

---

# Définition et exploitation des méta-rôles des utilisateurs pour la recherche d'information collaborative

**Laure Soulier – Lynda Tamine**

*UPS - Laboratoire IRIT  
118 route de Nabonne - 31062 Toulouse Cedex 9  
soulier, tamine@irit.fr*

---

*RÉSUMÉ. La recherche d'information collaborative est un processus particulier impliquant un ensemble d'utilisateurs partageant un même besoin en information. Dans ce contexte, l'exploitation de la division du travail au travers des rôles est une des techniques utilisées pour structurer la session de recherche et optimiser son efficacité. Dans ce papier, nous proposons d'étudier les caractéristiques de comportement d'une paire de collaborateurs sur la base d'hypothèses de leur complémentarité. Nous définissons ainsi la notion de rôles latents qui sont (a) détectés en temps réel et (b) ensuite réinjectés dans un modèle d'ordonnancement de documents. Les expérimentations, menées sur des fichiers logs de sessions de collaboration réelles impliquant des paires d'utilisateurs, mettent en évidence l'efficacité de notre approche comparativement à des stratégies de recherche individuelles ou à celles qui considèrent des rôles fixes.*

*ABSTRACT. Collaborative information retrieval is a particular setting involving a set of users sharing the same information need. In this context, the application of the division of labor policy through collaborators' roles is generally used in order to structure the search session and enhance its retrieval effectiveness. In this paper, we propose to analyse the search features of pairwise collaborators allowing to identify their implicit roles according to research hypothesis based on their complementarity. We define the notion of collaborators' meta-role which is (a) identified in real time, and (b) reinjected within a collaborative document ranking model. The experimental evaluation performed on search logs of real collaborative search session involving pairs of users highlights the effectiveness of our model with respect to individual-based or fixed roles-based search sessions.*

*MOTS-CLÉS : recherche d'information collaborative, rôles, modèle d'ordonnancement, complémentarité des comportements*

*KEYWORDS: collaborative information retrieval, roles, ranking model, behavior complementarity*

---

## 1. Introduction

La recherche d'information collaborative (RIC) est un domaine de recherche qui traite des problématiques sous-jacentes à la définition d'environnements permettant à un groupe d'utilisateurs de mener une activité de recherche d'information en vue de résoudre un besoin en information partagé. Ce besoin s'inscrit généralement dans le cadre d'une tâche exploratoire complexe (Pickens *et al.*, 2008 ; Shah et González-Ibáñez, 2011). Ainsi, le domaine est à la croisée de nombreuses disciplines, dont principalement les interactions homme-machine (Morris et Horvitz, 2007) et la recherche d'information (RI) (Foley et Smeaton, 2009). Sous ce dernier angle particulièrement et à la différence de la RI classique ou individuelle, les enjeux majeurs consistent à formaliser des distributions d'ordonnements de résultats en considérant les utilisateurs impliqués (Pickens *et al.*, 2008) puis à optimiser la pertinence collective (Foley et Smeaton, 2009). Dans cette perspective, l'exploitation de la complémentarité des utilisateurs est une des techniques utilisées (Sonnenwald *et al.*, 2004) qui repose sur le paradigme de *division du travail* et qui a pour objectif de structurer et d'optimiser l'efficacité d'une session de recherche (Kelly et Payne, 2013). Selon une approche similaire adoptée en RI classique qui consiste à exploiter le comportement de l'utilisateur pour prédire la pertinence des résultats (Agichtein *et al.*, 2006 ; Joachims, 2002), une direction de recherche en RIC s'appuie sur la caractérisation des comportements des différents utilisateurs, à travers la notion de rôle (Pickens *et al.*, 2008 ; Soulier *et al.*, 2014b). Cette notion fait l'objet d'une taxonomie (Golovchinsky *et al.*, 2009) et est à la base de différents modèles comme celui reposant sur la paire de *prospecteur* et de *mineur* (Pickens *et al.*, 2008). Dans ces travaux, les rôles sont cependant assignés *a priori* aux utilisateurs et des modèles d'ordonnements appropriés sont utilisés à la base de la médiation qui permet de calculer les ordonnements personnalisés à chaque utilisateur. Cependant, ces modèles de médiation orientés-rôles sont basés sur l'hypothèse forte que les utilisateurs ont un comportement défini *a priori*. Dans ce papier, nous relâchons cette hypothèse en proposant d'identifier les rôles des utilisateurs au travers de caractéristiques de comportement, telles que le temps de visite d'une page. D'une façon globale, nous proposons un modèle de RIC, appelé *MineRank*, qui apprend en temps réel les rôles implicites des utilisateurs durant une session de RIC à partir de leur comportement afin d'estimer la pertinence collective des documents. Plus précisément, notre modèle repose sur deux étapes : (1) l'apprentissage des rôles latents des collaborateurs d'une façon non-supervisée à partir des caractéristiques de comportement et (2) l'intégration des rôles des collaborateurs dans l'ordonnement collaboratif des documents. L'évaluation expérimentale basée sur des fichiers *logs* de deux collections issues d'activité de recherche collaborative entre paires d'utilisateurs montre l'efficacité de notre modèle.

Le papier est organisé comme suit. La section 2 présente les travaux liés à la prise en compte des rôles en RI et en RIC. La section 3 introduit les motivations et les questions de recherche. La section 4 détaille notre modèle de RIC basé sur les méta-rôles des collaborateurs. L'évaluation expérimentale est présentée dans la section 5. Enfin, la section 6 conclut le papier.

## 2. Synthèse des travaux connexes

Nous synthétisons dans un premier temps les travaux portant sur la modélisation du comportement et notamment des rôles des utilisateurs en RI. Ensuite, nous présentons spécifiquement une synthèse des travaux de la RIC basés sur les rôles.

**Analyse du comportement des utilisateurs pour la détection des rôles.** En RI, le principal objectif de la modélisation d'un utilisateur est la prédiction de la pertinence des résultats en réponse à sa requête (Agichtein *et al.*, 2006 ; Joachims, 2002). Les caractéristiques de comportement utilisées sont les données utilisateurs, telles que les clics (Joachims, 2002), les mouvements des yeux (Lorigo *et al.*, 2008) ou les interactions sociales (Kwak *et al.*, 2010 ; Henderson *et al.*, 2012). Ce dernier aspect a été particulièrement traité dans l'identification de rôles sociaux, tels que les rôles de "Leaders" (Pal et Counts, 2011) ou de "Celebrities" (Golder et Donath, 2004). Pour cela, des techniques statistiques (Golder et Donath, 2004), probabilistes (Pal et Counts, 2011) ou de classification (Kwak *et al.*, 2010) sont utilisées. D'autres auteurs se concentrent sur l'apprentissage des rôles latents (Henderson *et al.*, 2012 ; McCallum *et al.*, 2007) en proposant une méthode formelle qui repose sur l'analyse des similarités ou des oppositions des utilisateurs en tenant compte de leurs interactions. Par exemple, Henderson *et al.* (2012) réalisent une transformation d'une matrice multidimensionnelle de caractéristiques de comportement afin d'extraire les rôles structurels des utilisateurs du réseau tandis que McCallum *et al.* (2007) revisitent l'algorithme du LDA pour apprendre les rôles évolutifs des acteurs d'un réseau de communication en utilisant comme source d'évidence le contenu des messages.

