A Bayesian Framework for Modeling Price Preference in Product Search

Yinan Zhang
University of Illinois at Urbana-Champaign
Urbana, IL 61801
yzhng103@illinois.edu

Parikshit Sondhi
WalmartLabs
San Bruno, CA 94066
psondhi@walmartlabs.com

Anjan Goswami
WalmartLabs
San Bruno, CA 94066
agoswami@walmartlabs.com

Chengxiang Zhai
University of Illinois at Urbana-Champaign
Urbana, IL 61801
czhai@illinois.edu

Abstract

Product search is an emerging search application where optimization of search results relies critically on an accurate model of a user's price preference. In this paper, we propose a Bayesian framework for modeling a user's price preference with a particular focus on developing a smart price filter model for inferring a user's price preference based on the user's selection of price filters and optimizing ranking of products accordingly. Preliminary experiment results with product search log show promise of the framework, which opens up interesting opportunities for new research in the intersection of machine learning, information retrieval and economics.

1 Introduction

Product search is an emerging search application that differs from web search in that a user often needs to make purchasing decisions based on the search results. Recent research work proposed frameworks for optimizing the search results in product search with an emphasis on the products' utility to a user (e.g. [3],[5]), but little effort has been made to formally take a user's *price preference* into consideration. Price preference reflects a user's preference on the products' price, which is a critical aspect in the context of product search. However, the challenges of dynamically modeling a user's price preference in the interactive process of product search and optimizing product ranking based on the inferred price preference model has not yet been addressed in the existing literature.

In this paper, we propose a Bayesian framework for modeling a user's price preference with a particular focus on developing a smart price filter model for inferring a user's price preference based on the user's selection of price filters and optimizing ranking of products accordingly. Price filters (e.g., [$150-$200]) are natural navigation tools that users use to interact with search results. Traditionally, a price filter acts as a "hard" filter in the sense that, when the user applies the filter, all products priced outside of the filter are excluded while all products with prices falling within the filter are displayed and follow their original ranking order. Such an approach essentially assumed a naïve price preference model where the user's preferred prices are always exactly represented by the selected range, which is generally not accurate for the following reasons:

- Users may not know precisely the price of their interested products in the first place;
- When users select a price range, they might be more interested in products at prices around the middle of the range rather than those priced near the boundary of the range;
- Some product priced a little above the price range might also be appealing to the user, e.g. if it has a substantial value relative to its price.

In this work, we attempt to address all these issues in a principled way and propose the "smart price filter" (SPF) model to more accurately formalize the user's price preference and optimize the ranking of products in response to a user's selection of a price filter. We show a simple realization of the SPF model based on utility theories in classic economics studies [2], and demonstrate that the traditional price filter is a degenerated instantiation of the SPF model. We further show that the SPF model parameters can be estimated based on product search log data, and the produced ranking based on SPF is significantly better than the ranking produced by the standard Boolean price filter.

2 "Smart price filter" (SPF) model

The main problem we study is how to infer a user's price preference based on the selection of a price filter such as [$150-$200]. The price preference model implicitly used in a product search engine nowadays assumes that (1) the user does not accept any price outside this range, and (2) the user equally prefers any value inside the price range. As discussed before, neither assumption is true. Our main goal is to design a probabilistic model for price preference that could potentially allow prices outside the selected range to have non-zero probabilities and prices inside the range to have unequal probabilities in a reasonable way.

In a product search application, the price preference model has to be embedded in a product ranking model since from users' perspective, what matters is the ranking of the products that the system would show them after they select a filter. Thus, instead of explicitly modeling a user's price preference by a distribution over prices, we introduce a novel user action model that characterizes the probability that a user would select a particular price range filter r given that the user likes a particular product entity e (of known price). Such an implicit price model enables optimizing the product ranking directly based on a user's price filter selection. Below, we first present a Bayesian framework for optimally updating the product ranking based on a user's selected filter.

Formally, we propose that, when observing the user's selection of price filter r, the system ranks the products in descending order of the user's *posterior propensity* $p(e|r)$ in each product e that is derived from the user's *prior propensity* $p(e)$ and the user's *action model* $p(r|e)$ via Bayes' theorem:

$$p(e|r) \propto p(e)\, p(r|e)$$

According to this Bayesian ranking framework, the ranking of product entities can be seen to depend on two component models: the prior propensity $p(e)$ and the user action model $p(r|e)$. Since we focus on studying the price preference model, the prior propensity $p(e)$ is not our focus. In practice, the prior propensity could be estimated in different ways from the initial product ranking as well as additional personalization information of the user if available, and is generally available to us in a product search engine which ranks products based on probability of relevance.

