Expert Systems With Applications 144 (2020) 113088

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A hybrid recommendation system for Q&A documents

Ming Li*, Ying Li, Wangqin Lou, Lisheng Chen

School of Economics and Management, China University of Petroleum-Beijing, Beijing, 102249, China

ARTICLE INFO

Article history:
Received 6 August 2019
Revised 10 October 2019
Accepted 12 November 2019
Available online 13 November 2019

Keywords:
Q&A documents
Recommendation systems
Knowledge management

ABSTRACT

Question and answer (Q&A) documents are a new type of knowledge document composed of a question part and an answer part. The questions represent knowledge needs, and the answers contain the knowledge that meets these knowledge needs. An overload of accumulated Q&A documents decreases the reuse of valuable knowledge. In this paper, we propose a novel hybrid system to recommend Q&A documents to alleviate overload. First, knowledge needs are partitioned, and current knowledge needs are identified by sequentially clustering the Q&A documents. Second, a content-based (CB) recommendation method, a collaborative filtering (CF) recommendation method and a complementarity-based recommendation method are used to find the Q&A documents that are potentially helpful for the user. Third, the three initial recommendation lists of Q&A documents derived from the three recommendation methods are combined to form a more comprehensive recommendation based on the Fermat point. Because reading all Q&A documents in the recommendation list consumes an enormous amount of time and users prefer to read Q&A documents one by one starting from the top, a novel ranking mechanism is proposed to ensure that users obtain comprehensive knowledge to the greatest extent possible from the limited number of Q&A documents at the top of the list. The proposed approach is evaluated and compared based on an experimental dataset. Our experimental results show that the approach is feasible, performs well, and provides a more effective way to recommend Q&A documents.

1. Introduction

Knowledge is an important resource (Chen, Chen & Wu, 2012). It not only enhances the individual capability for problem solving but also improves the competitiveness of enterprises (Blondet, Duigou & Boudaoud, 2019; Wang & Wang, 2012). Knowledge management refers to the process of collecting, storage, retrieval, disseminating, and using knowledge resources (Lee, Foo, Leong & Ooi, 2016). It enables knowledge to reach the required people at the right time. People use knowledge to enhance their capability. Knowledge management systems (KMS) are IT-enabled systems that support knowledge management (H. Wang & Meng, 2019). Specifically, the processes of knowledge management, such as collecting, storage, disseminating and application, are enhanced with KMS (Centobelli, Cerchione & Esposito, 2018). Knowledge recommendation is a core part of KMS (Karna, Supriana & Maulidevi, 2017). It provides knowledge to users based on their knowledge needs. Because knowledge recommendation promotes the sharing of knowledge, it has attracted the attention of many researchers, and great progress has been made. However, previous studies have focused on formal documents such as papers (Zhao, Wu & Liu, 2016), documents (Lai, Liu & Lin, 2013) and learning materials (Salehi & Nakhai Kamalabadi, 2013).

A question and answer (Q&A) document is a new type of knowledge document that includes not only knowledge content but also knowledge needs (Rodrigo, Herrera & Peñas, 2019; Wen, Tu, Cheng, Xie & Yin, 2019). The knowledge content resides in the answers, and the knowledge needs are represented by the question. Q&A documents are attracting increasing attention from both Internet users and specialists in organizational knowledge management (Ji, Xu, Wang & He, 2012, 2019; Vuori & Okkonen, 2012). For example, Q&A websites such as Yahoo Answers[1] and Zhihu[2] are platforms on which knowledge is shared by posting and answering questions. A large number of Q&A documents have accumulated with the rapid increase in the number of questions and answers posted over time (Palomera & Figueroa, 2017; Zhang, Liu et al., 2014). These Q&A documents are helpful not only for community members but also for organizations. This accumulation leads to the problem of information overload of the archived Q&A documents, which, in turn, decreases the reuse of the valuable knowledge held by these documents. Thus, the issue arises of how to use these Q&A documents efficiently.

* Corresponding author.
E-mail address: limingzyq@cup.edu.cn (M. Li).

1 https://answers.yahoo.com/.
2 https://www.zhihu.com/.

https://doi.org/10.1016/j.eswa.2019.113088
0957-4174/

Considerable effort has been made to overcome the problem of archived Q&A document overload. Question search is the most commonly used method. Questions are retrieved based on the keywords input as a query (Zhang, Wu, Wu, Li & Zhou, 2014); then, the user chooses the questions from the returned list of questions. The major challenge of question search is the lexical gap between the questions and the keywords in the query (Wu et al., 2014). Various advanced models have been applied to prevent this lexical gap (Chen, Zhang, Zhao, Yao & Cai, 2018; Figueroa, 2017; Figueroa & Neumann, 2016; Muthmann & Petrova, 2014). The performance of the question search is closely related to the keywords. It is challenging for users to input accurate keywords (Mao, Hao, Wang & Huang, 2018); the keywords need to represent the user's knowledge needs fully and accurately. The keywords must also be consistent with the words of the answered questions (Wang, Ming, Hu & Chua, 2010). Moreover, after the suggested questions are selected and browsed, new keywords must be input for the next search. This continuous repeated input significantly decreases efficiency.

In addition to the retrieval of documents via searching, recommending documents based on knowledge needs is another way to ease the overload problem (Liu, Lai & Chiu, 2011). Rated documents reflect the user's knowledge needs (Lai & Liu, 2009). Based on the identified knowledge needs, providing these documents actively and automatically will help the user find the required documents. No extra effort, such as inputting keywords, is necessary (Huang, Lu, Duan & Zhao, 2012; Zhao et al., 2016). Consequently, recommending Q&A documents is a better way to overcome the problem of Q&A document overload.

Rated documents, especially documents with higher rates, represent user needs. Thus, the other Q&A documents that are relevant to the rated documents are probably also helpful for the user. Therefore, the content-based recommendation method is used to recommend Q&A documents based on the similarity between Q&A documents. Other topics may be relevant to the knowledge needs of the current user, but they may not be found. The Q&A documents on these topics are probably rated by other users with similar knowledge needs. Because the current user has not found these documents, they will not be recommended to the user with content-based (CB) methods. Therefore, the collaborative filtering (CF) recommendation algorithm is used to recommend these Q&A documents based on the similar knowledge needs of other users. Q&A documents have the characteristic of fragmentation; that is, the answer to a question may provide only partial information. This is the main difference between Q&A documents and formal documents (Liu, Chen & Huang, 2014). Complementary Q&A documents are needed for complete information (Liu et al., 2014). Therefore, complementary Q&A documents should also be recommended to the user. The Q&A documents that are recommended by the three algorithms are not the same, but they are all potentially relevant to the knowledge needs of the user. To make a comprehensive recommendation, the three recommendation methods need to be combined.

The premise of effective recommendation by the three recommendation methods is accurate knowledge needs. The knowledge needs of users are always goal- or topic-oriented and change over time. The documents at a particular stage focus on particular topics. Therefore, documents related to the current knowledge needs will be more relevant. However, in most studies, changing knowledge needs are not considered, and knowledge needs are not differentiated. Confusing different types of knowledge needs will lead to the recommendation of irrelevant documents in content-based recommendations, collaborating filtering-based recommendations and complementary recommendations. Therefore, knowledge needs must be distinguished.

The outcome of document recommendation algorithms is a list of documents (Lai et al., 2013). Users do not know how many documents they need to read, and they must browse the full list to choose the required Q&A documents. This is likely to trigger information overload, especially when Q&A documents are recommended because of the enormous amount of these documents. Therefore, a new mechanism is needed to rank Q&A documents in a recommendation list. Regardless of how many Q&A documents are browsed, the topics of these documents are covered to the greatest extent possible.

To resolve these problems, a novel hybrid recommendation system for Q&A documents is proposed. First, considering changes in knowledge needs, a sequential clustering method is used to divide knowledge needs and identify current knowledge needs. Because the documents are rated in chronological order, changing the topic of the documents reflects changing knowledge needs. With the sequential clustering method, the rated documents are classified into stages. The rated documents within a stage share similar topics. The documents in the last stage indicate the current knowledge needs. Second, the three recommendation methods are used to make three initial recommendation lists based on the specific knowledge needs. Content-based recommendation should make recommendations according to the derived current knowledge needs. In complementary-based recommendation, only the documents that are complementary to documents at the current stage should be recommended. Considering the changes in knowledge needs, in the findings of similar users in collaborating filtering recommendations, the similarity between rated documents within the stage are used instead of the similarity between all the rated documents. Third, the three initial recommendation lists are combined because they are not the same. In the combined list, the rankings of the Q&A documents in the initial recommendation lists should be maintained to the greatest extent possible. That is, the combined recommendation list should be similar to the three initial lists, and the distance from the initial three recommendation lists to the combined list should be minimized. The Fermat point (Spain, 1996) is the point that has the smallest distances to the three vertices of a triangle. The three recommendation lists can be regarded as the three vertices of a triangle, and the Fermat point is the combined recommendation list that minimizes the distance from the initial three recommendation lists. Fourth, a novel ranking method is proposed to form the final recommendation list. The Q&A documents are re-ranked according to the coverage of content along with the original rankings. Because of information processing ability and limited time, only a few documents can be read. Users prefer to read Q&A documents one by one starting from the top. Therefore, a new ranking mechanism is proposed in which the top items in the list are those that can cover the main contents of the recommendation list to the greatest extent. Regardless of how many Q&A documents are read, these are the Q&A documents that cover the topics in the recommendation list to the greatest extent. An experimental evaluation is conducted to verify the proposed hybrid recommendation system. The proposed hybrid recommendation system performs better than the single recommendation methods and the hybrid recommendation based on conventional combination approaches of posterior weighting on the metrics of precision, recall and F1. In the statistical test, all the hypothesis tests are statistically significant. This indicates that the proposed method performs significantly better than the other methods.

The main contributions of this paper are as follows.

(1) The hybrid recommendation system for the Q&A documents is a novel proposed. Most studies on recommending documents to individual users focus on formal documents but neglect recommendations for Q&A documents.
(2) The division of knowledge needs and the identification of current knowledge needs are proposed. Sequence clustering

is used to divide knowledge needs for a more accurate recommendation.

(3) This is the first synthesis of content-based recommendation, collaborating filtering recommendation and complementary recommendation. In the hybrid recommendation, the Fermat point is used to combine the initial recommendation lists.

(4) The ranking mechanism proposed is novel in that it captures the most content of recommended Q&A documents with any number of Q&A documents at the top of the recommendation lists to alleviate information overload.

The remainder of this paper is organized as follows. Section 2 presents related works, including those on Q&A documents and recommendation systems. Our hybrid recommendation system for Q&A documents is described in Section 3. In Section 4, the experiments and evaluation results are illustrated. Conclusions and possible future research directions are discussed in Section 5.

2. Related works

2.1. Q&A documents

Documents are carriers of knowledge (Cerchione & Esposito, 2017). More precisely, in this study, knowledge refers to explicit knowledge. Knowledge is acquired by reading documents such as books, reports and manuals. Except for systematic learning, in most conditions, only part of the knowledge that is relevant to the user's needs is required. To obtain this appropriate part, full texts must be browsed to locate the required knowledge (Fujii & Atsushi, 2008; Wu, Hawking, Turpin & Scholer, 2012). Q&A documents are a new type of document that include a question part and an answer part (Liu et al., 2014; Xue, Jeon & Croft, 2008). The answers store knowledge; however, the knowledge in the answers is not systematic and is only part of formal knowledge that aims to directly resolve the question posed. Because Q&A documents share knowledge in a convenient manner, they are attractive for both Internet users and specialists in organizational knowledge management (Ong, Day & Hsu, 2009; Roy et al., 2018; Wen et al., 2019) Q&A websites that focus on knowledge sharing by posting questions and providing answers to these questions are flourishing. With the continued accumulation of Q&A documents, information overload may arise, decreasing the reuse of the valuable knowledge held by these documents. Many approaches have been proposed to alleviate information overload. Searching Q&A documents is used to help find the required Q&A documents (Ji et al., 2019.; Zhang, Wu et al., 2014; Zhou, Liu, Liu, Zeng & Zhao, 2013). Q&A documents are found by matching the Q&A documents and the input keywords, and the performance of the search is largely affected by the accuracy of the input keywords.

