Expert Systems With Applications 91 (2018) 442–451

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Automatic detection of false positive RFID readings using machine learning algorithms

Haishu Ma[a], Yi Wang[b], Kesheng Wang[a,*]

[a] Norwegian University of Science and Technology, Department of Production and Quality Engineering, S.P. Andersens veg 5, 7031 Trondheim, Norway
[b] The university of Manchester, School of Materials, Sackville St Blg, Manchester M13 9PL, UK

A R T I C L E I N F O

Article history:
Received 10 January 2017
Revised 3 August 2017
Accepted 9 September 2017
Available online 13 September 2017

Keywords:
RFID
Machine learning
Classification
False positive readings

A B S T R A C T

Radio frequency identification (RFID) has been widely used for the automatic identification, tracking and tracing of goods throughout the supply chain from the manufacturer to the customer. However, one technological problem that impedes the productive and reliable use of RFID is the constraint of false positive readings, which refers to tags that are detected accidentally by the reader but not the ones of interest. This paper focuses on the use of machine learning algorithms to identify such RFID readings. A total of 11 statistical features are extracted from received signal strength (RSS) and phase rotations derived from the raw RFID data. Each of the features is highly statistically different to distinguish the false positive readings, but satisfactory classification cannot be achieved when these features are considered individually. Classifiers based on logistic regression (LR), support vector machine (SVM) and decision tree (DT) are constructed, which combine all of the extracted features to classify the RFID readings more effectively. The performance of the classifiers is evaluated in a real-world factory. Results show that SVM provides the highest accuracy of up to 95.3%. DT shows slightly better accuracy (92.85%) than LR (92.75%), while LR has the larger area under the curve (0.976) than DT (0.949). Overall, machine learning algorithms could achieve accuracy of 93% on average. The proposed methodology provides a much more reliable RFID application as false-positive readings are detected immediately without human intervention, which enables a significant potential of fully automatic identification and tracking of goods throughout the supply chain.

1. Introduction

Radio frequency identification (RFID) can be used to enhance visibility and traceability of supply chain. Once the product is attached with the RFID tag from the beginning of the supply chain, it is assigned a unique electronic product code (EPC) and then can be automatically identified, tracked and traced from the supplier to the customer. Shipping from distribution centers to retailers is one of the key processes in the supply chain. In order to ensure the smooth flow of goods, RFID has been commonly deployed in the warehouse and distribution centers (Keller, Thiesse, Ilic, & Fleisch, 2012). The goods attached with tags are automatically registered by the RFID reader when loaded to a truck after passing through the portal.

However, one technological constraint hinders the effectiveness of the RFID application in the warehouse. RFID reader can detect any object attached with the tag that appears in the reading range of the radio frequency field (Ju Tu & Piramuthu, 2008). But the reader cannot distinguish between tagged tags that actually pass through the portal and the ones that appear in the reading field by chance, it is highly likely that incorrect invoices will be issued and the retailer stores will pay for goods that they neither received nor purchased. A false positive RFID read corresponds to the tag that is detected by the RFID portal but not loaded to the truck. Therefore, the reliable and productive RFID application in the distribution center process remains doubtful without solving the false positive RFID readings (Bong, Chang, & Oh, 2014). Various reasons contribute to the false positive problem. Physical conditions in the warehouses are complex. Metallic object and the truck itself may cause multipath reflections of RF signals, which can extend the read range of the reader (Keller, Thiesse, & Fleisch, 2015). As a result, the tag assumed to be away from the reader can be unintentionally read. Other reasons include that a warehouse man might buffer a tag temporarily near the portal or he is passing by the portal with another tag when the RFID portal is reading (Keller, Thiesse, Kungl, & Fleisch, 2010).

Many efforts have been devoted to resolving the false positive problem. The available measures can be grouped into three

* Corresponding author.
E-mail addresses: haishu.ma@ntnu.no (H. Ma), yi.wang@gmail.com (Y. Wang), kesheng.wang@ntnu.no (K. Wang).

http://dx.doi.org/10.1016/j.eswa.2017.09.021
0957-4174/

categories, i.e. sliding-window, extra hardware, and RSS based method. Bai, Wang, and Liu (2006) proposed the sliding window method. The false positive readings with the occurrence rate below a noise threshold are filtered. This method considers the RFID data stream to be uniform flow, which is the ideal situation in practice. Jeffery, Garofalakis, and Franklin (2006) introduced the adaptive sliding window, SMURF, to compensate for the inherent unreliability of RFID data streams. A variant of sliding window mechanism was proposed by Bian, Peng, and Zhang (2013) to eliminate the false readings in a RFID tracking system. The algorithm has demonstrated its superior performance than the traditional mechanism. Tu and Piramuthu (2011) proposed using extra readers to determine the false positive reading. When the tag is read by the two readers at the same time, it is considered to be actually present. Otherwise, the false positive readings are confirmed if the tag is read by none of the readers. Krigslund, Popovski, Pedersen, and Olesen (2012) presented a novel method focusing on a two-device setup, a reader and an interference source. The setup can impose intentional interference between the reader and tag. Experimental results showed that the false positive read was reduced by imposing interference.

