
Suggestion contextuelle composite

Thibaut Thonet* — **Romain Deveaud**** — **Iadh Ounis***** — **Craig Macdonald*****

* *IRIT, Université Paul Sabatier, Toulouse, France.*
thibaut.thonet@irit.fr

** *Yellow Pages Group, Montréal, Canada. Travail effectué lorsque l'auteur était affilié à University of Glasgow.*
romain.deveaud@pj.ca

*** *School of Computing Science, University of Glasgow, Glasgow, Royaume-Uni.*
{iadh.ounis, craig.macdonald}@glasgow.ac.uk

RÉSUMÉ. La suggestion contextuelle consiste à recommander à un utilisateur un ensemble de lieux d'activités adaptés à ses préférences et à son contexte. La plupart des approches existantes considèrent uniquement ces deux caractéristiques pour constituer leur liste de suggestions. Cependant, les recherches en systèmes de recommandation ont récemment souligné l'importance de la diversité des suggestions. Cet article présente un modèle novateur de suggestion contextuelle inspiré de la recherche composite qui consiste à regrouper les suggestions en différentes grappes thématiquement cohésives. L'évaluation réalisée dans le cadre de la piste Contextual Suggestion de TREC 2013 et 2014 montre que notre approche est compétitive et permet d'améliorer la diversité des suggestions sans dégrader leur pertinence.

ABSTRACT. Contextual suggestion aims to provide a user with venues that are relevant to his/her preferences and context. Most existing approaches only take into account those two features in order to build their suggestion list. However, research in recommender systems recently stressed the importance of diversified suggestions. This paper introduces a novel model for contextual suggestion based on composite retrieval, which consists of regrouping suggestions into different, topically cohesive bundles. The evaluation conducted through the TREC 2013 and 2014 Contextual Suggestion track shows that our approach is competitive and improves diversity within suggestions without degrading relevance.

MOTS-CLÉS : Suggestion contextuelle, Recherche composite, Grappe, Diversité.

KEYWORDS: Contextual suggestion, Composite retrieval, Bundle, Diversity.

1. Introduction

La création d'un ensemble de dispositifs portables tels que les *smartphones* et les tablettes a impacté considérablement les possibilités de recherche d'information en situation de mobilité. L'atelier SWIRL 2012 (*Second Strategic Workshop on Information Retrieval in Lorne*) a mis en évidence ce besoin en nouveaux systèmes de recherche d'information (Allan *et al.*, 2012), qui doivent être capables d'anticiper le besoin de l'utilisateur et renvoyer des informations pertinentes vis-à-vis du contexte, sans nécessiter la formulation explicite d'une requête. Un exemple de scénario dans lequel un tel besoin apparaît est le suivant : un utilisateur se trouve dans une ville donnée et il souhaite se divertir. Il serait alors profitable pour cet utilisateur de disposer d'un outil qui lui propose des lieux d'activités (tels que des restaurants, des musées, des monuments) adaptés à la fois à ses goûts personnels, à sa position géographique, au moment de la journée, à la météo, etc. Cette problématique constitue l'objet d'étude de la piste de recherche *Contextual Suggestion* (CS) de TREC¹, qui propose un cadre d'évaluation à la tâche de suggestion contextuelle (Dean-Hall *et al.*, 2013).

La majorité des systèmes de suggestion contextuelle se focalisent sur la modélisation des profils d'utilisateurs et la représentation des lieux d'activités afin d'obtenir un classement des lieux les plus pertinents pour un contexte et un utilisateur donnés. Cependant, la diversité des suggestions n'a jamais été le point de focalisation principal des approches précédentes. Pourtant, il a été suggéré que la diversification a un effet largement positif sur les systèmes de recommandation et de suggestion contextuelle de lieux d'activités (Ziegler *et al.*, 2005 ; Albakour *et al.*, 2014).

Notre contribution dans cet article est la création d'un modèle de suggestion contextuelle promouvant la diversité et inspiré de la recherche composite (Amer-Yahia *et al.*, 2014 ; Bota *et al.*, 2014) : l'approche que nous proposons consiste à regrouper les suggestions en différentes grappes thématiquement cohésives. L'ensemble des grappes couvre alors une grande diversité de thèmes. L'évaluation de notre système a été effectuée dans le cadre d'expérimentation de TREC CS 2013 et 2014. Elle montre que notre approche composite de suggestion contextuelle est compétitive et permet d'améliorer la diversité des suggestions sans dégrader leur pertinence.

La structure de l'article est la suivante. En section 2, nous présentons les travaux existants dans les domaines de la suggestion contextuelle et de la diversification en Recherche d'Information (RI). La section 3 détaille ensuite notre modèle composite de suggestion contextuelle en introduisant notre mesure de similarité thématique entre lieux d'activités, ainsi que le processus de création de grappes de lieux d'activités. Dans la section 4, nous décrivons l'évaluation réalisée dans le cadre de TREC CS 2013 et 2014, ainsi que les résultats obtenus. Finalement, nous concluons et évoquons les perspectives soulevées par cet article dans la section 5.

1. <https://sites.google.com/site/treccontext/>

2. État de l’art

Les travaux présentés dans cet article se rapportent au domaine des systèmes de recommandation, dont l’objectif est de renvoyer à l’utilisateur des résultats personnalisés, adaptés à ses préférences. Pour plus d’informations sur ce sujet, le lecteur pourra consulter Bobadilla *et al.* (2013). En particulier, nous nous intéressons à la suggestion contextuelle, consistant à proposer à un utilisateur une liste de lieux d’activités adaptés à ses préférences et à son contexte. La section 2.1 présente les travaux existants dans ce domaine. Dans la section 2.2, nous décrivons les contributions en RI et en systèmes de recommandation qui promeuvent la diversité des résultats.

2.1. Suggestion contextuelle à TREC

La piste TREC CS fournit aux participants un ensemble de contextes géographiques (représentés par des villes) et un ensemble de profils d’utilisateurs constitués de notes d’intérêt portant sur un échantillon de lieux d’activités prédéfinis. À partir de ces données, les participants suggèrent pour chaque paire utilisateur/contexte une liste de lieux d’activités adaptés au profil et situés dans le contexte.

