Decision Support Systems xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

Decision Support Systems

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

ExUP recommendations: Inferring user's product metadata preferences from single-criterion rating systems

Alfred Castillo [a], Debra Vander Meer [b,*], Arturo Castellanos [c]

[a] Department of Management, Human Resources, and Information Systems, Orfalea College of Business, Cal Poly San Luis Obispo, San Luis Obispo, CA 93407, United States
[b] Department of Information Systems and Business Analytics, College of Business, Florida International University, 11200 SW 8th Street, RB 256B, Miami, FL 33199, United States
[c] Department of Information Systems and Statistics, Zicklin School of Business, Baruch College, New York, NY 10010, United States

ARTICLE INFO

Article history:
Received 20 February 2017
Received in revised form 16 February 2018
Accepted 17 February 2018
Available online xxxx

Keywords:
Recommendation system
E-commerce
Single-rating
Product metadata
Multi-valued attributes

ABSTRACT

Recommendation systems make use of complex algorithms and methods to provide recommendations to consumers. Typically, online rating schemes use a single rating metric that captures the overall user experience with a product. Nevertheless, this might hinder the intricacies of how a product's attributes influence an individual's preferences. While it is possible to use sentiment and semantic analysis to interpret free text in user reviews, if available, to gain insight into a user's reasons for a product rating, these methods are expensive to implement and error prone, and rely on significant data input from the user. To overcome these challenges, we propose a method for inferring user preferences and generating recommendations without relying on the availability or quality of text reviews. Specifically, our method is designed to use existing product metadata and user rating patterns to shed light on how the attributes of a product correspond to individual preferences. Our method uses only the user's history of ratings and the corresponding product attributes to generate predicted ratings for products a user has not yet experienced. This work extends existing work in this area by focusing on multi-valued attributes, and considering the distinct impact of each attribute value in a user's preferences. In terms of computational complexity, our method runs in linear time, making it feasible for real-time implementations. Our experimental results showed that, compared with the two best-performing existing state of the art methods, our method provided review score predictions with up to: 47.7% greater precision, 6.9% greater recall, and 20.5% greater F-measure than existing methods.

1. Introduction

Organizations use recommender systems (RS) to better understand their customers' preferences and to leverage that understanding for filtering, recommending, and cross-selling products that the users will like. Techniques in RS use customer-provided data about the various products that they have experienced (e.g., ratings and comments).

The amount of information gathered from users for analysis differs across organizations. The most common implementation is single-dimensional, where a customer is simply asked how much he or she liked a purchased product or service on a 5-star Likert-like scale. A one-star rating means that the customer did not like the product, and a five-star rating indicates that the customer really liked it. This approach has the benefit of simplicity (it requires little effort on the part of the customer and system), which may explain its popularity in use in high-traffic internet sites such as TripAdvisor and Yelp. Netflix previously used the five-star rating system, but recently adopted an even simpler dichotomous thumbs up/down system [1]. These systems have the limitation of producing a single-point data with no dimensional information. Any richness of data related to ratings must be discovered via other means, e.g., text-mining the user provided rating summary. On the other hand, multi-criteria systems are multi-dimensional, where customers are asked for more specific preference information along a range of categories [2]. For example, Zagat asks for information across four criteria describing the user's experience in a restaurant: food, décor, service, and cost. Table 1 shows examples of single-dimension and multiple-dimension rating systems for comparison.

Intuitively, one would expect that providing more rating dimensions for the user to fill in would be beneficial and should therefore represent the majority of implementations. However, a survey of sites that collect user experience data will show that single-dimensional ratings (the focus of this paper) are by far the most common. This begs the question of why organizations would decide against having a multi-dimensional rating scale. It is possible that organizations are hoping to avoid survey fatigue [3], where the number of people willing to provide data is generally inversely proportional to the number of data points they are expected to provide (e.g., trading off richness of data to gain a larger volume of customer input). This reduces the transaction time and provides a better user experience [4].

* Corresponding author.
E-mail addresses: acasti63@calpoly.edu (A. Castillo), vanderd@fiu.edu (D. Vander Meer), Arturo.Castellanos@baruch.cuny.edu (A. Castellanos).

https://doi.org/10.1016/j.dss.2018.02.006
0167-9236/

Table 1
Domain examples with single and multiple dimension rating systems.

	Single dimension		Multiple dimension	
Restaurants	Yelp	User rating with comments	Zagat	Food, decor, service, cost
Movies	Netflix	Thumbs up/down rating	Kids-In-Mind.com	Sex & nudity, violence & gore, profanity
Books	Barnes and noble	User rating with comments	CompassBookRatings.com	Recommended age, overall rating, profanity/language, violence/gore, sex/nudity

A closer look at a typical single-dimension rating scenario reveals opportunities for preference analysis. To illustrate this, consider a scenario where two different sets of people dine in the same restaurant on the same day. One, a mother accompanied by her young children, was impressed with the child-friendliness of the establishment, but did not particularly enjoy the food. Another, an older couple, may have loved the food, but were perhaps less excited about the child-friendly atmosphere. These differences impact the perceptions of their experiences at the restaurant. Suppose that each of these customers were to rate the restaurant a 4 (out of 5), based on their overall experiences. They would do so for vastly different reasons, which are not captured in the single-dimension rating. Here, a collaborative filtering approach would consider these customers to be similar, even though they likely have different preferences, making a content-based approach seem appropriate.

