Document recommendation based on interests of co-authors for brain science

Zhong, Han; Huang, Zhisheng

published in
Health Information Science
2019

DOI (link to publisher)
10.1007/978-3-030-32962-4_11

document version
Publisher's PDF, also known as Version of record

document license
Article 25fa Dutch Copyright Act

Document Recommendation Based on Interests of Co-authors for Brain Science

Han Zhong[1(✉)] and Zhisheng Huang[2]

[1] College of Information Technology and Network Security, People's Public Security University of China, Beijing 100038, China
z.h0912@163.com

[2] Knowledge Representation and Reasoning Group, Vrije University Amsterdam, 1081 HV Amsterdam, The Netherlands
huang@cs.vu.nl

Abstract. Personalized knowledge recommendation is an effective measure to provide individual information services in the field of brain science. It is essential that a complete understanding of authors' interests and accurate recommendation are carried out to achieve this goal. In this paper, a collaborative recommendation method based on co-authorship is proposed to make. In our approach, analysis of collaborators' interests and the calculation of collaborative value are used for recommendations. Finally, the experiments using real documents associated with brain science are given and provide supports for collaborative document recommendation in the field of brain science.

Keywords: User and co-author · Interests · Recommendation · Semantic technology · Brain science

1 Introduction

With the development of big data and artificial intelligence, the field of brain science based on digital resources has become a hot topic in recent years [1]. The style of people's life and knowledge renewal speed are accelerated with the ever-growing of brain science data. How to make the users to find interesting contents from large scale of data resources quickly and accurately has become an inevitable problem in the development of brain science. However, personal recommendation of data and interests provides an ideal way to solve this problem. Personalized recommendation is a mode of information service to provide information for users based on the needs of users [2, 3]. Establishing the interest model of author-topic and analyzing the interest degree of co-author' topics is a hot research interest to predict that the target users may be interested in the theme of resources.

There are many recommendation methods such as recommendation algorithm based on contents, recommendation algorithm based on rules [4] and collaborative recommendation algorithm [5]. Despite there are advantages and disadvantages, these algorithms are capable of mining users' potential interests and providing new

H. Wang et al. (Eds.): HIS 2019, LNCS 11837, pp. 108–118, 2019.
https://doi.org/10.1007/978-3-030-32962-4_11

learning resources. Shehata et al. [6] discovered users' interests and model to study the semantic relations between sentences through the concept map of ontology. Zhang et al. gave a kind of interest points of attention degree from semantic and structural features [7]. Chen et al. realized the theme recommendation based on the users' interests through the combination of the graph abstract method with the similarity algorithm based on the contents [8]. Cai et al. introduced the mechanism of trust of collaborative filtering to make the recommendation [9]. Guo et al. gave a novel social recommendation method and incorporate item relations using a probabilistic matrix factorization framework from the items' perspective [10]. Chen et al. introduced a novel attention mechanism in collaborative filtering to address the challenging item- and component-level implicit feedback in multimedia recommendation [11]. Jiang et al. proposed an author topic model-based collaborative filtering method to facilitate comprehensive points of interest recommendations for social users [12]. However, existing recommendation approaches often ignored the relationship between users and recommended objects. There are some problems among the literature recommendation because many users have access to record data in large scale. At the same time, the results need to be sorted and optimized based on a huge number of results from the reasoning relationships.

Meanwhile, knowledge service based on the semantic technology is becoming a new technology of information system from new generation of Web and many systems of semantic technology have been put into applications [13]. We are interested in a series of semantic technology application systems based on the platform of LarKC, especially based on the platform of brain informatics knowledge service. The platform of LarKC is the major semantic technology research and development project in the European Union's seventh research framework project. The name LarKC stands for the Large Knowledge Collider, which commits to develop a platform for massive semantic data processing and reasoning [14, 15]. It is important to use the method of recommendation systems among the process of personalized recommendation. This paper introduced the method of calculating co-author' interests and constructed the model of authors - topic interests.

The remainder of this paper is organized as follows. Sections 2 and 3 describe the model of interests between authors and topics and the model of interests between co-author and topics, respectively. Section 4 describes the model of interests between author and co-author. Section 5 provides an experiment and discussion. Section 6 discusses the knowledge service system of brain science based on semantic technology. Finally, Sect. 7 gives conclusions and future work.

