HAL
open science

L'intégrale de Choquet discrète pour l'agrégation de pertinence multidimensionnelle

Bilel Moulahi, Lynda Tamine, Sadok Ben Yahia

To cite this version:

Bilel Moulahi, Lynda Tamine, Sadok Ben Yahia. L'intégrale de Choquet discrète pour l'agrégation de pertinence multidimensionnelle. Conférence francophone en Recherche d'Information et Applications - CORIA 2013, Apr 2013, Neuchâtel, Suisse. pp. 1 -16. hal-01143227

HAL Id: hal-01143227

https://hal.science/hal-01143227v1

Submitted on 17 Apr 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Open Archive TOULOUSE Archive Ouverte (OATAO)
OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : http://oatao.univ-toulouse.fr/
Eprints ID : 12662

To cite this version : Moulahi, Bilel and Tamine, Lynda and Ben Yahia, Sadok *L'intégrale de Choquet discrète pour l'agrégation de pertinence multidimensionnelle*. (2013) In: Conférence francophone en Recherche d'Information et Applications - CORIA 2013, 3 April 2013 - 5 April 2013 (Neuchâtel, Switzerland).

Any correspondance concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

L'intégrale de Choquet discrète pour l'agrégation de pertinence multidimensionnelle

Bilel Moulahi[*,**] — **Lynda Tamine**[*] — **Sadok Ben Yahia**[**,***]

[*] *Université Paul Sabatier, Institut de Recherche en Informatique de Toulouse, France*

bilel.moulahi@irit.fr, lynda.tamine@irit.fr

[**] *Université de Tunis El Manar, Faculté des Sciences de Tunis, LIPAH, 2092, Tunis, Tunisie*

sadok.benyahia@fst.rnu.tn

[***] *Institut Mines-TELECOM, TELECOM SudParis, UMR CNRS Samovar, 91011 Evry Cedex, France*

RÉSUMÉ. *Dans ce papier, nous nous intéressons à étudier le problème de l'agrégation multi-critères dans le domaine de la recherche d'information (RI). Nous proposons une nouvelle approche basée sur l'intégrale de Choquet pour l'agrégation de pertinence multidimensionnelle. La principale originalité de cet opérateur, outre sa capacité à modéliser des interactions entre les différentes dimensions de pertinence, est sa capacité à généraliser de nombreuses fonctions d'agrégation classiques. L'évaluation de l'efficacité de notre approche est effectuée dans une tâche de recherche de tweets, où les critères conjointement utilisés sont, la pertinence thématique, l'autorité et la fraîcheur. Les résultats expérimentaux obtenus sur la collection de test fournie par la tâche Microblog de TREC* 2011 *montrent la pertinence de notre proposition.*

ABSTRACT. *In this paper, we are particularly interested in addressing the multi-criteria aggregation problem in the information retrieval (IR) field. We propose a Choquet integral based approach for multidimensional relevance aggregation. The main originality of this operator is its capability in modelling interactions between the different relevance dimensions. The evaluation of the appropriateness of the method is carried out within the TREC Microblog search task, where topicality, authority and recency criteria are jointly considered. Experimental results using TREC 2011 Microblog track dataset emphasize the relevance of our proposal.*

MOTS-CLÉS : *Agrégation, pertinence multidimensionnelle, intégrale de Choquet, mesure floue.*

KEYWORDS: *Aggregation, multidimensional relevance, Choquet integral, fuzzy measure.*

1. Introduction

Le problème de la combinaison multicritère a été abondamment étudié dans plusieurs domaines de recherche tels que la théorie du choix social (Arrow, 1974), les problèmes de prise de décision multicritères (Steuer, 1986) et la fusion de donnés (Ah-Pine, 2008). La combinaison multicritères s'impose généralement lorsque, pour une tâche donnée, nous disposons d'un ensemble d'alternatives qui devraient être jugées par rapport à un nombre bien défini de critères, et nous sommes amenés à produire un score global pour chacune de ces alternatives selon chaque critère. Ce problème, appelé également agrégation *multidimensionnelle*, a récemment connu une attention particulière dans le domaine de la recherche d'information (RI). Son émergence dans la communauté de RI est liée principalement au caractère multidimensionnel du concept de pertinence. En effet, de nombreux travaux (Borlund, 2003, Taylor *et al.,* 2007, Saracevic, 2007) ont identifié plusieurs critères, au-delà de la dimension thématique, qui peuvent affecter la perception de pertinence des utilisateurs. Une direction de recherche prometteuse, qui a émergé par conséquence, consiste à considérer, à la fois, tous les critères ayant un impact sur la notion de pertinence et pouvant influencer les jugements de pertinence des utilisateurs, afin d'améliorer la précision des systèmes de RI. Ceci s'est imposé de manière cruciale dans de nombreuses applications tels que la RI mobile (Cong *et al.,* 2009, Bouidghaghen *et al.,* 2011b), la RI sociale (Ounis *et al.,* 2011, Chen *et al.,* 2012) et la RI personnalisée (Daoud *et al.,* 2010). Ces travaux ont mis l'accent sur l'utilité de considérer plusieurs dimensions de pertinence autres que la dimension *thématique*, et ont proposé différentes stratégies de combinaison de critères. Plus récemment, d'autres approches se sont intéréssées à l'aggrégation de pertinence multidimensionnelle en RI (da Costa Pereira *et al.,* 2012, Gerani *et al.,* 2012) et ont proposé des cadres de combinaison de pertinence. Une autre direction de recherche qui s'intéresse aussi à l'agrégation multicritères en RI, appelée agrégation de *"ranking"*, s'est aussi récemment imposée dans la communauté de RI (Farah *et al.,* 2007, Wei *et al.,* 2010). La différence entre cette dernière et l'agrégation de pertinence est le fait qu'en agrégation de *"ranking"*, l'aspect multidimensionnel de la pertinence est quasiment ignoré et la combinaison ne repose, très souvent, que sur le critère *thématique*, en se basant sur un ensemble hétérogène d'ordonnancements de documents.

Dans cet article, nous proposons une approche basée sur l'intégrale de Choquet discrète pour l'agrégation de pertinence multidimensionnelle. Fondée sur les intégrales floues et basée sur le concept de la mesure floue, la principale originalité de l'intégrale de Choquet, en plus de sa généralisation des opérateurs d'agrégation classiques tels que les moyennes arithmétiques et ses variantes, réside dans sa capacité à modéliser toutes les interactions et les dépendances qui peuvent exister entre les différentes dimensions de pertinence. Les contributions majeures présentées dans ce papier sont comme suit :

– Proposition d'une nouvelle approche multicritère basée sur l'intégrale de Choquet pour l'agrégation de pertinence multidimensionnelle dans le domaine de RI.