The *action model* $p(r|e)$ is our focus, which characterizes how likely the user selects a price filter if the user is interested in a particular product and serves as a probabilistic linkage between the user's selected filter and the propensity. In a more general setting, the action model can probabilistically characterize a much wider range of user actions in addition to price filter selection, e.g. other facet selection actions, and even more broadly, additional user behavior categories such as query typing, conversational interactions with the system, etc. The Bayesian probabilistic framework we propose here could accordingly serve to provide formal guidance to many more ranking update problems in a variety of other scenarios, thus opening up many interesting directions for future research.

We now turn to the question of how to instantiate the action model $p(r|e)$ for our SPF model. We first note that the traditional price filter can be easily shown to be equivalent to using the following Boolean action model (i.e., "hard" price filter).

$$p^{(H)}(r|e) = \mathrm{I}_{\{a_r \le c_e \le b_r\}} = \begin{cases} 1 & \text{if } a_r \le c_e \le b_r \\ 0 & \text{otherwise} \end{cases}$$

where a_r and b_r are the lower and upper bounds of the filter r and c_e is the price of product e. As discussed earlier, such a model is inaccurate. Below we propose a more accurate action model. Specifically, we introduce a novel economics-inspired perspective of seeing a user's decision on selecting a price filter as dependent on the utility of the product of their interest, and derive the *utility action model* based on utility models in economics studies.

To characterize how likely the user selects a price filter, one might think of the multinomial distribution, where each price filter is considered as one possible outcome class and the probabilities of resulting in each outcome class are the goals for estimation. Such simple discrete choices would be a nice way to model the user's selection actions in cases of other discrete-valued facet filters, but *not* price filters due to their distinct characteristics. Firstly, the multinomial model ignores the essential ordinal relationship among the price filters. Moreover, many contemporary e-commerce search systems generate dynamic price filters or even allow users to type in price filters themselves; in such cases, the price filters may vary across different search sessions, but the multinomial model only allows for fixed outcome classes. Thus, we need to address these issues by developing a special method to realize the action model.

In this work, we relate price filters to the products' *utility* [2]. Generally, the utility of a product to a user is the amount of value the user sees in the product for which they're willing to pay. The product's utility in an ideal world should be fixed at around the price tag of the product, but this is often not the case in reality, so it is typically modeled as a random variable following a certain parametric distribution, termed the *utility model*. Our key postulation is that a rational user would always select a price filter that covers the utility of their interested product, and the user's action model could thus be computed based on the *utility model*. Now, we first give an instantiation of the utility model. We make the following assumptions based on classic economic theories:

- The utility u_e of product e to a user is composed of an observed component μ_e and an unobserved component ε: $u_e = \mu_e + \varepsilon$;
- The observed component μ_e represents the intrinsic value of the product;
- The unobserved component ε is modeled as a random variable that captures all unobserved factors in affecting the product's utility to the user.

Various probabilistic models have been proposed in economics studies to characterize the distribution of ε, including Gaussian distribution, logistic distribution, Type I extreme value distribution, etc [4]. In this work, we choose Gaussian distribution since it belongs to Exponential family and thus has a closed form conjugate prior, which is a desired property from the perspective of parameter estimation (to be discussed later). [1] Assuming $\varepsilon \sim \mathcal{N}(0, \sigma_e^2)$, where σ_e^2 represents the variance of the utility value of product e to users, we have the *Gaussian utility model*:

$$u_e \sim \mathcal{N}(\mu_e, \sigma_e^2)$$

Again, if we use a_r, b_r to denote the lower and upper bounds of price filter r, then, given that the user would select a price filter that covers the utility of the product they are interested in, we obtain the *utility action model*:

$$p(r|e) = \Phi\left((b_r - \mu_e)/\sigma_e\right) - \Phi\left((a_r - \mu_e)/\sigma_e\right)$$

where Φ represents the cumulative distribution function of the standard Gaussian distribution.

It could easily be observed that the utility action model could reduce to the "hard" price filter if we set $\sigma_e^2 \to 0$ and $\mu_e = c_e$ (the price of product e). In a more general case, when $\mu_e \neq c_e$, i.e. when the observed utility component may deviate from the product's price, we are able to capture the situations where the users may not have a precise idea of the price of their interested product, either due to over- / under- pricing on the market side or some general misinterpretation about the product's value on the user side. Further, when $\sigma_e^2 > 0$, we trigger the randomness of the utility value. In other words, users may not always see in the product the exact same utility value as the product's observed utility, though values around the observed utility are more likely than values deviating a lot from it. Learning that the user selects a particular price filter, conversely, would "hint" to the system via our Bayesian ranking framework that the user might be more interested in products with their observed utility value in the middle of the selected price range than those whose observed utility is near the boundary, and that there is also some slight chance that the user is interested in some other product with a observed utility value outside of the range. Thus, the SPF model nicely captures all of our motivating intuitions for improving the traditional price filter.

