Electronic Commerce Research and Applications 9 (2010) 29–37

Contents lists available at ScienceDirect

Electronic Commerce Research and Applications

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

Inferring preference correlations from social networks

Tad Hogg

Hewlett-Packard Laboratories, MS 1139, 1501 Page Mill Road, Palo Alto, CA 94304, United States

ARTICLE INFO

Article history:
Received 28 May 2008
Received in revised form 3 April 2009
Accepted 7 April 2009
Available online 22 April 2009

Keywords:
Consumer preferences
Social network
Homophily
Network topology

ABSTRACT

Identifying consumer preferences is a key challenge in customizing electronic commerce sites to individual users. The increasing availability of online social networks provides one approach to this problem: people linked in these networks often share preferences, allowing inference of interest in products based on knowledge of a consumer's network neighbors and their interests. This paper evaluates the benefits of inference from online social networks in two contexts: a random graph model and a web site allowing people to both express preferences and form distinct social and preference links. We determine conditions on network topology and preference correlations leading to extended clusters of people with similar interests. Knowledge of when such clusters occur improves the usefulness of social network-based inference for identifying products likely to interest consumers based on information from a few people in the network. Such estimates could help sellers design customized bundles of products and improve combinatorial auctions for complementary products.

1. Introduction

In social networks, people with similar characteristics tend to associate (McPherson et al. 2001). This property allows vendors to use available online social networks to plan targeted marketing via offline word-of-mouth interactions (Domingos and Richardson 2001). The similarity also enables social network administrators to improve collaborative filtering by weighting user preference correlations based on distance in their social network (Lam 2004, Ben-Shimon et al. 2007, Woerndl and Groh 2007), and enhance online reputation mechanisms (Hogg and Adamic 2004, Swamynathan et al. Wang and Chiu 2008) such as the one used on eBay (Resnick et al. 2006).

Explicit or estimated knowledge of these networks (Garton et al. 1997) can also help vendors infer products or bundles of products likely to interest specific individuals based on observed behavior of a few people in the network. In particular, learning the preferences of one person allows inference of similar preferences for others nearby in the network (McPherson et al. 2001), that is, people who are connected via a small number of links in the network. This use of social network structure allows:

- targeting a single product to members of the social network based on observations of others in the network, a form of collaborative filtering (Lam 2004),
- identifying a bundle of products of likely interest to the group based on observations of interest in the bundle as a whole by some members of the network,
- suggesting possible bundles for further evaluation from observed preferences for individual products from several people who are close together in the network, and
- devising a product bundle for the collective consumption by the group, generalizing the concept of a bundle of products of interest to a single person.

Most discussion of product bundles focuses on purchases by individual consumers, primarily for complementary products. However, the last application in this list, identifying bundles for collective consumption, is especially beneficial for situations with allocative externality (Jehiel and Moldovanu 2005, Ranger 2005), where the value of an item to an individual depends on what items others in the network have. An example is a vacation with friends having higher value for people in the network than each person taking the same vacation separately. The large datasets of shared preferences from online networks enable vendors to identify such groups with low cost, even when such situations are relatively rare.

Product bundles are usually created by the sellers, based on their expectations of how consumers value different combinations and the incremental cost of adding items to a bundle. With information goods, the marginal costs of additional items can be quite low (aside from possible licensing or royalty costs). Sellers will usually choose to offer only a few of the exponentially many possible bundles in the hope that these will be sufficient to allow price discrimination among consumers (e.g., between casual and dedicated users of a set of software tools).

If consumer preferences are sufficiently diverse or change rapidly, a fixed set of bundles will not efficiently capture the market

E-mail address: tad.hogg@hp.com

1567-4223/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.elerap.2009.04.006

potential. For example, the bundles may be priced too high and thereby eliminate casual users from the market. On the other hand, low pricing may obtain high market share but may miss the potential revenue from dedicated users' higher values.

An alternative is to allow individual consumers to select bundles of interest. One approach to identifying high-value product bundles is a combinatorial auction (Cramton et al. 2006). These auctions accept bids on arbitrary bundles of the products for sale, thereby allowing bidders to express valuations for complementary and substitutable products (i.e., cases in which the value of the bundle is either greater or less than the sum of the individual item values). Even losing bids reveal consumer bundle preferences (Huberman et al. 2001). Unfortunately, combinatorial auctions involve considerable overhead to arrange and computational cost to run due to the difficulty of evaluating preferences over a potentially exponentially large number of combinations of bundles. Thus using combinatorial auctions to determine bundle preferences for large numbers of consumers and items can be prohibitive. An intermediate scenario is to allow consumers to select items with a simple bundle pricing scheme (e.g., a fixed price for any ten songs from a large set of songs). In this scheme of customized bundles (Wu et al. 2008), each consumer gains flexibility to select according to their individual interests, but without the complex overhead of adjusting prices for arbitrary bundles through a combinatorial auction.

