Purchase Influence Mining: Identifying Top-k Items Attracting Purchase of Target Item

Sungchul Kim[1], Jinyoung Yeo[2], Eunyee Koh[1], Nedim Lipka[1]
[1]Adobe Research, San Jose, CA, United States
[2]Pohang University of Science and Technology (POSTECH), Pohang, South Korea
sukim@adobe.com, jinyeo@postech.ac.kr, {eunyee, lipka}@adobe.com

ABSTRACT

Web logs in e-commerce sites consist of user actions on items such as visiting an item description page, adding an item to a wishlist, and purchasing an item. Those items could be represented as nodes in a graph while viewing their relationships as edges according to the user actions. Based on the item graph, identifying items that attract users to purchase the target item could be practically used for supporting business decisions. To do this, we introduce a new task, called 'Purchase Influence Mining', that finds the top-k items (PIM-items) maximizing the estimated purchase influence from them to a target item. We solve this problem by modeling the purchase influence as the shortest path between item pair. According to the result, our approach more consistently finds the k PIM-items than the baseline.

Keywords

Data mining; Purchase influence mining; E-commerce

1. INTRODUCTION

In e-commerce sites, there is a wealth of information including item descriptions and user action logs such as visiting an item description page, adding an item to a wishlist, purchasing an item, and so on. We assume that a collection of user actions on items reflects purchase influence on the items. Accordingly, identifying items that attract users to purchase a target item could be helpful for establishing appropriate marketing strategy and maximizing the revenue or sales. Previously, the association rule mining used to extract latent item-item relationships from transaction logs [1], but it focuses only on typical transaction logs which have the limited information only compared to the session logs of the e-commerce sites. By incorporating item-item similarities, recommender systems have been introduced to identify items which will be preferred by individual users [2]. From the perspective of the user-level influences on the social graph, there have been many trials to find the most influential users who could be selected for viral marketing, which is also called as influence maximization (IM) problem [4, 3].

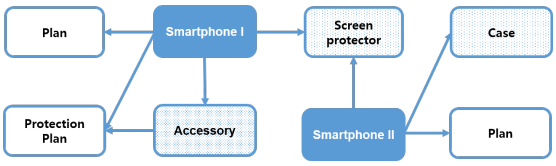

Figure 1: An example of item graph (The background pattern represents an item category)

However, to the best of our knowledge, among user action history, mining purchase influence between items has not been fully explored. In this work, we assume that the item-item relationships can be extracted from user action logs on items, and they indicate purchase influence between items. More specifically, we assume that given a pair of the visited item v and the purchase item p by a user, v has influence on purchasing p. For example, if a user is interested in purchasing a game console thus visited the description page of the game console, (s)he might be interested in purchasing its game titles or accessories such as a game pad, a storage device, and even a huge TV, and thus has visited their description pages as well. Compared to this user, if someone visited the description page of the game console but has not visited the other products' description pages which are highly related to the game console, (s)he may not purchase the game console in near future. Based on this intuition, we formalize the purchase influence mining that finds the k items (PIM-items) that maximizes the purchase influence from them to a query item. To solve this problem, we first construct an item graph by considering the visit-purchase relationship between items as their edges on the graph[1]. Then, we model the purchase influence from item to item as the shortest path between them. In experiments, we evaluate our approach in terms of the consistency, and the result shows that our approach more consistently find the k PIM-items than the baseline.

2. METHOD AND MATERIAL

Dataset: We collected session logs of a real-world e-commerce site, in which each session log includes user action history on items. Based on the dataset, for each user, we generated a list of the visited item-purchased item pair, (v_i, v_j), where v_i was an item visited by a user before (s)he

[1]Fig. 1 shows a part of the constructed item graph

WWW'16 Companion, April 11–15, 2016, Montréal, Québec, Canada.
ACM 978-1-4503-4144-8/16/04.
http://dx.doi.org/10.1145/2872518.2889364 .

bought v_j. From one month session logs, we gathered the dataset of 311K users and 2K items.

Construction of Item Graph: An item graph is represented as a directed graph $G(V, E, W)$ where nodes V are items, edges E are connections between items and edge weights W are the probabilities that nodes' influence the purchase of another item. The weights W can be defined based on the user action logs on item pairs. In this work, we define the weight of an edge (v_i, v_j) by aggregating all session logs of purchasing v_j. Formally, among a set of users who bought v_j, U_{v_j}, where $|U_{v_j}| = n_{v_j}$, the weight can be computed as the number of users who visit v_i over n_{v_j}.