2.2. Recommendation systems

Recommendation systems have been proposed to overcome the problem of information overload (Isinkaye, Folajimi & Ojokoh, 2015). In a recommendation system, relevant information is actively provided to users according to their preferences. Recommendation systems have been widely used in many areas, such as to recommend news (Okura, Tagami, Ono & Tajima, 2017; Xiao, Ai, Hsu, Wang & Jiao, 2015), movies (Armentano, Schiaffino, Christensen & Boato, 2015; Subramaniyaswamy, Logesh, Chandrashekhar, Challa & Vijayakumar, 2017), blogs (Subramaniyaswamy & Pandian, 2014; Zhang, Zhang, Yen & Zhu, 2017), articles (Ohta, Hachiki & Takasu, 2011; Sugiyama & Kan, 2015) and products (Cao & Li, 2007; Juheng Zhang & Piramuthu, 2018). Different types of algorithms have been proposed to generate recommendations, such as content-based (CB) recommendations (Lops, de Gemmis & Semeraro, 2011), collaborative filtering (CF) recommendations (Schafer, Frankowski, Herlocker & Sen, 2007; Su & Khoshgoftaar, 2009) and hybrid recommendations (Burke, 2002, 2007). CB recommendations derive the preferences of users by analyzing the attributes of the rated or browsed items (Lops et al., 2011). They recommend new items based on the derived preferences. The major drawbacks of CB recommendations are that they offer little variety and do not provide serendipitous recommendations. CF recommendations are based on human judgments (ratings), which can address the shortcomings of CB recommendations to some extent (Su & Khoshgoftaar, 2009). CF recommendations identify the relationship between items or users by analyzing the similarity between ratings instead of analyzing the items. The CB recommendation and CF recommendation methods have also been combined as hybrid approaches to make recommendations (Burke, 2002). The CB and CF approaches are combined for the recommendation of TV programs (Barragáns-Martínez et al., 2010). However, this combination does not consider changes in the preferences of users. The two filtering approaches are also combined for the recommendation of news topics (Lu, Dou, Lian, Xie & Yang, 2015) and for the recommendation of sports news (Lenhart & Herzog, 2016). They emphasize timeliness because people are interested in the latest news. However, although current knowledge needs should be considered in knowledge recommendation, the previous knowledge needs may also be the current knowledge needs of others. Moreover, both TV programs and news differ from the form of Q&A documents, and users may still be confused by the number of items to be read in the recommendation lists.

3. Proposed hybrid recommendation system for Q&A documents

The proposed approach includes five main parts, as shown in Fig. 1. First, a Q&A document is modeled for further processing. Considering the varying lengths of texts, the biterm topic model (BTM) is used. Second, knowledge needs are partitioned, and the current knowledge needs are identified by sequential clustering. Third, a CB recommendation, CF recommendation and complementarity-based recommendation are proposed to find the three initial recommendation lists of Q&A documents. Finally, the three recommendation lists are combined, and the novel ranking mechanism is proposed.

3.1. Modeling Q&A documents

Q&A documents are in textual form, which cannot be directly calculated. Therefore, the textual form needs to be transformed into a numerical form for calculation. Because the lengths of questions and answers vary, the BTM (Cheng, Yan, Lan & Guo, 2014), which performs better in dealing with short and long texts, is used.

3.1.1. BTM modeling

In BTM, the biterm (i.e., unordered co-occurring words) is used instead of individual words. The entire corpus is modeled by a multinomial distribution θ over topics. The biterm u in the set of biterms $\boldsymbol{U}$ is derived by choosing a topic t_u from the topic distribution θ and obtaining the words in the biterm u from the topic-word distribution ϕ_{t_u}.

Suppose α and β are the Dirichlet priors. The generative process of BTM can be given as follows (Cheng et al., 2014).

Step 1. Sample a distribution over topics: $\boldsymbol{\theta} \sim Dirichlet(\alpha)$.

Step 2. For each topic t, sample a distribution over words: $\phi_t \sim Dirichelet(\beta)$.

Step 3. For each of the $\boldsymbol{U}$ biterms u, sample a topic, $t_u \sim Multinominal(\boldsymbol{\theta})$, and sample words, $w_{u1}, w_{u2} \sim Multinominal(\phi_{t_u})$.

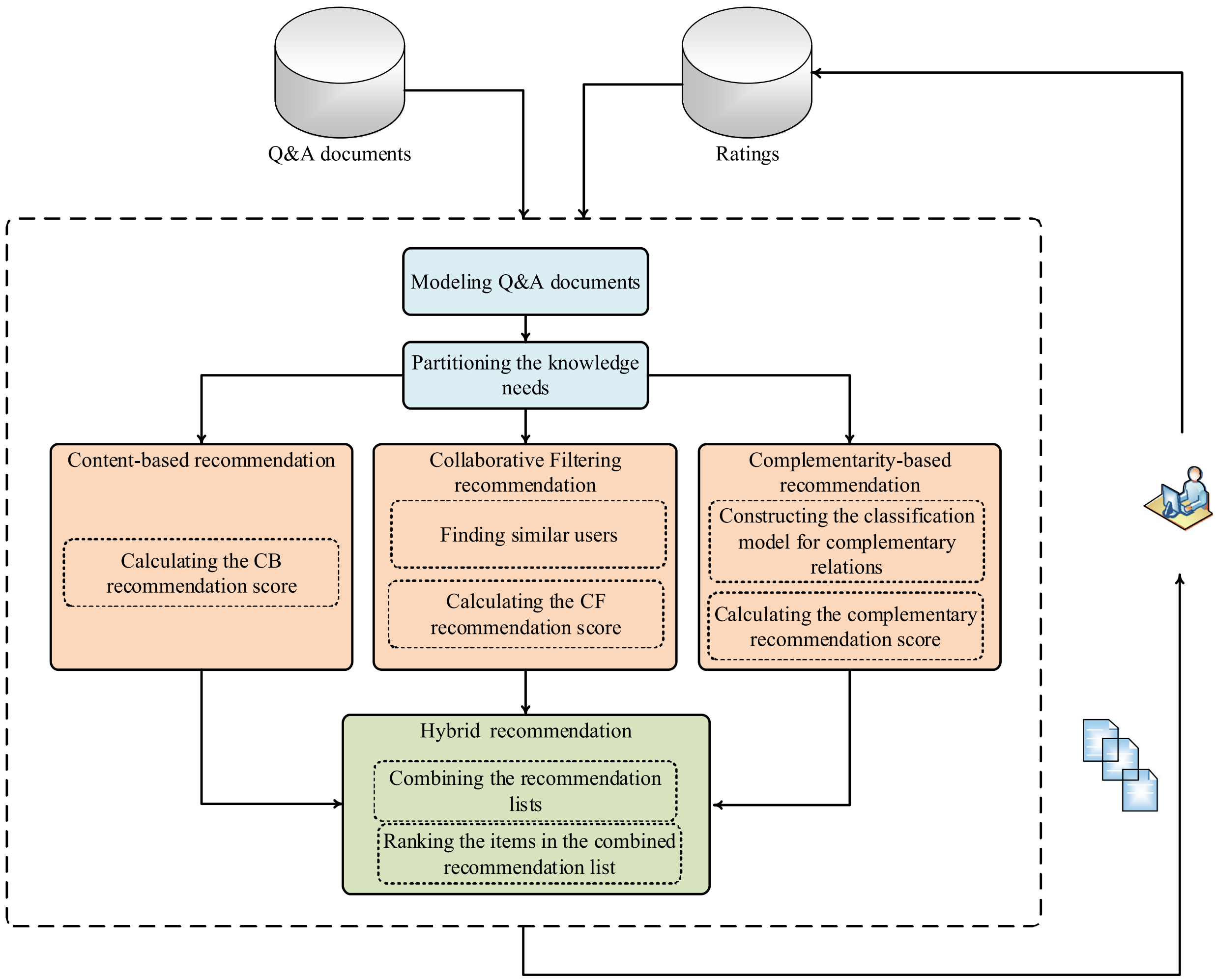

Fig. 1. Architecture of the proposed hybrid recommendation system.

Then, the likelihood of the entire corpus is

$$P(U) = \prod_{(u1,u2)} \sum_t P(t)P(w_{u1}|t)P(w_{u2}|t) = \prod_{(u1,u2)} \sum_t \theta_t \phi_{u1|t}\phi_{u2|t}. \tag{1}$$

The topic proportions of a document can be represented by the expectation of the topic proportions of biterms generated from the document:

$$P(t|d) = \sum_u P(t|u)P(u|d), \tag{2}$$

where $P(t|u)$ can be derived using Bayes' formula with the parameters estimated in BTM and $P(u|d)$ is the distribution of biterms in the document.

3.1.2. Modeling questions and answers

Question q_i can be represented by vector $\{t^q_{i1}, t^q_{i2}, \cdots, t^q_{ik}\}$, where k is the number of topics in the corpus of question set C_q. Topic proportion t^q_{il} can be inferred as follows:

$$\begin{aligned} t^q_{il} &= P\left(t^q_l|q_i\right) \\ &= \sum_{b\in C_q}\left(\frac{P\left(t^q_l\right)P\left(w^q_i|t^q_l\right)P\left(w^q_j|t^q_l\right)}{\sum_{l=1}^k P\left(t^q_l\right)P\left(w^q_i|t^q_l\right)P\left(w^q_j|t^q_l\right)} \times \frac{n_{q_i}(b)}{\sum_{b\in C_q} n_{q_i}(b)}\right) \\ &= \sum_{b\in C_q}\left(\frac{\theta_{t^q_l}\phi_{i|t^q_l}\phi_{j|t^q_l}}{\sum_{l=1}^k \theta_{t^q_l}\phi_{i|t^q_l}\phi_{j|t^q_l}} \times \frac{n_d(b)}{\sum_{b\in C_q} n_{q_i}(b)}\right) \end{aligned} \tag{3}$$

where $n_{q_i}(b)$ is the number of times of the occurrence of biterm $b = (w^q_i, w^q_j)$ in question q_i, and the parameters $\theta_{t^q_l}$, $\phi_{i|t^q_l}$ and $\phi_{j|t^q_l}$ can be inferred by Gibbs sampling (Cheng et al., 2014; Geman & Geman, 1993).

In the same way, the number of topics in the corpus of answers C_a is v. Thus, answers a_i can be modeled as $\{t^a_{i1}, t^a_{i2}, \cdots, t^a_{iv}\}$, and topics t^a_{il} can be inferred as follows:

$$\begin{aligned} t^a_{il} &= P\left(t^a_l|a_i\right) \\ &= \sum_{b\in C_a}\left(\frac{P\left(t^a_l\right)P\left(t^a_i|z^a_l\right)P\left(w^q_j|t^a_l\right)}{\sum_{l=1}^t P\left(t^a_l\right)P\left(w^a_i|t^a_l\right)P\left(w^a_j|t^a_l\right)} \times \frac{n_{a_i}(b)}{\sum_{b\in C_a} n_{a_i}(b)}\right) \\ &= \sum_{b\in C_a}\left(\frac{\theta_{t^a_l}\phi_{i|t^a_l}\phi_{j|t^a_l}}{\sum_{l=1}^t \theta_{t^a_l}\phi_{i|t^a_l}\phi_{j|t^a_l}} \times \frac{n_{a_i}(b)}{\sum_{b\in C_a} n_{a_i}(b)}\right) \end{aligned} \tag{4}$$

where $n_{a_i}(b)$ is the frequency of biterm $b = (w^a_i, w^a_j)$ in answers a_i, and the parameters $\theta_{t^a_l}$, $\phi_{i|t^a_l}$ and $\phi_{j|t^a_l}$ can also be inferred from Gibbs sampling (Cheng et al., 2014; Geman & Geman, 1993).