Certaines approches consistent à former un index des éléments textuels associés à chaque lieu d’activités (titre, description, critiques, etc.) et représentent les lieux sous la forme de vecteurs dans un espace de termes (Rikitianskii *et al.*, 2013 ; Drosatos *et al.*, 2013 ; Yang et Fang, 2013). La représentation vectorielle des lieux d’activités et du profil permet ensuite d’ordonner les suggestions en fonction de leur pertinence en calculant une similarité entre suggestion candidate et profil (Rikitianskii *et al.*, 2013 ; Hubert *et al.*, 2013). La représentation des lieux d’activités est parfois également basée sur la (ou les) catégorie(s) thématique(s) (par exemple, « restaurant cantonais », « musée d’art ») associée(s) aux lieux (Hubert *et al.*, 2013 ; Roy *et al.*, 2013 ; Albakour *et al.*, 2014). De nombreux systèmes décomposent le profil de l’utilisateur en un profil positif (associé aux lieux que l’utilisateur a apprécié) et un profil négatif (associé aux lieux que l’utilisateur a trouvé inintéressants) (Yang et Fang, 2013 ; Rikitianskii *et al.*, 2013 ; Hubert *et al.*, 2013). À partir des profils positifs et négatifs de l’utilisateur, Yang et Fang (2013) s’inspirent du filtrage collaboratif en se basant sur l’opinion exprimée vis-à-vis des lieux d’activités et en considérant qu’un utilisateur qui possède des opinions semblables à un autre utilisateur aura tendance à apprécier les mêmes lieux.

2.2. Diversification

La diversification est un problème qui a été largement abordé en RI *ad-hoc*. Diversifier les résultats associés à une recherche de documents permet de couvrir les différentes interprétations de la requête (Clarke *et al.*, 2008). Par exemple, un utilisateur qui soumet la requête « jaguar » à un moteur de recherche peut chercher des informations sur l’animal ou la marque de voitures. Intégrer des documents traitant des

deux aspects à la liste des résultats permet alors d'augmenter les chances de répondre au besoin en information de l'utilisateur.

Le domaine des systèmes de recommandation a également mis en avant la préférence des utilisateurs pour les résultats diversifiés (Bradley et Smyth, 2001 ; Ziegler *et al.*, 2005) et a ainsi proposé de nombreux systèmes prenant en compte cette dimension (Di Noia *et al.*, 2014 ; Vargas et Castells, 2013). Di Noia *et al.* (2014) constituent tout d'abord un premier classement de pertinence des éléments à recommander non-notés en estimant la note que leur accorderait l'utilisateur, puis ré-ordonnent les recommandations en utilisant une fonction de coût intégrant à la fois les notions de diversité et de préférence de l'utilisateur. Vargas et Castells (2013) modélisent chaque utilisateur par un ensemble de sous-profils représentant une partition de l'ensemble des centres d'intérêt de l'utilisateur. Chaque sous-profil permet de générer par filtrage collaboratif un ensemble de recommandations. Les recommandations issues de l'ensemble des sous-profils sont ensuite réunies dans une liste commune et ordonnées en fonction de leur score inspiré de xQuAD (Santos *et al.*, 2010), combinant pertinence des recommandations et diversité des sous-profils représentés dans la liste.

2.3. Bilan et contribution

Malgré son importance reconnue dans le domaine des systèmes de recommandation, la question de la diversification en suggestion contextuelle a été largement ignorée par les approches existantes. La raison principale est probablement le cadre actuel d'évaluation de TREC CS, qui ne tient pas compte de la diversité dans ses métriques (Dean-Hall *et al.*, 2013). Albakour *et al.* (2014) prennent en compte cet aspect en proposant un système de suggestion contextuelle promouvant la diversité thématique des lieux d'activités en se basant également sur la technique de diversification xQuAD.

Un des problèmes associés aux méthodes de diversification classique en recommandation est la détérioration de la personnalisation et de la pertinence des suggestions en cherchant à couvrir un plus large spectre de thématiques (Bradley et Smyth, 2001). Pour cette raison, nous utilisons dans notre approche une méthode de diversification originale, inspirée des récents travaux en recherche composite (Amer-Yahia *et al.*, 2014 ; Bota *et al.*, 2014). Ces approches consistent à remplacer la liste de résultats classique en RI *ad-hoc* par une liste de grappes (*bundles*) thématiquement cohésives, et où deux grappes différentes traitent de thèmes différents. Ce paradigme permet à la fois à l'utilisateur d'explorer un thème défini – en parcourant une grappe donnée – et de consulter des résultats diversifiés, grâce à la variété de thèmes couverts par l'ensemble des grappes. Appliquée à la suggestion contextuelle, cette approche composite se traduit par la suggestion de grappes de lieux d'activités adaptés au contexte de l'utilisateur, telles que chaque grappe regroupe des lieux thématiquement similaires et telles que deux grappes différentes contiennent respectivement des lieux associés à des thèmes différents. Notre approche composite de la suggestion contextuelle est décrite en détail dans la section suivante.

3. Modèle de suggestion contextuelle composite

Le problème à résoudre s'apparente à une recommandation contextuelle de lieux d'activités. Nous disposons d'un ensemble d'utilisateurs et de contextes géographiques ; le profil préférentiel de chaque utilisateur, représenté sous la forme d'un ensemble de notes de pertinence attribuées à un échantillon de lieux d'activités, est également connu. L'objectif est alors, à partir de ces données, de suggérer, pour chaque paire utilisateur/contexte, une liste de lieux d'activités qui respectent les préférences de l'utilisateur et la position géographique du contexte. Notre approche se focalise en particulier sur la suggestion de lieux d'activités couvrant des thèmes diversifiés. Nous prenons en compte la position géographique en pré-filtrant les lieux d'activités de manière à considérer comme suggestions potentielles uniquement les lieux localisés dans le contexte donné. Cette étape de pré-filtrage n'étant pas primordiale à la compréhension de notre approche composite, elle n'est pas détaillée dans cet article. La section 3.1 présente notre mesure de similarité thématique entre lieux d'activités. Dans la section 3.2, nous introduisons les critères de qualité utilisés lors de la formation des grappes, qui est décrite en section 3.3.

