RFID Data Mining: Opportunities and Challenges

Joshua Huang
E-Business Technology Institute
The University of Hong Kong

25 August 2007
WSIETA2007
Chengdu, Sichuan, China
Agenda

• Overview of RFID Technology and Applications
• RFID in Logistics and Supply Chain Management
• Research Areas
  – Data Collection and Processing
  – Data Management
  – Event Processing
  – Data Mining
• Summary
RFID Technology Components

- **RFID Tag**
- **RF Antenna**
- **RFID Reader**
- **Network**
- **Workstation**
Tags and Readers

- **Passive Tags**
  - Without battery.
- **Active Tags**
  - With battery

**Readers**

- "Quietized" Fully Enclosed Conveyor Tunnel
- Symbol Reader
RFID middleware

Source: Forrester Research: RFID Middleware
RFID applications

RFID Revenues by Application Market

2006 = $2.3 Billion

- Toll
- Ticketing
- Supply Chain
- Shop-Floor
- Sensing
- Security
- RTLS
- Retail POS
- Rental Tracking
- Auto Immobilizer
- Baggage Handling
- Asset Tracking
- Animal Tracking

Other

354.6

137.5

438.3

49.3

454.0

98.6

45.8

35.6

148.4

24.5

362.9

100.8

Smart Shelves
Library Example

Book search and inventory

Security control and check out
Bridging Physical and Digital Worlds

![Image of warehouse management system](image-url)

---

**Warehouse Management**

<table>
<thead>
<tr>
<th>Location</th>
<th>Description</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
<th>Status</th>
<th>Qty</th>
</tr>
</thead>
<tbody>
<tr>
<td>1A Top</td>
<td>003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1A Bottom</td>
<td>002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1B Top</td>
<td>004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1B Bottom</td>
<td>005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

![Image of warehouse stacks](image-url)
Agenda

• Overview of RFID Technology and Applications
• RFID in Logistics and Supply Chain Management
• Research Areas
  – Data Collection and Processing
  – Data Management
  – Event Processing
  – Data Mining
• Summary
Logistics and Supply Chain Management

• **Logistics** - ...the process of planning, implementing, and controlling the efficient, effective flow and storage of goods, services, and related information from point of origin to point of consumption for the purpose of conforming to customer requirements. -- (Council of Logistics Management, http://www.clm1.org/mission.html, 12 Feb 98)

• **Supply Chain Management (SCM)** - ...the oversight of materials, information, and finances moving in a process from supplier to manufacturer to wholesaler to retailer to consumer. SCM involves coordinating and integrating these flows both within and among companies.
Logistics and Supply Chain

• **Supply Chain Model**

  - Requirements information flow
  - Fulfillments information flow

  - Customers
  - Physical Distribution
  - Manufacturing
  - Purchasing
  - Suppliers

• **Supply types**
  - One-origin/one-destination
  - One-origin/multiple-destinations
  - Multiple-origins/one-destination
  - Multiple-origins/multiple-destinations
Business Objectives

- Cost reduction
- Inventory control
- Reduce lead time
- Supply chain optimization

Problems
- Mismatch between physical inventory and inventory database
- Lack of information sharing for entire supply chain optimization
RFID Applications in SCM

• Bridge the gaps between physical inventory and digital inventory
• Real time data collection to update inventory databases
• Enable information sharing, e.g., through EPC network
• Real time product tracking and positioning
• Enable global supply chain optimization
**Food Supply Chain Example**

1. Radio tag placed on carton.
2. Dairy ships carton to grocery store.
3. Consumer purchases tagged carton.
5. Carton arrives at recycling center. Manufacturer produces replacement.

A. Manufacturer tracks product through wireless radio communication.

*Products are tracked through their entire lifetime*

Source: How Stuff Works
Challenges and Research Opportunities

• RFID enabled supply chain operation will generate massive data that needs to be stored, managed and analyzed

• **How much data?**
  – Consider a supermarket chain implementing RFID:
  – 12 bytes EPC + Reader ID + Time = 18 bytes per tag
  – Average number of tags in a neighborhood store = 700,000
  – Data generated per second = 12.6 GB
  – Data generated per day = 544 TB
  – Assuming 50 stores in the chain,
    • *data generated per day = 2720 TB*
• Overview of RFID Technology and Applications
• RFID in Logistics and Supply Chain Management
• Research Areas
  – Data Collection and Processing
  – Data Management
  – Event Processing
  – Data Mining
• Summary
Research Areas

