Interprétation possibiliste d'une analyse de sentiments multicritère pour un système de recommandation sur le web

Possibilistic interpretation of a sentiment analysis for a web recommender system

A. Imoussaten[1] B. Duthil[2] F. Trousset[1] J.Montmain[1]
[1] LGI2P, Ecole des Mines d'Alès
[2] L3i, Université de La Rochelle

Site EERIE – Parc Scientifique G. Besse, 30035 Nîmes Cedex,
abdelhak.imoussaten@mines-ales.fr, francois.trousset@mines-ales.fr, jacky.montmain@mines-ales.fr
Avenue M. Crépeau, 17042 La Rochelle cedex 01,
benjamin.duthil@univ-lr.fr

Résumé :
Cet article propose un système de recommandation (SR) original basé sur l'extraction automatique d'opinions et l'aide à la décision multicritère multi acteurs. Notre SR cherche à optimiser l'automatisation du traitement de l'information disponible sur le web pour éviter aux utilisateurs les phases d'évaluation par critère et d'identification des préférences généralement complexes et fastidieuses dans les SR. Pour exposer notre approche, nous nous focalisons sur la recommandation de films, mais elle resterait valide dans de nombreux domaines. Elle s'applique dès lors que l'on dispose i) d'évaluations globales par la presse spécialisée des entités entre lesquelles il faut choisir et ii) de critiques d'internautes portant sur les différentes caractéristiques des entités rédigées en texte libre. La recommandation repose alors sur l'association d'un internaute et d'un journal spécialisé réputé avoir les mêmes valeurs que l'utilisateur et donc constituer une référence fiable et personnalisée. L'extraction automatique d'opinions multiples ne pouvant fournir que des évaluations imprécises et incertaines sur chaque critère, notre modèle se place ainsi dans le cadre de la théorie des possibilités.

Mots-clés :
Théorie de possibilités, Fusion d'intervalles, Agrégation multicritère, Système de recommandation, Extraction d'opinions.

Abstract:
This paper proposes an original recommender system (RS) based upon an automatic extraction of trends from opinions and a multicriteria multi actors assessment model. Our RS tries to optimize the use of the available information on the web to reduce as much as possible the complex and fastidious steps for multicriteria assessing and for identifying users' preference models. The particular case of RS for moviegoers is dealt to simplify our presentation, but this approach is applicable in many other domains. It may be applied as soon as i) overall assessments of competing entities are provided by trade magazines and ii) web users' critics in natural languages and related to some characteristics of the assessed entities are available. Recommendation is then based on the capacity of the RS to associate a web user with a trade magazine that conveys the same values as the user and thus represents a reliable personalized source of information. Precise assessments related to multiple criteria cannot be expected from automatic opinion mining. Our RS thus relies on possibility theory elements.

Keywords:
Possibility theory, Intervals merging, Multicriteria aggregation, Recommender system, Opinion-mining.

1 Introduction

Ces dernières années, de nombreuses entreprises et sites Web ont mis en place des systèmes permettant l'analyse des préférences de leurs utilisateurs afin de mieux répondre aux attentes de ces derniers. À ce jour, les systèmes de recommandation sont présents dans de multiples domaines comme le tourisme/les loisirs [17], la publicité [3], le commerce électronique [13], les films [19], etc. En raison de l'explosion de la quantité de données diffusées sur l'Internet ces dernières années, rechercher et trouver des produits, des services ou des contenus pertinents est devenu une tâche difficile pour l'utilisateur souvent noyé par la masse d'informations. Ceci

explique l'intérêt croissant porté aux systèmes de recommandation tant par les utilisateurs que par les sites commerciaux.
La problématique de recommandation a été identifiée comme étant le moyen d'aider les individus à trouver des informations, ou des éléments, qui sont susceptibles de les intéresser. Généralement, on considère un ensemble d'utilisateurs *Users* et un ensemble d'éléments *Items* (services ou produits) pouvant être recommandés à chacun des utilisateurs. Par ailleurs, une recommandation multicritère contribue à améliorer la qualité d'un SR car elle permet d'expliciter les caractéristiques pour lesquelles un item a été proposé à l'utilisateur [10], [1], [2] et [18].

