User Authority Ranking Models for Community Question Answering*

Xuebo Liu[a], Haoran Xie[b], Yanghui Rao[a†], Qing Li[c], Fu Lee Wang[b], Tak-Lam Wong[d]

[a] School of Mobile Information Engineering, Sun Yat-sen University, China

[b] Caritas Institute of Higher Education, Hong Kong

[c] Department of Computer Science, City University of Hong Kong, Hong Kong

[d] Department of Mathematics and Information Technology, Hong Kong Institute of Education, Hong Kong

Abstract

The proliferation of knowledge-sharing communities has generated large amounts of data. Prominent examples of how user-generated content can be harnessed include the IBM Watson question answering sytem and Siri, the question answering application in iPhones. Faced with massive data, user authority ranking is important to the development of question answering and other e-commerce services. In this study, we propose three probabilistic models to rank the user authority of each question. Compared to the existing approaches focused on the user relationship primarily, our method is more effective because we consider the link structure and topical similarities between users and questions simultaneously. We use a real-world data set from *Zhihu*, a popular community question answering website in China to conduct experiments. Experimental results show that our model outperforms other methods in ranking the user authority.

Keywords: User authority ranking; topic modeling; community question answering

* A preliminary version of this paper has been published in [1].

† Corresponding author. Tel.: +86 15989787592. E-mail address: raoyangh@mail.sysu.edu.cn

1. Introduction

Question answering (QA) systems involve different techniques for answer extraction, answer presentation, question analysis, and so forth. The successes of IBM's Watson and Apple's Siri have highlighted Q/A research and commercialization opportunities [2]. Community Question Answering (CQA) websites, such as *Quora* (www.quora.com), *Stackoverflow* (stackoverflow.com), *Yahoo!Answer* (answers.yahoo.com) and *Zhihu* (zhihu.com), are extremely popular in recent years. In CQA websites, users can raise their own questions, answer the questions posted by others and read the answers to a question. It is interesting that CQA websites have strong social characteristics, which are different from the traditional ones such as *Baidu Knows* (zhidao.baidu.com). Take *Zhihu* as an example, the content read by user *u* in his/her home page depends on the people *u* have followed (i.e., "friends" of *u*). Therefore, a user is able to see the questions from all his/her friends, their answers and their "liked" answers. By allowing the interaction between users, CQA websites have accumulated millions of questions and their answers over time [3].

Although CQA websites have attracted many users to contribute knowledge, they bring several new challenges to provide high quality services [4]: (1) **Poor expertise matching**: Many questions are difficult to be recommended to the expert with the best-matching interest and ability, which results in suboptimal answers and delay of satisfaction. (2) **Answer sparisity**: It is hard for users to find questions that they are interested in, thus there are some questions with few answers. (3) **Duplicated questions**: If a user can not find a satisfied answer for one question, he/she may post a new duplicated question, the answers of which may be offered in previous similar questions. Thus, it is necessary to develop user authority ranking models in CQA websites to solve these problems.

Traditional methods of user authority ranking are based on the relationship between users primarily [5, 6], which are insufficient to get a unique ranking list for each question. For example, the appropriate users to answer different questions may differ. If we rank the user authority based on the relationship between users only, the user who is a movie star will have many followers and achieve a high ranking. However, he/she may not be appropriate to answer the questions about computers. We here propose a new framework based on the link structure and topical similarities between users and questions, which can get a unique user authority ranking list for each question. In our framework, all answers posted by a user are considered as a document, the descriptions and all answers to a question are also treated as a document; we then apply Latent Dirichlet Allocation (LDA) [7] to extract topics from both users and questions. After generating these topics, we can measure the topical similarities between questions and answers, and rank users based on the user relationship and topical similarities.

The rest of the paper is organized as follows. Section 2 introduces the related work on user authority ranking in CQA websites. In section 3, the proposed framework is described in detail. Section 4 illustrates the data set used and the experimental results. We conclude our paper and make a future plan in section 5.

