MineRank: Leveraging Users' Latent Roles for Unsupervised Collaborative Information Retrieval

Laure Soulier*, Lynda Tamine** and Chirag Shah***

* *Sorbonne Universités, UPMC Univ Paris 06, UMR 7606, LIP6, F-75005, Paris,France*
** *Toulouse University UPS IRIT 118 route de Narbonne, 31062 Toulouse Cedex 9*
*** *School of Communication & Information (SC&I) Rutgers University 4 Huntington St, New Brunswick, NJ 08901, USA*

Email address: laure.soulier@lip6.fr*, tamine@irit.fr** and chirags@rutgers.edu*** (Laure Soulier*, Lynda Tamine** and Chirag Shah***)

MineRank: Leveraging Users' Latent Roles for Unsupervised Collaborative Information Retrieval

Laure Soulier*, Lynda Tamine** and Chirag Shah***

* *Sorbonne Universités, UPMC Univ Paris 06, UMR 7606, LIP6, F-75005, Paris,France*
** *Toulouse University UPS IRIT 118 route de Narbonne, 31062 Toulouse Cedex 9*
*** *School of Communication & Information (SC&I) Rutgers University 4 Huntington St, New Brunswick, NJ 08901, USA*

Abstract

Research on collaborative information retrieval (CIR) has shown the positive impact of collaboration on retrieval effectiveness in the case of complex or exploratory tasks. The synergic effect of accomplishing something greater than the sum of its individual components is reached through the gathering of collaborators' complementary skills. However, these approaches often lack the consideration that collaborators might refine their skills and actions throughout the search session, and that a flexible system mediation guided by collaborators' behaviors should dynamically optimize the search effectiveness. In this article, we propose a new unsupervised collaborative ranking algorithm which leverages collaborators' actions for (1) mining their latent roles in order to extract their complementary search behaviors; and (2) ranking documents with respect to the latent role of collaborators. Experiments using two user studies with respectively 25 and 10 pairs of collaborators demonstrate the benefit of such unsupervised method driven by collaborators' behaviors throughout the search session. Also, a qualitative analysis of the identified latent role is proposed to explain an over-learning noticed for one of the dataset.

Keywords: Collaborative information retrieval, unsupervised role mining,

Email address: `laure.soulier@lip6.fr`*, `tamine@irit.fr`** and `chirags@rutgers.edu`*** (Laure Soulier*, Lynda Tamine** and Chirag Shah***)

latent role, user study

1. Introduction

In the recent years, researchers have argued that in addition to creating better algorithms and systems for individualized search and retrieval, a substantial leap can be taken by incorporating collaborative aspects in Information Retrieval (IR) (Twidale et al., 1997), referred to as Collaborative Information Retrieval (CIR) (Fidel et al., 2000). However, simply allowing multiple people collaborate on a search task does not guarantee any advantages over a single searcher. For that, one needs to look deeper into the aspects of collaboration that make it successful and investigate how those aspects can be incorporated in a search setting. Many have found that when the collaborators bring a diverse set of skills to a project, they could achieve something more than what they could using their individual skills and contributions (Soulier et al., 2014). But how does one ensure the use of such diverse skills in search? One approach could be asking the searchers involved in a CIR project about the roles (e.g., query constructor, information assessor) they would like to play. However, these collaborators may either not know about such skills they have or be unable to specify any preferences. Therefore, one may need to mine their skills automatically through behavioral features. Recently, this approach (Soulier et al., 2014) has been proposed, aiming at dynamically identifying, through search features, the possible roles of collaborators according to a role taxonomy (Golovchinsky et al., 2009). However, the labeled roles are predefined regardless of the users, and accordingly, one lack of the proposed approach is that it does not ensure that the identified roles exactly fit with collaborators' search skills.

To tackle this gap, this current article presents a new approach - called *MineRank* - that mines in real time the unlabeled role that collaborators could play in a CIR context. The objective is to leverage the diversity of the collaborators' search skills in order to ensure the division of labor policy and to optimize the overall performance. Instead of following predefined labels or taxonomy of roles, *MineRank* is an unsupervised algorithm that (1) learns about the complementarity of collaborators' unlabeled roles in an unsupervised manner using various search behavior-related features for each individual involved, and (2) re-injects these unlabeled roles for

collaboratively ranking documents. The algorithm is used for various experiments and evaluated using retrieval effectiveness at various levels. The results show that the model is able to achieve synergic effect in CIR by learning the latent role of collaborators.

The remainder of the article is structured as follows. Section 2 presents the related work. In Section 3, we motivate our approach and introduce the problem definition. Section 4 focuses on our two-step unsupervised CIR model relying on the collaborators' unlabeled roles. The experimental evaluation and results are described in Section 5. Section 6 concludes the article.

2. Related Work

2.1. Collaborative Information Retrieval

Collaborative information retrieval (CIR) is defined as the search process involving multiple users solving a shared information need (Golovchinsky et al., 2009). Research has found that this setting is particularly beneficial in the case of complex or exploratory information tasks (Morris and Horvitz, 2007) in which an individual alone would suffer from insufficient knowledge or skills. Indeed, collaboration in search improves its retrieval effectiveness by providing the opportunity to gather complementary skills and/or knowledge in order to solve an information need as well as satisfying mutual benefits of collaborators through the synergic effect of the collaboration (Shah and González-Ibáñez, 2011b).

Collaboration between users is supported by three main principles: (1) avoiding redundancy between users' actions (division of labor), either at the document level (Foley and Smeaton, 2009) or the role level (Pickens et al., 2008; Shah et al., 2010); (2) favoring the information flow among users (sharing of knowledge), either implicitly by search inference (Foley and Smeaton, 2010) or explicitly by collaborative-based interfaces (Morris and Horvitz, 2007); and (3) informing users of other collaborators' actions (awareness) (Dourish and Bellotti, 1992). Supporting these three principles remains a challenge in CIR, often tackled with adapted interfaces, revisited IR techniques, or collaborative document ranking models (Joho et al., 2009).

In this article, we focus more specifically on the third aspect dealing with CIR models. In previous work, the focus has been on the mediation of collaborators' actions and complementarity skills in order to enhance the synergic

effect within collaboration towards the satisfaction of the shared information need (Shah and González-Ibáñez, 2011b). Henceforth, division of labor is a pivotal issue for coordinating collaborators in terms of search actions with respect to their complementary skills. Assigning roles is one way of tackling this challenge since roles give a structure to the search process (Kelly and Payne, 2013). Beyond simply considering collaborators as peers and focusing on inferring the global relevance of documents towards all collaborators (Foley and Smeaton, 2009) or personalizing document scores for each collaborator (Morris et al., 2008), several works (Pickens et al., 2008; Shah et al., 2010; Soulier et al., 2013) propose assigning asymmetric roles to users in order to optimize the collaborative search effectiveness. Golovchinsky et al. (2009) suggested these roles in a role taxonomy.

Pickens et al. (2008) proposed a pair of roles, namely Prospector-Miner that involved splitting a search task between the collaborators. The Prospector was responsible for formulating search request for ensuring search diversity, whereas the Miner was devoted to identifying highly relevant documents. Similarly, Shah et al. (2010) proposed a CIR model relying on the couple of Gatherer-Surveyor where the former's goal was to quickly scan search results and the latter focused on diversity. For these models, users' roles ensured a task-based division of labor.

Different from these works, Soulier et al. (2013) ensured the division of labor among collaborators by considering their domain expertise as the core evidence source of a collaborator's role aiming to solve a multi-faceted information need. For this purpose, the authors structured collaborators' actions by assigning documents to the most likely suited users as well as allowing them to simultaneously explore distinct document subsets.

A new role-based approach has been proposed in (Soulier et al., 2014) by considering that search behaviors of collaborators were dynamic and that their role might evolve throughout the search session. With this in mind, collaborators' predefined and labeled roles, namely Prospector-Miner and Gatherer/Surveyor, were identified in real time assuming a task-based division of labor policy based on their search behavior oppositions. Then, documents are ranked according to the CIR models associated to the mined roles (Pickens et al., 2008; Shah et al., 2010).

2.2. User Behavior Models for Document Retrieval

The user behavior modeling domain focuses on the understanding of the user model within the search session. On one hand, some work (Evans

and Chi, 2010; Yue et al., 2014) only focus on the user modeling in a high abstract level in order to build generative behavioral models. For instance, Yue et al. (2014) analyze temporal sequential data of collaborative search through a hidden Markov model. On the other hand, other research attempt to model user behavior and to re-inject them in a retrieval model in order to enhance the search effectiveness (Agichtein et al., 2006). In this last research domain, more close to our contribution, we distinguish three main lines of work based on feature-based document relevance prediction models (Agichtein et al., 2006; Radinsky et al., 2013), personalization approaches through users' preferences (Bennett et al., 2012; Teevan et al., 2005), or role extraction-based ranking models (Henderson et al., 2012; McCallum et al., 2007).

The first category of research dealing with prediction models analyzes several dimensions of user behaviors. In most of these works, a simplistic approach is usually followed that consists of integrating clickthrough data within the document scoring (Joachims, 2002) since this source of evidence expresses user's search behaviors. In addition, some authors (Agichtein et al., 2006) suggest a further abstraction level by proposing a robust user behavior model which takes into account the collective behaviors for reducing noise within an individual search session. Instead of smoothing individual behaviors with collective search logs, Radinsky et al. (2013) proposed another analysis dimension that refines individual search log through a temporal aspect within search behaviors for predicting queries and click frequencies. The user model relied on time-series and dynamically extracted topical trends re-injected within the ranking or the query auto-suggestion.

Beyond analyzing search behaviors for document ranking, another line of work (Heath and White, 2008; White and Dumais, 2009) exploits search behaviors for predicting search engine switching events. Findings of these works may be used for enhancing the retrieval effectiveness and coverage of the information need for discouraging switching activities. For instance, dealing with the personalization approach, user profiles might be extracted considering users' relevance feedback (Bennett et al., 2012; Leung et al., 2008; Soulier et al., 2013; Teevan et al., 2005). In an individual search setting, Bennett et al. (2012) proposed to combine short-term and long-term search behaviors for mining users' interests. They modeled a multi-feature profile based on search history, query characteristics, document topic and

users' search actions. In contrast, Leung et al. (2008) modeled users' profiles through a concept-based representation inferred from clickthrough data. The profile is then used to learn users' preferences with an SVM algorithm and personalizing their search results. Search personalization is also proposed in collaborative search settings (Morris et al., 2008; Soulier et al., 2013). For instance, Morris et al. (2008) integrated a personalized score (Teevan et al., 2005) within (1) a document smart-splitting over collaborators' rankings for retrieving individual rankings; and (2) a relevance summarization of relevance feedback for building the final document list capitalizing the collective relevance.

In the last category, previous works have proposed to model and/or mine users' roles from their search behaviors. In this context, contributions aim at either statistically identifying predefined roles (Golder and Donath, 2004; Kwak et al., 2010) or mining latent roles through probabilistic models (Henderson et al., 2012; McCallum et al., 2007). The first perspective relies on social network interactions for identifying labeled or predefined roles through a statistical analysis, such as "Celebrities" or "Ranters" (Golder and Donath, 2004), or the "Network Leaders" using a PageRank-like algorithm (Pal and Counts, 2011) or a clustering method (Kwak et al., 2010). The second perspective offers a formal way to identify latent roles through the analysis of users' interactions similarities/dissimilarities (Henderson et al., 2012; McCallum et al., 2007). For instance, Henderson et al. (2012) focused on the transformation of a feature-based multidimensional matrix for identifying the users' behavior model while McCallum et al. (2007) revised the LDA algorithm within a communication social network for mining the evolving roles of users according to message contents.