We additionally extracted category information of the items such as phone or accessory. Then, we set one constraint on item pairs that items in the same category cannot have an edge. Intuitively, customers will not buy two phones within a short time period. Although we have to more precisely consider the category-based relationships among items, in this work we simplify the task since the primary goal is to gain insight of the purchase intention from the item graph.

Shortest Path Model for Purchase Influence Mining: With the item graph $G(V, E, W)$, the purchase influence mining can be formalized as follows:

$$arg \max_{T \in V, |T|=k} F(x, T, G) \quad (1)$$

where $F(x, T, G)$ is a function of a node set that provides the estimated purchase probability of the target item node x from a seed item node set T in the item graph G. Then, $F(x, T, G) = \sum_{t \in T} p(x|t, E_{(x,t)})$ where $E_{(x,t)}$ is a subset of edge set $E \in G$ can be manually defined to represent the purchase influence from t to x. We model this probability as the shortest path between x and t on the graph G, $p(x|t, E_{(x,t)}) = \prod_{(v_i,v_j) \in E_{(x,t)}} p(v_i|v_j)$ where $E_{(x,t)}$ is the edge set of the shortest path connecting x and t, and $p(v_i|v_j)$ is the edge weight computed as previously described. That is, given the target item x and the number of the items k, the PIM-items are the k items that maximize the total probabilities of $p(x|t, E_{(x,t)})$.

3. EVALUATION RESULT

To evaluate the proposed approach, we experimentally validate the consistency of identifying the k PIM-items for query items. To do this, we divide the session logs into two disjoint sets and construct two item graphs according to them, respectively. Then, for each graph, we apply our approach to model the purchase influence and extract the k PIM-items for each target item. The consistency is computed as the average number of co-occurred items in the two PIM-item sets. As a baseline, for each query item, we extract the PIM-items based on the co-purchased frequency. That is, the baseline selects k items that are the most frequently purchased with the target item by the same user.

Our method more consistently extracts the PIM-items than the baseline method (Fig. 2). Specifically, the average number of the PIM-items which are consistently identified is 2.7102, 5.4134, and 10.7939 out of 5, 10, 20 PIM-items. In contrast, the baseline method finds 1.3048, 2.9366, and 7.6579 PIM-items. It indicates that the visit-purchase relationships between items are more meaningful than the co-purchase relationships, and the proposed approaches are useful to consistently extract the PIM-items from the item graph constructed based on session logs. As examples of the

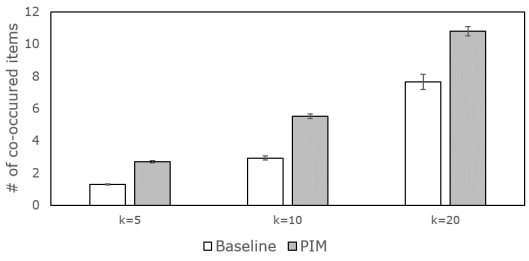

Figure 2: Consistency of the number of co-occurred PIM items (The error bar represents standard error)

PIM-items of mobile devices, there are equipment protection plan, a screen protector, a smartphone case and so on.

4. CONCLUSION AND FUTURE WORK

In this work, we newly introduce the purchase influence mining based on the item graph, where nodes are items and edges are latent purchase influence between items. Based on the real-world dataset, we construct the item graph based on the visit-purchase relationships between items, and model the purchase influence as the shortest path between item pairs on the graph. According to the result, our approach more consistently finds the PIM-items than the baseline.

Similar to the associaion analysis, the PIM-items could be the basis for business decisions about marketing activities like product placements or pricing strategy in market place. However, the association analysis is based on typical transaction logs that generally contain limited information of the item-relationships, however we could extract various relationships between items from session logs according to user actions. Thus, for future work, we have a plan to suggest more complex models to capture and estimate diverse item-item relationships such as visit-visit, visit-purchase, purchase-purchase and so on. In addition, we will apply our method to datasets that have more category information. By doing this, we can mine relationships between not only specific items, but also item categories. In this work, the purchase influence mining handles a single target item only. However, intuitively we could think the same problem by targeting a set of items, not a single item. Thus, we will formalize the purchase influence mining with large k and find solutions to solve it.

5. REFERENCES

[1] R. Agrawal, T. Imieliński, and A. Swami. Mining association rules between sets of items in large databases. *SIGMOD Rec.*, 22(2):207–216, June 1993.

[2] R. M. Bell and Y. Koren. Lessons from the netflix prize challenge. *SIGKDD Explor. Newsl.*, 9(2):75–79, 2007.

[3] S. Bhagat, A. Goyal, and L. V. Lakshmanan. Maximizing product adoption in social networks. WSDM '12, pages 603–612, 2012.

[4] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. KDD '03, pages 137–146, 2003.