A Collaborative Document Ranking Model for a Multi-Faceted Search

Laure Soulier, Lynda Tamine and Wahiba Bahsoun

IRIT Laboratory - 118 route de Narbonne, 31062 Toulouse Cedex 9

{soulier, tamine, wbahsoun}@irit.fr

Abstract. This paper presents a novel collaborative document ranking model which aims at solving a complex information retrieval task involving a multi-faceted information need. For this purpose, we consider a group of users, viewed as experts, who collaborate by addressing the different query facets. We propose a two-step algorithm based on a relevance feedback process which first performs a document scoring towards each expert and then allocates documents to the most suitable experts using the Expectation-Maximisation learning-method. The performance improvement is demonstrated through experiments using TREC interactive benchmark.

Keywords: collaborative information retrieval, multi-faceted search

1 Introduction

It is well-known in information retrieval (IR) domain that one critical issue is the understanding of users' search goals hidden behind their queries, in attempt to get a better insight of how to explore the information space towards relevant documents [2]. Most retrieval approaches consider that a query addresses mainly a single topic whereas multi-faceted search-based approaches [1, 13] have highlighted the need of considering the topical coverage of the retrieved documents towards different aspects of the query topic, named facets. One multi-faceted query example, extracted from the TREC Interactive dataset [17], is "Hubble Telescope Achievements". For this query, users have manually identified, for instance, several aspects: "focus of camera", "age of the universe space telescope" and "cube pictures". The key emerging challenges from multi-faceted search are how to infer the different query facets and how to exploit them jointly to select relevant results. To tackle the underlying issues, a first category of work attempts to enhance the query representation in order to identify the query facets [5, 22], whereas a second line of work in the same category [4, 12] considers result diversification towards query facets.

Another category of work [15, 19] arisen from the Collaborative IR (CIR) domain underlines that collaboration could benefit complex tasks and more particularly exploratory queries. Indeed, complex problems can be difficultly solved within ad-hoc IR due to the single searcher's knowledge or skills inadequacy [19]. A collaborative framework enables overcoming this lack considering that a group of users may analyze more in-depth the different query facets in contrast to a single user who performs individually the first level of the information need.

Shah [19] introduces the notion of the synergic effect of collaborative search in so far as *"the whole (is) greater than the sum of all"*. Therefore, CIR is another response to tackle multi-faceted search issues in which multiple users search documents together in response to a shared information need considering their different knowledge expertise and points of view with respect to the same query.

In this paper, we present a collaborative document ranking model suited to solve a multi-faceted query. Our approach allows to leverage the users' different knowledge expertise and assigns them implicit knowledge-based roles towards at least one query facet. These facets are modeled through document and user topical-based representations using the LDA generative model. Documents are allocated to experts using the Expectation-Maximization algorithm.

In section 2, we discuss related work. Section 3 describes our collaborative ranking approach involving an expert group. Section 4 presents our experimental evaluation. Finally, section 5 concludes the paper and introduces future work.

2   Related Work

The multi-faceted search issue can arise within two search settings, namely the individual-based search and the collaborative based-one.

Individual-based search is a classic IR setting in which one user aims at satisfying an information need. We address particularly in this paper a complex information need expressed by means of a multi-faceted query. There are two lines of work in this area. The first one remains on identifying explicit query facets prior to performing a facet-based retrieval model [5, 6, 9]. Authors use different tools or methods addressed at the query level such as terminological resources for building a hierarchy among the detected facets [5]. Other methods [9] remain on the analysis of the user navigation for classifying query facets using document features. In contrast, generative probabilistic-based models [3, 6] aim at modeling documents as topical-based vectors in which each component expresses a query facet. Another line of work [4, 24] dealing with multi-faceted search focuses on the diversification of the search results without highlighting the query facets. The key idea is to select a diversified document subset either using a term-frequency distribution [4] or a graph modeling approach [24].

