

Sujet de stage inter-équipes CEDRIC

Detect and burst the bubbles in Recommender Systems

Détection et éclatement des bulles d'information dans les systèmes de recommandation

Cédric du Mouza (ISID), Nicolas Travers (Vertigo)

Mots-clés : Systèmes de recommandation, bulle filtrante, Big Data, scalability, détection de communautés

1 Contexte

Après l'élection de Donald Trump, un journaliste déclarait « il y a 58 millions d'électeurs pro-Trump et je n'en ai vu aucun »[1]. Cette phrase illustre un phénomène déjà décrit par Eli Pariser en 2011 [2], dont les journalistes s'étaient en parti emparés à la suite du Brexit, celui de la « **bulle filtrante** ».

Qu'il s'agisse de l'algorithme de recommandation de Facebook (EdgeRank [3]) qui décide du contenu à afficher aux utilisateurs, ou des méthodes de Deep Learning [4] de Youtube qui choisi quelles vidéos mettre en avant en fonction de l'historique du visiteur, notre expérience sur ces plateformes est totalement personnalisée. C'était l'idée de départ des systèmes de recommandation, un champ de recherche qui a plus de 20 ans, de maximiser la pertinence du contenu présenté à chaque utilisateur.

On trouve dans cette littérature bien avant le livre d'Eli Pariser un questionnement sur la possibilité de ces algorithmes de nous enfermer dans une zone de confort et d'ainsi transformer le « village global » en « tribus » [5][6]. Le sujet ne se limite cependant pas aux élections puisque le gouvernement français a décidé de lancer en 2016 une concertation avec ces mêmes plateformes pour limiter la recommandation de propagande djihadistes par ces mêmes algorithmes [7]. Le problème prend donc de l'ampleur, et il semble crucial de pouvoir efficacement mesurer si un utilisateur est pris dans une boucle afin d'essayer de lui proposer autre chose. La tendance naturelle à l'enfermement de ces systèmes a été étudiée et partiellement prouvée par un groupe de chercheurs [8] et des solutions [9] très spécifiques ont été envisagées et expérimentées [10][11].

Cependant ces recherches s'ancrent davantage dans la communauté du traitement de l'information et ce sujet reste encore très marginal au sein de la communauté des systèmes de recommandation et peu de modèles se focalisent sur ce point précis d'éclatement de bulle. On peut néanmoins citer un modèle qui mélange des affinités un peu aléatoirement pour ajouter de la diversité dans les recommandations [12]. Il devient ainsi pertinent de trouver une méthode pour détecter efficacement l'enfermement produit par les recommandations et d'y ajouter de la diversité tout en restant pertinent.

2 Sujet de stage

Ce stage s'appuie sur un large jeu de données déjà collecté issu de Twitter [13] comportant plus de **3 milliards de messages**, **2 182 867 utilisateurs** et **125 451 980 arcs**. Le but est de proposer une amélioration du système de recommandation RReco implémenté par *Quentin Grossetti* en thèse dans l'équipe. Basé sur ce contexte et ce jeu de données, le but du stage est triple :

1. Choisir une méthode qui permette de **détecter** efficacement des **communautés** telle que la méthode de Louvain par exemple [14];
2. Proposer des méthodes pour **calculer la distance** entre celles-ci (Topologique et Sémantique) et

également de quantifier l'**aspect communautaire d'une information** (i.e., bulle filtrante) ;

3. Faire passer à l'échelle ce calcul dans le système **RTReco** avec à terme un flux de 500 millions de messages.

La particularité de la méthode est de s'intégrer non pas dans un réseau d'utilisateur classique, mais sur un réseau de similarité entre utilisateur apportant une pertinence plus fortes entre les utilisateurs. Ce méta-graphe d'utilisateur permet non seulement de faciliter les recommandations, mais surtout de réduire considérablement le temps de calcul de celles-ci. Ainsi, ce stage devra exploiter cet aspect méta-graphe pour en extraire des communautés d'information et d'évoluer sous différents aspects ces *bulles filtrantes*.

Références

- [1] Julien Cadot. Bulles de filtrage : il y a 58 millions d'électeurs pro-trump et je n'en ai vu aucun, 2016.
- [2] Eli Pariser. *The Filter Bubble : What the Internet Is Hiding from You*. Penguin Group , The, 2011.
- [3] Jason Kincaid. The secret sauce that makes facebook's news feed tick, 2010.
- [4] Paul Covington, Jay Adams, and Emre Sargin. Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems, RecSys '16*, pages 191–198, New York, NY, USA, 2016. ACM.
- [5] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. Grouplens : An open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work, CSCW '94*, pages 175–186, New York, NY, USA, 1994. ACM.
- [6] Marshall Van Alstyne and Erik Brynjolfsson. Global village or cyber-balkans ? modeling and measuring the integration of electronic communities. *Management Science*, 51(6) :851–868, 2005.
- [7] Guillaume Champeau. Terrorisme : le gouvernement s'attaque aux bulles filtrantes de facebook, youtube,..., 2016.
- [8] Tien T. Nguyen, Pik-Mai Hui, F. Maxwell Harper, Loren Terveen, and Joseph A. Konstan. Exploring the filter bubble : The effect of using recommender systems on content diversity. In *Proceedings of the 23rd International Conference on World Wide Web, WWW '14*, pages 677–686, New York, NY, USA, 2014. ACM.
- [9] Paul Resnick, R. Kelly Garrett, Travis Kriplean, Sean A. Munson, and Natalie Jomini Stroud. Bursting your (filter) bubble : Strategies for promoting diverse exposure. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work Companion, CSCW '13*, pages 95–100, New York, NY, USA, 2013. ACM.
- [10] Siamak Faridani, Ephrat Bitton, Kimiko Ryokai, and Ken Goldberg. Opinion space : A scalable tool for browsing online comments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10*, pages 1175–1184, New York, NY, USA, 2010. ACM.
- [11] Travis Kriplean, Michael Toomim, Jonathan Morgan, Alan Borning, and Andrew Ko. Is this what you meant ?: Promoting listening on the web with reflect. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12*, pages 1559–1568, New York, NY, USA, 2012. ACM.
- [12] Panagiotis Adamopoulos and Alexander Tuzhilin. On over-specialization and concentration bias of recommendations : Probabilistic neighborhood selection in collaborative filtering systems. In *Proceedings of the 8th ACM Conference on Recommender Systems, RecSys '14*, pages 153–160, New York, NY, USA, 2014. ACM.
- [13] Camelia Constantin, Ryadh Dahimene, Quentin Grossetti, and Cédric Du Mouza. Finding Users of Interest in Micro-blogging Systems. In *International Conference on Extending Database Technology, EDBT*, Bordeaux, France, March 2016.
- [14] P. De Meo, E. Ferrara, G. Fiumara, and A. Provetti. Generalized louvain method for community detection in large networks. In *2011 11th International Conference on Intelligent Systems Design and Applications*, pages 88–93, Nov 2011.