Un SR qui utilise un Modèle Multicritère basé sur la découverte des Préférences à partir d'Évaluations (MMPE) permet d'obtenir des recommandations détaillées plus pertinentes vis-à-vis de l'utilisateur [1], [2]. Cependant, la mise en place d'un tel modèle nécessite une base de connaissances où les items sont évalués par rapport à un ensemble de critères. Cette contrainte imposée par le modèle est très lourde pour l'utilisateur. Des approches ont tenté d'automatiser l'évaluation multicritère d'items par des techniques de datamining [20], mais elles restent spécifiques à un domaine d'utilisation, car une adaptation à un autre domaine nécessite un nouvel apprentissage supervisé quasi rédhibitoire (annotation manuelle d'un nouveau corpus d'apprentissage). Cette charge de travail nuit largement à la diffusion des MMPE.
Dans [11] et [12], une méthode non-supervisée d'extraction d'opinions multicritères est proposée. Elle permet aux utilisateurs de s'affranchir de la contraignante évaluation partielle sur chaque critère des items à recommander : les utilisateurs déposent simplement leurs critiques comme à l'accoutumé, sous la forme de textes en langage naturel exprimant leurs opinions, et le système se charge de les analyser automatiquement. Il les décompose d'abord selon les critères d'évaluation (segmentation thématique) avant de calculer les polarités ou opinions pour chacun des extraits résultant de la segmentation (opinion-mining/sentiment analysis). En combinant cette méthode à un système interactif d'aide à la décision multicritère (SIAD), il devient possible d'avoir un système automatisé pour la recommandation. Cependant l'affectation automatisée d'un score à un item pour un critère reste une opération au résultat imprécis fourni sous la forme d'un intervalle. Le SIAD doit donc gérer ensuite les multiples évaluations imprécises issues de l'analyse de sentiments sur chaque critère d'évaluation (fusion d'intervalles), puis agréger sur l'ensemble des critères. Pour exposer notre approche, nous nous focalisons sur la recommandation de films, mais elle resterait valide dans de nombreux domaines. De plus, l'évaluation des films sera limitée à deux critères *scénario* et *acteur* dans cet article.
La section 2 résume l'approche de détection d'opinions que nous avons développée dans [10]. Des critiques d'internautes sont donc extraits et calculés des scores d'opinions relativement à un groupe de critères. La section 3 explique comment fusionner ces scores imprécis pour un critère donné. La section 4 détaille comment agréger les scores obtenus à l'étape précédente pour évaluer les films dans un espace d'évaluation multicritère. La section 5 montre comment ces approches peuvent être combinées pour répondre à la problématique de la recommandation multicritère.

2 Extraction d'opinions

Sur la base de méthodes statistiques, notre approche construit un dictionnaire de descripteurs d'opinion pour une thématique donnée. Ce dictionnaire est ensuite utilisé pour extraire automatiquement la polarité d'extraits de texte. Deux étapes sont distinguées dans cette évaluation multicritère :
– dans un premier temps, les portions de texte se rapportant à l'un des critères d'évaluation sont extraites avec notre approche *Synopsis* décrite dans [9]. Il s'agit donc de segmenter le texte par thème (ici, les critères) ;

– ensuite, pour chaque critère, la polarité des segments qui lui ont été affectés par *Synopsis* est calculée. Il s'agit là de l'étape d'opinion-mining à proprement parler [12].

2.1 Segmentation thématique par critère

L'approche non supervisée *Synopsis* est utilisée pour identifier les extraits de texte qui traitent des critères considérés et se compose de 3 étapes :

1. Construction automatique du corpus d'apprentissage utilisé pour apprendre les mots caractéristiques d'un critère ;
2. Apprentissage automatique de ces mots caractéristiques, appelés descripteurs, et construction d'un lexique associé au critère ;
3. Segmentation du texte en utilisant le lexique associé au critère.

L'approche utilise un ensemble de mots appelés « mots germes ». Ces mots germes permettent d'une part de caractériser sémantiquement le critère considéré, et d'autre part d'amorcer l'apprentissage des descripteurs caractéristiques d'un critère [9]. La construction du corpus d'apprentissage est automatique. Très schématiquement, plus un mot est fréquemment repéré dans le voisinage d'un mot germe du critère (comptage sur fenêtre glissante), plus on considère comme « forte » l'appartenance (sémantique) de ce mot au champ lexical du critère.

Synopsis construit son corpus d'apprentissage à partir de documents du web. La phase de segmentation permet également de gérer l'imprécision du langage en proposant non pas une segmentation unique pour un texte donné, mais plusieurs segmentations possibles pour des granularités de langage différentes [9]. Cette granularité correspond à l'idée que les passages identifiés comme traitant d'un thème (critère ici) peuvent varier en fonction du niveau d'expertise de l'utilisateur. Les méthodes statistiques qui supportent notre approche permettent de construire un dictionnaire de descripteurs caractéristiques d'une thématique donnée et peuvent être paramétrées afin de contrôler la taille de ce dictionnaire (i.e., le niveau d'expertise). Plus l'utilisateur veut être certain de l'appartenance des segments de texte à un critère, plus le vocabulaire appris devra être sélectif. L'appartenance d'un descripteur au champ lexical du critère résulte du calcul d'une fonction de discrimination ; un seuil sur cette fonction permet de définir l'étiquette « forte » sur l'appartenance au critère. Plus ce seuil est haut, plus la phase d'apprentissage des descripteurs du dictionnaire de *Synopsis* est sélective (figure 1). Les seuils retenus sont calculés automatiquement à partir de l'analyse des paliers de la fonction d'appartenance au critère [10].

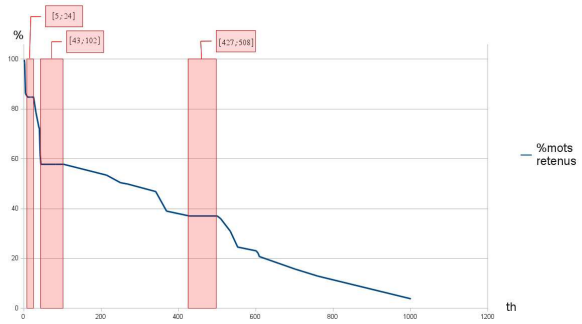


Figure 1 : pourcentage de mots retenus en fonction du seuil de discrimination

2.2 Analyse de sentiments

L'approche d'extraction d'opinions (ou d'analyse de sentiments) est une adaptation de l'approche *Synopsis* à l'extraction d'opinions pour les phases de construction automatique du corpus d'apprentissage et d'apprentissage des descripteurs. Notons que les descripteurs d'opinion appris sont généralement propres au domaine d'analyse (ici le cinéma) [14]. Les mots germes deviennent des mots germes d'opinion [12]. Deux ensembles de mots d'opinion sont initialement distingués les positifs (P) et les négatifs (N) :

P = {good, nice, excellent, positive, fortunate, correct, superior};
N = {bad, nasty, poor, negative, unfortunate, wrong, inferior}.

