Knowledge-Based Systems 000 (2016) 1–11

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Item-based relevance modelling of recommendations for getting rid of long tail products

Daniel Valcarce*, Javier Parapar, Álvaro Barreiro

Information Retrieval Lab, Computer Science Department, University of A Coruña, Campus de Elviña s/n, 15071 A Coruña, Spain

ARTICLE INFO

Article history:
Received 2 December 2015
Revised 25 February 2016
Accepted 22 March 2016
Available online xxx

Keywords:
Recommender systems
Collaborative filtering
Relevance models
Long tail

ABSTRACT

Recommender systems are a growing research field due to its immense potential application for helping users to select products and services. Recommenders are useful in a broad range of domains such as films, music, books, restaurants, hotels, social networks, news, etc. Traditionally, recommenders tend to promote certain products or services of a company that are kind of popular among the communities of users. An important research concern is how to formulate recommender systems centred on those items that are not very popular: the long tail products. A special case of those items are the ones that are product of an overstocking by the vendor. Overstock, that is, the excess of inventory, is a source of revenue loss. In this paper, we propose that recommender systems can be used to liquidate long tail products maximising the business profit. First, we propose a formalisation for this task with the corresponding evaluation methodology and datasets. And, then, we design a specially tailored algorithm centred on getting rid of those unpopular products based on item relevance models. Comparison among existing proposals demonstrates that the advocated method is a significantly better algorithm for this task than other state-of-the-art techniques.

1. Introduction and motivation

Recommender systems are becoming increasingly popular to help users with the finding of relevant items. On the one hand, users are becoming more demanding and, on the other hand, more amount of information is available. For this reason, many e-commerce companies have started to use these systems for product recommendation with the intention of increasing the number of sales. From the business perspective, recommenders are effective tools to improve user satisfaction and, thus, sales revenue. The former is achieved through high quality recommendations whilst the latter is more dependent of the seller circumstances. It is hard to establish what is a good recommendation. Research efforts have focused on accuracy since the very beginning of the field of recommender systems. However, recent studies have pointed out the importance that other properties such as serendipity, novelty or diversity may have as key features to deliver good recommendations [14,23].

There exist several aspects that should be taken into account with respect to revenue enhancement. Fleder and Hosanagar have analysed thoroughly the effect of recommender systems in sales diversity [12]. These authors concluded that recommenders can increase sales if they discount popularity appropriately generating more diverse suggestions. The key idea behind this study is that recommenders should not focus solely on popular products but also on long-tail items. This proposition is the main idea of Anderson's book *The Long Tail: Why the Future of Business Is Selling Less of More* [2]. He coined the term long tail to refer to those less popular products that have a low demand in large catalogues. In the current context of e-commerce, he declared that promoting long tail items is crucial for both the user and the business: customers may discover new and unexpected relevant products while the companies may sell the majority of their stock. He claimed that a retailing strategy based on selling a large number of products in small quantities is more profitable than a business centred on selling large amounts of a small set of popular products.

Traditionally, recommendation approaches in the long tail [27,39] have explored ways for promoting long tail items in the recommendation process. However, addressing the long tail problem in that way is not particularly novel because a high performance recommender should recommend both popular and long tail items according to what fits best to the users' tastes. A growing body of literature has focused on improving the diversity of recommender systems [1,34]. Since more diverse recommendations are expected to lead to a larger catalogue coverage, more long tail products are likely to be recommended. In the same way, many

* Corresponding author. Tel.: +34 881 01 1276; fax: +34 981 167 160.
E-mail addresses: daniel.valcarce@udc.es (D. Valcarce), javierparapar@udc.es (J. Parapar), barreiro@udc.es (Á. Barreiro).

http://dx.doi.org/10.1016/j.knosys.2016.03.021
0950-7051/

studies have analysed the importance of novelty in recommendation. A recommendation is considered novel when the recommended item is unpopular (such as a long tail). Although enhancing diversity and novelty is a way for recommending more long tail products, we argue that there will still be items that vendors will be unable to sell.

The long tail nature of a product can be produced by very different factors: it is rarely sold and therefore has very few ratings, it is sold but it has almost no ratings given its embarrassing nature (e.g., sex toys) or it is barely recommended by the existing recommender. All these aspects make recommending in the long tail an extremely difficult task. Thus, effectively suggesting all the products of companies' catalogues is almost impossible following a unified approach.

In business terminology, the items from the inventory of a company that cannot be successfully sold are called excessive stock or *overstock*. Overstock can be the result of poor prediction of product demand but it can also be a consequence of market fluctuations.Thus, an effective process management and a diverse recommender may minimise overstock effects. However, as we stated before, despite the efforts oriented towards improving sales of long tail products, there will be situations when companies will not be able to sell some items. In this case, we think that a specially designed recommender may provide interesting benefits. Therefore, in this paper, we develop a formal recommender model for dealing with these products which may help companies to liquidate their excessive stock.

