A Data Mining Framework for Massive RFID Data Based on Apriori Algorithm

Fan BU
WMG, International Manufacturing Centre, University of Warwick, Coventry, CV4 7AL, United Kingdom
502318400@qq.com

Abstract. In modern manufacturing industries, machines are connected and shared within the manufacturing network. Thus, IoT (Internet of Things) and information technologies are widely applied in the smart factory, among which RFID (Radio Frequency Identification) devices are the most important elements to collect real time data and track important objects. With the continuously developing and widely applications of RFID technology, huge amount of RFID data would be generated during manufacturing processes. In this paper, an intelligent factory framework based on RFID is proposed and massive RFID data is produced. Then the unique characteristics of RFID data in intelligent factory are analyzed, and an algorithm of mining frequent patterns based on Apriori is designed to mine the frequent path knowledge. The algorithm is helpful for the production planning and task scheduling of intelligent factory.

1. Introduction
In recent years, the concept of IoT (Internet of Things) has been widely spread among manufacturing industries all over the world due to its power to readily connect “anything at anytime and anywhere” [1]. Many researches have been conducted on the development and implementation of IoT. It is believed that IoT would be a promising system to achieve auto execution for intelligent plants [2] and also to provide vital information for management to make decisions. In order to realize the connectivity and to build the IoT inside manufacturing plants, RFID has been applied in the workshops as it is a vital technique to help collecting real time data of materials, components, robots and other devices [2] and track important objects throughout the manufacturing processes [3].

However, with the continuously developing and more widely applications of RFID technology, huge amount of RFID data would be generated during manufacturing processes. A lot of useful information, or so called potential patterns with great value are contained in these data. By extracting useful information from RFID database, production planning, task scheduling and abnormal condition detection of intelligent plants could be benefited. By combining the utilization of RFID technology in intelligent plants, this article analyses the unique characters of RFID path data, argues the mining methods of RFID path data and forwards an algorithm of mining frequent patterns based on Apriori.

The rest of this paper is organized as follows. Section 2 reviews previous studies related to this work. The architecture and contents of the RFID-based intelligent factory are described in Section 3. Section 4 details the development of RFID frequent patterns mining based on the Apriori algorithm. Finally, conclusions and future work are discussed in Section 5.

2. Literature review
2.1. Implementation of RFID technology in intelligent factory

2.1.1. A comparison of identification technologies

Huang et al. [4] argued that monitoring efficiency is one of the biggest challenges in collecting data and tracking components in intelligent plants. Wi-Fi and Bluetooth equipment simply cannot be used for identification due to the high level installing cost and built-in application issues [5]. However, there are three other well-proven technologies—QR codes, barcodes, and RFID that can be applied in identification tasks [6-7].

| Table 1: Comparison of barcode, QR code and RFID in identification technology |
|-------------------------------------------------|----------------|----------------|
| Barcode | QR code | RFID |
| Multiple tag reading | NO | NO | YES |
| Ability to copy with harsh environments | LOW | LOW | HIGH |
| Distance limit | 0-4M | 0-4M | 30CM-AT LEAST 100M |
| Specific item identification | NO | NO | YES |
| Information update | NO | NO | YES |
| Maximum data storage amount | 20 Digits | 3KB | 128KB |
| Monitor presence of parts | NO | NO | YES |

Based on the comparison made in table 1[6]-[7], the advantages of RFID are shown clearly. RFID would achieve multiple tags reading in a long effective range while compare to barcode and QR code which can only be readable in 4 meters and incompatible for multiple tags reading.

RFID is more adaptive to harsh environment than the other two. This means the RFID would work more accurately and uninterrupted than barcode and QR code under unexpected weather conditions.

RFID tags could readily be upgraded due to the larger amount storage capacity and the ability to update the information that kept inside the tags.

Using RFID would achieve a higher monitoring accuracy by enabling to track and identify specific items and monitoring presence of parts. While these cannot be realized when using barcode or QR code.

2.1.2. Current situation of RFID implementation in manufacturing plants

- Logistic:

Since the fast development of information technology in supply chain and logistics, RFID has been used to tackle with new challenges. RFID has provided an efficient method to track product throughout the logistic processes by recognizing unique IDs in products tags [8]. In addition, RFID technology could automatically collect real-time data instead of manual operations during the whole transportation [9]. With the visibility of products, RFID has increased the productivity while decreasing the operation costs and tracking inaccuracy [10].

- Warehousing

In recent years, RFID has become a popular method to help managing warehouses in intelligent plants. Each RFID tag is carrying a unique ID, a microchip and a battery so that suppliers could write the information (such as type, batch info, producer and etc.) of the product into the tag [11]. While an active RFID tag is matched with the information system, the position and quantity information would be automatically updated and then transported to the management system [12]. As Hao et al. [12] stated, the multiple-read function of RFID has avoided the man-made errors during both products tracking and shipment selection with a lower labor costs at the same time.