2. Related Work

In this section, we review previous studies on user authority ranking by relationships, similar question recommendation, and topic-sensitive expert recommendation, which will shed light on user authority ranking in CQA websites.

2.1 User Authority Ranking by Relationships

The existing algorithms computed the user authority in CQA websites by the relationship between users primarily. *Bouguessa et. al.* [8] proposed a method to find the expert based on the best answer the user posted. *Jurczyk and Agichtein* [9] applied HITS algorithm [6] to rank the user authority by the followed-following relationship. *Zhang et. al.* [10] proposed an algorithm based on the users' specialty. Although the algorithms that analyzing the relationship between users have achieved the desired results, there are some challenging problems to be addressed by these algorithms, e.g., the identified expert could not give the satisfied answer to the field he/she is not skilled in. Thus, some other algorithms based on the extra information were proposed. *Guo et. al.* [11] proposed an algorithm to explore the similarity between users and askers by the tags of users. *Liu et. al.* [12] proposed a probabilistic language model to predict the best answerer of questions.

2.2 Similar Question Recommendation

For similar question recommendation, *Jeon et. al.* [13] proposed a statistical approach that explored the semantic features to measure the question similarity. *Li et. al.* [14] designed a strategy via machine translation rather than the simple cosine similarity approaches. *Hao et. al.* [15] developed a pattern-based algorithm by seed patterns and a semi-supervised approach. *Wu et. al.* [16, 17] proposed an approach that exploited both the user interest and feedback, in which the historical data was employed to generate the user interest. These studies showed the benefits of exploiting topic model over the conventional methods. Similar to some extant research, we explore the semantics of questions using topic models. However, our method also considers the link structure of users, which was neglected in most approaches for the similar question recommendation.

2.3 Topic-sensitive Expert Recommendation

Recently, some algorithms of expert recommendation have been proposed to consider both the relationship between users and the similarity of topics. TwitterRank [18] is a typical algorithm developed to estimate the influence of Twitter users based on the followed-following relationship and topical similarities. *Zhou et. al.* [19] proposed a topic-sensitive page ranking (TSPR) algorithm for expert recommendation in CQA websites. The algorithm first employed LDA to extract the topics of users, and then recommended experts by the number of answers a user posted and the similarity between the user and the question initiator [20]. *Zhao et. al.* [21] designed a method to model experts and topics jointly by incorporating each user's contribution dynamically. *Chen et. al.* [22] presented a rating system of user reputation based on user comments. *Pal et. al.* [23] proposed an algorithm based on Gaussian mixture models to identify topical authorities in microblogs. *Liu et. al.* [24] proposed a novel topic model for expert recommendation in CQA websites.

Different from these algorithms, our method considers both the link structure of users and the topical similarity between users and questions. To the best of our knowledge, this is the first piece of work to consider the topical similarity between users and questions for user authority ranking in CQA websites.

3. Methodology

The proposed user authority ranking (UAR) models aim to rank the user authority for each question, by exploiting the relationship between users and the topical similarities between users and questions. Conventionally, the more "likes" a user receives, the higher authority he/she will achieve, which is similar to the ranking approach of websites. It is important to note that the weight of each "likes" given to an answer is different from the others. Assume that both user *B* with higher authority and user *C* with lower authority give "likes" to user *A*, the "likes" from user *B* is often more powerful to improve the authority of user *A* than user *C*. Therefore, we estimate of influence of users based on the iteration method of PageRank [5]. In addition, due to that each question is consisted of different topics and the topics those different users are familiar with are various, we also take the similarity between users and topics into consideration.

3.1 Topic Extraction of Users

We apply LDA [7] to perform the topic extraction of users. LDA is an unsupervised topic extraction model based on the bag-of-word assumption. It treats the text as a vector whose characteristic of each dimension is the frequency of words appeared in the text. Each text can be expressed as the probability distribution of a series of topics and each topic can be represented by the probability distribution of a series of words. LDA is a nature model for topic extraction of documents, in which the probability distributions of topics for each text, and words for each topic can be estimated by the Gibbs sampling algorithm [25]. The detailed process of user topic extraction in our UAR models is as follows:

Firstly, we consider all answers of a user posted as a text, and the one-to-one mapping between a user and its text is established. All answers to a question are considered as another text. The one-to-one mapping between a question and its text is also established in a similar way.

