La Classification d'Opinion comme préambule à la Recommandation Automatique de Contenus

Damien Poirier

La Classification d'Opinion comme préambule à la Recommandation Automatique de Contenus

Damien Poirier

Orange Labs, 2 avenue Pierre Marzin, 22300 Lannion, FRANCE
LIFO - Université d'Orléans, rue Léonard de Vinci, 45000 Orléans, FRANCE
Lip6 - Université P. et M. Curie, 104 av. du président Kennedy, 75016 Paris, FRANCE
damien.poirier@orange-ftgroup.com

RÉSUMÉ. *Les systèmes de recommandation automatique sont devenus, à l'instar des moteurs de recherche, un outil incontournable pour tout site contenant un catalogue riche, que ce soit en termes de produits, de musiques, de news ou encore simplement de pages web. Le bon fonctionnement de ces systèmes repose sur une grande quantité d'information. Parallèlement à cela, les données présentes sur Internet ne cessent de s'enrichir depuis l'apparition du Web 2.0, notamment grâce au contenu généré par les utilisateurs (User Generated Content). L'étude de ces données peut être un gain pour la recommandation automatique. Nous proposons d'exploiter des textes provenant d'un site communautaire, en les classant selon l'opinion qu'ils expriment, afin d'alimenter un moteur de recommandation basé sur une technique de filtrage collaboratif. Nous présentons toute la chaîne de traitement ainsi que son évaluation.*

ABSTRACT. *Recommender systems have become, like search engines, a tool that cannot be ignored by a website with a large selection of products, music, news or simply webpages. The performance of this kind of system depends on a large amount of information. At the same time, the amount of information on the Web is continuously growing, especially due to increased User Generated Content since the explosion of Web 2.0. The study of this data could bring a significant contribution to the field of recommendation. In this paper, we propose a method providing recommendations from only textual data. We propose to label texts according to their expressed opinion in order to supply a recommender system based on a collaborative filtering technique. We describe the entire processing chain and its evaluation.*

MOTS-CLÉS : *Classification d'opinion, Recommandation Automatique, Filtrage Collaboratif.*

KEYWORDS: *Opinion Mining, Automatic Recommendation, Collaborative Filtering.*

1. Introduction

L'augmentation des données présentes sur le Net oblige les outils de navigation à s'adapter. Outre les moteurs de recherche, les systèmes de recommandation deviennent maintenant incontournables sur certains grands sites internet. L'objectif de ces systèmes est de sélectionner, parmi un catalogue, les articles (vidéos, musiques, livres, pages web, etc..) les plus susceptibles d'intéresser un utilisateur en particulier. Le filtrage collaboratif est une technique de recommandation automatique qui cherche à prédire l'appréciation d'un utilisateur sur un item à l'aide d'informations sur les goûts d'autres utilisateurs. Ces informations sont très souvent des notes et il en faut un nombre très important pour obtenir de bonnes recommandations. Le manque d'information est souvent le principal frein aux systèmes de recommandation, notamment lors du lancement d'un nouveau service (Schein *et al.*, 2002).
Les données textuelles, qui sont majoritaires sur le Web, sont encore très peu utilisées en recommandation. Le contenu généré par les utilisateurs contient énormément de commentaires subjectifs d'opinions sur les contenus de nombreux catalogues de toutes sortes. La fouille de données d'opinion (Pang *et al.*, 2008), est un domaine qui s'est beaucoup développé ces dernières années dans le but d'extraire des informations de tous ces textes d'utilisateurs.
Nous avons donc d'un côté des systèmes de navigation qui peuvent manquer d'information, et d'un autre côté, des textes remplis d'information non encore exploitée. Dans cet article, nous proposons une méthode permettant d'adapter les données textuelles aux techniques de filtrage collaboratif. Pour ce faire, nous proposons de construire des données d'usages en inférant des notes à partir de commentaires utilisateurs à l'aide de la classification d'opinion. Ces données d'usages sont ensuite exploitables par un moteur de type filtrage collaboratif pour la recommandation.
Dans une première partie nous décrivons les données utilisées pour les expérimentations. Puis nous décrivons et évaluons la méthode utilisée pour attribuer une polarité d'opinion à un texte. Enfin, nous présentons la méthode utilisée pour établir des recommandations à partir de données d'usages et nous présentons différents résultats.