**L'exploitation des rôles en RIC.** L'enjeu principal de la RIC consiste en l'optimisation de la collaboration par le biais de la division du travail basée sur la différenciation des utilisateurs afin d'éviter la redondance de leurs actions (Kelly et Payne, 2013). Deux types de médiation peuvent répondre à cet enjeu : (1) la *médiation utilisateur*, reposant sur des interfaces adaptées à la collaboration (Morris et Horvitz, 2007) où les collaborateurs sont libres et ont la possibilité de coordonner leurs actions afin de faire émerger leurs différences et leurs rôles implicites et (2) la *médiation système* basée sur des techniques ou des modèles de RI qui exploitent les différences entre les collaborateurs afin de prédire la pertinence des documents (Pickens *et al.*, 2008 ; Shah *et al.*, 2010 ; Soulier *et al.*, 2014b). Dans cette dernière approche, plus proche de notre contribution, les différences et les oppositions de comportement entre les collaborateurs peuvent être caractérisées par des rôles (Kelly et Payne, 2013 ; Sonnenwald *et al.*, 2004). Ces derniers sont définis dans une taxonomie (Golovchinsky *et al.*, 2009) et exploités en tant que rôles fixes et prédéfinis dans des modèles de RIC comme ceux de Prospecteur et de Mineur (Pickens *et al.*, 2008), de Rassembleur et de Surveilleur (Shah *et al.*, 2010) ou d'Expert et de Novice de domaine (Soulier *et al.*, 2014b). En amont de ces modèles et dans l'objectif de répondre à la problématique d'assignation des rôles, nous avons proposé d'identifier de façon supervisée les rôles prédéfinis des utilisateurs en temps réel et d'adapter en conséquence le modèle d'ordonnement collaboratif des documents (Soulier *et al.*, 2014a). Cette approche définit la notion de session hybride permettant aux utilisateurs de structurer librement la collaboration

(médiation utilisateur), tout en étant guidés par un modèle d'ordonnement basé sur les rôles détectés (médiation système).

### 3. Motivations et questions de recherche

L'analyse de l'état de l'art permet de mettre en évidence l'apport des rôles dans la collaboration afin de structurer et d'exploiter la complémentarité des collaborateurs. Une première réponse orientée système a été fournie dans la littérature en proposant des modèles de RIC basés sur des rôles prédéfinis et fixes sur l'ensemble de la session (Pickens *et al.*, 2008; Shah *et al.*, 2010; Soulier *et al.*, 2014b). La problématique d'assignation des rôles aux utilisateurs a été soulignée dans ce contexte (Soulier *et al.*, 2014a). Cependant, nous faisons l'hypothèse que la modélisation des comportements selon des rôles prédéfinis sans aucun *a priori* des compétences des utilisateurs peut être réductrice de la latitude de leurs actions au cours d'une session de recherche. Ainsi, nous proposons de bénéficier des avantages des deux types de médiation (médiation système et médiation utilisateur) et d'étendre notre précédente contribution (Soulier *et al.*, 2014a) en relâchant la contrainte des rôles prédéfinis afin d'identifier les rôles latents des collaborateurs. Pour cela, nous définissons le concept de méta-rôle en nous appuyant sur deux hypothèses de recherche principales basées sur le comportement des collaborateurs : (1) les collaborateurs ont des comportements différents et (2) ces différences de comportement évoluent tout au long de la session.

A notre connaissance, nos travaux sont les premiers à considérer les deux types de médiation (utilisateur et système) dans un contexte de RIC. Pour cela, nous exploitons la différence et la complémentarité des collaborateurs en temps réel afin d'identifier les compétences de recherche pour lesquelles ils sont les plus efficaces dans le but d'optimiser l'efficacité de la session de recherche. Plus particulièrement, notre contribution s'appuie sur les questions de recherche suivantes :

- QR1 : Comment inférer le méta-rôle des collaborateurs à partir de leurs actions ?
- QR2 : Comment exploiter le méta-rôle des collaborateurs afin de prédire la pertinence collective des documents ?

### 4. Modèle non supervisé de RIC basé sur une médiation hybride

Nous présentons dans cette section notre modèle de RIC basé sur les méta-rôles des collaborateurs. Dans la mesure où notre modèle est une extension d'un de nos précédents travaux (Soulier *et al.*, 2014a), nous utilisons les concepts et les notations sous-jacents à la définition des complémentarités des comportements des utilisateurs.

#### 4.1. Notions de base

*Session de recherche utilisateur.* La session de recherche collaborative synchrone  $S$  implique une paire d'utilisateurs  $u_1$  et  $u_2$  qui interagissent afin de résoudre un besoin en information partagé  $I$  pendant une durée  $T$ . Chaque utilisateur  $u_j \in \{u_1, u_2\}$  effectue des actions séparément et formule ses propres requêtes  $\{q_{u_j}^1, q_{u_j}^2, \dots, q_{u_j}^T\}$ , ordonnées de façon chronologique. Pour chacune d'elles, des ensembles de documents-résultats  $\{D_{u_j}^1, D_{u_j}^2, \dots, D_{u_j}^T\}$  sont respectivement retournés. Nous caractérisons les actions des utilisateurs par un ensemble de caractéristiques de

comportement  $F = \{f_1, \dots, f_k, \dots, f_n\}$ , telles que le temps de visite d'une page ou la note moyenne attribuée aux pages visitées, sauvegardées pendant la session de recherche  $S$ . Nous modélisons le comportement de l'utilisateur  $u_j$  à un instant  $t_l$  de la session de recherche au travers d'une matrice temporelle  $S_{u_j}^{(t_l)} \in \mathcal{R}^{t_l \times n}$ . Chaque élément  $S_{u_j}^{(t_l)}(t_l, f_k)$  de la matrice  $S_{u_j}^{(t_l)}$  représente la valeur moyenne de la caractéristique  $f_k$  pour l'utilisateur  $u_j$  agrégée pour l'intervalle de temps  $[0..t_l]$  compte tenu des documents sélectionnés  $\mathcal{D}(u_j)^{(t_l)}$  durant cette période. La modélisation temporelle du comportement des utilisateurs  $S_{u_j}^{(t_l)}$  permet de prendre en compte le fait que leur comportement s'affine au cours de la session et de les caractériser d'une façon globale à un instant  $t_l$  en réduisant le biais d'actions bruitées dans la session.