Secondly, we adopt LDA to train the text of all users and questions, in which, the probability distribution of the topics corresponding to each text θ and probability distribution of the words corresponding to each topic φ can be estimated.

Thirdly, we keep θ unchanged and carry out the Gibbs sampling only with the input of user text.

Finally, we get a new φ_u and let $UZ = \varphi_u$, which represents the topic distribution of a text. The definition of UZ is shown as follows:

Definition 1 *UZ: the matrix of* $U \times Z$. *U is the number of users. Z is the number of topics.* UZ_{ij} *denotes the number of words assigned to topic* z_j *as appeared in all answers posted by user* u_i.

3.2 User Authority Transition Matrix

In the past researches on social networks, the relationships of following and followed between users [17], and the number of answers provided by each user [18, 23] are often employed to generate the user authority transition matrix by iterative computation [25]. Different from the traditional social networks, the approval mechanism is introduced into the CQA websites. Conventionally, the more "likes" a user receives, the higher authority he/she will achieve. Meanwhile, the weights of a "likes" given by different users are varied, i.e., the approval by the expert in a certain field is more powerful to improve the authority of the user whom he/she give "likes" to in this field. In addition, the "likes" given by a user who seldom make approval is of higher value in comparison to those who often delivers "likes". Therefore, we consider the user's authority ranking as a *Markov Chain* [26], and the transition matrix is calculated as follows:

$$T_{i,j} = \frac{V_{j\to i}}{\sum_{for\ every\ user\ k} V_{j\to k}} * sim_{i,j}, \quad (1)$$

where T represents the user authority transition matrix. $T_{i,j}$ represents the influence of user i to user j, $V_{j\to i}$ represents the number of "likes" that user j gives to user i, and the denominator is the summation of the number of "likes" that user j gives to all users, $sim_{i,j}$ represents the similarity between user i and j, which can be estimated as follows:

$$sim_{i,j} = 0.5 * \frac{UZ'_i \cdot UZ'_j}{\|UZ'_i\| \times \|UZ'_j\|} + 0.5, \quad (2)$$

where UZ' is the row-normalized form of UZ, i.e., the L_1-norm of each row is 1, UZ_i and UZ_j represent the degree of interest of user i and user j in all topics, respectively. The above equation is the normalized *cosine similarity* between UZ_i and UZ_j. As each element in the transition matrix of Markov Chain, i.e., the transition probability ranges from 0 to 1, yet the value of cosine similarity ranges from -1 to 1, we apply $0.5 + 0.5 * cosine\ similarity$ to normalize the value.

Another way to measure the similarity between user i and user j is based on the normalized Euclidean distance between UZ_i and UZ_j, as follows:

$$sim_{i,j} = \frac{1}{1+\sqrt{\sum_{z=1}^{Z}(UZ'_{iz}-UZ'_{jz})^2}}, \tag{3}$$

where Z is the number of topics, UZ'_{iz} reperesents user i's interest in topic z, which is the z-th element of UZ_i, $\sqrt{\sum_{z=1}^{Z}(UZ'_{iz} - UZ'_{jz})^2}$ is the Euclidean distance ED between UZ_i and UZ_j, and we use $\frac{1}{1+ED}$ to normalize the value.

We also propose a topic-sensitive transition matrix to rank the user's authority, as follows:

$$T_z(i,j) = \frac{V_{j\to i}}{\sum_{for\ every\ user\ k} V_{j\to k}} * sim_z(i,j), \tag{4}$$

where T_z represents the user authority transition matrix of topic z. $T_z(i,j)$ represents the influence of user i to user j in topic z. $V_{j\to i}$ represents the number of "likes" that user j gives to user i, and the denominator is the summation of the number of "likes" that user j gives to all users. $sim_z(i,j)$ denotes the similarity between user i and j in topic z.