2. Description des données

La chaîne de traitement de cette méthode (voir figure 1) nécessite l'utilisation de trois corpus différents. Les deux premiers corpus sont des corpus textuels et le troisième est un corpus de notes :

– Le corpus de notes, que nous appellerons *Corpus 3*, est une sélection des données mises à disposition par le site de location de DVD Netflix[1] lors du lancement de son challenge qui avait pour objectif d'améliorer les résultats de ses recommandations. Ces données de test, composées de 400 000 triplets utilisateur-film-note sont très utilisées dans le domaine de la recommandation automatique.

1. www.netflix.com

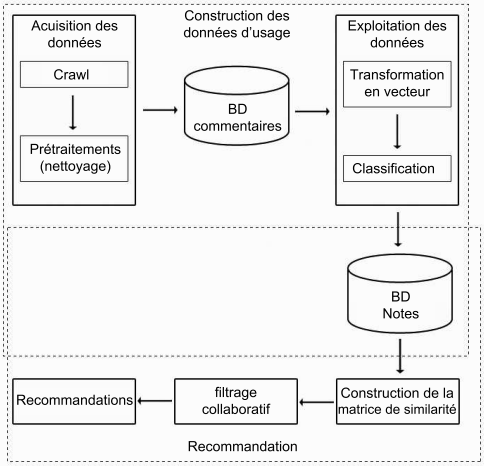

Figure 1. *Chaîne de traitement*

– Le *Corpus 2* a été construit spécialement pour ces expérimentations à partir de données extraites du site communautaire Flixster[2]. Ce site, de type Web 2.0, est un réseau social offrant aux amateurs de films l'opportunité de se créer un espace personnel afin de partager leurs goûts et opinions cinématographiques avec d'autres amateurs. La plupart des commentaires présents sur Flixster sont associés à un utilisateur, à un film, ainsi qu'à une note laissée par l'auteur résumant l'opinion exprimée dans le commentaire. Le *Corpus 2* est composé de quadruplets utilisateur-film-note-commentaire et est dédié au test pour la tâche de classification d'opinion. Il est composé d'environ 3 300 000 quadruplets dont tous les films, au nombre de 10 000, sont présents dans le *Corpus 3*. Le nombre d'utilisateurs concernés est d'environ 100 000.

– Le *Corpus 1* est également extrait du site Flixster et contient environ 175 000 quadruplets utilisateur-film-note-commentaire. Ce corpus est dédié à l'apprentissage pour la classification d'opinion. Il a été construit de façon à ce que les classes soient équilibrées et que tous les utilisateurs et les films présents soient différents de ceux présents dans le *Corpus 2*.

Les textes peuvent être assez variés, s'agissant de leur taille ou de leur style d'écriture. La taille moyenne des commentaires (environ 15 mots) est pratiquement équivalente pour les deux corpus textuels. Les seuls traitements appliqués aux textes sont un passage des lettres en minuscules et la suppression de tous les mots ayant moins de 3 occurrences dans le corpus d'apprentissage (*Corpus 1*). En ce qui concerne la distribution des corpus, il apparait très nettement que les notes positives sont plus présentent que les notes négatives, et ce quelque soit la provenance des données.

2. www.flixster.com

3. Première étape : Construction des données d'usages

Nous proposons d'utiliser les textes afin d'alimenter un système de recommandation basé sur du filtrage collaboratif, en adoptant la démarche expliquée par la figure 1. Cette première étape consiste donc à inférer des notes à partir de textes dans le but d'obtenir des triplets utilisateur-film-note en lieu et place des triplets utilisateur-film-commentaire. Cette tâche est exactement une tâche de classification d'opinion. Dans la littérature, on distingue deux grandes approches :

– La première est une approche linguistique consistant à construire des lexiques de mots d'opinions et à comptabiliser la présence ou l'absence de ces mots dans un texte afin de prédire l'opinion qu'il exprime (Yu *et al.*, 2003).