3.1. Distance thématique entre lieux d'activités

Nous avons basé notre distance thématique entre lieux d'activités sur une taxonomie de catégories thématiques (par exemple, « restaurant cantonais », « musée d'art ») organisée selon une structure d'arbre. Soit V l'ensemble des lieux d'activités (*venues*) constituant les suggestions potentielles. Chaque lieu d'activités peut être associé à un ensemble de catégories différentes. Nous définissons la distance thématique $\text{tdist} : V \times V \rightarrow \mathbb{R}^+$ entre $v \in V$ et $w \in V$ comme la longueur du plus court chemin dans l'arbre des catégories liant les deux plus proches catégories de v et w , respectivement. Formellement, en notant C_v et C_w l'ensemble des catégories thématiques de v et w , respectivement, et sp le plus court chemin (*shortest path*) entre deux catégories, tdist est définie par :

$$\text{tdist}(v, w) = \min_{\substack{c_v \in C_v \\ c_w \in C_w}} \text{sp}(c_v, c_w).$$

3.2. Critères de qualité des grappes de lieux d'activités

Afin de créer des grappes de lieux d'activités qui soient thématiquement cohésives et qui respectent les préférences de l'utilisateur, il est nécessaire de disposer de critères estimant dans quelle mesure un lieu d'activités est thématiquement similaire à un autre, et dans quelle mesure un lieu d'activités serait apprécié par l'utilisateur. La popularité d'un lieu d'activités constitue également un indicateur fondamental de l'appréciation générale de ce lieu. En effet, les propriétés dépendant uniquement du lieu d'activités se montrent souvent plus importantes que la similarité avec le profil de l'utilisateur pour la tâche de suggestion contextuelle (Deveaud *et al.*, 2014).

Soit $B \subseteq \mathcal{P}(V)$ l'ensemble des grappes (*bundles*) de lieux d'activités qu'il est possible de former. Soit U l'ensemble des utilisateurs dont les profils sont connus.

3.2.1. Popularité globale

La popularité globale (*overall popularity*) évalue la popularité d'un lieu d'activités. La popularité globale d'un lieu d'activités $v \in V$ est définie par :

$$\text{opop}(v) = \frac{\text{pop}(v)}{\max_{w \in V} \text{pop}(w)} \in [0; 1],$$

où les w désignent les lieux d'activités de V et $\text{pop} : V \rightarrow \mathbb{N}$ représente un indicateur de popularité d'un lieu (par exemple, nombre de « likes », note appréciative ou nombre de visiteurs). Par extension, la popularité globale d'une grappe $b \in B$ est définie par la moyenne de la popularité globale de ses constituants :

$$\text{opop}(b) = \frac{\sum_{v \in b} \text{opop}(v)}{|b|} \in [0; 1],$$

où les v désignent les lieux d'activités contenus dans la grappe b .

3.2.2. Similarité et cohésion thématiques

La similarité thématique (*topical similarity*) évalue dans quelle mesure deux lieux d'activités traitent de thèmes semblables. La similarité thématique entre deux lieux d'activités $v \in V$ et $w \in V$ est définie par :

$$\text{tsim}(v, w) = \frac{1}{1 + \text{tdist}(v, w)} \in]0; 1],$$

où $\text{tdist} : V \times V \rightarrow \mathbb{R}^+$ désigne la distance thématique définie dans la section 3.1.

La cohésion thématique (*topical cohesion*) évalue dans quelle mesure les lieux d'activités composant une grappe sont thématiquement similaires. La cohésion thématique d'une grappe $b \in B$ est définie par :

$$\text{tcoh}(b) = \frac{\sum_{v, w \in b} \text{tsim}(v, w)}{|b|^2} \in]0; 1],$$

où les v et w désignent les lieux d'activités contenus dans la grappe b .

3.2.3. Appréciation estimée

L'appréciation estimée (*estimated appreciation*) évalue l'appréciation d'un lieu d'activités par un utilisateur donné. L'estimation est basée sur le profil de l'utilisateur et est calculée à partir des notes données par cet utilisateur à un échantillon de lieux d'activités, pondérées par la similarité thématique entre le lieu d'activités à estimer et les lieux d'activités de l'échantillon. L'appréciation estimée d'un utilisateur $u \in U$ pour un lieu d'activités $v \in V$ est définie par :

$$\text{eapp}_u(v) = \frac{\sum_{w \in RV_u} \text{rating}_u(w) \cdot \text{tsim}(v, w)}{\sum_{w \in RV_u} \text{tsim}(v, w)} \in [0; 1],$$

où les w désignent les lieux d'activités de RV_u , ensemble des lieux d'activités de l'échantillon notés par u (*rated venues*), et $\text{rating}_u : RV_u \rightarrow [0; 1]$ associe à un lieu d'activités de RV_u la note donnée par l'utilisateur u à ce lieu, divisée par la note maximale possible. Par extension, l'appréciation estimée de u pour une grappe $b \in B$ est définie par la moyenne de l'appréciation estimée de ses constituants :

$$\text{eapp}_u(b) = \frac{\sum_{v \in b} \text{eapp}_u(v)}{|b|} \in [0; 1],$$

où les v désignent les lieux d'activités contenus dans la grappe b .

3.2.4. Score d'une grappe

Le score évalue la qualité des lieux d'activités contenus dans une grappe pour un utilisateur donné en utilisant les critères suivants : popularité globale, cohésion thématique et appréciation estimée. Le score d'une grappe $b \in B$ pour un utilisateur $u \in U$ est défini par :

$$\text{score}_u(b) = \text{opop}(b)^{C_{\text{opop}}} \cdot \text{tcoh}(b)^{C_{\text{tcoh}}} \cdot \text{eapp}_u(b)^{C_{\text{eapp}}} \in [0; 1],$$

où C_{opop} , C_{tcoh} et C_{eapp} sont des réels positifs modulant respectivement la popularité globale, la cohésion thématique et l'appréciation estimée dans la fonction de score.

3.3. Formation des grappes de lieux d'activités

La construction des grappes se déroule en deux étapes : tout d'abord, des grappes de lieux d'activités sont produites en grande quantité et formées par agrégation autour d'un lieu d'activités pivot en prenant en compte les critères de popularité globale (opop), cohésion thématique (tcoh) et appréciation estimée (eapp). Ensuite, les grappes sont classées en fonction de leur score, et les lieux d'activités pivots des meilleures grappes sont sélectionnés pour former les futures suggestions. La figure 1 fournit un exemple de grappes de lieux d'activités générées par notre système.

