

# Résumé de l'article sur le filtrage collaboratif sensible à la récence pour la recommandation du prochain panier

IFERROUDJENE Mouloud

## 1-Présentation

### 1.1-Problématique

La problématique à résoudre est la prédiction du prochain panier d'achat, cela veut dire qu'on a au préalable l'ensemble des articles du panier d'un client et qu'on souhaite prédire son futur panier. Particulièrement ce problème est soumis aux contraintes des services d'épicerie, qui sont :

- **La relations** des achats aux **besoins essentiels** de du client en prenant en compte **la saisonnalité** des produits.
- **La répétitivité** des achats ainsi que **la fidélité** du client aux différentes marques.
- **Multipeople** : le contenu du panier répondant à un groupe d'individus.
- **Multitasks** : le panier du client contient divers produits pour satisfaire des besoins non essentiellement reliés.

### 1.2-Solution proposée

La conception d'un système de recommandation est considérée de grande utilité car il permet aux utilisateurs de gagner du temps, en leur suggérant des produits qu'ils achètent habituellement en apprenant les besoins des utilisateurs tout en introduisant de la diversité et de la dynamique de consommation.

Dans cet article, les auteurs proposent une solution sous forme de Framework qui exploite la tendance naturelle de l'être humain, sous forme d'un seul prédicteur appelé « recency aware user-wise popularity ». Celui-ci est basé sur un ensemble de techniques de filtrage collaboratif (CF) pour la recommandation des Top-N articles. Ce Framework repose essentiellement sur la popularité et la récence des produits pour capturer les habitudes d'achats d'un consommateur ; plus particulièrement dans les services d'épicerie en vue de leurs popularité grandissante et les différentes contraintes qu'il soulèvent.

Par la suite, les auteurs étudient l'influence de la dernière date d'achat du client sur ses futures paniers par le biais d'une expérimentation sous forme d'analyse empirique détaillée. Celle-ci vise à montrer la robustesse de la solution proposée tout en soulignant/justifiant la prise en compte de la récence comme variable pertinente dans la construction du modèle.

Enfin, Ils concluent l'analyse par une comparaison du modèle proposé avec différents autres modèles d'état de l'art en s'appuyant sur la métrique d'évaluation **nDCG** (Gain cumulé actualisé normalisé). Ils abordent finalement les futures améliorations possibles qu'ils comptent apporter à la solution.

## 2-Travaux en relation

La plupart des travaux fondateurs dans ce domaine s'appuient sur l'utilisation de règles d'association, en trouvant des articles (groups d'articles) qui témoignent de la présence d'autres articles dans le même panier et souvent dans la littérature on trouve une décomposition des ces méthodes en deux paradigmes : les systèmes de recommandations séquentielle et générale.

Parmi les méthodes citées par l'auteur, on trouve :

- **Le model FPMC** <sup>1</sup> : (Stat-Of-The-Art Model) Basé sur le schéma de transition des chaînes de Markov, où l'algorithme apprend une factorisation de rang inférieur de la matrice de transition, en utilisant les SBPR comme critère d'optimisation.
- **Les approches basées sur l'apprentissage de la représentation**, Ils tentent d'apprendre une représentation pour les utilisateurs et les articles, parmi elles sont citées :
  - **Triple2vec** <sup>2</sup> : se base sur divers principes tel que la complémentarité, compatibilité et la fidélité.
  - **Adaloyal** : Ce modèle construit des probabilités pondérées par la fidélité de l'utilisateur et propose une représentation hybride des utilisateurs et des articles.

<sup>1</sup> Steffen Rendle, Christoph Freudenthaler, and Lars Schmidt-Thieme. 2010. Factorizing Personalized Markov Chains for Next-basket Recommendation. In Proceedings of the 19th International Conference on World Wide Web (WWW '10). ACM, New York, NY, USA, 811–820. <https://doi.org/10.1145/1772690.1772773>

<sup>2</sup> A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko. Translating embeddings for modeling multi-relational data. In NIPS, 2013. Triple2Vec: Learning Triple Embeddings from Knowledge Graph

### 3-Approche Proposée

L'approche proposée identifie d'abord les catégories de produits qui sont souvent vendues ensemble, puis elle prend en compte le panier partiel de l'utilisateur et classe les articles en fonction de leur complémentarité avec les articles déjà présent dans le panier. Enfin, elle tient compte de la périodicité, en suggérant des produits qui sont plus susceptibles d'être achetés par l'utilisateur, en raison de l'intervalle de temps écoulé depuis l'achat précédent du produit.