Data Mining

Event Processing

Data Management

Data Collection and Processing
Agenda

• Overview of RFID Technology and Applications
• RFID in Logistics and Supply Chain Management
• Research Areas
  – Data Collection and Processing
  – Data Management
  – Event Processing
  – Data Mining
• Summary
RFID Data Collection and Preprocessing

• Data collection problems
  – False readings, i.e., read unexpected tags
  – Duplicate readings, e.g., same reader reading many times, multiple readers reading the same tag
  – Missed readings, i.e., tags undetected due to RF interference, or malfunctions of tags or readers
  – High speed and large volume, e.g., many tags present to many readers

• Solutions
  – Multiple cycles reading to reduce missed readings
  – Data filtering and cleansing to remove duplicates and false readings
  – Data abstraction and compression
Cleansed RFID Data Records

- Raw Data
  - (EPC, location, time)
  - Duplicate records due to multiple readings of a product at the same location
  - \((r_1, l_1, t_1) \ (r_1, l_1, t_2) \ldots (r_1, l_1, t_{10})\)

- Cleansed Data: Minimal information to store and removal of raw data
  - (EPC, Location, time_in, time_out)
  - \((r_1, l_1, t_1, t_{10})\)

- Fill-up missing records and correct wrongly-registered information through multiple data collection points (e.g., manufacturing process, different storages at large distribution centers)
Agenda

• Overview of RFID Technology and Applications
• RFID in Logistics and Supply Chain Management
• Research Areas
  – Data Collection and Processing
  – Data Management
  – Event Processing
  – Data Mining
• Summary
Object Identification

• Identify an object at a specific time and a specific location
  – Three essential data elements:
    • Object identification, location and time: (ID, Loc, T)
  – Data of other attributes of identified objects are stored in a database and can be matched through ID

• A lot of ID data are collected from readers in the business process
Electronic Product Code
96 bits Standard

With 96 bit code, 268 million companies can each categorize 16 million different products where each product category contains up to 687 billion individual units.
RFID Data Modeling

- Different applications may require different data schema designs
- Data models must support data queries and analysis
- In supply chain management, modeling the process flow is important

www.aeroid.co.uk/rfidbasics.html
Research Problems

• Indexing on EPC code
  – How to build the EPC code index to support query and analysis
• EPC code based fast query processing
  – Query related to different parts of EPC code
• Distributed data management
Agenda

• Overview of RFID Technology and Applications
• RFID in Logistics and Supply Chain Management
• Research Areas
  – Data Collection and Processing
  – Data Management
  – Event Processing
  – Data Mining
• Summary
RFID Event Processing

- Data query and analysis are often event-oriented
- Events are defined in different ways and at different levels
- Event processing generates different events from raw RFID data
Grouping Data Into Different Events with GID

- Objects often move and stay together (e.g., purchase orders, shipments)
- If 1000 packs of soda stay together at the distribution center, register a single record
  - \((GID, \text{distribution center}, \text{time}_{\text{in}}, \text{time}_{\text{out}})\)
- GID is a group identifier that represents the 1000 packs that stayed together at the distribution center
Grouping Table

#EPCs | #GIDs
---|---
n | 1

\{r_1, \ldots, r_i\} \quad \{r_{i+1}, \ldots, r_j\} \quad \{r_{k+1}, \ldots, r_l\} \quad \{r_{l+1}, \ldots, r_m\} \quad \{r_{m+1}, \ldots, r_n\}

3n | 10+n
Different Events Generated for Different Purposes

Store View:

Transportation → backroom → shelf → checkout

dist. center → truck → backroom → shelf → checkout

Transportation View:

dist. center → truck → Store
Agenda

- Overview of RFID Technology and Applications
- RFID in Logistics and Supply Chain Management
- Research Areas
  - Data Collection and Processing
  - Data Management
  - Event Processing
  - Data Mining
- Summary
Data Mining Tasks

- Classification
- Prediction
- Clustering
- Association Analysis
RFID Data Mining Research

- Data mining problems
  - RFID data cleansing
  - Event based data mining
    - Event classification
    - Event prediction
    - Event clustering
    - Event association
  - Outlier event detection and analysis
- New techniques
  - Efficient algorithms
  - Distributed data mining
- Applications
  - Supply chain analysis and optimization
  - Inventory control
  - Supply chain management
Path Event Database:

