RELIEFS : Un système pour le filtrage adaptatif de documents textuels basé sur la notion de résonance

Brouard Christophe, Nie Jian-Yun

Laboratoire RALI
Département d'Informatique et de Recherche Opérationnelle
Université de Montréal
CP. 6128 succ. Centre Ville
Montréal QC H3C3J7, Canada

{brouardc,nie}@iro.umontreal.ca

Résumé :
Nous présentons le système RELIEFS pour RELevance Information Extraction Fuzzy System dédié au filtrage adaptatif de documents textuels. Ce système s'inspire très largement de mécanismes cognitifs intervenant dans les processus de sélection d'information. En particulier, il implémente l'idée de résonance introduite dans la théorie cognitive ART [GROS 76] qui rend compte de la façon dont notre système cognitif selectionne l'information provenant de notre environnement. L'évaluation de la pertinence d'un document est ainsi réalisée par l'estimation d'un degré de résonance du document vis-à-vis d'une requête. Nous montrons que cette idée n'est pas étrangère à la recherche d'information mais qu'elle s'identifie bien avec la notion de double implication que l'on retrouve sous une forme plus ou moins explicite dans les formalisations de la pertinence proposées dans des cadres logiques ou probabilistes. L'application de ce système à la tâche de filtrage adaptatif de TREC9 [BROU 00] a donné des résultats proches des meilleurs systèmes de filtrage. Ceci suggère que l'approche proposée est valide.

Mots-clés: recherche d'information, pertinence, résonance, cognition, filtrage.
Keywords : information retrieval, relevance, resonance, cognition, filtering.
Catégorie : chercheur.

1 Introduction

Dès les premières réalisations de systèmes informatiques utilisés pour la recherche d'information (RI), on s'est interrogé sur la notion centrale de pertinence (on pourra se reporter à [SARA 75] pour un historique) ; l'objectif ultime étant de formaliser cette notion pour proposer une fonction qui évalue automatiquement la pertinence d'un document vis-à-vis d'une requête et qui sélectionne ainsi de façon satisfaisante des documents pour l'utilisateur. Cependant, il n'existe encore à l'heure actuelle aucune définition précise de cette notion [FROE 94].

Avec le système RELIEFS, nous avons expérimenté l'idée selon laquelle la notion de *résonance* s'identifierait bien avec celle de pertinence. Nous considérons cette notion plus particulièrement dans le sens de la théorie cognitive et connexioniste ART (Adaptative Resonance Theory) proposée par Grossberg [GROS 76] qui par ce mécanisme rend compte de la façon dont certaines informations provenant de notre environnement seraient sélectionnées au détriment d'autres informations. ART permet par exemple de modéliser la façon dont il nous est possible de suivre une conversation dans un environnement bruyant c'est-à-dire de sélectionner les sons relatifs à la conversation au détriment des autres sons qui apparaissent alors en fond. Il est intéressant de constater que la notion de résonance n'est pas complètement nouvelle pour la RI. En effet, il existe une analogie avec des formalisations de la pertinence en RI. Ces considérations théoriques font l'objet de la première partie de cet article. Dans la seconde partie,

nous décrivons le système RELIEFS qui met en œuvre l'idée de résonance dans la tâche de filtrage de documents textuels. Les résultats du système dans le cadre de la tâche de filtrage adaptatif de TREC9 [BROU 00] sont ensuite présentés. Enfin, nous terminons en discutant l'apport de notre approche en mettant l'accent sur le nouveau regard qu'elle permet de poser sur la notion de pertinence.

2 Vers une formalisation de la notion de pertinence

Nous abordons ici le problème de la formalisation de la notion de pertinence. Nous proposons d'établir un lien entre cette notion et celle de résonance sous la forme où elle apparaît dans la théorie cognitive ART. Nous donnons d'abord les grandes lignes de cette théorie. Nous présentons ensuite les formalisations de la pertinence en RI en mettant en évidence la convergence autour de l'idée de double implication entre document et requête. Enfin nous établissons de façon plus précise l'analogie entre la double implication et la résonance.

2.1. La théorie de la résonance adaptative (ART)

Cette théorie part de l'analyse d'un système en interaction avec son environnement. Le système (S) prend en compte les informations du monde externe (E) tout en appliquant ses connaissances pour la sélection de ces informations (figure 1). La sélection d'une information du monde externe par le système fait donc intervenir deux processus de nature différente, un processus ascendant (E→S) et un processus descendant (S→E).