*Meta-rôle.* A chaque instant  $t_l$ , le méta-rôle  $MR_{1,2}^{(t_l)}$  met en évidence les différences de comportement entre les collaborateurs  $u_1$  et  $u_2$  pendant la fenêtre temporelle  $[0..t_l]$ . Plus particulièrement, le méta-rôle souligne les complémentarités des collaborateurs vis-à-vis de leurs compétences de recherche sans aucun *a priori* de rôles prédéfinis ou d'appartenance à une taxonomie (Golovchinsky *et al.*, 2009). Similairement à la notion de rôle introduite précédemment (Soulier *et al.*, 2014a), un méta-rôle  $MR_{1,2}^{(t_l)}$  est construit pour une paire d'utilisateurs  $u_1$  et  $u_2$  et est constitué d'un noyau  $\mathcal{K}_{1,2}^{(t_l)} = \{f_1, f_2, \dots, f_p\} \in F^p$  de  $p$  caractéristiques de comportement  $f_k \in F$  ainsi que d'une matrice de corrélations  $C_{1,2}^{(t_l)} \in \mathcal{R}^{p \times p}$  qui traduit les complémentarités et les similarités entre les collaborateurs  $u_1$  et  $u_2$  compte tenu de leur comportement de recherche. Les corrélations sont estimées sur la base des différences  $\Delta_{1,2}^{(t_l)}(f_k)$  et  $\Delta_{1,2}^{(t_l)}(f_{k'})$  de paires de caractéristiques  $f_k \in \mathcal{K}_{1,2}^{(t_l)}$  et  $f_{k'} \in \mathcal{K}_{1,2}^{(t_l)}$ , où  $\Delta_{1,2}^{(t_l)}(f_k) = S_{u_1}^{(t_l)}(f_k) - S_{u_2}^{(t_l)}(f_k)$ . Notre objectif étant d'identifier les complémentarités entre collaborateurs, nous faisons l'hypothèse que les collaborateurs se comportent d'une façon complémentaire vis-à-vis de deux caractéristiques  $f_k$  et  $f_{k'}$  si la corrélation  $\rho(\Delta_{1,2}^{(t_l)}(f_k), \Delta_{1,2}^{(t_l)}(f_{k'}))$  de leur différence est proche de -1. A l'inverse, une corrélation proche de 1 traduit un comportement similaire des collaborateurs vis-à-vis des caractéristiques  $f_k$  et  $f_{k'}$ . L'estimation sur la base des corrélations, plutôt que des valeurs intrinsèques des collaborateurs permet de mettre en évidence (a) les compétences pour lesquelles un collaborateur est le plus efficace et (b) les compétences pour lesquelles il est plus efficace que son collaborateur (Soulier *et al.*, 2014a). De plus, nous ajoutons la contrainte que les corrélations doivent être estimées sur la base des caractéristiques qui ont permis de souligner des différences significatives entre les collaborateurs vis-à-vis d'une caractéristique  $f_k$ , grâce au test statistique de Kolmogorov-Smirnov ( $p(\Delta_{1,2}^{(t_l)}(f_k)) < \theta$ , avec  $\theta = 0.05$ ). Chaque élément  $C_{1,2}^{(t_l)}(f_k, f_{k'})$  de la matrice de corrélations est estimé comme suit :

$$C_{1,2}^{(t_l)}(f_k, f_{k'}) = \begin{cases} \rho(\Delta_{1,2}^{(t_l)}(f_k), \Delta_{1,2}^{(t_l)}(f_{k'})) & \text{si } p(\Delta_{1,2}^{(t_l)}(f_k)) < \theta \\ & \text{et } p(\Delta_{1,2}^{(t_l)}(f_{k'})) < \theta \\ 0 & \text{sinon} \end{cases} \quad [1]$$

#### 4.2. Apprentissage du méta-rôle en contexte collaboratif

L'enjeu principal de l'apprentissage des méta-rôles réside dans la construction de son noyau  $\mathcal{K}_{1,2}^{(t_l)} \in F^p$ . Ce dernier inclut les caractéristiques discriminantes  $f_k \in F$

qui maximisent la différence entre les collaborateurs vis-à-vis de leurs compétences afin de caractériser leurs comportements. Nous présentons le problème d'extraction des caractéristiques sous la forme d'un problème d'optimisation. Puis, nous proposons une méthode de résolution de ce problème.

#### 4.2.1. Formalisation du problème d'apprentissage du méta-rôle

Afin de construire le noyau  $\mathcal{K}_{1,2}^{(t_l)}$  du méta-rôle, nous identifions le plus petit sous-ensemble de  $p$  caractéristiques sur la base de trois hypothèses, inspirées des hypothèses formulées par Geng *et al.* (2007) et adaptées à notre contexte collaboratif :

- **H1** : l'importance des caractéristiques de comportement mesure la qualité de l'ordonnement collaboratif des documents réalisés en considérant comme source d'évidence une caractéristique de comportement. Comme proposé par Shah *et al.* (2010), nous formalisons l'ordonnement au travers d'une classification des documents basée sur les jugements de pertinence effectués par les collaborateurs depuis le début de la session jusqu'à l'instant  $t_l$  où l'on souhaite apprendre le méta-rôle. Chaque groupe de documents est associé à un utilisateur et représente les documents alloués à ce dernier. Ainsi, nous proposons de classifier, par le biais d'un algorithme de classification 2-means, l'ensemble  $\mathcal{D}^{(t_l)} = \mathcal{D}(u_1)^{(t_l)} \cup \mathcal{D}(u_2)^{(t_l)}$  des documents sélectionnés par les deux collaborateurs  $u_1$  et  $u_2$  jusqu'à l'instant  $t_l$  en considérant comme critère de classification la caractéristique  $f_k$ . La classe dont le centre de gravité est le plus élevé est assignée au collaborateur  $u_j$  caractérisé par la valeur  $S_{u_1}^{(t_l)}(f_k)$  de la caractéristique  $f_k$  la plus élevée tandis que l'autre classe est associée à l'autre collaborateur. Dans l'objectif de vérifier la qualité de la caractéristique  $f_k$  sur la classification des documents vis-à-vis des deux collaborateurs  $u_1$  et  $u_2$ , nous proposons d'estimer la mesure d'importance  $\mathcal{A}_{1,2}^{(t_l)}(f_k)$  comme suit :

$$\mathcal{A}_{1,2}^{(t_l)}(f_k) = \frac{TP1_{f_k}^{(t_l)} + TP2_{f_k}^{(t_l)}}{TP1_{f_k}^{(t_l)} + FN1_{f_k}^{(t_l)} + TP2_{f_k}^{(t_l)} + FN2_{f_k}^{(t_l)}} \quad [2]$$

où  $TPj_{f_k}^{(t_l)}$  représente le nombre de documents assignés à l'utilisateur  $u_j \in \{u_1, u_2\}$  par l'algorithme de classification basé sur la caractéristique  $f_k$  et jugés pertinents par ce même utilisateur à l'instant  $t_l$ .  $FNj_{f_k}^{(t_l)}$  correspond au nombre de documents non assignés, par classification selon le critère  $f_k$ , à l'utilisateur  $u_j \in \{u_1, u_2\}$  qui a jugé comme pertinents ces mêmes documents à l'instant  $t_l$ .

- **H2** : la redondance entre les caractéristiques doit être évitée dans l'objectif de considérer celles qui sont les plus discriminantes permettant de caractériser le méta-rôle au travers de la complémentarité des compétences de recherche des collaborateurs. Ainsi, nous nous appuyons sur la mesure de corrélation  $C_{1,2}^{(t_l)}(f_k, f_{k'})$  à partir des différences  $\Delta_{1,2}^{(t_l)}(f_k)$  et  $\Delta_{1,2}^{(t_l)}(f_{k'})$  des caractéristiques de comportement  $f_k$  et  $f_{k'}$ .