Unlike individual-based search, the collaborative-based one is a retrieval setting in which several users collaborate for satisfying a shared information need. Considering a multi-faceted information need, the integration of collaboration within a search task enables to benefit from the *synergic effect* of collaboration [19] in so far as people with different backgrounds and points of view search relevant documents together with respect to a multi-faceted shared information need. Collaboration principles [10], namely *awareness*, *sharing of knowledge* and *division of labor*, enable to leverage from users their different skills and search abilities by warming collaborators of actions of their pairs, transmitting the information flow among users and avoiding redundancy. Previous work in the CIR domain [10, 16, 18, 21] rely on different retrieval strategies, such as designing user roles and relevance feedback techniques, in order to explore the different query facets. We distinguish two main categories of work depending on if they integrate

user roles or not. The first category remains on a relevance feedback process without integrating user roles [10, 16], providing iteratively a ranked list of documents either at the user level [10, 16] or at the group level [16]. The second category of work, more close to ours, consists in integrating a supplementary layer based on user roles [18, 21] by assigning to each collaborator a specific task for avoiding redundancy. Pickens et al. [18] model the *prospector* and *miner* roles which respectively aim at ensuring the quality of the selected documents and favoring the diversity of the search results in response to the multi-faceted information need by means of query suggestion. This model is guided by the division of labor and sharing of knowledge principles throughout a relevance feedback process.Shah et al. [21] introduce another couple of roles, namely the *gatherer* and the *surveyor* relying on different predefined tasks: respectively, performing a quick overview of documents for detecting relevant ones or a better understanding of the topic to explore other subtopic areas in accordance to the multi-faceted query.

In this paper, we focus on a collaborative context in which a multi-faceted information need is solved by a group of users. Unlike previous work based on explicit user roles [18, 21], our model assigns to users implicit roles considering their domain knowledge mined through a relevance feedback process. Therefore, our model enables to leverage users' skills in which they are the most effective, in opposition to other models [18, 21] based on explicit roles which assign predefined users roles or domain expertise. For this purpose, we model topical-based profiles for estimating the knowledge expertise of users and propose an iterative document ranking model using a learning-based algorithm in order to assign documents to the most likely suited users according to their domain knowledge.

3   The Model

Several underlying concerns of multi-faceted search within a CIR context may arise, such as identifying the most suitable users for solving a multi-topical information need or performing a collaborative retrieval task according to the different query facets. We focus here on the second concern. More specifically, we address two main research questions: 1) How to mine the query facets? 2) How to build collaborative document rankings?

3.1   Mining Query Facets

First, given an initial document dataset $\mathcal{D}$ and a multi-faceted query Q, we extract a subset $\mathcal{D}^*$ of n documents which satisfies the condition of being relevant with respect to the query topic and ensuring a broad topical coverage. For this purpose, we use the *Maximal Marginal Relevance* (MMR) [4] that selects document $D^* \in \mathcal{D}$ with the highest relevance and marginal diversity score considering subset $\mathcal{D}^*$:

$$D^* = \arg \max_{D_i \in \mathcal{D}}[\gamma sim_1(D_i, Q) - (1 - \gamma) \max_{D_{i'} \in \mathcal{D}^*} sim_2(D_i, D_{i'})] \qquad (1)$$

where $sim_1(D_i, Q)$ and $sim_2(D_i, D_{i'})$ express the similarity between document D_i and query Q, respectively document $D_{i'}$. $\gamma \in [0, 1]$ is a weighting factor.

Second, we use the generative probabilistic algorithm named LDA [3] applied on the diversified document dataset $\mathcal{D}^*$ in order to identify latent topics. Each topic is assimilated to a facet of query Q. The algorithm computes a word-topic distribution $\phi_{w|t}$ and document-topic distribution $\theta_{D_i|t}$ for respectively estimating the probability of word w and document D_i given topic t. The optimal number of topics T for the document dataset $\mathcal{D}^*$ is generally tuned using a likelihood measure:

$$l(T^*|w,t) = \arg \max_{T} \sum_{w \in W} log(\sum_{t \in T} p(w|t)) \qquad (2)$$

where W is the set of words extracted from document dataset $\mathcal{D}^*$. The probability $p(w|t)$ corresponds to the word-topic distribution $\phi_{w|t}$.

Each document $D_i \in \mathcal{D}^*$ is represented using a topical distribution $D_i = (w_{1i}, \ldots, w_{ti}, \ldots, w_{Ti})$ where w_{ti} represents the weight of mined topic t by the LDA algorithm for document D_i expressed by the probability $\theta_{D_i|t}$. Each expert $E_j \in E$ is characterized by a knowledge profile built using the previously selected documents and inferred on the document representation using the LDA inference method. The profile of expert E_j is noted $\pi(E_j)^{(k)} = (w_{1j}^{(k)}, \ldots, w_{tj}^{(k)}, \ldots, w_{Tj}^{(k)})$, where $w_{tj}^{(k)}$ represents the expertise of expert E_j towards topic t at iteration k.