The similarity between two questions $q_i = \{t^q_{i1}, t^q_{i2}, \cdots, t^q_{ik}\}$ and $q_j = \{t^q_{j1}, t^q_{j2}, \cdots, t^q_{jk}\}$ is derived based on the Jensen-Shannon divergence (Ma, Zhang & Zhang, 2017) and is defined as follows:

$$\begin{aligned} \mathrm{sim}\left(q_j, q_i\right) &= 1 - JSD\left(q_j|q_i\right) \\ &= 1 - \frac{D_{KL}\left(q_j||\frac{q_j+q_i}{2}\right) + D_{KL}\left(q_i||\frac{q_j+q_i}{2}\right)}{2} \end{aligned}$$

$$= 1 - \frac{\sum_{l=1}^{k} z_{jl}^{q} \log\left(\frac{2z_{jl}^{q}}{z_{jl}^{q}+z_{il}^{q}}\right) + \sum_{l=1}^{k} z_{il}^{q} \log\left(\frac{2z_{il}^{q}}{z_{jl}^{q}+z_{il}^{q}}\right)}{2}. \quad (5)$$

Similarly, the similarity between two answers $a_i = \{t_{i1}^a, t_{i2}^a, \cdots, t_{iv}^a\}$ and $e_j = \{t_{j1}^a, t_{j2}^a, \cdots, t_{jv}^a\}$ is defined as follows (J. Ma et al., 2017):

$$\begin{aligned} \mathrm{sim}\left(a_j, a_i\right) &= 1 - JSD\left(a_j | a_i\right) \\ &= 1 - \frac{D_{KL}(a_j || \frac{a_j+a_i}{2}) + D_{KL}(a_i || \frac{a_j+a_i}{2})}{2} \\ &= 1 - \frac{\sum_{l=1}^{v} t_{jl}^{a} \log\left(\frac{2t_{jl}^{a}}{t_{jl}^{a}+t_{il}^{a}}\right) + \sum_{l=1}^{v} t_{il}^{a} \log\left(\frac{2t_{il}^{a}}{t_{jl}^{a}+t_{il}^{a}}\right)}{2}. \end{aligned} \quad (6)$$

The similarity between Q&A documents qa_i and qa_j can be derived as follows:

$$\mathrm{sim}\left(qa_j, qa_i\right) = \mathrm{sim}\left(a_j, a_i\right) \times \mathrm{sim}\left(q_j, q_i\right). \quad (7)$$

3.2. Partitioning of knowledge needs

Typically, knowledge needs are target oriented. Knowledge is often used to resolve a problem or perform a task. When a problem is resolved or a task is completed, new knowledge needs arise for the next tasks or problems; therefore, knowledge needs change over time. To distinguish knowledge needs and to identify current knowledge needs, sequential clustering (Fisher, 1958) is used to cluster rated Q&A documents. Sequential clustering is different from classical clustering methods, such as K-means clustering (Jain, 2010). In K-means clustering, only the similarity between documents is considered. However, in sequential clustering, documents are clustered according to similarity based on maintenance of the original rating order. The main idea of sequential clustering is to find dividing points and divide ordered documents into several segments, each of which is regarded as a class. The best dividing points should reduce the inner differences of each class and increase the differences between the classes (Fisher, 1958). The details of the steps involved in the partitioning of knowledge needs are given as follows (Fisher, 1958).

The set of ordered rated Q&A documents is $C = \{X_1, X_2, \ldots, X_i, \ldots, X_n\}$, where X_i is a Q&A document.

Step 1: Define the class diameter. Class $C_k = \{X_i, X_{i+1}, \ldots, X_j\}, C_k \subseteq C$, and $D(i, j)$ is the diameter of class C_k.

$$\begin{aligned} D\ (i,\ j) &= \sum_{t=i}^{j} \left(1 - sim\left(X_t, \bar{X}_k\right)\right) \\ &= \sum_{t=i}^{j} \left(1 - Sim\left(q_t, \overline{q^k}\right) \times Sim\left(a_t, \overline{a^k}\right)\right), \end{aligned} \quad (8)$$

where q_t and a_t are the question vector and the answer vector of Q&A document X_t, respectively, $\bar{X}_k$ is the mean vector of class C_k, and $\overline{q^k}$ and $\overline{a^k}$ are the mean vectors of questions and answers in class C_k, respectively.

$$\overline{q^k} = \frac{1}{j-i+1} \sum_{t=i}^{j} q_t \quad (9)$$

$$\overline{a^k} = \frac{1}{j-i+1} \sum_{t=i}^{j} a_t \quad (10)$$

Step 2: Define the loss function of the classification results.

Q&A documents are divided into k classes, which are represented by $\{X_{t_1}, X_{t_1+1}, \ldots, X_{t_2-1}\}$, $\{X_{t_2}, X_{t_2+1}, \ldots, X_{t_3-1}\}, \ldots,$ *and* $\{X_{t_k}, X_{t_k+1}, \ldots, X_n\}$ or, in shorthand, $\{t_1, t_1+1, \ldots, t_2-1\}$, $\{t_2, t_2+1, \ldots, t_3-1\}$, …, *and* $\{t_k, t_k+1, \ldots, n\}$. The sequence number of the first Q&A document of each class is the dividing point. The dividing points satisfying $1 = t_1 < t_2 < \ldots < t_k < n = t_{k+1} - 1$. Thus, the loss function of the classification results is defined as follows:

$$L[b(n,k)] = \sum_{i=1}^{k} D(t_i, t_{i+1} - 1). \quad (11)$$

When n and k are fixed, the lower the value of $L[b(n, k)]$ is, the smaller the sum of squares of deviations, and the more reasonable the classification. Therefore, it is necessary to find a classification method to minimize loss function $L[b(n, k)]$. $b(n, k)$ is a classification method to minimize the loss function.

Step 3: Recursive formula to find the optimal solution. According to the Fisher algorithm, the recursive formula of optimal segmentation is obtained as follows:

$$L[b(n,2)] = \min_{2 \le j \le n} \left\{D\left(t_1, t_{j-1}\right) + D\left(t_j, t_n\right)\right\} \quad (12)$$

$$L[b(n,k)] = \min_{k \le j \le n} \left\{L\left[b\left(t_{j-1}, t_{k-1}\right)\right] + D\left(t_j, t_n\right)\right\}. \quad (13)$$

According to the above two recursive formulas, we find the segmentation points from the back to the front to minimize $L[b(n, k)]$ and $L[b(t_{k-1}, k-1)]$. As a result, the optimal solution $b(n, k) = \{C_1, C_2, \ldots, C_k\}$ is obtained.

3.3. CB recommendation

The other documents that share the same topic as the documents previously preferred by the user are probably also of interest to the user. CB recommendation recommends Q&A documents that are similar to those previously preferred by the user based on the similarity between documents (Pazzani & Billsus, 2007). Because knowledge needs change over time, we focus on current knowledge needs. The last cluster that is derived in partitioning the knowledge needs is the current knowledge needs. The recommendation score is derived by summing the similarity between the new Q&A documents and each rated document along with the ratings.

The recommendation score of the new Q&A document qa_t for current knowledge needs k_{i*} of user u_i is derived as follows:

$$score_{CB}(qa_t, k_{i*}) = \sum_{qa_l \in k_{i*}} sim(qa_t, qa_l) \times r_l^{i*}, \quad (14)$$

where r_l^{i*} is the rating given by user u_i for Q&A document qa_l.

Then, the recommendation list of Q&A documents based on CB recommendation is derived. In this list, the recommended Q&A documents are ranked in descending order of the recommendation score.

3.4. CF recommendation

The Q&A documents that a similar user preferred are probably of interest to the target user. CF recommendation recommends Q&A documents that similar users preferred. The first step consists of finding similar users (Huang et al., 2012). Then, the recommendation scores of documents are derived by accumulating the ratings of similar users.

Similarity between users is determined by the current knowledge needs of the target user and the most similar knowledge needs of other users. Thus, the similarity between target user u_p and user u_m can be derived as follows:

Table 1
Determination of the complementary score based on a decision tree.

Input: training dataset, criterion, splitter, max_depth, min_samples_split, min_samples_leaf
Output: A decision tree
1: Create the initial root node associating the whole dataset.
2: Choose question similarity as the best attribute and split the dataset accordingly.
3: For each subset obtained after splitting, do as follows:
4: a) All instances in the class with the lower question similarity are deemed to be uncomplementary.
5: b) For instances in the class with higher question similarity, choose answer similarity as the second best attribute and split the dataset accordingly.
6: For each subset obtained after second splitting, do as follows:
7: a) All instances in the class with lower answer similarity are deemed to be complementary.
8: b) All instances in the class with higher answer similarity are deemed to be uncomplementary.
9: End For
10: End For

$$sim(u_p, u_m) = \max_n \sum_{qa_{mnl} \in k_{mn}} \sum_{qa_{p*j} \in k_{p*}} sim\left(qa_{p*j}, qa_{mnl}\right) \times \frac{1}{|k_{mn}| \times |k_{p*}|}, \tag{15}$$

where k_{mn} is the nth knowledge needs of user u_m and k_{p^*} is the current knowledge needs of target user u_p.

The recommendation score of document qa_t for target user u_p is obtained by accumulating the ratings of similar users as follows:

$$score_{CF}(qa_t, u_p) = \sum_{u_m \in U_p} sim(u_p, u_m) * r_l^{mn}, \tag{16}$$

where U_p is the set of users who are similar to u_p.

In the same way, the recommendation list based on CF recommendation is derived. In the list, the Q&A documents are also ranked in descending order of the recommendation score.

3.5. Complementarity-based recommendation

Because Q&A documents are characterized by fragmentation, the knowledge provided in a single Q&A pair may be partial and incomplete. Recommending complementary Q&A documents will ensure that users obtain more comprehensive knowledge. Q&A documents that have similar questions but dissimilar answers are deemed to be complementary (Liu, Chen, Shen & Lu, 2015), and a classification problem is posed. Decision tree classification is used to predict the complementary relationships between two Q&A documents based on question similarity and answer dissimilarity (Liu et al., 2015). First, the decision tree needs to be constructed, as shown in Table 1.

Based on the derived decision tree, the complementary Q&A documents for user u_p can be found. The set of Q&A documents for current knowledge needs is denoted R_p^*, and the set of complementary Q&A documents whose degree of complementary is not less than the threshold for each document qa_i^p in R_p^* is denoted C_p^i. Thus, the recommendation score of Q&A documents qa_j for user u_p is calculated as follows:

$$score_{COMP}\left(qa_j, u_p\right) = \sum_{qa_i^p \in R_p^*} comp\left(qa_j, qa_i^p\right) * r_i^{p*}, \tag{17}$$

where r_i^{p*} is the rating of document qa_i^p given by user u_p and $comp$ is the complement degree.

3.6. Hybrid recommendation

The three initial recommendation lists are derived by the CB recommendation, CF recommendation and complementarity-based recommendation. The CB recommendation increases the depth of knowledge by recommending similar Q&A documents. The CF recommendation expands the scope of knowledge by recommending Q&A documents that similar users preferred. The complementarity-based recommendation ensures that the Q&A documents are complete. To form an integrated list for comprehensive recommendation, the three lists need to be combined.

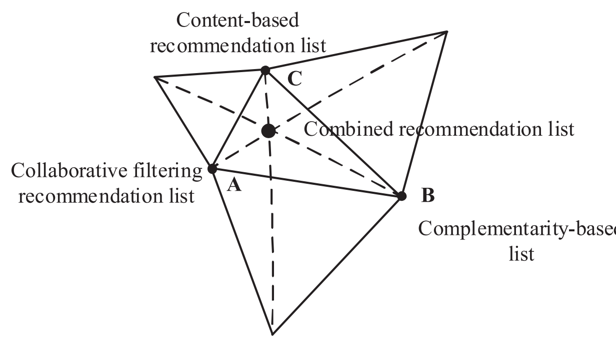

Fig. 2. Combining the recommendation lists based on the Fermat point.

(1) **Combination of the three initial recommendation lists**

In the combined list, the rankings of the Q&A documents in the initial recommendation lists should be maintained to the greatest extent possible. Therefore, the combined recommendation list should be similar to the three initial lists. In other words, the distance from the initial three recommendation lists should be minimized. The Fermat point (Spain, 1996) is the point that has the smallest distances to the three vertices of a triangle. The three recommendation lists can be regarded as three vertices of a triangle denoted by A, B and C, and the Fermat point is the combined recommendation list that minimizes the distance from the initial three recommendation lists, as shown in Fig. 2.

Because the recommendation scores are at different scales, the scores are normalized by max normalization. The three vertices of the triangle are represented by the normalized recommendation scores. The Euclidean distance is used to calculate the inner angle of a triangle. In the triangle, if the inner angle is not less than one hundred and twenty degrees, then the vertex of the obtuse angle is the Fermat point; otherwise, the calculation formula of the Fermat point is as follows:

$$\overrightarrow{OF} = \frac{\sin(A)}{\sin\left(A + \frac{\pi}{3}\right)} \overrightarrow{OA} + \frac{\sin(B)}{\sin\left(B + \frac{\pi}{3}\right)} \overrightarrow{OB} + \frac{\sin(C)}{\sin\left(C + \frac{\pi}{3}\right)} \overrightarrow{OC}, \tag{18}$$

where F is the Fermat point, O is the coordinate origin, and the value of each dimension of $\overrightarrow{OF}$ represents the combined recommendation score of the corresponding Q&A document.

