

Warehousing and Mining Massive RFID Data Sets

Jiawei Han

Department of Computer Science
University of Illinois at Urbana-Champaign

Abstract

Radio Frequency Identification (RFID) applications are set to play an essential role in object tracking and supply chain management systems. In the near future, it is expected that every major retailer will use RFID systems to track the movement of products from suppliers to warehouses, store backrooms, and eventually to points of sale. The volume of information generated by such systems can be enormous as each individual item (a pallet, a case, or an SKU) will leave a trail of data as it moves through different locations. As a departure from the traditional data cube, we propose a new RFID data warehouse model that preserves object transitions while providing significant compression and path-dependent aggregates, based on the following observations: (1) items usually move together in large groups through early stages in the system (e.g., distribution centers) and only in later stages (e.g., stores) do they move in smaller groups, and (2) although RFID data is registered at the primitive level, data analysis usually takes place at a higher abstraction level. Techniques for summarizing data, query processing in RFID data warehouse, RFID flow-cube construction, and data mining based on this framework are developed. We also illustrate a few promising research topics for mining such massive RFID data warehouses.