2.3. Research Objectives

From the literature review, one can infer that the key challenge in CIR relies on the difficulty of ranking documents in order to both satisfy individual and mutual goals with respect to the shared information need. Therefore, this challenge assumes that users are different and guided by complementary skills or knowledge (Sonnenwald, 1996). In order to consider users' differences, one possible way might be to assign different roles with respect to their skills. However, CIR models based on predefined roles (Pickens et al., 2008; Shah et al., 2010; Soulier et al., 2013) raise two main concerns (Soulier et al., 2014):

1. The role assignment assumes that users behave the same way throughout the session by assigning roles to users at the beginning of the search session.

2. A role might not particularly be in accordance with the users' intrinsic skills, and more particularly in which they are the most effective.

One solution is to derive users' roles from their differences and complementarities toward their interactions in order to exploit these roles within the ranking. For this purpose, two main approaches can be traced, which, unlike works focusing on user behavior models that mainly deal with intrinsic users' values (Agichtein et al., 2006; Bennett et al., 2012; Leung et al., 2008), consider users relatively to their peer in order to highlight how they are the most effective. The first one operates on a pool of predefined labeled roles, and consists of a dynamic role assignment monitored by a supervised learning based on features inferred from users' interactions handled in the search system (Soulier et al., 2014). Once predefined roles have been identified, the associated state of the art CIR models are used to solve the query. However, one limitation that could be raised from this work is that the labeled roles are predefined regardless of the users, restricting the likelihood of ensuring that the identified roles exactly fit with collaborators' search skills. Therefore, two main challenges could be raised: 1) might a user be assigned to the most likely role fitting with his/her behavior, even if it is not particularly in accordance with his/her skills?, and 2) what if users can be fitted to multiple roles?

The second approach, which we state in this article, dynamically characterizes latent roles, not labeled as a role belonging to a taxonomy, with an unsupervised learning method. More particularly, in contrast to well-know CIR models (Pickens et al., 2008; Shah et al., 2010) and in accordance to the limitations raised by Soulier et al. (2014), we approach here the problem of predefined roles, and propose to dynamically mine the unlabeled roles of collaborators throughout the search session in an unsupervised manner, and accordingly adapt the collaborative document ranking. As shown in Figure 1, unlabeled roles of both collaborators are mined at each time a user submits a query. Then, features modeling these unlabeled roles are re-injected within the collaborative document ranking model in order to provide a ranked list of documents displayed to this user. In order to

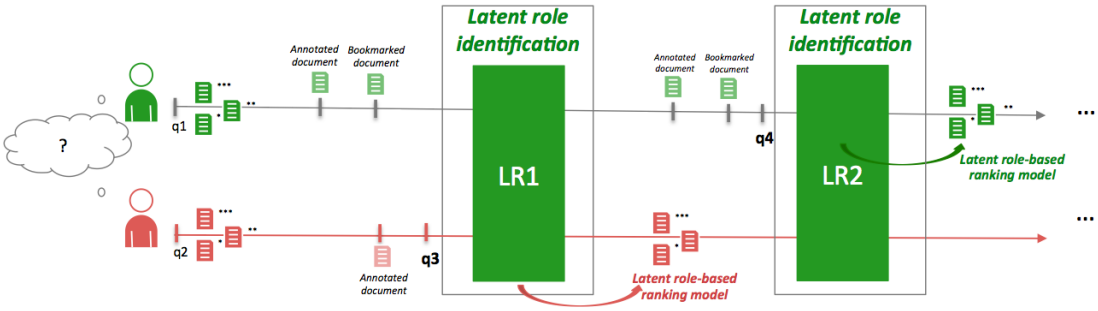


Figure 1: **Unsupervised latent role learning methodology.**

ensure our twofold objective of (1) mining unlabeled roles with respect to collaborators' behaviors and (2) collaboratively ranking documents, we rely on a feature set estimated at the document level. Therefore, our intuition is that if we consider documents receiving a relevance feedback as good indicator of users' search behaviors and preferences (Agichtein et al., 2006), those features estimated on the relevant document set would enable to (1) mine latent roles of collaborators and (2) re-inject the mined latent roles within a CIR model.

Therefore, we aim to address the following research questions in this article:

- **RQ1**: How to infer collaborators' unlabeled roles through the differences and complementarities in their behaviors?
- **RQ2**: How to leverage these unlabeled roles for collaboratively ranking documents with respect to the shared information need?

We introduce the concept of latent role that captures in real time the roles of collaborators according to the complementarity of their search skills, without any assumptions of predefined roles labeled or belonging to a role taxonomy (Golovchinsky et al., 2009). More particularly, guided by the division of labor policy, the users' latent roles leverage the skills in which collaborators are different, the most complementary and the most effective for enhancing the retrieval effectiveness of the search session. Also, we assume that search skills of collaborators might be inferred within the persistence of their search behaviors, since collaborators might have noisy search actions

which might be due to the task, the topic, the interface design or collaborators' engagement within the task. With this in mind, we are aware that this concept depends on two main hypothesis:

- H1: The search session is synchronous, enabling users to coordinate their actions and to exhibit their skills at the same time.
- H2: Users are both engaged in the information need solving, avoiding as best as possible noisy search actions or inactivity behaviors.

3. The model

We present here our model based on the latent role of collaborators assuming that users might refine their search strategies and behaviors throughout the session while they interact with their collaborators or assess search results.

Our model, called *MineRank*, considers search features modeling collaborators' behaviors and aims at ranking documents in a collaborative manner at each query submission by leveraging collaborators' search skill complementarities. For convenience, we call an iteration associated to timestamp t_l, the time-window beginning at each time user u submits query q and ending while document list $\mathcal{D}_u^{t_l}$ is retrieved to user u. More particularly, an iteration of *MineRank* relies on two main steps illustrated in Figure 2: (1) learning across time the most discriminant feature set which maximizes the differences between users' behaviors in search results in order to dynamically mine the latent role of collaborators (section 3.2), and (2) re-injecting latent roles for collaboratively ranking documents. For this purpose, we aim at predicting, through a learning model, the document relevance towards collaborators by taking into account their latent roles (section 3.3).

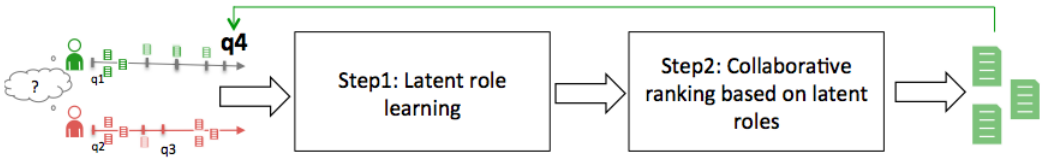


Figure 2: **Minerank methodology for an iteration.**

3.1. Notations

We consider a synchronous collaborative search session S involving a pair of users u_1 and u_2 for solving a shared information need I during a time interval T. Each user u browses separately and formulates his/her queries for accessing their respective document result sets. As shown in Figure 1, users have the possibility to perform different actions throughout the search session. Beyond formulating query, they interact with the retrieved documents by visiting their content, annotating web pages with comments, bookmarking documents or snipping pieces of information. Therefore, users' actions might be characterized by search behavior-based features, noted $F = \{f_1, \ldots, f_k, \ldots, f_n\}$. The latter expresses the set of n features captured during the search session, detailed in Table 1. These features, based on the literature (Agichtein et al., 2006), include two types of features:

- *Submitted query features* that capture collaborators' search experience with respect to the query topic. For instance, we integrate features based on the overlap between the query and pieces of document (title, content, annotations/snippets generated by a user).
- *Selected page features* that capture collaborators' browsing behaviors in the search session in order to highlight time spent on webpages/on a specific domain as well as the specificity or readability of documents visited/annotated/snipped, and bookmarked by a given user.

We highlight that the feature set is slightly different from the one used in (Soulier et al., 2014) since the intuition of the proposed model here is to reinject the behavioral features within the collaborative document ranking. Furthermore, the feature set can be extended with no impact on the model.

Following Soulier et al. (2014), we represent a temporal feature-based user's behavior matrix $S_u^{(t_l)} \in \mathcal{R}^{t_l \times n}$, where t_l is the timestamp. Each element $S_u^{(t_l)}(t_j, f_k)$ represents the average value of feature f_k for user u aggregated over the set $\mathcal{D}(u_j)^{(t_j)}$ of documents visited/annotated/snipped/bookmarked during the time interval $[0 \ldots t_j]$. Assuming that users' search behaviors might be refined throughout the session, the temporal modeling enables the characterization of the overall behavior of the user at timestamp t_l avoiding the bias induced by noisy search actions.

According to Soulier et al. (2014), users' search skill difference toward a particular search feature $f_k \in F$ is referred to as $\Delta_{1,2}^{(t_l)}(f_k)$, where

Table 1: **Search behavior features**

	Feature	Description
Query features	TitleOverlap (TiO)	Fraction of shared words between query and page title
	TextOverlap (TeO)	Fraction of shared words between query and page content
	AnnotationOverlap (AO)	Fraction of shared words between query and page annotation
	SnippetOverlap (SO)	Fraction of shared words between query and snippet of the page
	VisitedPosition (VP)	Position of the URL in visited page order for the query
Page features	TimeQueryToPage (TQTP)	Time between the query submission and the visit of the page
	TimeOnPage (TOP)	Page dwell time
	TimeOnDomain (TOD)	Cumulative time for this domain
	Readability (Read)	Document content readability
	Specificity (Spec)	Document content specificity
	Rating (Ra)	Rate of this page

$\Delta_{1,2}^{(t_l)}(f_k) = S_{u_1}^{(t_l)}(f_k) - S_{u_2}^{(t_l)}(f_k)$. In addition, to ensure that both behaviors are found different in both users, and in order to identify those search behaviors each user is better than the other one, we would like to know in which of those search behaviors each user is better than the other one and highlight the skills in which he/she is the most effective with respect to his/her intrinsic skills as well as his/her collaborator's skills. With this in mind, correlations between collaborators' search feature differences were estimated, pair by pair, by adding the constraint that the difference between users for the implied features is significant, using the Kolmogorov-Smirnov test (p-value $p(\Delta_{1,2}^{(t_l)}(f_k))$). Therefore, complementarities and similarities between collaborators u_1 and u_2 with respect to their search behaviors are emphasized through a correlation matrix $C_{1,2}^{(t_l)} \in \mathcal{R}^{p \times p}$ in which each element $C_{1,2}^{(t_l)}(f_k, f_{k'})$ is estimated as (Soulier et al., 2014):

$$C_{1,2}^{(t_l)}(f_k, f_{k'}) = \begin{cases} \rho(\Delta_{1,2}^{(t_l)}(f_k), \Delta_{1,2}^{(t_l)}(f_{k'})) & if\ p(\Delta_{1,2}^{(t_l)}(f_k)) < \theta\ and\ p(\Delta_{1,2}^{(t_l)}(f_{k'})) < \theta \\ 0 & otherwise \end{cases} \tag{1}$$

As the goal is to focus on search behavior complementarities between unlabeled roles of collaborators u_1 and u_2, we assume that two features f_k and $f_{k'}$ behave similarly if the correlation $\rho(\Delta_{1,2}^{(t_l)}(f_k), \Delta_{1,2}^{(t_l)}(f_{k'}))$ of their difference is close to 1. The closer to -1 the correlation is, the more collaborators' skills towards features f_k and $f_{k'}$ are complementary. Focusing on users' difference $\Delta_{1,2}^{(t_l)}(f_k)$ towards search feature f_k is not enough since it does not ensure that collaborators' roles are complementary with respect to two search features: one user could be better for both features (Soulier et al., 2014), and, in this case, there is no need to leverage the other collaborator as a division of labor actor.