– L'évaluation de l'opérateur proposé dans une tâche spécifique de recherche de *tweets* (Ounis *et al.,* 2011), où les critères conjointement considérés sont : la dimension *thématique de recherche*, la *fraîcheur* des *tweets* et l'*autorité* des auteurs. Notre approche est comparée à d'autres mécanismes d'agrégation multicritères telles que la moyenne arithmétique, la moyenne arithmétique pondérée et la stratégie de combinaison linéaire (Vogt *et al.,* 1999). Les expérimentations ont été menées sur la collection de test fournie par la tâche Microblog de TREC 2011.

La suite du papier est organisée comme suit, la section 2 présente une synthèse des travaux antérieurs sur la pertinence multidimensionnelle et donne un aperçu des opérateurs d'agrégation multicritères proposés. Notre approche d'agrégation de pertinence multidimensionnelle basée sur l'intégrale de Choquet est décrite dans la section 3. Dans la section 4, nous présentons les résultats expérimentaux obtenus et nous discutons l'efficacité de notre opérateur d'agrégation dans la tâche de recherche de *tweets*. La section 5 conclut le papier et donne un aperçu des travaux futurs.

2. De la pertinence thématique à la pertinence multidimensionnelle : constat et synthèse des travaux

2.1. *Constat*

La littérature concernant le domaine de la RI a connu un très grand nombre de publications portant sur le concept clé de pertinence durant les deux dernières décennies (Saracevic *et al.,* 1974, Schamber, 1991, Barry, 1994, Mizzaro, 1998, Cosijn *et al.,* 2000, Saracevic, 2007). La dimension de pertinence la plus couramment utilisée est la pertinence thématique qui traduit la proximité des sujets des requêtes et de documents. Ce constat est revendiqué dans plusieurs travaux qui ont souligné que le critère thématique est la dimension principale de pertinence et que tous les autres critères en sont dépendants (Saracevic, 2007). L'évaluation de la pertinence des documents est alors basée sur le calcul des scores individuels séparément selon un critère unique. Cependant, un nombre considérable d'études sur la pertinence ont mis l'accent sur l'aspect multidimensionnel de ce concept (Saracevic *et al.,* 1974, Schamber, 1991, Barry, 1994, Cosijn *et al.,* 2000, Borlund, 2003, Duan *et al.,* 2010, Saracevic, 2007). Ces études montrent qu'outre la dimension thématique, qui joue un rôle important, mais non exclusif, dans les jugements de pertinence des utilisateurs, un certain nombre de facteurs ou d'autres attributs entrent en jeu dans le processus de jugement. Ce constat est expérimenté dans de nombreux travaux et a fait l'objet de plusieurs applications réelles et récentes telles que la RI sociale (Becker *et al.,* 2011, Ounis *et al.,* 2011, Chen *et al.,* 2012), la RI mobile (Cong *et al.,* 2009, Bouidghaghen *et al.,* 2011b) et la RI personalisée (Gauch *et al.,* 2003, Daoud *et al.,* 2010, Bouidghaghen *et al.,* 2011a). Parmi ces critères, on cite à titre d'exemple, la *fraîcheur de l'information*, les *centres d'intérêt*, la *géolocalisation*, l'*autorité* des auteurs. (Cooper, 1973) a indiqué que pour juger un document, plusieurs propriétés et critères pourraient être considérés à l'égard de l'utilisateur tels que l'*originalité (novelty)*, le *caractère informatif (informativeness)*, la *crédibilité*, l'*importance* et la *clarité*. Barry a conduit une

étude expérimentale dans laquelle elle a identifié environ 23 catégories de pertinence, chacune d'entre elles comprend plusieurs facteurs affectant la perception de pertinence des utilisateurs. (Cosijn *et al.,* 2000) affirment que durant une interaction utilisateur avec un système de RI, la pertinence peut être affectée par la dimension temporelle. Les auteurs ont tracé une table d'attributs où ils ont classé la pertinence en trois catégories : *topicale, cognitive, socio-cognitive*. Récemment, Saracevic montre, dans son travail pionnier (Saracevic, 2007), la présence de plusieurs critères qui entrent en jeu dans le processus de jugement de pertinence des documents, telles que la *validité*, la *correspondance cognitive*, ou des facteurs basés sur le contenu comme la *topicalité*, la *qualité* et la *clarité*. L'auteur a montré aussi que ces critères ne sont pas indépendants et que la dimension *thématique* s'est vue attribuer la plus grande importance mais elle reste toujours en interaction avec les autres critères. (Pasi *et al.,* 2007) ont proposé un système multicritères pour le filtrage d'information. Le système PENG proposé est basé sur des critères tels que le *contenu*, le *coverage*, la *fiabilité*, la *nouveauté* et la *timeliness*. (Duan *et al.,* 2010) ont exploré les effets des critères sur les scores globaux de pertinence des documents dans *Twitter*. Ils ont identifié les critères basés sur le *contenu* et d'autres qui ne le sont pas comme l'*autorité*, la *longueur des tweets* et la présence ou non des *URL*s, et ont analysé l'effet de chaque facteur à part.

La considération de l'aspect multidimensionnel de la pertinence pose alors le problème d'identifier une méthode de combinaison efficace des différentes dimensions pour aboutir aux scores globaux de pertinence. On distingue deux directions de recherche dans ce sens, la première s'intéresse à l'agrégation d'ordonnancements (Farah *et al.,* 2007, Wei *et al.,* 2010), et la deuxième s'intéresse à l'agrégation multidimensionnelle de pertinence. Tandis que la première catégorie des travaux repose, en très grande partie, sur la pertinence *thématique* et se base sur la combinaison des listes ordonnées des résultats de recherche depuis un nombre de système de RI différents, la deuxième catégorie exploite l'aspect multidimensionnel de pertinence au-delà des facteurs *thématiques* classiques. Les deux catégories ont proposé des stratégies de combinaison et des fonctions d'agrégation appropriées. Dans cette section, nous focaliserons sur les les travaux traitant de la pertinence multidimensionnelle.

2.2. *Agrégation de pertinence multidimensionnelle*

Tandis que les recherches antérieures sur la pertinence ont été axées sur ce concept du point de vue *thématique* (*topical*), les recherches les plus récentes ont mis l'accent sur plusieurs dimensions différentes : représentation multidimensionnelle côté utilisateur (profil, centres d'intérêts), environnement de recherche (dispositifs utilisés, localisation géographique) et aspect temporels (Cooper, 1973, Barry, 1994, Cosijn *et al.,* 2000, Borlund, 2003, da Costa Pereira *et al.,* 2012, Taylor, 2012, Gerani *et al.,* 2012). Les études les plus récentes dans ce cadre s'imposent de manière cruciale dans de nombreuses applications telles que la RI mobile (Bouidghaghen *et al.,* 2011b, Cong *et al.,* 2009), la RI sociale (Ben Jabeur *et al.,* 2010, Becker *et al.,* 2011, Ounis *et al.,* 2011, Chen *et al.,* 2012) et la RI personnalisée (Gauch *et al.,* 2003, Daoud *et al.,*