The similarity function in the topic-sensitive transition matrix is:

$$sim_z(i,j) = 1 - \left(UZ'_{iz} * ln\left(\frac{UZ'_{iz}}{UZ'_{jz}}\right) + UZ'_{jz} * ln\left(\frac{UZ'_{jz}}{UZ'_{iz}}\right)\right), \tag{5}$$

where UZ' is the row-normalized form of matrix UZ, i.e., the L_1-norm of each row is 1. UZ'_{iz} indicates the degree of interest of user i in topic z. The above equation is the normalized *relative entropy* between UZ_i and UZ_j. If the degrees of interest of user i and j in topic z are close, both $ln\left(\frac{UZ'_{iz}}{UZ'_{jz}}\right)$ and $ln\left(\frac{UZ'_{jz}}{UZ'_{iz}}\right)$ tend to approximate 0, while the value of sim approximately tends to 1. Otherwise, the value of sim will be quite small. The larger the value of sim is, the more similar user i and j in topic z will be.

3.3 User Authority Ranking for Each Topic

In Section 3.2, we get the user authority transition matrix iteratively. Next, UAR models take the approval relationship between users and the topical similarity into account to compute the authority ranking of users in topic *z*:

$$UR_z = \lambda T \cdot UR_z + (1-\lambda) * UZ_z'', \tag{6}$$

where UR_z represents the user authority ranking of topic z, λ is a weighting parameter between 0 and 1. A lager value of λ indicates that the approval relationship between users has a greater influence on the authority ranking. While a smaller value of λ indicates that the degree of interest of the user to topic *z* has a greater influence on the authority ranking. T is the transition matrix described in Section 3.2. UZ_z'' is the column-normalized form of matrix UZ, i.e., the L_1-norm of each column is 1. It represents the degree of interest of each user to topic *z*.

After convergence, we get the final result of the user authority ranking for each topic.

3.4 Topic Extraction of Questions

To extract the topics of questions, we first treat all answers to a question as a document. Then, we apply LDA trained in Section 3.1, and carry out Gibbs sampling on the documents of questions, in which θ is kept unchanged. Finally, we get φ_q and let $QZ = \varphi_q$. The definition of QZ is described as follows:

Definition 2 *QZ: the matrix of Q × Z. Q is the number of questions. Z is the number of topics.* QZ_{ij} *represents the number of words that assigned to topic* z_j *in all the answers of question* q_i.

3.5 User Authority Ranking for Each Question

Since we have estimated both *QZ* (the topic distribution of every question) and *UR* (the user authority ranking of every topic), we get the user authority ranking of each question by multiply the two matrices (i.e., *Bayes*'s rule):

$$QR = QZ \cdot UR, \tag{7}$$

where QZ represents the topic distribution of each question, UR denotes the user authority ranking of each topic. The multiplication result is the user authority ranking of every question. The detail of our UAR models is shown in Algorithm 1.

ALGORITHM 1. The UAR Models

Input: The user's information (answers, "likes" received and the content of all answers he/she posted), and the question's information (its description, the content of all its answers).
Output: The user authority ranking of every question QR.
Parameters: topic number Z, damping factor parameter λ, iteration times C

Use LDA to train the documents of all users and questions and get UZ
for *each user i* **do**
  for *each user j* **do**
    Estimate $T_{i,j}$ according to Eq. 1 and Eq. 2/3
    Estimate $T_z(i,j)$ according to Eq. 4 and Eq. 5
  end for
end for
for *each topic z* **do**
  for $i = 1, \ldots, C$ **do**
    Update UR_z according to Eq. 6
  end for
end for
Use the trained LDA to infer the documents of questions and get QZ
Estimate QR according to Eq. 7

4. Experiments

In this section, we introduce the preliminaries (i.e., data set, parameters settings and baselines) for experiments and report the experimental results.

4.1 Data Set

A real-world data set from *Zhihu* was employed for the experiment. *Zhihu* is one of the most popular question answering communities in China. Different from *StackOverflow* and *Yahoo!Answer*, we can get all users who "like" an answer, thus we can get user authority transition matrix based on users' "like" relationships.