With this in mind, we introduce the concept of latent role based on the following hypothesis:

- H1: A latent role models the most significant similarities and complementarities between collaborators with respect to their search behaviors or skills throughout the search session in order to identify skills in which collaborators are the most effective.

- H2: Complementarities and similarities are respectively expressed by negative and positive correlations between search behavior features.

Therefore, at each timestamp, the latent role $LR_{1,2}^{(t_l)}$ highlights, for a pair of collaborators u_1 and u_2, their search skill differences and complementarities during the time period $[0..t_l]$; where search skills of a user are inferred from his temporal feature-based behavior matrix $S_u^{(t_l)}$ to highlight the persistence of unlabeled roles with respect to the task, the topic, the interface design or users' engagement within the task. Accordingly, the latent role $LR_{1,2}^{(t_l)}$ involves:

- A kernel $\mathcal{K}_{1,2}^{(t_l)}$ of a subset $F_{k1,2}^{(t_l)} \subset F$ of p behavioral features F, where p is automatically defined by the latent role mining algorithm (see Section 3.2). In other words, p expresses the number of the most significant features used to characterize the latent role according to hypothesis H1 and H2.

- A correlation matrix $C_{1,2}^{(t_l)} \in \mathcal{R}^{p \times p}$ which emphasizes complementarities and similarities between unlabeled roles of collaborators u_1 and u_2 with respect to their search behaviors.

3.2. Learning Users' Latent Roles in Collaborative Search

The underlying issue of the latent role mining consists of identifying the most discriminant features for characterizing collaborators' search behaviors which maximize, for a pair of collaborators, their complementarity. This leads us to propose a collaboration-oriented latent role mining approach based on a feature selection. The intuition behind our contribution is illustrated in Figure 3. The feature selection operates on the analysis of users' search behaviors. Once users have been identified as different towards search features, their complementary behaviors are modeled through a weighted network, for identifying the most important and discriminant features for characterizing collaborators' latent roles over the search session.

In what follows, we express first the optimization problem framework and the underlying assumptions. Then, we propose a solving method.

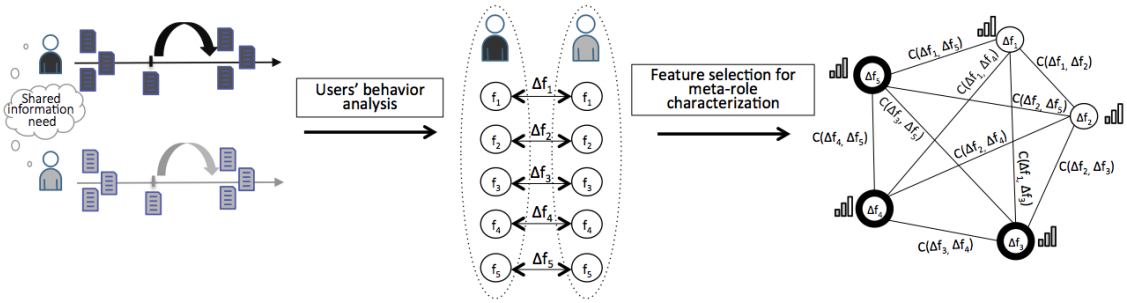


Figure 3: **Unsupervised latent role learning methodology.**

3.2.1. Latent Role Design

Inspired from work proposed by Geng et al. (2007) and adapted to our collaborative latent role mining, the feature selection consists in building the latent role kernel $\mathcal{K}_{1,2}^{(t_l)}$ by identifying the smallest subset $F_{k1,2}^{(t_l)}$ of p features (p is undefined) according to three assumptions:

- **A1**: the importance $\mathcal{R}ec_{1,2}^{(t_l)}(f_k)$ of features $f_k \in F_{k1,2}^{(t_l)}$ is dependent on their abilities to provide a good indicator in the document assignment to users within the collaborative document ranking. We assume that a CIR model might support the division of labor, and that a document might be assigned to the most likely suited collaborator. As proposed

by Shah et al. (2010), we formalize this principle through a document classification relying on relevance feedback collected throughout the search session until timestamp t_l where each document cluster represents documents allocated to one of the collaborators. With this goal, we propose to cluster, using a 2-means classification, the set $\mathcal{D}^{(t_l)}$ of selected documents (through annotations/snippets/bookmarks) by both collaborators until timestamp t_l according to the value of feature f_k. The cluster with the highest centroid is assigned to the collaborator u_j with the highest value $S_{u_1}^{(t_l)}(f_k)$ whereas the other cluster is assigned to the other collaborator. We measure the quality of the classification based on feature f_k towards each collaborator u_1 and u_2 using the recall measure $\mathcal{R}ec_{1,2}^{(t_l)}(f_k)$:

$$\mathcal{R}ec_{1,2}^{(t_l)}(f_k) = \frac{TP_{f_k}^{(t_l)}}{TP_{f_k}^{(t_l)} + FN_{f_k}^{(t_l)}} \tag{2}$$

where $TP_{f_k}^{(t_l)}$ is the number of documents assigned to the cluster associated to the user who selected those documents using the 2-means classification based on feature f_k. For instance, if u_1 selected document d_1 before timestamp t_l, we consider a "True Positive" action if the classification algorithm attributes document d_1 to user u_1. Accordingly, $TP_{f_k}^{(t_l)}$ is incremented of 1. Inversely, $FN_{f_k}^{(t_l)}$ expresses the number of documents not assigned to the cluster associated to the user who selected those documents. For instance, if document d_1 is attributed to cluster of user u_2, $FN_{f_k}^{(t_l)}$ is incremented of 1.

- **A2**: the redundancy between features might be avoided in order to consider only the most discriminant ones for characterizing latent roles through complementary search behaviors among users, modeled using feature correlations $C_{1,2}^{(t_l)}(f_k, f_{k'})$. We investigate here how to identify the most discriminant features for characterizing users' roles, and more particularly features highlighting complementary search behaviors among users. The main assumption is to identify for which skills collaborators are the most suited with respect to their collaborators for solving the shared information need. We used the correlation $C_{1,2}^{(t_l)}(f_k, f_{k'})$ between collaborators' differences $\Delta_{1,2}^{(t_l)}(f_k)$ and $\Delta_{1,2}^{(t_l)}(f_{k'})$ towards a pair of search behavior features f_k and $f_{k'}$.

- **A3**: the feature selection must maximize the importance $\mathcal{R}ec_{1,2}^{(t_l)}(f_k)$ of the selected features $f_k \in F_{k1,2}^{(t_l)}$ within the collaborative document ranking and minimize the redundancy $C_{1,2}^{(t_l)}(f_k, f_{k'})$ between the pairwise selected features. Thus, we formalize the feature selection algorithm as the following optimization problem:

$$\begin{array}{rc} max_\alpha & \sum_{k=1}^{n} \mathcal{R}ec_{1,2}^{(t_l)}(f_k) \cdot \alpha_k \\ min_\alpha & \sum_{k=1}^{n} \sum_{k'=1}^{n} C_{1,2}^{(t_l)}(f_k, f_{k'}) \cdot \alpha_k \cdot \alpha_{k'} \\ subject\ to & \alpha_k = \{0,1\};\ k = 1, ..., n \\ and & \sum_{k=1}^{n} \alpha_k = p \end{array} \tag{3}$$

where α is the vector of size n where each element α_k is a boolean indicator specifying if feature f_k is included in the feature subset $F_{k1,2}^{(t_l)}$ at timestamp t_l.

This optimization problem with multi-objectives might be transformed as a unique objective optimization problem by linearly combining the both optimization functions.

$$\begin{array}{rc} max_\alpha \sum_{k=1}^{n} \mathcal{R}ec_{1,2}^{(t_l)}(f_k) & \cdot\alpha_k - \gamma(\sum_{k=1}^{n} \sum_{k'=1}^{n} C_{1,2}^{(t_l)}(f_k, f_{k'}) \cdot \alpha_k \cdot \alpha_{k'}) \\ subject\ to & \alpha_k = \{0,1\};\ k = 1, ..., n \\ and & \sum_{k=1}^{n} \alpha_k = p \end{array} \tag{4}$$

where γ is a decay parameter expressing the level of behavior complementarity taken into account in the latent role mining algorithm. This parameter is fixed over the session since we hypothesize that the ratio between the feature importance and complementarity does not depend on the collaborators' current latent roles at timestamp t_l.

3.2.2. Latent Role Optimization

Our optimization problem defined in Equation 4 might be resolved by undertaking all the possible feature combinations of size p, where $p = 2, \ldots, n$. Although optimal, this method is time-consuming with a complexity of up to $O(\sum_{p=1}^{n} C_n^p)$.

We propose, here, a graph-based resolution algorithm attempting to identify the best feature subset which may provide a locally optimal solution but

more practically applicable with complexity which could reach a maximum of $O(3^{\frac{n}{3}})$. The main objective is to extract the smallest feature node set which enhances the importance of the set of retained features while maximizing the differences of collaborators within their search behaviors. In this context, we represent features through a collaboration-based graph $G_{1,2}^{(t_l)}$ modeling search behaviors of collaborators u_1 and u_2 at timestamp t_l. The graph $G_{1,2}^{(t_l)} = (A_{1,2}^{(t_l)}, C_{1,2}^{(t_l)})$, illustrated in Figure A.10, involves nodes $A_{1,2}^{(t_l)}$ which represent each feature $f_k \in F$, weighted by an importance measure $\mathcal{R}ec_{1,2}^{(t_l)}(f_k)$ within the collaborative document ranking, and undirected weighted edges $C_{1,2}^{(t_l)} : \mathcal{R}^{F \times F}$ which represent collaborators' search behavior similarities or complementarities by considering the correlation $C_{1,2}^{(t_l)}(f_k, f_{k'})$ between differences of pairwise features f_k and $f_{k'}$.

The used notations are detailed in Table 2. In what follows, we describe the algorithm, called *Coll-Clique*, for solving the optimization problem (Algorithms 1 and 2). Please, note that an illustration of our algorithm is presented in Appendix A.

Table 2: **Notations**

Notation	Description
C	The feature graph representing the growing clique
P	The evolving candidate graph
K	The maximum clique satisfying the optimization problem
$Nbhd(C)$	The function that returns in a decreasing order the neighboring features of all features belonging to C with non null weight
$Nds(K)$	The function that returns all the features belonging to K

In order to solve the optimization problem, we extended the maximum clique algorithm (Carraghan and Pardalos, 1990) in order to fit with our feature selection problem in a collaborative context. Our intuition is that a weighted graph is complete since it models search behavior complementarities through correlations between pairwise features. The *Coll-Clique* algorithm, rather than focusing on a node level for identifying the biggest complete subgraph (Carraghan and Pardalos, 1990), also called the maximum clique, aims at extracting the subgraph which maximizes the node weights, namely the search feature importances (assumption A1), and minimizes the relationship weights between nodes, namely pairwise search behavior correlations for both collaborators (assumption A2).

Algorithm 1: Main

Data: $G_{1,2}^{(t_l)} = (A_{1,2}^{(t_l)}, C_{1,2}^{(t_l)})$, γ
Result: $F_{sel}^{(t_l)}$
begin
$C = \{\}$
$K = \{\}$
$P = G_{1,2}^{(t_l)}$
$K = Coll - Clique(C, P, \gamma, K)$
$F_{sel}^{(t_l)} = Nds(K)$
Return $F_{sel}^{(t_l)}$

As shown in Algorithm 1, *Coll-Clique* relies on two feature graphs:

- The growing clique C, candidate to be the maximum clique K.
- The feature graph P, which includes candidate features to be added to the growing clique C. Nodes in P are obtained through the function $Nds(P)$.