(2) **Ranking Q&A documents in the combined recommendation list**

Users are often confused by the recommendation lists of documents because they do not know the number of documents they

Table 2
Ranking algorithm of Q&A documents.

Input: recommendation list L
Output: recommendation list L^* with the new ranking
1: For each $qa_i \in L$, do
//Find the Q&A document that has the highest accumulative similarities to the Q&A documents in L
2: $qa^* = \{qa_k | \max_{qa_k \in L} \sum_{qa_l \in L} (sim(qa_k, qa_l) * \omega^{\frac{\max r - r_l + 1}{\max r} * \frac{\max r - r_k + 1}{\max r}})\}$
3: add qa^* to L^*;
4: End For
5: count++;
6: When the count is lower than $|L|$, do
//Find the Q&A document that has the lowest accumulative similarities to the Q&A documents in L^*.
7: $qa' = \{qa_k | \max_{qa_k \in (L - L^*)} \sum_{qa_l \in L^*} (1 - sim(qa_k, qa_l)) * \omega^{\frac{\max r - r_k + 1}{\max r}}\}$
8: add qa' to L^*;
9: count++;
10: End While

need to read. Because of information processing ability and limited time, only a few documents can be read. Users prefer to read Q&A documents one by one starting from the top. Therefore, a new ranking mechanism is needed in which the top items in the list are those that can cover the main contents of the recommendation list to the greatest extent possible. Regardless of how many documents are read, they should be the documents that can most effectively cover the topics in the recommendation list to. The novel ranking algorithm is proposed in Table 2. The main idea of the algorithm is to identify the highest coverage document first and then continue to find the highest coverage document from the remaining documents (Zhang, Liu & Ren, 2016). In this process, the recommendation score in the original ranking is also considered. With the algorithm, the Q&A documents are ranked according to the coverage of contents integrated with the original ranking.

4. Experimental evaluations

In this section, we evaluate the performance of the proposed hybrid recommendation system. Because we use only the Q&A documents, without losing generality, the Q&A documents are collected from the famous Chinese CQA website Zhihu. The quality of the Q&A documents on the website is very high. The website attracts many users; as of the end of November 2018, the number of users had exceeded 220 million. We collected 2147 Q&A documents from the website for the experiment.

The prototype system was developed, and the collected Q&A documents were stored in the system. Users used the system to search for the required Q&A documents and provided feedback by rating the relevance from 1 to 10 after reading the Q&A documents. Thirty-two graduate and senior undergraduate students were invited to participate in the evaluation. They were registered in the system, and each user was assigned 4 to 6 topics. The assignments provided the scenario of their search to enable the users to understand the topics and to generate clear knowledge needs. Taking the topic "food safety" as an example, the scene was simulated as follows: recently, a series of food safety incidents have occurred that have seriously threatened public health. Users need knowledge in this area to understand food safety and to avoid being affected by food safety incidents.

The screen captures of the experimental results are shown in Fig. 3.

Fig. 3. Screen captures of the experimental results.

First, the decision tree used to identify the complementary relationship was constructed. Hierarchical 5-fold cross validation was used to test the performance of the decision tree. The results are shown in Table 3. From the results, we can see that the values of the three metrics are high; thus, the performance of the decision tree is good. Moreover, the values of the five areas are close to each other, which indicates that the performance of the decision tree is stable.

Table 3
The results of hierarchical 5-fold cross validation of the decision tree.

Evaluation metrics	fold-1	fold-2	fold-3	fold-4	fold-5	Average
Precision	0.839	0.768	0.771	0.815	0.773	0.793
Recall	0.886	0.885	0.906	0.893	0.930	0.900
F1	0.862	0.822	0.833	0.852	0.844	0.843

In predicting the complementary relationship using a decision tree, when the degree of the complementary relationship exceeds 0.7, the complementary relationships are more obvious. Therefore, the threshold δ^{comp} is set to 0.7. In the re-ranking of the recommendation lists, the Q&A documents ranked first in the original recommendation list should be selected as a higher priority; only when the ω is greater than 1 is it an incremental function. A higher value leads to a greater influence of the original rankings but reduces the weight of the similarity between Q&A documents. Through multiple experiments, we set the parameter $\omega = 1.1$ by balancing the weight of the similarity and the original rankings.

The evaluation includes three parts. The first part consists of evaluating the performance of the CB recommendation, the CF recommendation and the complementarity-based recommendation in conditions that take into account the partition of knowledge needs. The second part consists of evaluating the performance of the hybrid method compared with the single recommendation method and the hybrid method based on conventional combination approaches of posterior weighting. The third part consists of assessing the effect of the recommendation list with new rankings.

4.1. Evaluation metrics

With regard to the evaluation of recommendation systems, the precision, recall and F1 are used as evaluation metrics

Table 4
Comparison of the CB recommendation.

Top-K	CB recommendation	Precision	Recall	F1
5	With partitioning of knowledge needs	0.655	0.053	0.098
	Without partitioning of knowledge needs	0.139	0.010	0.019
10	With partitioning of knowledge needs	0.576	0.093	0.158
	Without partitioning of knowledge needs	0.145	0.022	0.037
15	With partitioning of knowledge needs	0.524	0.126	0.202
	Without partitioning of knowledge needs	0.137	0.030	0.049
20	With partitioning of knowledge needs	0.477	0.153	0.229
	Without partitioning of knowledge needs	0.130	0.038	0.058
25	With partitioning of knowledge needs	0.447	0.179	0.252
	Without partitioning of knowledge needs	0.126	0.046	0.067
30	With partitioning of knowledge needs	0.415	0.200	0.265
	Without partitioning of knowledge needs	0.122	0.054	0.073

(Tewari & Barman, 2018). In the experiments, these metrics were used to evaluate the CB recommendation, the CF recommendation, the complementarity-based recommendation, and the hybrid recommendation based on conventional combination approaches of a posteriori weighting and the proposed hybrid recommendation. Coverage is used to assess the effect of the re-ranking of the recommendation list (Ma, Wei & Chen, 2011).

Precision is determined by the fraction of recommended Q&A documents that are actually relevant to the user's current knowledge needs. Recall involves the fraction of relevant Q&A documents that are correctly identified by the recommendation method. Because the two metrics are mutually exclusive, precision and recall are combined into the F1 measure. Coverage reflects the coverage of the subset of recommended Q&A documents to the whole list. The four metrics can be defined as follows (Ma et al., 2011; Tewari & Barman, 2018):

$$\text{Precision} = \frac{\text{number of correctly recommended Q A documents}}{\text{number of recommended Q A documents}} \tag{19}$$

$$\text{Recall} = \frac{\text{number of correctly recommended Q A documents}}{\text{number of relevent Q A documents}} \tag{20}$$

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{21}$$

$$\text{Coverage}\left(L', L\right) = \frac{1}{|L|} \sum\nolimits_{qa_i \in L} \left(\max_{qa_j \in L'} \left(sim\left(qa_i, qa_j\right)\right)\right), \tag{22}$$

where L is the original recommendation list and L' is the subset at the top of list L.

4.2. Experimental results and discussion

4.2.1. Comparison of the three recommendation methods with and without the partitioning of knowledge needs

(1) Performance of the CB recommendation

In the experiment, Q&A documents are recommended by the CB recommendation algorithm both with and without the partitioning of knowledge needs. The experimental results are shown in Table 4. In the table, each row represents the first K Q&A documents in the recommended list, and the two sub-rows indicate taking and not taking the partitioning of knowledge needs into consideration. The columns include the metrics. Each cell of the table stores the corresponding experimental result. For example,

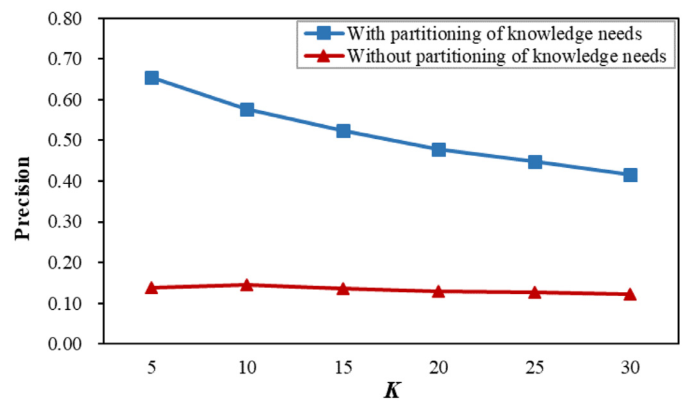

Fig. 4. Comparison of the CB recommendation based on precision.

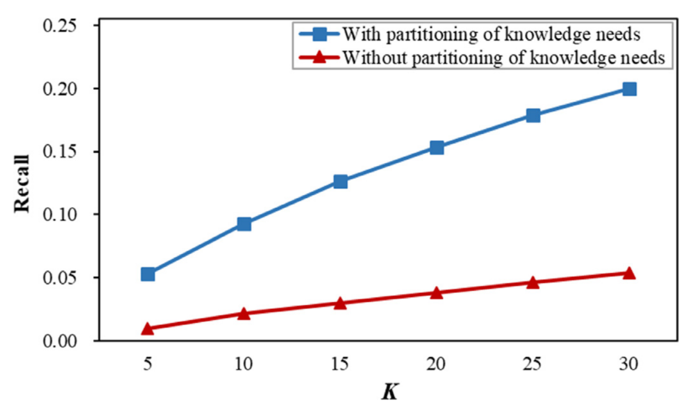

Fig. 5. Comparison of the CB recommendation based on recall.

0.655, which resides in the first sub-row of the second row and the third column, is the average value of the precision of the first 5 Q&A documents in the recommendation list derived by the CB recommendation with partitioning of knowledge needs. The results are also depicted in Figs. 4–6. The CB recommendation algorithm with the partitioning of knowledge needs performs better. Fig. 4 shows that the average precision of the CB recommendation is only 0.133. However, with the partitioning of knowledge needs, the average precision of the CB recommendation improves to 0.516. In particular, when K is smaller, the effect of integrating this partitioning is more obvious. This method is more practical because users prefer to read a limited number of Q&A documents at the top of the list. In Fig. 5, recall of the CB recommendation improves with the increase in the K value both with and without the partitioning of knowledge needs. With the partitioning of

Table 5
Statistical results of all the hypotheses of CB recommendation.

	Measure	Hypothesis	T value	P value
$N=5$	Precision	With partition >Without partition	8.18970	3.00E-09
	Recall	With partition >Without partition	6.79626	1.30E-07
	F1	With partition >Without partition	6.90459	9.64E-08
$N=10$	Precision	With partition >Without partition	7.34010	2.91E-08
	Recall	With partition >Without partition	6.06613	1.02E-06
	F1	With partition >Without partition	6.24157	6.18E-07
$N=15$	Precision	With partition >Without partition	7.56960	1.56E-08
	Recall	With partition >Without partition	6.36017	4.42E-07
	F1	With partition >Without partition	6.59790	2.27E-07
$N=20$	Precision	With partition >Without partition	7.67236	1.18E-08
	Recall	With partition >Without partition	6.53451	4.15E-09
	F1	With partition >Without partition	6.80539	3.83E-05
$N=25$	Precision	With partition >Without partition	7.40360	2.45E-08
	Recall	With partition >Without partition	6.46559	3.29E-07
	F1	With partition >Without partition	6.71851	1.62E-07
$N=30$	Precision	With partition >Without partition	7.52460	1.76E-08
	Recall	With partition >Without partition	6.56612	2.48E-07
	F1	With partition >Without partition	6.86912	1.06E-07

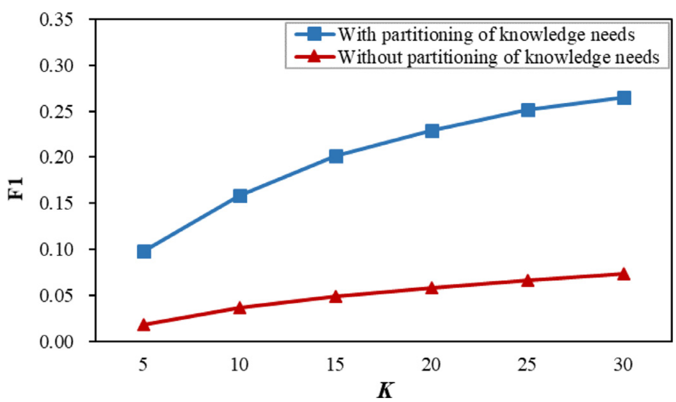

Fig. 6. Comparison of the CB recommendation based on F1.

