International Journal of Control and Automation
Vol.7, No.10 (2014), pp.145-154
http://dx.doi.org/10.14257/ijca.2014.7.10.14

An Algorithm for Moving Path Discovery using Frequent Graphical Mining Approach

Xuexuan Zhu

School of Mechanical Engineering Xinyu University, Jiangxi Xinyu 338004, China
xinyucollege@163.com

Abstract

RFID (Radio Frequency Identification) has played great important role in tracing and tracking the goods in supply chain management (SCM, where control and automation are very important). The application of RFID technology is able to record the moving path of different objects when they carrying the tags and readers. Thus, large number of information and knowledge are generated when the objects moved from one location to another. Such information and knowledge could be used for decision making such as automation and control. Traditional database does not consider the relationship of different elements of the RFID attributes. This paper introduces an algorithm to discover the frequent moving path of the RFID-equipped objects using graphical mining approaches. A set of experiments are carried out to evaluate the algorithm by using real-life data from a warehouse. It is observed that with the increasing of k *, more information will be shown in the graph and the moving path could be discovered according to the weight of the edge.*

Keywords: *Moving Path, RFID, Graphical Mining, Algorithms*

1. Introduction

RFID (Radio Frequency Identification) has played great important role in tracing and tracking the goods in supply chain management (SCM) [1, 2]. And the control and automation are very significant in SCM [3]. That because the SCM could be efficient and effective when the logistics operations are real-time controlled and the decisions within it may be carried out smoothly. To this end, large number of research has been focused on the RFID application on SCM. In food supply chain, RFID technology has been used to meet the requirements of traceability and visibility [4]. This paper examines the technological approach used which has great implications in relation to the cost related to the traceability system and to ease the deployment. Zhong et al (2008) introduces an RFID-based manufacturing execution system (MES) which integrates the production and SCM in an enterprise [5]. In this system, hardware of RFID devices and software of control system are reported to address the discrete manufacturing environment. Angeles (2006) comprehensively studied the RFID technology in SCM application and implementation issues are focused [6]. As a mobile technology, process freedoms and real-time visibility into SCM could be enabled. In the Italian FMCG supply chain, RFID technology with EPC network plays significant role in real-time visibility of product flows, reflecting with the quantitative assessment of the potential reduction in the bullwhip effect [7].

The application of RFID technology is able to record the moving path of different objects when they carrying the tags and readers. Thus, large number of information and knowledge are generated when the objects moved from one location to another. Such information and knowledge could be used for decision making such as automation and control. In order to

ISSN: 2005-4297 IJCA
Copyright ⓒ 2013 SERSC

achieve this purpose, lots of studies are carried out by researchers. Many large retailers use RFID as their new supply chain technology. As a result, six in-depth case studies analyzing 88 reported RFID applications, reports a clear view of the implementation landscape, aiming to guide the organizations to ease their supply chain strategies [8]. The insights into how organizations can use the RFID technology are developed from this paper. An RFID-based traceability approach to improve production scheduling was presented through an in-depth study for a manufacturer is conducted to explore scheduling options enabled by an RFID-based traceability system with a novel information visibility-based scheduling (VBS) rule that utilizes information generated from the real-time traceability systems for tracking work in processes (WIPs), parts and components, and raw materials to adjust production schedules [9]. A topology of indoor spaces as well as the deployment of RFID readers, and propose the transition probabilities that capture how likely objects move from one RFID reader to another in probabilities, together with the characteristics of indoor topology and RFID readers, into a probabilistic distance-aware graph model were proposed to study the objects' movement indoor [10]. A complex dynamic of environmental decentralized supply chains and how these dynamics can affect environmental and economic outcomes was studied, considering a supply chain with a manufacturer and two different suppliers: a recycled-material and a raw-material supplier [11]. As the new concept of Internet of Things, great myriad of studies are carried out for the moving path discovery [12-14].