2 The Model of Interests Between Authors and Topics

Users' interests on the topic of the literature can be understood that providing appropriate resources and content services is important for users. Selecting the topics of possible interests is the most favorable way for recommendation systems [16, 17, 18]. In this paper, we provide potential interests based on the interests of author and co-author.

In this paper, we consider that the authors interested in the topics instead of the relationship between the documents' topics in order to describe the model of interests. The user-topic model reflects the degree of interests in different concept topics.

A document can be described through some attributes, such as the title of a document, authors, the date of publication, the press, the form of document and so on. We simplified this document into a kind of three basic attributes of DOI, the author (authors) and topic (Topics) to identify it in this paper. So, some definitions are made as follows:

Definition 1 (Document). *A document d = <DOI, Authors, Topics> where Authors are a sequence of* $<author_1, author_2, ..., author_n>$ *, Topics =* $\{topic_1, topic_2, ..., topic_m\}$

A document is composed of classification number, author and topics in general. The authors of a document constitute a sequence according to the order of authorship and the topic is the set of the keywords. However, one document usually owns more than one author. So the collection consisting of some documents is defined as follows:

Definition 2 (Document Repository). *A document repository D is a set of documents* $\{d_1, d_2, ..., d_k\}$*, among them,* $di \in D$

Each document includes some authors in the collection of documents. So, the set of all authors in a document is defined as follows:

Definition 3 (Author Set). *Given a document repository D, the author set of D is defined as:*

$$Author\ Set\,(D) = \{author_i |\, \exists d = \ <DOI,\ Authors,\ Topics> \ \in D, \exists i,\ Authors = <\ldots,\ author_i,\ \ldots>\}$$

Each document includes some topics. So, the set of topics among the documents is defined as follows:

Definition 4 (Topic Set). *Given a topic repository D, the topic set of D is defined as:*

$$Topic\ Set\,(D) = \{topic_i |\, \exists d = \ <DOI,\ Authors,\ Topics> \ \in D, \exists i,\ topics = <\ldots,\ topic_i,\ \ldots>\}.$$

The interest degree of topic can be measured by the number of the topic publication and this kind of measure does not take new interests and the order of this author into account. Because the first author and the last author of this paper may be different on the level of interests in the topics. We will discuss other forms of improvements in this metric mode. So, the interest of topics among the authors is defined as follows:

Definition 5 (Interest). *A set of documents is D, the interest of topics of an author is defined as follows:*

$$\text{Interest}\,(\text{author}, \text{topic}) = \frac{|\{\text{d} : \text{d} = \ <\text{DOI}, <\ldots, \text{author}, \ldots> , \{\ldots,\ \text{topic},\ \ldots\} > \}|}{|\text{All Published Papers}|} \tag{1}$$

Among them, $|AllPublishedPapers|$ *refers to the number of all the published articles, that is* $All\ Published\ Papers = \{d : d = <DOI, <\ldots, author, \ldots>, Topics> \ \in D\}$.

The value of interest in a topic is actually a regularized (Normalized) value (i.e. it belongs to [0,1]) through above formal definition. It indicates that this author interested in this topic extremely if the value of interests is 1. And it shows that this author is not interested in this topic if the value of interests is 0.

3 The Model of Interests Between Co-author and Topics

The measurement of the author's interest value is often adopted among the literature recommendation. And this simple method can not describe that the author may generate new interests. We believe that a researcher often extended his personal interest to a new topic. A researcher can generate new interests due to the influence of his/her friends, teachers and others. Using the information related to one's social relationships can estimate the new interests.

The relationship of network cooperation is the important part of the social relation network, which has an important significance during the process of the scientific research [19, 20, 21]. The model of author-topic interest is extended to the model of co-author and topics to carry out the recommendation. This paper puts forward the degree of interest of author-topic and finds the topics of co-authors to build the model of interests of co-author-topic.

There are some documents which have more than one authors in one document, the co-author is defined as follows:

Definition 6 (Co-author). *Given a document repository D,* $Author_i$ *and* $Author_j$ *are co-author in a D, as Coauthor (D,* $author_i$*,* $author_j$*, d), if and only if d = <DOI, Authors, Topics> ∈D and (Authors = <…,* $author_i$*, …,* $author_j$*, …> or Authors = <…,* $author_j$*, …,* $author_i$*, …>)*

The co-author refers to the co-author that is described in a specific document. In the same way, there is co-author when not being special documents.

