Pushing the limits of RFID: Empowering RFID-based Electronic Article Surveillance with Data Analytics Techniques

Research-in-Progress

Matthias Hauser
University of Würzburg
Emil-Fischer-Straße 32
97074 Würzburg, Germany
matthias.hauser@uni-wuerzburg.de

Daniel Zügner
University of Würzburg
Emil-Fischer-Straße 32
97074 Würzburg, Germany
daniel.zuegner@gmail.com

Christoph Flath
University of Würzburg
Emil-Fischer-Straße 32
97074 Würzburg, Germany
christoph.flath@uni-wuerzburg.de

Frédéric Thiesse
University of Würzburg
Emil-Fischer-Straße 32
97074 Würzburg, Germany
frederic.thiesse@uni-wuerzburg.de

Abstract

False-positive classification is a central issue for RFID environments with limited process control, such as in-store settings. In the case of electronic article surveillance, false positives not only lead to incorrect inventory data but also trigger false alarms, which impair customer satisfaction. A typical counter measure is to reduce antenna power, which in turn leads to greatly diminished detection rates. In contrast, the present study investigates the applicability of data analytics to achieve high detection rates while retaining low false positives. In contrast to prior research, our test setting acknowledges the lack of process control in retail environments. We consider various walking paths and speeds as well as RFID tags concealed by shopping bags. To distinguish theft from non-theft events, we derive predictors, which are not just aggregations of the signal strength. Rather, individual reads are put into temporal relation to one another and are augmented with antenna information.

Keywords: Machine learning, Data analysis, Radio Frequency Identification (RFID), Logistics

Introduction

In recent years, Radio Frequency Identification (RFID) has become a widespread alternative to traditional auto-id technologies, particularly to the barcode. Although the enthusiastic expectations from the early days of RFID diffusion have not been met so far (Thiesse et al. 2011), RFID is now in use in an ever increasing number of supply chains, such as the fashion industry and elsewhere. However, the quality of data collected by means of RFID transponder labels is still a common issue in many real-world implementations (Keller et al. 2014). Owing to its nature as a fully automatic data collection technology, RFID suffers from different phenomena surrounding the capturing of RF signals (Brusey et al. 2003). On the one hand, so-called 'false-negative' reads denote RFID-equipped objects that remain undetected by a reader device, for example, due to electromagnetic shielding. The usual countermeasures to prevent false negatives include improved antenna configurations or increasing the RF power of the reader. This, in turn, leads to the second phenomenon of 'false-positive' reads, that is, undesired detections of RFID-equipped objects within range of the electromagnetic field.

The presence of false positives poses a major challenge to any RFID-based information system that relies on precise measurements of inventory levels and supply chain operations. Against this backdrop, the present paper investigates the application of data analytics to eliminate false positives from RFID data streams. We particularly consider the use of RFID in the context of electronic article surveillance (EAS), which depends on a high level of data quality for several reasons. First, the cost-efficient substitution of classical proprietary EAS solutions by RFID has become an important foundation for the business case underlying many RFID implementations in retail (GS1 Germany 2007). However, if the performance of RFID-based EAS lags behind its established counterpart, the economic justification of RFID as a whole is at risk. Second, the value of article security in retail stores is evident as it is here, at the very end of the supply chain, where a major fraction of total shrinkage in retail occurs due to theft (Bamfield 2011). Finally, insufficient quality of data on the physical flow of goods exerts a direct influence on the availability of products in stores and, as a consequence, decreasing customer satisfaction because of stockouts (DeHoratius and Raman 2008).