A typical RFID application needs to keep the whole historical information of every object which carried the RFID devices. The information includes the data from periodical feedback in a same location. Thus, the moving path of individual object is consisted of different locations and sub-paths, which are represented from the form of location and duration. The moving path could be expressed by the RFID moving trajectory so that large graphical diagram could be used for visualizing the moving paths. With the visualization, decision on moving objects would be easily carried out.

In this paper, an algorithm for moving path discovery using frequent graphical mining approach is proposed. Based on the dimensions selected from users, the layers are controlled and the RFID data could be compressed. It aims to achieve visualized control and automation so that the decision makings are much easier in the logistics and supply chain. Traditional database does not consider the relationship of different elements of the RFID attributes. However, RFID data contains the path structure information, which could be used for the frequent path discovery under the OLAP architecture to build up the graphical model of RFID-tagged objects.

The rest of this paper is structured as follows. Section 2 presents the RFID-enabled logistics graph by giving some definitions and operations. Under the definition and operations, Section 3 reports on the algorithm through several layers. Section 4 talks about the experiments and discusses the results. Conclusions and findings are organized in Section 5. Future research directions are in this section as well.

2. RFID-enabled Logistics Graph

The RFID application in supply chain could be abstracted as a graph. RFID database is consisted by the set $< EPC, (a_1, ..., a_n), (m_1, ..., m_k), path >$, where EPC is the electronic product code that is unique globally, $(a_1, ..., a_n)$ and $(m_1, ..., m_k)$ have the same meaning of traditional database, whose attributes are converted into non-path attribute value and measurement value. $path$ presents the path information. The

 Copyright ⓒ 2014 SERSC

difference of the RFID database and traditional database is the path data, which has the unique EPC .

The RFID-enabled logistics graph is based on some definitions:

Definition 1. Graph model. The RFID data could be expressed as a set, $D = \{G_1, G_2, ..., G_n\}$, each $G_i = (I_{1i}, I_{2i}, ..., I_{ki}; G_i)$, $I_{1i}, I_{2i}, ..., I_{ki}$ are the attribute. Thus, the $G_i = (V_i, E_i)$ is a graph model of the RFID data. V_i and E_i are the vertex and the edge attribute. Table 1 takes the moving objects for example to illustrate the definition. The moving objects come from a warehouse from the year of 2001 to 2010. Figure 1 shows the graph established from the data in table 1. The information includes the product and time. The attributes of the vertex are location and quantity. The edge indicates the moving path of the objects.

Table 1. RFID Data Sample

Product	Year	Quantity	Path
A	2001	300	Y,Z,N
A	2001	700	Y,S,H
A	2002	200	Y,Z,N
A	2002	115	Y,Z,H
B	2001	100	Y,Z
B	2001	230	Y,S,H
…	…	…	…
A	2010	100	Y,Z,H,S
…	…	…	…