En explorant les différentes caractéristiques de système, deux aspects importants sont à distinguer ; la popularité (l'indicateur le plus fiable) et les achats non-fréquent, relié a des situations spécifiques. Pour surpasser ces limites, un indicateur est proposé pour le modèle appelé « **User-Wise Popularity** » qui est insensible aux relations entre les différents articles ainsi que la rareté des achats et qui se base essentiellement sur la fréquence d'achats de produit sans prendre en considération la tendance particulière des utilisateurs envers certains articles.

Un autre aspect important à pris en considération de cette étude est la **notion de récence** qui est définie comme étant la date du dernier achat d'un article donné. Les auteurs posent l'**hypothèse** que si un article est acheté récemment alors il est plus probable qu'il sera inclus dans le futur panier. Cette caractéristique temporelle du système permet de détecter le changement de popularité (tendance) d'un article particulier en définissant une métrique nommée « **recency aware user-wise popularity** » qui encode à la fois la recense ainsi que la popularité d'un produit.

Considérant cette dernière métrique, deux nouvelles approches, **Item-popularity-based CF** et **user-popularity-based CF** de filtrage collaboratif ont été définies en remplaçant les évaluations « rating » par l'indicateur « **recency aware /user-wise popularity** » dans les deux approches classiques de CF, celle basée sur l'utilisateur et celle basée sur l'article « Item ».

Ces deux dernières permettent de mesurer la similarité entre les deux différents produits respectivement utilisateurs ayant acheté le même produit en se basant sur des fonctions de calcul de similarité « **Asymmetric cosine similarity** » et de **score** pour bien mettre en évidence les produits les plus susceptibles d'être sélectionnés pour le prochain panier.

### 4-Expérimentation et résultats

Pour évaluer les trois déférentes approches, deux ensembles de données (de taille et distribution différentes) pré-traités ont été utilisés. Elles sont tirées de « **Dunnhumby The Complete Journey** <sup>3</sup> » et « **Instacart** <sup>4</sup> ».

Après les avoir divisés en deux sous-ensembles : validation et test, ils sont ensuite mis à l'étude en variant les différents hyperparamètres des modèles. Par la suite une comparaison, basé sur la métrique **nDCG** <sup>5</sup>, est faite entre d'autres méthodes existantes (quatre méthodes de base et deux méthodes d'état de l'art FPMC et Triple2vec +AdaLoyal <sup>6</sup>).

Cette comparaison a fini par confirmer les hypothèses de départ des auteurs tout en montrant que la méthode proposée (**UP-CF@r**) obtenait les meilleures performances relatives.

### 5-Conclusion

En conclusion, dans cet article un Framework pour la résolution du problème de prédiction du prochain panier est proposé, elle se base sur une approche de filtrage collaboratif, l'étude montre que les approches proposées permettent d'atteindre de meilleures performances que les méthodes de base.

Ce Framework se base essentiellement sur deux principes, la popularité et la notion de récence. Ces derniers permettent de palier au problème de variabilité de la popularité des articles ainsi que le phénomène la rareté de choix d'article dû au changement des habitudes du consommateur.

### 6-Futurs développements

Les auteurs de cet article proposent d'étudier, prochainement, l'influence de la notion de consecutivité d'achats d'articles ainsi que d'approfondir leurs compréhensions de la corrélation temporelle des achats de produit.

---

<sup>3</sup> <https://www.dunnhumby.com/careers/engineering/sourcefiles>

<sup>4</sup> Accessed on 2018. The Instacart Online Grocery Shopping Dataset 2017. <https://www.instacart.com/datasets/grocery-shopping-2017>.

<sup>5</sup> Normalized DCG, [https://en.wikipedia.org/wiki/Discounted\\_cumulative\\_gain#Normalized\\_DCG](https://en.wikipedia.org/wiki/Discounted_cumulative_gain#Normalized_DCG)

<sup>6</sup> T2V+Ada's implementation, <https://github.com/MengtingWan/grocery>