- Production line

In intelligent workshops, production line are equipped with the walking-workers (robots with RFID readers). Production materials are brought to each working center which is located in fixed positions
while working robots flexibly move to the center to assemble the parts [20]. Passive RFID tags are deployed on the ceiling of the environment while the tag readers are installed inside robots. When the robot is moving inside the building, RFID reader could measure the different phases of signal to help locating the robot position [13] in order to collect the path data to track the robots. In addition, by using RFID tags to predict obstacles in the path of robots, moving correction system is developed more effectively [14]. As a result, RFID technology could improve the effectiveness of production line and also the convenience of managing the production line by collecting path data of robots.

2.2. Data mining methods of RFID in intelligent plants

2.2.1. Attributes of RFID data
RFID data consists some simple information. After the tags being read by sensors, the information of scanning location, time and EPC(Electronic Product Code) are transmitted in the enterprise information system[15]-[16]. As Derakhshan et al.[17] discussed, although there is not much information carried in single RFID data, the manufacturing companies should also put efforts on transforming the raw data into an available form for enterprise-level applications to use.

Data amount of RFID is always large. As it is known, RFID sensors are continuously transmitting real time data to the system. Some sensors could read more than 100 tags within one second so that even a medium size RFID-based workshop could generate gigabytes of data just in a single day without counting the chances of a great number of data flood events happening in the system [17].

Another issue of RFID data is the inaccuracy. Jeffery et al.[18] pointed out that reading errors and unreliable data could always be found in RFID data streams. Many researchers have discussed the average accuracy of RFID reading data lies between 60% to 70% which means RFID equipment would generate a large part of both inaccuracy and useless data during the work time [17][19]. Hence, the cleaning and extracting of RFID data has drawn much attention of researchers [19].

2.2.2. Current methods of RFID path data mining
In modern manufacturing business, data mining is a technology tool which combines modeling and algorithms to analyze huge amount of data in order to extract more useful and comprehensive information for management to make their decisions [21]. With the increasing amount of both RFID technology and smart robots that implemented in intelligent workshops, the amount of path data generated would be more than ever [22]. Thus, RFID path data mining becomes a crucial tool for manufacturers.

Here are three methods that are commonly used in RFID path data mining:
- Association rules:
  Speaking of data mining, many scholars have stated that association rules is one of the most studied methods [23]. As Kim et al. [23] mentioned, association rules includes both positive associations and negative associations. For example, the location data and the time data appeared together is a positive association while locations which robots did not appear is a negative association. However, with the increasing amount of location and time data of RFID, current association rules should be modified in order to extract more useful data. As Won et al. [24] discussed, they combined the ontology-driven rule with hierarchically clustering association rules to reduce the data duplication during RFID data mining to improving data relevance. Also, Kim et. al [23] modified multiple association rules to reduce the level of abstraction of RFID data. In addition, Li-min [25] has reviewed the pros and cons of Apriori algorithms to point out a new research direction of launching it into RFID path data mining.
- Cluster analysis:
  Another useful data mining tool for RFID data is clustering. Clustering mechanisms are algorithms that designed for grouping the data of high similarity or according to different measurements related to each situation [28]. The main purpose of clustering analysis implement in mining RFID path data is to decrease the repetition and to further filter the collected path data to database [26]. While the data tags are clustered, each manufacturing activity would be linked with the data of location, time and the order
of the actions. Thus, the data flow can be taken into a probability model for further analysis [27].

-Frequent path data mining:

In this years, RFID technology has been widely used to provide an effective and relatively inexpensive method of monitoring walking-walkers in intelligent workshops as the frequent path is always expected information of business manufacturer management. However, the efficiency of this method is based on the data mining tool to extract path patterns [29]. Because the problem formulations and path data mining algorithms are different due to the application situations, a number of algorithms has been introduced during these years [30][31]. With the increasing utility of RFID technology, frequent path mining methods would attract more attention from scholars.

3. An intelligent workshop architecture based on RFID

3.1. Introduction of the framework

As it is shown above, figure 1 has presented an intelligent workshop framework that based on the implement of RFID technology. Generally, there are objects with RFID tags and objects with RFID readers. Those objects implemented with RFID technology would generate massive amount of data in daily manufacturing processes. Data could be stored in local warehouse or updated in cloud service for further processing.

3.1.1. Fixed or mobile RFID objects

In this framework, raw material, unfinished goods, finished goods and logistic robots are moving RFID objects. Besides the last one, other objects are passively tagged. Moving robots are equipped with active reader so that they can scan the RFID tags of other moving objects during the whole process.

