Bayesian Product Suggestions

There is a growing need for Bayesian probability techniques and how they can be used to develop a set of related product suggestions for customers in the e-commerce marketplace. Since electronic shopping has rapidly made its way into millions of households, developing an artificial intelligent personal shopping assistant is crucial. Amazon.com, the world’s largest online retailer, has over 30 million customers purchasing items from the website on a 24/7 basis (Amazon.com). By taking a user’s past purchases into account, suggestions can then be presented in order to expose the user to a product that they might not have purchased otherwise. This method, and many others, is becoming a crucial aspect to the online shopping experience – so much so that customers are beginning to shy away from websites that don’t offer the feature and give their loyalty over to companies that do. This paper intends to present techniques (in technical detail) that are being used today as well as the usefulness, known struggles and rewards of an artificial intelligent personal shopper.

Utilizing Related Products for Post-Purchase Recommendation in E-commerce

Unlike traditional shopping, purchasing through an online marketplace presents more opportunities for second chance purchases. In physical stores, customers tend to shop with a goal item in mind, find their way to the checkout, purchase the item and leave. However, with the introduction of e-commerce, companies have the opportunity to present users with additional purchases that would fit well with the items they had already chosen. This “post-purchase” stage has three essential challenges that, if met correctly, can become an additional sale for the seller: relevance of recommendation, recommendation coverage and time sensitivity (329).

The algorithm design presented in this paper focuses on determining recommended purchases based on the “active product” i.e., the product for which all recommendations are computed (330). The first and second challenges are encountered here by determining the relevance of other potential purchases to the active product and making sure the most relevant of these products are recommended. This is done by calculating two different factors – the category-level relevance factor and the product-level relevance factor. The category-level relevance factor is computed by matching sets of keywords into groups of related categories. All matched categories are given a probability of the likelihood of such category containing a plausible post-purchase product. Each probability is then ranked in order with the
top ones referenced as “top-related post-purchase categories for the active product” (330). Then, by matching related products that share key attributes with the active product, the product-level relevance factor is calculated in order to determine the likelihood that a customer would purchase a product after previously purchasing the active product. There is a direct correlation between purchase history and well predicted purchase recommendations. The calculated factor is used to take advantage of this history data and provide recommendations of products that other users bought after originally purchasing the active product.

The final challenge involves “recommending the right products at the right time” (331). This temporal factor is much more prevalent when added to the ease of use of an e-commerce shopping experience. Because users are more susceptible to make a purchase when it can be done at the press of a button, selling a related product is one click away. The algorithm presented in this paper claims that the use of historical purchase data and their category/product-level relevance factors can achieve a 4% to 37% coverage gain of relevant products while producing a 12% to 30% purchase rate gain depending on the time frame in which the recommendations are made (332).

**Selectively Acquiring Ratings for Product Recommendation**

The author proposes ways to fully embrace the importance of a “recommendation system” in an e-commerce marketplace by presenting strategies to “improve the accuracy of the predicted ratings generated by a collaborative filtering recommendation algorithm for the entire customer population,” (Huang 379). Not only does the paper provide specific mathematical computations and algorithm examples, it also explains the importance of using the algorithms correctly when used within a marketplace.

The paper brings to light the idea of active learning and the large possibilities it presents to recommendation systems. Huang claims that, “to the best of [their] knowledge, no previous study has investigate the generic active learning method that is applicable to any recommendation algorithms, especially the heuristic memory-based algorithms such as the popular user-based and item-based neighborhood algorithms” (381). In other words, an active learning method that is based upon user input (more specifically, ratings) could have a much larger impact in providing relevant recommendations to the user. The issue of obtaining this user-inputted data immediately arises and three different sampling methods are given: active learning sampling, sequential active learning sampling, and benchmark sampling. Active learning sampling involves providing the customer with products to rate
whose ratings would be useful to future recommendations. In its simplest form, to gather useless information would be useless to the problem at hand. One of the issues with active learning is that the amount of data acquired at each step is relatively small. Sequential active learning sampling is founded on the previous idea but increases the amount of gathered data by allowing the method “to be performed sequentially to further improve the sampling performance” (383). Finally, benchmark sampling is to ask the user to provide random recommendations of products that may or may not be related to purchase history.

