Summary of Recency Aware Collaborative Filtering for Next Basket Recommendation paper

IFERROUDJENE Mouloud

1. Presentation

1.1. Challenges & Goal

The paper addresses the problem of the next basket recommendation, which is described as the prediction of the future item while knowing the users' basket contents and other information as well. Particularly this problem is exposed to the constraints of the grocery services, which are:

- **The relation of the purchases** to the customer's essential needs, taking into account the seasonality of the products.
- **The repetitiveness** of the purchases of the same group of items over time.
- The customer's **loyalty** to different brands.
- **Multi-people:** meaning that the purchases are results of a single user's tastes, but it can instead be a complex compound of different tastes of many people.
- **Multitasking:** meaning that the user often has various shopping tasks that are not necessarily related to each other.

Thus, the complete grocery cannot always be guessed by a few products, which makes the task very different from the standard recommendations.

1.2. Proposed solution

The concept of a recommender system is considered to be very useful in the case of saving time by reminding the users of their natural products and also in the case of allowing dynamic consumption through diversity and proposing new products.

The authors propose a framework that exploits the natural human tendency to be repetitive as well as the popularity aspect. These two are merged and represented as a single predictor called ***recency aware user-wise popularity*** and used by the framework built on collaborative filtering (CF) for top-N item recommendation.

This framework relies mainly on the popularity and recency of products to capture a consumer's buying habits, more specifically in grocery services, because of their recently growing popularity and the various constraints they raise.

Next, the authors study the influence of time, in particular recency, on future shopping baskets through a detailed empirical analysis and experiments. The aim is to show the robustness of the proposed solution while emphasizing the inclusion of such a relevant variable in the model design.

Finally, the analysis through a comparison between the proposed model and various other baseline and states of the art models using the **nDCG**[5] (Normalized Discounted Cumulative Gain) evaluation metric. They also discussed further possible improvements to the proposed framework.

2. Related work

Most of the original work in this field relies on the use of association rules, finding articles (or groups of articles) that show the presence of other articles in the same basket. Often in the literature, we find a breakdown of these methods into two paradigms: sequential and general recommendation systems.

The following methods, among others, are mentioned by the authors:

- **The FPMC**[1] **model**: a state-of-the-art model based on the Markov chain transition scheme, where the algorithm learns a lower order factorization of the transition matrix, using SBPR as an optimization criterion.
- **Representation learning approaches**, methods that learn a representation for users and articles, as they cited:
    - **Triple2vec**[2]: based on various principles such as complementarity, compatibility, and fidelity.
    - **Adaloyal**: This model builds probabilities weighted by user loyalty and provides a hybrid representation of users and items.

[1] 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

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

3. Proposed Framework

The proposed approach first identifies the product categories that are often sold together, to classify the items according to their complementarity with the items already in the basket, taking into consideration both the user's partial basket and the items' purchase periodicity. Furthermore, it suggests products that the customers are more likely to buy, due to the time interval since the last purchase of the product.

While reviewing the different characteristics of the system, two important aspects are to be distinguished, **popularity** (the most reliable indicator) and **non-frequent purchases** related to specific situations. To deal with the latter, an indicator is proposed for the model called "**User-Wise Popularity**" which is insensitive to the correlations between different items as well as the rarity of purchases and mainly based on the frequency of product purchases without taking into account the particular tendency of users towards specific items.

Another critical aspect to consider in this study is the **recency's** notion, which is defined as the date of the last purchase of a given item. **The authors hypothesize** that if an item is purchased recently, then it is more likely to be included in the future basket. This **temporal feature** of the system allows detecting the change in popularity (trend) of a particular item by defining a metric called "**recency aware user-wise popularity"**, which encodes both the recency and the popularity of a product.

For that matter, they defined two new collaborative filtering approaches: **Item-popularity-based CF** and **user-popularity-based CF**, by replacing the "rating" evaluations respectively, with the **recency aware** and **user-wise popularity**" indicators.

The latter defined approaches allow measuring the similarity between two different products, and respectively users who bought the same product, using "**asymmetric cosine similarity**" scoring functions to highlight the products most likely to be selected for the next basket.

4. Experimentation and results

To evaluate the three different approaches, they used two pre-processed data sets of varying size that come from a different distribution: « **Dunnhumby The Complete Journey** [3] » and « **Instacart** [4] ». After splitting the data between train/Dev sets and doing some hyperparameters tuning of the proposed models, they came up with impressive comparison results, based on the **nDCG**[5] metric, with some chosen baseline and state-of-the-art existing methods (FPMC et Triple2vec, +AdaLoyal [6])

This comparison eventually confirmed the authors' initial hypotheses while showing that the proposed method (UP-CF@r) obtained the best relative performance.

5. Conclusion

In conclusion, the authors proposed a set of collaborative filtering approaches for solving the problem of predicting the next basket. The study shows that the proposed framework achieves state-of-the-art performance.
This framework is mainly based on two principles: (i) popularity and (ii) recency, which deals with the problem of variability in item popularity (e.g., concept drift, seasonality) as well as the phenomenon of item choice's rarity due to consumer's varying habit over time.

6. Future works

The authors of this article plan to study the correlation between products and time and to understand better the notion of consecutiveness among products' purchases as some literature suggests, and possibly add it to the proposed model.

---

[3] https://www.dunnhumby.com/careers/engineering/sourcefiles

[4] Accessed in 2018. The Instacart Online Grocery Shopping Dataset 2017. https: //www.instacart.com/datasets/grocery-shopping-2017.

[5] Normalized DCG, https://en.wikipedia.org/wiki/Discounted_cumulative_gain#Normalized_DCG

[6] T2V+Ada's implementation, https://github.com/MengtingWan/grocery