3.2 Collaborative Document Ranking

Expert-based Document Scoring. The expert-based document scoring aims at reranking documents with respect to each expert's domain expertise towards the query facets. It provides a first attempt for document ranking given experts which is, then, used within the collaborative learning method. We estimate, at iteration k, the probability $p^{(k)}(d_i|e_j, q)$ of document D_i given expert E_j and query Q as follows:

$$p^{(k)}(d_i|e_j, q) = \frac{p^{(k)}(e_j|d_i, q)p(d_i|q)}{p(e_j|q)} \qquad (3)$$

where d_i, e_j and q are random variables associated respectively with document D_i, expert E_j and query Q. Considering the probability $p(e_j|q)$ is not discriminant and assuming that expert e_j and query q are independent, we obtain:

$$p^{(k)}(d_i|e_j, q) \propto p^{(k)}(e_j|d_i, q)p(d_i|q) \qquad (4)$$
$$\propto p^{(k)}(e_j|d_i)p(d_i|q)$$

We first estimate the probability $p(d_i|q)$ of document D_i given query Q as:

$$p(d_i|q) = \frac{\frac{p(d_i).p(q|d_i)}{p(q)}}{\sum_{D_{i'} \in \mathcal{D}^*} \frac{p(d_{i'}).p(q|d_{i'})}{p(q)}} \propto \frac{p(d_i).p(q|d_i)}{\sum_{D_{i'} \in \mathcal{D}^*} p(d_{i'}).p(q|d_{i'})} \qquad (5)$$

with $\mathcal{D}^*$ is the reference dataset. The probability $p(d_i)$ of document D_i is independent of the query and can be estimated as a uniform weight: $p(d_i) = \frac{1}{|\mathcal{D}|}$.

Considering the facet-based distribution of document D_i, we estimate $p(q|d_i)$ by combining two similarity scores $RSV_{LDA}(Q|D_i)$ and $RSV_{BM25}(Q, D_i)$ respectively based on the LDA-based document ranking model detailed in [11] and a BM25-based document scoring:

$$p(q|d_i) = \lambda RSV_{LDA}(Q|D_i) + (1 - \lambda)RSV_{BM25}(Q, D_i) \qquad (6)$$

$$with \ RSV_{LDA}(Q|D_i) = \prod_{w \in Q} \sum_{t=1}^{T} p(w|t).p(t|d_i)$$

where $p(w|t)$ represents the probability of term w given topic t estimated by $\phi_{w|t}$ and $p(t|d_i)$ is the probability of topic t given document D_i, previously noted w_{ti} in the document representation. We estimate the probability $p^{(k)}(e_j|d_i)$ as a cosine similarity sim_{cos} comparing the two topical distributions of document D_i and knowledge profile $\pi(E_j)^{(k)}$ associated to expert E_j at iteration k as:

$$p^{(k)}(e_j|d_i) = \frac{sim_{cos}(D_i, \pi(E_j)^{(k)})}{\sum_{D_{i'} \in \mathcal{D}^*} sim_{cos}(D_{i'}, \pi(E_j)^{(k)})} \qquad (7)$$

Expert-based Document Allocation. Here, we aim at allocating documents to the most suitable experts considering document scores computed in Equation 3. For this aim, we use the learning *Expectation Maximization* algorithm [7]. The EM-based collaborative document allocation method runs into two stages, as detailed in Algorithm 1. Notations are detailed in Table 1.