- **H3** : la sélection des caractéristiques permettant de faire émerger le méta-rôle doit maximiser l'importance  $\mathcal{A}_{1,2}^{(t_l)}(f_k)$  des caractéristiques sélectionnées  $f_k \in \mathcal{K}_{1,2}^{(t_l)}$  dans le modèle d'ordonnement collaboratif et minimiser la redondance  $C_{1,2}^{(t_l)}(f_k, f_{k'})$  entre les paires de caractéristiques sélectionnées. Ainsi, nous formalisons l'algorithme

de sélection des caractéristiques pour la construction des méta-rôles selon le modèle d'optimisation suivant :

$$\begin{aligned}
& \max_{\alpha} \quad \sum_{k=1}^n \mathcal{A}_{1,2}^{(t_l)}(f_k) \cdot \alpha_k & [3] \\
& \min_{\alpha} \quad \sum_{k=1}^n \sum_{k'=1}^n C_{1,2}^{(t_l)}(f_k, f_{k'}) \cdot \alpha_k \cdot \alpha_{k'} \\
& \text{sous la contrainte } \alpha_k = \{0, 1\}; k = 1, \dots, n \text{ et } \sum_{k=1}^n \alpha_k = p
\end{aligned}$$

avec  $\alpha = \{\alpha_1, \dots, \alpha_k, \dots, \alpha_n\}$  étant le vecteur de taille  $n$  où chaque élément  $\alpha_k$  est un indicateur booléen spécifiant si la caractéristique  $f_k$  est incluse dans le sous-ensemble de caractéristiques  $\mathcal{K}_{1,2}^{(t_l)}$  à l'instant  $t_l$ .

Ce problème d'optimisation à plusieurs objectifs peut être transformé en un problème d'optimisation à unique objectif comme suit :

$$\begin{aligned}
& \max_{\alpha} \quad \sum_{k=1}^n \mathcal{A}_{1,2}^{(t_l)}(f_k) \cdot \alpha_k - \gamma (\sum_{k=1}^n \sum_{k'=1}^n C_{1,2}^{(t_l)}(f_k, f_{k'}) \cdot \alpha_k \cdot \alpha_{k'}) & [4] \\
& \text{sous la contrainte } \alpha_k = \{0, 1\}; k = 1, \dots, n \text{ et } \sum_{k=1}^n \alpha_k = p
\end{aligned}$$

où  $\gamma$  est un paramètre de contrôle.

#### 4.2.2. Optimisation des méta-rôles

Notre problème d'optimisation défini dans l'Equation 4 peut être résolu par une simple exploration de l'ensemble des combinaisons possibles des caractéristiques de comportement de taille  $p$ , où  $p = 2, \dots, n$ . Bien que permettant d'identifier la solution optimale selon une approche exacte, cette méthode peut s'avérer coûteuse en temps avec une complexité pouvant atteindre  $O(\sum_{k=1}^p C_n^k)$ . Nous proposons dans cette contribution une approche de résolution approximée basée sur la théorie des graphes qui permet d'identifier le meilleur sous-ensemble de caractéristiques. Dans ce contexte, nous proposons de représenter les caractéristiques de comportement au travers d'un graphe de collaboration  $G_{1,2}^{(t_l)}$  qui modélise le comportement des collaborateurs  $u_1$  et  $u_2$  à l'instant  $t_l$ . Le graphe  $G_{1,2}^{(t_l)} = (A_{1,2}^{(t_l)}, C_{1,2}^{(t_l)})$  est constitué de noeuds  $A_{1,2}^{(t_l)}$  qui représentent chaque caractéristique  $f_k \in F$ , pondérée par une mesure d'importance  $\mathcal{A}_{1,2}^{(t_l)}(f_k)$  en tant que source d'évidence dans un modèle de RIC, et des relations non orientées  $C_{1,2}^{(t_l)} : F \times F$  qui représentent les similarités ou les complémentarités des comportements des collaborateurs vis-à-vis des paires de caractéristiques de comportement. Chaque relation  $(f_k, f_{k'})$  est pondérée par une mesure de corrélation  $C_{1,2}^{(t_l)}(f_k, f_{k'})$  des différences entre caractéristiques de comportement.

Dans l'objectif de résoudre notre problème d'optimisation décrit dans l'Equation 4, nous proposons un algorithme, appelé *Coll-Clique*, qui se base sur un algorithme d'extraction de clique maximale (Fahle, 2002). Cependant, au lieu de se concentrer sur le niveau des noeuds du graphe, notre méthode a pour objectif d'extraire le sous-graphe qui maximise le poids des noeuds (Hypothèse H1), à savoir l'importance des caractéristiques, et minimise la relation entre les noeuds (Hypothèse H2), à savoir la

corrélation entre paires de caractéristiques afin de résoudre notre problème d'optimisation (Hypothèse H3). Comme présenté dans l'Algorithme 1, *Coll-Clique* repose sur deux graphes : la clique en croissance  $C$ , candidate à être identifiée comme la clique maximale  $K$ , et le graphe des caractéristiques  $P$ , qui comporte les caractéristiques candidates à être ajoutées à la clique en croissance  $C$ . Les noeuds du graphe  $P$  sont extraits grâce à la fonction  $Nodes(P)$ . Initialement,  $C$  est vide et  $P$  est le graphe  $G_{1,2}^{(t_l)}$  qui inclut toutes les caractéristiques. L'algorithme est récursif et incrémente la clique en croissance  $C$  à partir des caractéristiques  $f_h$  incluses dans le graphe  $P$ . Ce dernier est extrait grâce à la fonction  $Nbhd(C)$  qui crée un nouveau graphe  $P$  comportant uniquement les caractéristiques reliées à celles impliquées dans la clique  $C$  dont le poids déprécié est positif. L'opération d'agrégation entre la clique  $C$  et une caractéristique  $f_h$  est notée  $C \oplus f_h$ . Notons  $W(K)$  la fonction qui estime la somme des poids des caractéristiques incluses dans la clique maximale  $K$ . Nous faisons l'hypothèse que le poids d'un graphe correspond à l'indicateur que nous souhaitons maximiser dans la mesure où le poids  $\mathcal{A}_{1,2}^{(t_l)}(f_h)$  des noeuds  $f_h$  est déprécié de façon récursive selon la corrélation  $C_{1,2}^{(t_l)}(f_h, f_{k'})$  des noeuds adjacents  $f_{k'}$ . Ainsi, si le poids  $W(C) + W(P)$  est plus faible que  $W(K)$ , il n'est alors pas nécessaire de lancer une nouvelle itération car l'agrégation d'une caractéristique de  $P$  avec la clique en croissance  $C$  ne permettra pas de construire une clique de poids plus élevé que celui de la clique courante  $K$ . Au final, le sous-ensemble des  $p$  caractéristiques est inféré de la clique maximale  $K$ , obtenue grâce à notre algorithme *Coll-clique*, et constitue le noyau  $\mathcal{K}_{1,2}^{(t_l)}$  du méta-rôle.