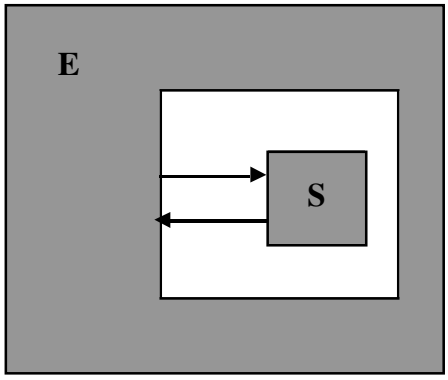

Figure 1: Interaction d'un système (S) avec son environnement (E). Extrait de [CLAN 97].

ART modélise cette interaction à partir d'un réseau de neurones à deux couches. La couche de neurone d'entrée (F1) permet de représenter le stimulus provenant de l'environnement. L'autre couche (F2) correspond à une représentation interne du système. Les deux couches sont liées par des connections de nature associatives. ART modélise la façon dont une information de l'environnement peut être sélectionnée par le système en considérant un principe de résonance. D'une façon globale, la sélection d'une information dépend de son degré de résonance avec l'attente du système. Dans le réseau, la résonance se traduit par une propagation de l'activation de F1 vers F2 et une rétro-propagation de l'activation de F2 (après compétition interne) vers F1 (figure 2). Si l'activation retro-propagée à F1 s'apparie bien avec la distribution d'activation codant le stimulus dans F1, un état de résonance est atteint (on peut parler de résonance dans le sens où cet appariement conduit à une amplification de l'activation du stimulus). Grossberg présente sa théorie comme une solution au problème inhérent à tout système en interaction avec un environnement à savoir le dilemme entre stabilité et plasticité [GROS 99]. Ce dilemme peut

s'exprimer par l'interrogation suivante : Comment préserver une connaissance précédemment acquise tout en continuant d'apprendre ? Les connections orientées de F1 vers F2 permettent de prendre en compte l'environnement (si elles n'existaient pas le système serait coupé du monde externe) et les connections orientées de F2 vers F1 permettent de prendre en compte ses connaissances (s'il elles n'existaient pas, le système ne pourrait tenir compte des connaissances emmagasinées au cours de ses précédentes expériences). La prise en compte simultanée de ces deux types de connections garantit donc au système une plasticité tout en lui conférant une certaine stabilité.

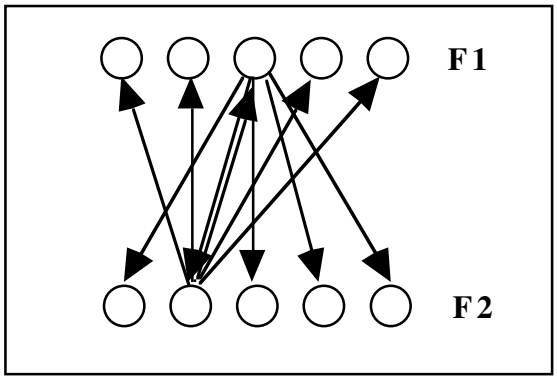

Figure 2. Propagation et rétro-propagation de l'activation entre F1 et F2.

2.2. Formalisation de la notion de pertinence en recherche d'information

On peut distinguer deux grands types de recherche portant sur la formalisation de la notion de pertinence en RI. Le premier type concerne l'identification de la relation qui unit un document pertinent à un besoin d'information. Les recherches de Green [GREE 95a, b] sont des exemples de ce type de recherche. Dans ce premier type d'étude, on se focalise exclusivement sur le contenu du document, soit le critère thématique. Cependant de nombreux autres critères peuvent entrer dans un jugement de pertinence. Le second grand type de recherche effectué sur la pertinence porte précisément sur la mise en évidence de ces critères d'évaluation. En procédant à l'interrogation des utilisateurs sur leurs jugements de pertinence et en analysant leurs réponses, Barry [BARR 94] a par exemple distingué 7 grandes catégories de critères dont, le contenu du document, les expériences de l'utilisateur (connaissances par exemple), les croyances et préférences de l'utilisateur, les autres informations et autres sources (consensus, vérification externe), les sources des documents (qualité de la source), les documents comme entité physique (coût, facilité pour se le procurer), le contexte de l'utilisateur (contraintes de temps par exemple).

Nous nous plaçons ici dans le cadre du premier type de recherche, c'est-à-dire dans la problématique de l'identification de la relation. Un certain nombre d'études ont porté sur une formalisation mathématique de cette notion. La logique et les probabilités sont deux cadres dans lesquels des propositions ont été faites.