2010, Bouidghaghen *et al.,* 2011a, da Costa Pereira *et al.,* 2012). Ces travaux ont mis l'accent sur l'utilité de considérer plus qu'un critère dans les jugements de pertinence et ont suggéré divers opérateurs d'agrégation. Parmi ces mécanismes d'agrégation, nous citons principalement la méthode de combinaison linéaire, le produit de facteurs, les opérateurs arithmétiques pondérés et les opérateurs d'agrégation prioritaires. Ainsi, le type d'agrégation le plus communément utilisé est le mécanisme de combinaison linéaire. Parmi les approches s'inscrivant dans ce cadre, le travail de (Gauch *et al.,* 2003), dans lequel les auteurs ont proposé un modèle de combinaison des scores originaux et des scores personnalisés des documents, calculés en fonction de leurs degré de similarité avec les profils utilisateurs représentant ses *centres d'intérêts*. La méthode d'agrégation utilisée est une combinaison linéaire des deux dimensions ainsi citées. (Cantera *et al.,* 2008) proposent d'utiliser le modèle MCI (Multiplicative Competitive Interaction) comme modèle de combinaison des scores de la correspondance thématique d'un document, de la *localisation géographique* de l'utilisateur et de ses *centres d'intérêts* et ce, dans un cadre de RI mobile. Ainsi, l'expression générale d'utilité d'un document selon le modèle MCI est donnée par une combinaison linéaire des scores individuels. Les attributs contextuels considérés sont principalement la *localisation* et le contexte du *dispositif mobile* qui sont combinés avec le score textuel des documents. Dans la même direction de recherche, (Cong *et al.,* 2009) ont proposé un modèle de RI basé sur la *localisation géographique* et le critère *thématique* dans lequel les documents sont classés par le biais d'une combinaison linéaire des deux dimensions. Plus tard, (Gerani *et al.,* 2012) ont proposé un mécanisme de combinaison linéaire pour l'agrégation de pertinence multicritères. La méthode proposée permet de générer des scores globaux de pertinence qui ne requierent pas le fait que les scores individuels à combiner ne soient pas forcément comparables. Les auteurs se sont basés sur l'algorithme d'Espérance Conditionnelle Alternée et l'algorithme BoxCox pour analyser le problème d'incomparabilité et effectuer une transformation de score quand il est nécessaire. Parmi les approches utilisant un produit de facteurs pour l'agrégation des différents critères de pertinence, le travail de (White *et al.,* 2005), dans lequel les auteurs exploitent le profil utilisateur dans la chaîne d'accès à l'information. Pour intégrer les centres d'intérêts dans la fonction d'appariement du modèle de RI et calculer un score global de pertinence, les auteurs utilisent un produit de facteurs entre le score de correspondance des *centres d'intérêts* et le score de correspondance *thématique* entre la requête utilisateur et les documents. Par ailleurs, (Bouidghaghen *et al.,* 2011b) ont suggéré un modèle multicritères de pertinence basé sur trois dimensions de pertinence à savoir, la *thématique* de recherche, les *centres d'intérêts* et la *localisation géographique*. Pour agréger ces dernières, les auteurs ont exploité des opérateurs d'agrégation prioritaires définissant une sorte de priorité sur les trois critères. Plus tard, dans la même lignée, (da Costa Pereira *et al.,* 2012) ont proposé une représentation multidimensionnelle de la pertinence et ont exploité 4 critères : *aboutness*, *coverage*, *appropriateness* et *reliability*. Les auteurs ont aussi utilisé un schéma d'agrégation prioritaire en se basant sur deux opérateurs à savoir, *and* et *scoring*.

Dans ce papier, nous proposons une nouvelle approche pour l'agrégation de pertinence multidimensionnelle en se basant sur l'intégrale de Choquet discrète. Cette

approche, qui en plus d'être générale, est capable d'échapper au problème d'indépendance des critères posée dans les opérateurs d'agrégation classiques. À notre connaissance, ce type d'agrégation floue n'a jamais été proposé dans un cadre applicatif de RI.

3. L'intégrale de Choquet pour l'agrégation de pertinence multidimensionnelle

3.1. *Formalisation du problème et motivations*

Nous considérons le problème d'agrégation de pertinence multidimensionnelle comme un problème de prise de décision multicritère. Plus précisément, dans le contexte de RI, nous sommes confrontés à trouver un consensus sur le classement d'un ensemble de documents d_j à partir d'un ensemble total $\mathcal{D} = \{d_1, d_2, \ldots, d_M\}$ selon un ensemble de critères c_i, en réponse à une requête utilisateur. Chaque critère c_i appartient à l'ensemble $\mathcal{C} = \{c_1, c_2, \ldots, c_n\}$ des dimensions de pertinence. La combinaison des scores partiels des documents, obtenus sur chaque dimension de pertinence est appelée *agrégation*. Ceci est effectué par le biais d'un opérateur d'agrégation approprié ou une fonction ayant, généralement, la forme suivante :

$$\mathcal{F} : \begin{cases} \mathbb{R}^N \longrightarrow \mathbb{R} \\ (C_{1j} \times C_{2j} \times \ldots \times C_{Nj}) \longrightarrow \mathcal{F}(C_{1j}, C_{2j}, \ldots, C_{Nj}) \end{cases}$$

où $\mathcal{F}$ est la fonction d'agrégation calculant les scores globaux des documents et C_{ij} (ou $C_i(d_j)$) est le score partiel obtenu sur le document d_j selon le critère c_i. C_{ij} peut être interprété comme le degré de satisfaction du document d_j selon le critère c_i. En outre, des préférences sur les critères peuvent être considérées, ces préférences peuvent être exprimées selon un vecteur de poids normalisés $W = (w_l, w_2, \ldots, w_n)$ (avec $0 \leq w_i \leq 1$) ou par une relation de préférence binaire $\succ_C$. La relation $d_1 \succ_C d_2$ peut être interprétée par *"d_1 est plus pertinent que d_2 selon l'ensemble $\mathcal{C}$ des dimensions de pertinence"*[1]. Nous montrons à travers l'exemple qui suit une des insuffisances de ces mécanismes d'agrégation basés sur la combinaison linéaire.