Definition 7 (Co-author without specific document). $Author_i$ *and* $Author_j$ *have the same interests as Coauthor (D,* $author_i$*,* $author_j$*) if and only if* $\exists\, d \in D$ *such that, Coauthor(D,* $author_i$*,* $author_j$*, d).*

The method above defines the common interest of two authors only, which is a partial order relationship. In order to describe the similar degree of the co-author in the same topic, we introduced the co-author's distance, which is defined as follows:

Definition 8 (Authorship distance). *Authorship distance is a partial mapping AD from* Author Set × Author Set → Real Number *which is defined as:*

AD(a, a) = 0;
AD(a, b) = 1; if coauthor(D, a, b)
AD(a, b) = n; if exists an author c such that AD(a, c) = n − 1 and co-author(c, b) and there exists no other c′ such that AD(a, c′) < n − 1.

There is the nearest distance (distance of 0) between each author and his/her interest. The distance is 1 if two authors were collaborators. Then the distance is 2 if there are other authors, and so on. Above authorship-distance is considered without the number of co-author. If considering the number of co-authors, here is another definition:

Definition 9 (Authorship distance with co-authored number). *Authorship distance is a partial mapping ADN from* Author Set × Author Set → Real Number *which is defined as:*

ADN(a, a) = 0;

$ADN(a,\ b) = \frac{1}{|\{\mathrm{d}|\mathrm{coauthor}(\mathrm{D,a,b,d})\}|}$*; if coauthor (D, a, b)*

$ADN(a,\ b) = \mathrm{d}_1 + \mathrm{d}_2$*; if exists an author c such that ADN (a, c) =* d_1 *and ADN (c, b) =* d_2 *and there exists no other c′ such that ADN (a, c′) <* d_1 *and ADN (c′, b) <* d_2*.*

Axiom 1. *Authorship distance AD is a metric distance, which owns follow characters:*

(1) *Nonnegativeness:* $|AD(a,\ b)| > \ = 0$
(2) *Identity:* $AD(a, b) = 0$ *if and only if* $a = b$
(3) *Symmetry:* $AD(a, b) = AD(b, a)$
(4) *The triangle inequality:* $AD(a, b) + AD(b, c) > \ = AD(a, c)$

Axiom 2. *Authorship distance with co-authored number ADN is a metric distance, which owns follow characters:*

(1) *Nonnegativeness:* $|ADN(a,\ b)| > \ = 0$
(2) *Identity:* $ADN(a, b) = 0$ *if and only if* $a = b$
(3) *Symmetry:* $ADN(a, b) = ADN(b, a)$
(4) *The triangle inequality:* $ADN(a, b) + ADN(b, c) > \ = ADN(a, c)$

4 The Model of Interests Between Author and Co-author

It will makes recommendation by combining the author's interested topics and co-author's interested topics and it is worthy to recommend if there is a close distance between a topic and its own network. Based on this idea, this corresponding recommended formula is defined as follows:

$$recommendationValue(a, topic, p) = Interest(author, topic), \quad \text{if p = my own interest only} \tag{2}$$

$$recommendationValue(a, topic, p) = \sum_{b=1,2,\ldots,k} \frac{Interest(b, topic)}{ADN(a, b)}, \quad \text{if p = my author network,}$$

k is the number of authors

(3)

The way of single measure can not describe the author's interests. We can combine the topic's interest and co-author's interest to make recommendation. So, the recommended formula is defined as follows:

$$\mathrm{Re}commendationValue(a, topic, p) = k_1 \cdot Interest(author, topic) + k_2 \cdot \sum_{b=1,2,\ldots,k} \frac{Interest(b, topic)}{ADN(a, b)} \quad (4)$$

Here, k is the number of authors. k_1, k_2 is the weight of measurement respectively. Normally $k_1 = k_2 = 0.5$. We do not need to consider the influence of other authors because it requires a lot of overhead in computation. Meanwhile, we need to consider a threshold of an author's distance instead of the author of more distance. The author's influence can be ignored if it is more than the threshold. If the distance of co-author is less than the threshold, the recommended formula with threshold of t is defined as follows:

$$recommendationValue(a, topic, p) = Interest(author, topic), \quad \text{if p} = \text{ my own interest only} \quad (5)$$

$$recommendationValue(a, topic, p) = \sum_{b=1,2,\ldots,k} \frac{Interest(b, topic)}{ADN(a, b)}, \quad \text{if p} = \text{my author network and AND(a,b)} < t \quad (6)$$

$$\mathrm{Re}commendationValue(a, topic, p) = k_1 \cdot Interest(author, topic) + k_2 \cdot \sum_{b=1,2,\ldots,k} \frac{Interest(b, topic)}{ADN(a, b)}$$
$$(\text{k is the number of authors, } k1 = k2 = 0.5 \text{ and } ADN(a,\ b) < t) \quad (7)$$