Dans un cadre logique, déjà Cooper [COOP 71] insiste explicitement sur le lien étroit entre pertinence et conséquence logique « One might, on first reflection, suppose that logical consequence could be distantly related to relevance, but it will be argued shortly that when problem is posed in terms of declarative sentences, logical consequence and relevance are very intimately connected » ([COOP 71] p. 22). Plus récemment, dans le même ordre d'idée, Van Rijsbergen [VAN 86] propose aussi de considérer la pertinence comme une implication logique entre un document et une requête, soit $D \rightarrow Q$. Pour illustrer cette proposition, considérons trois documents D1, D2 et D3, respectivement représentés par A, B&C et A&B&C (où A,B et C sont des mots clefs et "&" désigne une conjonction). Considérons également une requête Q

représentée par A&B. Le document pertinent est D3, car D1 est trop général, D2 ne porte pas sur A et D3 est bien le seul document tel que D→Q (A&B&C→A&B). Nie [NIE 88] ajoute le rôle important joué par l'implication inverse (Q→D) et propose une agrégation de l'évaluation des deux implications, soit F(P(D→Q),P(Q→D)). Pour illustrer cette idée, en reprenant l'exemple précédent, considérons le document D4 représenté par A&B. Il est manifestement plus en adéquation avec la requête que D3, et le seul moyen de départager D4 de D3 est de considérer D→Q.

Cette formulation logique de la pertinence n'est pas déconnectée des modèles probabilistes utilisés en RI. Van Rijsbergen [VAN 86] suggère que P(D→Q) soit évalué comme une probabilité conditionnelle P(Q|D) et non comme une implication matérielle[1]**.** Même dans les modélisations probabilistes les plus classiques, les deux implications jouent un rôle central [FUHR 92]. Par exemple, on trouve très souvent dans la fonction d'évaluation de la pertinence la mesure (ou une forme derivée) de P(t|Q) - la probabilité de trouver le terme t dans un document pertinent vis-à-vis de Q; et P(t|D) - la probabilité qu'un terme soit dans un document. En supposant l'indépendance entre les termes t, le produit de P(t|Q) pour tous les termes de D est utilisé pour calculer P(D|Q). Fuhr appelle cette approche, l'approche orientée requête (c'est-à-dire que P(t|Q) est estimée pour une requête fixe). D'une façon similaire on obtient P(Q|D) en composant P(t|D) pour tous les termes de la requête dans une approche orientée document.

Même le schéma de pondération tf*idf utilisé dans le modèle de l'espace vectoriel est une forme de ces implications. La fréquence d'un terme dans un document donné (tf) est une forme de P(t|D) et un composant de P(Q|D). L'indice idf qui est basé sur la proportion de documents contenant un terme donné est une forme de P(D|t), un composant de P(D|Q).

On est donc amené à émettre l'hypothèse selon laquelle, la double implication entre D et Q formerait l'essence de la notion de pertinence. Même si ces deux implications ont été mesurées et combinées de différentes façons dans différents modèles.

2.3. L'analogie entre pertinence et résonance

Il est intéressant de noter l'analogie qui existe entre l'idée de résonance et les formalisations de la notion de pertinence proposées en RI. En effet, en considérant que le document D correspond au stimulus (codé dans F1) et que la requête Q correspond aux représentations internes du système (codées dans F2). L'implication D→Q correspond à la propagation d'activation de F1 vers F2 et l'implication inverse à la propagation de F2 vers F1. L'analogie entre la propagation d'activation et l'implication n'est pas une simple analogie de surface. La propagation est réalisée en tenant compte du poids des connections liant les neurones des deux couches. Or, ces connections sont calculées selon une règle d'association qui correspond en quelque sorte à une implication subsymbolique [SMOL 92]. Nous reviendrons dans notre discussion sur les implications de cette identification entre résonance et pertinence.

[1] L'implication matérielle qui est vraie lorsque la prémisse est fausse quelle que soit la valeur de vérité de la conclusion, est souvent remise en cause pour l'évaluation de la pertinence. En particulier, une de ses conséquence est que lorsqu'un document est incohérent, il devient pertinent pour toutes les requêtes Pour faire face à ce problème, certaines logiques comme la logique de la pertinence [ANDE 75] or les logiques conditionnelle [STAL 68] proposent d'autres représentations de l'implication dans lesquelles le désavantage de l'implication matérielle disparaît. Pour des exemples de l'utilisation de ces logiques dans le cadre de la Recherche d'Information on pourra respectivement se reporter à [MEGH 96] et [NIE 96].