3.3.1. Production des grappes

La première étape de l'approche est inspirée de l'algorithme « BOBO » (*Bundles One-By-One*) introduit par Amer-Yahia *et al.* (2014). Nous avons adapté cet algorithme afin de prendre en compte les critères de qualité des grappes qui ont été définis dans la section 3.2. L'objectif de cet algorithme est de produire un nombre donné de grappes de qualité suffisante. Typiquement, le nombre de grappes générées dans cette étape est choisi 10 à 100 fois supérieur au nombre de suggestions destinées à être renvoyées à l'utilisateur. Le pseudo-code est donné dans l'algorithme 1. Tout d'abord, une liste ordonnée de lieux d'activités pivots est constituée. Cette liste classe par ordre décroissant de leur popularité globale (opop) les lieux d'activités d'une liste de suggestions potentielles V . Puis, tant que le nombre de grappes formées est inférieur au nombre requis c , l'algorithme itère sur la liste des pivots et construit une grappe autour

<p>Algorithme 1 : BOBO</p> <p>Input : un utilisateur u, une liste V de lieux d'activités, un nombre c de grappes souhaité, un nombre maximum β de lieux d'activités par grappe</p> <p>Output : un ensemble d'au plus c grappes</p> <p>$Cand \leftarrow \emptyset$ $Pivots \leftarrow \text{descending_sort}(V, \text{opop})$ while $Pivots \neq \emptyset$ and $Cand < c$ do $v \leftarrow Pivots[0]$ $b \leftarrow \text{pick_bundle}(u, v, V, \beta)$ $V \leftarrow V \setminus b$ $Pivots \leftarrow Pivots \setminus b$ $Cand \leftarrow Cand \cup \{b\}$ end return $Cand$</p>	<p>Grappe de pizzerias</p> <ol style="list-style-type: none"> 1. Waterstone Pizza 2. Arosto 3. Sbarro 4. Downtown Pizza
	<p>Grappe de lieux historiques/musées</p> <ol style="list-style-type: none"> 1. National Air and Space Museum 2. Old Post Office Pavilion 3. Petersen House 4. United States Capitol Visitors Center

Figure 1 – Exemple de grappes générées par notre système. À l'intérieur des grappes, les lieux d'activités sont ordonnés par pertinence décroissante. Les étiquettes des grappes ont été choisies manuellement.

<p>Algorithme 2 : pick_bundle</p> <p>Input : un utilisateur u, un lieu d'activités pivot v, une liste V de lieux d'activités, un nombre maximum β de lieux d'activités par grappe</p> <p>Output : une grappe b d'au plus β lieux d'activités, formée autour du lieu d'activités pivot v</p> <p>$b \leftarrow \{v\}$ $active \leftarrow V$ while $active \neq \emptyset$ and $b < \beta$ do $w \leftarrow \text{argmax}_{w \in active} \lambda \cdot \text{tsim}(v, w) + (1 - \lambda) \cdot \text{eapp}_u(w)$ $b \leftarrow b \cup \{w\}$ $active \leftarrow active \setminus \{w\}$ end return b</p>
--

de chaque pivot à l'aide de la procédure `pick_bundle`, qui est détaillée ci-dessous. À chaque itération, le pivot et les lieux d'activités constituant la grappe sont retirés de V afin d'éviter qu'ils apparaissent dans plusieurs grappes différentes.

La procédure `pick_bundle`, décrite dans l'algorithme 2, construit une grappe autour d'un lieu d'activités central (le pivot v). Cela est réalisé en cherchant dans la liste de suggestions potentielles V le lieu d'activités w qui maximise une combinaison linéaire de sa similarité thématique (`tsim`) au pivot v et de son appréciation estimée (`eapp`). Dans cette combinaison linéaire, $\lambda \in [0; 1]$ module l'impact de chacun des deux critères. Lorsque le lieu d'activités qui maximise cette fonction est trouvé, il est ajouté à la grappe et l'itération continue tant que la taille de la grappe est inférieure à la taille maximale β .

3.3.2. Sélection des grappes et choix des suggestions

Une fois qu'un nombre défini de grappes a été produit, les grappes sont ordonnées suivant leur score, défini dans la section 3.2.4. Ensuite, nous extrayons le lieu d'activités pivot de chacune des k grappes en tête de la liste ordonnée. Ce procédé fournit une liste ordonnée de k suggestions de lieux d'activités pour un utilisateur donné. La qualité des suggestions ainsi générées est évaluée dans la section suivante.

4. Évaluation

4.1. Cadre d'expérimentation

4.1.1. Lieux d'activités de Foursquare

Afin de disposer d'un ensemble de lieux d'activités constituant des suggestions potentielles, nous avons tiré parti du réseau social géolocalisé Foursquare², qui est doté d'une base de lieux d'activités associés à un ensemble de catégories thématiques de granularité fine³. Ces catégories, organisées en arbre, permettent d'obtenir une mesure de similarité qui estime la proximité thématique entre lieux d'activités, comme présenté dans la section 3.1. De plus, Foursquare fournit un ensemble d'indicateurs sociaux sur les lieux d'activités, comme le nombre de « likes », la note appréciative ou le nombre d'utilisateurs de Foursquare ayant fréquenté le lieu. Ces informations constituent de précieux indices de popularité – facteur essentiel en suggestion contextuelle (Deveaud *et al.*, 2014). En particulier, nous avons utilisé le nombre de « likes » pour estimer la valeur de la fonction $\text{pop} : V \rightarrow \mathbb{N}$, introduite dans la section 3.2.1.

4.1.2. TREC CS 2013

Notre cadre d'expérimentation est basé sur la piste CS de la compétition TREC 2013. Dans la piste CS 2013, chaque participant à la compétition dispose des données suivantes : un ensemble de 50 contextes géographiques (villes américaines), un échantillon de 50 lieux d'activités et un ensemble de 562 profils d'utilisateurs. Les lieux de l'échantillon sont associés à une URL vers un site Web décrivant le lieu. Pour chaque lieu de l'échantillon, l'utilisateur a donné une note entre 0 et 4 (0 signifiant un désintérêt total, 2 l'indifférence et 4 un intérêt profond) indiquant son intérêt pour le lieu après visite du site Web. L'ensemble des notes données par un utilisateur constitue ainsi son profil. Ces notes de pertinence ont été utilisées pour estimer la valeur des fonctions $\text{rating}_u : RV_u \rightarrow [0; 1]$, introduites dans la section 3.2.3.