To the best of our knowledge, the specific problem of getting rid of long tail items (i.e., liquidating stock) has never been addressed systematically. Thus, first, we state the overstock clearance problem formally, we study how to evaluate this task and how this process differs from the classic recommendation problem. Next, we propose three methodologies for estimating which items are part of the excessive stock of businesses given a standard recommendation dataset. Then, we adapt a collaborative filtering approach whose roots lie in pseudo-relevance feedback (a well-known task in the Information Retrieval field) to the overstock liquidation problem. This approach builds a statistical relevance model of each of the long tail items which enables the identification of target users for selling those particular items. Finally, we conduct a series of thorough experiments to analyse and compare the performance of our proposal with other collaborative filtering algorithms. The results confirm our intuition that a probabilistic item-based relevance model enables to build an effective recommender to get rid of the long tail products.

In brief, the contributions presented in this article are: (1) the formulation of a novel recommendation task consisting in liquidating long tail items, (2) the proposal of three different methodologies for estimating the overstock products of a dataset and its critical analysis, (3) the design of a collaborative filtering algorithm based on relevance modelling for the task of liquidating long tail items and (4) the empirical comparison of different collaborative filtering algorithms (including the proposed one) under this novel task.

2. Getting rid of the long tail

The objective of a recommender system is to elaborate personalised rankings of products for each user. Every recommendation task involves a set of products or items (we will use these two terms interchangeably), which we denote as $\mathcal{I}$, and a set of possible customers or users, denoted as $\mathcal{U}$. Recommendation in the long tail refers to the generation of item recommendations to users including not only popular products but also long tail ones. However, here we propose a different approach: for those items in the long tail we want to get rid of, we want to identify those potential users that will buy the product, even when that item is not on the top of the users' preferences. Of course, this approach does not replace the classic recommendation. On the contrary, it is specially designed to address a business concern and should be used as a complement to current methods in a operational setting. We claim that recommenders may help to get rid of long tail products which can be seen as a proxy representation of the overstock phenomena.

Recommender systems are usually classified in three main categories: content-based, collaborative filtering and hybrid techniques [6,28]. While content-based recommenders employ the properties of the items to recommend similar items to those appreciated by the user [20], collaborative filtering techniques rely on data about past interactions between users and items (ratings, purchases, clicks, etc.) [11]. And, finally, hybrid methods combine algorithms from both families. Content-based algorithms are usually preferred for improving novelty because they find similar items based on content, not on popularity [24]. However, sometimes the available information about items is not adequate for using this family of methods. Additionally, content-based recommenders may lead to over-specialisation suggesting only items that are very similar to those rated by the target user [11]. Therefore, we propose a collaborative filtering approach specially designed for the task of getting rid of the long tail.

Since we are focusing on a pure collaborative filtering approach, we are interested in the users' feedback. We are assuming that we have explicit feedback, i.e., ratings from users to items. However, our proposal is also suitable for dealing with implicit feedback such as clicks or purchase information that can be represented with binary ratings. We indicate the relationship between the user u and the item i with the rating $r_{u,i}$. We can organise these ratings in a user-item matrix $\mathbf{R}$. Additionally, $\mathcal{U}_i$ refers to the subset of users that rated the item i. Similarly, $\mathcal{I}_u$ denotes the subset of items that have been rated by the user u.

Collaborative filtering algorithms exploit the past interactions between users and items to create meaningful recommendations. We can classify these techniques in two main categories [11]. Model-based approaches build a predictive model from the ratings. In contrast to these systems, which learn patterns from historical data, memory-based methods (also known as neighbourhood-based) actually employ all the ratings stored in the system. Memory-based algorithms can compute suggestions using the information of like-minded people (user-based) or, on the other hand, recommend items that are similar to the ones the user liked in the past (item-based).

The classic task in the field of recommender systems consists in estimating, given a particular user, the top items for which this user is most likely to be interested in. This task can be formulated as finding a scoring function $s : \mathcal{U} \times \mathcal{I} \to \mathbb{R}$ such that, for each user u, we can generate a ranked list of k items $L_u^k \in \mathcal{I}^k$ sorted by decreasing score order.

2.1. Problem statement

In this article, we propose a novel recommendation problem: instead of generating the best item suggestions for each user, we aim to find the best users for each long tail product. The inversion of the classic recommendation task (recommending users to items in the place of suggesting items to users) was recently studied to improve sales diversity [34]. These authors explored the inverted task and proposed a probabilistic approach that enhances sales diversity. Still, at the end, their intent was to improve the original recommendation problem (suggesting items to users). In contrast, our intention is to address a very different problem: how to get rid of the excessive stock suggesting the most suitable users for each item.

Please cite this article as: D. Valcarce et al., Item-based relevance modelling of recommendations for getting rid of long tail products, Knowledge-Based Systems (2016), http://dx.doi.org/10.1016/j.knosys.2016.03.021