Copyright ⓒ 2014 SERSC

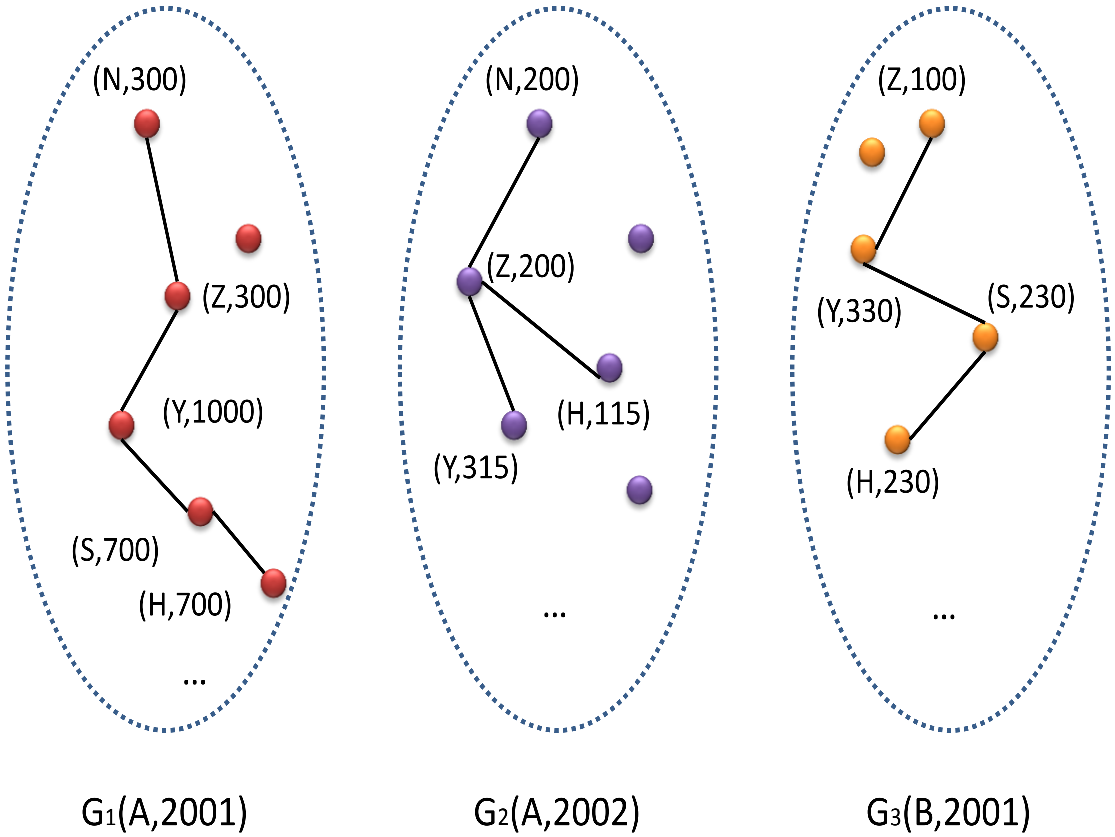

Figure 1. Established Graph Model from Table 1

Definition 2. Vertex set. Given a graph model G , vertex set $\phi = \{g_1, g_2, ..., g_k\}$ only when (1) $\forall g_i \in \phi, g_i \subseteq V(G)$ and $g_i \neq \varnothing$; (2) $\cup_{g_i \in \phi} g_i = V(G)$; (3) $\forall g_i, g_j \in \phi (i \neq j)$, $g_i \cap g_j = \varnothing$. Thus, ϕ is a vertex set of G .

Definition 3. Priority sequence. ϕ and ϕ' are the two vertex set of G , if $\forall g_i^{'} \in \phi, \exists g_i \in \phi$, s.t. $g_i^{'} \subseteq g_i$, then ϕ has the priority sequence of ϕ': $\phi \succ \phi'$.

Definition 4. Attribute compatible group. A is a sub-set of vertex attributes in G . If one of the vertex set ϕ satisfies: $\forall u, v \in V$, if $\phi(u) = \phi(v)$ that means u and v are in the same set. And $\forall a_i \in A, a_i(u) = a_i(v)$ u and v have the same attribute value. Thus, ϕ is an attribute compatible group, which is expressed as ϕ_A . If all the attribute compatible groups is S_A , which has a maximum ϕ_A^{max} , s.t. $\forall \phi_A^{'} \in S_A, \phi_A^{'} \prec \phi_A$.

Definition 5. A-R compatible group. $A(R)$ is a sub-set of vertex (edge) in G . If ϕ satisfies (1) ϕ is attribute compatible; (2) the neighbor edge set from the vertex in a same group, $NG_{\phi, E_I}(u) = NG_{\phi, E_I}(v)$, $E_I \in R$. Then, ϕ is an A-R compatible group,

 Copyright ⓒ 2014 SERSC

which is marked as $\phi_{(A,R)}$. Let $S_{(A,R)}$ denotes the A-R compatible group. There is a maximum group $\phi^{\max}_{(A,R)}$, s.t. $\forall \phi'_{(A,R)} \in S_{(A,R)}, \phi'_{(A,R)} \prec \phi_{(A,R)}$.