– La deuxième est une approche issue de l'apprentissage automatique qui consiste, dans un premier temps, à apprendre un modèle de classification à l'aide d'exemples déjà classés et, dans un deuxième temps, à classer de nouveaux textes à l'aide du modèle appris. La plupart des outils d'apprentissage automatique ont été testés sur cette tâche et les deux techniques les plus adaptées et les plus utilisées sont les Machine à Vecteurs de Support (SVM) et les classifieurs Naïfs Bayésiens (NB) (Kobayakawa *et al.*, 2009, Abbasi *et al.*, 2008, Pang *et al.*, 2002).

Après différents essais (Poirier *et al.*, 2009), nous avons choisi d'utiliser un classifieur naïf bayésien à sélection de variables pour cette tâche (Boullé, 2007) avec une représentation des textes sous forme de vecteur de fréquences des mots. L'apprentissage a été effectué sur le *Corpus 1* et le test sur le *Corpus 2*. Nous avons effectué une classification sur deux classes : les textes positifs et ceux négatifs. Le F-score obtenu en test, calculé sur la matrice de confusion (voir tableau 1), est de 0,71.

		Vraies Notes	
		HT	TTC
Notes	NEG	**79,08**	36,78
Prédites	POS	20,92	**63,22**

Tableau 1. *Matrice de confusion obtenue pour la classification d'opinion (les résultats sont des pourcentages)*

Avec les résultats de la classification d'opinion, nous avons construit une matrice d'usages exploitable par un moteur collaboratif.

4. Deuxième étape : Recommandation

La deuxième étape a pour but d'établir des recommandations à l'aide de la matrice d'usages obtenue précédemment. Les outils dédiés à cette tâche sont les moteurs de filtrage collaboratif. Leur principe est de remplir les cases vides dans la matrice d'usages en prédisant des notes. Pour ce faire, il existe deux grandes approches. l'approche basée sur les utilisateurs (Resnick *et al.*, 1994) consiste à comparer les

utilisateurs entre eux et à retrouver ceux ayant des goûts en communs, les notes d'un utilisateur étant ensuite prédites selon son voisinage. L'approche basée sur les items (Sarwar *et al.,* 2001) consiste à rapprocher les items appréciés par des personnes communes et à prédire les notes des utilisateurs en fonction des items les plus proches de ceux qu'il a déjà notés. La méthode utilisée ici est une approche basée sur les items avec une mesure de similarité qui est une distance de Jaccard pondérée par la distance de Pearson (Candillier *et al.,* 2008). La mesure choisie pour comparer les résultats est la Root Mean Squared Error (RMSE). La formule est la suivante :

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\bar{X}_i - X_i)^2}{n}}$$

avec X_i la note originale et $\bar{X}_i$ la note prédite.

Pour information, la RMSE est la mesure qui a été utilisée lors du challenge Netflix. L'objectif était d'obtenir un score inférieur à 0,85. Concernant le moteur utilisé ici, la meilleure RMSE obtenue est de 0,913. Celle-ci a été obtenue en apprenant sur un corpus riche provenant de Netflix, soit la même origine que le corpus de test. La première observation concerne donc le fait d'apprendre et de tester sur des corpus de provenances différentes (Flixster et Netflix dans ce cas). Les résultats obtenus en construisant la matrice de similarité entre items à l'aide des notes originales du *Corpus 2* donne une RMSE de 0.933. Il y a donc une perte d'information qui était prévisible sachant que les données Netflix sont beaucoup plus riches avec une moyenne de 200 notes par utilisateur alors qu'elle n'est que de 30 pour les données Flixster. Le deuxième test consiste à vérifier si le fait d'inférer des notes grâce à la classification d'opinion entraîne une perte d'information supplémentaire. Avec les notes obtenues à la suite de la tâche de classification d'opinion, la RMSE obtenue est de 0,936, soit une différence très faible. La RMSE obtenue avec le *Corpus 2*, en remplaçant les vraies notes par des notes aléatoires, est de 0,95. Les résultats des différentes expérimentations sont résumés dans le tableau 2.