Partant de l'hypothèse qu'un document qui contient au moins un mot de P (resp. N) et aucun de N (resp. P) est porteur d'une opinion positive (resp. négative), un ensemble de

documents associé à un mot germe est construit à l'aide d'un moteur sur le web comme dans *Synopsis* (requête pour le mot germe « good » dans le cinéma : *movie + good − bad − nasty − negative − poor − unfortunate − wrong − inferior*). Les descripteurs d'opinions sont limités aux adjectifs et groupes adjectivaux [10], [12]. Les techniques de filtrage statistiques par fenêtre glissante de *Synopsis* sont adaptées au comptage des descripteurs d'opinions.

Enfin, notre approche calcule pour un texte segmenté, un score d'opinion pour un critère sous la forme d'une somme pondérée des appartenances des segments au critère et aux descripteurs positifs ou négatifs. Ainsi, puisque l'extraction des segments thématiques dépend du seuil de précision exigé, le score d'opinion du texte pour le critère sera lui aussi affecté par ce seuil. Nous avons mis au point un algorithme de segmentation suffisamment performant pour calculer plusieurs segmentations d'un texte en fonction du seuil (échantillonnage) sur la fonction d'appartenance au critère. Ce résultat nous permet donc de calculer l'expression d'opinion du texte sur le critère d'analyse pour plusieurs valeurs du paramètre modélisant le niveau de connaissance de l'utilisateur. Il existe une borne inférieure et une borne supérieure pour ce score. Par conséquent, le score d'opinion est une entité imprécise que nous représentons par un intervalle dont les extrémités sont les bornes précédentes [10]. Cet intervalle sera noté par la suite I_i^j où *i* est le $i^{\text{ème}}$ critère et *j* la $j^{\text{ème}}$ critique (texte).

3 Fusion d'intervalles

A l'issue de l'analyse de sentiments automatique, une critique fournit un score imprécis relativement à un critère sous la forme d'un intervalle dont les valeurs correspondent à différentes segmentations possibles du texte (obtenues en faisant varier la précision de l'extraction comme décrit dans la section 2). Cette imprécision permet d'intégrer la subjectivité des critiques dans le processus d'évaluation d'un critère. Cette « subjectivité » techniquement liée à la précision de l'extraction est « homogénéisée » par le traitement automatique de la segmentation. La multiplicité d'avis imprécis recueillis automatiquement pour un critère est par ailleurs source d'incertitude sur l'évaluation du critère. Pour synthétiser ces avis, faciliter leur représentation et leur manipulation dans la suite du processus de recommandation tout en respectant l'imprécision et l'incertitude des scores, la théorie des croyances fournit le cadre approprié. Les distributions de possibilité donnent une bonne approximation des fonctions de croyances et sont très pratiques des points de vue de la représentation et du calcul.

Soient un espace d'évaluation discret et fini Ω et 2^Ω l'ensemble des parties de Ω. Nous proposons un bref rappel sur la théorie des possibilités [7] et la théorie des croyances [4], [23].

3.1 Théorie des possibilités

Une distribution de possibilité π est une fonction définie sur l'ensemble Ω à valeurs dans [0,1] telle que : $\sup_{\omega\in\Omega}\pi(\omega)=1$. A partir de π, sont définies respectivement deux mesures, une mesure de possibilité : $\Pi(A)=\sup_{\omega\in A}\pi(\omega)$ et une mesure de nécessité : $N(A)=1-\Pi(A^c)$ $\forall A\subseteq\Omega$. $\Pi(A)$ quantifie à quel degré l'événement A est plausible alors que $N(A)$ quantifie la certitude de A. Une $\alpha-$coupe d'une distribution de possibilité π est l'ensemble : $E_\alpha(\pi)=\{\omega\in\Omega \ / \ \pi(\omega)\geq\alpha\}$, $\alpha\in]0,1]$.

3.2 Théorie des croyances

Une *masse de croyance* notée m, est une fonction de 2^Ω dans [0,1] telle que : $\sum_{A\in P(\Omega)} m(A)=1$ et $m(\varnothing)=0$. Un ensemble E

de Ω tel que $m(E) > 0$ est appelé *élément focal*. L'ensemble des éléments focaux est noté $\mathbb{F}$. La fonction m peut être représentée par deux mesures : la *fonction de croyance* $Bel(A) = \sum_{\{E \in \mathbb{F} / A \supseteq E\}} m(E)$ où $\{E \in \mathbb{F} / A \supseteq E\}$ est l'ensemble des témoignages en faveur de *A* et la *fonction de plausibilité* $Pl(A) = \sum_{\{E \in \mathbb{F} / A \cap E \neq \varnothing\}} m(E)$ où $\{E \in \mathbb{F} / A \cap E \neq \varnothing\}$ est l'ensemble des éléments focaux qui ont une relation avec *A* et le rendent possible. Les éléments focaux peuvent représenter des observations imprécises sous forme d'intervalles. Dans ce cas, la probabilité d'un événement $A \subseteq \Omega$ noté $\Pr(A)$, est imprécise et elle appartient à l'intervalle $[Bel(A), Pl(A)]$ ($Bel(A)$ et $Pl(A)$ représentent respectivement la probabilité inférieure et la probabilité supérieure de *A*).