3. Description du système RELIEFS

Nous avons expérimenté l'idée de résonance dans la tâche de filtrage de document textuels. Cette tâche consiste à sélectionner dans un flux de documents continu, les documents pertinents vis-à-vis d'une requête préalablement donnée. Nous avons plus particulièrement considéré la tâche de filtrage adaptatif dans laquelle le système s'adapte à partir des jugements de pertinence donnés par l'utilisateur pour tout document sélectionné. On notera que RELIEFS qui reprend les grands principes de ART se situe néanmoins à un niveau de description plus élevé (l'analogie avec le fonctionnement neuronal étant moins poussée).

3.1. Architecture générale

Dans le cadre de la sélection de documents textuels, nous considérerons que les stimuli correspondent aux différents mots des documents et nous considèrerons une unité abstraite désignant la pertinence d'un document (représentant en quelque sorte l'objectif du système c'est-à-dire la requête). On notera qu'une unité de pertinence distincte est créée pour chaque requête différente. Nous allons tenir à jour un réseau liant les unités représentant les différents mots rencontrés dans les documents et l'unité "pertinence" (figure 3).

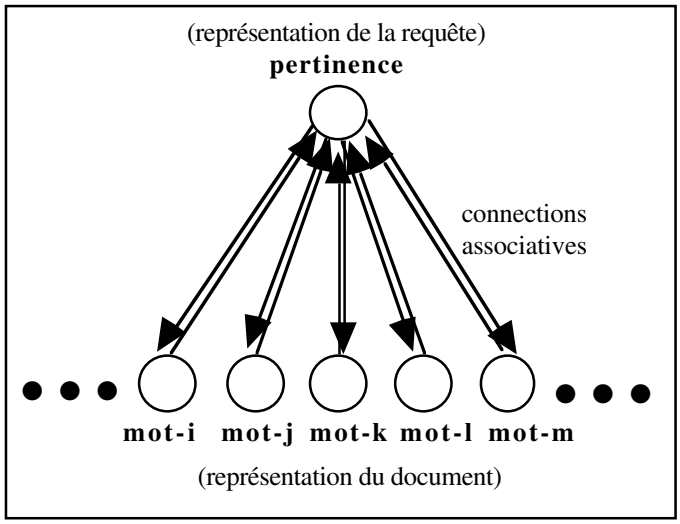

Figure 3. Réseau construit et mis à jour par RELIEFS

On peut résumer le fonctionnement global du système en quelques lignes. Ainsi, lorsqu'un document se présente, les mots présents dans ce document agissent comme des indices de la pertinence du document. Un bon indice est un mot dont la résonance avec l'unité de pertinence est importante (les deux implications sont fortes). La pertinence globale du document est calculée sur la base de la résonance de l'ensemble de ses mots vis-à-vis de l'unité pertinence. Enfin, pour adapter le système au besoin des utilisateurs, le système doit apprendre à partir des jugements fournis par l'utilisateur. Cet apprentissage consiste à modifier la valeur des connexions entre les mots et l'unité pertinence.

3.2. La règle d'apprentissage

Etant donné un document et un jugement de pertinence, nous souhaitons mettre à jour les connections présentes dans la structure. Soient $W_{PQ_{(k-1)}}$ et W_{PQ_k} les évaluations de la connection orientée de l'unité P vers l'unité Q, avant et après la prise en compte de la k-ième observation (un couple document-jugement de pertinence, que l'on note O_k). Soit $\mu_P(O_k)$ une valeur indiquant la

présence ou l'absence de l'objet représenté par l'unité P dans O_k (1 si présent et 0 si absent). La règle d'apprentissage proposée est :

$$W_{PQ_k} = \frac{\alpha\, W_{PQ_{(k-1)}} + \mu_P(O_k)\,\mu_Q(O_k)}{\alpha + \mu_P(O_k)} \qquad avec \;\; \alpha = \sum_{i=1}^{k-1} \mu_P(O_k)$$

Dans notre cas, P et Q corrrespondent respectivement à un mot et à l'unité "pertinence" ou inversement. La règle fonctionne de la façon suivante :

- Si P et Q sont présents dans l'observation, la connection de P vers Q et la connection inverse sont renforcées. On prend ainsi en compte le jugement de l'utilisateur en augmentant le poids des mots apparaissant dans le document jugé pertinent.

- Si P est présent et Q ne l'est pas, la connection de P vers Q est affaiblie et la connection inverse reste inchangée. Si P est un mot et Q l'unité pertinence ceci revient à affaiblir le poids attribué à un mot qui apparaît dans un document jugé comme non pertinent par l'utilisateur. Si P est l'unité pertinence et Q est un mot, ceci revient à affaiblir le poids attribué à un mot qui n'apparaît pas dans un document jugé comme pertinent par l'utilisateur.