L'objectif des participants à la piste CS est de soumettre, pour chaque paire utilisateur/contexte, une liste de cinq suggestions de lieux d'activités, adaptées à la fois au profil de l'utilisateur et au contexte géographique. Chaque suggestion comprend une URL associée au lieu d'activités. Lors du processus d'évaluation des participants, les

2. <https://foursquare.com/>

3. <https://developer.foursquare.com/categorytree>

utilisateurs dont le profil était fourni aux participants notent la pertinence des suggestions qui leur sont adressées. Pour chaque suggestion le concernant, l'utilisateur donne une note d'intérêt entre 0 et 4 indiquant son attrait pour le lieu après visite du site Web. La pertinence géographique des suggestions est évaluée à la fois par les utilisateurs et les évaluateurs du NIST,⁴ et représentée par une note entre 0 et 2. Une suggestion est considérée comme pertinente si sa note d'intérêt est supérieure ou égale à 3, et si sa position géographique a obtenu une note supérieure ou égale à 1.

4.1.3. *Adaptation de la collection de TREC CS 2013*

Les sites Web des lieux d'activités suggérés par les participants à TREC CS 2013 proviennent généralement de l'*Open Web* mais ne sont pas toujours issus du réseau social géolocalisé que nous utilisons (Foursquare). Or, comme il l'a été évoqué précédemment, dans notre cadre d'expérimentation, l'estimation de la popularité et de la similarité thématique se basent sur des informations issues de Foursquare – le nombre de « *likes* » et l'arbre de catégories Foursquare. La collection de jugements de pertinence de TREC CS 2013, portant sur des lieux d'activités qui ne sont pas nécessairement issus de Foursquare, n'est donc pas utilisable dans l'état pour évaluer notre approche. Pour pallier ce problème, nous avons associé les lieux d'activités suggérés par les candidats à leur lieu d'activités Foursquare, lorsque ceux-ci existent. Ce procédé nous permet alors de disposer d'une collection d'évaluation contenant des jugements de pertinence portant sur un ensemble de lieux d'activités Foursquare. Cette adaptation de la collection de TREC CS 2013 présente cependant l'inconvénient d'empêcher la comparaison des performances de notre système avec celui des participants réels à TREC CS 2013. Nous avons donc également participé à TREC CS 2014 pour confronter notre approche aux meilleurs systèmes existants.

4.2. *Métriques d'évaluation*

Nous utilisons lors de notre évaluation les trois mesures officielles de TREC CS : la précision à 5 (P@5), le rang inverse moyen (*Mean Reciprocal Rank* – MRR) et le gain avec biais temporel (*Time-Biased Gain* – TBG) (Dean-Hall *et al.*, 2013).

De plus, dans le but d'évaluer la diversité des suggestions, nous adaptons la métrique de similarité intraliste introduite par Ziegler *et al.* (2005). Pour une liste de lieux d'activités $l \subseteq V$, la diversité intraliste (*Intralist Diversity* – ILD) est définie par :

$$\text{ILD}(l) = \frac{\sum_{v,w \in l} 1 - \text{tsim}(v,w)}{|l|^2},$$

4. Le NIST (*National Institute of Standard and Technology* – <http://www.nist.gov/>) est l'organisme en charge de l'organisation de TREC.

où $\text{tsim} : V \times V \rightarrow [0; 1]$ est la similarité thématique introduite dans la section 3.2.2. Par extension, pour un ensemble de listes de lieux d'activités $L \subseteq \mathcal{P}(V)$, la diversité intraliste moyenne (*Mean Intralist Diversity – MILD*) est définie par :

$$\text{MILD}(L) = \frac{\sum_{l \in L} \text{ILD}(l)}{|L|}.$$

4.3. Expérimentations

4.3.1. Classement multicritère

Nous avons été amenés à comparer plusieurs versions de notre système. Chaque version est définie par un jeu de paramètres donné. Il était alors nécessaire de classer ces versions, pour savoir laquelle était la plus efficace. Cependant, nous mesurons les performances de notre système à l'aide de trois métriques (P@5, MRR et MILD), ce qui pose problème pour constituer un classement unique des versions.

Pour pallier cette difficulté, nous nous sommes inspirés de Montague et Aslam (2002), qui décrivent un algorithme nommé *Borda Count* permettant de fusionner en un classement unique des éléments issus de plusieurs classements différents en fonction du rang de ces éléments dans chaque classement. Par exemple, supposons que l'on souhaite fusionner en un classement multicritère unique un ensemble de N versions de notre système, déjà classées selon un ensemble de critères représentés par les métriques P@5, MRR et MILD. Pour chaque classement associé à P@5, MRR et MILD, les versions se voient attribuées un nombre de points dépendant de leur rang : la première version dans le classement obtient N points, la seconde obtient $N - 1$, et ainsi de suite. Le classement multicritère est ensuite obtenu par le total de points des versions : la version avec le plus grand nombre de points est positionnée en tête du classement multicritère, et ainsi de suite.

4.3.2. Optimisation des paramètres λ , C_{opop} , C_{tcoh} , C_{eapp} et β

Nous avons cherché à trouver une combinaison des paramètres λ , C_{opop} , C_{tcoh} , C_{eapp} et β améliorant les performances de notre approche en terme de P@5, MRR et MILD. λ module l'importance de la similarité thématique et de l'appréciation estimée lors de l'agrégation des grappes autour des lieux d'activités pivots dans l'algorithme 2. C_{opop} , C_{tcoh} et C_{eapp} définissent respectivement l'importance de la popularité générale (opop), la cohésion thématique (tcoh) et l'appréciation estimée (eapp) dans le calcul du score des grappes. β correspond au nombre maximum de lieux contenus dans chaque grappe. Les résultats de l'expérience, réalisée sur la collection adaptée de TREC CS 2013 décrite en section 4.1.3, sont indiqués dans le tableau 1.