3.3 Fusion possibiliste

Les règles de combinaison classiques présupposent que les sources sont toutes fiables (mode conjonctif) ou au moins une source est fiable (mode disjonctif) [4]. Plusieurs autres règles dites de *compromis* entre les modes conjonctifs et disjonctifs ont été proposées [8]. La règle utilisée ici pour construire les distributions de possibilité est de forme *compromis*.

Soit $\{I_j \mid j = 1, NI\}$ l'ensemble des intervalles distincts en sortie de l'analyse de sentiments pour un critère. La probabilité d'occurrence d'un intervalle I_j est considérée comme sa masse de croyance $m(I_j)$.

Dans le cas de singletons, les évaluations sont précises, nous avons :

$\forall A \in P(\Omega), \Pr(A) = Bel(A) = Pl(A)$.

Lorsque les intervalles sont emboîtés $I_1 \subseteq I_2 \subseteq .... \subseteq I_{NI}$, une distribution de possibilités π est calculée à partir de la mesure de *plausibilité* : $\forall \omega \in \Omega$, $\pi(\omega) = Pl(\{\omega\}) = \sum_{j=1,NI} m(I_j).1_{I_j}(\omega)$.

Lorsque les intervalles sont seulement cohérents ($\bigcap_{j=1,NI} I_t = I \neq \varnothing$: les opinions partagent au moins une valeur), mais pas emboîtés, deux distributions de possibilité sont construites. D'abord, $\forall \omega \in \Omega$, $\pi_1(\omega) = \sum_{j=1,NI} m(I_j).1_{I_j}(\omega)$. Ensuite, on note $E_s = E_{s-1} \cup E_{\alpha_s}(\pi_1)$ $(s = 2, r)$ $(E_1 = I)$ les nouveaux intervalles emboités construits à partir de $\{I_j \mid j = 1, NI\}$. Dans [6] les masses de croyances $m(E_s)$ sont calculées selon $m(E_s) = \sum_{\{I_j \text{ associé à } E_s\}} m(I_j)$ (chaque opinion I_j est associée de façon unique au plus petit ensemble E_s qui peut la contenir). Ainsi, une distribution de possibilité π_2 peut être définie : $\forall \omega \in \Omega$, $\pi_2(\omega) = \sum_{s=1,r} m(E_s).1_{E_s}(\omega)$. Les distributions π_1 and π_2 sont les meilleures approximations inférieure et supérieure pour les opinions $\{I_j \mid j = 1, NI\}$) [6].

Les données ne sont généralement ni précises, ni cohérentes. Les probabilités d'une part et la possibilité-nécessité, d'autre part, correspondent à des situations extrêmes et idéales [6]. Dans notre problématique, les opinions exprimées par les internautes peuvent évidemment correspondre à des évaluations contradictoires. Cela signifie que la contrainte de cohérence n'est pas toujours vérifiée dans la pratique, *i.e.*, $\bigcap_{j=1,NI} I_j = \varnothing$. Pour faire face à cette situation, [16] propose une méthode qui consiste à considérer la cohérence des groupes d'intervalles avec une intersection non vide à partir des intervalles initiaux $\{I_j, j = 1, NI\}$, ce qui revient à chercher les sous-ensembles $K_\beta \subset \{1, ..., NI\}$, $\beta = 1, g$ tels que :

$\bigcap_{j \in K_\beta} I_j \neq \varnothing$, où g est le nombre de sous ensembles K_β. Chaque sous-ensemble d'intervalles est appelé Sous ensemble Maximal Cohérent (SMC). Cette notion a été initialement introduite dans [21], comme

moyen de raisonnement dans le cas de bases logiques inconsistantes et a été étendue aux cas de représentations probabilistes imprécises dans [24] et [5]. Pour chaque groupe K_β, deux distributions sont construites π_1^β et π_2^β, les approximations possibilistes inférieure et supérieure de $\{I_j \mid j \in K_\beta\}$. Soit π_1 (resp. π_2) la distribution de possibilité union des distributions de possibilité $\pi_1^\beta \mid \beta = 1, g$ (resp. $\pi_2^\beta \mid \beta = 1, g$) :

$$\pi_1 = \tilde{\bigcup}_{\beta=1,g} \pi_1^\beta \text{ (resp. } \pi_2 = \tilde{\bigcup}_{\beta=1,g} \pi_2^\beta \text{)}, \quad (1)$$

alors π_1 et π_2 sont les approximations (possibilistes) inférieure et supérieure des données $\{I_j \mid j = 1, NI\}$. Les distributions π_1 et π_2 sont dites *g*-modales.

Pour de simples raisons pratiques de gestion de la complexité et de l'interprétation, on peut ne garder qu'une seule distribution, synthèse approchée des deux distributions π_1 et π_2 (ex.: une approximation affine par morceaux de type moyenne). L'idéal est de garder les deux distributions inférieure et supérieure tout au long du processus de fusion et d'introduire une fonction de décision sur la conclusion portée par chacune des deux distributions au moment du choix final (ex.: un ordre de Pareto sur la valeur moyenne et l'imprécision moyenne des distributions).