On the other hands, some RFID readers are fixed at the exit of raw material warehouse, some are fixed at the entrance of finished goods warehouse and others are fixed with the machines in each working process stage. These fixed readers scan the objects come to them and collect information to communicate with data warehouse in wireless environment.

3.1.2. Wireless environment

In this intelligent workshop, the communication mechanism is based on the wireless environment constructed in the workshop due to the mobility of many moving RFID objects. However, to secure the accuracy and the continuity of information transmission, Bluetooth and Wi-Fi are considered to be better choice than other wireless telecommunication technologies. In addition, these data collect from each working process would be put into local data warehouse and cloud service.

3.1.3. Data storage

The local data warehouse is a traditional place for data storage, cleansing, processing and mining. While considering the fast development of cloud computing, this framework also made it possible to upload data to cloud service. Cloud computing could make production plan, material management and some relative services become more efficiency and also save computing energy for workshops.
4. Frequent patterns mining of massive RFID data

4.1. RFID data form
During a typical RFID application, when an object with RFID tags passes a reader, the information of EPC code of the object, reader ID and the event time will be recorded as the following format:

<EPC code, reader ID, event time>

If this object passes some different readers in different locations, to show the path information, primary data would be processed as:

Object ID<location 1, location 2, ..., location n>

This simplified format of path data could meet the requests in most kinds of path data mining problems. Although the formats shown above can be quite easy for readers to understand, the correlations between those path patterns would be much difficult to express. Thus, the mining methods of this basic type of path data is also important for further studying.

4.2. The objective of RFID data mining
In modern intelligent plants, work load during peak-time can be significantly high. However, some processing stages can be repeated more times than others due to the difference of products. Thus, there is always a highly standard request of logistic routing planning within the plants to help reduce logistic pressure and also increase manufacturing efficiency.

By using Apriori algorithm, our mining task is to correlate event time with object path data in order to identify the most frequent path patterns inside the plants throughout off-time to peak-time. The frequent patterns would represent the most frequent traffic routes with time information and the routes that are not so occupied relatively. By analyzing this, management could readjust their path planning, dispatch plan of their Automatic Guided Vehicle(AGV) robots, and even the layout of manufacturing machines to increase working efficiency.

By mining the RFID path patterns, most frequent patterns and recommended new path patterns which could increase logistic performance would be output from the programs as expected.

4.3. Flowchart of frequent patterns mining processes
Figure 2 Flowchart of frequent patterns mining processes

Key concepts of applying Apriori algorithm:
1. All the sets that includes items more than the min_supp are considered as frequent sets.
2. Any non-empty subsets of those frequent sets are also frequent.
3. Usually, the value of min_supp is a percentage. However, in this case, the min_supp must be an integer because the item actually stands for locations which can only be counted by using integers.
4. The confidence value is used to select strong rules. If the confidence C is bigger than the min_conf, it means the potential correlations between those items could be more valuable to analyze.

5. Conclusion and future work
In recent years, RFID technology has been widely applied in many areas. Meanwhile, due to the fast increasing amount of data produced from RFID applications, how to extract useful information from RFID database has become a significant researching topic. When objects with passive tags are moving inside the plant, the moving tracks between locations with path data are generated and kept by the implemented RFID devices. By combining the utilization of RFID technology in intelligent plants, this article analyzed the unique characters of RFID path data, argued the mining methods of RFID path data and presented an algorithm of mining frequent patterns. This algorithm could facilitate business management in fast building production plan, assembling routes in order to increase manufacturing efficiency.

In our future research, more dimensions of RFID data should be take into account. Besides, mining and analyses of real time RFID data also require more further research.

Reference
[1] Li, Shancang, Li Da Xu, and Shanshan Zhao. "The internet of things: a survey." Information Systems Frontiers 17.2 (2015): 243-259.
[2] Atzori, Luigi, Antonio Iera, and Giacomo Morabito. "The internet of things: A survey." Computer networks 54.15 (2010): 2787-2805.

[3] Gusmeroli, Sergio, Salvatore Piccione, and Domenico Rotondi. "A capability-based security approach to manage access control in the internet of things." Mathematical and Computer Modelling 58.5 (2013): 1189-1205.

[4] Huang, Yi, Brian C. Williams, and Li Zheng. "Reactive, model-based monitoring in RFID-enabled manufacturing." Computers in Industry 62.8 (2011): 811-819.

[5] Makris, Sotiris, George Michalos, and George Chryssolouris. "RFID driven robotic assembly for random mix manufacturing." Robotics and Computer-Integrated Manufacturing 28.3 (2012): 359-365.

[6] Wyld, David C. "RFID 101: the next big thing for management." Management Research News 29.4 (2006): 154-173.

[7] Arendarenko, Ernest. "A study of comparing RFID and 2D barcode tag technologies for pervasive mobile applications." Master, Department of Computer Science and Statistics, University of Joensuu (2009).