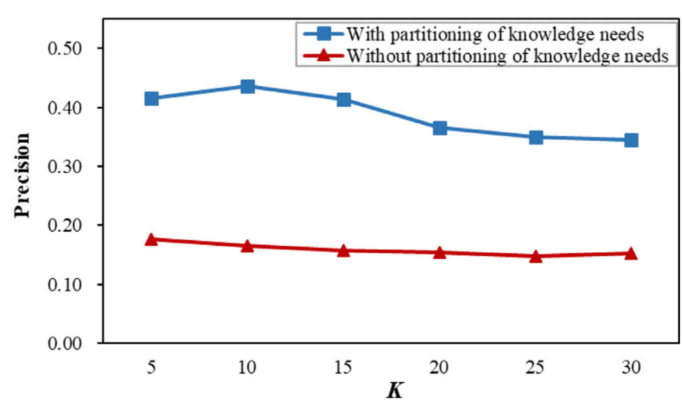

Fig. 7. Comparison of the CF recommendation based on precision.

knowledge needs, a more obvious improvement is derived. When the number of recommended Q&A documents is 30, an average of 20.0% of all relevant Q&A documents are recommended with the partitioning of knowledge needs. Without partitioning, the value of recall is only 5.4%. The F1 score combines precision and recall into a single measure. Fig. 6 shows that the F1 score of the CB recommendation with the partitioning of knowledge needs is significantly higher. In conclusion, the CB recommendation with the partitioning of knowledge needs is superior to the traditional algorithm.

The paired T-test is used to test the significance of differences between with and without partitioning of knowledge needs in CB recommendation. The statistical results of all the hypotheses are shown in Table 5. All the hypothesis tests are statistically significant. This indicates that CB recommendation with the partitioning of knowledge needs performs significantly better.

(2) Performance of the CF recommendation

In the same way, the CF recommendation is evaluated with and without the partitioning of knowledge needs. The experimental results are listed in Table 6. In the table, each row represents the first K Q&A documents in the recommended list, and the two sub-rows indicate taking and not taking the partitioning of knowledge needs into consideration. The columns include the metrics. Each cell of the table stores the corresponding experimental result. For example, 0.415, which resides in the first sub-row of the second row and the third column, is the average value of the precision of the first

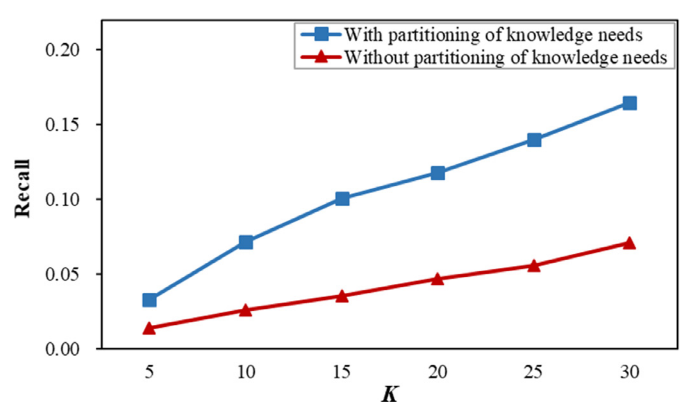

Fig. 8. Comparison of the CF recommendation based on recall.

5 Q&A documents in the recommendation list derived by the CF recommendation with partitioning of knowledge needs. The results are also depicted in Figs. 7–9. The partitioning of knowledge needs also improves the performance of the CF recommendation. The precision, recall and F1 values all improve. The average precision improves from 0.159 to 0.388, recall improves from 0.042 to 0.105, and the F1 value improves from 0.062 to 0.156. The improved effects of the partitioning of knowledge needs on the performance of the CF recommendation are obvious. With the improvement of K, precision is deceased as the ratio of irrelevant Q&A documents

Table 6
Comparison of the CF recommendation.

Top-*K*	CF recommendation	Precision	Recall	F1
5	With partitioning of knowledge needs	0.415	0.033	0.061
	Without partitioning of knowledge needs	0.176	0.014	0.025
10	With partitioning of knowledge needs	0.437	0.071	0.122
	Without partitioning of knowledge needs	0.166	0.026	0.044
15	With partitioning of knowledge needs	0.414	0.101	0.160
	Without partitioning of knowledge needs	0.157	0.036	0.057
20	With partitioning of knowledge needs	0.366	0.118	0.175
	Without partitioning of knowledge needs	0.155	0.047	0.071
25	With partitioning of knowledge needs	0.350	0.140	0.197
	Without partitioning of knowledge needs	0.148	0.056	0.080
30	With partitioning of knowledge needs	0.345	0.165	0.219
	Without partitioning of knowledge needs	0.153	0.071	0.095

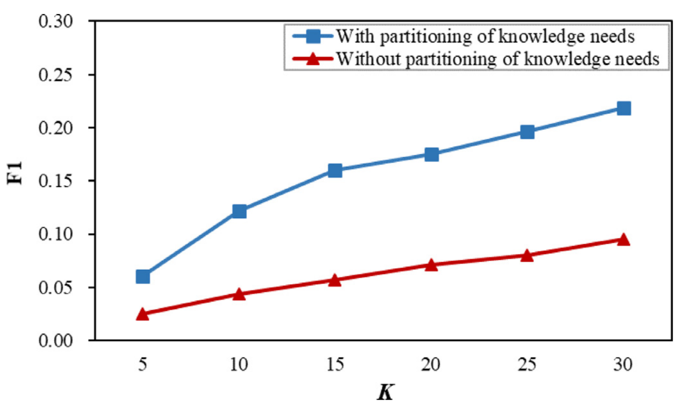

Fig. 9. Comparison of the CF recommendation based on F1.

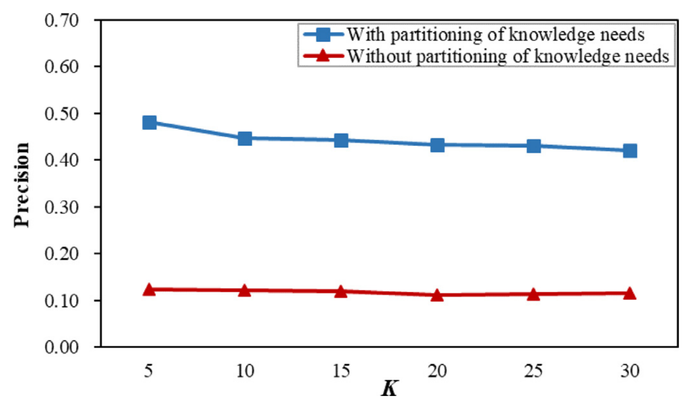

Fig. 10. Comparison of the complementarity-based Q&A recommendation based on precision.

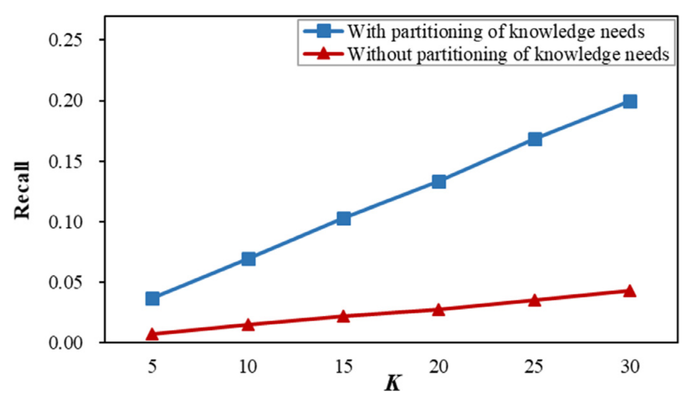

Fig. 11. Comparison of the complementarity-based Q&A recommendation based on recall.

improves. Recall is improved as more relevant Q&A documents are obtained.

According to the experimental results, the CF recommendation algorithm improves with the partitioning of knowledge needs. Regarding the most important effects, because the knowledge needs of users change over time, discriminating their knowledge needs will make the identified knowledge needs accurate. Moreover, by determining the similarity between users between stages instead of directly through the Q&A documents, sparse results will be avoided.

The paired T-test is used to test the significance of differences with and without partitioning of knowledge needs in CF recommendation. The statistical results of all the hypotheses are shown in Table 7. All the hypothesis tests are statistically significant. This indicates that CF recommendation with the partitioning of knowledge needs performs significantly better.

(3) Performance of the complementarity-based Q&A recommendation

In the same way, the complementarity-based Q&A recommendation is evaluated with and without the partitioning of knowledge needs. The experimental results are listed in Table 8. In the table, each row represents the first *K* Q&A documents in the recommended list, and the two sub-rows indicate taking and not taking the partitioning of knowledge needs into consideration. The columns include the metrics. Each cell of the table stores the corresponding experimental result. For example, 0.482, which resides in the first sub-row of the second row and the third column, is the average value of the precision of the first 5 Q&A documents in the recommendation list derived by the complementarity-based recommendation with partitioning of knowledge needs. The results are also depicted in Figs. 10–12. The improvement of the complementarity-based Q&A recommendation by partitioning the knowledge needs is especially obvious. The precision, recall and F1 values all improve significantly. Fig. 10 shows that the average precision without the partitioning of knowledge needs is 0.117. However, with the partitioning of knowledge needs, the value reaches 0.443. As shown in Fig. 11, the average recall of the complementarity-based Q&A recommendation improves from 0.025 to 0.118. Fig. 12 indicates that the average F1 value improves from 0.039 to 0.176. The recall and F1 values are both improved with the increase in *K*. In particular, when the partitioning of knowledge needs is used, the improving trend is more obvious. Therefore, the complementarity-based Q&A recommendation improves significantly with the partitioning of knowledge needs compared to the traditional algorithm.

Table 7
Statistical results of all the hypotheses of CF recommendation.

	Measure	Hypothesis	T value	P value
$N = 5$	Precision	With partition > Without partition	5.89073	1.68E-06
	Recall	With partition > Without partition	5.28379	9.52E-06
	F1	With partition > Without partition	5.33924	8.12E-06
$N = 10$	Precision	With partition > Without partition	6.46716	3.27E-07
	Recall	With partition > Without partition	5.71564	2.76E-06
	F1	With partition > Without partition	5.82581	2.02E-06
$N = 15$	Precision	With partition > Without partition	9.30327	1.75E-10
	Recall	With partition > Without partition	7.52819	1.75E-08
	F1	With partition > Without partition	7.86183	7.13E-09
$N = 20$	Precision	With partition > Without partition	8.26079	2.49E-09
	Recall	With partition > Without partition	7.03501	6.72E-08
	F1	With partition > Without partition	7.34566	2.87E-08
$N = 25$	Precision	With partition > Without partition	8.57399	1.10E-09
	Recall	With partition > Without partition	7.07992	5.94E-08
	F1	With partition > Without partition	7.48496	1.96E-08
$N = 30$	Precision	With partition > Without partition	9.36711	1.49E-10
	Recall	With partition > Without partition	7.45378	2.14E-08
	F1	With partition > Without partition	8.06953	4.11E-09

Table 8
Comparison of the complementarity-based Q&A recommendation.

Top-*K*	Complementarity-based Q&A recommendation	Precision	Recall	F1
5	With partitioning of knowledge needs	0.482	0.037	0.068
	Without partitioning of knowledge needs	0.124	0.008	0.014
10	With partitioning of knowledge needs	0.446	0.069	0.119
	Without partitioning of knowledge needs	0.122	0.015	0.027
15	With partitioning of knowledge needs	0.444	0.103	0.165
	Without partitioning of knowledge needs	0.119	0.022	0.036
20	With partitioning of knowledge needs	0.433	0.134	0.202
	Without partitioning of knowledge needs	0.112	0.028	0.044
25	With partitioning of knowledge needs	0.430	0.169	0.239
	Without partitioning of knowledge needs	0.113	0.035	0.053
30	With partitioning of knowledge needs	0.421	0.200	0.266
	Without partitioning of knowledge needs	0.115	0.043	0.061

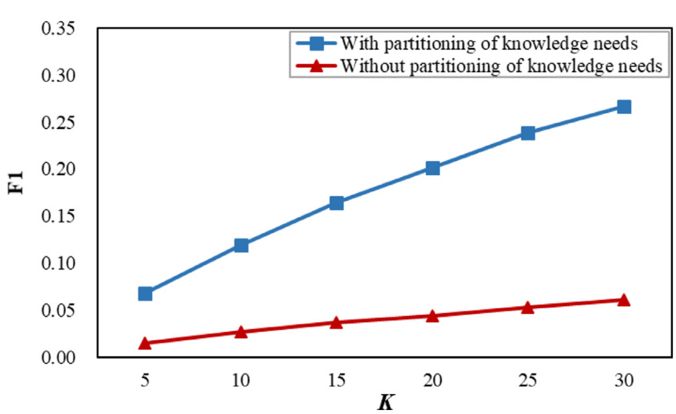

Fig. 12. Comparison of the complementarity-based Q&A recommendation based on F1.