Exemple *Considérons un problème de prise de décision multicritère avec deux critères (c_1, c_2) et trois documents d_1, d_2 et d_3 ayant les scores partiels suivants :*
$C_1(d_1) = 0.45$, $C_1(d_2) = 0$, $C_1(d_3) = 1$
$C_2(d_1) = 0.45$, $C_2(d_2) = 1$, $C_2(d_3) = 0$.
Supposons que ces scores varient entre 0 et 1 et que le document d_1 est préféré à d_2 et d_3 (i.e., $d_1 \succ d_2 \sim d_3$)[2]*. La question majeure qui se pose ici est : est ce que les*

1. De la même façon, une relation de préférence partielle $\succ_{c_i}$ peut être aussi introduite sur l'ensemble des documents. En effet, $d_1 \succ_{c_i} d_2$ peut être interprété par : *"d_1 est plus pertinent que d_2 selon la dimension de pertinence c_i"*
2. Ceci dit, les documents qui sont jugés de façon balancée, en terme de score, sur les deux critères, sont préférés aux documents satisfaisant un seul critère.

opérateurs de combinaison linéaire et les moyennes aritmétiques pondérées sont en mesure de modéliser ce type de préférence ?
Pour ce faire, il serait nécessaire de trouver les poids w_1 et w_2 de c_1 et c_2 respectivement, de telle sorte que les préférences puissent être modélisées par ces opérateurs d'agrégation. Nous aurons donc :
$d_2 \sim d_3 \Leftrightarrow w_1(C_1(d_2)) + w_2(C_2(d_2)) = w_1(C_1(d_2)) + w_2(C_2(d_3)) \Rightarrow w_1 = w_2$
$d_1 \succ d_2 \Leftrightarrow w_1(C_1(d_1)) + w_2(C_2(d_1)) > w_1(C_1(d_2)) + w_2(C_2(d_2))$
$\Rightarrow 0.45 \times (w_1 + w_2) > w_2$
Par conséquent, nous obtenons : $0.9 \times w_2 > w_2$*, ce qui est absurde.*

Pour pallier à cet inconvénient, des opérateurs d'agrégation prioritaires ont été récemment proposés (da Costa Pereira *et al.,* 2009). Néanmoins, il existe toujours des préférences qui ne peuvent pas être représentées par ces derniers. Formellement, le processus de priorisation est considéré uniquement sur les critères individuels (*e.g.,* $c_i \succ c_j \succ c_k$). Toutefois, si on considère, par exemple, quatre critères : *aboutness* (Ab), *fraîcheur* d'information (Fr), *centres d'intérêt* (Ci) et *autorité* (Au), un utilisateur pourrait préférer les documents *récents* traitant un *topic* donné, plutôt que les documents *autoritaires* réprésentant au mieux le sujet recherché par l'utilisateur. Ainsi, on aura la préférence $\{Ab, Fr\} \succ_{\mathcal{C}} \{Au, Ci\}$, qui ne peut pas être modélisée par les opérateurs d'agrégation ainsi cités. Pour prendre en compte ce type de contrainte, il serait judicieux de considérer des poids non seulement sur les critères individuels, mais aussi sur les sous ensembles de critères (*e.g.,* $w_{\{Ab,Fr\}}$, $w_{\{Au,Ci\}}$). Dans l'exemple, comme les documents qui satisfont de manière équitable les deux critères sont les documents les plus préférés, on devrait attribuer un poids w_{12} au sous ensemble $\{c_1, c_2\}$. Idéalement, on doit assigner à w_{12} un score elevé (*e.g.*, w_{12} = 1) et attribuer à w_1 et w_2 des scores faibles (*e.g.*, $w_1 = w_2$ = 0.35) puisque les critères c_1 et c_2 ne répondent pas aux préférences quand ils agissent de manière indépendante.

Pour répondre à ce problème, nous proposons de traiter le problème d'agrégation de pertinence multidimensionnelle à l'aide de l'intégrale de Choquet (Choquet, 1953). Cet opérateur permet de définir des poids d'importance non seulement sur des critères uniques mais aussi sur des combinaisons de critères. En reprenant l'exemple déjà mentionné, les préférences souhaitées peuvent être simplement modélisées par une mesure floue μ tel que $\mu_{\{Ab,Fr\}} > \mu_{\{Au,Ci\}}$ où $\mu_{\{.\}}$ représente le degré d'importance d'un critère (*resp.,* un sous ensemble de critères). Notre choix de l'intégrale de Choquet comme opérateur d'agrégation est en outre motivé par sa capacité à généraliser plusieurs opérateurs d'agrégation classiques (Grabisch *et al.,* 2000) tels que la moyenne arithmétique pondérée ou les opérateurs OWA (Ordered Weighted Averaging operator) (Yager, 1988) et même les approches basées sur la combinaison linéaire (Vogt *et al.,* 1999). De plus, comme il a été démontré dans plusieurs travaux de RI, que les dimensions de pertinence interagissent et sont souvent dépendants (Saracevic, 2007, Carterette *et al.,* 2011), l'intégrale de Choquet se basant sur le concept de mesure floue, s'avère un bon candidat pour ce type d'agrégation.

3.2. *Spécification formelle de l'opérateur d'agrégation de Choquet*

Nous introduisons dans cette section notre approche pour l'agrégation de pertinene multidimensionnelle. Nous présentons dans ce qui suit les définitions d'une mesure floue (appelée aussi capacité) et de l'intégrale de Choquet dans une taĉhe de RI.

***Notion de mesure floue* :** Soient $\mathcal{C}$ l'ensemble des critères de pertinence et $I_{\mathcal{C}}$ l'ensemble de tous les sous-ensembles de ces dimensions de pertinence. Une mesure floue (ou capacité) est une fonction μ de $I_{\mathcal{C}}$ dans $[0, 1]$ tel que :
$\forall\, I_{C_1}, I_{C_2} \in I_{\mathcal{C}}$, si $(I_{C_1} \subseteq I_{C_2})$ alors $\mu(I_{C_1}) \leq \mu(I_{C_2})$, avec $\mu(I_{\varnothing})$ = 0 et $\mu(I_{\mathcal{C}})$ = 1.
$\mu(I_i)$ (qu'on notera μ_{I_i}) peut être interprété comme le degré d'importance de la combinaison I_i d'un ensemble de critères. En effet, le poids μ_{I_i}, considéré indépendemment des poids usuels définis individuellement sur chaque critère, est attribué à chaque combinaison I_i de critères.

***Principe d'agrégation de Choquet* :** Soient $\mathcal{D}$ la collection des documents, et $d_j \in \mathcal{D}$. Le score global de d_j, donné par l'intégrale de Choquet selon une mesure floue μ et un ensemble $\mathcal{C}$ de critères de pertinence, est défini par :

$$Ch_{\mu}(C_{1j}, ..., C_{Nj}) = \sum_{i=1,\ldots,N} (c_{(i)j} - c_{(i-1)j}).\mu_{C_{(i)}} \qquad [1]$$

avec $c_{(i)j}$ est le score obtenu selon un critère donné[3], $C_{(i)}$ = $\{c_i, ..., c_N\}$ est un ensemble de dimensions de critère de pertinence, avec $C_{(0)} = \varnothing$, $\mu_{C_{(0)}} = 0$ et $\mu_{C_{(1)}} = 1$. Comme mentionné dans la section 3.1, C_{ij} est le score partiel [4] de d_j selon le critère c_i et $\mu_{C_{(i)}}$ est le degré d'importance de la combinaison $\{c_i, ..., c_N\}$ de critères. En effet, de cette manière, l'intégrale de Choquet est bien capable de prendre en compte tout type d'interaction entre les critères et représenter toutes les préférences qui n'ont pas pu être capturées par un opérateur d'agrégation simple (Grabisch *et al.,* 2000). Notons qu'il existe trois types d'interactions qui peuvent être modélisées par la mesure floue :

– *Intéraction positive*, appelée aussi *redondance*, quand le poids global de deux critères est supérieure à leur poids individuels : $\mu_{\{c_i,c_j\}} > \mu_{c_i} + \mu_{c_j}$. Cette inégalité peut être interprétée comme *"la contribution de c_j à toute combinaison de critères contenant c_i est strictement supérieure à la contribution de c_j à la même combinaison quand c_i est exclu"*. Dans ce cas, c_i et c_j sont négativement corrélés.

