interaction with other participants within the social network, and/or the relationship between the content and tags that the user contributed to the social tagging system. Given a user trust score, the user might be flagged as a legitimate user or spammer

EVALUATION

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. Some researchers dealing with bookmarks used a public data set released by BibSonomy as a part of the ECML PKDD Discovery Challenge 2008 on Spam Detection in Social Bookmarking Systems.

To model trust in other types of tagging systems, where spam is introduced through videos, tweets, or user profiles, data are usually crawled from the corresponding social network, like YouTube, Twitter, or MySpace, respectively. For example, Lee et al. [28] collected around 215,000 users and 4 million tweets from Twitter. Since this raw data are missing ground truth for evaluation, they manually labeled a small portion of users distinguishing between legitimate users,

To model trust in other types of tagging systems, where spam is introduced through videos, tweets, or user profiles, data are usually crawled from the corresponding social network, like YouTube, Twitter, or MySpace, respectively Since this raw data are missing ground truth for evaluation, they manually labeled a small portion of users distinguishing between legitimate users,

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 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.

VI.EXPERIMENTAL RESULTS

CONCLUSION AND FUTURE RESEARCH

In this article, we dealt with one of the key issues in social tagging systems: combating 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 retagging. 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] Wikimedia Foundation Inc. (2011, Dec.). Flickr. [Online]. Available: http://en.wikipedia.org/wiki/Flickr.

1) Pingdom Blog. (2011, Jan.). Internet 2010 in numbers. [Online]. Available: http://royal.pingdom.com/2011/01/12/internet-2010-in-numbers
2) 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.
3) 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.
4) L. S. Kennedy, S.-F. Chang, and I. V. Kozintsev, ―To search or to label?: Predicting the performance of search-based automatic image classifiers,‖ in Proc.ACM MIR, Oct. 2006, pp. 249–258.
5) 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.
6) 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.
7) 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.
8) Yahoo, Inc. (2011, Dec.). Flickr—Tags. [Online]. Available: http://www.flickr.com/help/tags.
9) 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.
10) 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.
11) D. Fetterly, M. Manasse, and M. Najork, ―Spam, damn spam, and statistics:Using statistical analysis to locate spam Web pages,‖ in Proc. ACM WebDB, June 2004, pp. 1–6.
12) S. Marti and H. Garcia-Molina, ―Taxonomy of trust: Categorizing P2P reputation systems,‖ Comput. Netw., vol. 50, no. 4, pp. 472–484, Mar. 2006.
13) A. Thomason, ―Blog spam: A review,‖ in Proc. CEAS, Aug. 2007.