### 4.3. Ordonnancement collaboratif basé sur le méta-rôle des collaborateurs

Dans cette étape, nous réinjectons le méta-rôle identifié précédemment afin d'ordonner collaborativement les documents. Pour cela, nous utilisons le classifieur basé sur la régression logistique qui, à chaque instant  $t_l$  repose sur deux étapes :

(1) *L'étape d'apprentissage* qui estime le modèle de classification en estimant la probabilité  $P_j(x_i^{(t_l)})$  d'assigner un document  $d_i \in \mathcal{D}^{(t_l)}$  à une utilisateur  $u_j$ , modélisé par une classe  $c_j \in \{0; 1\}$  en déterminant le vecteur prédicteur  $\beta_j^{(t_l)}$  comme suit :

$$\max_{\beta_j^{(t_l)}} \sum_{d_i \in \mathcal{D}^{(t_l)}} (c_j \cdot \ln(P_j(x_i^{(t_l)})) + (1 - c_j) \ln(1 - P_j(x_i^{(t_l)}))) \quad [5]$$

$$\text{avec } P_j(x_i^{(t_l)}) = \frac{\exp(x_i^{(t_l)} \cdot \beta_j^{(t_l)})}{1 + \exp(x_i^{(t_l)} \cdot \beta_j^{(t_l)})}$$

Dans cet objectif, l'ensemble  $\mathcal{D}^{(t_l)}$  des documents sélectionnés par les deux collaborateurs  $u_1$  et  $u_2$  avant l'instant  $t_l$  sont considérés comme données d'entraînement, où chaque document  $d_i$  est modélisé par un vecteur  $x_i^{(t_l)} \in \mathcal{R}^m$  des caractéristiques du noyau  $\mathcal{K}_{1,2}^{(t_l)}$  du méta-rôle.

(2) *L'étape de test* qui considère l'ensemble  $\mathcal{D}_{n_{sel}}^{(t_l)}$  des documents non sélectionnés par les deux collaborateurs  $u_1$  et  $u_2$  avant l'instant  $t_l$ . Le modèle prédit grâce au classifieur de régression logistique, la probabilité  $P_j(x_i^{(t_l)})$  que le document  $d_i \in \mathcal{D}_{n_{sel}}^{(t_l)}$  soit alloué à la classe  $c_j$  du collaborateur  $u_j$  qui obtient la probabilité la plus élevée



---

**Algorithm 1: Coll-Clique**

---

**Données:**  $C, P, \gamma$

**Résultat:**  $K$

Fonction Coll-Clique( $C, P, \gamma$ ) {

**Pour chaque**  $f_h \in P$  **faire**

$W(\mathcal{G}) = \sum_{f_k \in Nodes(C)} \mathcal{A}_{1,2}^{(t_i)}(f_k)$  for  $\mathcal{G} = \{C, P, K\}$

**Si** ( $W(C) + W(P) \leq W(K)$ ) **alors**

        /\* Retourne la clique maximale \*/

**Retourne**  $K$

    /\* Incrémente la clique en croissance  $C$

$C = C \oplus f_h$  \*/

**Pour chaque**  $f_{k'} \in P$  **faire**

        /\* Déprécie les poids des noeuds \*/

$\mathcal{A}_{1,2}^{(t_i)}(f_{k'}) = \mathcal{A}_{1,2}^{(t_i)}(f_{k'}) - C_{1,2}^{(t_i)}(f_h, f_{k'}) * 2\gamma$

    /\* Construit l'ensemble des noeuds candidats \*/

$P' = Nbdh(C)$

**Si** ( $Nbdh(C) = \{\}$  et  $W(C) > W(K)$ ) **alors**

        /\* Stocke l'optimum local \*/

$K = C$

**Si**  $P' \neq \{\}$  **alors**

        /\* Lance une nouvelle récursion \*/

        Coll-Clique( $C, Nbdh(C), \gamma$ )

    /\* Supprime un noeud pour une nouvelle récursion \*/

$C = C \setminus f_h; P = P \setminus f_h$

}

---

$P_j(x_i^{(t_i)})$ . De plus, nous considérons la division du travail en nous assurant que les listes de documents simultanément retournées aux collaborateurs soient distinctes.

## 5. Evaluation expérimentale

Nous proposons une expérimentation du type *log-study*, reposant sur des *logs* de sessions de recherche collaborative réelles. Deux objectifs principaux sont identifiés :

- 1) Valider les hypothèses de recherche liées au comportement des collaborateurs.
- 2) Evaluer l'efficacité de notre modèle de RIC.

### 5.1. Données expérimentales

Afin de répondre aux objectifs de notre évaluation, nous avons utilisé deux jeux de données *EU1* et *EU2*, comportant respectivement 25 et 10 paires de participants. Ces derniers ont participé à une tâche de recherche exploratoire en naviguant sur le Web et en soumettant des requêtes sur des moteurs de recherche publics. Les participants devaient identifier le maximum de pages Web pertinentes en rapport avec un sujet durant une session de 30 minutes. Afin de sauvegarder les informations pendant la session de RIC, les participants disposent d'un système de collaboration implémenté

	Description	Mesure (temps exprimé en secondes)
$N_q$	Nombre de requêtes	Nombre de requêtes soumises
$L_q$	Longueur des requêtes	Nombre moyen de mots par requête
$R_q$	Recouvrement entre requêtes	Ratio moyen de mots communs entre deux requêtes successives
$N_{pq}$	Nombre de pages par requête	Nombre moyen de pages visitées par requête
$D_p$	Durée de visite	Durée moyenne d'une visite sur une page
$N_j$	Nombre de jugements de pertinence	Nombre de jugements de pertinence (annotations, notations et snippets)
$N_m$	Nombre de messages	Nombre de messages échangés entre les participants

**Tableau 1.** *Caractéristiques de comportement utilisées pour l'analyse comparative*

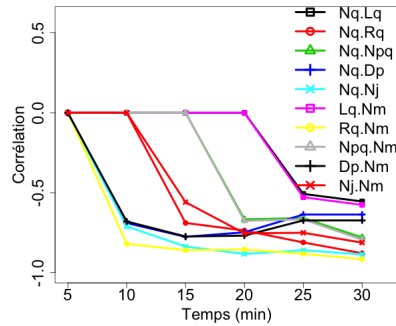
sous la forme d'une barre d'outils et de tâches compatible sous Firefox. Ce logiciel est développée à partir du code du système Coagmento (Shah et González-Ibáñez, 2011), mis à disposition<sup>1</sup>, et que nous avons adapté à nos besoins. Il est composé (a) d'une barre d'outils permettant aux collaborateurs de noter, d'annoter des pages Web ou de sauvegarder des informations (appelées *snippets*) et (b) d'une barre latérale composée d'un système de messagerie instantanée ainsi que d'un espace de visualisation des documents sélectionnés par les participants. Les protocoles des expérimentations utilisateurs étaient similaires, hormis trois différences : (1) la langue de recherche, français pour *EU1* et anglais pour *EU2* ; (2) le sujet de la tâche qui était "tempêtes tropicales" et "global warming", respectivement pour les expérimentations *EU1* et *EU2* ; et (3) les objectifs de la tâche dans la mesure où les participants à l'expérimentation *EU2* avaient également l'objectif d'écrire un rapport à partir des informations sauvegardées au cours de la session.