Definition 6. Establishment operation. Given a graph model G, attribute sub-set A, relation sub-set R, k is the group amount. Then $|\phi^{\max}_{A}| \le k \le |\phi^{\max}_{(A,R)}|$, an aggregation graph G_{k-s} is established. The vertex of G_{k-s} is the sub-group set of establishment operation on ϕ_A. Then, k is the vertex amount of G_{k-s}.

3. Algorithm for Moving Path Discovery

3.1. Group Tree Data Structure

The algorithm proposed in this paper is based on some data structures like group tree (T), neighbor graph of binary (NG-B) graph, and participating array (PA). The group tree data structure is the main part which is able to connect the NG-B and PA. The following Figure 2 shows the relationship among these three data structures.

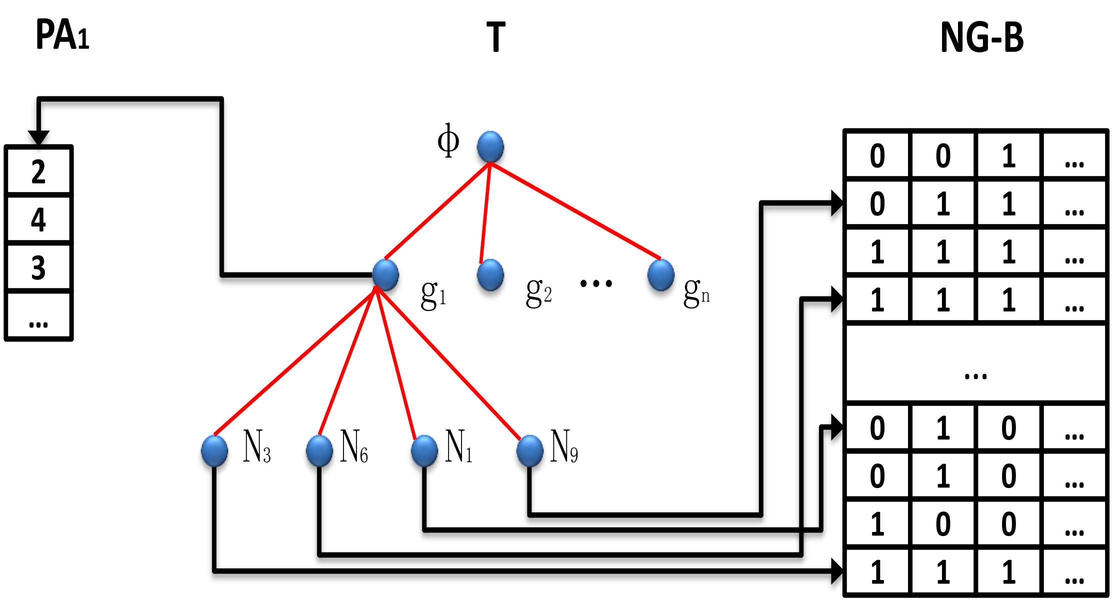

Figure 2. Data Structure Based on T

Figure 2 shows the data structure established from Table 1. There are three main parts. T is used for storing the graph from ϕ. The root of T is the current group ϕ. The first layer is $g_1, g_2, ..., g_n$. Nodes from different g_i are built as its children. $NG-B$ records the relation of vertex and edges. If a vertex N_i connects with the edge from g_j, the value in the corresponding position is '1'. Otherwise is '0'. PA is got from $NG-B$. Each g_i has a PA_i which connects with the g_i. The array keeps the vertex amount of the connection between g_i and g_j, marked as $|P_{E,g_i}(g_i)|$.

3.2. Proposed Algorithm

It is very important to find out the frequent moving path of the objects. Thus, based on the graph model, we proposed an algorithm for the frequent path discovery.

