Using Abstract Attributes for Capturing Item Relationships and Making Product Recommendations

Yash Talreja

Introduction

Capturing relationships among items and using those relationships for recommending related products is an important aspect of any retail business.

Displaying related items together not only enriches a customer's buying experience, it also increases incremental sales for the retailer or the e-commerce marketplace operator.

Most e-commerce systems capture item relationships primarily using two techniques:

1. Explicit capturing of item relationships at each individual stock-keeping unit (SKU) level, which needs to be specified by the merchandisers (for an e-tailer) or a merchant (for a marketplace)

2. Statistical correlation based on purchase history

In the first technique, related items are identified and their relationship is specified at the level of individual items.

This is usually done by the staff working in the merchandising department of a retailer or a seller in a marketplace such as eBay or Amazon.com.

Since explicit identification of item relationships needs be manually done, it is resource intensive and does not scale. It does not work well for retailers that are rapidly expanding their product selections, adding new products and product categories, since any time a new product is added, there will be a need to identify its item relationships with existing products.

This becomes more laborious as the number of related products increase. For instance, a store that has 200 brands of laser printers and 100 brands of printer paper will, as a rule, need to capture all 20,000 relationships explicitly.

The second technique, statistical correlation, is highly automated and is extensively used by retailers for recommending related items ("Customers who bought this book also bought . . .").

Statistical correlation works very well for correlating items that have ample purchase data and have significant velocity. However, this method still fails to help jumpstart a new product category by associating the new products with existing products. Also, it's not every effective at recommending related products that have higher margins but low velocity—for example, recommending a wedding ring to someone buying a book on wedding planning.

What we are proposing is a new approach for recommending products based on abstract attributes. These abstract attributes will be associated with each item by the merchandiser or the seller. Then, the relationships between items will be automatically inferred based upon these abstract attributes.

The abstract attributes will form a knowledge framework, which can be represented in a "thesaurus"—a semantic

Email: info@the-technology-gurus.com
Phone: 650 339 6240

framework that captures generic item relationships, their common uses, the various classes of users, and other item-merchandising properties.

Abstract attributes can greatly improve cross-selling opportunities for new items added to the catalog, thus overcoming the chicken-and-egg-like situation caused by statistical recommendation: the new items do not have any purchase history, so they are never recommended, and so, they never build up enough correlation with related items.

Automatically identifying related items using abstract attributes can accelerate adoption of newer categories and provide a way to narrow down a large set of candidate items for personalized recommendations.

Here are some examples that illustrate how this can be achieved:

1. There are twenty inkjet printers in a catalog, and a new brand of inkjet printer paper is introduced as a new item.

   In the traditional way of cross-linking items, an accessory relationship would be manually created between the paper and the existing printers in some sort of “item master table” at the individual SKU level.

   In contrast, an abstract-attribute-based system would automatically determine the supplies relationship between the paper and all the inkjet printers.

   This system can automatically display the paper on the website as a supply item along with all the inkjet printers. It can also remember when a customer buys an inkjet printer from us; it will display the inkjet paper to them on their next visit.

   The system will be able to do this even if the paper is a brand new item and has no purchase data.

   Other rules can be specified to govern and fine tune the display of the paper or other accessories.

2. A food retailer is adding a new category to his wares: Gourmet Food. Since this is a new category, none of the items in this category will have a purchase history.

   Relationships derived from abstract attributes can be used for recommending, say, Chinese gourmet food items to people who have purchased books that were written by Chinese authors, or are about China, or were written in Chinese.

3. Inference can also help us cross-sell items that have low velocity but high margins.

   For example, a retailer that sells both books and jewelry could cross-sell expensive wedding jewelry to people looking at wedding-planning books.

Email: info@the-technology-gurus.com
Phone: 650 339 6240

4. Abstract attributes do not have to be used in isolation—they can also be used in conjunction with statistical correlation for personalization/recommendations.

   Based on the prior purchase history of our customers' accounts, we can infer what abstract themes they are interested in and which target audience they usually buy items for. We can use this knowledge to refine the list of suggested items produced by statistical correlation.

The second section, "Item Relationships: From Physical to Abstract," describes how item relationships based on abstract attributes can greatly expand the traditional item relationships based on physical attributes—color, shape, print, size—which needed be explicitly defined at the SKU level for each SKU.

In the third section, "Capturing Abstract Relationships," we give more details on how to capture and use abstract attributes for product recommendations and improving customer experience.

Item Relationships: From Physical to Abstract

The goal of this section is to describe item relationships in a generalized manner, independent of the tools and technologies used by the retailer or e-commerce system to implement them. We also introduce the notion of abstract attributes and abstract relationships, which can be used for making product recommendations and cross-selling.

Item relationships vary across a wide spectrum, from coupling items based on origin or physical attributes, to broader abstract relationships.

At one end of the spectrum, we have items whose relationships are based on physical attributes such as color, size, or style—this includes variations (e.g., the same item in different colors and sizes) or matching pieces of furniture, bed sheets, or apparel.