## 5.2. Résultats expérimentaux

### 5.2.1. Validation d'hypothèses

Notre analyse repose sur les deux hypothèses formulées dans la Section 3 : (1) les collaborateurs sont caractérisés par des comportements différents et complémentaires vis-à-vis des caractéristiques de comportement et (2) leur comportement évolue au cours de la session. Notre première expérimentation a pour objectif de valider ces hypothèses. Afin de modéliser le comportement des participants, nous avons exploité le jeu de données *EU1*. De plus, nous avons considéré les caractéristiques présentées dans le Tableau 1 qui traduisent les actions des collaborateurs aussi bien au niveau des requêtes soumises que des pages visitées et sauvegardées. Nous proposons une analyse temporelle des comportements des collaborateurs en modélisant la session de recherche des utilisateurs à partir des caractéristiques estimées à intervalles réguliers d'une minute. Afin de mettre en exergue la complémentarité des comportements des collaborateurs, nous calculons des corrélations entre les différences de leur comportement vis-à-vis des caractéristiques décrites dans le Tableau 1 et en nous basant sur l'Equation 1. Plus particulièrement, nous avons calculé, puis représenté dans la Figure 1, selon des intervalles de 5 minutes, les corrélations significatives et négatives

1. <https://github.com/InfoSeeking/CoagmentoCollaboratory>



**Figure 1.** Analyse des corrélations significatives des caractéristiques de comportement des collaborateurs pour le scénario EUI

de ces différences pour les paires de participants du scénario EUI. Par exemple, la paire de caractéristiques  $N_q - N_{pq}$  n'est pas significativement différente jusqu'à la 15<sup>ième</sup> minute mais devient négative ensuite avec des valeurs proches de  $-0.66$  jusqu'à la 25<sup>ième</sup> minutes et  $-0.77$  en fin de session. Ces variations mettent en évidence la convergence des participants dans la coordination de leurs actions dans la session en termes de compétences de recherche. Nous remarquons également que les collaborateurs convergent dans la complémentarité de leur comportement pour 8 parmi les 10 paires de caractéristiques impliquées dans des corrélations négatives 15 minutes après le début de la session. Ce constat laisse supposer que, malgré un intervalle de temps nécessaire pour identifier leurs compétences complémentaires, les utilisateurs arrivent à structurer la collaboration au fur et à mesure de leurs actions. Ces analyses nous permettent de mettre en évidence la différence et la complémentarité des collaborateurs (hypothèse 1) qui évoluent au cours du temps (hypothèse 2).

### 5.2.2. Evaluation de l'efficacité de notre modèle de médiation

#### 5.2.2.1. Protocole d'évaluation

Afin d'analyser l'efficacité de notre modèle de RIC et d'évaluer la consistance de notre modèle par rapport aux utilisateurs, à l'objectif et au sujet de la tâche, nous avons effectué une validation croisée qui (1) apprend le modèle afin d'optimiser ses paramètres à partir d'un jeu de données, e.g. EUI, et (2) teste l'efficacité du modèle sur le deuxième jeu de données, e.g. EU2, en tenant compte des valeurs optimales des paramètres identifiés lors de l'étape d'apprentissage. Compte tenu de l'objectif de notre modèle, nous proposons de considérer des caractéristiques qui traduisent implicitement la pertinence des documents afin d'effectuer un ordonnancement, comme indiqué dans le Tableau 2. Ces caractéristiques, extraites de la littérature (Agichtein *et al.*, 2006), regroupent des facteurs de comportement liés aux requêtes  $C_R$  et aux pages  $C_P$ , traduisant respectivement l'expérience de recherche et le comportement de navigation des collaborateurs.

Etant donné la distinction des langues des participants, nous avons construit deux index de documents. Pour chaque jeu de données, nous avons agrégé les pages Web

	Description		Mesure (temps exprimé en secondes)
$C_R$	$R_{Ti}$	Recouvrement Titre- Requête	Ratio moyen de mots communs entre le titre de la page et la requête
	$R_{Te}$	Recouvrement Texte- Requête	Ratio moyen de mots communs entre le texte de la page et la requête
	$R_{An}$	Recouvrement Annotation- Requête	Ratio moyen de mots communs entre l'annotation de la page et la requête
	$R_{Sn}$	Recouvrement Snippet- Requête	Ratio moyen de mots communs entre le snippet de la page et la requête
	$RV_R$	Rang de visite	Rang de visite de la page par rapport à la requête
	$D_{PR}$	Durée page-Requête	Temps entre la soumission de la requête et la visite de la page
$C_P$	$D_p$	Durée de visite	Durée moyenne d'une visite sur une page
	$D_{Do}$	Durée de visite du domaine	Durée cumulée du temps de visite d'un domaine
	$Lis$	Lisibilité	Lisibilité du contenu du document
	$Spec$	Spécificité	Spécificité du contenu du document
	$Note$	Note	Note de la page

**Tableau 2.** *Caractéristiques de comportement*

visitées par les participants ainsi que les 100 premières pages retournées par le moteur de recherche Google pour les requêtes soumises par les participants. Dans l'objectif d'augmenter la taille des index, nous avons propagé ce protocole à d'autres expérimentations, non prises en compte dans ce travail. Au final, nous avons obtenu deux index de 24 266 et 74 844 documents, respectivement pour les expérimentations *EU1* et *EU2*. L'ensemble des documents pertinents a été construit à partir des actions des participants, à savoir les pages Web annotées/commentées/sauvegardées, en ajoutant une contrainte d'accord en s'assurant qu'au moins deux participants ont validé ces pages. Nous obtenons ainsi 38 et 20 documents pertinents pour les expérimentations utilisateurs *EU1* et *EU2* respectivement.

Afin de comparer l'efficacité de notre modèle, nous avons généré, pour chaque requête, les ordonnancements associés aux modèles de référence suivants :

- *BM25-RIC* : le modèle probabiliste BM25 qui est apparenté à un modèle individuel.
- *Logit-RIC* : une version de notre modèle de RIC basé sur la dernière étape d'ordonnement collaboratif en considérant toutes les caractéristiques de comportement (Tableau 2) comme facteur de pertinence, plutôt que le méta-rôle des collaborateurs.
- *PM-RIC* : le modèle de RIC basé sur les rôles prédéfinis de Prospecteur et de Mineur (Pickens *et al.*, 2008).
- *GS-RIC* : le modèle de RIC basé sur les rôles prédéfinis de Rassembleur et de Surveillleur (Shah *et al.*, 2010).
- *MineRank* : notre modèle collaboratif basé sur les méta-rôles. Lorsque les utilisateurs ne sont pas significativement différents vis-à-vis des caractéristiques, nous utilisons le modèle *BM25-RIC* similairement à Soulier *et al.* (2014a).