On peut aussi choisir l'une des distributions π_1 et π_2 en fonction de l'attitude souhaitée par rapport au risque inhérent à la décision : raisonner avec la distribution inférieure (resp. la distribution supérieure) peut correspondre à une attitude d'aversion au risque lié à la croyance dans l'information (resp. une attitude flexible d'acceptation du risque).

Dans tous les cas de figures, les calculs de la section 4 restent les mêmes pour chacune de ces distributions. Ainsi, le traitement n'est illustré dans cet article que pour une distribution indifférenciée notée π.

4 Agrégation et modèle de préférences

Lors d'une évaluation multicritère (*n* critères), l'agrégation est une opération qui permet de modéliser le système de préférence du décideur (ordre partiel sur $\mathbb{R}^n$) par une forme analytique et l'ordre classique sur $\mathbb{R}$. L'objectif est alors de déterminer la fonction d'agrégation que le décideur utilise, et qui synthétise l'expression de ses préférences lors de l'évaluation. L'opérateur le plus communément utilisé est la moyenne pondérée, elle permet de donner des rôles non symétriques aux critères de l'agrégation. Pour une moyenne pondérée MP_ω il s'agit d'identifier le vecteur des poids relatifs à chacun des critères $\omega = (\omega_1, ..., \omega_n) \in [0,1]^n$ qui la caractérise.

La méthode des moindres carrés est classiquement utilisée pour les quantités agrégées précises. Imaginons que le décideur ait affecté une note globale à un ensemble d'items constituant la base d'apprentissage pour l'identification de son système de préférence. Soit x^k la note agrégée des évaluations pour un item k. La méthode des moindres carrés recherche le vecteur de poids ω^* qui minimise l'écart entre x^k et la valeur calculée avec le modèle linéaire MP_ω sur l'ensemble des items de la base d'apprentissage :

$\omega^* = Arg \min_{(\omega_1,...,\omega_n)\in[0,1]^n} \sum_{k \in items} (MP_\omega(x_1^k, ..., x_n^k) - x^k)^2$ où

$MP_\omega(x_1^k, ..., x_n^k) = \sum_{i=1}^{n} \omega_i . x_i^k$, n est le nombre de critères et ω_i le poids associé au i^{ème} critère. La solution de ce problème d'optimisation fournit le jeu de poids des critères modélisant *au mieux* la stratégie du décideur.

La presse spécialisée fournit en général une simple évaluation globale x^k d'un film accompagnée d'une critique plus ou moins détaillée selon le journal. Il est donc difficile pour l'utilisateur de savoir si les raisons

(critères) pour lesquelles un film est bien perçu par un journal correspondent à ses critères personnels. Comme nous l'évoquions en introduction, fournir les scores partiels x_i^k est une tâche qui rebute les internautes. Ceux-ci veulent bien participer à un RS collaboratif, mais préfèrent exprimer leur opinion en langage naturel sous la forme d'une critique unique, évoquant dans leur texte les seuls critères qui les ont marqués en bien ou en mal pour le film qu'ils analysent. C'est la raison pour laquelle nous avons voulu automatiser l'extraction de critères et l'analyse d'opinions. En revanche, les scores établis automatiquement à partir des critiques sont imprécis. Cette section propose une technique d'identification des poids adaptée à la nature des données dont on dispose dans notre problématique.

La stratégie que l'on cherche à identifier est celle d'un journal spécialisé : quels sont les critères qui caractérisent les valeurs véhiculées par ce journal ? Un utilisateur saura ensuite quel journal est le plus à même de lui fournir des recommandations adaptées à son propre système de choix.

Un journal spécialisé attribue des évaluations globales à des films à l'aide d'échelles du type : nombre d'étoiles, nombre de barres, etc. Ces scores sont imprécis et nous les modélisons à l'aide d'une partition de $\Omega = [0, 20]$ en distributions de possibilité trapézoïdales : $\tilde{\pi}^k, k \in items$.

Selon la méthode de fusion présentée dans la section précédente, nous calculons à partir des opinions extraites du web, pour chaque critère i et chaque item k, une distribution de possibilité $\pi_i^k, i \in \{1,...,n\}, k \in items$ (ou $\pi_{1,i}^k$ et $\pi_{2,i}^k$ si l'on ne fait pas d'approximation des distributions inférieure et supérieure). Notons $\bar{\pi}^k = MP_\omega(\pi_1^k,...,\pi_n^k)$ l'agrégation des distributions $\pi_1^k,...,\pi_n^k$ pour un item k par l'opérateur d'agrégation MP_ω. On cherche alors pour un journal donné la distribution de poids ω qui fera correspondre au mieux $\tilde{\pi}^k$ et $\bar{\pi}^k$ sur l'ensemble des items évalués. Autrement dit, la question que l'on se pose, est : connaissant les $\pi_1^k,...,\pi_n^k$ quelle distribution ω permettrait d'inférer au mieux l'avis du journal $\tilde{\pi}^k$? Mathématiquement, la question se ramène à : quelle distribution de poids maximise $\bar{\pi}^k \Rightarrow \tilde{\pi}^k$ pour un journal donné sur l'ensemble des items évalués ?