Initially, C is empty and P is the graph including all the features. The algorithm, as shown in Algorithm 2, recursively increments the growing clique C using features f_h involved within graph P built upon the function $Nbhd(C)$ which creates a new candidate feature graph P only retrieving in a decreasing order features characterized by a positive depreciated weight. This operation is noted $C \oplus f_h$. At each recursion, the weight $\mathcal{R}ec_{1,2}^{(t_l)}(f_{k'})$ of the other remaining features $f_{k'}$ are depreciated by the correlation $C_{1,2}^{(t_l)}(f_j, f_{k'})$ with respect to the last selected feature f_j.

Let us denote $W(K)$ as the sum of feature weights within the maximum clique K. We assume that this sum refers to the indicator we would like to maximize (Equation 4) since the feature weight (importance $\mathcal{R}ec_{1,2}^{(t_l)}(f_{k'})$) is recursively depreciated with respect to the adjacent edge weight (correlation $C_{1,2}^{(t_l)}(f_j, f_{k'})$). If the weight $W(C) + W(P)$ of features within C and P is lower than the weight $W(K)$ within K identified until the current iteration, there is no way to build a clique from C by adding features from P with a higher weight than the weight of features in K. Finally, the set of selected

Algorithm 2: Coll-Clique

```
Data: C, P, γ, K
Result: K
begin
    forall the f_j ∈ P do
        W(C) = Σ_{f_k ∈ Nds(C)} Rec_{1,2}^{(t_l)}(f_k)
        W(P) = Σ_{f_k ∈ Nds(P)} Rec_{1,2}^{(t_l)}(f_k)
        W(K) = Σ_{f_k ∈ Nds(K)} Rec_{1,2}^{(t_l)}(f_k)
        if (W(C) + W(P) ≤ W(K)) then
            /* Return the maximum clique */
            Return K
        /* Increment the growing clique C */
        C = C ⊕ f_j
        /* Depreciate node weights */
        forall the f_k' ∈ P do
            Rec_{1,2}^{(t_l)}(f_k') = Rec_{1,2}^{(t_l)}(f_k') − C_{1,2}^{(t_l)}(f_j, f_k') * 2γ
        /* Build the candidate node set */
        P' = Nbhd(C)
        if (P' = {} and W(C) > W(K)) then
            /* Save the local optimum */
            K = C
        if P' ≠ {} then
            /* Launch new recursion */
            Coll-Clique(C,P',γ, K)
        /* Remove node for new recursion */
        C = C\f_j
        P = P\f_j
```

features $F_{k1,2}^{(t_l)}$ of size p inferred from the p nodes within the maximum clique K builds the latent role kernel $\mathcal{K}_{1,2}^{(t_l)}$.

We highlight that our algorithm strictly follows the framework of the maximum clique extraction algorithm proposed by Carraghan and Pardalos

(1990) which is the initial version of the branch-and-bound algorithm category, well-known to ensure the guarantee of the optimal solution (Wu and Hao, 2015). We add one heuristic in order to consider a weighted graph aiming at solving an optimization problem, initially proposed by Geng et al. (2007) and adapted to our problem. Therefore, given that the candidate clique C in incremented by positively-weighted nodes in a decreasing order, one could resume that the equation $(W(C) + W(P) \leq W(K))$ aims at maximazing the weight of the maximum clique K where its weight could be estimated as follows:

$$W(K) = \sum_{k=1}^{|K|} \mathcal{R}ec_{1,2}^{(t_l)}(f_k) - \gamma(\sum_{k=1}^{|K|} \sum_{k' \neq k; k'=1}^{|K|} C_{1,2}^{(t_l)}(f_k, f_{k'})) \tag{5}$$

The first part of the equation refers to the initial weight of nodes f_k in the initial graph $\mathcal{G}$ while the second part expresses the depreciation of the node weight with respect to the nodes belonging to the maximum clique. Therefore, maximizing the weight of the maximum clique K is equivalent to solving the optimization problem presented in Equation 4.

3.3. Latent Role-based Collaborative Document Ranking

In this section, we re-inject the latent role kernel identified in the previous section in order to collaboratively rank documents to users. The idea is to use the most discriminant features, characterizing search behavior complementarities of both collaborators, for first assigning documents to the most likely suited collaborator, in order to ensure the division of labor, and then ranking the documents assigned to each user. For this purpose, we use a classifier learning algorithm which operates on the document representation restricted to features implied within the latent role $M_{1,2}^{(t_l)} \in \mathcal{R}^{m \times m}$ at timestamp t_l. We highlight that only the document list $\mathcal{D}_u^{t_l}$ associated to the class of the user u who submitted query q is displayed. Indeed, as explained in Figure 1, we only attempt to satisfy the information need of the user who submitted the query. We assume that the other collaborator u' might not be interested in query q and is already examining a document list $\mathcal{D}_{u'}^{t'_l}$ retrieved with respect to a previously submitted query at timestamp $t'_l < t_l$. We choose to use the Logistic Regression as the classifier learning algorithm which, as usual, runs into two stages illustrated in Figure 4:

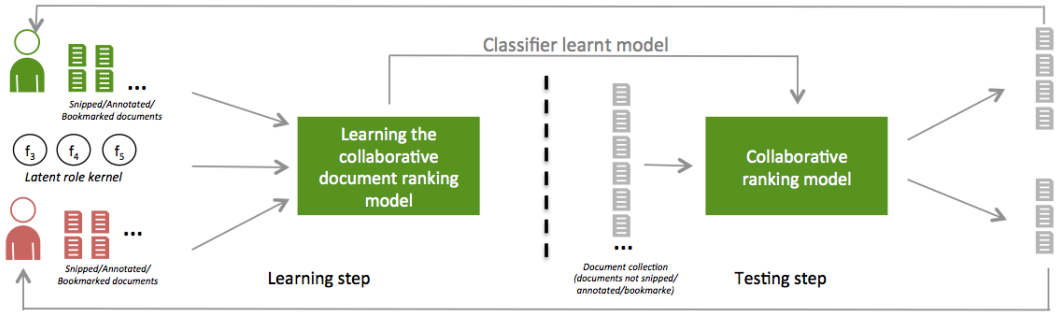


Figure 4: **Overview of the collaborative document ranking using latent role of collaborators**

- Stage 1. The *learning step* considers the set $\mathcal{D}^{(t_l)}$ of snipped/bookmarked/annotated documents by either collaborator u_1 or u_2 before timestamp t_l. Documents selected by both collaborators are removed from this set, since they are not discriminant for the collaborative-based document allocation to collaborators. Each document $d_i \in \mathcal{D}^{(t_l)}$ is modeled by a feature vector $x_i^{(t_l)} \in \mathcal{R}^m$, estimated according to value of the feature $f_k \in \mathcal{K}_{1,2}^{(t_l)}$ for document d_i with respect to collaborators' actions and timestamp t_h of its assessment, with $t_h \leq t_l$. Document d_i also receives a classification variable $c_i^{(t_l)} \in \{0;1\}$ where values 0 and 1 express the class of collaborators, respectively u_1 and u_2, who have selected this document.

  The objective of the document ranking learning function is to identify the predictor weight vector $\beta_j^{(t_l)} \in \mathcal{R}^m$ to estimate the probability of allocating documents to collaborator $u_j \in \{u_1, u_2\}$. The logistic regression aims at maximizing the likelihood l detailed in Equation 6 which relies on the logit function formalized in Equation 7. The latter models the probability $P_j(x_i^{(t_l)})$ for document d_i belonging to user class $c_i \in \{0;1\}$ with respect to feature vector $x_i^{(t_l)}$.

$$max_{\beta_j^{(t_l)}, \beta_j^{(t_l)}} \quad \sum_{d_i \in \mathcal{D}^{(t_l)}} (c_i \cdot \ln(P_j(x_i^{(t_l)}))) + (1 - c_i) \ln(1 - P_j(x_i^{(t_l)})) \tag{6}$$

$$where\ P_j(x_i^{(t_l)}) = \frac{\exp(x_i^{(t_l)} \cdot \beta_j^{(t_l)})}{1 + \exp(X_j^{(t_l)} \cdot \beta_j^{(t_l)})} \tag{7}$$

- Stage 2. The *testing step* considers the set $\mathcal{D}_{nsel}^{(t_l)}$ of documents not selected by both collaborators u_1 and u_2 before timestamp t_l. The feature vector $x_i^{(t_l)}$ is estimated according to feature average values of document d_i with respect to the search logs collected before timestamp t_l, not necessarily by both collaborators. Indeed, there is no available value for action-based features, such as AnnotationOverlap, for the pair of collaborators considering that the document has not been collected by the pair of users. The fitted model learnt through the logistic regression algorithm estimates the probability $P_j(x_i^{(t_l)})$ of assigning document $d_{i'} \in \mathcal{D}_{nsel}^{(t_l)}$ to the collaborator class c_i with respect to the predictor weight $\beta_j^{(t_l)}$. Document $d_{i'}$ is allocated to the collaborator class c_j with the highest probability $P_j(x_i^{(t_l)})$; $\forall j \in \{0, 1\}$, which is also used for ranking documents within the collaborator class.

Moreover, we add a supplementary layer of division of labor by ensuring that result lists $\mathcal{D}_{u'}^{t_l}$ and $\mathcal{D}_{u'}^{t'_l}$ simultaneously displayed (even if retrieved at different timestamps t_l and t'_l) to collaborators u and u' include distinct documents.

4. Experimental Evaluation

We performed an experimental evaluation investigating the impact of mining latent roles of collaborators on the retrieval effectiveness of a collaborative document ranking model. The hypothesis which guides our investigation are the following ones:

1. A CIR model should fit with collaborators' complementary skills in which they are the most effective with respect to collaborative setting taking into account their whole behaviors regardless of their skills or predefined roles.

2. A CIR model should achieve a greater effectiveness than users working separately.

3. A CIR model should dynamically tune collaborators' roles in an unsupervised manner in the search session instead of assigning roles regardless of their skills.

In what follows, we describe the experimental protocol and present the obtained results.

4.1. Protocol Design

4.1.1. User studies

Since no well-established benchmark exists in the CIR domain, we used search logs collected from two different collaborative-based user studies *US1* and *US2* supported by a collaborative search system (Shah and González-Ibáñez, 2011a) based on user mediation. The system allows users to browse the web and submit queries on independent search engines, mainly Google. The system includes a toolbar and a sidebar providing a functionality to users to interact with their peers through an instant messaging system as well as bookmarking, annotating and snipping web pages. Moreover, the sidebar ensures awareness by displaying to a user what he/she, as well as his/her collaborator, have bookmarked/snipped/annotated during the session. In addition, the system tracks collaborators' activities and records their search logs, such as visited pages, submitted queries and relevance feedback, all over the session. An overview of the used system is illustrated in Figure 5. We outline that this system ensure the awareness paradigm since the sidebar allows collaborators to be aware of other relevance feedback.