1. Learning the document-expert mapping. The aim is to learn through an EM-based algorithm how experts are likely to assess the relevance of a document. This method is divided into two steps:
- The *E-step* estimates the probability $p(c_j = 1|X_i^{(k)})$ of relevance of document D_i towards expert E_j considering all experts at iteration k considering document score vector $X_i^{(k)}$. The probability $p(c_j|X_i^{(k)})$ is estimated using a mixture model that considers Gaussian probability laws ϕ_j to model the relevance c_j of documents for expert E_j at iteration k:

$$p(c_j = 1|X_i^{(k)}) = \frac{\alpha_j p_{ij}^{(k)}}{\alpha_j p_{ij}^{(k)} + (1 - \alpha_j)\bar{p}_{ij}} = \frac{\alpha_j p_{ij}^{(k)}}{\sum_{l=1}^{m} \alpha_l p_{il}^{(k)}} \qquad (8)$$

$$with \ \begin{cases} p_{ij}^{(k)} = p(c_j = 1)p(x_{ij}^{(k)}|c_j = 1) \\ \bar{p}_{ij}^{(k)} = p(c_j = 0)p(x_{ij}^{(k)}|c_j = 0) \end{cases}$$

If we consider the fact that document irrelevance towards expert E_j, noted $c_j = 0$, can be formulated by the probability of being relevant for another expert, noted $c_l = 1 \ \forall l = \{1, \ldots, m\} \ with \ l \neq j$, the denominator corresponds to the sum of the probabilities expressing the document relevance, noted $c_l = 1$, towards expert E_l. Thus, we replace the probabilities $pij^{(k)}$ and $p_{il}^{(k)}$ by the Gaussian

$X_i^{(k)} = \{x_{i1}^{(k)}, ..., x_{im}^{(k)}\}$	The scoring vector where each element $x_{ij}^{(k)}$ is estimated by equation 4.
$X^{(k)} \in \mathcal{R}^{n \times m}$	The matrix including the n vectors $X_i^{(k)}$.
$c_j = \{0, 1\}$	The hidden variable referring to the irrelevance, respectively relevance, of a document of belonging to category c_j of expert E_j.
ϕ_j	The Gaussian probability density function of relevant documents respectively according to expert E_j.
θ_j	The parameters for the Gaussian score distribution ϕ_j related to expert E_j, namely μ_j and σ_j.
α_j	The coefficient of the mixture model assuming that $\sum_{j=1}^m \alpha_j = 1$.

Table 1: Notations used in Algorithm 1

Algorithm 1: EM-based collaborative document ranking

Data: $\mathcal{D}$, $X^{(k)}$, E
Result: $M^{EM,k} \in \mathcal{R}^{n \times m}$
begin

    /* Stage 1: Learning the document-expert mapping   */
    while *nonconvergence* **do**

        /* E-step   */
        forall the *documents $D_i \in \mathcal{D}$* **do**
            forall the *experts $E_j \in E$* **do**

$$p(c_j = 1|X_i^{(k)}) = \frac{\alpha_j \phi_j(x_{ij}^{(k)})}{\sum_{l=1}^m \alpha_l \phi_l(x_{il}^{(k)})}$$

$$M_{ij}^{EM,k} = p(c_j = 1|X_i^{(k)})$$

        /* M-step   */
        forall the *experts $E_j \in E$* **do**

$$S_j = \sum_{h=1}^n p(c_j = 1|X_h^{(k)})$$

$$\alpha_j = \frac{1}{n} \sum_{h=1}^n p(c_j = 1|X_h^{(k)})$$

$$\mu_j = \frac{1}{S_j} \sum_{h=1}^n p(c_j = 1|X_h^{(k)}).x_{hj}^{(k)}$$

$$\sigma_j = \frac{1}{S_j} \sum_{h=1}^n p(c_j = 1|X_h^{(k)}).(x_{hj^{(k)}} - \mu_j)^2$$

    /* Stage 2: Allocating documents to experts   */
    $M^{EM,k} = odds(M^{EM,k})$
    Return $M^{EM,k}$

---

density values $\phi_j(x_{ij}^{(k)})$ and $\phi_l(x_{il}^{(k)})$ for obtaining the final estimated probability.
- The *M-step* updates the parameters θ_j and allows estimating the "Expected Complete Data Log Likelihood":

$$\mathcal{L}(c_j = 1|X_j^{(k)}, \theta_j) = \sum_{h=1}^n \sum_{l=1}^m log(p(c_l = 1|X_h^{(k)}))p(c_l = 1|X_h^{(k)}) \qquad (9)$$

The algorithm convergence is reached when the log-likelihood is maximized.