Finding these underlying differences using single-rating data alone is challenging. In cases where users provide textual data describing their rating rationale, the unstructured data may be difficult to analyze due to incompleteness or irrelevance, as well as challenges related to context-dependent word meanings.

Ideally, we would like to have the best of both worlds data-wise: a rich multiple-dimension dataset to build inferences from, and minimal customer effort in required data entry. The products themselves provide an undervalued rich source for additional data, which is clean, as they are objects stored in a database so that they can be dynamically populated on the company website pages. Consider Yelp, for example, where a restaurant is described across a number of features, including the type of food, relative cost, the availability of parking, whether alcohol is served, and a number of other attributes. By understanding the formative relationship of a user's rating to the attributes of a product, it may be possible to infer the diversity of preferences among customers, and improve recommendation accuracy. The challenge is using these product attributes, in conjunction with user recommendation scores, to develop user preferences that are effective at generating recommendations.

In this work, we propose an approach for inferring the embedded dimensionality in user ratings, based on users' rating histories. Our method uses this embedded dimensionality to aid in explaining the variance in ratings users provide across similar products. Because these preferences differ on a person-to-person basis, our approach is based on modeling individuals rather than groups.

Cacheda and colleagues proposed an initial approach to use product attributes and individual preferences to generate recommendations [5,6] (referred to throughout the paper as the "Cacheda" approach). Their work provided recommendations within a specific product label, and considered combinations of product attributes within a product feature as one categorization. Consider the movie *Pineapple Express*, for example. It is labeled as a crime, action, and comedy movie in the Genre product feature. Any relative evaluation of that movie using the original Cacheda method would have to compare that movie with other movies that were also labeled crime, action, and comedy. This significantly reduces the sample size to draw inferences from, and can create an artificial "cold start" problem even when ample data is available. By not considering how well-received a movie is relative to the multiple populations (i.e., multiple attribute values) it belongs to, Cacheda can only base inferences with the population that shares the full label set. We extend Cacheda's work by viewing a product as a member of possibly many populations, and compare that product relative to the other members that also belong to that categorization. Our approach generates predictions given that a user has provided a rating for other products that also contain some given categorization. This is because preferences are generated for each categorization so that they can be analysed in isolation from the confounding effects inherent to nominal multi-valued attributes. Our methodology, Ex Uno Plures (ExUP) – meaning "out of one, many" in Latin – makes the following contributions:

First, we propose a method for generating individual recommendations based on an individual's tendencies in rating products in combination with multi-valued nominal product description attributes. This method extends Cacheda's original work in this area by considering the contribution of each descriptive attribute relative to each user's preferences, rather than as one set of product attributes. Our method is generalizable to any domain where users provide a single-dimension rating for products associated with multi-valued nominal dimensional data.

Second, we performed an analysis of the computational complexity of our method. We show that it maintains linear complexity, updateing for each new product rating in near real-time.

Third, we implemented our method as a prototype to demonstrate the feasibility of our approach. We created a large 10 M record database of movie data and user ratings from online sources to sample from and serve as a testbed dataset for accuracy comparison experiments.

Finally, we ran a set of experiments to compare the accuracy of our method to that of existing methods, and demonstrate that we achieve 6.74% to 47.74% greater precision, 0.76% to 6.85% greater recall, and an F-measure that is 4.79% to 20.54% greater than existing methods. We also ran a set of experiments to show that our method performs well with datasets with various unique attribute set sizes.

The remainder of the paper is organized as follows: In Section 2, we discuss background and related work. In Section 3, we discuss the solution approach and introduce our method. In Section 4, we describe the instantiated ExUp artifact and the experimental setup, and present a discussion of the results. The paper then concludes with a discussion of implications, limitations, and future work.

2. Related work

Recommendation systems (RS) are responsible for selecting a subset of items or products that may be of interest for users from a large pool of alternatives, with the assumption that latent user preferences can be inferred from a user's demographic data (e.g., gender, age, income, zip code), explicit transactional data (e.g., product ratings, comments), implicit transactional data (e.g., adding/removing items from carts), or an item's attribute data (e.g., product brand, product price) [7]. RS is used in many web-based implementations from e-learning to e-commerce to e-government [8] via three primary categories: collaborative filtering (CF), content-based filtering, and hybrid approaches [7,9,10]. From an information retrieval perspective, there are also knowledge-based implementations of RS [11]; however these differ from the former three categories of RS in that they rely less on generating latent user preferences, and more on explicitly defined knowledge as the primary mechanism for providing recommendations (e.g., a user profile populated with specific preferences) [12].

The various implementations of RS are similar in goal, matching users with products, but differ in types of algorithm(s) used, the focus of preferential analysis in utility functions, and the type of data used in analysis or generated as part of the process. The most common RS technique is collaborative filtering (CF) [13], which can be: (1) memory-

Please cite this article as: A. Castillo, et al., ExUP recommendations: Inferring user's product metadata preferences from single-criterion rating systems, Decision Support Systems (2018), https://doi.org/10.1016/j.dss.2018.02.006