Pour limiter le temps de recherche des paramètres optimaux, et, par conséquent, le nombre de configurations à tester, nous avons optimisé chaque paramètre individuellement, à partir du jeu de paramètres de référence $\lambda = 0,5$, $C_{opop} = 1$, $C_{tcoh} = 1$, $C_{eapp} = 1$ et $\beta = 5$. Par exemple, nous avons fait varier le paramètre λ indépendamment des autres paramètres afin de savoir quelle valeur donnait les meilleures

performances. Pour comparer les versions du système associées aux différentes valeurs de λ , nous avons ensuite utilisé la méthode de classement multicritère présentée précédemment. Dans les colonnes P@5, MRR et MILD du tableau, le nombre en exposant correspond au nombre de points attribués en fonction du classement suivant les métriques P@5, MRR et MILD, respectivement. Les nombres en gras correspondent à la meilleure performance suivant le critère associé à leur colonne, pour un paramètre donné. La colonne Points contient la somme des points obtenus par les différentes versions pour chaque métrique. La version du système possédant la somme de points la plus élevée est alors considérée comme la version « gagnante », c'est-à-dire la version qui a la meilleure combinaison de P@5, MRR et MILD. Pour l'optimisation du paramètre λ , les différentes versions comparées sont telles que $\lambda \in \{0; 1/3; 0,5; 2/3; 1\}$. Les autres paramètres C_{opop} , C_{tcoh} , C_{eapp} et β sont fixés aux valeurs de référence : $C_{opop} = 1$, $C_{tcoh} = 1$, $C_{eapp} = 1$ et $\beta = 5$. La version gagnante est la version telle que $\lambda = 2/3$, car le total de points pour cette version est de 12, supérieur aux totaux de toutes les autres versions faisant varier λ (11 pour $\lambda = 0,5$; 10 pour $\lambda = 1/3$; 8 pour $\lambda = 1$; 4 pour $\lambda = 0$). En appliquant cette méthodologie à l'ensemble des paramètres, nous avons finalement obtenu le jeu de paramètres suivant : $\lambda = 2/3$, $C_{opop} = 5$, $C_{tcoh} = 1$, $C_{eapp} = 10$ et $\beta = 7$. Ses performances sont indiquées dans la dernière ligne du tableau 1. Les valeurs optimales obtenues pour ces paramètres suggèrent un ordre d'importance sur les critères à amplifier lors de l'estimation de la qualité individuelle des grappes réalisée en 3.3.2. Ordonnés du plus important au moins important, ces critères sont : l'appréciation estimée, la popularité globale et la cohésion thématique.

4.3.3. Impact de la diversité, personnalisation et popularité sur les suggestions

Nous avons également testé l'impact de la diversité, de la personnalisation et de la popularité sur les suggestions. Pour ce faire, nous avons comparé plusieurs versions de notre système, correspondant à toutes les combinaisons possibles des facteurs de diversité (influencée par λ , C_{tcoh} et β), de personnalisation (influencée par λ et C_{eapp}) et de popularité (influencée par C_{opop}). Les versions comparées ont été dérivées du jeu de paramètres optimal, établi en section 4.3.2. Le nom de chaque version indique la prise en compte des différents facteurs lors de la constitution de la liste de suggestions. Par exemple, la version *div + perso* prend en compte uniquement la diversité et la personnalisation, en ignorant la popularité. Ces versions ont été évaluées en terme de P@5, MRR et MILD. Les résultats de cette expérience, réalisée sur la collection adaptée de TREC CS 2013 décrite en section 4.1.3, sont indiqués dans le tableau 2.

Afin d'analyser les effets de la diversité, de la personnalisation et de la popularité, nous avons comparé les différentes versions en utilisant également la méthode de classement multicritère. Comme précédemment, le nombre en exposant correspond au nombre de points attribués en fonction de son classement. Les nombres en gras correspondent à la meilleure performance suivant le critère associé à leur colonne. La colonne Points contient la somme des points obtenus par les différentes versions pour chaque métrique. Nous pouvons constater que la version qui obtient la meilleure combinaison de P@5, MRR et MILD (17) est la version *div + perso + pop*. Cela

λ	C_{opop}	C_{tcoh}	C_{eapp}	β	P@5	MRR	MILD	Points
0,5	1	1	1	5	0,3139(2)	0,5463(4)	0,5384 (5)	11
0	1	1	1	5	0,2906(1)	0,4944(1)	0,5347(2)	4
1	1	1	1	5	0,3229 (5)	0,5422(2)	0,5271(1)	8
2/3	1	1	1	5	0,3139(3)	0,5466 (5)	0,5375(4)	12
1/3	1	1	1	5	0,3193(4)	0,5451(3)	0,5361(3)	10
0,5	1	1	1	5	0,3139 (3)	0,5463(1)	0,5384(1)	5
0,5	5	1	1	5	0,3121(2)	0,5535 (3)	0,5458 (3)	8
0,5	10	1	1	5	0,3103(1)	0,5510(2)	0,5455(2)	5
0,5	1	1	1	5	0,3139 (3)	0,5463 (3)	0,5384 (3)	9
0,5	1	5	1	5	0,2924(2)	0,5436(2)	0,5169(1)	5
0,5	1	10	1	5	0,2709(1)	0,5298(1)	0,5208(2)	4
0,5	1	1	1	5	0,3139(1)	0,5463(1)	0,5384(1)	3
0,5	1	1	5	5	0,3184(2)	0,5676 (3)	0,5391(2)	7
0,5	1	1	10	5	0,3247 (3)	0,5612(2)	0,5404 (3)	8
0,5	1	1	1	5	0,3139 (8)	0,5463(3)	0,5384(3)	14
0,5	1	1	1	1	0,3103(6)	0,5114(1)	0,5346(2)	9
0,5	1	1	1	3	0,3130(7)	0,5558(6)	0,5348(1)	14
0,5	1	1	1	7	0,3067(5)	0,5617 (8)	0,5502(4)	17
0,5	1	1	1	8	0,3067(4)	0,5564(7)	0,5578(5)	16
0,5	1	1	1	9	0,3067(3)	0,5535(5)	0,5588(6)	14
0,5	1	1	1	10	0,2933(2)	0,5519(4)	0,5644(7)	13
0,5	1	1	1	12	0,2924(1)	0,5463(2)	0,5725 (8)	11
2/3	5	1	10	7	0,3139	0,5635	0,5602	-

Tableau 1 – Résultats de l'expérience d'optimisation des paramètres de notre système. Chaque paramètre a été optimisé individuellement à l'aide d'un classement multicritère. Les exposants correspondent aux points attribués aux versions en fonction de leur classement. Les meilleures performances pour chaque métrique sont indiquées en gras. La dernière ligne du tableau montre les résultats obtenus pour la version combinant les meilleures valeurs de chaque paramètre.

