Computers in Industry 65 (2014) 721–732

Contents lists available at ScienceDirect

Computers in Industry

journal homepage: www.elsevier.com/locate/compind

A real time process management system using RFID data mining

Kyunglag Kwon, Daehyun Kang, Yeochang Yoon, Jong-Soo Sohn, In-Jeong Chung *

Korea University, Republic of Korea

ARTICLE INFO

Article history:
Received 21 October 2013
Received in revised form 16 December 2013
Accepted 11 February 2014
Available online 7 March 2014

Keywords:
Process management
RFID
Data mining
Procedure Tree
Enterprise Resource Planning
Real time system

ABSTRACT

Recently, there have been numerous efforts to fuse the latest Radio Frequency Identification (RFID) technology with the Enterprise Information System (EIS). However, in most cases these attempts are centered mainly on the simultaneous multiple reading capability of RFID technology, and thus neglect the management of massive data generated from the RFID reader. As a result, it is difficult to obtain flow information for RFID data mining related to real time process control. In this paper, we propose an advanced process management method, called 'Procedure Tree' (PT), for RFID data mining. Using the suggested PT, we are able to manage massive RFID data effectively, and perform real time process management efficiently. Then we evaluate the efficiency of the proposed method, after applying it to a real time process control system connected to the RFID-based EIS. For the verification of the suggested system, we collect an enormous amount of data in the Enterprise Resource Planning (ERP) database, analyze characteristics of the collected data, and then compute the elapsed time on each stage in process control. The suggested system was able to perform what the traditional RFID-based process control systems failed to do, such as predicting and tracking of real time process and inventory control.

1. Introduction

The Enterprise Information System (EIS), Enterprise Resource Planning (ERP), Supply Chain Management (SCM) and e-commerce are frequently used to manage business resources in most enterprises. Nowadays, many attempts are made by small and medium enterprises to use Radio Frequency Identification (RFID) technology. Compared with traditional barcodes and labels, RFID technology can provide high efficiency, accuracy and fast reading speed, since human hands are not needed to scan the items one-by-one [1]. Much research on the many advantages of RFID system, such as untouchability, convenience and data storage capability, has been reported for efficient product management [2].

Since the traditional RFID systems have a few drawbacks in obtaining real time valuable information, such as analysis of workflow, and prediction of next stages, they are vulnerable to real time process management. The reason is that their functions are focused mainly on simultaneous multiple reading capabilities, and thus they cannot handle the massive raw data to accommodate real time process management.

Therefore, we need to save production workflow information, and refine RFID raw data into a usable form. For instance, we need some additional processes (e.g. cleansing, filtering or grouping the data), to obtain more valuable information in each process stage, or the workflow of goods. However, the following problems should be solved beforehand, to obtain such information about the real time process management.

- Many occurrences of overhead in the database server, caused by redundant, massive RFID data.
- Difficulty of abstracting human-legible information from the raw data only.
- Incapability of predicting the next sequence, or determining whether the current process is functioning properly or not.
- Difficulty of making correct judgment and flow analysis on the current situation (i.e. human judgment should be involved in recognizing circumstances of a certain time, to provide relevant context-aware information at the specific time).

To overcome these limitations, we suggest a novel data management method for RFID data mining, called Procedure Tree (PT), a more appropriate way for the enterprise to perform effective real time process management. In order to collect process flow information for RFID data mining, PT provides a mechanism to efficiently manage static information related to procedure data, and convert them into a useful form, for effective real time process management. With the suggested PT, we can easily work out the above problems efficiently, and manage dynamic information, such as flow data in process management.

* Corresponding author at: #302, 2nd Science and Technology Building, 2511 Sejong-ro, Sejong-si, Republic of Korea. Tel.: +82 44 860 1342; fax: +82 44 860 1584.
E-mail address: chung@korea.ac.kr (I.-J. Chung).

http://dx.doi.org/10.1016/j.compind.2014.02.007
0166-3615/

We then apply the suggested idea to RFID-based EIS, and implement an intelligent real time process management system. For the verification of the system, we collect a huge amount of data in the ERP database, analyze characteristics of the collected data, and then compute the elapsed time from the preparation stage to the information acquisition stage of process management. Finally, we show that the proposed system is more efficient in executing real time process management effectively than the traditional ones, which could hardly perform the task at all [2].