The paired T-test is used to test the significance of differences with and without partitioning of knowledge needs in complementarity-based recommendation. The statistical results of all the hypotheses are shown in Table 9. All the hypothesis tests are statistically significant. This indicates that the complementarity-based recommendation with the partitioning of knowledge needs performs significantly better.

4.2.2. Performance of the hybrid recommendation

The CB recommendation, CF recommendation and complementarity-based Q&A recommendation all have unique advantages in different respects. Therefore, the three lists are combined to form a comprehensive recommendation. In the experiments, the performance of the three single recommendation methods, the proposed hybrid recommendation based on the Fermat point and the hybrid recommendation based on the conventional combination approach of a posteriori weighting are compared. The experimental results for precision, recall and F1 under different k values are obtained, as shown in Table 10. In the table, each row represents the first K Q&A documents in the recommended list used for the calculation of the experimental results, and the five sub-rows indicate the five recommendation methods. The columns include the metrics. Each cell of the table stores the corresponding experimental result. For example, 0.755, which resides in the fifth sub-row of the second row and the third column, is the average value of the precision of the first 5 Q&A documents in the recommendation list derived by the proposed hybrid recommendation.

The results are also illustrated in Figs. 13–15. The precision, recall and F1 values of the proposed hybrid recommendation are all higher than those of any other single recommendation and the hybrid recommendation based on the conventional combination approach of a posteriori weighting under all k values.

Table 9
Statistical results of all the hypotheses of complementarity-based recommendation.

	Measure	Hypothesis	T value	P value
$N = 5$	Precision	With partition >Without partition	6.11111	8.95E-07
	Recall	With partition >Without partition	5.57511	4.13E-06
	F1	With partition >Without partition	5.62698	3.56E-06
$N = 10$	Precision	With partition >Without partition	6.05371	1.05E-06
	Recall	With partition >Without partition	5.64808	3.35E-06
	F1	With partition >Without partition	5.73072	2.65E-06
$N = 15$	Precision	With partition >Without partition	7.11329	5.42E-08
	Recall	With partition >Without partition	6.26080	5.85E-07
	F1	With partition >Without partition	6.48765	3.09E-07
$N = 20$	Precision	With partition >Without partition	7.78782	8.69E-09
	Recall	With partition >Without partition	6.77170	1.39E-07
	F1	With partition >Without partition	7.03585	6.71E-08
$N = 25$	Precision	With partition >Without partition	7.51454	1.81E-08
	Recall	With partition >Without partition	6.50127	2.97E-07
	F1	With partition >Without partition	6.78902	1.33E-07
$N = 30$	Precision	With partition >Without partition	7.25424	3.68E-08
	Recall	With partition >Without partition	6.32554	4.88E-07
	F1	With partition >Without partition	6.61429	2.16E-07

Table 10
Comparison of the hybrid recommendation with the other recommendations.

Top-K	Recommendation algorithm	Precision	Recall	F1
5	CB recommendation	0.655	0.053	0.098
	CF recommendation	0.415	0.033	0.061
	Complementarity-based recommendation	0.482	0.037	0.068
	Hybrid recommendation based on posteriori weighting	0.529	0.041	0.077
	Proposed hybrid recommendation	0.755	0.062	0.114
10	CB recommendation	0.576	0.093	0.158
	CF recommendation	0.437	0.071	0.122
	Complementarity-based recommendation	0.446	0.069	0.119
	Hybrid recommendation based on posteriori weighting	0.507	0.078	0.135
	Proposed hybrid recommendation	0.652	0.104	0.179
15	CB recommendation	0.524	0.126	0.202
	CF recommendation	0.414	0.101	0.160
	Complementarity-based recommendation	0.444	0.103	0.165
	Hybrid recommendation based on posteriori weighting	0.496	0.115	0.185
	Proposed hybrid recommendation	0.643	0.154	0.246
20	CB recommendation	0.477	0.153	0.229
	CF recommendation	0.366	0.118	0.175
	Complementarity-based recommendation	0.433	0.134	0.202
	Hybrid recommendation based on posteriori weighting	0.484	0.149	0.227
	Proposed hybrid recommendation	0.620	0.199	0.297
25	CB recommendation	0.447	0.179	0.252
	CF recommendation	0.350	0.140	0.197
	Complementarity-based recommendation	0.430	0.169	0.239
	Hybrid recommendation based on posteriori weighting	0.465	0.179	0.257
	Proposed hybrid recommendation	0.598	0.238	0.336
30	CB recommendation	0.415	0.200	0.265
	CF recommendation	0.345	0.165	0.219
	Complementarity-based recommendation	0.421	0.200	0.266
	Hybrid recommendation based on posteriori weighting	0.481	0.221	0.300
	Proposed hybrid recommendation	0.577	0.273	0.364

As shown in Fig. 13, the precision of the proposed hybrid recommendation algorithm in the top 5 Q&A documents recommended is 75.5%; that is, users consider at least four Q&A documents to be helpful. With the increase in the k value, precision decreases. The main reason for this is that the algorithm ranks the relevant Q&A documents at the top of the list. Therefore, when the k value increases, precision decreases slightly. When the top 30 Q&A documents are recommended, the accuracy of the proposed hybrid recommendation is 57.7%, which is still higher than that of the other algorithms.

As shown in Fig. 14, recall of the proposed hybrid recommendation algorithm also increases with the increase in the k value. The average value of recall is 0.172, which is higher than that of the CF recommendation (0.105), the complementarity-based recommendation (0.118), the hybrid recommendation based on the conventional combination approach of a posteriori weighting (0.131), and the CB recommendation (0.134). Fig. 15 shows that the F1 value of the hybrid recommendation is higher than that of any of the single recommendation algorithms under different values of k.

To summarize, the proposed hybrid recommendation algorithm can identify Q&A documents that are particularly relevant to knowledge needs and provide more accurate recommendations.

The paired T-test is also used to test the significance of differences between the proposed hybrid recommendation method and the other recommendation methods. The statistical results of all the hypotheses on the three metrics, including precision, recall and

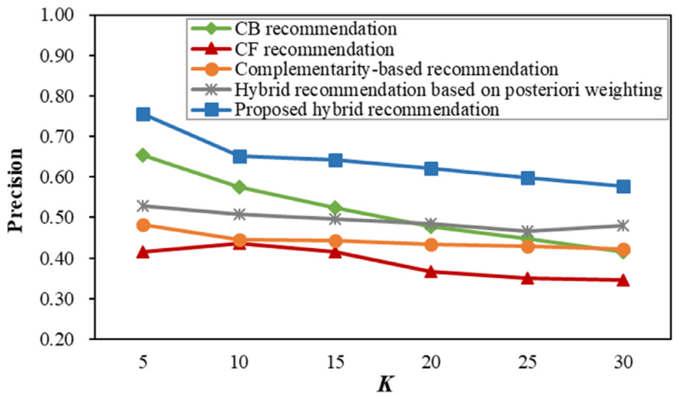

Fig. 13. Comparison of the hybrid recommendation with other recommendations based on precision.

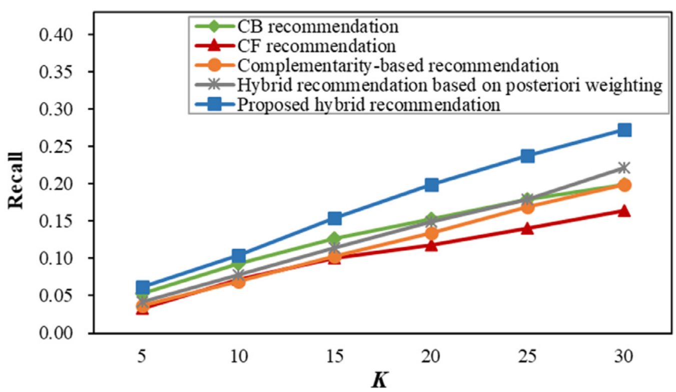

Fig. 14. Comparison of the hybrid recommendation with other recommendations based on recall.

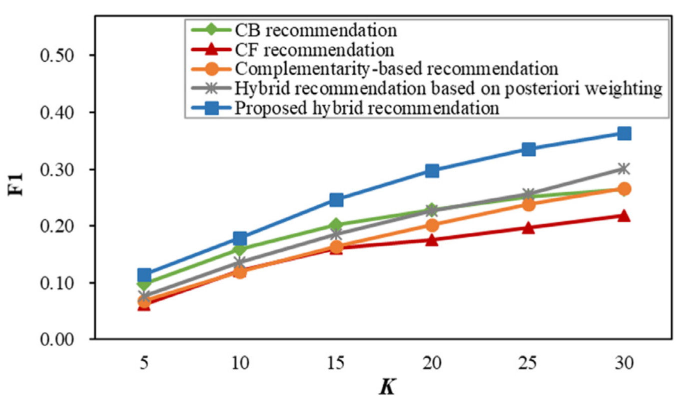

Fig. 15. Comparison of the hybrid recommendation with other recommendations based on F1.

F1, are shown in Tables 11–13. All the hypothesis tests are statistically significant. This indicates that the proposed hybrid recommendation method performs significantly better than other recommendation methods.

4.2.3. Performance of the new ranking algorithm

In this section, we calculate the coverage of Q&A documents for the combined recommendation list and re-rank the combined recommendation list. The results are shown in Table 14. Each row in the table indicates that the top K Q&A documents in the recommendation lists are used in the calculation of the experimental results. The other two columns are coverage of the recommendation list and coverage of the re-ranked recommendation list. Each cell is the corresponding value. For example, 0.963, which resides in the

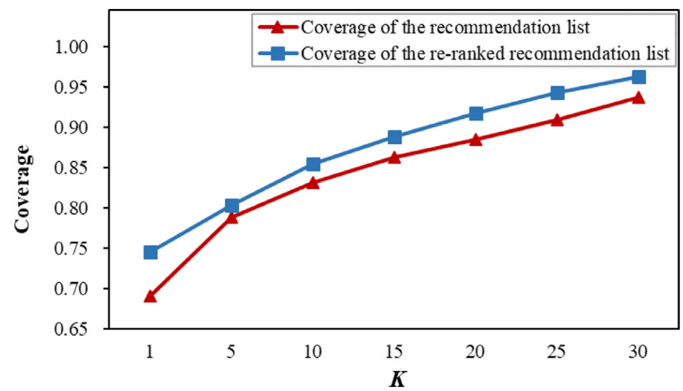

Fig. 16. Evaluation results of the new ranking algorithm.

third column and the eighth row, is the coverage of the top 30 Q&A documents to the entire list in the re-ranked recommendation list.

The results are also illustrated in Fig. 16. The experimental results indicate that the performance of the new ranking is better. For example, after re-ranking, the top 15 Q&A documents cover nearly 89% of the recommended list; that is, users need to read only the top 15 Q&A documents to grasp the main content of the entire recommendation list. However, users must read 20 Q&A documents to grasp 89% of the content of the total recommended list before re-ranking. Therefore, the new ranking algorithm can re-order the Q&A documents that cover the main content of the recommendation list at the top of the list, so users can read only a small number of Q&A documents to obtain sufficient knowledge and improve their learning efficiency.

4.3. Discussion

The results show that the proposed approach performs better with respect to the metrics. The experimental results of the CB recommendation, CF recommendation and complementarity-based recommendation show that the experimental results with partitioning of knowledge needs are better. This is mainly because the knowledge needs of users change, and partitioning ensures that the identified knowledge needs are up to date. Therefore, the recommendation methods perform better.

The performance of the combined list is better than that of any single list because different aspects are considered simultaneously. Regarding the re-ranking of the Q&A documents in the recommendation list, the coverage of the combined list is better than that of the original combined recommendation list. This method is useful in practical recommendation applications. It is difficult for users to browse all the Q&A documents in a list. Furthermore, users do not know how many Q&A documents they should read. With the proposed method, regardless of how many Q&A documents they read, these documents are the ones that cover most of the topics among the Q&A documents in the list.