We collected 576 questions and 209309 answers from *Zhihu*, and the total number of users is 9043. The *Jieba Chinese Text Segmentation* (http://github.com/fxsjy/jieba) was employed to perform the Chinese word segmentation. The detail process of preparing the above dataset is as follows:

- For each question, its description, contents of all its answers and the real ranking of all answers were crawled.

- For each user, the number of friends, followers, answers, “likes” received and the contents of all answers he/she posted were crawled.

4.2 Parameters Setting

We have several parameters in our models, i.e., the Dirichlet hyper-parameters α and β, the number of topics Z, and the damping factor parameter λ used in PageRank. In this paper, we set the Dirichlet priors $\alpha = 50/Z$, and $\beta = 0.05$ as in [27]. We run LDA with 1000 iterations of Gibbs sampling. After trying several different numbers of topics, we empirically set $Z = 50$. We choose these parameter settings because they give coherent and meaningful topics for our data set. Table 1 shows top five words of 20 topics generated by LDA.

Table 1: Top five words of different topics discovered by LDA.

TOPIC 1	英雄 (Hero)	Dota (Dota)	技能 (Skill)	比赛 (Contest)	游戏 (Game)
TOPIC 2	时间 (Time)	学习 (Study)	事情 (Matter)	很多 (Many)	工作 (Work)
TOPIC 3	情况 (Situation)	很多 (Many)	来说 (In terms of)	时间 (Time)	两个 (Two)
TOPIC 4	知乎 (Zhihu)	http (http)	用户 (User)	内容 (Content)	网站 (Website)
TOPIC 5	用户 (User)	产品 (Product)	互联网 (Internet)	微信 (WeChat)	需求 (Requirement)
TOPIC 6	人类 (Human)	科学 (Science)	研究 (Research)	数学 (Math)	理论 (Theory)
TOPIC 7	中国 (China)	政治 (Politics)	社会 (Society)	台湾 (Taiwan)	国家 (Country)
TOPIC 8	工作 (Work)	北京 (Beijing)	很多 (Many)	公司 (Company)	几个 (Few)
TOPIC 9	历史 (History)	中国 (China)	皇帝 (Emperor)	时代 (Era)	文化 (Culture)
TOPIC 10	经济 (Economy)	发展 (Develop)	中国 (China)	政府 (Government)	市场 (Market)
TOPIC 11	喜欢 (Like)	生活 (Life)	世界 (World)	人生 (Life)	爱情 (Love)
TOPIC 12	新闻 (News)	时间 (Time)	媒体 (Media)	视频 (Video)	节目 (Program)
TOPIC 13	设计 (Design)	设计师 (Designer)	照片 (Photo)	效果 (Effect)	图片 (Picture)
TOPIC 14	汽车 (Car)	司机 (Diver)	车辆 (Vehicle)	开车 (Car Driving)	驾驶 (Drive)

TOPIC 15	中国 (China)	美国 (USA)	日本 (Japan)	国家 (Country)	国内 (Domestic)
TOPIC 16	公司 (Company)	企业 (Enterprise)	投资 (Invest)	业务 (Business)	保险 (Insurance)
TOPIC 17	建筑 (Architecture)	设计 (Design)	房子 (House)	城市 (Urban)	保护 (Protect)
TOPIC 18	电影 (Movie)	作品 (Works)	动画 (Animation)	故事 (Story)	导演 (Director)
TOPIC 19	比赛 (Contest)	球员 (Player)	球队 (Team)	防守 (Defense)	足球 (Football)
TOPIC 20	用户 (User)	软件 (Software)	功能 (Function)	系统 (System)	手机 (Cellphone)

When computing the user authority ranking of each topic, the damping factor parameter λ is set to be 0.85 based on cross-validation.