– *Intéraction négative* ou *complémentarité*, quand le poids global de deux critères est plus petit que leurs poids individuels : $\mu_{\{c_i,c_j\}} < \mu_{c_i} + \mu_{c_j}$.

– *Indépendence*, quand il n'existe aucune corrélation entre l'ensemble des critères. Dans ce cas on dit que la mesure floue est additive : $\mu_{\{c_i,c_j\}} = \mu_{c_i} + \mu_{c_j}$.

Comme illustré dans l'exemple, quand il s'agit de l'agrégation multicritère, ces interactions doivent être prises en compte dans n'importe quel processus d'agrégation, d'où

3. $c_{(\cdot)j}$ indique que les indices sont permutés de façon à ce que $0 \leq c_{(1)j} \leq ... \leq c_{(N)j}$.
4. La différence entre $c_{(i)j}$ et C_{ij} est que les scores partiels $c_{(i)j}$ sont permutés avant de calculer le score global, alors que C_{ij} est le score partiel du document d_j, obtenu sur le critère c_i.

la puissance de l'intégrale de Choquet dans ce contexte. Cependant, quand on traite de l'agrégation multidimensionnelle avec l'intégrale de Choquet en se basant sur les mesures floues, le défi majeur qui se pose est l'identification des capacités, surtout quand le nombre de critères est élevé (Grabisch *et al.,* 2008). Pour répondre à ce défi, nous allons nous baser sur la méthode des moindres carrés, qui est la méthode d'optimisation la plus utilisée dans la littérature pour cette fin[5] (Grabisch *et al.,* 2008). Pour ce faire, supposons tout d'abord que nous avons un petit sous ensemble de l documents $\mathcal{D}$, que nous allons considérer comme ensemble d'apprentissage. Après la sélection de $\mathcal{D}$, une réalité de terrain (*ground truth*) permet d'exprimer des préférences sur l'ensemble des documents selon les critères choisis. Supposons que les scores partiels C_{ik}, qui devront être assignés à chaque document $d_k \in \mathcal{D}$, sont donnés selon chaque critère c_i. L'objectif de la méthode des moindres carrés est donc de minimiser l'erreur quadratique totale E^2 entre le score désiré $y(d_k)$ donné sur chaque document d_k, et les scores globaux calculés par l'intégrale de Choquet. La formule de calcul de l'erreur quadratique est donné par : $E^2 = \sum_{k=1}^{l}(Ch_\mu(C_{1k}, C_{2k}, \ldots, C_{Nk}) - y(d_k))^2$.

4. Évaluation expérimentale

Dans cette section, nous évaluons notre approche d'agrégation de pertinence dans un cadre de RI sociale. Plus particulièrement, nous considérons une tâche de recherche de *tweets* où nous nous sommes basés sur trois dimensions de pertinence à savoir, la *dimension thématique*, la *fraîcheur* des *tweets*, et l'*autorité* des auteurs des *tweets*. Les objectifs de ces expérimentations sont : *(i)* Montrer les corrélations existantes entre les critères considérés par le biais de la mesure floue, *(ii)* Évaluer l'efficacité de l'intégrale de Choquet comparativement à d'autres opérateurs d'agrégation standards tels que les moyennes arithmétiques (pondérées) et les mécanismes de combinaison linéaire.

4.1. *Cadre expérimental*

Tâche de recherche des *tweets* : Comme défini dans la tâche Microblog de TREC 2011 (Ounis *et al.,* 2011), la recherche de *tweets* est une tâche en temps réel dans laquelle les utilisateurs s'intéressent à l'information pertinente et récente, à la fois. Des travaux récents s'intéressant à cette tâche de recherche ont identifié plusieurs facteurs ayant un impact considérable dans le classement final des documents (Nagmoti *et al.,* 2010). Parmi ces critères nous citons : la *topicalité*, la *longueur* des *tweets*, la *présence des URLs* dans un *tweet* et l'*autorité* (Duan *et al.,* 2010, Chen *et al.,* 2012). En effet, la spécificité de la tâche Microblog de TREC 2011 (Ounis *et al.,* 2011) qui définit la recherche de *tweets* comme étant une tâche de recherche dans laquelle les utilisateurs s'intéressent aux informations *récentes* et *pertinentes*, motive bien l'utilisation de l'intégrale de Choquet, où les interactions entre les critères est bien prise

5. Nous avons appliqué cette méthode au sein du package KAPPALAB se basant sur le langage R (Grabisch *et al.,* 2008).

en considération. Ceci est d'autant plus important que les utilisateurs préfèreront les *tweets* qui sont à la fois pertinents et récents, *i.e.,* dont les scores sont balancés entre ces deux critères et non pas été biaisés par l'un des deux. Pour notre part, nous exploitons trois dimensions de pertinence à savoir, la *topicalité (thématique de recherche)* (To), la *fraîcheur des tweets* (Fr), et l'*autorité* (Au). L'agrégation de ces critères avec l'intégrale de Choquet selon une mesure floue μ, en réponse à une requête utilisateur Q, est définie par :

$$Ch_\mu(To(T_j, Q), Au(T_j), Fr(T_j)) = \sum_{i=1,\ldots,3} (c_{(i)j} - c_{(i-1)j}).\mu_{C_{(i)}} \qquad [2]$$

où T_j est un *tweet*, $c_{(i)}$ indique qu'un score partiel[6] obtenu sur un critère donné, en réponse à la requête Q, sont permutés tel que : $0 \leq c_{(1)j} \leq c_{(2)j} \leq c_{(3)j}$, et $C_{(i)} = \{c_i, ..., c_3\}$. Ch_μ est le score global qui définit le classement final de chaque document selon les trois dimensions de pertinence. Nous donnons dans ce qui suit les mesures de ces différents critères.

– *Topicalité :* nous proposons d'utiliser la fonction Okapi BM25 (Robertson *et al.,* 1976) pour le classement des *tweets* en réponse à une requête Q donnée :

$$To(T, Q) = BM25(T, Q) = \sum_{q_i \in Q} \frac{Idf(q_i) \cdot tf(q_i, T) \cdot (k_1 + 1)}{tf(q_i, T) + k_1(1 - b + b\frac{Length(T)}{avg_{length}})} \qquad [3]$$

où q_i est un terme de la requête, $Idf(q_i)$ est la fréquence inverse du document, $Length(T)$ dénote la longueur d'un *tweet* T et avg_{length} représente la moyenne des longueurs des *tweets* dans la collection.