2. Allocating documents to experts. The key issue is to detect which expert is more likely to assess the relevance of a document. Similarly to the probabilistic model assumption, we assign to each matrix element $M_{ij}^{EM,k}$ the odds value $odds(M_{ij}^{EM,k})$, computed as the ratio between the probability of relevance $M_{ij}^{EM,k}$ and the probability of irrelevance, estimated by $\sum_{\substack{l=1 \\ l \neq j}}^{m} M_{il}^{EM,k}$. The classification is done using this output matrix $M^{EM,k}$ where document D_i is allocated to expert E_j by maximizing the probability $p(c_j|X_i^{(k)})$ that document D_i is relevant for expert E_j considering its scores $X_i^{(k)}$:

$$\forall D_i, \ \exists \ E_j^*; \ E_j^* = \arg \max_{E_j \in E} M_{ij}^{EM,k} \tag{10}$$

Moreover, we propose an additional layer which ensures division of labor by removing from document allocation towards user u_j, documents already displayed within collaborators' lists.

4 Experimental Evaluation

Considering that it does not exist online collaborative search log, except proprietary ones [15, 19], the retrieval effectiveness of our model was evaluated through a simulation-based framework which is an extension of the experimental framework proposed in Foley et al [10]. We used the same dataset, namely the TREC 6-7-8 Interactive one[1] which models users' interactions within an interactive-based IR task. One of the goals of users who perform this task is to identify several instances, namely aspects, related to the information need [17].

4.1 Experimental Setup

Dataset. We use the TREC Financial Times of London 1991-1994 Collection (Disk 4 of the TREC ad-hoc Collection) which includes 210 158 articles. We performed our effectiveness evaluation on 277 collaborative query search sessions extracted from dataset runs and built upon 20 TREC initial topics. The latter are analyzed for checking whether they are multi-faceted by estimating the diversity coverage of the top 1000 documents. We use the Jaccard distance between documents in pairs which avoids document length bias. We retained the whole set of 20 TREC topics, characterized by diversity coverage values very close to 1. Considering our collaborative model, we retain the 10 participants who have provided "rich format data" including the list of documents selected by the user and their respective selection time-stamp label. For each TREC topic, we retained as relevance judgments the respective feedback provided by TREC participants. For improving their reliability, we ensured the agreement between users by considering only documents which have been assessed twice as relevant. The agreement level is tuned in section 4.2 for testing the stability of our model.

[1] http://trec.nist.gov/data/interactive.html

Collaboration Simulation. Here, we extend the experimental protocol proposed in [10] by building groups of experts. Considering that experts use generally a more specific vocabulary [21, 23], we analysed the expertise level $Expertise(U_j, Q)$ of each user U_j with respect to query Q using relevance feedback expressed through their respective TREC runs. For this purpose, we estimate the average specificity of the selected document set $\mathcal{D}^{S_Q(U_j)}$ using the specificity indicator $Pspec$ [14] for search session S_Q related to query Q:

$$Expertise(U_j, Q) = \frac{\sum_{D_i \in \mathcal{D}^{S_Q(U_j)}} \mathcal{L}_s(D_i)}{|\mathcal{D}^{S_Q}(U_j)|} \tag{11}$$

with $\mathcal{L}_s(D_i) = \underset{t \in D_i}{avg} \, Pspec(t) = \underset{t \in D_i}{avg} (-log(\frac{df_t}{N}))$. The document frequency of term t is noted df_t and N is the number of documents in the collection.

For each query Q, we performed a 2-means classification of users from all the participant groups who have achieved an interactive task considering their respective expertise level as criteria. We identified as experts users who belong to the class with the highest average expertise level value. Within each participant group and for each query, we perform all the combinations of size m, with $m \geq 2$, for building the set of 277 groups of experts. We notice that 19 TREC topics enable to form at least one group of experts of size $m \geq 2$. For each group of experts, we identified the time-line of relevance feedback which represents the whole set of selected documents by all the experts of the group synchronized chronologically by their respective time-stamp feature. We carefully consider the time-line of the collaborative search session by ensuring that every document assessed as relevant is displayed in the user's document list. Moreover, assuming that a user focuses his attention on the 30 top documents within a ranked list [10], only those are displayed to the user.

