A Text-Based Framework for Dynamic Shopping-Cart Analysis

Kamakura, Wagner
Rice University, USA.

Abstract

Market Basket Analysis (MBA), also known as Association Rule Mining (ARM), is already widely known and utilized by traditional and online retailers. This practice has its origin in the data-mining literature, with the introduction of association-rule mining. Despite its popularity, MBA/ARM has been criticized for its assumption that joint occurrence implies complementarity. Moreover, despite being further optimized for large-scale implementation, MBA/ARM suffers from a "curse of dimensionality," with the problem size and data sparsity growing with the square of available items. A common solution is to first classify items into pre-defined categories and carry the analysis at the category level, considerably reducing problem size. However, this simplification is prone to problems because all items within each category are automatically assumed as perfect substitutes, and categories must be mutually-exclusive, preventing an item (e.g., almonds) from belonging to more than one category (snacks, baking goods and/or bulk sales).

The main purpose of our study is to incorporate a longitudinal component into Market Basket Analysis, looking at the sequential formation of the basket, rather than its final composition only, while also reducing the dimensionality of the problem down from the number of SKU's to the most common descriptors of a shopping cart, through the text-mining of all SKU descriptors. We demonstrate empirically how dynamic MBA provides valuable insights into how the purchase of one product leads to the purchase of another, which cannot always be properly inferred from the final basket compositions in traditional MBA. Given that sequential data on shopping-cart formation is now widely available to online retailers, there is no good reason for overlooking the additional insights embedded in purchase sequences.

Keywords: *Market Basket Analysis; Dynamic Shopping Cart Analysis; Text-Mining; Hidden Markov.*