– *Autorité :* représente l'influence des auteurs des *tweets* dans Twitter. Nous la définissons comme présentée dans (Nagmoti *et al.,* 2010) : $Au(T) = Au_{nb}(T) + Au_{me}(T)$, où *(i)* $Au_{nb}(T)$ est le *nombre total des tweets*, pour favoriser les *tweets* publiés par des utilisateurs influents. Nous la définissons par : $Au_{nb}(T) = N(a_i(T))$, avec $a_i(T)$ représentant l'auteur du *tweet* T, et $N(a_i(T))$ dénotant le nombre de *tweets* publiés par a_i. *(ii)* $Au_{me}(T)$ est le *nombre de mentions d'un auteur*, *i.e.,* plus un auteur est mentionné, plus il est populaire. Elle est définie par : $Au_{me}(T) = N(ma_i)(T))$, avec $N(ma_i)(T))$ estimant le nombre de fois l'auteur du *tweet* T est mentionné.

– *Fraîcheur des tweets :* c'est la différence entre le temps de publication du *tweet* $Tp(T)$ et la date de soumission de la requête $Ts(Q)$: $Fr(T) = T_s(Q) - T_p(T)$.

Collection de test : Nous exploitons la collection de test TREC Microblog 2011. Les statistiques de la collection sont données dans le tableau 1.

Mesures d'évaluation : Nous utilisons la mesure de précision à 30 documents (P@30) ainsi que la moyenne des précisions (MAP), qui sont les mesures officielles des tâches Microblog de TREC 2011 et 2012. Nous avons calculé ces mesures avec l'outil standard $trec_eval$[7].

6. Tous les scores partiels sont normalisés et sont tous dans $[0, 1]$.
7. *http ://trec.nist.gov/trec_eval*

Nombre de *tweets*	$16,141,812$
tweets nulls	$1,204,053$
Termes uniques	$7,781,775$
Nombre de Twitters	$5,356,432$
Requêtes	49

Tableau 1 – Statistiques de la collection de test de la tâche Microblog de TREC 2011.

4.2. *Évaluation des performances de recherche*

4.2.1. *Protocole d'évaluation*

Une étape cruciale préalable au calcul des scores globaux par l'opérateur de Choquet est l'identification des mesures floues.

Algorithm 1 Identification des mesures floues

Entrées: Q_{app}, Q_{test}, l'ensemble des *tweets* $\mathcal{D}$, les jugements de pertinence $qrels$.
Sortie: Les valeurs des capacités $\mu_{\{i\}}$ de tous les critères et sous ensemble de critères.

1. **Pour** chaque topic $T_i \in Q_{app}$ **Faire**
2. Calculer la $P@30$ pour chaque combinaison de capacités (parmi les 19 combinaisons expérimentées) $\mu_{c_i} \in [0,1]$, avec un pas de 0.1, pour chaque critère c_i et pour chaque combinaison de critères. *//Voir les exemples de capacité dans le tableau 2.*
3. **Fin Pour**
4. Sélectionner la combinaison de capacités ayant donné la meilleure moyenne en termes de $P@30$ sur l'ensemble Q_{app}.
5. Sélectionner un sous ensemble de documents (D_{app}) retournés (pertinents et non pertinents) avec leurs scores partiels et globaux donnés en se basant sur la combinaison de capacités choisie (sélectionner tous les *tweets* pertinents ainsi que le même nombre de *tweets* non pertinents pour chaque requête).
6. Assigner à tous les documents pertinents des scores (partiels et globaux) plus élevés que les documents non pertinents (même si ces derniers ont un meilleur classement).
7. Appliquer la méthode des moindres carrés sur l'ensemble D_{app} des *tweets*.

Comme nous avons les jugements de pertinence (*qrels*) associés au 49 requêtes de la tâche, nous allons nous baser sur la méthode des moindres carrés pour identifier la meilleure combinaison de capacités qui doit être associée à chaque critère et à chaque sous ensemble de critères. Les capacités sont apprises sur un sous ensemble d'apprentissage selon la méthode détaillée dans l'algorithme 1. En effet, nous avons expérimenté 19 combinaisons de capacités $\mu^{(i)}(c_{topicalite}, c_{autorite}, c_{fraicheur})$ obtenues avec un pas de 0.1, comme précisé dans l'étape 3 de l'algorithme 1, et tel que la somme des trois scores individuels partiels soit égale à 1. Pour ce faire, nous avons procédé à une division aléatoire de l'ensemble des requêtes en deux sous ensembles Q_{app} et Q_{test}. Le premier ensemble est utilisé pour l'apprentissage et le deuxième

(25 requêtes) pour le test. La méthode d'identification des capacités est détaillée dans l'algorithme 1.

Le tableau 2 présente des exemples de capacités avec les résultats obtenus en terme de $P@30$ pour chacune des combinaisons, après application de l'intégrale de Choquet sur les trois dimensions de pertinence (Tr, Au, Fr). La combinaison de capacités μ^6 donnée dans le tableau 2, représente la meilleure combinaison obtenue dans la phase d'apprentissage donnant ainsi la meilleure valeur de $P@30$ sur l'ensemble Q_{app}.

Combinaison	μ_{To}	μ_{Au}	μ_{Fr}	$\mu_{\{To,Au\}}$	$\mu_{\{To,Fr\}}$	$\mu_{\{Au,Fr\}}$	$P@30$
μ^1	0.3	0.4	0.3	0.7	0.6	0.7	0.033
μ^2	0.3	0.5	0.2	0.8	0.5	0.7	0.138
μ^3	0.4	0.2	0.4	0.6	0.8	0.6	0.033
μ^4	0.4	0.3	0.3	0.7	0.7	0.6	0.1
μ^5	0.4	0.4	0.2	0.8	0.6	0.6	0.13
μ^6	0.8	0.1	0.1	0.9	0.9	0.2	0.281
$\mu^\star$	**0.612**	**0.164**	**0.138**	**-0.09**	**0.91**	**-0.01**	**0.37**

Tableau 2 – Un exemple des valeurs de capacités sélectionnées dans la phase d'apprentissage sur Q_{app}. $\mu^\star$ est la combinaison retournée par la méthode des moindres carrés et utilisée sur Q_{test}.