Here the related items share one or more physical attributes.

Similarly, items with the same origin—such as books by the same author, songs by the same singer, items with the same brand name or by the same manufacturer, or clothing by the same designer—share origin attributes.

Then, there are relationships based on functionality, such as base-accessory or consumer-supplies.

An *accessory* is defined as "a subordinate or supplementary item." Accessories have life spans similar to the base item; for example, a person who buys a PC is likely to buy one or two printers during the lifetime of the PC.

Other examples of accessories are matching shoes and jewelry, and cellular phone batteries.

The consumer-supply relationship is similar to the base-accessory relationship, with the difference that supplies are consumed and need to be replenished on a regular basis—for example, a printer needs a regular supply of printer paper—while accessories are not.

We now introduce the notion of *abstract relationships*, which can be implemented by introducing abstract attributes to item definitions. These relationships are based on nonphysical, abstract similarities such as a common theme, audience, or purpose.

Email: info@the-technology-gurus.com
Phone: 650 339 6240

the TECHNOLOGY Gurus

Here are some examples of items with abstract relationships between them:

- a golf ball, a golf-instruction video, and a book about golf. All of these are related to abstract notion of *golf*.
- a mini backpack, a sleek cellular phone, a pink iPod, a rhinestone belly button ring, and a pair of low-rise jeans—all popular among teenage girls and other young females
- a Japanese cookbook, shoji screens, kimonos, and a Japanese DVD

So by associating abstract attributes with items, one could capture abstract relationships such as having a common theme, targeting a common audience, or being meant to be used together. New products added to the item catalog would automatically get associated with existing items and get recommended. The items could belong to different product categories and could be made by different manufacturers.

Capturing Abstract Relationships

Abstract item relationships can be inferred by associating abstract attributes with each item and using a semantic framework. The semantic framework is represented using a thesaurus and rules, which allow us to combine the knowledge represented in the thesaurus with Boolean logic.

A thesaurus captures relationships between items indirectly by providing a controlled vocabulary of abstract concepts such as *golf*, *men*, *toaster*, *shirt*, *Christmas*, and *Seattle Mariners*. Items are mapped to terms from this controlled vocabulary.

The thesaurus contains the relationships among terms, in addition to the terms themselves. The relationships among items can be inferred by tracking the relationships between terms associated with the items.

The contexts in which the terms are mapped to an item also need to be tracked: the term *children*, for instance, has one associating context in “books *for* children” and another in “books *about* children.”

Some examples of abstract item attributes:

- target user
- usage
- type
- “about”

The terms used for abstract item attributes are meaningful entities themselves. Furthermore, the meanings of these terms are independent of any specific product line or merchandising effort.

For example, the term *children* can be used to describe a book, video, toy, sporting goods, and even content. Using a common language to describe all entities breaks down our traditional barriers between product lines, and even between contents, features, and items.

When people are associating terms with an item, they are not doing it for a specific application. That knowledge becomes available to all kinds of possible applications, including future applications. So the person associating an item with the concept *children* does not need to know whether the marketplace operator intends to use the data to create the browse tree, cross-sell, or make personalized product recommendations.

Email: info@the-technology-gurus.com
Phone: 650 339 6240

the TECHNOLOGY Gurus

Summary

There are two popular methods for cross-selling and recommending products to customers in e-commerce systems: (1) specifying and identifying item relationships at SKU levels, and (2) using statistical correlation.

Specifying and identifying item relationships at SKU levels can be resource intensive. It also does not provide a mechanism to cover relationships with future items, since relationships need to be specified after the items have already been created in the catalog.

Statistical correlation provides us with a very powerful and highly scalable way for identifying related items. However, it relies upon the existence of prior purchase data, so it is not useful for new items added to the catalog or for new products categories. It cannot be used for cross-selling low velocity items with higher margins either.

Using abstract attributes and relationships based on abstract concepts such as type, usage, target user, and "about" provide a scalable and flexible way for improving our current capabilities for cross-selling, personalization, and recommendation.

The automatic inference based on abstract attributes eliminates the manual work needed for identifying relationships at Amazon Standard Item Number (ASIN) level.

It can be especially useful for identifying cross-selling opportunities for new items, accelerating the adoption of newer categories, capturing broad relationships, and providing a way to narrow down a large set of candidate items for personalized recommendations.

About Us

The Technology Gurus is a boutique consulting company that assists technology companies with their interim CTO, software development and outsourcing management, and M&A-related technology-due-diligence needs.

To find out more about our services, visit us at www.The-Technology-Gurus.com or call us at 650-339-6240.

Yash Talreja is the principal consultant at The Technology Gurus. Yash has more than 17 years of senior/executive management-level experience at both large companies such as Oracle, PeopleSoft, and Amazon.com and startups such as Televoke and MessageGate.

Yash is also the founder of the Lean Agile Group, an organization of more than 4,500 practitioners of Lean/Agile software development methodologies.

Yash can be reached at yash@the-tech-gurus.com.

Email: info@the-technology-gurus.com
Phone: 650 339 6240