5 Experiment and Discussion

For this experiment, the authors' interests of the topics and the co-author's interests of the same topic were represented respectively by using a large number of documents. At the same time, we also needed to calculate the distance of the co-authors to measure the interest value of the recommendation. In the field of brain science, for example, Dr. Liang wants to query the literature or researches about inductive reasoning of human cognitive function. There will be many results of literatures or resources as shown in Table 1 when he queried something in the knowledge service platform of brain science. However, the recommendation of literatures based on the interests of co-authors can improve the search efficiency. The system automatically puts the other two kinds of query conditions as the interests of co-authors when Dr. Liang chooses cognitive function as a query condition. The interest recommendation of topic is shown as Fig. 1. Through some topics of interests, we calculated the value of some topics recommendation. For example, fMRI among the fMRI, ERP, Eye-movement, PET, Behavior has the biggest value of interest recommendation. Healthy college-student among the healthy college-student, healthy young adults, healthy older people, healthy middle-aged, patients have the biggest value of interest recommendation. Inductive reasoning among inductive reasoning, problem-solving, visual research, discovery learning, computation has the biggest value of interest recommendation. In the brain science data system, the redefined query strategy algorithm of using co-author's interest is shown in Table 2 and we got the query results as shown in Table 3.

Table 1. The results of simple query

ID	Title	Cognitive function	Experimental-type	Subjects type
1	ERP characteristics of sentential inductive reasoning in time and frequency domains	Inductive Reasoning	ERP	Normal-Subject
2	An fMRI study of the numerical stroop task in individuals with and without minimal cognitive impairment	Inductive Reasoning	fMRI	Patient-Subject
…	…	…	…	…
35	The Role of Category Label in Adults' Inductive Reasoning	Inductive Reasoning	fMRI	Normal-Subject

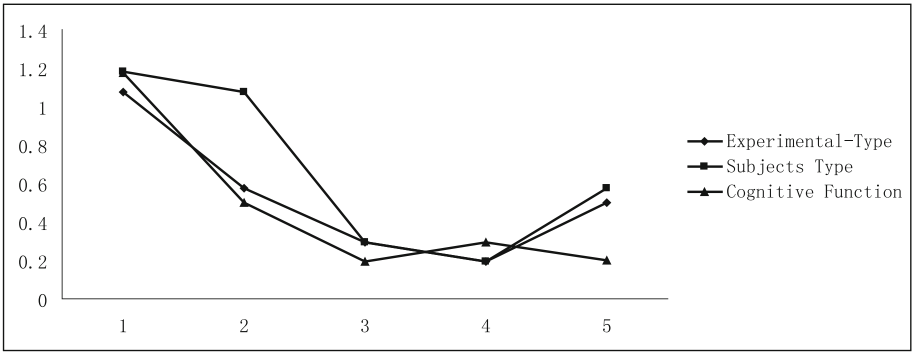

Fig. 1. The recommendation value of interests (the horizontal ordinate indicates the topic of interest and the vertical ordinate indicates the recommended value of interest)

According to the results, it is easy to see that the number of results is 10 when the query is without co-author's interests, however, when the query is defined with co-author's interests, the number of results is 35. In view of the number of results, the co-author recommendation greatly shortens the document filtering process for researchers. And the conclusion of experts' evaluation is that when the query isn't considered with co-author's interests, the accuracy rating is 50.8%; but when the query is refined using co-author's interests, the accuracy rating is 80%. This recommendation method improves the accuracy rating significantly and makes researchers find more suitable literatures and resources.