4.3 Evaluation Metrics

To measure the accuracy of various methods, two metrics commonly used in information retrieval are adopted:

- *Mean Reciprocal Rank (MRR)*: This index is the multiplicative inverse of the rank of the first retrieved expert for each topic.
- *nDCG@K*: This index measures the performance of a recommendation system based on the graded relevance of the recommended entities. It varies from 0.0 to 1.0, with 1.0 representing the ideal ranking of all entities:

$$nDCG@K = \frac{1}{|Q|} \sum_{q \in Q} \frac{\sum_{j=1}^{K} \frac{1}{log_2(j+1)} score(M_{q,j})}{IdealScore(K,q)}. \tag{8}$$

In the above, Q is the set of questions. $M_{q,j}$ is the *j*-th expert generated by method M for question q. $score(M_{q,j}) = 2^{v(M_{q,j})} - 1$, where $v(M_{q,j})$ is the ground truth score for the expert $M_{q,j}$. $IdealScore(K,q)$ is the ideal ranking score of the top-*K* experts for question q.

4.4 Comparison with Baselines

In this subsection, we denote the proposed model of user authority ranking using the topic-sensitive transition matrix as **UART**, and employ the following baselines for comparison:

- In-degree by number of followers (**IDF**): this algorithm measures the authority of users according to the number of followers. The more followers a user has, the higher value of the user authority will be.
- In-degree by number of “likes” (**IDV**): the algorithm measures the authority of users according to the number of “likes” received. The more “likes” a user owns, the higher value of the user authority will be.
- PageRank by number of followers (**PRF**): the algorithm generates the user ranking by applying PageRank with the number of followers.
- PageRank by number of “likes” (**PRV**): the algorithm generates the user ranking by applying PageRank with the number of “likes”.
- Topic-Sensitive PageRank (**TSPR**): the algorithm generates the user ranking of each question according to the following aspects: (1) the user topical similarity between the asker and other users; (2) the number of times that users answered the questions raised by the asker [19].

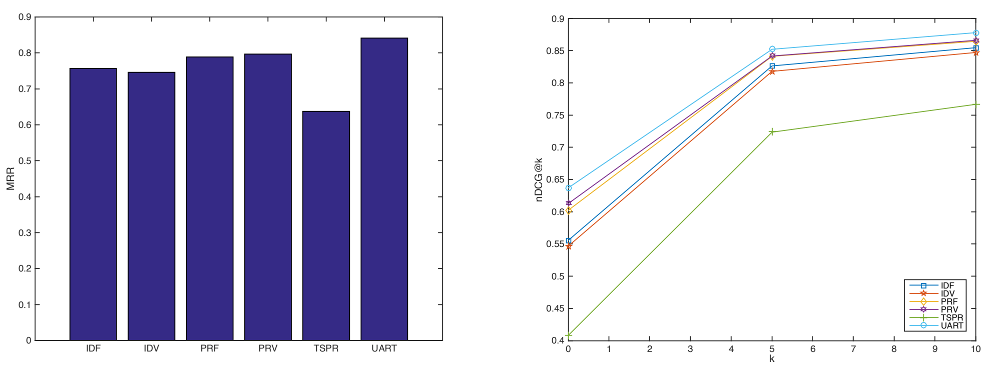

Fig. 1. The performance of different methods.

The performance of all methods is presented in Figure 1, from which we can observe that the proposed UART outperformed other methods for both MRR and nDCG metrics. The results have verified our observation that it is more effective to take the topical similarity between questions and users into consideration when ranking the user authority. Futhermore, we investigated the performance of all methods in terms of nDCG@k by varying k from 0 to 10. We can observe that the gap among various methods become smaller when k becomes larger. A possible reason is that the UART can recommend high quality results in top-5 and then ranking results are getting similar when k is larger than 10.

4.5 Comparison with Topic-independent Models

As stated in Section 3.2, we also propose two models of user authority ranking using the topic-independent transition matrix. For these two methods, the one of estimating the similarity between user i and user j by the normalized *cosine similarity* is denoted as **UARC**, and the one of using the normalized *Euclidean distance* to measure the similarity is denoted by **UARE**.

Table 2: The performance and efficiency of different models.