The two parameters in the utility action model, μ_e and σ_e^2, are typically unknown in the real world, so we need to make inference from observations of past user activities. In our experiments, we use Bayesian inference theories to estimate these parameters based on the search log data by using the Normal-Inverse-Gamma distribution as a conjugate prior to the Gaussian distribution [1]. The hyper-parameters could be heuristically set based on any available prior knowledge about the product's utility. For example, we can set the prior mean of μ_e at the product's price c_e.

In the search log data of a contemporary e-commerce search system, the users' eventual purchases (if any) together with their selected price filters (if any) are typically recorded for each search session.

[1] Type I extreme value distribution is also in Exponential family, but its conjugate prior is overly complicated.

Therefore, for each product e, the price filters selected in all the sessions that resulted in an eventual purchase of e could be collected to form the set of observations for making inference on the utility of e. To make the inference computation tractable, we can pick the mid point m_r of the price range in each selected filter r as an approximation to the whole range, so that each m_r is treated as one observed value of the product's utility. Since the price ranges of the price filters in e-commerce search engines are often relatively short segments as compared to the magnitude of product prices, it is typically reasonable to approximate the whole ranges by their mid points.

Note that such a Bayesian posterior update procedure could continue on and on when new observations are obtained from the search log, due to the property of conjugacy. Also, as the inference result, the utility action model could be computed either based on a point estimate or a posterior predictive distribution of the parameters μ_e and σ_e^2, both easily obtainable from the posterior distribution. [2]

3 Experiments

To demonstrate the effectiveness of the SPF model, we performed extensive experiments using the search log data of Walmart e-commerce search engine [3] to compare the SPF model against the traditional "hard" price filter. Due to space limitations, we only briefly outline the experiment set-up and results.

We collected around 62,000 search sessions in the past month in which the user (a) selected at least one price filter (which has been implemented in the "hard" filter manner) and (b) purchased one product at the end of the session. Without doing any advanced computation, we already noticed clear evidence in support of our motivating intuitions regarding the weaknesses of "hard" price filters: around 43% of the users' selected price filters do *not* cover the price of their eventual purchased product, and in such cases the user had to de-select the filter and either scanned the original ranking list again or tried other price filters and/or other facet selections so as to navigate to the product of their interest, which greatly affected the navigational efficiency as well as the overall user experience. We obtained the set of all price filters ever selected by any user that have led to the purchase of each product (around 6.0 filters per product), and used this set as observations to make inference on the utility action model of the product.

Then, we collected 50 popular queries together with the corresponding 50 ranked lists of top 128 products most relevant to each query as determined by the search engine, and carried out a simulated user evaluation. In particular, for each of the 50 rankings, we treated every price filter ever selected by a user that led to an eventual purchase of some product on the ranking as a simulated user search session: we respectively applied the "hard" price filter model and our SPF model to update the ranking, and then compare the ranks of the user's eventually purchased product in the two updated rankings. In cases where the "hard" price filter simply "missed" the eventually purchased product, we computed the rank as the total number of filtered products plus the original rank of the purchased product, emulating the scenario where the user scans the entire filtered list without finding the product, de-selects the filter, and then scans the original list to look for the product. The p-values as obtained from a one-sided Wilcoxon signed-rank test are less than 0.001 for *all* 50 ranked lists, strongly indicating that our SPF model is significantly superior than the traditional "hard" price filter in terms of its efficiency in helping users navigate to the products of their interests.

4 Conclusions and Future Work

We proposed a novel Bayesian framework for modeling a user's price preference and optimizing ranking of products. With the framework, a smart price filter model is developed based on utility theory in economics to model a user's price preference based on the user's selection of price filters. Experiment results show that the proposed new model is more effective than the traditional naïve Boolean price preference model. The proposed framework and model opens up interesting new research opportunities in the intersection of machine learning, information retrieval and economics. For example, we can further develop more accurate user action models and formalize more user activities that go beyond price filter selection. Another interesting extension is to introduce active learning for optimal preference elicitation (e.g., dynamically adjust the price ranges to focus on the most uncertain range of prices).

[2] Despite their different mathematical forms, the utility action models derived via these alternative ways in the real world do not differ much in effect, so we use the one coming from the point estimate in our experiment.

[3] http://www.walmart.com/

References

[1] M. K. Cowles. *Applied Bayesian statistics*, volume 98. Springer, 2013.

[2] R. Frank and B. Bernanke. *Principles of Microeconomics*. Mcgraw-Hill.

[3] B. Li, A. Ghose, and P. G. Ipeirotis. Towards a theory model for product search. In *Proceedings of the 20th international conference on World wide web*, pages 327–336. ACM, 2011.

[4] D. McFadden and K. Train. Mixed mnl models for discrete response. *Journal of applied Econometrics*, 15(5):447–470, 2000.

[5] J. Wang and Y. Zhang. Utilizing marginal net utility for recommendation in e-commerce. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 1003–1012. ACM, 2011.