	Données Netflix	Données Flixster		
	avec vraies notes	avec vraies notes	avec notes issues de la classification	avec notes aléatoires
RMSE	0,913	0,933	0,936	0,95

Tableau 2. *Résultats obtenus en recommandation selon les données d'apprentissage*

5. Conclusion

Nous venons de présenter une méthode permettant d'établir des recommandations en se basant uniquement sur des textes non structurés provenant de blogs n'ayant aucun lien avec le service recevant les recommandations, si ce n'est le sujet (les films). Les résultats observés sont encourageants. On observe que l'apport de données extérieures pour l'apprentissage du moteur de recommandation entraîne une perte d'information. Il faut tout de même préciser que le corpus Netflix est d'une richesse rare avec une moyenne de 200 films notés par utilisateur (sur les 400 000 utilisateurs répertoriés). L'autre information de ces expérimentations est qu'avec les notes provenant

de la classification d'opinion, la perte d'information est minime par rapport aux notes données explicitement par les utilisateurs (0,936 à 0,933). Ces résultats semblent donc valider l'idée d'alimenter un moteur de recommandation avec des textes extraits d'Internet (blogs, forums, etc.). Toutefois, de nouvelles méthodes de classification d'opinion doivent encore être évaluées sur cette chaîne de traitement, d'autres échelles de notes (3 classes, 5 classes, régression), d'autres représentations (TF-idf), etc. afin de vérifier leurs influences sur la recommandation.

6. Bibliographie

Abbasi A., Chen H., Salem A., « Sentiment analysis in multiple languages : Feature selection for opinion classification in Web forums », *ACM Trans. Inf. Syst.*, vol. 26, n° 3, p. 1-34, 2008.

Boullé M., « Compression-Based Averaging of Selective Naive Bayes Classifiers », *Journal of Machine Learning Research*, vol. 8, p. 1659-1685, 2007.

Candillier L., Meyer F., Fessant F., « Designing Specific Weighted Similarity Measures to Improve Collaborative Filtering Systems », *ICDM '08 : Proceedings of the 8th industrial conference on Advances in Data Mining*, Springer-Verlag, p. 242-255, 2008.

Kobayakawa T. S., Kumano T., Tanaka H., Okazaki N., Kim J.-D., Tsujii J., « Opinion classification with tree kernel SVM using linguistic modality analysis », *CIKM '09 : Proceeding of the 18th ACM conference on Information and knowledge management*, ACM, New York, NY, USA, p. 1791-1794, 2009.

Pang B., Lee L., « Opinion Mining and Sentiment Analysis », *Found. Trends Inf. Retr.*, vol. 2, n° 1-2, p. 1-135, 2008.

Pang B., Lee L., Vaithyanathan S., « Thumbs up ? : sentiment classification using machine learning techniques », *EMNLP '02 : Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, Association for Computational Linguistics, Morristown, NJ, USA, p. 79-86, 2002.

Poirier D., Fessant F., Bothorel C., Guimier de Neef E., Boullé M., « Approches Statistique et Linguistique Pour la Classification de Textes d'Opinion Portant sur les Films », *RNTI*p. 147-169, 2009.

Resnick P., Iacovou N., Suchak M., Bergstorm P., Riedl J., « GroupLens : An Open Architecture for Collaborative Filtering of Netnews », *Proceedings of ACM 1994 Conference on Computer Supported Cooperative Work*, ACM, p. 175-186, 1994.

Sarwar B., Karypis G., Konstan J., Reidl J., « Item-based collaborative filtering recommendation algorithms », *WWW '01 : Proceedings of the 10th international conference on World Wide Web*, ACM, New York, NY, USA, p. 285-295, 2001.

Schein A. I., Popescul A., H. L., Popescul R., Ungar L. H., Pennock D. M., « Methods and Metrics for Cold-Start Recommendations », *In Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM Press, p. 253-260, 2002.

Yu H., Hatzivassiloglou V., « Towards answering opinion questions : separating facts from opinions and identifying the polarity of opinion sentences », *Proceedings of the 2003 conference on Empirical methods in natural language processing*, Association for Computational Linguistics, Morristown, NJ, USA, p. 129-136, 2003.