The paper concludes with the analysis of an experiment run by the active learning algorithm on Netflix data and what the results mean to the usefulness of the product as a whole. It was found that “as more ratings were acquired to be used for recommendation, the prediction error of both algorithms generally decreased as expected, for all three sampling methods” (384). This new method of using active learning in recommending products proved to be an effective way in predicting desired products for the user.

**On-demand Feature Recommendations Derived from Mining Public Product Descriptions**

Unlike the previous two papers, this one strives to explain profitable techniques to acquire data of product descriptions, ratings and a user’s past purchases to then produce recommendations relevant to the accumulated data. Instead of asking the user to input ratings of their own, this algorithm takes advantage of the massive amount already in existence by mining product descriptions from major e-commerce websites to then give insight into what key words are used in the descriptions of products and what reviewers say about those products in general. From the collected data, insights can be drawn as to what other products are most likely related and thus would have the highest probability of being purchased by a user with the same interests.

The paper is spends most of its time on techniques used to mine useful data from other sources into “product profiles” (182). These product profiles are a set of core features that help to describe a product by keyword. Each product profile is used to generically describe a product based on these features and also give recommendations to other product profiles in which the same features are present. Once data is accumulated, the algorithm evaluates the relative closeness of the acquired data to the product at hand. If the mined data is deemed useful enough, it is then presented to the user as a recommendation.
How to Interpret the Helpfulness of Online Product Reviews: Bridging the Needs between Customers and Designers

With the incredible growth of information being presented to shoppers in e-commerce marketplaces a new problem has surfaced – how do shoppers distinguish between useful and un-useful information when faced with making informed consumer decisions? This paper proposes a way to help distil the massive amount of public information into succinct and relevant recommendations for users to relate. This is done by focusing “on how to automatically build the connection between online customer’s voting and a designer’s rating and predict the customer reviews’ helpfulness based on the review content,” (Jin 87). The author understands that all information is not relevant and there needs to be some sort of breakdown of the data to its most useful parts. This sorting of helpful and non-helpful product reviews is referred to as opinion mining.

The availability of product reviews has become a double edged sword. Opinion based recommendations are only useful if they are accessible to the customer, while a large amount tends to oversaturate the actual usefulness of the reviews as a whole. The algorithm expressed in the paper uses a “helpfulness rating” for each product review made on a site. The idea is simple; each product has a set of reviews that contain user opinions. Within each review among the set, a rating is given by shoppers based upon the usefulness of the review. Customers are presented with the reviews with the highest usefulness ratings first, in order to give them the most relevant information about the product. Unlike other recommendation systems that are given from an artificial intelligent mechanism, this method relies solely on the cooperation of a customer’s peers and the e-commerce community.

A Personalized Recommendation System Based on Product Attribute-specific Weights and Improved User Behavior Analysis

The amount of information a customer is presented with when participating in e-commerce shopping is continually increasing and persistently making it more difficult for a customer to be presented with only the most relevant and effective suggestions. The paper faces the problem as follows – instead of mining for information within product reviews and descriptions, the author proposes a way to solely follow the previous actions made by individual users in the past to give insight to what decisions they would make in the future. Presented as a naïve Bayesian algorithm that uses weights to differentiate different product attributes, the algorithm then applies these weights to give a more personalized recommendation system to individual users based on their specific tastes.
To evaluate how effective the Bayesian algorithm is at providing customers with the most relevant information, an on-line shopping mall was created and chosen users were given access (6). The algorithm was run on small and large sets of available review data and was found to provide nearly the same recommendations regardless of the initial amount of mined data. This was able to prove that the Bayesian algorithm did not relay on massive amounts of data but rather used probability to help give ranking to the most relevant and useful products for the customer.

The information presented helps to solidify the importance of implementing a useful and correct recommendation system for e-commerce marketplaces. The continual increase in electronic shopping opens up doors to exciting opportunities for companies to provide customers with even more information about products they are interested in. Each recommendation system described above provides an even easier shopping experience, complete with the confidence and satisfaction that the purchased product is one tailored specifically to the customer. Whether using a post-purchase recommendation system or one that is based upon reviews from the e-commerce community, it has never been easier for a shopper to find exactly what they want and be pleased with the online shopping experience.
Works Cited