Table 2. The algorithm of research recommendation

Algorithm of Research Recommendation

Input: username

Output: literatureResults

```
literatures = getLiterature(username)
For each topic
For each interest of topic and Co-author
interests = getInterest(Topics) or [getInterest(Topics)/AND(author and   Co-author)]
RecommendationValue = getValue(interests)
End For
Initialize maxInterestValue(j) = InterestValue(1)
For each interest of topic
If(InterestValue(i)> InterestValue(i-1)) then
maxInterestValue(j) = InterestValue(i)
End If
End For
End For
literatureResults = getResult(username, maxInterestValue(1),
maxInterestValue(2), maxInterestValue(3))
return literatureResults
```

Table 3. The recommendation results of co-author

ID	Title	Cognitive function	Experimental-type	Subjects type
1	the fMRI research: the inductive reasoning of figure	Inductive Reasoning	fMRI	Normal-Subject
2	Dynamics of frontal, striatal, and hippocampal systems during rule learning	Inductive Reasoning	fMRI	Normal-Subject
…	…	…	…	…
10	The Role of Category Label in Adults' Inductive Reasoning	Inductive Reasoning	fMRI	Normal-Subject

6 The Knowledge Service System of Brain Science Based on Semantic Technology

The knowledge service system of brain science based on semantic technology mainly contains three level called web server, business logic and data process. The architecture of our system is depicted as Fig. 2. In the system, users use the web interface to post operation requirements to the server. The server sends the SPARQL queries to the SPARQL end point, which is launched by the workflow on the LarKC platform. At the same time, the system also permits users to write their own SPARQL queries and submit them by a submitting interface to the server. Meanwhile, the system can carry out the query, reasoning and so on. And a interface of this system is shown as Fig. 3.

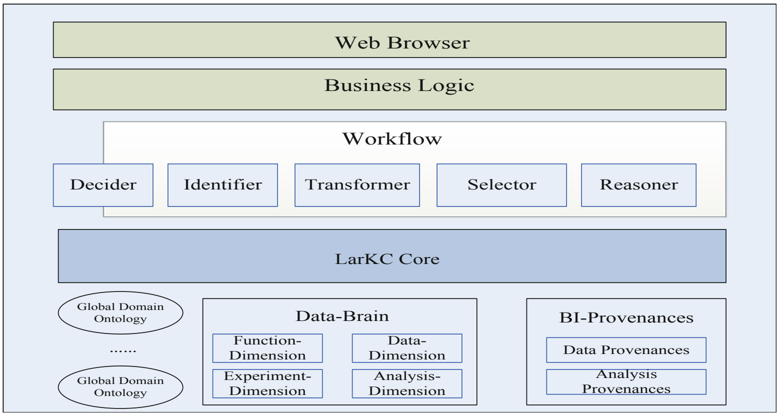

Fig. 2. System architecture

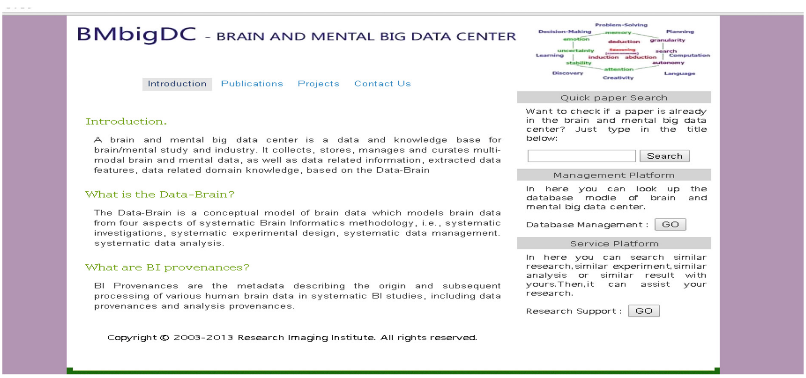

Fig. 3. System interface

Meanwhile, researchers need to query similar studies from internet to understand the development of research in the knowledge service system of brain science. However, the results may be millions among the studies of knowledge service of brain science. It is very difficult to find wanted literature from these results. Therefore, the literature recommendation has an important significance during the system of knowledge service system of brain science. We will put brain science provenances into this platform and conduct the class of data-brain as interests to make recommendations to improve the accuracy for finding similar researches.

7 Conclusions and Future Work

This paper mainly focused on the recommendations for the interesting topics, and put the author's research interest and the co-author's research interest together to research the recommendations. We had been able to quantify the author's interest in a certain topic, which takes a co-author's interest on the same topic and the distance between the co-authors into account. Based on these recommendations in the potential interests, the recommendation algorithm's efficiency and accuracy were improved greatly by combining the co-author's interest on the same topic.