Version	λ	C_{opop}	C_{tcoh}	C_{eapp}	β	P@5	MRR	MILD	Points
<i>div + perso + pop</i>	2/3	5	1	10	7	0,3139(5)	0,5635 (7)	0,5602(5)	17
<i>div + perso</i>	2/3	0	1	10	7	0,1336(2)	0,2741(2)	0,5780(6)	10
<i>div + pop</i>	1	5	1	0	7	0,3148(6)	0,5466(6)	0,5549(4)	16
<i>perso + pop</i>	-	5	-	10	1	0,3157 (7)	0,5152(5)	0,5341(2)	14
<i>div</i>	1	0	1	0	7	0,1309(1)	0,2428(1)	0,5960 (7)	9
<i>perso</i>	-	0	-	10	1	0,1605(3)	0,3074(3)	0,4720(1)	7
<i>pop</i>	-	5	-	0	1	0,3094(4)	0,5113(4)	0,5350(3)	11

Tableau 2 – Résultats de l'expérience comparant les effets de la diversité, de la personnalisation et de la popularité sur les suggestions. La comparaison est réalisée par classement multicritère. Les exposants correspondent aux points attribués aux versions en fonction de leur classement. Les meilleures performances pour chaque métrique sont indiquées en gras. Les valeurs indiquées par un tiret signifie que, dans cette version, le paramètre en question n'a pas d'impact.

Système	λ	C_{opop}	C_{tcoh}	C_{eapp}	β	P@5 (rang)	TBG (rang)	MRR (rang)
<i>UDInfoCS2014_2</i>	-	-	-	-	-	0,5585 (1)	2,7021 (1)	0,7482 (1)
<i>RAMARUN2</i>	-	-	-	-	-	0,5017 (2)	2,3718 (2)	0,6846 (2)
<i>BJUTa</i>	-	-	-	-	-	0,5010 (3)	2,2209 (4)	0,6677 (4)
<i>BJUTb</i>	-	-	-	-	-	0,4983 (4)	2,1949 (5)	0,6626 (6)
<i>uogTrBunSumF</i>	2/3	1	1	10	5	0,4943 (5)	2,1526 (7)	0,6704 (3)

Tableau 3 – Résultats des meilleures participations à TREC CS 2014. Notre système *uogTrBunSumF* a respectivement obtenu les rangs de 5^{ème}, 7^{ème} et 3^{ème} sur 25 en P@5, TBG et MRR.

montre que, comme nous l'avions supposé, la combinaison de la diversité, de la personnalisation et de la popularité est bénéfique à la qualité globale des suggestions. Sur l'ensemble des versions, on peut noter la grande influence de la popularité sur la précision, ce qui se traduit par un bon classement des versions *div + pop*, *perso + pop* et *pop*. Il est intéressant de remarquer que la popularité semble être un facteur plus important que la personnalisation des suggestions : *div + pop* (16) a obtenu un meilleur score que *div + perso* (10). Ces résultats sont en accord avec Deveaud *et al.* (2014), qui soulignent l'importance de la popularité des suggestions, et la prévalence de ce facteur sur la personnalisation. Cette expérience montre également que notre système est capable de promouvoir la diversité des suggestions, sans dégrader la précision. En effet, si nous comparons notre meilleure version (*div + perso + pop*) à la version qui a la meilleure précision (*perso + pop*), on constate une amélioration en MILD de 6% et une amélioration en MRR de 9% contre une perte en P@5 de seulement 2%.

Il est cependant important de noter que la collection de jugements de pertinence utilisée lors des expériences précédentes est incomplète. En effet, les utilisateurs n'ont pas exprimé leur appréciation vis-à-vis de l'intégralité des lieux d'activités extraits de Foursquare et utilisés comme suggestions candidates. Un lieu d'activités qui n'a été suggéré par aucun participant peut malgré tout être pertinent. Or l'évaluation présentée ici considère qu'un lieu d'activités qui n'est pas associé à un jugement de pertinence est nécessairement non pertinent. Nous sommes donc conscients des limites de la collection adaptée de TREC CS 2013 que nous utilisons.

4.3.4. Participation à TREC CS 2014

Afin d'évaluer la compétitivité de notre système de suggestion contextuelle et le confronter à des utilisateurs réels, nous avons participé à TREC CS 2014. Cette édition de TREC CS compte un ensemble de 50 contextes géographiques et 299 profils d'utilisateurs, composés de notes portant sur un échantillon de 100 lieux d'activités. Les 12 participants ayant utilisé l'*Open Web* pour constituer leur base de suggestions candidates ont soumis un total de 25 systèmes. Les lieux d'activités suggérés par notre système⁵ ont été extraits de Foursquare. Le jeu de paramètres utilisé par notre système est le suivant : $\lambda = 2/3$, $C_{opop} = 1$, $C_{tcoh} = 1$, $C_{eapp} = 10$ et $\beta = 5$. Ce jeu de paramètres diffère légèrement du jeu optimal déterminé en section 4.3.2, car la soumission à TREC CS 2014 a été effectuée antérieurement à l'optimisation rigoureuse des paramètres. Les systèmes ont été évalués sur un ensemble de 299 paires utilisateur/contexte. Les résultats de l'évaluation pour les 5 meilleurs systèmes sont indiqués dans le tableau 3. Notre système, nommé *uogTrBunSumF*, a respectivement obtenu les rangs de 5^{ème}, 7^{ème} et 3^{ème} sur 25 en terme de P@5, TBG et MRR. Il est à noter que les différences en P@5 entre les systèmes classés de la 2^{ème} à la 5^{ème} position sont extrêmement faibles. Cela suggère que notre système est capable de renvoyer des suggestions qui sont non seulement diversifiées, mais également de pertinence comparable à celle atteinte par les meilleurs systèmes de suggestion contextuelle existants.