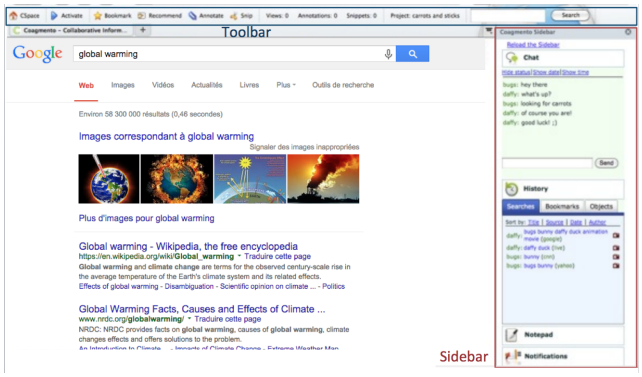


Figure 5: **Coagmento system**

The user studies, *US1* and *US2*, involved respectively 25 non native and 10 native English user pairs (a total of 70 people) who were recruited from university campuses and received compensation for their involvement within the experiments ($20 per person, with an additional $50 for the three best performing groups). Accordingly, these participants performed the task of exploratory search problem within a 30 minutes in a co-located setup in their mother tongue. During the task the collaborators interacted with each other in order to identify as many relevant documents as possible. Their interaction

Table 3: **Collaborative tasks in user studies US1 and US2**

Tasks	Guidelines
US1	*The mayor of your countryside village must choose between building a huge industrial complex or developing a nature reserve for animal conservation. As forest preservationists, you must raise awareness about the possibility of wildlife extinction surrounding such an industrial complex. Yet, before warning all citizens, including the mayor, you must do extensive research and collect all the facts about the matter. Your objective is to create a claim report together, outlining all the possible outcomes for wildlife should the industrial complex be built. Your focus is on wildlife extinction. You must investigate the animal species involved, the efforts done by other countries and the association worldwide to protect them and the reasons we, as humans, must protect our environment in order to survive. You must identify all relevant documents, facts, and pieces of information by using bookmarks, annotations, or saving snippets. If one document discusses several pieces of useful information, you must save each piece separately using snippets. Please assume that this research task is preliminary to your writing, enabling you to provide all relevant information to support your claims in your report.*
US2	*A leading newspaper agency has hired your team to create a comprehensive report on the causes, effects, and consequences of the climate change taking place due to global warming. As a part of your contract, you are required to collect all the relevant information from any available online sources that you can find. To prepare this report, search and visit any website that you want and look for specific aspects as given in the guideline below. As you find useful information, highlight and save relevant snippets. Later, you can use these snippets to compile your report, no longer than 200 lines. Your report on this topic should address the following: Description about global warming, scientific evidence about global warming affecting climate change, causes of global warming, consequences of global warming causing climate change, measures that different countries around the globe have taken over the years to address this issue including recent advancements. Also describe different view points people have about global warming (specify at least three different view points you find) and relate those to the the aspects controversies on this topic.*

behaviors generally involve discussions about their search strategies or link sharing exchange, always through the chat system. For user study *US2*, participants also had to write a report using web pages saved throughout the search session, letting them less time to browse on the web. Topics of the tasks are respectively "tropical storms" and "global warming". Guidelines of these two tasks are expressed in Table 3.

Statistics of both user studies are shown in Table 4. An analysis of submitted queries by participants shows that queries are mainly a reformulation of topics since that each topic word often occurs in the queries. Indeed, among the 1174 submitted queries for *US1*, the terms "tropical" and "storm" are used 1077 and 1023 times, respectively whereas the terms "global" and "warming" are used 254 and 247 times over the 313 submitted queries in *US2*. During the task, participants examined respectively 91 and 73 webpages for *US1* and *US2*. We highlight that the number of submitted queries and visited pages by collaborative groups are higher for *US1*; this could be explained by the additional objective of participants in *US2* which consisted of writing a report. The latter left participants less time for browsing the web.

	US1	US2
Topic	Tropical storm	Global warming
Number of dyads	25	10
Total number of visited pages	4734	1935
Total number of bookmarked/rated pages	333	-
Total number of snipped pages	306	208
Total number of submitted queries	1174	313
Average number of terms by query	3.65	4.73

Table 4: Statistics of user studies US1 and US2

4.1.2. Data

Given the distinct languages of both user studies, we build two separate document indexes. Respectively, for each user study *US1* and *US2*, we aggregated the respective web pages seen by the whole participant set as well as the top 100 search engine result pages (SERPs) from Google for the submitted query set. We highlight that the SERPs were extracted later in order to avoid processing overload. Each web page was processed for extracting

<title> and <p> tags. In order to increase the size of document indexes, we carried out this protocol for other proprietary user studies performed in other collaborative settings, not considered for our experiments (Shah and González-Ibáñez, 2011a). In the end, the indexes included 24, 226 and 74, 844 documents for *US1* and *US2* respectively.

4.1.3. Evaluation protocol

In order to avoid a bias that could be involved by a re-ranking approach relying on features measuring the similarity of the document with respect to the query, we highlight that the collaborative ranking step is carried out on the whole document collection. With this in mind, we consider two versions of our model:

- **MineRank(q)**: our proposed unsupervised ranking model in which the latent role mining and the collaborative ranking steps are successively launched at each query submission.
- **MineRank(t)**: our proposed unsupervised model in which the collaborators' latent role mining (section 3.2) is performed at regular timestamps t similarly to Soulier et al. (2014) while the collaborative ranking step (section 3.3) is launched at each query submission by considering the latent role mined at the last timestamp t.

We highlight that, similarly to Soulier et al. (2014), the BM25 model is launched when no search skill differences between collaborators are detected.

For effectiveness comparison, we ran the following baselines at each query submission:

- *BM25*: the BM25 ranking model which refers to an individual setting. This setting simulates a search session in which collaborators perform their search task on independent search engines.

  The BM25 ranking model (Robertson and Walker, 1994) estimates the similarity score between document d_i and query q_h as:

$$RSV(d_i, q_h) = \sum_{t_v \in q_h} \frac{N - n_v + 0.5}{n_k + 0.5} \frac{f_{iv} \cdot (k_1 + 1)}{f_{iv} + k_1 \cdot (1 - b + b \cdot \frac{|d_i|}{avg_{dl}})} \quad (8)$$

  where N expresses the collection size, n_v the number of documents including term t_v. The frequency of term t_v in document d_i is noted

f_{iv}. $|d_i|$ represents the length of document d_i whereas the average document length is noted avg_{dl}. k_1 and b are model parameters.

- *Logit*: the CIR model which only involves the last step of our algorithm (section 3.3) considering the whole set of features in order to measure the effectiveness of a personalized search without any consideration of the latent roles.

- *PM*: the CIR model refers to a system-based mediation guided by predefined fixed roles of Prospector-Miner (Pickens et al., 2008).

  According to collaborators' roles, this model relies on two different ranking functions:

  1. The query term suggestion function aims at favoring the Prospector's search diversity. For each term t_k belonging to documents previously retrieved in lists L, its score is estimated as:

     $$score(t_k) = \sum_{L_h \in L} w_r(L_h) w_f(L_h) rlf(t_k, L_h) \tag{9}$$

     where $rlf(t_k, L_h)$ expresses the number of documents with term t_k in list L_h.

  2. The document ranking function ensures the relevance of document not examined by the Prospector towards the topic. The score of document d_i is estimated as follows:

     $$score(d_i) = \sum_{L_h \in L} w_r(L_h) w_f(L_h) borda(d_i, L_h) \tag{10}$$

     where $borda(d_i, L_h)$ is a voting function.

  These two functions are based on relevance and freshness factors estimated as follows:

  - The relevance factor $w_r(L_h)$ which estimates the ratio of relevant documents in list L_h retrieved for query q_h, noted $|rel \in L_h|$ with the number of non-relevant documents in the same list, noted $|nonrel \in L_h|$:

    $$w_r(L_h) = \frac{|rel \in L_h|}{|nonrel \in L_h|} \tag{11}$$

- The freshness factor $w_f(L_h)$ which estimates the ratio of documents not visited in list L_h, noted $|nonvisit \in L_h|$, with the number of visited documents in L_h, noted $|visit \in L_h|$:

$$w_f(L_h) = \frac{|nonvisit \in L_h|}{|visit \in L_h|} \tag{12}$$

- *GS*: the CIR model refers to a system-based mediation guided by predefined fixed roles of Gatherer-Surveyor (Shah et al., 2010).

  This model is launched after query submissions (q_h and $q_{h'}$) of each collaborator (respectively, u_j and $u_{j'}$) and consists in two steps:

  1. The *Merging step* in which document lists are merged using the CombSUM function.
  2. The *Splitting step* in which documents in the merging list are classified using a 2-means algorithm. Each cluster is assigned to a collaborator using the following criteria: the cluster with the highest gravity center is assigned to the gatherer whereas the remaining cluster is assigned to the Surveyor.

- *RoleMining*: the user-driven system-mediated CIR model which mines predefined roles of collaborators in real time and rank documents according to the associated state of the art CIR models (Soulier et al., 2014). In contrast to our proposed approach, this setting consider roles (namely Gatherer/Surveyor and Prospector/Miner) predefined in a role taxonomy Golovchinsky et al. (2009) that could not exactly fits with users' skills.

  This model exploits the correlation matrix denoting collaborators' behaviors in order to assign users predefined roles, modeled through a role pattern. More particularly, according to a role pattern pool, the role-based identification assigns the role pattern correlation matrix $F^{R_{1,2}}$ which is the most similar to the collaborators' correlation one $C^{(t_l)}_{u_1,u_2}$ obtained for the pair of users (u_1, u_2) at given timestamp t_l.

$$\begin{aligned} & argmin_{\ R_{1,2}} ||F^{R_{1,2}} \ominus C^{(t_l)}_{u_1,u_2}|| \\ & subject\ to: \\ & \forall_{(f_j,f_k)\in\mathcal{K}^{R_{1,2}}}\ F^{R_{1,2}}(f_j,f_k) - C^{(t_l)}_{u_1,u_2}(f_j,f_k)) > -1 \end{aligned} \tag{13}$$

where $||.||$ represents the Frobenius norm and $\ominus$ is the minus operator defined as:

$$F^{R_{1,2}}(f_j, f_k) \ominus C^{(t_l)}_{u_1,u_2}(f_j, f_k) = \begin{cases} F^{R_{1,2}}(f_j, f_k) \ominus C^{(t_l)}_{u_1,u_2}(f_j, f_k) \\ \quad if\ F^{R_{1,2}}(f_j, f_k) \in \{-1; 1\} \\ 0 \quad otherwise \end{cases}$$

4.1.4. Ground truth and Metrics

We built the ground truth using clickthrough data following the assumption that implicit relevance derived from clicks are reasonably accurate (Joachims et al., 2005). More particularly, the ground truth relies only the clicked documents and includes an agreement level, as suggested in (Shah and González-Ibáñez, 2011b). However, in contrast to Soulier et al. (2014) which consider an agreement level involving two users, we reinforce the agreement level condition by the constraint that participants might belong to different groups in order to take into account the bias of intra-group collaboration interactions. Indeed, collaborators are likely to interact through the chat system in order share document links, as suggested in Section 4.1.1. This results in a small relevant document set, namely 38 and 20 for user studies *US1* and *US2* respectively.

We used well-known collaborative-based metrics proposed to evaluate the search outcomes of collaborative search using (Shah and González-Ibáñez, 2011b). These metrics are precision and recall oriented and are estimated at the group level considering documents selected by collaborators throughout the search session in which submitted queries constitutes a whole instead of independant actions. In order to evaluate the retrieval effectiveness of our proposed model, the metrics are applied on a document set agregating rankings retrieved throughout the session and estimated at the group level (as donne by Shah and González-Ibáñez (2011b)) over all queries submitted by all collaborators. More particularly, we (a) consider rankings with respect to their top 20 ranked documents as usually done in the information retrieval domain by the rank R, (b) merged the top 20 documents of all rankings retrieved with respect to queries submitted throughout the search session by all collaborators, and then (c) estimated the collaborative metrics over this merged document set, namely at the group level.