Metrics. We highlight here that CIR implies a different effectiveness evaluation approach compared to ad-hoc IR. Indeed, even if the goal of a collaborative document ranking model is to select relevant documents, the main objective remains on supporting the collaboration within the group [20]. For estimating the retrieval effectiveness of our model at the session level, we used metrics proposed in [20]. We consider measures at rank 30 considering the length of the displayed document lists. The evaluation metrics are the following:

- $Cov@R$: the coverage ratio at rank R for analysing the diversity of the search results displayed during the whole search session:

$$Cov@R = \frac{1}{|\theta|} \sum_{Q \in \theta} \frac{1}{|E_Q|} \sum_{e \in E_Q} \frac{Coverage(L_{e,Q})}{\sum_{l \in L_{e,Q}} |l|} \tag{12}$$

where θ is the set of TREC topics, E_Q represents the set of groups e of experts who have collaborated for solving query Q and $L_{e,Q}$ is the set of displayed lists related to query Q for group e. $Coverage(L_{e,Q})$ corresponds to the number of distinct documents displayed to expert group e for query Q. The total number of documents displayed throughout the same session is noted $|l|$.

- *RelCov@R*, the relevant coverage ratio at rank R which adds the supplementary layer to the coverage measure by including the condition that distinct displayed documents should be relevant:

$$RelCov@R = \frac{1}{|\theta|} \sum_{Q \in \theta} \frac{1}{|E_Q|} \sum_{e \in E_Q} \frac{RelevantCoverage(L_{e,Q})}{\sum_{l \in L_{e,Q}} |l|} \tag{13}$$

where $RelevantCoverage(L_{e,Q})$ corresponds to the number of distinct relevant documents displayed to expert group e for query Q.
- *P@R*: the average precision at rank R:

$$P@R = \frac{1}{|\theta|} \sum_{Q \in \theta} \frac{1}{|E_Q|} \sum_{e \in E_Q} \frac{1}{|L_{E,Q}|} \sum_{l \in L_{e,Q}} \frac{D_{SelRel}(Q,l)}{\sum_{l \in L_{e,Q}} |l|} \tag{14}$$

where $D_{SelRel}(Q,l)$ represents the number of relevant documents retrieved within document list l related to query Q.

Baselines. We performed four scenarios considering either an individual-based search or a collaborative-based one:
- *W/oEMDoL* is the individual version of our model which integrates only the expert-based document scoring presented in section 3.2.
- *W/oDoL* is our collaborative-based model detailed in section 3 by excluding the division of labor principle.
- *W/oEM* is our collaborative-based model by excluding the expert-based document allocation, detailed in section 3.2.
- *OurModel* is our collaborative-based model detailed in section 3.

4.2 Results

Here, we present the results obtained throughout our experimental evaluation. We adopt a learning-testing method through a two-cross validation strategy in order to tune the retrieval model parameters and then test its effectiveness. For this purpose, we randomly split the 277 search sessions into two equivalent sets, noted Q_A and Q_B. In what follows, we detail the parameter tuning, for both baselines and our model, and the retrieval effectiveness results.

Parameter Tuning. First, we tune the weighting parameter γ in Equation 1, used for the diversification of the search results using the MMR score [4]. The diversity criteria considered for building the subset of documents $\mathcal{D}^*$ of size $n = 1000$ is inversely proportional to the value of γ. For both subsets of queries Q_A and Q_B, the retrieval effectiveness is optimal for a γ value equal to 1.

Second, we estimate the optimal number of topics used for the topical distribution of documents. For each multi-faceted query, we perform the LDA algorithm with a number of topics T from 20 to 200 with a step of 20. The number of topics is tuned using the likelihood, presented in Equation 2. We retain $T = 200$ as the optimal value for both query sets Q_A and Q_B. Considering that previous work [8] found that the number of query subtopics is lower than 10 for most of

the queries, we add a supplementary layer for modeling query facets by considering the top f facets among the 200 ones of the topical distribution of the query valued by the probability $\theta_{Q|t}$. Documents and experts are therefore represented by extracting these f facets from their topical vector.

Third, we jointly tune the number f of the top facets and the parameter λ which combines a similarity score based on the topical modeling and the BM25 algorithm within the document scoring step in equation 6. In order to tune $\lambda \in [0;1]$ and $f \in [1;10]$, we rank, for each TREC topic $Q \in \theta$, the top 1000 diversified documents $D_i \in \mathcal{D}^*$ according to their topical distribution. We retain the value $\lambda = 0.6$ and $f = 5$ which maximizes the retrieval effectiveness for both sets of queries Q_A and Q_B.