5. Le code source de notre système de suggestion contextuelle est disponible à l'adresse suivante : <https://github.com/tthonet/composite-contextual-suggestion>.

5. Conclusion et perspectives

Afin de réaliser un système de suggestion contextuelle promouvant la diversité, nous nous sommes inspirés de la recherche composite en formant des grappes de lieux d'activités. Chaque grappe individuelle est thématiquement cohésive et l'ensemble des grappes est thématiquement diversifié. Les lieux d'activités constituant les suggestions potentielles sont associés à un ensemble de catégories organisées en une hiérarchie arborescente qui permet de définir une mesure de similarité thématique entre lieux d'activités. Le système construit et ordonne les grappes en fonction de critères de popularité globale, de cohésion thématique et d'appréciation estimée. L'évaluation du système réalisée dans le cadre de TREC CS 2013 et 2014 montre que notre approche composite mène à des suggestions diversifiées en conservant une pertinence comparable aux meilleurs systèmes de suggestion contextuelle.

Dans des travaux futurs, il serait intéressant de réaliser une étude auprès d'utilisateurs réels en situation de mobilité pour analyser leur préférence entre un système de suggestion contextuelle classique, fournissant une simple liste de lieux d'activités, et un système de suggestion contextuelle composite, qui renvoie une liste de grappes de lieux d'activités. Cette expérience permettrait également d'étudier la dimension exploratoire apportée par l'approche composite et son impact sur les utilisateurs.

Remerciements

Nous tenons à remercier Guillaume Cabanac et Karen Pinel-Sauvagnat pour leurs relectures et les conseils prodigués lors de l'écriture de cet article.

6. Bibliographie

- Albakour M.-D., Deveaud R., Macdonald C., Ounis I., « Diversifying Contextual Suggestions from Location-based Social Networks », *Proceedings of the 5th Information Interaction in Context Symposium, IIX '14*, ACM Press, New York, New York, USA, p. 125-134, 2014.
- Allan J., Croft B., Moffat A., Sanderson M., « Frontiers, Challenges, and Opportunities for Information Retrieval : Report from SWIRL 2012 the Second Strategic Workshop on Information Retrieval in Lorne », *SIGIR Forum*, vol. 46, n° 1, p. 2-32, 2012.
- Amer-Yahia S., Bonchi F., Castillo C., Feuerstein E., Mendez-Diaz I., Zabala P., « Composite Retrieval of Diverse and Complementary Bundles », *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, n° 11, p. 2662-2675, 2014.
- Bobadilla J., Ortega F., Hernando A., Gutiérrez A., « Recommender Systems Survey », *Knowledge-Based Systems*, vol. 46, p. 109-132, 2013.
- Bota H., Zhou K., Jose J. M., Lalmas M., « Composite Retrieval of Heterogeneous Web Search », *Proceedings of the 23rd International Conference on World Wide Web, WWW '14*, ACM Press, New York, New York, USA, p. 119-130, 2014.
- Bradley K., Smyth B., « Improving Recommendation Diversity », *Proceedings of the 12th Irish Conference in Artificial Intelligence and Cognitive Science, AICS '01*, p. 75-84, 2001.

- Clarke C. L. A., Kolla M., Cormack G. V., Vechtomova O., Ashkan A., Büttcher S., MacKinnon I., « Novelty and Diversity in Information Retrieval Evaluation », *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, ACM Press, New York, New York, USA, p. 659-666, 2008.
- Dean-Hall A., Thomas P., Clarke C. L. A., Simone N., Kamps J., Voorhees E., « Overview of the TREC 2013 Contextual Suggestion Track », *Proceedings of the 22nd Text Retrieval Conference*, TREC '13, 2013.
- Deveaud R., Albakour M.-D., Macdonald C., Ounis I., « On the Importance of Venue-Dependent Features for Learning to Rank Contextual Suggestions », *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, CIKM '14, ACM Press, New York, New York, USA, 2014.
- Di Noia T., Ostuni V. C., Rosati J., Tomeo P., Di Sciascio E., « An Analysis of Users' Propensity toward Diversity in Recommendations », *Proceedings of the 8th ACM Conference on Recommender Systems*, RecSys '14, ACM Press, New York, New York, USA, p. 285-288, 2014.
- Drosatos G., Stamatelatos G., Arampatzis A., Efraimidis P. S., « DUTH at TREC 2013 Contextual Suggestion Track », *Proceedings of the 22nd Text Retrieval Conference*, TREC '13, 2013.
- Hubert G., Cabanac G., Pinel-Sauvagnat K., Palacio D., Sallaberry C., « IRIT, GeoComp, and LIUPPA at the TREC 2013 Contextual Suggestion Track », *Proceedings of the 22nd Text Retrieval Conference*, TREC '13, 2013.
- Montague M., Aslam J. A., « Condorcet Fusion for Improved Retrieval », *Proceedings of the 11th International Conference on Information and Knowledge Management*, CIKM '02, ACM Press, New York, New York, USA, p. 538-548, 2002.
- Rikitianskii A., Harvey M., Crestani F., « University of Lugano at the TREC 2013 Contextual Suggestion Track », *Proceedings of the 22nd Text Retrieval Conference*, TREC '13, 2013.
- Roy D., Bandyopadhyay A., Mitra M., « A Simple Context Dependent Suggestion System », *Proceedings of the 22nd Text Retrieval Conference*, TREC '13, 2013.
- Santos R. L., Macdonald C., Ounis I., « Exploiting query reformulations for web search result diversification », *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, ACM Press, New York, New York, USA, p. 881-890, 2010.
- Vargas S., Castells P., « Exploiting the Diversity of User Preferences for Recommendation », *Proceedings of the 10th Conference on Open Research Areas in Information Retrieval*, OAIR '13, p. 129-136, 2013.
- Yang P., Fang H., « Opinion-based User Profile Modeling for Contextual Suggestions », *Proceedings of the 2013 International Conference on the Theory of Information Retrieval*, ICTIR '13, ACM Press, New York, New York, USA, p. 80-83, 2013.
- Ziegler C.-N., McNee S. M., Konstan J. A., Lausen G., « Improving recommendation lists through topic diversification », *Proceedings of the 14th International Conference on World Wide Web*, WWW '05, ACM Press, New York, New York, USA, p. 22-32, 2005.