Models	MRR	nDCG	Complexity
UART	0.84114	0.87893	O(UUZ) + O(ZC)
UARC	0.81583	0.87119	O(UU) + O(ZC)
UARE	0.81944	0.87139	O(UU) + O(ZC)

The comparison of different models are shown in Table 2, in which U is the number of users, Z is the number of topics, and C is the iteration times. Although the performance of UARC and UARE is worse than that of UART, we can find that the topic-independent ones (UARC and UARE) are much faster than UART. The main reason is that the topic-independent models only need to compute once when estimating the transition matrix, but the topic-sensitive model need to estimate the transition matrix for each topic so that the complexity are increased from O(UU) to O(UUZ) (where Z is the number of topics). The trade-off here is that UART can achieve better accuracy in terms of MRR and nDCG. Therefore, the above three models can be employed in different applications. For those applications concerning more on accuracy, UART can be more suitable than the other two. Conversely speaking, both UARC and UARE could be more appropriate for systems requiring shorter response time.

5. Conclusion

Many promising question answering system application areas, including education, health, and defense have been identified [2]. This paper proposed an effective framework to estimate the user authority ranking in CQA social networks, which is helpful to product recommendations [28, 29] and educational module creation [30] in the big data era. We evaluated the framework by a real-world data set from *Zhihu*, and the experimental results demonstrated the effectiveness of our models when compared to other existing methods.

In the future, we plan to present a new probabilistic model for short text topic extraction when there are few answers to a question. The sparsity of content in short documents brings new challenges to topic modeling and user authority ranking. On one hand, the frequency of words, which is quite important to model long text, play limited discriminative role in short documents. On the other hand, inferring topics and ranking the user authority from large-scale short documents becomes a critical task for many areas. Therefore, user authority ranking in short text deserves further research.

Acknowledgments

The research described in this paper has been supported by “the Fundamental Research Funds for the Central Universities” (Project Number: 46000-31610009) and a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (UGC/FDS11/E06/14). This work has also been supported, in part, by a Strategic Research Grant (Project no. 7004218), and an Applied Research Grant (Project no. 9667095), both of City University of Hong Kong.

References

[1] X. Liu, S. Ye, X. Li, Y. Luo, Y. Rao, ZhihuRank: A topic-sensitive expert finding algorithm in community question answering websites. In Procceedings of the 14th International Conference on Web-based Learning (ICWL), 2015 (to appear).

[2] H. Chen, R.H.L. Chiang, V.C. Storey, Business intelligence and analytics: from big data to big impact. MIS Quarterly 36 (4) (2012) 1-24.

[3] Y. Liu, S. Li, Y. Cao, C.-Y. Lin, D. Han, Y. Yu, Understanding and summarizing answers in community-based question answering services. In Proceedings of the 22nd International Conference on Computational Linguistics (Coling), 2008, pp. 497-504.

[4] E. Agichtein, C. Castillo, D. Donato, A. Gionis, G. Mishne, Finding high-quality content in social media. In Proceedings of the International Conference on Web Search and Web Data Mining (WSDM), 2008, pp. 183-194.

[5] S. Brin and L. Page, The anatomy of a large-scale hypertextual web search engine. Computer Network and ISDN Systems, 30 (1-7) (1998) 107-117.

[6] J. M. Kleinberg, Authoritative sources in a hyperlinked environment. Journal of the ACM, 46 (5) (1999) 604-632.

[7] D. M. Blei, A. Y. Ng, M. I. Jordan, Latent dirichlet allocation. Journal of Machine Learning Research, 3 (2003) 993-1022.

[8] M. Bouguessa, B. Dumoulin, S. Wang, Identifying authoritative actors in question-answering forums: the case of yahoo! answers. In Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2008, pp. 866-874.

[9] P. Jurczyk and E. Agichtein, Discovering authorities in question answer communities by using link analysis. In Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management (CIKM), 2007, pp. 919-922.

[10] J. Zhang, M. S. Ackerman, L. Adamic, Expertise networks in online communities: structure and algorithms. In Proceedings of the 16th International Conference on World Wide Web (WWW), 2007, pp. 221-230.