However, we did not take the degree of the author's interest and the co-author's interest changes into consideration in the paper. And in the further study, we will consider to add the author's interest changes into the calculation of interest. It can recommend the interesting topics for researchers in brain science research efficiently.

Acknowledgements. The work is supported by the the JKF program of People's Public Security University of China (2019JKF334), and the National Key Research and Development Plan (2016YFC0801003).

References

1. Lane, R.D., Ryan, L.: Memory reconsolidation, emotional arousal, and the process of change in psychotherapy: New insights from brain science. Behav. Brain Sci. **38**, 1–64 (2015)
2. Ryan, P.B., Bridge, D.: Collaborative recommending using formal concept analysis. Knowl. Based Syst. **19**(5), 309–315 (2006)
3. Sarwa, B.S., Karypis, G., Konstan, J.: Item-based collaborative filtering recommendation algorithms. In: Proceedings of the 10th International Conference on World Wide Web, pp. 285–295. ACM, New York (2001)
4. Ma, H., Zhou, D., Liu, C., et al.: Recommender systems with social regularization. In: Proceedings of the 4th ACM International Conference on Web Search and Data Mining, Hong Kong, China, pp. 287–296 (2011)
5. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Trans. Knowl. Data Eng. **17**(6), 734–749 (2005)
6. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. Sci. Am. **284**(5), 34–43 (2001)

7. Fensel, D., van Harmelen, F.: Unifying reasoning and search to web scale. IEEE Internet Comput. **11**(2), 94–95 (2007)
8. Dan, B., Guha, R.V, Brian, M.: RDF Vocabulary Description Language 1.0: RDF Schema, W3C Recommendation, 10 February, 2004
9. Hao, C., Yubo, J., Chengwei, H.: Research of collaborative filtering recommendation based on user trust model. Comput. Eng. Appl. **46**(35), 148–151 (2010)
10. Guo, L., Ma, J., Chen, Z., Jiang, H.: Incorporating item relations for social recommendation. Chin. J. Comput. **37**(1), 219–228 (2014)
11. Chen, J., Zhang, H., He, X., Nie, L., Liu, W., Chua, T.-S.: Attentive collaborative filtering: multimedia recommendation with item- and component-level attention. In: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 335–344. ACM, New York (2017)
12. Jiang, S., Qian, X., Shen, J., Yun, F., Mei, T.: Author topic model-based collaborative filtering for personalized POI recommendations. IEEE Trans. Multimedia **17**(6), 907–918 (2015)
13. Efthymiou, K., Sipsas, K., Mourtzis, D.: On knowledge reuse for manufacturing systems design and planning: a semantic technology approach. CIRP J. Manuf. Sci. Technol. **8**, 1–11 (2014)
14. Zeng, Y., Zhong, N., Wang, Y., Qin, Y.L., Huang, Z.S., Zhou, H.Y.: User-centric query refinement and processing using granularity based strategies. Knowl. Inf. Syst. **27**(3), 419–450 (2010)
15. Zeng, Y., Zhou, E.Z., Qin, Y.L. Zhong, N.: Research interests: their dynamics, structures and applications in web search refinement. In: Proceeding of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence, pp. 639–646. IEEE Computer Society, Washington, DC, USA (2010)
16. Zhang, J., Tao, X., Wang, H.: Outlier detection from large distributed databases. World Wide Web **17**(4), 539–568 (2014)
17. Li, H., Wang, Y., Wang, H., Zhou, B.: Multi-window based ensemble learning for classification of imbalanced streaming data. World Wide Web **20**(6), 1507–1525 (2017)
18. Khalil, F., Wang, H., Li, J.: Integrating Markov model with clustering for predicting web page accesses. In: Proceeding of the 13th Australasian World Wide Web Conference (AusWeb 2007), pp. 63–74 (2007)
19. Khalil, F., Li, J., Wang, H.: An integrated model for next page access prediction. Int. J. Knowl. Web Intell. 1(1), 48–80 (2009)
20. Ma, J., Sun, L., Wang, H., Zhang, Y., Aickelin, U.: Supervised anomaly detection in uncertain pseudoperiodic data streams. ACM Trans. Internet Technol. (TOIT) **16**(1), 4–15 (2016)
21. Peng, M., Zeng, G., Sun, Z., Huang, J., Wang, H., Tian, G.: Personalized app recommendation based on app permissions. World Wide Web **21**(1), 89–104 (2018)