In order to estimate the collaborative-based metrics, we adapted the universe, relevant universe, coverage and relevant coverage sets defined in (Shah and González-Ibáñez, 2011b):

- The universe U of web pages represents the document dataset.
- The relevant universe U_r refers to the ground truth, with $U_r \subset U$.
- The coverage $Cov(g)$ of a collaborative group g expresses the total number of distinct documents retrieved for all queries submitted by collaborators of group g throughout the search session.
- The relevant coverage $RelCov(g)$ of a collaborative group g refers to the total number of distinct relevant documents retrieved for all queries submitted by collaborators of group g throughout the search session.

With this in mind, we used the following collaborative metrics measuring the synergic effect of a collaborative group g:

- The precision $Prec(g)$ estimated for collaborative group g:

$$Prec(g) = \frac{RelCov(g)}{Cov(g)} \tag{14}$$

- The recall $Recall(g)$ estimated for collaborative group g:

$$Recall(g) = \frac{RelCov(g)}{U_r} \tag{15}$$

- The F-measure $F(g)$ estimated for collaborative group g which combines both precision and recall metrics:

$$F(g) = \frac{2 * Prec(g) * Recall(g)}{Prec(g) + Recall(g)} \tag{16}$$

Finally, these measures are averaged over the collaborative groups of each user study, namely the 25 collaborative groups (*US1*) on one hand and the 10 ones (*US2*) on the other hand.

Please, note that the computation of these collaborative-oriented metrics is similar as it could be done by the precision and recall measures in classical information retrieval. However, while classical IR considers a ranking as evidence source, we rely here on a set built by a merging of the top 20 documents retrieved for all queries submitted over the search session by all collaborators.

4.2. Results

This section reports the obtained results with respect to several scopes. First, we present the parameter tuning step and then, we analyze the retrieval effectiveness.

4.2.1. Parameter Tuning

In order to highlight the consistency of our model regardless of users, tasks and topics, we performed a learning-testing approach in two steps, illustrated in Figure 6: (1) the learning step which optimizes the model parameter(s) using one of both datasets, e.g. *US1*, and (2) the testing step which estimates the retrieval effectiveness of our model on the other dataset, e.g. *US2*, according to parameter optimal value(s) found in the learning step.

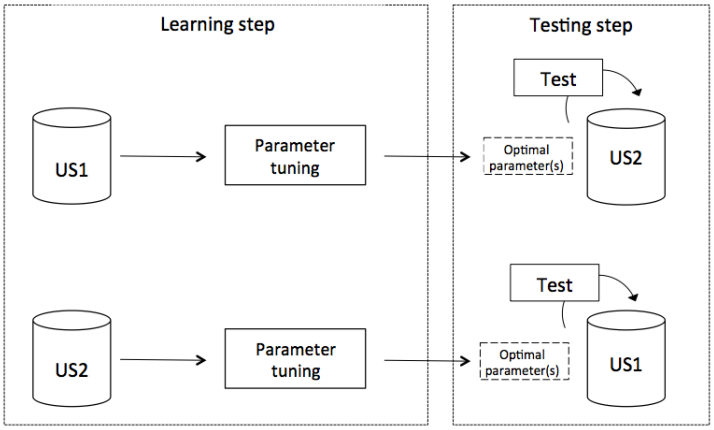


Figure 6: **Parameter tuning methodology**

For reminder, our model dynamically mine latent roles of collaborators according to a feature set by leveraging features' importance within the collaborative ranking model and collaborators' complementarities. With this in mind, we expressed the assumptions that the decay parameter γ combining these two aspects in the optimization problem (see Equation 4) might be fixed over the session. Accordingly, the tuning phase mainly concerns the γ parameter which expresses the collaborators' complementarity. Both versions of our proposed model, namely *MineRank(q)* and *MineRank(t)*, are concerned with this tuning and we consider the F-measure indicator as

the tuned effectiveness metric since it is a combination of precision and recall.

The first version of our model, namely *MineRank(q)*, launched at each query submission only depends on the parameter γ, used within the latent role mining step (Equation 4). The latter was tuned with a value range $\gamma \in [0..1]$, as illustrated in Figure 7. We can see that the optimal value for parameter γ is reached at 0.5 and 0.2 for respectively user studies *US1* and *US2* with a F-measure value respectively equals to 0.074 and 0.060. This difference suggests that the constraint of the report writing in *US2* does not allow collaborators to fully emphasize their search behavior complementarity.

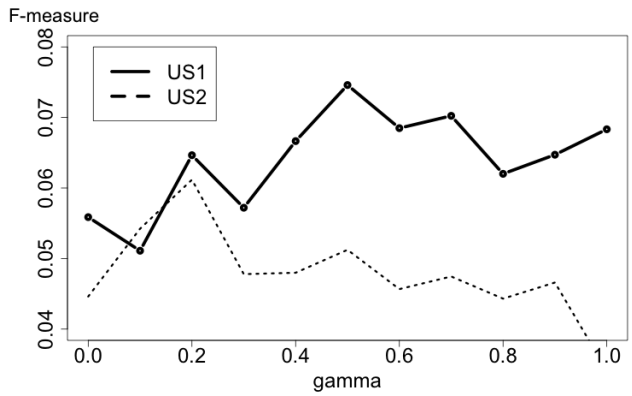
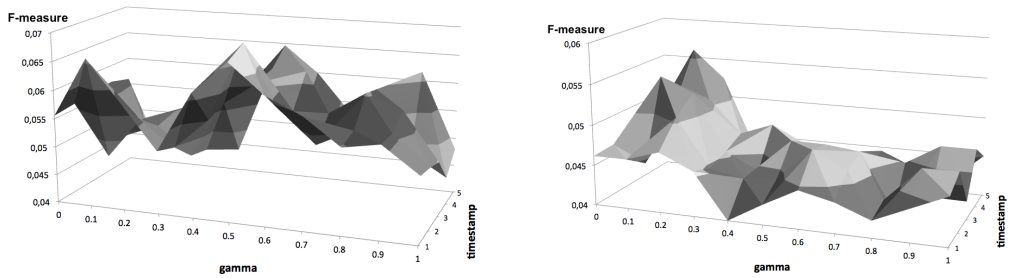


Figure 7: **Parameter tuning of *MineRank(q)***

The second version of our model, namely *MineRank(t)*, requires fixing the parameter $\gamma \in [0..1]$, as previously done, and also timestamp t where the latent role mining step is launched. We consider that a time-window for mining latent roles between 1 and 5 minutes is a reasonable range for our experiments in order to better fit with our model assumptions that search behaviors evolve throughout the search session. Figure 8a and Figure 8b illustrate the variation of the F-measure for our model *MineRank(t)* with respect to both parameters $\gamma \in [0..1]$ and $t \in [1..5]$, for respectively user studies *US1* and *US2*. The F-measure is optimal ($F = 0.069$) when $\gamma = 0.5$ and $t = 2$ for dataset *US1* while in dataset *US2* it reaches 0.056 when $\gamma = 0.1$ and $t = 3$.



(a) $MineRank(t)$ model for US1 (b) $MineRank(t)$ model for US2

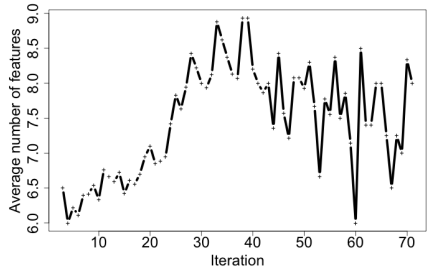
Figure 8: **Parameter tuning of *MineRank(t)***

The optimal values γ obtained for both versions highlight that the consideration of search skill complementarities within the role mining approach is higher in user study *US1* than in user study *US2*. Moreover, scores obtained for both versions highlight that the effectiveness of the model *MineRank(t)* is lower than those obtained for the model *MineRank(q)* in both user studies. This is consistent with the fact that our model relies on relevance feedback expressed after each submitted query, not particularly at regular timestamps. Therefore, for the remaining experiments, we only consider the version *MineRank(q)*.

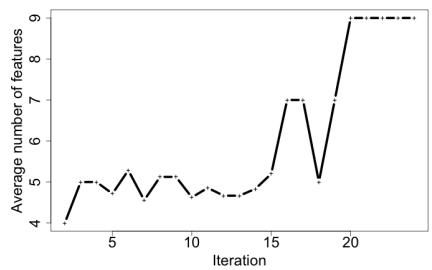
4.2.2. Analyzing the dynamics of collaborators' latent roles

In this section, our goal is to identify the evolution of search skills exploited by collaborators throughout the search session. For this purpose, we analyze, first, the average number of selected features for characterizing the latent role kernels of collaborators, and second, the average overlap between the feature set selected for two successive collaborators' latent role kernels. These indicators are estimated over time at each query submission, as illustrated in Figure 9. For reminder, we call an iteration t_l the time-window beginning at each time user u submits query q and ending while document list $\mathcal{D}_u^{t_l}$ is retrieved to user u.

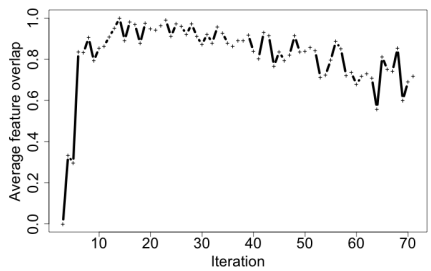
Since significant differences between features are required and a latent time interval is required for highlighting differences between collaborators (Soulier et al., 2014), we highlight that only 3 collaborative groups from the 25 in user study *US1* performed more than 60 iterations with respect to our model, and only one group performed more than 18 iterations in user study



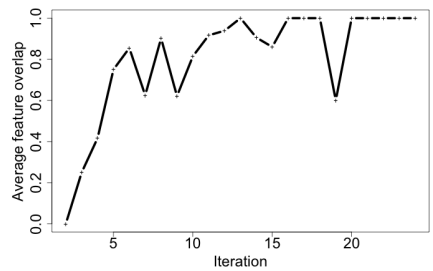
(a) Average number of features for user study US1



(b) Average number of features for user study US2



(c) Average feature set overlap for user study US1



(d) Average feature set overlap for user study US2

Figure 9: **Analysis of collaborators' latent role evolution throughout the whole search sessions.**

US2. Accordingly, the variations noticed afterwards are not significant.

From Figure 9a, respectively Figure 9b, we can see that the number of features increases over time and seems to reach an optimum between 8 and 9, respectively 7, search behavior features over the 11 for *US1*, respectively *US2*. First, the fact that the number of features increases over time might be explained by the amount of considered data (namely, relevance feedback) in the algorithm increases over time, increasing the likelihood to obtain significant p-value within the collaborators' behavior differences analysis. Second, the difference between both datasets can be due to the tasks of the participants, but also, this could suggests that a more intensive activity within the search process, through submitted queries for instance, gives a better landscape of collaborators complementarities.

In Figure 9c and Figure 9d, the overlap indicator is also close to 1 after the 20th submitted query for both datasets, which indicates that the kernel of the latent roles mined after this timestamp are almost stable over the remainder of the session. This suggests that participants naturally adopt the best behaviors with respect to their search skills, which tend to converge and be persistent over the time. Beforehand, participants take time to identify their best search strategies, and the latent role varies between successive query submissions.

4.2.3. Retrieval Effectiveness

In this section, we measure the retrieval effectiveness of our collaborative ranking approach based on latent role mining *MineRank(q)* with respect to state-of-the-art ranking models *BM25*, *Logit*, *GS*, *PM* and *RoleMining*. Table 5 presents the obtained results.