Le numérateur de la règle correspond précisément à la forme générale de la règle de Hebb [HEBB 49] dont le principe est associatif (si deux unités sont activées simultanément leurs connections sont renforcées). Le dénominateur permet de normaliser W_{PQ_k}. On peut aisément montrer que W_{PQ_n} représente la fréquence de l'objet représenté par Q sachant la présence de l'objet représenté par P pour les n premières observations [BROU Sous presse]. La règle associative présentée permet donc simplement de calculer de façon incrémentale la fréquence relative de documents pertinents sachant la présence d'un mot ainsi que la fréquence d'un mot dans les documents pertinents.

3.3. La fonction d'évaluation de la pertinence

Dans le but d'évaluer une résonance entre un document et une requête, nous proposons d'évaluer la résonance de chacun des mots présents dans le document. La qualité de résonance d'un mot correspond à sa capacité à propager l'activation vers l'unité "pertinence" et à recevoir une activation de la même unité. Soit W_{iP} le poids de la connections orientée du mot i vers l'unité pertinence et W_{Pi} le poids de la connection inverse. La résonance entre le mot i et la pertinence est évaluée par $W_{iP}.W_{Pi}$ (un produit est choisi pour évaluer la conjonction). L'absence d'un mot est donc d'autant plus pénalisante que ce mot est résonant avec la requête c'est-à-dire que $W_{iP}.W_{Pi}$ est grand. Elle correspond à l'idée que l'absence d'un relai de l'activation entrainerait une dissipation de l'activation et ne favoriserait pas l'émergence d'un état de résonance. La fonction d'évaluation retenue correspond finalement à la somme des produits $W_{iP}.W_{Pi}$ pour chacun des mots présents dans le document. De plus, de façon à normaliser cette somme nous l'avons rapportée à la même somme pour l'ensemble des mots (que ces mots soient présents ou pas dans le document).
Ainsi cette évaluation de la pertinence s'écrit :

$$R(d,q) = \frac{\sum_{i^*=1}^{k} W_{Pi}.W_{iP}}{\sum_{i=1}^{n} W_{Pi}.W_{iP}}$$

où i* sont les indices des mots présents dans le document et n est un nombre fixe de mots choisi pour le terme de normalisation.

4 Expérimentations

Nous avons expérimenté ce système dans la tâche de filtrage adaptatif dans le cadre de la conférence TREC-9. La particularité de cette sous-tache de filtrage est de partir d'un très petit nombre de documents pertinents et de tenter de trouver de nouveaux exemples de documents pertinents tout en essayant de minimiser les erreurs. L'ensemble des documents considéré pour la tâche de filtrage de TREC-9 était la collection OHSUMED extraite de MEDLINE (articles de médecine). Les documents des années 1988-1991 (environ 300000 documents) constituait l'ensemble test et seulement deux documents de 1987 par requête servaient à un pré-apprentissage. Dans le but de pouvoir comparer les systèmes, une fonction d'utilité et une fonction de précision étaient fournies par les organisateurs. Cette année, la fonction d'utilité avait la forme : T9P = 2*R-N, où R représente le nombre de documents pertinents sélectionnés et N représente le nombre de documents non pertinents sélectionnés. La fonction de précision avait la forme T9R = (R/max((R+N),50). L'objectif était donc d'optimiser l'une ou l'autre de ces deux fonctions sur les 63 requêtes fournies avec la collection.

4.1. Adaptation de RELIEFS à la tâche de filtrage

Les principes de RELIEFS tels que nous venons de les décrire dans la section précédente ont constitué l'essentiel du système qui a été utilisé dans la tâche de filtrage adaptatif de TREC-9. Néanmoins, d'autres composants spécifiques à cette tâche devaient être ajoutés. Il s'agit de la politique des seuils, dont l'objectif est de définir le score à partir duquel le système décide de sélectionner le document et d'un module servant à la sélection des mots du texte pour l'évaluation ainsi que pour l'apprentissage (on ne considère donc pas tous les mots du document).

4.1.1. Politique des seuils

Le seuil optimal varie selon la requête et l'apprentissage réalisé sur les précédents documents sélectionnés. Ces modules sont en général conçus de façon empirique. Pour notre part, nous avons considéré une mise à jour incrémentale considérant les deux règles suivantes:

- Si un document sélectionné n'est pas pertinent le seuil est augmenté.

(on considère que le système est trop tolérant)

- Si un document (qu'il soit pertinent ou pas) n'est pas sélectionné le seuil est abaissé.