Model effectiveness evaluation. Table 2 compares the obtained results using our collaborative model with those obtained by the three scenarios, detailed in section 4.1. The reported precisions show that our model generally overpasses the baseline models without the EM step, namely *W/oEM* and *W/oEMDoL*. More particularly, the improvement is significant compared to the scenario *W/oEM* for the query testing set Q_B with a value of $+20.61\%$. This highlights the importance of allocating documents to experts, as detailed in section 3.2, considering the whole set of experts in contrast to a ranking algorithm only based on a personalized document scoring towards a particular expert, as detailed in equation 4. We notice that the scenario performed without division of labor, namely *W/oDoL*, provides better results than our model considering the precision measures. On a first point of view, we could conclude that division of labor is inappropriate for solving collaboratively a multi-faceted information need. However, this statement is counterbalanced by the analysis of coverage-based measures, detailed in section 4.1. In fact, we notice that our model provides in most of the cases higher values of the coverage-based measures with several significant improvements around $+17.69\%$ for the $Cov@30$ metric and up to $+55.40\%$ for the $RelCov@30$ one. These results show that our model ensures both diversity of the displayed documents throughout the coverage measure $Cov@30$ and relevance of these diversified documents within the displayed lists by means of the $RelCov@30$ measure. This contrast between precision measures and coverage-based ones is explained by the fact that the latter takes into account the residual relevance feedback within the displayed document lists whereas the second one does not consider document redundancy between successive displayed lists. In summary, the results show that our collaborative ranking model is more appropriate for satisfying a multi-faceted information need compared to an individual one, namely through scenario *W/oEMDoL*; moreover, as reported by coverage-based measures we confirm that our model favors the topical diversity of the search results thanks to the division of labor principle.

Complementary Analysis. We perform further analysis to show the impact of two parameters, namely the relevance agreement level and the group size, on the model effectiveness. Figure 1 plots the retrieval effectiveness variation within these two parameters. We can see that our model curve overpasses baseline curves for both parameters. The decreasing trend of curves in Figure 1(a)

Learning set → Testing set	Scenario	$Cov@30$	$\%Ch$	$RelCov@30$	$\%Ch$	$P@30$	$\%Ch$
$Q_B \rightarrow Q_A$	W/oDoL	0.486	+12.01	0.131	**+19.58***	0.388	-0.30
	W/oEM	0.571	-4.60	0.124	**+26.73***	0.309	**+20.60***
	W/oEMDoL	0.454	**+17.56*****	0.101	**+55.40****	0.369	+1.13
	OurModel	0.545		**0.157**		0.372	
$Q_A \rightarrow Q_B$	W/oDoL	0.481	**+10.78*****	0.147	+4.29	0.414	-2.12***
	W/oEM	0.526	+1.38	0.134	+14.77	0.379	+7.01
	W/oEMDoL	0.453	**+17.69*****	0.131	+17.16	0.394	+ 2.73
	OurModel	**0.533**		**0.156**		0.406	

Table 2: Comparative analysis of the retrieval effectiveness of our model - %Ch: our model improvement. Student test significance *: $0.01 < t \leq 0.05$; **: $0.001 < t < 0.01$; ***: $t < 0.001$

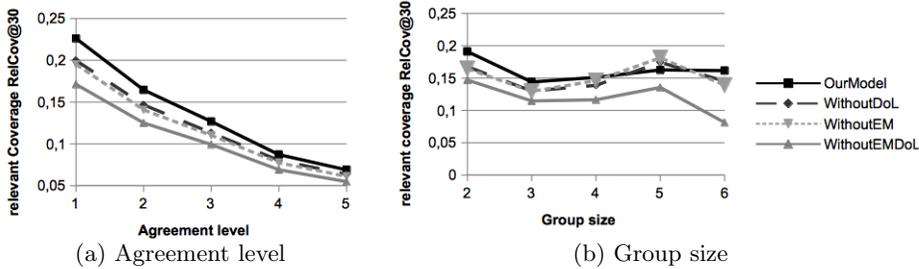(a) Agreement level   (b) Group size

Fig. 1: Impact of agreement level (a) and group size (b) on the retrieval effectiveness

can be explained by the fact that higher is the agreement level, fewer documents are assessed as relevant within the search session and this favors the search failure regardless of the retrieval model. From Figure 1(b), we notice that the curve of our model is generally stable even with the increasing size of the collaborator group. These statements confirm that the retrieval model improvements are stable within different configurations of collaborative search settings.