En logique floue, il existe plusieurs propositions pour l'implication. Les implications de Willmott, Mamdani, Rescher-Gaines, Kleene-Dienes, Brouwer-Gödel, Goguen, Lukasiewicz, Fodor, Reichenbach sont soit restrictives soit trop permissives [15], [22]. Un compromis consiste à calculer la surface commune entre $\tilde{\pi}^k$ et $\bar{\pi}^k$ telle que cette mesure représente leur pourcentage de recouvrement :

$$Imp(\bar{\pi}^k, \tilde{\pi}^k) = 100.\int \wedge(\bar{\pi}^k, \tilde{\pi}^k)) / \int \tilde{\pi}^k \,. \qquad (2)$$

Pour les distributions de possibilité multimodales, nous avons décidé de ne considérer que le mode qui représente l'opinion de la majorité des internautes dans le contexte de cet article.

Nous proposons maintenant de déterminer les poids de chacun des critères par l'intermédiaire d'une analyse de sensibilité. Cette dernière consiste à faire varier les poids de chacun des critères et d'évaluer, pour chaque variation des ω_i, l'implication de $\tilde{\pi}^k$ par $\bar{\pi}^k$ sur l'ensemble des items. La distribution de poids ω^* retenue pour un journal donné est :

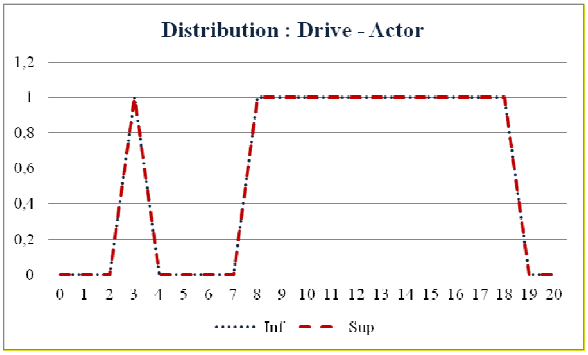
$$\omega^* = \max_{(\omega_1,...,\omega_n)\in[0,1]^n} \sum_{k \in items} (\bar{\pi}^k \Rightarrow \tilde{\pi}^k) \qquad (3)$$

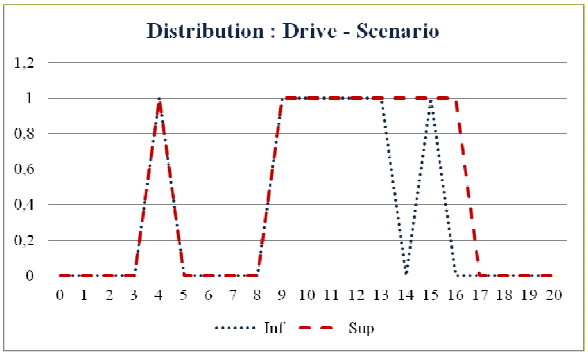
Ce résultat donne les critères prioritaires pour un journal spécialisé donné.

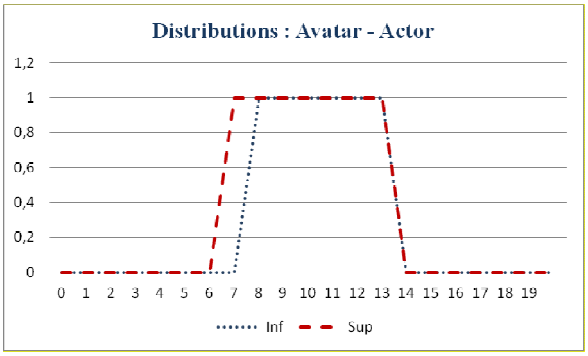
5 Cas d'étude

Le prototype logiciel qui supporte notre système de recommandation de films repose sur l'association d'un internaute avec un journal spécialisé qui porte les mêmes priorités ou valeurs que lui. Ce prototype utilise le module d'extraction d'opinions multicritères

de la section 2, ainsi qu'une base de critiques de films écrites en langage naturel, issues du célèbre site de critiques cinématographiques IMDB. Les critiques fournies par IMDB (environ 3000 critiques par film) fournissent assez d'information pour avoir une idée représentative de la diversité des opinions des internautes sur un film. Chaque critique de film a été scorée par notre système d'extraction d'opinions multicritères selon deux critères : *acteur* et *scénario*. Avec le processus de fusion de la section 3, nous obtenons par exemple pour les films « *Drive* » et « *Avatar* », à partir de 324 opinions extraites pour le premier et 1191 pour le second, les distributions des deux critères (section 3) :







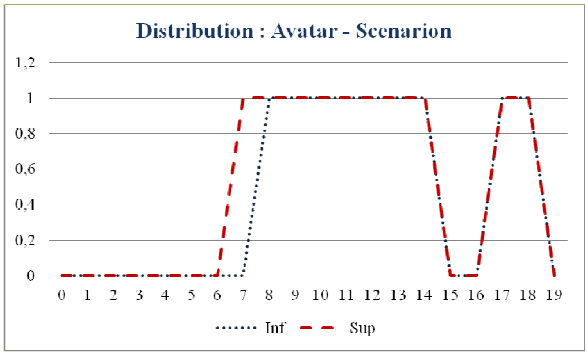


Figure 2 – Distributions de possibilité des critères pour les films « Drive » et « Avatar »

Nous disposons par ailleurs, pour chacun des films de la base, d'évaluations globales sous la forme d'un nombre d'étoiles (transformées en $\tilde{\pi}^k$) attribuées par la presse spécialisée. Nous avons retenu 20 journaux dans cette application.