Nous avons utilisé les mesures d'évaluation orientées collaboration, définies par Shah et González-Ibáñez (2011) et décrites dans le Tableau 3. Ces mesures ont été appliquées sur les ordonnancements au rang  $R = 20$ . Les notations utilisées sont les

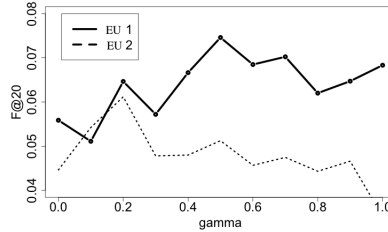
suivantes :  $Cov@R(g)$ , resp.  $RelCov@R(g)$ , correspond au nombre total de documents distincts, resp. distincts pertinents, parmi les  $R$  retournés sur l'ensemble de la session au groupe  $g$ .  $|RelDoc|$  exprime le nombre total de documents pertinents.

Métrique	Description
Précision	$Prec@R(g) = \frac{RelCov@R(g)}{Cov@R(g)}$
Rappel	$Rappel@R(g) = \frac{RelCov@R(g)}{ RelDoc }$
F-mesure	$F@R(g) = \frac{2 * Prec@R(g) * Rappel@R(g)}{Prec@R(g) + Rappel@R(g)}$

**Tableau 3.** Mesures d'évaluation orientées collaboration

### 5.2.2.2. Paramétrage et analyse comparative de notre modèle

Afin d'évaluer la valeur optimale du paramètre  $\gamma$  utilisé dans l'étape d'apprentissage des méta-rôles (Equation 4) de notre modèle *MineRank*, nous considérons la F-mesure au rang 20 ( $F@20$ ), comme illustré dans la Figure 2. L'amplitude de la valeur du paramètre  $\gamma$  est définie telle que  $\gamma \in [0..1]$  et obtient une valeur maximale à 0.5 et 0.2 respectivement pour les expérimentations utilisateurs *EU1* et *EU2* avec un score respectif de 0.074 et 0.060.



**Figure 2.** Paramétrage de notre modèle *MineRank*

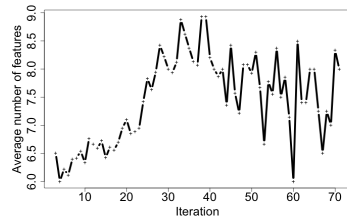
Le Tableau 4 présente les valeurs obtenues pour notre modèle *MineRank* et les modèles de référence *BM25-RIC*, *Logit-RIC*, *GS-RIC* et *PM-RIC*. Nous soulignons le fait que les mesures d'efficacité sont faibles, compte tenu du faible nombre de documents pertinents pour les deux expérimentations utilisateurs. Nous observons que notre modèle fournit généralement de meilleurs résultats que les modèles de référence. De plus, les améliorations obtenues grâce à la mesure de précision, par rapport à la mesure de rappel, prouvent que notre modèle est plus particulièrement orienté rappel. Une analyse plus fine des résultats montre que, pour le jeu de données *EU1*, notre modèle dépasse significativement les modèles de référence orientés RI individuelle et collaborative sur l'ensemble des trois mesures d'évaluation, avec des augmentations comprises entre 43.24% et 650.15%. Par contre, pour l'expérimentation utilisateur *EU2*, les résultats sont significatifs seulement par rapport au modèle *BM25-RIC* sur les trois mesures et au modèle *PM-RIC* pour la mesure de rappel. Ces résultats attestent de trois conclusions principales : (1) notre modèle permet de bénéficier de l'effet synergique de la collaboration en fournissant de meilleurs résultats que le modèle individuel *BM25-RIC* ; (2) l'ordonnancement collaboratif sur la base des méta-rôles des collaborateurs bénéficie d'une valeur ajoutée par rapport à un modèle collaboratif

<i>Apprentissage</i>		<i>Prec@20</i>		<i>Rappel@20</i>		<i>F@20</i>	
$\rightarrow$ <i>Test</i>		valeur	%Tx	valeur	%Tx	valeur	%Tx
<i>EU2</i>	BM25-RIC	0,009	285,79 ***	0,066	236,51 ***	0,017	280,84 ***
	Logit-RIC	0,031	20,66	0,155	43,24 *	0,052	23,64
$\rightarrow$ <i>EU1</i>	GS-RIC	0,009	306,78 ***	0,045	400,10 ***	0,015	324,29 ***
	PM-RIC	0,014	178,55 ***	0,029	650,15 ***	0,018	254,85 ***
	<b>MineRank</b>	<b>0,038</b>		<b>0,223</b>		<b>0,065</b>	
<i>EU1</i>	BM25-RIC	0,015	85,92 *	0,163	74,19 *	0,027	85,29*
	Logit-RIC	0,025	10,67	0,252	12,50	0,046	10,92
$\rightarrow$ <i>EU2</i>	GS-RIC	0,030	-6,24	0,184	54,28	0,051	0
	PM-RIC	0,056	-50,04 ***	0,205	38,46*	0,088	-41,87 ***
	<b>MineRank</b>	<b>0,028</b>		<b>0,284</b>		<b>0,051</b>	

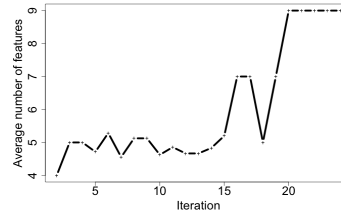
**Tableau 4.** Comparaison de notre modèle de RIC basé sur les méta-rôles en terme d'efficacité. %Tx : amélioration de MineRank. Test de significativité de student \* :  $0.01 < t \leq 0.05$ ; \*\* :  $0.001 < t \leq 0.01$ ; \*\*\* :  $t \leq 0.001$ .

basé sur une analyse intrinsèque des comportements utilisateurs (modèle *Logit-RIC*) ; (3) l'apprentissage du méta-rôle des collaborateurs semble plus efficace qu'un modèle de RIC basé sur des rôles prédéfinis (modèles *GS-RIC* et *PM-RIC*). Cependant, une analyse des deux jeux de données souligne des améliorations plus faibles pour *EU2*, et particulièrement des diminutions significatives par rapport au modèle de référence *PM-CIR* pour la précision et la F-mesure. Cela peut être expliqué par le fait que, pour cette expérimentation utilisateur, le paramétrage est réalisé sur un jeu de données plus important (*EU1*), pouvant provoquer un sur-apprentissage du modèle en termes de caractéristiques sélectionnées et du paramètre  $\gamma$ , et, par conséquent, impactant sur l'efficacité globale du modèle.