[8] Erkan, Turan Erman, and Gulin Feryal Can. "Selecting the best warehouse data collecting system by using AHP and FAHP methods." Tehnicki vjesnik 21.1 (2014): 87-93.

[9] Spekman, Robert E., and Patrick J. Sweeney. "RFID: from concept to implementation." International Journal of Physical Distribution & Logistics Management 36.10 (2006): 736-754.

[10] Varila, Mikko, Marko Seppänen, and Petri Suomala. "Detailed cost modelling: a case study in warehouse logistics." International Journal of Physical Distribution & Logistics Management 37.3 (2007): 184-200.

[11] Jie, SunYu, and Julian Clive. "Research of Logistics Product Intelligent System Distribution Based on Internet of Things." International Journal of Digital Content Technology and its Applications 7.6 (2013): 979.

[12] Hao, H., and Bin Zhao. "Application of RFID technology in development of warehouse logistics management system by Internet of things." International Conference on Education, Management, Commerce and Society (EMCS-15). Atlantis Press. 2015.

[13] DiGiampaolo, Emidio, and Francesco Martinelli. "Mobile robot localization using the phase of passive UHF RFID signals." IEEE Transactions on Industrial Electronics 61.1 (2014): 365-376.

[14] Nakanishi, Tadashi, et al. "Moving correction method of a mobile robot using passive RFID system based on obstacle prediction." Indoor Positioning and Indoor Navigation (IPIN), 2014 International Conference on. IEEE, 2014.

[15] Gonzalez, Hector, et al. "Warehousing and analyzing massive RFID data sets." Data Engineering, 2006. ICDE'06. Proceedings of the 22nd International Conference on. IEEE, 2006.

[16] Generation, E. P. C. "Tag Data Standards Version 1.1 Rev. 1.27." Standard Specification, EPCglobal Inc TM (2005).

[17] Derakhshan, Roozbeh, Maria E. Orlovska, and Xue Li. "RFID data management: challenges and opportunities." RFID, 2007. IEEE International Conference on. IEEE, 2007.

[18] Jeffery, Shawn R., Minos Garofalakis, and Michael J. Franklin. "Adaptive cleaning for RFID data streams." Proceedings of the 32nd international conference on Very large data bases. VLDB Endowment, 2006.

[19] Floerkemeier, Christian, and Matthias Lampe. "Issues with RFID usage in ubiquitous computing applications." International Conference on Pervasive Computing. Springer Berlin Heidelberg, 2004.

[20] Huang, George Q., Y. F. Zhang, and P. Y. Jiang. "RFID-based wireless manufacturing for walking-worker assembly islands with fixed-position layouts." Robotics and Computer-Integrated Manufacturing 23.4 (2007): 469-477.

[21] Wang, Kesheng. "Applying data mining to manufacturing: the nature and implications." Journal of Intelligent Manufacturing 18.4 (2007): 487-495.
[22] Aggarwal, Charu C., and Jiawei Han. "A Survey of RFID Data Processing." Managing and Mining Sensor Data 2013 (2013): 349-382.

[23] Kim, Younghee, and Ungmo Kim. "Mining Multilevel Association Rules on RFID data." Intelligent Information and Database Systems, 2009. ACIIDS 2009. First Asian Conference on. IEEE, 2009.

[24] Won, Dongwoo, and Dennis McLeod. "Ontology-driven rule generalization and categorization for market data." Data Engineering Workshop, 2007 IEEE 23rd International Conference on. IEEE, 2007.

[25] Li-min, N. I. U. "Review of the optimization and analysis of Apriori algorithm [J]." Journal of Guilin University of Electronic Technology 1 (2007): 006.

[26] Bashir, Ali Kashif, et al. "Energy efficient in-network RFID data filtering scheme in wireless sensor networks." Sensors 11.7 (2011): 7004-7021.

[27] Gonzalez, Hector, Jiawei Han, and Xiaolei Li. "Mining compressed commodity workflows from massive RFID data sets." Proceedings of the 15th ACM international conference on Information and knowledge management. ACM, 2006.

[28] Jain, Anil K. "Data clustering: 50 years beyond K-means." Pattern recognition letters 31.8 (2010): 651-666.

[29] Liu, Yunhao, et al. "Mining frequent trajectory patterns for activity monitoring using radio frequency tag arrays." IEEE Transactions on Parallel and Distributed Systems 23.11 (2012): 2138-2149.

[30] Gonzalez, Hector, Jiawei Han, and Xiaolei Li. "Mining compressed commodity workflows from massive RFID data sets." Proceedings of the 15th ACM international conference on Information and knowledge management. ACM, 2006.

[31] Han, Jiawei, et al. "Warehousing and mining massive RFID data sets." Advanced Data Mining and Applications (2006): 1-18.