Overall, the experimental results show that the partitioning of knowledge needs improves the performance of recommendations. The combined recommendation method further improves its performance. The re-ranking mechanism provides a novel and practical way to list the Q&A documents in a recommendation.

The absolute value of the experimental results of the three types of recommendation algorithms is related to the context of investigation to some extent. For example, each user has his own opinions and judgments. One user may find a document very relevant, whereas another user may not. Therefore, the values of the individual experimental results with respect to the evaluation metrics are related to the users. However, this does not influence the conclusions deduced from the experimental results because we fo-

Table 11
Statistical results of the proposed recommendation method on precision.

Top-K	Recommendation algorithm	T value	P value
5	Proposed recommendation > CB recommendation	2.62467	0.013344
	Proposed recommendation > CF recommendation	6.99733	7.46E-08
	Proposed recommendation > Complementarity-based recommendation	4.01819	0.000347
	Proposed recommendation > Hybrid recommendation based on posteriori weighting	5.14117	1.43E-05
10	Proposed recommendation > CB recommendation	2.25421	0.031395
	Proposed recommendation > CF recommendation	5.53622	4.62E-06
	Proposed recommendation > Complementarity-based recommendation	5.29829	9.13E-06
	Proposed recommendation > Hybrid recommendation based on posteriori weighting	3.68914	0.00086
15	Proposed recommendation > CB recommendation	4.80969	3.70E-05
	Proposed recommendation > CF recommendation	7.41521	2.37E-08
	Proposed recommendation > Complementarity-based recommendation	5.40223	6.78E-06
	Proposed recommendation > Hybrid recommendation based on posteriori weighting	4.19040	0.000215
20	Proposed recommendation > CB recommendation	6.03004	1.13E-06
	Proposed recommendation > CF recommendation	8.50008	1.34E-09
	Proposed recommendation > Complementarity-based recommendation	5.63209	3.51E-06
	Proposed recommendation > Hybrid recommendation based on posteriori weighting	4.88792	2.96E-05
25	Proposed recommendation > CB recommendation	6.54139	2.66E-07
	Proposed recommendation > CF recommendation	10.12183	2.40E-11
	Proposed recommendation > Complementarity-based recommendation	6.77737	1.37E-07
	Proposed recommendation > Hybrid recommendation based on posteriori weighting	6.17135	7.54E-07
30	Proposed recommendation > CB recommendation	7.57131	1.55E-08
	Proposed recommendation > CF recommendation	12.00811	3.42E-13
	Proposed recommendation > Complementarity-based recommendation	11.46255	1.12E-12
	Proposed recommendation > Hybrid recommendation based on posteriori weighting	5.83749	1.95E-06

Table 12
Statistical results of the proposed recommendation method on recall.

Top-K	Recommendation algorithm	T value	P value
5	Proposed recommendation > CB recommendation	3.11857	0.003907
	Proposed recommendation > CF recommendation	7.18691	4.43E-08
	Proposed recommendation > Complementarity-based recommendation	4.25766	0.000178
	Proposed recommendation > Hybrid recommendation based on posteriori weighting	4.47429	9.63E-05
10	Proposed recommendation > CB recommendation	2.26582	0.030594
	Proposed recommendation > CF recommendation	5.47379	5.52E-06
	Proposed recommendation > Complementarity-based recommendation	4.87986	3.03E-05
	Proposed recommendation > Hybrid recommendation based on posteriori weighting	3.35422	0.002112
15	Proposed recommendation > CB recommendation	5.17540	1.30E-05
	Proposed recommendation > CF recommendation	7.38598	2.57E-08
	Proposed recommendation > Complementarity-based recommendation	5.34153	8.07E-06
	Proposed recommendation > Hybrid recommendation based on posteriori weighting	3.88530	0.000502
20	Proposed recommendation > CB recommendation	6.33399	4.76E-07
	Proposed recommendation > CF recommendation	8.04849	4.35E-09
	Proposed recommendation > Complementarity-based recommendation	5.30406	8.98E-06
	Proposed recommendation > Hybrid recommendation based on posteriori weighting	4.03539	0.000331
25	Proposed recommendation > CB recommendation	6.58064	2.38E-07
	Proposed recommendation > CF recommendation	9.31338	1.70E-10
	Proposed recommendation > Complementarity-based recommendation	5.95951	1.38E-06
	Proposed recommendation > Hybrid recommendation based on posteriori weighting	4.56812	7.38E-05
30	Proposed recommendation > CB recommendation	7.49326	1.92E-08
	Proposed recommendation > CF recommendation	11.49983	1.03E-12
	Proposed recommendation > Complementarity-based recommendation	9.55659	9.36E-11
	Proposed recommendation > Hybrid recommendation based on posteriori weighting	4.04011	0.000327

cus on the comparison of the performances of different recommendation results. That is, our focus is on performance differences instead of absolute values. The bias of individual users will influence the experimental results of all the algorithms simultaneously. However, the difference values are stable. Moreover, in the experiments, the results of statistical tests show that the performance of the proposed hybrid recommendation system is significantly better.

This research has practical and managerial implications. It provides a better way to share Q&A documents and provides a reference for constructing Q&A document recommendation systems. It can also be adopted in current CQA websites to further help users find the required Q&A documents. In this study, the feedback of users was derived explicitly by the manual ratings. Implicit feedback, such as residence time and additions to favorites, can be used in automated settings. Q&A documents are a new carrier of knowledge. As an important source of knowledge, their sharing will provide an impetus for knowledge management, especially in the Web2.0 era. In the Web2.0 era, the knowledge to be managed in organizations is provided by answering questions. The knowledge is then organized as Q&A documents. This study provides an efficient way to obtain and share this knowledge. Consequently, the achievement of knowledge management objectives can be promoted.

The research also has research implications. This study provides a way to share the Q&A form of documents, which enriches

Table 13
Statistical results of the proposed recommendation method on F1.

Top-*K*	Recommendation algorithm	T value	P value
5	Proposed recommendation > CB recommendation	3.09551	0.004145
	Proposed recommendation > CF recommendation	7.20101	4.26E-08
	Proposed recommendation > Complementarity-based recommendation	4.25232	0.00018
	Proposed recommendation > Hybrid recommendation based on posteriori weighting	4.54917	7.78E-05
10	Proposed recommendation > CB recommendation	2.28434	0.029355
	Proposed recommendation > CF recommendation	5.49858	5.14E-06
	Proposed recommendation > Complementarity-based recommendation	4.94649	2.50E-05
	Proposed recommendation > Hybrid recommendation based on posteriori weighting	3.41268	0.001809
15	Proposed recommendation > CB recommendation	5.13811	1.45E-05
	Proposed recommendation > CF recommendation	7.44994	2.16E-08
	Proposed recommendation > Complementarity-based recommendation	5.40810	6.66E-06
	Proposed recommendation > Hybrid recommendation based on posteriori weighting	3.98793	0.000378
20	Proposed recommendation > CB recommendation	6.31401	5.04E-07
	Proposed recommendation > CF recommendation	8.24863	2.57E-09
	Proposed recommendation > Complementarity-based recommendation	5.41807	6.48E-06
	Proposed recommendation > Hybrid recommendation based on posteriori weighting	4.29369	0.00016
25	Proposed recommendation > CB recommendation	6.61077	2.19E-07
	Proposed recommendation > CF recommendation	9.64229	7.59E-11
	Proposed recommendation > Complementarity-based recommendation	6.18370	7.28E-07
	Proposed recommendation > Hybrid recommendation based on posteriori weighting	5.05686	1.82E-05
30	Proposed recommendation > CB recommendation	7.59810	1.45E-08
	Proposed recommendation > CF recommendation	11.91311	4.19E-13
	Proposed recommendation > Complementarity-based recommendation	10.27320	1.68E-11
	Proposed recommendation > Hybrid recommendation based on posteriori weighting	4.65040	5.84E-05

Table 14
Evaluation results of the new ranking algorithm.

Top-*K*	Coverage of the recommendation list	Coverage of the re-ranked recommendation list
1	0.691	0.745
5	0.789	0.803
10	0.831	0.854
15	0.862	0.888
20	0.884	0.917
25	0.909	0.942
30	0.937	0.963

research on overcoming information overload problems in CQA. Moreover, it enriches research on recommendation systems. It is a novel proposal to divide knowledge needs with sequence clustering. This method can also be used in research on other recommendation systems to provide a more accurate modeling of knowledge needs. It provides a new way to combine the initial recommendation lists via the Fermat point. The ranking mechanism can also be used in research on other recommendation systems, especially document recommendation systems, to make recommendation lists easier to use because users will not be confused by the number of documents to be read.

5. Conclusion and future works

In this paper, a novel hybrid approach to recommending Q&A documents is proposed. Knowledge needs are partitioned by sequential clustering, and current knowledge needs are identified. CB recommendation, CF recommendation and complementarity-based recommendation are used to find the Q&A documents that the user is most likely to prefer. To provide a comprehensive recommendation list, the three initial lists are combined, and a novel ranking mechanism is proposed to ensure that users obtain the most comprehensive knowledge with a limited number of Q&A documents to the greatest extent possible. The comprehensive experimental results show that the proposed method performs well. There are also limitations to be resolved in future research. The proposed hybrid recommendation system can be put into practice, and extensive experiments can be conducted to further verify the proposed approach. Although updating can be performed at fixed intervals, to provide more accurate recommendations, timely updating should be considered. In the experiments, feedback was given explicitly by ratings. In future research, implicit feedback can also be used, such as residence time and additions to favorites in automated settings. The DoE (Design of Experiment) will also be used for experimental analysis to obtain more accurate results.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Credit authorship contribution statement

Ming Li: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing. **Ying Li:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. **Wangqin Lou:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft. **Lisheng Chen:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - review & editing.

Acknowledgments

The research is supported by the National Natural Science Foundation of China under [grant numbers 71571191, 91646122], the Humanity and Social Science Youth Foundation of Ministry of Educationin China [grant number 15YJCZH081].

References

Armentano, M. G., Schiaffino, S., Christensen, I., & Boato, F. (2015). *Movies recommendation based on opinion mining in twitter* (pp. 80–91). Cham: Springer. doi:10.1007/978-3-319-27101-9_6.

Barragáns-Martínez, A. B., Costa-Montenegro, E., Burguillo, J. C., Rey-López, M., Mikic-Fonte, F. A., & Peleteiro, A. (2010). A hybrid content-based and item-based collaborative filtering approach to recommend TV programs enhanced with singular value decomposition. *Information Sciences, 180*(22), 4290–4311. doi:10.1016/J.INS.2010.07.024.

Blondet, G., Duigou, J. Le, & Boudaoud, N. (2019). A knowledge-based system for numerical design of experiments processes in mechanical engineering. *Expert Systems with Applications, 122*, 289–302. doi:10.1016/J.ESWA.2019.01.013.

Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction, 12*(4), 331–370. doi:10.1023/A:1021240730564.

Burke, R. (2007). Hybrid web recommender systems. In *The adaptive web* (pp. 377–408). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-540-72079-9_12.

Cao, Y., & Li, Y. (2007). An intelligent fuzzy-based recommendation system for consumer electronic products. *Expert Systems with Applications, 33*(1), 230–240. doi:10.1016/J.ESWA.2006.04.012.

Centobelli, P., Cerchione, R., & Esposito, E. (2018). Aligning enterprise knowledge and knowledge management systems to improve efficiency and effectiveness performance: A three-dimensional fuzzy-based decision support system. *Expert Systems with Applications, 91*, 107–126. doi:10.1016/J.ESWA.2017.08.032.

Cerchione, R., & Esposito, E. (2017). Using knowledge management systems: A taxonomy of SME strategies. *International Journal of Information Management, 37*(1), 1551–1562. doi:10.1016/J.IJINFOMGT.2016.10.007.

Chen, Y.-. J., Chen, Y.-. M., & Wu, M.-. S. (2012). An empirical knowledge management framework for professional virtual community in knowledge-intensive service industries. *Expert Systems with Applications, 39*(18), 13135–13147. doi:10.1016/J.ESWA.2012.05.088.

Chen, Z., Zhang, C., Zhao, Z., Yao, C., & Cai, D. (2018). Question retrieval for community-based question answering via heterogeneous social influential network. *Neurocomputing, 285*, 117–124. doi:10.1016/J.NEUCOM.2018.01.034.