(on considère que le système est trop sélectif)

L'amplitude de l'augmentation est plus importante que celle de la diminution car cette dernière est réalisée beaucoup plus fréquemment (la majorité des documents ne sont pas sélectionnés). Nous considérons cette diminution de façon à éviter que le système reste silencieux dans le cas où le seuil initial fixé (que nous basons sur le score réalisé par les documents donnés en apprentissage) est trop haut. Dans les deux cas, nous avons considéré différents critères pour adapter l'amplitude de la variation à la situation. Il s'agit de critères locaux comme :

- le nombre de documents non pertinents consécutifs sélectionnés sans documents pertinents. Plus ce nombre est important, plus l'augmentation est forte et la diminution faible.

- le nombre de documents pertinents consécutifs sans documents non pertinents (uniquement pour la diminution). Plus ce nombre est important, plus la diminution est forte.

- le nombre de documents traités (on prend plus de risque au début qu'à la fin). Plus ce nombre est important plus l'amplitude de la variation est forte.

Nous avons aussi considéré un critère plus global portant sur les probabilités de pertinence estimées pour chaque seuil. Le seuil correspondant à une probabilité de 0.33 (qui correspond à la probabilité optimisant la mesure d'utilité) est une valeur de référence vers laquelle le système tend à se régler. L'optimisation des résultats du système consiste principalement à choisir les différentes constantes qui paramètrent ces variations.

4.1.2. Module de sélection des mots

Dans le but d'alléger le traitement réalisé pour chaque document (évaluation et apprentissage) et dans le but d'éliminer le bruit que constitue la présence fortuite d'un mot dans un document pertinent, nous avons considéré d'abord les meilleurs mots du document (toujours sur le critère du produit des fréquences relatives). Si moins de N mots pouvaient être selectionnés de cette façon, cette sélection était complétée par les mots du document les mieux reliés aux mots de la requête. Pour préciser ce dernier point, l'évaluation des mots les mieux reliés à la requête a été calculée sur la mesure de leur double implication évaluée à partir des documents de l'ensemble d'apprentissage (documents de 1987) vis-à-vis de l'ensemble des mots de la requête. La double implication qui correspond ici au produit de la fréquence du mot sachant la présence du mot de la requête avec la fréquence inverse a donc une nouvelle fois été utilisée pour sélectionner les mots "pertinents". Le calcul des fréquences relatives liant les mots de l'ensemble d'apprentissage correspond à la construction d'un thésaurus flou [RADE 76], [MIYA 90], [LUCA 91]. Et la complétion par les mots reliés à la requête est une forme d'extension de requête. Finalement, lorsque moins de N mots restaient sélectionnés à la suite de cette seconde phase les mots du document étaient pris dans leur ordre de lecture. Dans nos expériences nous avons pris N=20.

4.2. Premiers résultats

Pour l'évaluation de TREC-9, nous avons soumis deux résultats, un pour l'utilité et un pour la précision. Après un premier examen, la comparaison sur le critère d'utilité avec les autres systèmes sont favorables puisque environ 80 % de nos scores sont supérieurs ou égaux à la valeur médiane (60 % supérieur) et le système figure parmi les cinq systèmes à avoir une utilité positive (+1.1).

Depuis les tout premiers résultats, nous avons amélioré notre gestion des seuils et nous avons éliminé le bruit correspondant à la présence fortuite de mots dans les documents pertinents en ne considérant que les 50 meilleurs mots pour le calcul du dénominateur de la fonction d'évaluation. Ces simples améliorations nous ont permis d'obtenir un score d'utilité de +8 et un score de précision très proche des résultats obtenus par les meilleurs systèmes à TREC-9. Ces résultats ne constituent pas totalement une surprise puisque comme nous l'avons dit, les autres approches prennent aussi en compte de façon plus ou moins explicite la double implication. Cependant notre mesure isole cet aspect. Nos résultats montrent que l'essence de la pertinence a été bien été prise en compte par la double implication.

4.3. Expériences complémentaires

4.3.1. Utilisation de conjonctions

Pris isolément, certains mots peuvent ainsi être de mauvais indices de la pertinence d'un document alors qu'en conjonction, ils peuvent devenir des indices très efficaces. Nous avons donc expérimenté l'utilisation de conjonctions de mots considérant que des conjonctions de mots permettraient, mieux que des mots simples, de rendre compte du sens contextuel des mots. Globalement, on peut penser que la prise en compte de conjonctions de mots permet de mieux décrire le besoin d'information. Dans cet objectif, nous avons utilisé un algorithme incrémental [BROU Sous presse] pour déterminer si une combinaison doit ou non être ajoutée. Cet algorithme utilise le fait que si, l'on a m1→m2 pour deux mots m1 et m2 (si m1 est présent, alors m2 l'est aussi), il est inutile de construire la conjonction m1&m2 puisque les implications (au sens de fréquences relatives) valables pour m1 le sont aussi pour la conjonction. La prise en compte des conjonctions n'a cependant pas amélioré les résultats (mais ne les a pas dégradé non plus) mais a augmenté considérablement le temps de traitement des documents.