2. Backgrounds and related work

2.1. RFID data mining

Data mining is the process of extracting and analyzing data, and summarizing it into more useful information, for users to figure out the correlations or patterns among dozens of fields in large data [3]. The following are typical conventional data mining techniques: A priori algorithm [4], Association Rule mining [5], Iterative Dichotomiser 3 (ID3), C4.5 [6], Support Vector Machine (SVM) [7], and variations to each technique [8], etc. However, it is undesirable to apply conventional data mining techniques to RFID data due to the following characteristics of RFID data.

- Simple type of data: data read from RFID readers at different positions are composed of an ordered set of several sources such as Electronic Product Code (EPC), location and time. EPC identifies an item; location is the place where the RFID reader scans the item; and time represents the moment when the reading takes place [9].
- Large quantity of data: because tag data are read periodically, massive data are produced regularly. Thus tagging every item in such a retail giant as Wal-Mart involves an enormous amount of data (approximately 7 terabytes of RFID data per day [10]), which requires periodic management of massive data.
- Accuracy of data: because recognition rate of data varies according to the actual circumstances of the work place, the real recognition rate of RFID is often in the 60–70% range [11,12]. This means that data read from readers must be cleansed.
- Spatial and temporal data [1]: by tracking and monitoring tagged products, information on location and time change is generated. Therefore data model suitable for application level which dynamically generates location and time change is required.

The characteristics of RFID are not appropriate for numeration and categorization of data; whereas we can numerate and categorize data easily, with the conventional methods. For this reason, it is difficult to extract valuable information for real time process management, using traditional data mining techniques.

In order to deal with RFID data, and to mine the valuable information from RFID data effectively, many research studies have recently been conducted [13,14]. We can categorize RFID data mining as follows [10]: RFID data cleaning by data mining [15], RFID data flow analysis, path-based classification and cluster analysis [16], frequent pattern and sequential pattern analysis [17], and outlier analysis in RFID data [18]. The author in [10] presented the Path Tree in RFID Cuboid for RFID data analysis and data mining. The RFID Cuboid is composed of three tables: Info table, Stay table, and Map table. In the RFID Cuboid, the author uses Info table to store object information, and Stay and Map tables to manage group information on items at the same location. To create Stay and Map tables for construction of the RFID Cuboid, the author builds Path Tree, which consists of nodes to identify a location, and arcs to represent flow information. In the Path Tree, an arc is called a Generalized Identifier (GID), and is a unique identifier that encodes the path traversed by the items. Path Tree is based on its tendency for items to move in groups. Since items do not move individually, but tend to move in groups of a specific unit, management of items by group becomes necessary. Therefore the author uses GIDs in Path Tree for Stay and Map tables, to group items. Additionally, the author proves efficiency of query processing and improvement of data compression, through data generalization.

However, it is inadequate to obtain individual process information for each item for process management, since the main purpose of Path Tree is to manage bulky object movements in groups, using GID. Therefore, the following problems are likely to arise in direct use of a Path Tree for process management.

- Applying the Path Tree directly is inadequate for the enterprises where individual movement of items is frequent, since Path Tree is based on multiple object movements.
- Because Path Tree is only used for the change of location, not for managing process information, it is inappropriate to apply for process management such as tracking previous process, detecting current process, and estimating next processes.
- Since Path Tree is difficult for tracing the workflow of items, we are unable to analyze the movement trend of items or errors during the production workflow.

Among five categories, we mainly concentrate on the second and fourth topics since these are much more important for the prediction and tracing of processes in the real time process management. Therefore, we suggest the Procedure Tree (PT), to obtain valuable information from massive RFID data and procedure data. We also demonstrate the real time characteristics of process management, by applying the suggested method to a real entrepreneurial environment.

2.2. Manufacturing process management in the RFID-based EIS

Completion of a product needs various working processes. There are constant sequences and orders in manufacturing processes. Since manufacturing processes in most enterprises are composed of complex processes, even partial errors in the manufacturing process can damage the entire working processes, thus critically affecting the quality of manufactured products, and drastically reducing the productivity.

The central aim of process management is smooth control of the entire production workflow, by regulating each work process, based on production result according to definite time plan [19]. Thus process management plays a significant role in enhancing productivity, through segmentation and standardization of each working phase. As shown in Table 1, a lot of studies have been done in the enterprise and academia.

Table 1 shows some typical RFID applications for process management and their limitations. Additionally, few systems are directly applicable to the real production environment due to the following reasons.

- It is difficult to procure a sufficient amount of data in the real environment.
- Technical problems exist in conducting experiments in a limited environmental scenario, because securing a sufficient number of items is difficult [22].

3. Problem statement

As described in the previous section, studies so far still have a few shortcomings to be solved, and few cases are readily applicable to the actual production environment. Since many experiments have been carried out in limited environments, certain