Pour chaque journal est calculée la distribution de poids qui synthétise le mieux la stratégie d'évaluation du journal, c'est-à-dire la distribution qui explique au mieux les notes $\tilde{\pi}^k$ attribuées par le journal aux films de la base d'apprentissage. Sur cet exemple, la stratégie d'évaluation d'un journal est modélisée par les poids qu'il accorde aux deux critères (figure 3). On remarque qu'il existe de grandes différences entre les politiques d'évaluation des journaux (par exemple pour *Les Cahiers du Cinéma*, les poids sont de 0.7 et 0.3 alors qu'à l'inverse ils sont de 0.3 et 0.7 pour *Charlie Hebdo*).
Il est évident que la conception d'un système de recommandation opérationnel nécessite que bien d'autres critères que les deux retenus *acteur* et *scénario* soient intégrés dans notre modèle pour une recommandation plus fiable. L'automatisation des phases de segmentation et d'opinion-mining que nous proposons laisse néanmoins envisager sereinement cette extension dans notre démonstrateur.
L'utilisation de ces résultats pour de la recommandation dépend de l'information dont on dispose sur l'utilisateur (préférences, évaluations antérieures, etc.). A minima, la figure 3 fournit à l'internaute les critères importants selon les 20 journaux de la base. Il

peut alors simplement choisir le journal qui véhicule les valeurs qui s'approchent le plus de son humeur du moment et aller consulter le hitparade de ce journal pour une recommandation personnalisée. Il n'y a ainsi pas de modèle de préférence de l'utilisateur à identifier, c'est lui-même qui repère le journal qui lui convient le mieux. Notons que ce principe simple permet à l'utilisateur de changer de « système de préférence » selon son humeur chaque fois qu'il va au cinéma !

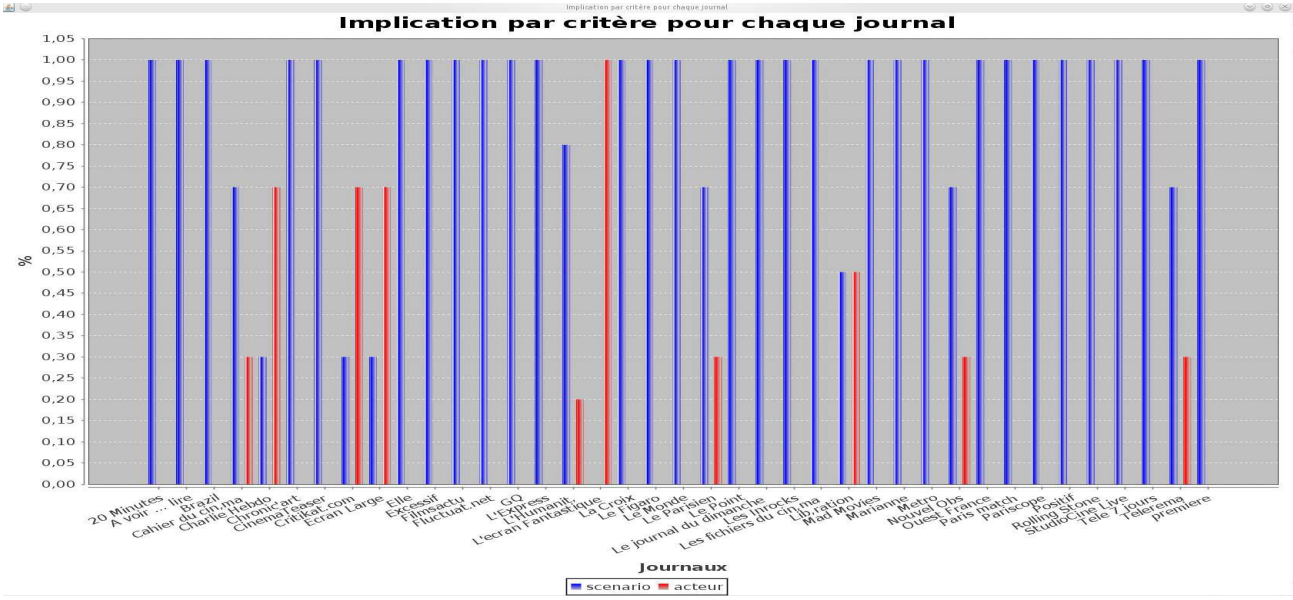


Figure 3 – Stratégies d'évaluation des 20 journaux de la base

Il est également possible d'imaginer que l'on identifie un modèle de préférence de l'utilisateur sur une base de films mythiques réduite ou sur des critiques qu'il aurait postées antérieurement. Tout dépend de l'investissement que l'on souhaite des utilisateurs du système collaboratif de recommandation. On calculerait ensuite une distance entre les distributions de poids des journaux et de l'utilisateur pour la recommandation.

Dans nos futurs travaux, nous projetons de garder tous les modes calculés par la méthode de fusion pour chaque critère et mettre en place un calcul des poids qui tienne compte des modes afin de préserver au mieux la disparité des avis des internautes.

6 Conclusion

Dans cet article, nous avons proposé un système de recommandation qui permet d'extraire automatiquement de critiques en langage naturel des opinions (scores) relativement à un ensemble de critères. Nous avons expliqué pourquoi les scores restitués étaient imprécis. Nous avons alors développé une méthode de fusion pour calculer l'opinion collective des internautes sur un film au regard d'un critère donné. Enfin, nous avons proposé une méthode d'identification de la stratégie d'évaluation d'un journal spécialisé. La recommandation la plus simple qui soit avec notre système consiste alors à montrer les critères prioritaires de la presse spécialisée à l'utilisateur afin qu'il choisisse le journal selon son humeur.