Algorithm $MPD(G, A, R, k)$

Input: Graph G , attribute set A , Edge set R , parameter k

Output: frequent moving path graph with k vertex

1. Compute $\phi_A^{\max}$;
2. Initialize the data structures;
3. $\phi_c = \phi_A^{\max}$;
4. $l = |\phi_c|$;
5. If $l < k$ then
6. Establish (ϕ_c);
7. Else if $l > k$ then
8. Build-graph (ϕ_c);
9. Endif
10. Return Graph-k;

```
Establish (φc)
{
    While PA of gi in T not '0' or '1' do
    {
        For(each child of gi ; l < k )
        {
            Group nodes which are lined together;
            l++ ;
            Update the data structures;
        }
        If l = k then
        Return;
    }
    Endwhile;
}
Build-graph (φc)
{
    Compute MD for pairs of groups;
    While l > k do
    {
        Select the pair of groups with the smallest MD value from T ;
        Combine the two groups;
        l++ ;
        Update the data structure;
    }
    Endwhile;
```

 Copyright ⓒ 2014 SERSC

Return;
}

In the proposed algorithm, according to the attribute set given by the end-users, the ϕ_A^{max} is calculated. The initial data structure is established. The output is a frequent graph which indicates the frequent moving path. The graph has k vertex. In the algorithm, the value of PA in g_i with '0' or '$|g_i|$' is regarded as the measure condition. The frequent group is build up since the vertex is connected to the T which means there is a link or edge of this two vertex. So a moving path exists between these two vertexes. The group has the relation $P_{E,g_i}(g_i)$. Thus, the frequent graph has the vertex participating degree which indicates the frequency of the moving path in the graph. The frequency is expressed as:

$$p_{i,j} = \frac{|P_{E,g_j}(g_i)| + |P_{E,g_i}(g_j)|}{|g_i| + |g_j|}$$

In this algorithm, a measurement index of the frequent moving path is proposed which is MD.

$$MD(g_i, g_j) = \sum_{k=i,j} |p_{ik} - p_{j,k}|$$

MD reflects the frequency measurement of the difference of g_i and g_j.

4. Experiments and Results Analysis

In order to evaluate the effectiveness and feasibility of the proposed algorithm, experiments are carried out. The experiments are under a computing environment as: CPU Pentium IV 1 GHz, 1G Memory, the operation system is Windows XP Professional SP2. Experimental data comes from a warehouse which distributes goods equipped with RFID technology to different locations. We select a yearly data for the experiments. Figure 3 shows the abstracted data for the experiments.

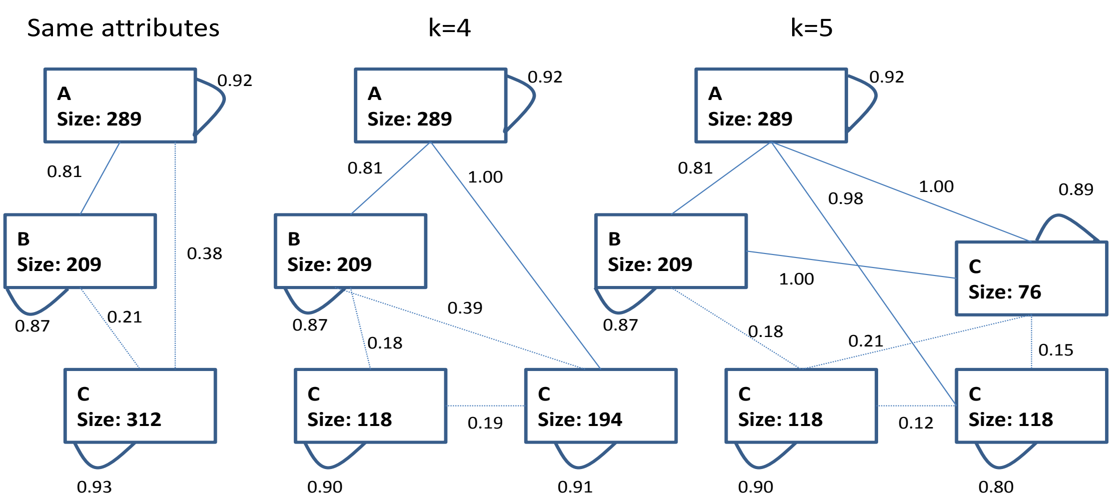

Figure 3. Graphical Results from Experimental Data

From Figure 3, it can be observed that with the increasing of k, more information will be shown in the graph and the moving path could be discovered according to the weight of the edge. Take $k = 4$ for example, the location c is divided into two groups. And both groups are connected to location B. Thus, both location has the moving path. However, the frequent moving path is with the weight 0.39. According to the logic, this proposed algorithm will figure out the frequent moving path using the graphical mining approach.

