AFFORDANCE-BASED INTERESTINGNESS IN AUGMENTEDASSOCIATION RULE MINING

RUPAL SETHI

Abstract

With vast amounts of data available, commercial organizations are orienting themselves towards developing strategies to identify unusual and "interesting" purchase patterns. These interesting patterns may be put to action and used as a source of competitive advantage. Thus, there is a need for an intelligent system that identifies hidden patterns in huge data there by empowering managers to make more effective decisions. Resulting knowledge discovery from an intelligent system may be classified as additional new knowledge (novelty) for the manager or knowledge contrary to his prior beliefs (unexpectedness). This notion of interestingness in pattern recognition literature has led to the development of objective as well as subjective measures. However most of these interestingness measures (Geng & Hamilton, 2006) result only in finding interesting associations between items. They do not go beyond the occurrence of such interesting patterns to address plausible reasons behind product purchase. Thus, there is a need to shift the investigation from "What products do customers buy together?" to "Why do customers buy certain products together?". The overarching objective of this dissertation is to address issues related to the second question. Our approach to subjective interestingness compares manager's expectations and customer perceptions through the lens of affordance (Gibson, 1977). In the first essay, we plan to study interestingness through a multi-stage affordance model comprising object (product), user (customer) and observer (manager). We add observer's perspective to the formalization ofaffordance (Chemero, 2003). Our intention is to study the deviation of a manager's (observer's) expectations of purchase patterns from customers' (users') actual purchase behavior. The research question that we address in this essay is: What is the nature of subjective interestingness (among associations of items) in the context of manager's expectations and customers' actual purchase patterns? We address this by studying four new facets of subjective interestingness: conformance, substitutability, unexpectedness and novelty. We propose an affordance-based association rule (AAR) and proceed to develop a Product-Affordance framework for AARs. We also evolve measures for each of the facets. These measures are used in ranking and clustering of AARs. Information provided by the measures help addressing plausible reasons behind product purchase. Thus the framework has potential to enhance the quality of discovered knowledge and thus support managerial learning and decision making.

We validate the proposed Product-Affordance framework and the measures corresponding to the facets of interestingness through a full-fledged application involving products purchased in a supermarket. The methodology adopted is a mixed-method approach comprising analytical/empirical analysis and qualitative analysis. Qualitative analysis is beneficial in tapping affordances associated with purchase of products. Since affordance is a relational concept, each customer may have a distinct set of affordance for the same object based on intended goals or applications (Leonardi, 2013). Qualitative analysis is also be required for tapping manager's expectations about customer purchase patterns in the form of beliefs.

Pattern Recognition and Association Rule Mining literature has so far focused on generating and pruning positive associations using various interestingness measures. However there are very few studies that explore the mining process of indirect association rules. Indirect

associations pertain to two items that are indirectly related to each other through a third item (Tan, Kumar & Srivastava, 2000). These associations are important since they connect seemingly unrelated items present in disjoint transactions which cannot be captured through positive ARs. AR Mining literature is limited to studying indirect associations in two categories: substitution rules and negative association rules. Previous research studies have incorporated a limited definition (either in statistical terms or on the basis of manager's static knowledge) of substitution and negative rules. In the second essay, we provide a customer centric model of substitution rule mining using the lens of affordance. We investigate the research question: How can affordance-based subjective interestingness be used to understand and mine indirect relations (substitution and negative) between items? We address this by adopting graph-based machine learning. We propose an approach based on dynamic ontology wherein items are positioned based on the affordances they are purchased for. This arrangement contrasts with the traditional static taxonomy that highlights manager's static knowledge of product classification. For example, a hose pipe and a water bottle can be considered siblings based on affordances leading to providing water for plants. However in a static taxonomy these two products may be far from each other. Thus dynamic ontology would help improve the quality of a manager's knowledge by observing changing customer purchase patterns on the fly. This could in turn lead to uncovering different kinds of substitution relations and negative associations.

In the third essay, we develop a Product-Affordance (P-A) network drawing parallels from the literature on Social Network Analysis (SNA). We address the research question: What are the features of interestingness when products and affordances are combined together as social networks? The concentration is on three aspects of P-A networks – centrality, brokerage and community detection. Network analyses are performed on the one-mode projection of the P-A network. Centrality measures aid in determining products that are powerful and central in the network. This knowledge is essential to get an idea of products that dominate the flow of affordance information. We also study brokerage (Burt, 2002) between product networks. We define "brokerage products" as those linking two disjoint networks or cliques of products. For example, paper cups can serve as a brokerage item between networks of soft drinks (Soda, Coke, Pepsi) and a network having milk and cereals. Community Detection (Raeder & Chawla, 2011) will help finding "interesting communities" of products that are linked with similar or dissimilar affordances. These three aspects of P-A networks have the potential to provide better managerial insights into understanding customer purchase patterns.