Les capacités $\mu^\star$ retournées par la méthode des moindres carrés sont données dans la dernière ligne du tableau 2. Nous remarquons que les valeurs de capacités obtenues sur les sous ensembles $\{To, Au\}$ et $\{Au, Fr\}$ sont négatives. Selon les trois types d'interaction que nous avons présentés dans la section 3, la contribution du critère de pertinence *topicalité* à toute combinaison de critère contenant le critère *autorité* est supérieure à sa contribution quand le critère *autorité* est exclu. Il en est de même pour les deux critères *fraîcheur des tweets* et *autorité*. Ceci est probablement dû au fait que les assesseurs de la tâche Microbog de TREC n'ont pas donné une très grande importance à l'autorité des auteurs des *tweets* pendant leur évaluation de la collection de test. Partant de ce constat, le critère *autorité*, malgré son importance dans *Twitter* (Chen *et al.,* 2012), peut biaiser les scores globaux des *tweets*, et ceci explique bien les capacités négatives assignées à $\mu_{\{To,Au\}}$ et $\mu_{\{Au,Fr\}}$. Les valeurs élevées de capacités associées à $\{To, Fr\}$ indiquent une interaction positive entre les deux critères, ce qui convient à l'objectif de la tâche de recherche de *tweets* et au cadre de RI que nous avons considéré dans cet article. De plus, étant donné que les utilisateurs de *Twitter* préfèrent les *tweets* qui sont à la fois pertinents et récents (*i.e.,* dont les scores sont balancés entre ces deux critères et non pas biaisés par l'un des deux), l'intégrale de Choquet a bien répondu à cette préférence en associant une capacité élevée à la combinaison du sous ensemble de critères {*topicalité, fraîcheur des tweets*} par rapport aux scores individuels de chacun d'entre eux. Ainsi, toutes les capacités obtenues sur toutes les combinaisons de critères montrent que les dimensions de pertinence interagissent, et ceci doit être considéré quand il s'agit de l'agrégation des critères dépendants.

4.2.2. *Évaluation des performances de recherche*

L'étude de l'efficacité de notre approche a été menée en se basant sur l'ensemble des 25 requêtes de test (Q_{test}) et en se comparant aux opérateurs de référence (*baseline*) : la moyenne arithmétique (MA), la moyenne arithmétique pondérée ($MAPond$) et la méthode de combinaison linéaire (MCL). La formule utilisée pour le calcul des score globaux avec la méthode de combinaison linéaire est donnée comme suivant : $MCL(T) = \sum_{i=1}^{3}(\alpha_i mcl_i(T))$, où $mcl_i(T)$ est le score partiel du *tweet* T selon le critère individuel c_i, avec $c_i \in$ *{topicalié, autorité, fraîcheur de tweet}*. Il est à noter que les poids α_i, des critères utilisés par les opérateurs de référence $MAPond$ et MCL, sont paramétrés pendant la phase d'apprentissage. Nous leur avions assigné les valeurs optimales donnant les meilleurs résultats en terme de $P@30$ pendant cette phase : $\alpha_{fraicheur} = 0.05$, $\alpha_{autorite} = 0.15$ and $\alpha_{topicalite} = 0.79$, avec α_i est le poids du critère c_i.

Mesure	MA	$MAPond$	MCL	*Notre approche*	*Améliorations sur les 3 méthodes*
$P@5$	0.184	0.184	0.376 ‡†	**0.512**	‡†⋆
$P@10$	0.18	0.188	0.38 ‡†	**0.488**	‡†⋆
$P@30$	0.152	0.16 ‡	0.344 ‡†	**0.4347**	‡†⋆
$P@100$	0.124	0.1308 ‡	0.2244 ‡†	**0.2932**	‡†⋆
MAP	0.0973	0.106 ‡	0.2173 ‡†	**0.2815**	‡†⋆

Tableau 3 – Efficacité de l'agrégation de pertinence multidimensionnelle de notre approche dans la tâche de recherche de *tweets*. Les symboles ‡, † et ⋆ indiquent les améliorations significatives sur MA, $MAPond$ et MCL respectivement.

Le Tableau 3 présente les résultats obtenus par notre approche *vs.* les trois opérateurs d'agrégation et ce, selon quatre points de précisions ($P@5$, $P@10$, $P@30$ et $P@100$). Ainsi, pour tester l'importance des améliorations obtenues par notre approche, nous avons effectué un *t*-test, et nous avons montré que ces améliorations sont statistiquement importantes avec des *p*-valeurs < 0.01 pour toutes les fonctions d'agrégation. Suivant la mesure officielle $P@30$ de TREC Microbog, notre approche est plus performante que les trois autres mécanismes d'agrégation, marquant à titre d'exemple, une amélioration de $23,60\%$ sur l'opérateur de moyenne arithmétique pondérée. La plus faible amélioration marquée est obtenue avec la méthode de combinaison linéaire MCL avec une valeur de 26.36% en terme de $P@30$, et ceci explique son utilisation dans de nombreuses approches d'agrégation, et même par les groupes participants à la tâche Microbog de TREC 2011.

5. Conclusion et perspectives

Dans cet article, nous avons présenté un aperçu des travaux portant sur l'agrégation multicritère dans le domaine de RI et nous avons proposé une nouvelle approche pour l'agrégation de pertinence multidimensionnelle. En se basant sur l'intégrale de Cho-

quet et sur le concept de mesure floue, l'approche proposée est capable de modéliser les interactions pouvant exister entre les critères de pertinence. En effet, nous avons effectué un apprentissage des mesures floues, donnant lieu ainsi à une méthode générale qui est capable de traiter le problème d'agrégation multicritère indépendamment du nombre de critères utilisés et quelque soit la le cadre d'agrégation. L'évaluation de notre approche, dans une tâche de recherche de *tweets* et sur la collection de test fournie par la tâche Microblog de TREC 2011, a montré sa supériorité par rapport aux méthodes d'agrégation standards.
En perspective, nous envisageons d'effectuer une analyse raffinée au niveau des requêtes pour un apprentissage dynamique des mesures floues, et ce en se basant sur les deux collections fournies par la tâche Microblog de TREC 2011 et 2012. Nous envisageons également l'étude de l'efficacité de l'approche dans d'autres cadres d'agrégation et dans d'autres contextes de RI, tel que la RI mobile avec des dimensions de pertinence différentes.

Remerciements

Ce travail est partiellement financé par le projet franco-tunisien PHC Utique n° $11G1417$, intitulé EXQUI.

6. Bibliographie

Ah-Pine J., « Data Fusion in Information Retrieval Using Consensus Aggregation Operators », *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, vol. 1, IEEE Computer Society, Washington, DC, USA, p. 662-668, 2008.

Arrow K. J., *Choix collectif et préférences individuelles*, vol. 1, Calmann-Lévy, 1974.

Barry C. L., « User-Defined Relevance Criteria : An Exploratory Study », *Journal of the American Society for Information Science*, vol. 45, n° 3, p. 149-159, 1994.

Becker H., Naaman M., Gravano L., « Beyond Trending Topics : Real-World Event Identification on Twitter », *ICWSM*, The AAAI Press, 2011.

Ben Jabeur L., Tamine L., Boughanem M., « A social model for literature access : towards a weighted social network of authors », *Adaptivity, Personalization and Fusion of Heterogeneous Information*, RIAO '10, Le centre de hautes etudes internationales d'Informatique Documentaire, Paris, France, p. 32-39, 2010.