The effectiveness of the proposed algorithm is examined with different data samples. Figure 4 shows the time cost of the proposed algorithm.

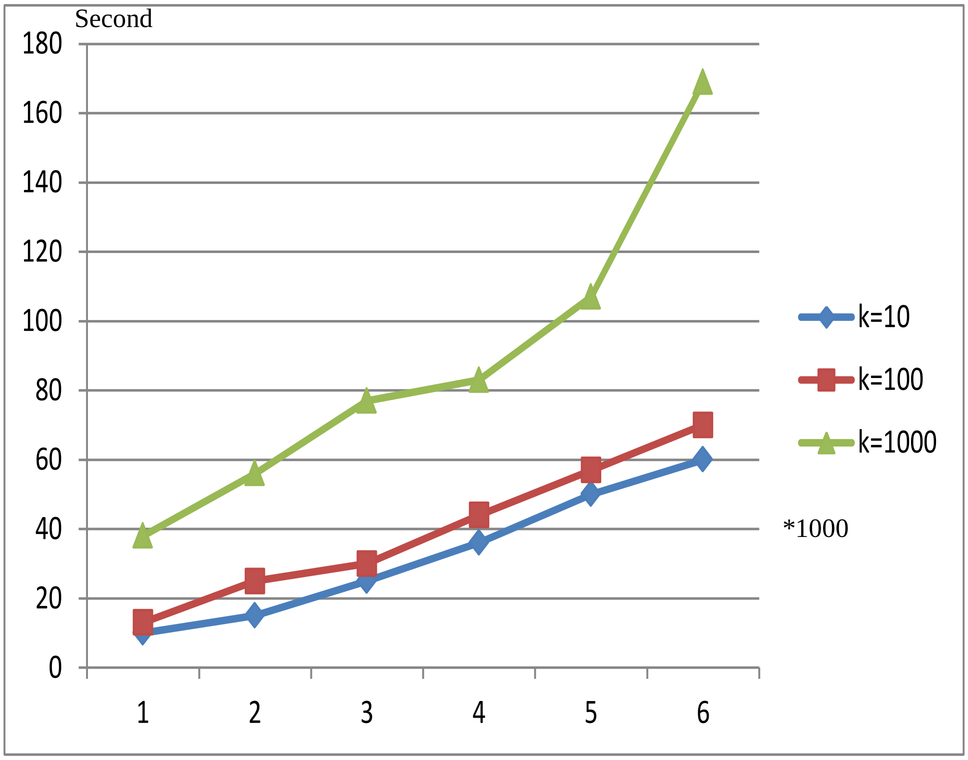

Figure 4. Time Cost of the Proposed Algorithm

From Figure 4, three experiments with different sample sizes: 10, 100, 1000 are carried out. It is observed that, with the increasing of sample sizes, the execution time increases. However, it increases gently at the first beginning. However, if the sample size is over 5000, that means the graphical nodes are too many, the time cost increases sharply. That because the iteration of the proposed algorithm aims to scan all the nodes from the data structure.

5. Conclusion

As the wide usage of RFID technology in logistics and supply chain management (SCM), large number of RFID-enabled data is collected. Such data carry rich moving path information and knowledge which could be used for advanced decision-makings like path planning and scheduling. This paper introduces an algorithm to discover the frequent moving path of the RFID-equipped objects using graphical mining approaches.

 Copyright ⓒ 2014 SERSC

Firstly, a set of definitions is proposed to establish a basic for the algorithm. Due to the differences between RFID database and traditional database, this paper defines a graphical model for using the RFID data. Secondly, the graphical groups are defined so that the graphical model from RFID data could be interpreted as the moving path for frequently objects delivery is figured out. Thirdly, the algorithm is proposed for discovering the moving path and the measurements of the frequency are calculated. Finally, experiments are carried out to evaluate the feasibility and effectiveness (time cost) of the proposed algorithm.