Cette approche permet donc de développer des systèmes de recommandation de type MMPE en s'affranchissant des tâches les plus contraignantes de ce type de systèmes collaboratifs. Ceci constitue une avancée importante car jusqu'à présent la nécessité d'évaluer manuellement un grand nombre de documents et cela selon plusieurs critères, représentait un frein majeur à la mise en place

de tels systèmes. La chaîne de traitement que nous proposons établit un processus d'automatisation cognitive qui laisse raisonnablement envisager un large déploiement et la généralisation des RS de type MMPE dans les années à venir pour traiter des données issues du web.

Références

[1] G. Adomavicius, Y. Kwon. New recommendation techniques for multicriteria rating systems, *Intelligent Systems, IEEE*, 2007, **22**(3), pp. 48-55.

[2] G. Adomavicius, N. Manouselis, Y. Kwon. Multi-criteria recommender systems. *In Recommender systems handbook*, 2011, pp.769-803, Springer US.

[3] K. W. Cheung, J. T. Kwok, M. H. Law, K. C. Tsui. Mining customer product ratings for personalized marketing. *Decision Support Systems*, 2003, **35**(2), pp.231-243.

[4] Dempster, A.P. Upper and Lower probabilities induced by multivalued mapping, *Annals of Mathematical Statistics, 1967*, **38**, pp. 325–339.

[5] S. Destercke, D. Dubois, E. Chojnacki. Possibilistic information fusion using maximal coherent subsets, *IEEE Transactions on Fuzzy Systems, 2009*, **17**, pp. 79–92.

[6] D. Dubois, H. Prade. Fuzzy sets and statistical data, *European Journal of Operational Research, 1986*, **25**, pp. 345–356.

[7] D. Dubois, H. Prade. Possibility Theory: An Approach to Computerized Processing of Uncertainty, *Plenum Press, New York*, 1988.

[8] D. Dubois, H. Prade. Representation and combination of uncertainty with belief functions and possibility measures. *Computational Intelligence, 1988*, **4**(3), pp. 244-264.

[9] B. Duthil, F. Trousset, M. Roche, G. Dray, M. Plantié, J. Montmain, P. Poncelet. Towards an automatic characterization of criteria, *In Database and Expert Systems Applications, DEXA*, pp. 457-465, Springer Berlin Heidelberg, 2011.

[10] B. Duthil. De l'extraction des connaissances à la recommandation. *Université de Montpellier II, Thèse de doctorat*, 2012.

[11] B. Duthil, F. Trousset, G. Dray, P. Poncelet, J. Montmain. Extraction d'opinions appliquée à des critères, *EGC, 2012*.

[12] B. Duthil, F. Trousset, G. Dray, J. Montmain, P. Poncelet. Opinion extraction applied to criteria. *DEXA, 2012*, Part II, LNCS 7447, pp. 489--496. Springer, Heidelberg, 2012.

[13] R. Ghani, A. Fano. Building recommender systems using a knowledge base of product semantics. *In Proceedings of the Workshop on Recommendation and Personalization in ECommerce, 2nd International Conference on Adaptive Hypermedia and Adaptive Web based Systems, 2002*, pp. 27-29.

[14] A. Harb, M. Plantié, G. Dray, M. Roche, F. Trousset, P. Poncelet. Web opinion mining: how to extract opinions from blogs? *International Conference on Soft Computing as Transdisciplinary Science and Technology, 2008*, pp. 211-217, Cergy, France.

[15] I. Iancu. A Mamdani Type Fuzzy Logic Controller, *In: E. P. Dadios, Ed., Fuzzy Logic: Controls, Concepts, Theories and Applications, InTech Croatia, Rijeka, 2012*, pp. 55-54.

[16] A. Imoussaten. Modélisation et pilotage de la phase de délibération dans une décision collective - vers le management d'activités à risques, *Mines ParisTech, Thèse de doctorat*, 2011.

[17] S. Loh, F. Lorenzi, R. Saldaña, D. Licthnow. A tourism recommender system based on collaboration and text analysis. *Information Technology & Tourism*, 2003, **6**(3), pp. 157-165.

[18] N. Manouselis, C. Costopoulou. Analysis and classification of multi-criteria recommender systems. World Wide Web, 2007, **10**(4), pp. 415-441.

[19] M. Plantié, J. Montmain, G. Dray. Movies recommenders systems: automation of the information and evaluation phases in a multi-criteria decision-making process. *16th International Conference on Database and Expert Systems Applications, DEXA'05, 2005*, Copenhague.

[20] M. Plantié. Extraction automatique de connaissances pour la décision multicritère, *Université Jean Monnet, Thèse de doctorat*, 2006.

[21] N. Rescher, R. Manor. On inference from inconsistent premises, *Theory and Decision, 1970*. **1**, pp. 179-219.

[22] D. Ruan, E. Kerre. Fuzzy implication operators and generalized fuzzy method of cases. *Fuzzy Sets and systems, 1993*, **54**(1), 23-37.

[23] G. Shafer. A mathematical theory of evidence, *Princeton University Press*, 1976.

[24] P. Walley. The elicitation and aggregation of beliefs. *PhD thesis, University of Warwick*, 1982.