4.3.2. Pondération du rôle relatif des deux implications

Considérant le principe général de notre méthode, nous avons tenté de faire varier l'importance relative des deux implications en considérant $(W_{Pi})^{\alpha}$ au lieu de W_{Pi} dans la fonction d'évaluation qui s'écrit alors :

$$R(d,q) = \frac{\sum_{i^*=1}^{k} (W_{Pi})^{\alpha}.W_{iP}}{\sum_{i=1}^{n} (W_{Pi})^{\alpha}.W_{iP}}$$

Si $\alpha = 1$, l'importance relative des deux implications est identique. Plus α est grand, plus l'importance W_{Pi} devient prépondérante. Il est aussi intéressant de constater que α règle directement la stabilité/plasticité du système. Plus α est grand, plus le système devient stable. Ceci s'explique simplement par le fait qu'en augmentant le poids des mots apparaissant le plus fréquemment dans les documents pertinents précédents, on continuera d'autant plus à les sélectionner (pour l'évaluation du document et l'apprentissage) au détriment des mots moins fréquemment sélectionnés. De nouveaux mots indices de la pertinence de document n'apparaîtront donc pas. Notre première idée était de donner initialement au système une grande plasticité et d'augmenter α avec le nombre de documents pertinents trouvés. Ceci correspond à l'objectif de découvrir de nouveaux indices de la pertinence lorsque la connaissance sur ces indices est très pauvre tout en permettant une fois que ces connaissances sont acquises de minimiser le nombre d'erreurs.

Cependant, nous devions considérer une stabilisation naturelle du système avec le nombre de document pertinents sélectionnés puisque, à chaque fois qu'un document pertinent est sélectionné, le poids W_{Pi} devient un peu moins sensible à la prise en compte d'une observation. Nous n'avons pas obtenu d'amélioration en faisant varier α n'a pas donné d'amélioration. Considérant que la stabilité/plasticité de notre système évoluait naturellement bien, nous avons cherché les valeurs de α (que nous avons donc gardées constantes) pour lesquelles les résultats étaient les meilleurs. Nos expériences ont montré qu'une valeur de α comprise entre 0.8 et 1.2 donnait les meilleurs résultats.

5 Discussion Générale

Pour la conception de ce système nous nous sommes inspirés de principes cognitifs dans un sens semblable à certaines recherches qui ont tenté de reproduire les mécanismes de propagation d'activation pour la RI [CRES 97]. D'une part, la règle d'apprentissage que nous utilisons est de type associatif et reprend un principe dont la réalité neurophysiologique a été mis en évidence par Hebb : "Quand un axone de la cellule A est assez proche pour exciter une cellule B et quand, de façon répétée et persistante, il participe à son activation, un certain processus de croissance ou un changement métabolique s'installe dans une cellule ou dans les deux, tel que l'efficacité de A, en tant qu'elle est une des cellules qui activent B, est augmentée". Par ailleurs, nous avons repris l'idée de résonance développée dans la théorie ART dont la plausibilité psychologique est appuyée par de nombreuses données neurologiques et de nombreux succès dans modélisation des différents processus cognitif [GROS 99]. Ce système apparaît donc cohérent avec les principes cognitifs d'association et de résonance. Cet aspect nous paraît important dans le sens où le rôle d'un tel système est de sélectionner des informations pour un utilisateur, et que l'intégration des principes de fonctionnement de l'utilisateur dans le système peut être profitable.

Par ailleurs, il semble qu'en identifiant résonance et pertinence, on apporte un éclairage nouveau sur la notion de pertinence. La pertinence d'une information apparaît de ce point de vue comme une conséquence du dilemme entre stabilité et plasticité. Une information est ainsi sélectionnée parce qu'elle garantit une réponse satisfaisante à ce problème. Notre approche permet d'une certaine façon de réconcilier les approches formelles de la pertinence avec des approches cognitives tournées vers l'utilisateur puisque nous montrons qu'un mécanisme cognitif de sélection de l'information (la résonance) est cohérent avec les formalisations de la pertinence en RI. Ce point de vue cognitif insiste sur l'aspect subjectif de la pertinence. Il la définie relativement à l'identité du système auquel l'information est présentée, cette identité évoluant elle-même avec l'apprentissage. En ce sens, il apporte un élément de réflexion par rapport au débat opposant "pertinence système" et "pertinence utilisateur" en déplaçant la question de l'objectivité vers celle de la différence de l'identité des systèmes.