The methods relying on extra hardware to identify false positive RFID readings can incur additional cost. Sliding window methods mainly utilize timestamps and tags readings to detect false positive readings. But there is more valuable information generated by the reader such as Received signal strength (RSS) and phase shift. RSS has been commonly used for indoor real time location system (Luo, O'Brien, & Julien, 2011; Ni, Liu, Lau, & Patil, 2004; Stella, Russo, & Begušić, 2014) and movement detection (Yao et al., 2015). The changes of object's orientation and location affect the RSS reflected by the RFID tag. These signal changes can be leveraged to detect the fluctuations caused by the tag motion. Parlak and Marsic (2013) extracted features from RSS and classified them as moving or still using statistical methods. A trauma resuscitation case study was used to evaluate their methods. Results showed that the accuracy achieved 80% in complex scenarios. Keller et al. (2010) proposed an algorithm that used the information gain criterion to separate tags that were loaded onto trucks and tags that were in range of the reader by accident. The algorithm was verified under real distribution center and results showed that RSS value was the most suitable tag characteristics than timestamps and antenna attributes. Keller et al. (2015) presented an empirical study using data mining techniques to detect false positive RFID tags based on the attributes derived from low level reader data. Moreover, they demonstrated that utilizing full spectrum of data reported by the reader hardware resulted in better performance than single-attribute classifier.

Machine learning has created new intelligent tool for automated extraction of useful information and knowledge from manufacturing systems and processes (Wang, 2007). RFID and artificial intelligence have been implemented together to enhance the responsiveness of the logistics workflow (Lee, Ho, Ho, & Lau, 2011). Zhong, Huang, Dai, and Zhang (2014) introduced decision tree and SVM to excavate practical standard operation times (SOTs) from RFID-enabled real-time shopfloor production data. Supervised pattern classification techniques, including k nearest neighbor and SVM, were used to differentiate the individual tag (Bertoncini, Rudd, Nousain, & Hinders, 2012). Chernbumroong, Cang, Atkins, and Yu (2013) presented an assisted living system that combines neural network and SVM to activities of an elderly person.

Phase value has attracted more attention in indoor localization (Hekimian-Williams, Grant, Liu, Zhang, & Kumar, 2010; Zhou & Griffin, 2012). But it is rarely used for the classification of RFID readings. This paper proposes a method that extract features from RSS and phase values reported by the RFID reader to determine the false positive readings using machine learning algorithms. The rest of this paper is organized as follows. Section 2 introduces the real-world factory where the RFID data are collected and our method is evaluated. Section 3 gives the technical background of RF signal propagation. Extracted features from RFID data are presented in Section 4. LR, DT, and SVM classifiers are constructed in Section 5. In Section 6 we describe the experimental results of the experiments, and discuss some of our findings. Finally, the conclusion and future research are given Section 7.

2. Empirical evaluation

An RFID enabled smart factory is set up by shanghai polytechnic university (SPU) in cooperation with knowledge discovery laboratory (KDL) in Norwegian University of Science and Technology (NTNU).

2.1. Experimental setup

Fig. 1 illustrates the process of the customized keychain from assembly to delivery. The keychain is assembled and produced in the smart factory to embody the typical characteristic of industry 4.0, i.e. mass personalization production. Each user is provided with the options to select the color of the keychain and print their names on it. All this information is transferred to the serial robot, which selects the raw materials, a bottom, a cover and a RFID tag, from the shelf and put them on the workbench equipped with a parallel robot, as shown in Fig. 1. Next the bottom, the cover, and the tag are assembled by the parallel robot. The finished keychain will then be put on the conveyor belt which can transfer the keychain to the printer workstation. The RFID tag will tell the printer what should be printed on the keychain. Finally, a mobile robot can deliver the keychain to the warehouse. All the processes above can be tracked and displayed in a RFID system, which is developed by APX systems in Norway. At each workstation, the assembly and the printer, the RFID reader antenna can read the tag. From the moment the tag is attached to the keychain, the product can be identified and tracked using RFID.

The functions of the RFID system are as follows. When the keychain is detected by the reader antenna installed at the assembly workstation, the RFID information will be displayed on the assembly list box. When the keychain arrives at the printer workstation, the printer list box will show the corresponding RFID information and the information in the assembly list box is cleared. This is real time tracking of RFID tags and all the RFID information during the production is displayed in the history list box and stored in the database of the RFID system for further research.

2.2. Data collection

Two reader antennas (Impinj Brickyard) are mounted on the portal. They are connected to Infinity 610 UHF reader which can report RSS, phase, timestamp and EPC. We recorded 2000 keychains tagged with RFID during our experiment, where 1000 of them were moved tags passing through the portal. Others were static tags that just appeared in the read range of the RFID portal by accident, i.e. the false positive readings. The period during which the RFID data are collected lasts 5 s, which is called the data gathering session. At each data gathering session, the tag is approximately interrogated by the RFID portal for 709 times. After the data collection is completed, we will extract features from the readings and classify them as moved tags or static tags using different machine learning methods.