<table>
<thead>
<tr>
<th>id</th>
<th>product</th>
<th>brand</th>
<th>path</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>tennis</td>
<td>nike</td>
<td>(f,10) (d,2) (t,1) (s,5) (c,0)</td>
</tr>
<tr>
<td>2</td>
<td>tennis</td>
<td>nike</td>
<td>(f,5) (d,2) (t,1) (s,10) (c,0)</td>
</tr>
<tr>
<td>3</td>
<td>sandals</td>
<td>nike</td>
<td>(f,10) (d,1) (t,2) (s,5) (c,0)</td>
</tr>
<tr>
<td>4</td>
<td>shirt</td>
<td>nike</td>
<td>(f,10) (t,1) (s,5) (c,0)</td>
</tr>
<tr>
<td>5</td>
<td>jacket</td>
<td>nike</td>
<td>(f,10) (t,2) (s,5) (c,1)</td>
</tr>
<tr>
<td>6</td>
<td>jacket</td>
<td>nike</td>
<td>(f,10) (t,1) (w,5)</td>
</tr>
<tr>
<td>7</td>
<td>tennis</td>
<td>adidas</td>
<td>(f,5) (d,2) (t,2) (s,20)</td>
</tr>
<tr>
<td>8</td>
<td>tennis</td>
<td>adidas</td>
<td>(f,5) (d,2) (t,3) (s,10) (d,5)</td>
</tr>
</tbody>
</table>

FlowGraph:

- **Duration Dist:**
  - 1: 0.67
  - 2: 0.33
- **Duration Exceptions:**
  - Given (f, 5) 1: 0.0
  - 2: 1.0
  - Given (f, 10) 1: 0.5
  - 2: 0.5

- **Transition Dist:**
  - shelf: 0.67
  - warehouse: 0.33
- **Transition Exceptions:**
  - Given (t, 1) shelf: 0.5
  - warehouse: 0.0
  - Given (t, 2) shelf: 1.0
  - warehouse: 0.0
Event Analysis

- Event analysis at each node
  - Distribution of durations at each node
  - Distribution of transition probabilities
  - Exceptions to duration and transition probabilities

- Path event analysis
  - Clustering and categorization
  - Classification and prediction
  - Outlier analysis
Event Association Analysis

- Frequent patterns and sequential patterns can be related to event movement and paths
  - Correlation analysis of different events
  - Sequential event patterns
Outlier Event Analysis

- Outlier event detection (by-product of event mining)
  - Event flow analysis: Detect those not in the major flows
  - Classification: Treat outliers and normal events as different class labels
  - Cluster analysis: Identify those that are deviate substantially in major clusters
  - Trend analysis: Those not following the major trend
Agenda

• Overview of RFID Technology and Applications
• RFID in Logistics and Supply Chain Management
• Research Areas
  – Data Collection and Processing
  – Data Management
  – Event Processing
  – Data Mining
• Summary
• In general, RFID data mining is still a young, largely unexplored field, therefore, a lot of opportunities.
• RFID data mining has close links with sensor data mining, moving object data mining, work flow data mining and stream data mining (where one can find related work and techniques)
• RFID data mining research should be closely related to applications
Thank You
RFID Enabled Supply Chain

Automatic Verification and Inventory Reconciliation

1. Raw material shipment:
   - Raw material A placed in drum 9
   - Raw material B placed in drum 10
   - Shipped on truck 23

2. Shipment of drugs:
   - Arrived July 11, 10:17 a.m. on truck 25
   - Shipped July 11, 6:45 p.m. on truck 76

3. Cancer Drug 1 stocked on shelves. History of Cancer Drug 1:
   - Arrived July 12, 3:32 p.m.
   - Source: Truck 78 from Wholesaler 2, truck 25 from Manufacturer Y, truck 33 from Chemical Plant X

1. EPC Verification
2. Track
3. Trace

Scan goods and check status:
- Cancer Drug 1: Okay
- Cancer Drug 2: Expired
- Cancer Drug 3: Recalled
- Etc.

Shipments received:
- Raw materials combined, drug created
- Vat placed on vat
- Vat placed into case and tagged
- Case placed on pallet and tagged
- Pallet shipped on truck 25