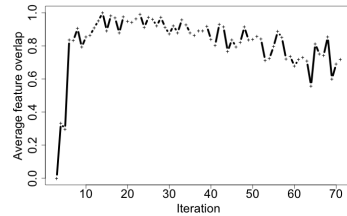
Afin de comprendre cette dernière observation, nous analysons le nombre de caractéristiques sélectionnées pour construire les noyaux des méta-rôles des collaborateurs. L'intuition est d'étudier la stabilité des méta-rôles grâce au taux de recouvrement des caractéristiques entre deux méta-rôles successifs. Ces indicateurs sont estimés à chaque soumission de requête, appelée aussi itération, comme illustré dans la Figure 3. Nous soulignons le fait que seulement 3 groupes sur 25 ont soumis au delà de 60 requêtes dans le jeu de données *EU1*, de même pour le jeu de données *EU2* avec un seul groupe après 18 requêtes. Au delà, nous considérons que les variations ne sont pas significatives. La Figure 3(a), respectivement la Figure 3(b), montre que le nombre de caractéristiques augmente avec le temps et semble atteindre un optimum entre 8 et 9, respectivement 7, caractéristiques parmi les 11 considérées, pour l'expérimentation utilisateurs *EU1*, respectivement *EU2*. Ce résultat laisse penser que le sur-apprentissage notifié précédemment peut être expliqué par la quantité de caractéristiques sélectionnées pour construire le méta-rôle des collaborateurs. Plus particulièrement, les meilleurs résultats obtenus avec les modèles *GS-RIC* et *PM-RIC* par rapport à notre modèle *MineRank* sont expliqués par le fait que ces modèles restreignent la complémentarité des compétences à un nombre limité de caractéristiques, à l'inverse du modèle *MineRank* qui propose une approche générique. Cette intuition est renforcée par le modèle *Logit-RIC*, qui implique l'ensemble des caractéristiques



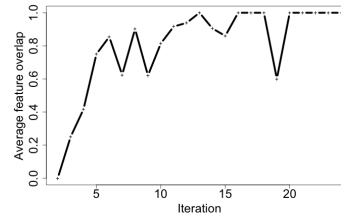
(a) Nombre moyen de caractéristiques pour EU1



(b) Nombre moyen de caractéristiques pour EU2



(c) Recouvrement moyen entre les ensembles de caractéristiques pour EU1



(d) Recouvrement moyen entre les ensembles de caractéristiques pour EU2

**Figure 3.** Analyse temporelle des caractéristiques de comportement impliquées dans l'étape d'ordonnancement collaboratif sur l'ensemble des sessions de recherche

et dont les résultats sont plus faibles que notre modèle *MineRank*. Dans les Figures 3(c) et 3(d), la mesure de recouvrement est proche de 1 après 20 requêtes soumises, ce qui indique que les participants adoptent naturellement la coordination en accord avec leur meilleure complémentarité des compétences. Avant cet instant, les participants ont besoin de temps pour déterminer leur stratégie de collaboration, faisant varier la structure du noyau du méta-rôle. Ainsi, ces conclusions donnent de la crédibilité à notre contribution qui a pour objectif d'adapter les ordonnancements à la complémentarité évolutive des compétences des collaborateurs.

## 6. Conclusion et perspectives

Dans ce papier, nous proposons une approche non-supervisée de RIC basée sur le comportement des utilisateurs qui apprend leur méta-rôle et le réinjecte dans le modèle d'ordonnancement collaboratif des documents. Les expérimentations basées sur des fichiers *log* de sessions de recherche collaborative réelles permettent de valider des hypothèses de comportement des collaborateurs et montrent l'efficacité de notre modèle. Ce travail comporte quelques limitations que nous souhaitons lever par la suite. Nous avons considéré un contexte de collaboration synchrone limité dans le temps et adapté à une paire de collaborateurs. Nous souhaitons élargir le cadre d'application en terme de taille du groupe et de la tâche de collaboration, telle qu'une tâche de type *fact-finding*. De plus, la prise en compte des actions des utilisateurs pourraient être complétée par des facteurs liés à des préférences à long-terme par la prise en compte de tâches de RI à multiples sessions.

## 7. Bibliographie

- Agichtein E., Brill E., Dumais S., Ragno R., « Learning User Interaction Models for Predicting Web Search Result Preferences », *SIGIR*, p. 3-10, 2006.
- Fahle T., « Simple and Fast : Improving a Branch-And-Bound Algorithm for Maximum Clique », *Algorithms - ESA 2002*, vol. 2461, p. 485-498, 2002.
- Foley C., Smeaton A. F., « Synchronous Collaborative Information Retrieval : Techniques and Evaluation », *ECIR*, p. 42-53, 2009.
- Geng X., Liu T.-Y., Qin T., Li H., « Feature Selection for Ranking », *SIGIR*, p. 407-414, 2007.
- Golder S. A., Donath J., « Social Roles in Electronic Communities », *Proc. of Association of Internet Researchers Conference*, 2004.
- Golovchinsky G., Qvarfordt O., Pickens J., « Collaborative Information Seeking », *IEEE Computer*, vol. 42, n° 3, p. 47-51, 2009.
- Henderson K., Gallagher B., Eliassi-Rad T., Tong H., Basu S., Akoglu L., Koutra D., Faloutsos C., Li L., « RoIX : Structural Role Extraction & Mining in Large Graphs », *SIGKDD*, p. 1231-1239, 2012.
- Joachims T., « Optimizing Search Engines Using Clickthrough Data », *SIGKDD*, p. 133-142, 2002.
- Kelly R., Payne S. J., « Division of Labour in Collaborative Information Seeking : Current Approaches and Future Directions », *CIS Workshop in CSCW*, 2013.
- Kwak H., Lee C., Park H., Moon S., « What is Twitter, a social network or a news media ? », *WWW*, p. 591-600, 2010.
- Lorigo L., Haridasan M., Brynjarsdóttir H., Xia L., Joachims T., Gay G., Granka L. A., Pellacini F., Pan B., « Eye tracking and online search : Lessons learned and challenges ahead », *JASIST*, vol. 59, n° 7, p. 1041-1052, 2008.
- McCallum A., Wang X., Corrada-Emmanuel A., « Topic and Role Discovery in Social Networks with Experiments on Enron and Academic Email », *J. Artif. Int. Res.*, vol. 30, n° 1, p. 249-272, 2007.
- Morris M. R., Horvitz E., « SearchTogether : an interface for collaborative web search », *UIST*, p. 3-12, 2007.
- Pal A., Counts S., « Identifying topical authorities in microblogs », *WSDM*, p. 45-54, 2011.
- Pickens J., Golovchinsky G., Shah C., Qvarfordt P., Back M., « Algorithmic mediation for collaborative exploratory search », *SIGIR*, p. 315-322, 2008.
- Shah C., González-Ibáñez R., « Evaluating the synergic effect of collaboration in information seeking », *SIGIR*, p. 913-922, 2011.
- Shah C., Pickens J., Golovchinsky G., « Role-based results redistribution for collaborative information retrieval », *IP&M*, vol. 46, n° 6, p. 773-781, 2010.
- Sonnenwald D. H., Maglaughlin K. L., Whitton M. C., « Designing to support situation awareness across distances : an example from a scientific collaboratory », *IP&M*, vol. 40, n° 6, p. 989 - 1011, 2004.
- Soulier L., Shah C., Tamine L., « User-driven System-mediated Collaborative Information Retrieval », *SIGIR*, p. 485-494, 2014a.
- Soulier L., Tamine L., Bahsoun W., « On domain expertise-based roles in collaborative information retrieval », *IP&M*, vol. 50, n° 5, p. 752-774, 2014b.