5 Conclusion and Future Work

In this paper, we propose a collaborative ranking model for satisfying a multi-faceted information need considering a group of experts. We propose an iterative relevance-feedback process for automatically updating expert's document list by means of the Expectation-Maximization learning method for collaboratively ranking documents. Our model was evaluated using a collaboration simulation-based framework and has shown effective results. Future work will focus on the design of other formal methods to emphasize division of labor and the modeling of user profile through his behavior in addition to his relevance feedback.

References

1. J. Allan and H. Raghavan. Using part-of-speech patterns to reduce query ambiguity. In *SIGIR'07*, pages 307–314, 2002.
2. A. Ashkan, C. L. Clarke, E. Agichtein, and Q. Guo. Classifying and characterizing query intent. In *ECIR'09*, pages 578–586, 2009.
3. D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, 2003.
4. J. Carbonell and J. Goldstein. The use of mmr, diversity-based reranking for reordering documents and producing summaries. In *SIGIR'98*, pages 335–336, 1998.
5. W. Dakka, R. Dayal, and P. G. Ipeirotis. Automatic discovery of useful facet terms. In *SIGIR'06*, pages 18–22, 2006.
6. S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman. Indexing by latent semantic analysis. *JASIST*, 41:391–407, 1990.
7. A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38, 1977.
8. R. Deveaud, E. SanJuan, and P. Bellot. Unsupervised latent concept modeling to identify query facets. In *RIAO'13*, 2013.
9. Z. Dou, S. Hu, Y. Luo, R. Song, and J.-R. Wen. Finding dimensions for queries. In *CIKM'11*, pages 1311–1320, 2011.
10. C. Foley and A. F. Smeaton. Synchronous collaborative information retrieval: Techniques and evaluation. In *ECIR'09*, pages 42–53, 2009.
11. M. Harvey, I. Ruthven, and M. Carman. Ranking social bookmarks using topic models. In *CIKM'10*, pages 1401–1404, 2010.
12. J. He, E. Meij, and M. de Rijke. Result diversification based on query-specific cluster ranking. *JASIST*, 62(3):550–571, 2011.
13. Y. Hu, Y. Qian, H. Li, D. Jiang, J. Pei, and Q. Zheng. Mining query subtopics from search log data. In *SIGIR'12*, pages 305–314, 2012.
14. G. Kim. Relationship between index term specificity and relevance judgment. *Inf. Process. Manage.*, 42(5):1218–1229, 2006.
15. M. R. Morris and E. Horvitz. Searchtogether: an interface for collaborative web search. In *UIST'07*, pages 3–12, 2007.
16. M. R. Morris, J. Teevan, and S. Bush. Enhancing collaborative web search with personalization: groupization, smart splitting, and group hit-highlighting. In *CSCW'08*, pages 481–484, 2008.
17. P. Over. The trec interactive track: an annotated bibliography. *Inf. Process. Manage.*, 37(3):369–381, 2001.
18. J. Pickens, G. Golovchinsky, C. Shah, P. Qvarfordt, and M. Back. Algorithmic mediation for collaborative exploratory search. In *SIGIR'08*, pages 315–322, 2008.
19. C. Shah. *Collaborative Information Seeking: The Art and Science of Making The Whole Greater Than The Sum of All*. Springer, 2012.
20. C. Shah and R. González-Ibáñez. Evaluating the synergic effect of collaboration in information seeking. In *SIGIR'11*, pages 913–922, July 2011.
21. C. Shah, J. Pickens, and G. Golovchinsky. Role-based results redistribution for collaborative information retrieval. *Inf. Process. Manage.*, 46(6):773–781, 2010.
22. Q. Wang, Z. Cao, J. Xu, and H. Li. Group matrix factorization for scalable topic modeling. In *SIGIR'12*, pages 375–384, 2012.

23. R. W. White, S. T. Dumais, and J. Teevan. Characterizing the influence of domain expertise on web search behavior. In *WSDM'09*, pages 132–141, 2009.
24. X. Zhu, A. B. Goldberg, J. Van, and G. D. Andrzejewski. Improving diversity in ranking using absorbing random walks. In *SIGIR'09*, 2009.