Inferring joint preferences from available social networks is a low-cost method of estimating preferences, though less specific in eliciting preferences than combinatorial auctions. Thus network-based inference could be especially useful to identify situations where deploying a combinatorial auction could be worthwhile, and others where simpler bundling schemes likely suffice to capture much of the value. This means finding situations where there appear to be considerable correlation among preferences for items among niche subgroups of consumers. This inference relies on a correlation between a pair of users having a link and the similarity of their interests. Such correlations arise because people with similar interests tend to form links, or conversely because people who are linked influence the interests of each other. In either case, the observation of a link correlates with similar interests. This correlation is sufficient for the inference procedure described in this paper, which does not utilize more detailed models, such as causal mechanisms, leading to the observed correlations.

In the remainder of this paper, we describe methods for identifying products and bundles likely to interest particular individuals in the context of an available social network. We then illustrate the benefits of using preference correlations with a simple model of social networks, showing how the usefulness depends on the structure of the underlying graph and preference correlation in the graph. Finally, we examine the influence of link semantics using an online system allowing people to both express preferences and explicitly form links with differing nominal semantics. This example illustrates the significance for e-commerce applications of eliciting link semantics from users of social networks.

2. Identifying preference correlations

Online activities, such as search and web site visits, reveal users' interests in a readily recordable form. When aggregated over sites related to multiple products, the activity records can indicate possible interest in bundles of these goods or services.

Consider a data set of visits by many people to web sites relevant to various products or services. This data can include clicks on links, search keywords and time spent with various product web pages. One approach to identifying potential product bundles is through correlations in observed behaviors on sites relevant to those products. A variety of statistical techniques could identify significant correlations (Rossi and Allenby 2003). One example is modeling the effectiveness of search keyword advertising (Ghose and Yang 2007, Rutz and Bucklin 2007). A simple method looks for statistically significant deviations from expected values among all users, combined with any available prior knowledge of bundles purchased by others. Setting a detection threshold based on the observed distribution of user visits allows further detailed investigation to focus on cases with potential complementary interests to the user. This method generalizes to correlations among many web sites. Furthermore, instead of treating each site individually, we could group sites into various types (e.g., those involving automobiles or travel) and then look for correlations among two or more types that are significantly greater than would be expected for independent choices.

Using correlations to identify potential product bundles relies on probing individual behavior. This is a relatively slow process, requiring extensive records tracking single individuals as they use various web sites or search terms. Collaborative filtering addresses this limitation by inferring correlations based on other people with similar observed behaviors (Goldberg et al. 1992, Huang et al. 2007), and can improve vendor performance (Chen and Wu 2007, Garfinkel et al. 2006).

A key aspect of collaborative filtering is selecting which other users' behaviors to use to infer preferences of a particular user. Social networks directly created by users can provide strong correlations in addition to or instead of inferring correlation from commonly observed behaviors (Lam 2004, Ben-Shimon et al. 2007, Woerndl and Groh 2007). That is, the homophily of social networks reflects people with similar interests tending to link to each other (McPherson et al. 2001). Provided the links indeed reflect common interests relevant to a particular vendor or class of products, available online social networks can improve on inferring links based on behavior.

We can use these correlations in two ways. First, once a bundle of interest is identified for one person, we can infer that others in the network neighborhood are more likely than average to also have that preference. Second, people close together in a network who express interest in different products may also be more likely than average to have an interest in those products as a bundle. This inference, while not as strong as actually observing a bundle preference for each of the individuals, could substantially expand the usefulness of the network inferences. Specifically, instead of waiting for a few people to directly express their bundle preferences we can use the combined behavior of several individuals known to be relatively close in the network.

As vendors gain experience with inference accuracy, they can identify more specific measures related to the actual value of the bundles to consumers, such as actual purchase histories. This procedure allows vendors, over time, to gain much of the same aggregate information as would be revealed through a series of combinatorial auctions, but without the cost of creating and running the auctions. A key question examined in the remainder of this paper is how network topology affects the usefulness of network-based inference, particularly the size of connected groups likely to have similar preferences.

3. Network inference performance

To illustrate the use of inference based on a social network, we consider a simple model as a theoretical benchmark for the empirical investigation discussed in Section 4. The social network among n people corresponds to a graph with n nodes. For simplicity, we consider undirected edges representing links between two people,