D'un point de vue plus pratique, nous avons présenté un nouveau système pour le filtrage adaptatif de documents textuels. Ce système a l'avantage de la simplicité puisqu'il se limite à l'évaluation de fréquences relatives et qu'il les compose très simplement. Les très bons résultats obtenus sur la tâche de filtrage adaptatif de TREC-9, nous confirme la validité de l'approche.

7 Références

[ANDE 75] Anderson, A. R., & Belnap, N. D. *Entailment. The logic of relevance and necessity*. Princeton: Princeton University Press, 1975.

[BARR 94] Barry, C. L. User-defined relevance criteria: An exploratory study. *Journal of the American Societuy for Information Science, 45(3),* 149-159, 1994.

[BROU Sous presse] Brouard, C. Une méthode incrémentale et associative pour l'extraction de règles floues. *Revue d'Intelligence Artificielle, N° spécial*, Sous presse.

[BROU 00] Brouard, C. & Nie, J-Y. *The system RELIEFS: a new approach for information filtering.* Proceedings of the Text Retrieval Conference (TREC-9), Nov 13-16, 2000.

[CLAN 97] Clancey, W. *Situated Cognition*. Cambridge: Cambridge University Press, 1997.

[COOP 71] Cooper, W. S. A definition of relevance for information retrieval. *Information Storage and Retrieval, 7*, 19-37, 1971.

[CRES 97] Crestani, F. Application of spreading activation techniques in information retrieval. *Artificial Intelligence Review, 11*, 453-498, 1997.

[FROE 94] Froehlich, T. J. Relevance reconsidered-toward an agenda for the 21st century: Introduction to special issue on relevance research. *Journal of the American Society for Information Science, 45(3)*, 124-134, 1994.

[FUHR 92] Fuhr, N. Probabilistic models in information retrieval. *The Computer Journal, 35(3)*, 243-255, 1992.

[GREE 95a] Green, R. Topical relevance relationships. I. Why topic matching fails. *Journal of the American Society for Information Science, 46(9)*, 646-653, 1995.

[GREE 95b] Green, R., & Bean, C. A. Topological relevance relationships. II. An exploratory study and preliminary typology. *Journal of the American Society for Information Science, 46(9)*, 654-662, 1995.

[GROS 76] Grossberg, S. Adaptative pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. *Biological Cybernetics, 23*, 121-134, 1976.

[GROS 99] Grossberg, S. The link between brain learning, attention and consciousness. *Consciousness and Cognition, 8*, 1-44, 1999.

[HEBB 49] Hebb, D. O. *The organisation of behaviour*. New York: John Wiley and Sons Inc, 1949.

[LUCA 91] Lucarella, D., & Morara, R. FIRST: Fuzzy information retrieval system. *Journal of Information Sciences, 17(1)*, 81-91, 1991.

[MEGH 96] Meghini, C., & Straccia, U. *A relevance terminological logic for information retrieval*. Proceedings of the 19th Annual ACM Conference on Research and Development in Information Retrieval, Zurich, 1996.

[MIYA 90] Miyamoto, S. Information retrieval based on fuzzy associations. *Fuzzy Sets and Systems, 38*, 191-205, 1990.

[NIE 88] Nie, J. Y. *An outline of a general model for information retrieval*. Proceedings of the 11th Annual ACM Conference on Research and Development in Information Retrieval, Grenoble, 1988.

[NIE 96] Nie, J. Y., & Brisebois, M. An inferential approach to information retrieval and its implementation using a manual thesaurus. *Artificial Intelligence Review, 10*, 409-439, 1996.

[RADE 76] Radecki, T. Mathematical model of information retrieval system based on the concept of fuzzy thesaurus. *Information Processing and Management, 12*, 313-318, 1976.

[SARA 75] Saracevic, T. Relevance: a review of the literature and a framework for thinking on the notion in information science. *Journal of the American Society for Information Science*, 321-343, 1975.

[SMOL 92] Smolensky, P. IA connexionniste, IA symbolique et cerveau. In D. Andler (Ed.), *Introduction aux sciences cognitives*, 79-106, Paris: Gallimard, 1992.

[STAL 68] Stalnaker, R. A theory of conditionals. In N. Rescher (Ed.), *Studies in Logical Theory*. Oxford, 1968.

[VAN 86] van Rijsbergen, C. J. A non-classical logic for information retrieval. *The Computer Journal, 29(6)*, 481-485, 1986.