Cheng, X., Yan, X., Lan, Y., & Guo, J. (2014). BTM: Topic modeling over short texts. *IEEE Transactions on Knowledge and Data Engineering, 26*(12), 2928–2941. doi:10.1109/TKDE.2014.2313872.

Figueroa, A. (2017). Automatically generating effective search queries directly from community question-answering questions for finding related questions. *Expert Systems with Applications, 77*, 11–19.

Figueroa, A., & Neumann, G. (2016). Context-aware semantic classification of search queries for browsing community question–answering archives. *Knowledge-Based Systems, 96*, 1–13. doi:10.1016/J.KNOSYS.2016.01.008.

Fisher, W. D. (1958). On grouping for maximum homogeneity. *Journal of the American Statistical Association, 53*(284), 789–798. doi:10.1080/01621459.1958.10501479.

Fujii, A. (2008). Modeling anchor text and classifying queries to enhance web document retrieval. In *Proceeding of the 17th international conference on World Wide Web - WWW '08* (p. 337). New York: ACM Press. doi:10.1145/1367497.1367544.

Geman, S., & Geman, D. (1993). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images*. *Journal of Applied Statistics, 20*(5–6), 25–62. doi:10.1080/02664769300000058.

Huang, Z., Lu, X., Duan, H., & Zhao, C. (2012). Collaboration-based medical knowledge recommendation. *Artificial Intelligence in Medicine, 55*(1), 13–24. doi:10.1016/J.ARTMED.2011.10.002.

Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal, 16*(3), 261–273. doi:10.1016/J.EIJ.2015.06.005.

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters, 31*(8), 651–666. doi:10.1016/J.PATREC.2009.09.011.

Ji, Z., Xu, F., Wang, B., & He, B. (2012). Question-answer topic model for question retrieval in community question answering. In *Proceedings of the 21st ACM international conference on Information and knowledge management* (pp. 2471–2474). ACM.

Ji, Z., Xu, F., Wang, B., & He, B. (2019). Question-Answer topic model for question retrieval in community question answering, 2471–2474.

Karna, N., Supriana, I., & Maulidevi, N. (2017). Recommendation system on knowledge management system via OAI-PMH. In *2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)* (pp. 1–5). IEEE. doi:10.1109/EECSI.2017.8239154.

Lai, C.-. H., & Liu, D.-. R. (2009). Integrating knowledge flow mining and collaborative filtering to support document recommendation. *Journal of Systems and Software, 82*(12), 2023–2037. doi:10.1016/J.JSS.2009.06.044.

Lai, C.-. H., Liu, D.-. R., & Lin, C.-. S. (2013). Novel personal and group-based trust models in collaborative filtering for document recommendation. *Information Sciences, 239*, 31–49. doi:10.1016/J.INS.2013.03.030.

Lee, V.-. H., Foo, A. T.-L., Leong, L.-. Y., & Ooi, K.-. B. (2016). Can competitive advantage be achieved through knowledge management? A case study on SMEs. *Expert Systems with Applications, 65*, 136–151. doi:10.1016/J.ESWA.2016.08.042.

Lenhart, P., & Herzog, D. (2016, September). Combining Content-based and Collaborative Filtering for Personalized Sports News Recommendations. In CBRecSys@ RecSys (pp. 3–10).

Liu, D.-. R., Chen, Y.-. H., & Huang, C.-. K. (2014). QA document recommendations for communities of question–answering websites. *Knowledge-Based Systems, 57*, 146–160. doi:10.1016/J.KNOSYS.2013.12.017.

Liu, D.-. R., Chen, Y.-. H., Shen, M., & Lu, P.-. J. (2015). Complementary QA network analysis for QA retrieval in social question-answering websites. *Journal of the Association for Information Science and Technology, 66*(1), 99–116. doi:10.1002/asi.23155.

Liu, D.-. R., Lai, C.-. H., & Chiu, H. (2011). Sequence-based trust in collaborative filtering for document recommendation. *International Journal of Human-Computer Studies, 69*(9), 587–601. doi:10.1016/J.IJHCS.2011.06.001.

Lops, P., de Gemmis, M., & Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. In *Recommender systems handbook* (pp. 73–105). Boston, MA: Springer US. doi:10.1007/978-0-387-85820-3_3.

Lu, Z., Dou, Z., Lian, J., Xie, X., & Yang, Q. (2015). Content-Based Collaborative Filtering for News Topic Recommendation. In *Proceedings of theTwenty-ninth AAAI conference on artificial intelligence* (pp. 217–223). Austin, Texas: AAAI press.

Ma, B., Wei, Q., & Chen, G. (2011). A combined measure for representative information retrieval in enterprise information systems. *Journal of Enterprise Information Management, 24*(4), 310–321. doi:10.1108/17410391111148567.

Ma, J., Zhang, Y., & Zhang, L. (2017). *Mobile spam filtering base on BTM topic model* (pp. 657–665). Cham: Springer. doi:10.1007/978-3-319-49109-7_63.

Mao, X. L., Hao, Y. J., Wang, D., & Huang, H. (2018). Query completion in community-based question answering search. *Neurocomputing, 274*, 3–7. doi:10.1016/j.neucom.2016.06.096.

Muthmann, K., & Petrova, A. (2014). An automatic approach for identifying topical near-duplicate relations between questions from social media Q/A sites. *Proceeding of WSDM 2014 Workshop: "Web-Scale Classification: Classifying Big Data from the Web."*.

Ohta, M., Hachiki, T., & Takasu, A. (2011). Related paper recommendation to support online-browsing of research papers. In *Fourth International Conference on the Applications of Digital Information and Web Technologies (ICADIWT 2011)* (pp. 130–136). IEEE. doi:10.1109/ICADIWT.2011.6041413.

Okura, S., Tagami, Y., Ono, S., & Tajima, A. (2017). Embedding-based news recommendation for millions of users. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '17* (pp. 1933–1942). New York: ACM Press. doi:10.1145/3097983.3098108.

Ong, C.-. S., Day, M.-. Y., & Hsu, W.-. L. (2009). The measurement of user satisfaction with question answering systems. *Information & Management, 46*(7), 397–403. doi:10.1016/J.IM.2009.07.004.

Palomera, D., & Figueroa, A. (2017). Leveraging linguistic traits and semi-supervised learning to single out informational content across how-to community question-answering archives. *Information Sciences, 381*, 20–32. doi:10.1016/J.INS.2016.11.006.

Pazzani, M. J., & Billsus, D. (2007). Content-Based recommendation systems. In *The adaptive web* (pp. 325–341). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-540-72079-9_10.

Rodrigo, A., Herrera, J., & Peñas, A. (2019). The effect of answer validation on the performance of question-answering systems. *Expert Systems with Applications, 116*, 351–363. doi:10.1016/J.ESWA.2018.09.014.

Roy, P. K., Ahmad, Z., Singh, J. P., Alryalat, M. A. A., Rana, N. P., & Dwivedi, Y. K. (2018). Finding and ranking high-quality answers in community question answering sites. *Global Journal of Flexible Systems Management, 19*(1), 53–68.

Salehi, M., & Nakhai Kamalabadi, I. (2013). Hybrid recommendation approach for learning material based on sequential pattern of the accessed material and the learner's preference tree. *Knowledge-Based Systems, 48*, 57–69. doi:10.1016/J.KNOSYS.2013.04.012.

Schafer, J. Ben, Frankowski, D., Herlocker, J., & Sen, S. (2007). Collaborative filtering recommender systems. In *The adaptive web* (pp. 291–324). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-540-72079-9_9.

Spain, P. G. (1996). The fermat point of a triangle. *Mathematics Magazine, 69*(2), 131–133. doi:10.1080/0025570X.1996.11996409.

Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*, 1–19 *2009*. doi:10.1155/2009/421425.

Subramaniyaswamy, V., Logesh, R., Chandrashekhar, M., Challa, A., & Vijayakumar, V. (2017). A personalised movie recommendation system based on collaborative filtering. *International Journal of High Performance Computing and Networking, 10*(1/2), 54. doi:10.1504/IJHPCN.2017.083199.

Subramaniyaswamy, V., & Pandian, S. C. (2014). Topic ontology-based efficient tag recommendation approach for blogs. *International Journal of Computational Science and Engineering, 9*(3), 177. doi:10.1504/IJCSE.2014.060682.

Sugiyama, K., & Kan, M.-. Y. (2015). A comprehensive evaluation of scholarly paper recommendation using potential citation papers. *International Journal on Digital Libraries, 16*(2), 91–109. doi:10.1007/s00799-014-0122-2.

Tewari, A. S., & Barman, A. G. (2018). Sequencing of items in personalized recommendations using multiple recommendation techniques. *Expert Systems with Applications, 97*, 70–82. doi:10.1016/J.ESWA.2017.12.019.

Vuori, V., & Okkonen, J. (2012). Knowledge sharing motivational factors of using an intra-organizational social media platform. *Journal of Knowledge Management, 16*(4), 592–603. doi:10.1108/13673271211246167.

Wang, H., & Meng, X. (2019). Transformation from IT-based knowledge management into BIM-supported knowledge management: A literature review. *Expert Systems with Applications, 121*, 170–187. doi:10.1016/J.ESWA.2018.12.017.

Wang, K., Ming, Z.-. Y., Hu, X., & Chua, T.-. S. (2010). Segmentation of multi-sentence questions: Towards effective question retrieval in cqa services. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval* (pp. 387–394).

Wang, Z., & Wang, N. (2012). Knowledge sharing, innovation and firm performance. *Expert Systems with Applications, 39*(10), 8899–8908. doi:10.1016/J.ESWA.2012.02.017.

Wen, J., Tu, H., Cheng, X., Xie, R., & Yin, W. (2019). Joint modeling of users, questions and answers for answer selection in CQA. *Expert Systems with Applications, 118*, 563–572. doi:10.1016/J.ESWA.2018.10.038.

Wu, H., Wu, W., Zhou, M., Chen, E., Duan, L., & Shum, H.-. Y. (2014). Improving search relevance for short queries in community question answering. In *Proceedings of the 7th ACM international conference on Web search and data mining - WSDM '14* (pp. 43–52). New York: ACM Press. doi:10.1145/2556195.2556239.

Wu, M., Hawking, D., Turpin, A., & Scholer, F. (2012). Using anchor text for homepage and topic distillation search tasks. *Journal of the American Society for Information Science and Technology, 63*(6), 1235–1255. doi:10.1002/asi.22639.

Xiao, Y., Ai, P., Hsu, C., Wang, H., & Jiao, X. (2015). Time-ordered collaborative filtering for news recommendation. *China Communications, 12*(12), 53–62. doi:10.1109/CC.2015.7385528.

Xue, X., Jeon, J., & Croft, W. B. (2008). Retrieval models for question and answer archives. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '08* (p. 475). New York: ACM Press. doi:10.1145/1390334.1390416.

Zhang, Jin, Liu, G., & Ren, M. (2016). Finding a representative subset from large-scale documents. *Journal of Informetrics, 10*(3), 762–775. doi:10.1016/J.JOI.2016.05.003.

Zhang, Juheng, & Piramuthu, S. (2018). Product recommendation with latent review topics. *Information Systems Frontiers, 20*(3), 617–625. doi:10.1007/s10796-016-9697-z.

Zhang, K., Wu, W., Wu, H., Li, Z., & Zhou, M. (2014). Question retrieval with high quality answers in community question answering. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management - CIKM '14* (pp. 371–380). New York: ACM Press. doi:10.1145/2661829.2661908.

Zhang, S., Zhang, S., Yen, N. Y., & Zhu, G. (2017). The recommendation system of micro-blog topic based on user clustering. *Mobile Networks and Applications, 22*(2), 228–239. doi:10.1007/s11036-016-0790-9.

Zhang, W.-. N., Liu, T., Yang, Y., Cao, L., Zhang, Y., & Ji, R. (2014). A topic clustering approach to finding similar questions from large question and answer archives. *PloS one, 9*(3), e71511. doi:10.1371/journal.pone.0071511.

Zhao, W., Wu, R., & Liu, H. (2016). Paper recommendation based on the knowledge gap between a researcher's background knowledge and research target. *Information Processing & Management, 52*(5), 976–988. doi:10.1016/J.IPM.2016.04.004.

Zhou, G., Liu, Y., Liu, F., Zeng, D., & Zhao, J. (2013). Improving question retrieval in community question answering using world knowledge. *Twenty-Third International Joint Conference on Artificial Intelligence*. Retrieved from https://www.aaai.org/ocs/index.php/IJCAI/IJCAI13/paper/viewPaper/6581.