There are several aspects to improve the algorithm. First of all, the data structure based on the group tree T could be improved. The relationships among T, PA, and NG-B could be compressed by a queue so that the data size of recording such datasets will be reduced. Secondly, the iterative loop in the algorithm could be improved by twining parallel computing technologies to reduce the time cost when the sample size is too large. For example, it is observed that the time cost increases sharply when the size is over 5000 in the experiments. How to reduce the time cost could be further considered.

References

[1] A. Brintrup, D. Ranasinghe and D. McFarlane, "RFID opportunity analysis for leaner manufacturing", International Journal of Production Research, vol. 48, no. 9, **(2010)**, pp. 2745-2764.

[2] R. Y. Zhong, *et al.*, "RFID-enabled Real-time Advanced Planning and Scheduling Shell for Production Decision-making", International Journal of Computer Integrated Manufacturing, vol. 26, no. 7, **(2013)**, pp. 649-662.

[3] R. Y. Zhong, *et al.*, "RFID-enabled Real-time Manufacturing Execution System for Mass-customization Production", Robotics and Computer-Integrated Manufacturing, vol. 29, no. 2, **(2013)**, pp. 283-292.

[4] T. Kelepouris, K. Pramatari and G. Doukidis, "RFID-enabled traceability in the food supply chain", Industrial Management & Data Systems, vol. 107, no. 2, **(2007)**, pp. 183-200.

[5] R. Y. Zhong, *et al.*, "Design and Implementation of DMES Based on RFID. Proceeding in 2nd International Conference on Anti-counterfeiting", Security and Identification, (2008) Aug. 20-23, pp. 475-477, Guiyang.

[6] R. Angeles, "RFID technologies: supply-chain applications and implementation issues", Information systems management, vol. 22, no.1, **(2005)**, pp. 51-65.

[7] E. Bottani, R. Montanari and A. Volpi, "The impact of RFID and EPC network on the bullwhip effect in the Italian FMCG supply chain", International Journal of Production Economics, vol. 124, no. 2, **(2010)**, pp. 426-432.

[8] J. Leung, W. Cheung and S.-C. Chu, "Aligning RFID Applications with Supply Chain Strategies", Information & Management, vol. 51, no. 2, **(2014)**, pp. 260-269.

[9] J. Chongwatpol and R. Sharda, "RFID-enabled track and traceability in job-shop scheduling environment", European journal of operational research, vol. 227, no. 3, **(2013),** pp. 453-463.

[10] A. I. Baba, *et al.*, "A graph model for false negative handling in indoor RFID tracking data", in Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, **(2013)**, pp.454-457, ACM.

[11] J. J. Nativi and S. Lee, "Impact of RFID information-sharing strategies on a decentralized supply chain with reverse logistics operations. International Journal of Production Economics", vol. 136, no. 2, **(2012)**, p. 366-377.

[12] R. Y. Zhong, *et al.*, "Mining SOTs and Dispatching Rules from RFID-enabled Real-time Shopfloor Production Data", Journal of Intelligent Manufacturing, DOI: 10.1007/s10845-012-0721-y, **(2013)**.

[13] Q. Y. Dai, *et al.*, "Radio frequency identification-enabled real-time manufacturing execution system: a case study in an automotive part manufacturer", International Journal of Computer Integrated Manufacturing, vol. 25, no. 1, **(2012)**, pp. 51-65.

[14] D. P. McCloskey, *et al.*, "From market baskets to mole rats: using data mining techniques to analyze RFID data describing laboratory animal behavior", **(2011)**, ACM.

Author

Xuexuan Zhu, is a lecturer in School of Mechanical Engineering Xinyu University, Jiangxi Xinyu. She graduated from Nanchang University with bachelor of engineering and master of philosophy in 2001 and 2006 respectively. She has published about 10 papers in international journals and conferences. Her research interests include automation and control, algorithms and analysis, as well as electrical applications.

 Copyright ⓒ 2014 SERSC