Borlund P., « The concept of relevance in IR », *Journal of the American Society for Information Science and Technology*, vol. 54, n° 10, p. 913-925, 2003.

Bouidghaghen O., Tamine L., Boughanem M., « Personalizing Mobile Web Search for Location Sensitive Queries », *International Conference on Mobile Data Management (MDM)*, IEEE Computer Society, p. 110-118, 2011a.

Bouidghaghen O., Tamine-Lechani L., Pasi G., Cabanac G., Boughanem M., da Costa Pereira C., « Prioritized Aggregation of Multiple Context Dimensions in Mobile IR », *Proceedings*

of the 7th Asia conference on Information Retrieval Technology, vol. 7097 of *AIRS'11*, Springer, Berlin, Heidelberg, p. 169-180, 2011b.

Cantera J. M., Arias M., Cabrero J., Garcia G., Zubizarreta A., Vegas J., de la Fuente P., « Mymose : Next generation search engine for mobile users », *the 3rd edition of the Future of Web Search Workshop*, 2008.

Carterette B., Kumar N., Rao A., Zhu D., « Simple Rank-Based Filtering for Microblog Retrieval : Implications for Evaluation and Test Collections », *Proceedings of the 20th Text REtrieval Conference (TREC 2011)*, 2011.

Chen K., Chen T., Zheng G., Jin O., Yao E., Yu Y., « Collaborative personalized tweet recommendation », *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, ACM, New York, NY, USA, p. 661-670, 2012.

Choquet G., « Theory of capacities », *Annales de l'Institut Fourier*, vol. 5, p. 131-295, 1953.

Cong G., Jensen C. S., Wu D., « Efficient retrieval of the top-k most relevant spatial web objects », *Proceedings of the VLDB Endowment*, vol. 2, p. 337-348, 2009.

Cooper W. S., « On selecting a measure of retrieval effectiveness », *Journal of the American Society for Information Science*, vol. 24, n° 2, p. 87-100, 1973.

Cosijn E., Ingwersen P., « Dimensions of relevance », *Inf. Process. Manage.*, vol. 36, n° 4, p. 533-550, 2000.

da Costa Pereira C., Dragoni M., Pasi G., « Multidimensional Relevance : A New Aggregation Criterion », *Proceedings of the 31th European Conference on IR Research, ECIR 2009*, p. 264-275, 2009.

da Costa Pereira C., Dragoni M., Pasi G., « Multidimensional relevance : Prioritized aggregation in a personalized Information Retrieval setting », *Inf. Process. Manage.*, vol. 48, n° 2, p. 340-357, 2012.

Daoud M., Tamine L., Boughanem M., « A Personalized Graph-Based Document Ranking Model Using a Semantic User Profile », *Proceedings of the 18th international conference on User Modeling, Adaptation, and Personalization, UMAP*, p. 171-182, 2010.

Duan Y., Jiang L., Qin T., Zhou M., Shum H.-Y., « An empirical study on learning to rank of tweets », *Proceedings of the 23rd International Conference on Computational Linguistics*, COLING '10, Stroudsburg, PA, USA, p. 295-303, 2010.

Farah M., Vanderpooten D., « An outranking approach for rank aggregation in information retrieval », *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM, New York, NY, USA, p. 591-598, 2007.

Gauch S., Chaffee J., Pretschner A., « Ontology-based personalized search and browsing », *Web Intelli. and Agent Sys.*, vol. 1, n° 3-4, p. 219-234, 2003.

Gerani S., Zhai C., Crestani F., « Score transformation in linear combination for multi-criteria relevance ranking », *Proceedings of the 34th European conference on Advances in Information Retrieval*, ECIR'12, Springer-Verlag, Berlin, Heidelberg, p. 256-267, 2012.

Grabisch M., Kojadinovic I., Meyer P., « A review of methods for capacity identification in Choquet integral based multi-attribute utility theory : Applications of the Kappalab R package », *European Journal of Operational Research*, vol. 186, n° 2, p. 766-785, 2008.

Grabisch M., Murofushi T., Sugeno M., Kacprzyk J., *Fuzzy Measures and Integrals. Theory and Applications*, Physica Verlag, Berlin, 2000.

Mizzaro S., « How many relevances in information retrieval ? », *Interacting with Computers*, vol. 10, n° 3, p. 303-320, 1998.

Nagmoti R., Teredesai A., De Cock M., « Ranking Approaches for Microblog Search », *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, vol. 1, Washington, DC, USA, p. 153-157, 2010.

Ounis T., Macdonald J., Soboroff I., « Overview of the trec-2011 microblog track », *Proceedings of the 20th Text REtrieval Conference (TREC 2011)*, 2011.

Pasi G., Bordogna G., Villa R., « A multi-criteria content-based filtering system », *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM, New York, NY, USA, p. 775-776, 2007.

Robertson S. E., Jones K. S., « Relevance weighting of search terms », *Journal of the American Society for Information Science*, vol. 27, n° 3, p. 129-146, 1976.

Saracevic T., « Relevance : A review of the literature and a framework for thinking on the notion in information science. Part III : Behavior and effects of relevance », *Journal of the American Society for Information Science*, vol. 58, n° 13, p. 2126-2144, 2007.

Saracevic T., Rothenberg D., Stephan P., « Study of information utility », *Proceedings of the American Society for information Science*, vol. 11, p. 234-238, 1974.

Schamber L., « Users' Criteria for Evaluation in a Multimedia Environment », *Proceedings of the 54th ASIS Annual Meetin*, vol. 28, p. 126-133, 1991.

Steuer R. E., *Multiple Criteria Optimization : Theory, Computation and Application*, John Wiley & Sons, New York, 1986.

Taylor A. R., « User relevance criteria choices and the information search process », *Inf. Process. Manage.*, vol. 48, n° 1, p. 136-153, 2012.

Taylor A. R., Cool C., Belkin N. J., Amadio W. J., « Relationships between categories of relevance criteria and stage in task completion », *Inf. Process. Manage.*, vol. 43, n° 4, p. 1071-1084, 2007.

Vogt C. C., Cottrell G. W., « Fusion Via a Linear Combination of Scores », *Information Retrieval*, vol. 1, n° 3, p. 151-173, 1999.

Wei F., Li W., Liu S., « iRANK : A rank-learn-combine framework for unsupervised ensemble ranking », *Journal of the American Society for Information Science and Technology*, vol. 61, n° 6, p. 1232-1243, 2010.

White R. W., Ruthven I., Jose J. M., « A study of factors affecting the utility of implicit relevance feedback », *Proceedings of the 28th annual international ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, p. 35-42, 2005.

Yager R. R., « On ordered weighted averaging aggregation operators in multicriteria decision making », *IEEE Transactions On Systems Man And Cybernetics*, vol. 18, n° 1, p. 183-190, 1988.