Table 5: **Comparison of the role mining impact on the retrieval effectiveness. *%Chg*: *MineRank* improvement. Student test significance *:** $0.01 < t \leq 0.05$ **; **:** $0.001 < t \leq 0.01$ **; ***:** $t \leq 0.001$**.**

training set → *testing set*		*Prec*@20			*Recall*@20			*F*@20		
		value	*%Chg*	*t*	value	*%Chg*	*t*	value	*%Chg*	*t*
Ground truth: Collaborative methodology										
$US2 \rightarrow US1$	BM25	0,009	285,79	***	0,066	236,51	***	0,017	280,84	***
	Logit	0,031	20,66		0,155	43,24	*	0,052	23,64	
	GS	0,009	306,78	***	0,045	400,10	***	0,015	324,29	***
	PM	0,014	178,55	***	0,029	650,15	***	0,018	254,85	***
	RoleMining	0.012	217.82	***	0.076	194.44	***	0.020	217.99	***
	MineRank(q)	**0,038**			**0,223**			**0,065**		
$US1 \rightarrow US2$	BM25	0,015	85,92	*	0,163	74.19	*	0,027	85.29	*
	Logit	0,025	10,67		0,252	12,50		0,046	10,92	
	GS	0,030	-6,24		0,184	54.28		0,051	0	
	PM	0,056	-50,04	***	0,205	38.46	*	0,088	-41.87	***
	RoleMining	0.024	18.91		0.216	31.70	*	0.046	20.13	*
	MineRank(q)	**0,028**			**0,284**			**0,051**		

From a general point of view, we can notice from Table 5 low evaluation metric values obtained using baselines and our proposed MineRank model as well. These low-level effectiveness results could be due to the association between different facts. Indeed, we hypothesize that precision measures are bounded de facto because of the small number of assumed relevant documents (see Section 4.1.4) for both user studies (38 and 20 for respectively $US1$

and $US2$). Moreover, since the SERPs were extracted asynchronously with participants' search task, as explained in section 4.1.2, the ground truth is more likely to be smaller since the documents clicked by the participants alongside the search task are not obviously in the top ranks of the SERPs.

In order to get a better insight of this statement, we used an extended ground truth to measure the retrieval effectiveness (see Table 6) and then analyzed the differences in the trends with the results presented in Table 5. The ground truth is extended with the top K similar documents to those clicked by the participants. Accordingly, we considered top 10 terms of these clicked documents to extend the task topic, as suggested in (Cui et al., 2002), and then extract top K similar documents obtained through a BM25 model, with K = 10, 20.

The comparison between the effectiveness results obtained using the initial ground truth (Table 5) and the extended ground truth (Table 6) highlights that precision, recall and F-measure values seem to increase if the ground truth includes more documents, particularly those considered as relevant through a query expansion methodology, while generally keeping the same trend over the effectiveness comparison. This observation reinforces our intuition about the fact that low metric values could be explained by the assumptions that support the methodology of building the ground truth.

When analyzing deeply the results, we can see that, regardless of the way used to build the ground truth, our model generally provides higher results than the four baselines. For instance, the F-measure values for $US2$ moved from 0.051 to 0.074 for respectively the initial ground truth and the extended one with $K = 10$, reaching 0.123 for $K = 20$. We note however that our model $MineRank(q)$ seems to be less effective than GS and PM baselines for the second user study $US2$ while introducing relevant documents in the ground truth, increasing the negative changes until obtaining significant decreases. We will address this issue later in the analysis. For convenience, we focus in what follows on the results presented in Table 5 obtained using the initial ground truth (built according to the clicked documents without expansion) which better fits with participants' relevance judgments.

In Table 5, we can see that the improvements with respect to the recall measure are generally higher than those obtained for the precision one, suggesting that our model is more recall-oriented. Indeed, this statement seems realistic since that the evaluation metrics are estimated at the group level,

Table 6: **Comparison of the role mining impact on the retrieval effectiveness. $\%Chg$: *MineRank* improvement according to the extended ground truth with top K similar documents. Student test significance *:** $0.01 < t \leq 0.05$ **; **:** $0.001 < t \leq 0.01$ **; ***:** $t \leq 0.001$**.**

training set → *testing set*		*Prec*@20			*Recall*@20			*F*@20		
		value	$\%Chg$	t	value	$\%Chg$	t	value	$\%Chg$	t
Extended ground truth ($K = 10$)										
	BM25	0,012	283,12	***	0,063	226,15	***	0,018	311,67	***
	Logit	0,041	9,12		0,173	18,44		0,067	9,45	
$US2 \rightarrow US1$	GS	0,010	339,35	***	0,038	437,68	***	0,016	348,90	***
	PM	0,016	187,76	***	0,025	733,95	***	0,018	308,62	***
	RoleMining	0,014	212,05	***	0,079	158,92	**	0,024	209,03	***
	MineRank(q)	**0,045**			**0,205**			**0,074**		
	BM25	0,026	87,45	*	0,193	75,00	*	0,045	86,64	*
	Logit	0,049	-2,41		0,341	-1,09		0,085	-1,95	
$US1 \rightarrow US2$	GS	0,059	-18,25		0,259	30,00		0,094	-11,06	
	PM	0,087	-45,08	*	0,226	49,18	*	0,125	-33,17	*
	RoleMining	0,041	16,84		0,256	31,88	*	0,070	18,91	
	MineRank(q)	**0,048**			**0,337**			**0,084**		
Extended ground truth ($K = 20$)										
	BM25	0,013	238,89	***	0,058	175,00	***	0,021	238,02	***
	Logit	0,044	4,16		0,141	12,97		0,066	5,54	
$US2 \rightarrow US1$	GS	0,011	331,31	***	0,031	422,50	***	0,016	346,61	***
	PM	0,016	179,35	***	0,022	633,95	***	0,019	275,74	***
	RoleMining	0,015	196,36	**	0,067	136,64	**	0,024	185,44	**
	MineRank(q)	**0,045**			**0,159**			**0,070**		
	BM25	0,041	17,55		0,208	62,43	*	0,067	24,01	
	Logit	0,078	-38,80	*	0,365	-7,66		0,128	-34,54	*
$US1 \rightarrow US2$	GS	0,098	-51,20	*	0,288	17,23		0,143	-41,66	*
	PM	0,137	-65,01	**	0,240	40,43	*	0,173	-51,72	*
	RoleMining	0,064	7,15		0,265	27,18	*	0,102	-18,09	
	MineRank(q)	0,075			0,360			0,123		

aggregating the different lists retrieved for all collaborators. Therefore, the coverage is more likely to be higher than the size of the ground truth, leading to recall measures over-passing precision ones. Moreover, for user study *US1* our model significantly overpasses both individual and collaborative baselines from 43.24% to 650.15% over the three metrics. For the second user study *US2*, we obtained significant improvements for the *BM25* and *RoleMining* baselines over at least two metrics as well as the collaborative ranking models *PM* for the recall measure. These results highlight four main contributions:

1. Our model enables users to benefit from the synergic effect of collaboration since the effectiveness of the collaboration (our model *MineRank*) is greater than the sum of the individual search session (baseline *BM25*);

2. Ranking documents with respect to latent roles gives an additional value to a CIR model based only on the behavior analysis of collaborators, more particularly in terms of recall (baseline *Logit*);

3. Mining latent roles for collaborators seems to be more effective than a CIR scenario in which roles are fixed throughout the search session (baselines *GS* and *PM*).

4. Leveraging complementarity in unlabeled roles of collaborators seems more effective than mining predefined roles with respect to their differences in search behaviors (baseline *RoleMining*).

However, by comparing both datasets, we observe that improvements are lower for user study *US2*, and more particularly with respect to the *PM* baseline for the precision measure, and accordingly the F-measure. This can be explained by two main reasons. First, the ratio between the number of submitted queries and the size of the document collection of both user studies suggests that the *US1* dataset seems to be more topic-concentrated than *US2*. Another reason might be that the observed difference in retrieval effectiveness between US1 and US2 is due to the difference in the ability of the latent role kernel, identified using *Coll-Clique* algorithm on both datasets, to the capture complementarities between the involved users. To investigate this hypothesis, we compared, for each user study, (1) the kernel tuned with the optimal setting M_O (γ =0.5 resp. 0.2 for *US1* and *US2* allowing to achieve the optimal retrieval effectiveness based on F-measure, as shown in Figure 7), and (2) the kernel identified in the tested model $MineRank(q)$. More particularly, we carried out a statistical analysis in order to determine, for each model (M_O and $MineRank(q)$), which features impact on the F-measure. For this purpose, we performed an ANOVA analysis between the F-measure obtained at each query submission and the whole features, noted as binary indicators referring to the presence or absence of the features within the kernel of the mined latent role. To obtain the best model, we consider first a full model including all explanatory variables (binary indicators of features), and performed a backward elimination to

Table 7: **Explanatory features for the F-measure according to the ANOVA statistical analysis. p-value *:** $0.01 < p \leq 0.05$ **; **:** $0.001 < p \leq 0.01$ **; ***:** $p \leq 0.001$**.**

	$MineRank(q)$		M_O	
	Feature	p-value	Feature	p-value
US1	TiO	0.033*	TiO	0.002**
	Spec	0.050*	TeO	0.034*
	SO	0.033*	TQTP	0.006**
			VP	0.000***
			Spec	0.015*
			SO	0.000***
US2	TOD	0.032*	TiO	0.036*
			TeO	0.001***
			VP	0.021*

remove iteratively the less significant factors with respect to the F-measure.

Table 7 presents the obtained models after the backward elimination for both models $MineRank(q)$ and M_O and both user studies *US1* and *US2*. We can see that, for both $MineRank(q)$ and M_O, the set of significant features is larger for user study *US1* than for *US2* as can be expected from Figure 9. This suggests that there is a wider range of differences between collaborators in *US1* than in *US2*, enabling to fully benefit from the collaborators' skills complementarity. Moreover, comparing the best explicative models obtained for $MineRank(q)$ with M_O, we can see that for *US1* half (3/6) of the features highlighted as significant for M_O are also significant for $MineRank(q)$, while for *US2* any (0/3) significant features for M_O have been highlighted as significant for $MineRank(q)$. Combining these observations, we can clearly explain the low results obtained on *US2*. Indeed, the difference between the number features significant in the *MineRank* setting and in the M_O one, resulting in a single feature (TOD), suggests that this set is insufficient for modeling behavioral differences/complementarities between users. This could be explained by the fact that the *Coll-Clique* algorithm is based on the γ parameter tuned on users of *US1* that behave much more differently.

5. Discussion and Conclusion

While several works claim that roles enable to structure collaboration among the participants (Kelly and Payne, 2013), some role-based approaches are inadequate to leverage the skills that individuals bring to a collaboration and to deal with the dynamics of the collaboration. Recently, a new approach (Soulier et al., 2014) combining a user-driven and a system-based mediation has been proposed, enabling to leverage collaborators' search behaviors in order to mine predefined roles belonging to a taxonomy.

In this article, we presented an unsupervised manner to mine latent roles of collaborators on the basis of their evolving search behavior complementarities. These latent roles, and more particularly the search features identified for characterizing their latent roles, are re-injected within the collaborative document ranking. Experiments on two user studies highlight that collaborators are able to benefit from the synergic effect of the collaboration while the real-time mining of their search behavior complements their latent roles.

However, this work is not without limitations leading to different perspectives.

First, we focus on the model design which relies on a particular collaborative setting between a pair of users working synchronously. We believe that our model could be enhanced as follows:

- Some work have highlighted that collaboration is generally performed in larger groups (Morris, 2013) and we believe that our contribution would gain in maturity if it could allow collaboration between more than two users. This extension would raise new challenges in the latent role modeling as well as the model steps. More particularly, a single correlation matrix is not sufficient to highlight search skill differences and similarities between users of a larger group since it is adapted for a dyad. Since our intuition is to identify in which skill each user is the most effective, an intrinsic analysis of search behaviors would lack the principles of division of labor and sharing of knowledge guiding collaboration. With this in mind, one intuition could be to build latent roles in a first time pairs by pairs which then could be aggregated and analyzed at the group level in order to identify each collaborator's skills.