[11] J. Guo, S. Xu, S. Bao, Y. Yu, Tapping on the potential of q&a community by recommending answer providers. In Proceedings of the 17th ACM Conference on Information and Knowledge Management (CIKM), 2008, pp. 921-930.

[12] X. Liu, W. B. Croft, M. Koll, Finding experts in community-based question answering services. In Proceedings of the 2005 ACM CIKM International Conference on Information and Knowledge Management (CIKM), 2005, pp. 315-316.

[13] J. Jeon, W. B. Croft, and J. H. Lee, Finding similar questions in large question and answer archives. In Proceedings of the 2005 ACM CIKM International Conference on Information and Knowledge Management (CIKM), 2005, pp. 84-90.

[14] S. Li and S. Manandhar, Improving question recommendation by exploiting information need. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference (ACL), 2011, pp. 1425-1434.

[15] T. Hao and E. Agichtein, Finding similar questions in collaborative question answering archives: toward bootstrapping-based equivalent pattern learning. Information Retrieval 15 (3-4) (2014) 332-353.

[16] H. Wu, Y. Wang, X. Cheng, Incremental probabilistic latent semantic analysis for automatic question recommendation. In Proceedings of the 2008 ACM Conference on Recommender Systems (RecSys), 2008, pp. 99–106.

[17] M. Qu, G. Qiu, X. He, C. Zhang, H. Wu, J. Bu, C. Chen, Probabilistic question recommendation for question answering communities. In Proceedings of the 18th International Conference on World Wide Web (WWW) 2009, pp. 1229–1230.

[18] J. Weng, E.-P. Lim, J. Jiang, Q. He, Twitterrank: finding topic-sensitive influential twitterers. In Proceedings of the Third International Conference on Web Search and Web Data Mining (WSDM), 2010, pp. 261-270.

[19] G. Zhou, S. Lai, K. Liu, J. Zhao, Topic-Sensitive Probabilistic Model for Expert Finding in Question Answer Communities. In Proceedings of the 21st ACM International Conference on Information and Knowledge Management (CIKM), 2012, pp. 1662-1666.

[20] T. H. Haveliwala, Topic-sensitive pagerank. In Proceedings of the 11th International World Wide Web Conference (WWW), 2002, pp. 517-526.

[21] T. Zhao, N. Bian, C. Li, M. Li, Topic-level expert modeling in community question answering. In Proceedings of the 13th SIAM International Conference on Data Mining (SDM), 2013, pp. 776-784.

[22] B.-C. Chen, J. Guo, B. Tseng, J. Yang, User reputation in a comment rating environment. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2011, pp. 159-167.

[23] A. Pal and S. Counts, Identifying topical authorities in microblogs. In Proceedings of the Forth International Conference on Web Search and Web Data Mining (WSDM), 2011, pp. 45-54.

[24] Y. Liu, M. Qiu, S. Gottipati, F. Zhu, J. Jiang, H. Sun, Z. Chen, CQARank: Jointly model topics and expertise in community question answering. In Proceedings of the 22nd ACM International Conference on Information and Knowledge Management (CIKM), 2013, pp. 99-108.

[25] S. Geman and D. Geman, Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6 (6) (1984) 721-741.

[26] J. R. Norris, Markov chains. Cambridge University Press, 1998.

[27] T. Griffiths and M. Steyvers, Finding scientific topics. The National Academy of Sciences, 101 (1) (2004) 5228-5235.

[28] E.-P. Lim, H. Chen, G. Chen, Business intelligence and analytics: Research directions, ACM Transactions on Management Information Systems 3 (4) (2013) 17-27.

[29] Y. Cai, R.Y.K. Lau, S.S.Y. Liao, C. Li, H.-F. Leung, L.C.K. Ma, Object typicality for effective Web of Things recommendations. Decision Support Systems 63 (2014) 52-63.

[30] J.L. Zhao, S. Fan, D. Hu, Business challenges and research directions of management analytics in the big data era. Journal of Management Analytics 1 (3) (2014) 169-174.