- Also, the synchronicity of the search actions could be considered dif-

ferently in the model assuming that users could work at different time-windows. This temporal asynchronicity raises the issue of multi-session management which could impact the search behavior modeling based on cumulative values of search features from the beginning of the search session to a given timestamps.

- Furthermore, we assumed that collaborators have complementary skills (Sonnenwald, 1996) which is not always the case in a CIR setting. For instance, in the retrieval model proposed by (Foley and Smeaton, 2009), document rankings are not personalized implying that the complementarity between collaborators is not necessarily required.

Second, experiments are somehow restricted to a particular framework which could limit the generalizability of results. We highlight below its peculiarities:

- We evaluated our model through user studies in which participants aimed at solving an exploratory task. We highlight that other types of tasks or also used in collaborative search, such as travel planning, shopping, fact-finding tasks, etc (Morris, 2013). Therefore, a future work would be to analyze if the latent role would benefit with these another type of tasks, more practical and less topic-oriented in which search skills are predominant.

- Last, our experimentation consists in a log-study in which latent role mining and collaborative document ranking are performed on search logs. We are aware that a user-study would be better appropriated to fit likely with collaborators search behaviors (interactions, communications or search strategies guided by ranking retrieved by the proposed model), but log-study-based evaluation protocol is less time-consuming and enables to compare with different baselines without additional costs. We highlight however that search logs are those collected during a real collaborative search session, making our evaluation as natural as possible. In the near future, we plan to evaluate our model through a user-study-oriented evaluation in which participants interacts with a system supporting the proposed CIR model based on latent roles of collaborators. We assume that this interactive and real-time user-study would also tackle the low metric value issue that we point out during the experiments.

Appendix A. Coll-Clique illustration

In order to illustrate our algorithm, we consider a collaborative search setting involving two users u_1 and u_2 whose behavioral activity is modeled by a set F of features $\{f_1, f_2, f_3, f_4\}$.

Figure A.10 illustrates the feature-based graph representing behavior of these two collaborators u_1 and u_2 at timestamp t_l. For instance, the weight of node f_1, namely 0.84, represents the feature importance f_1 estimated by Equation 2. The weight of the edge connecting f_1 and f_4 expresses the correlation between the collaborators' differences towards these two features, also perceived as the level of complementarity $C_{1,2}^{(t_l)}(f_1, f_4)$.

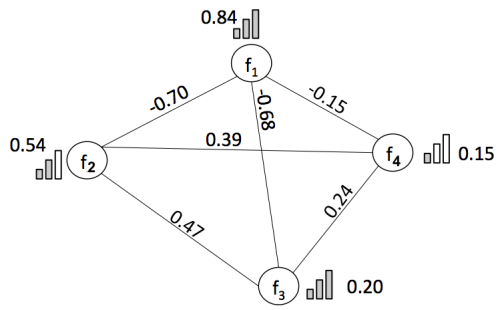


Figure A.10: Example of graph modeling collaborators behaviors according to four behavioral features

Table A.8 presents our Coll-Clique algorithm relying on the behavioral graph presented in Figure A.10. The "Rec." column expresses the degree of the recursion in the algorithm. Columns "C", "P" and "K" represent the recursion input data while columns "C $\oplus f_h$", "P'" and "K_{end}" represent the outputs of the recursion, respectively including the growing clique, the new candidate network in which feature weight are depreciated and the current maximum clique.

At the end of the recursion, the local clique consists of the feature set including f_1, f_2, and f_3 since that the sum of the node weight $W(C)$ is greater than the weight $W(K)$ of current clique K.

Rec.	C	P	K	W(C)	W(P)	W(K)	$C \oplus f_h$	P'	K_{end}
init	{}	1 (0.84) 2 (0.54) 3 (0.20) 4 (0.15)	{}	0	1.73	0	-	-	-
1	{}	1 (0.84) 2 (0.54) 3 (0.20) 4 (0.15)	{}	0	1.73	0	1 (0.84)	2 (1.24) 3 (0.88) 4 (0.3)	-
1.1	1 (0.84)	2 (1.24) 3 (0.88) 4 (0.3)	{}	0.84	2.42	0	1 (0.84) 2 (1.24)	3 (0.41)	-
1.1.1	1 (0.84) 2 (1.24)	3 (0.41)	{}	2.08	0.41	0	1 (0.84) 2 (1.24) 3 (0.41)	{}	1 (0.84) 2 (1.24) 3 (0.41)
1.1	1 (0.84)	2 (1.24) 3 (0.88) 4 (0.3)	1 (0.84) 2 (1.24) 3 (0.41)	0.84	2.42	2.49	1 (0.84) 3 (0.88)	2 (0.07) 4 (0.06)	1 (0.84) 2 (1.24) 3 (0.41)
1.1.1	1 (0.84) 3 (0.88)	2 (0.07) 4 (0.06)	1 (0.84) 2 (1.24) 3 (0.41)	1.72	0.13	2.53	STOP	ALGORITHM	

Table A.8: Illustration of a recursion of our Coll-Clique algorithm

References

Agichtein, E., Brill, E., Dumais, S., Ragno, R., 2006. Learning user interaction models for predicting web search result preferences, in: SIGIR, pp. 3–10.

Bennett, P.N., White, R.W., Chu, W., Dumais, S.T., Bailey, P., Borisyuk, F., Cui, X., 2012. Modeling the impact of short- and long-term behavior on search personalization, in: SIGIR, pp. 185–194.

Carraghan, R., Pardalos, P.M., 1990. An exact algorithm for the maximum clique problem. Operations Research Letters 9, 375 – 382.

Cui, H., Wen, J.R., Nie, J.Y., Ma, W.Y., 2002. Probabilistic query expansion using query logs, in: WWW, ACM. pp. 325–332.

Dourish, P., Bellotti, V., 1992. Awareness and coordination in shared workspaces, in: Proceedings of the ACM Conference on Computer-supported Cooperative Work, ACM. pp. 107–114.

Evans, B.M., Chi, E.H., 2010. An elaborated model of social search. Inf. Process. Manage. 46, 656–678.

Fidel, R., Bruce, H., Pejtersen, A.M., Dumais, S.T., Grudin, J., Poltrock, S., 2000. Collaborative Information Retrieval (CIR). The New Review of Information Behaviour Research 1, 235–247.

Foley, C., Smeaton, A.F., 2009. Synchronous collaborative information retrieval: Techniques and evaluation, in: ECIR, pp. 42–53.

Foley, C., Smeaton, A.F., 2010. Division of labour and sharing of knowledge for synchronous collaborative information retrieval. Inf. Process. Manage. 46, 762–772.

Geng, X., Liu, T.Y., Qin, T., Li, H., 2007. Feature selection for ranking, in: SIGIR, pp. 407–414.

Golder, S.A., Donath, J., 2004. Social roles in electronic communities, in: Proc. of Association of Internet Researchers Conference.

Golovchinsky, G., Qvarfordt, O., Pickens, J., 2009. Collaborative information seeking. IEEE Computer 42, 47–51.

Heath, A.P., White, R.W., 2008. Defection detection: Predicting search engine switching, in: WWW, pp. 1173–1174.

Henderson, K., Gallagher, B., Eliassi-Rad, T., Tong, H., Basu, S., Akoglu, L., Koutra, D., Faloutsos, C., Li, L., 2012. Rolx: Structural role extraction & mining in large graphs, in: SIGKDD, pp. 1231–1239.

Joachims, T., 2002. Optimizing search engines using clickthrough data, in: SIGKDD, pp. 133–142.

Joachims, T., Granka, L.A., Pan, B., Hembrooke, H., Gay, G., 2005. Accurately interpreting clickthrough data as implicit feedback, in: SIGIR, ACM. pp. 154–161.

Joho, H., Hannah, D., Jose, J.M., 2009. Revisiting ir techniques for collaborative search strategies, in: ECIR, pp. 66–77.

Kelly, R., Payne, S.J., 2013. Division of labour in collaborative information seeking: Current approaches and future directions, in: CIS Workshop at CSCW.

Kwak, H., Lee, C., Park, H., Moon, S., 2010. What is twitter, a social network or a news media?, in: WWW, pp. 591–600.

Leung, K.W.T., Lee, D.L., Ng, W., Fung, H.Y., 2008. A framework for personalizing web search with concept-based user profiles. ACM Trans. Internet Technol. 11, 17:1–17:29.

McCallum, A., Wang, X., Corrada-Emmanuel, A., 2007. Topic and role discovery in social networks with experiments on enron and academic email. J. Artif. Int. Res. 30, 249–272.

Morris, M.R., 2013. Collaborative Search Revisited, in: Proceedings of the Conference on Computer Supported Cooperative Work, ACM. pp. 1181–1192.

Morris, M.R., Horvitz, E., 2007. Searchtogether: an interface for collaborative web search, in: UIST, pp. 3–12.

Morris, M.R., Teevan, J., Bush, S., 2008. Enhancing collaborative web search with personalization: groupization, smart splitting, and group hit-highlighting, in: CSCW, pp. 481–484.

Pal, A., Counts, S., 2011. Identifying topical authorities in microblogs, in: WSDM, pp. 45–54.

Pickens, J., Golovchinsky, G., Shah, C., Qvarfordt, P., Back, M., 2008. Algorithmic mediation for collaborative exploratory search, in: SIGIR, pp. 315–322.

Radinsky, K., Svore, K.M., Dumais, S.T., Shokouhi, M., Teevan, J., Bocharov, A., Horvitz, E., 2013. Behavioral dynamics on the web: Learning, modeling, and prediction. ACM Trans. Inf. Syst. 31, 16:1–16:37.

Robertson, S.E., Walker, S., 1994. Some simple effective approximations to the 2-Poisson model for probabilistic weighted retrieval, in: Proceedings of the Annual International SIGIR Conference on Research and Development in Information Retrieval, ACM. pp. 232–241.

Shah, C., González-Ibáñez, R., 2011a. Coagmento - A System for Supporting Collaborative Information Seeking., in: Demo in Proceedings of Association for Information Science and Technology Annual Meeting, pp. 9–12.

Shah, C., González-Ibáñez, R., 2011b. Evaluating the synergic effect of collaboration in information seeking, in: SIGIR, pp. 913–922.

Shah, C., Pickens, J., Golovchinsky, G., 2010. Role-based results redistribution for collaborative information retrieval. Inf. Process. Manage. 46, 773–781.

Sonnenwald, D.H., 1996. Communication roles that support collaboration during the design process. Design Studies 17, 277–301.

Soulier, L., Shah, C., Tamine, L., 2014. User-driven system-mediated collaborative information retrieval, in: SIGIR, pp. 485–494.

Soulier, L., Tamine, L., Bahsoun, W., 2013. A collaborative document ranking model for a multi-faceted search, in: AIRS, pp. 109–120.

Teevan, J., Dumais, S.T., Horvitz, E., 2005. Personalizing search via automated analysis of interests and activities, in: SIGIR, pp. 449–456.

Twidale, M.B.T., Nichols, D.M.N., Paice, C.D., 1997. Browsing is a Collaborative Process. Inf. Process. Manage. 33, 761–783.

White, R.W., Dumais, S.T., 2009. Characterizing and predicting search engine switching behavior, in: CIKM, pp. 87–96.

Wu, Q., Hao, J.K., 2015. A review on algorithms for maximum clique problems. European Journal of Operational Research 242, 693 – 709.

Yue, Z., Han, S., He, D., 2014. Modeling search processes using hidden states in collaborative exploratory web search, in: CSCW, ACM. pp. 820–830.