Literature Review

The "RFID rush" in the mid-2000s spurred significant research activity in the operations management community. Liao et al. (2011) show that the most influential articles of this first wave of RFID research mostly fell in three categories, general reviews, prototype systems / case studies and analytic models. The first group (e.g., Angeles 2005; Roberts 2006) essentially provided basic RFID technology background and sketched basic application scenarios. The second group described concrete prototype setups (e.g., Ngai et al. 2007; Ni et al. 2004). The final group assessed the theoretical potentials of RFID systems in the operations management context using formal, analytic models. In general, this kind of analysis was prompted by Lee and Özer (2007) who noted that most RFID promises "are educated guesses at best and are not substantiated" because they were "not based on detailed, model-based analysis". Since then the literature has grown along these lines with the RFID implementations growing in size and scope. However, the necessary procedures for handling massive RFID data sets are still an under-researched issue. Researchers have hardly shifted away from an infrastructural and potential benefit focus to the question of how to actually draw value from large data collections by means of robust business processes.

Research in the area of RFID data quality includes various approaches for the detection and the removal of false positive reads. A number of authors consider threshold-based heuristics using the frequency of tag detections in a sliding-window approach (Bai et al. 2006; Brusey et al. 2003). Here, the number of times a particular tag was detected by the reader within a fixed time interval determines whether it is classified as 'false positive' or not. The underlying assumption is that false-positive reads occur only sporadically whereas tags that correctly pass the RF field are detected several times. Extensions of such algorithms were proposed by Jiang et al. (2006) and Tu and Piramuthu (2008), who consider the use of more than one antenna. In addition to the number of tag detections, both studies investigate complementary information on the number of antennas that detect a tag.

A common weakness of these examples of prior research may be seen in the lack of real-world data that is required to provide a solid estimate of the resulting detection rates. In contrast, the study by Keller et al. (2010) is based on a massive set of raw RFID data gathered at a distribution center equipped with more than 40 RFID gates. Moreover, the study is the first to include signal strength (RSSI) measurements in the data analysis. The authors compare the value of several timestamp-, antenna-, and RSSI-based indicators. The results indicate that RSSI information provides the best means for the classification of false-positive tag detections.

Models that go beyond the simple logic of threshold values were proposed by Goller and Brandner (2011) and Keller et al. (2014). The former present a probabilistic approach based on Hidden Markov Models. The resulting tag detection algorithm shows high detection accuracies for an automated transportation process under lab conditions. Similarly, Keller et al. (2014) use Dynamic Time Warping, a technique known from speech recognition, for the analysis of RFID time series. The corresponding model achieves a detection accuracy that surpasses earlier results by Keller et al. (2010) for the specific case of data sets with exactly one true positive (i.e., only one tag passes the RFID gate and all others are false positives).

Research Objective

Our research is concerned with the development and the evaluation of models for data cleansing in the context of RFID-based systems in retail. The research focus is set on "the last mile" of the retail supply chain (see Figure 1), which includes EAS as well as store processes and customer-oriented applications, for example, RFID-equipped fitting rooms. The application of RFID for monitoring store operations is generally error-prone and challenging because process control is low while the variety of processes on the part of employees and customers is very high. We specifically consider the case of RFID-based EAS for apparel and footwear to prevent shoplifting while minimizing the number of false alarms due to false-positive tag detections. Here, typical error sources are customers walking by the gate in close proximity but not through the gate or shielding of RFID tags. Shielding could be caused by customers themselves (e.g. by bunching up garments or putting them in a bag), or if two (or more) customers are in the reading field of the antennas, as the body fluids in persons significantly reduce the antennas' ranges. The purpose of our study is to go beyond prior research with regard to (i) the complexity and the performance of our classification approach and (ii) its generalizability under real-world conditions. To this end, we pursue a design-oriented research approach to develop a suitable classification artifact. The work presented here focuses on the design cycle (Hevner and Chatterjee 2010). In particular, we loosely follow the Cross-Industry Standard Process for Data Mining (CRISP-DM), a standard process model widely applied to approach data-mining problems (Chapman et al. 2000). The main outcome of our research is a data-mining model to improve data quality of RFID-based processes that is not limited to processes with full process control but is also applicable to processes with limited process control.

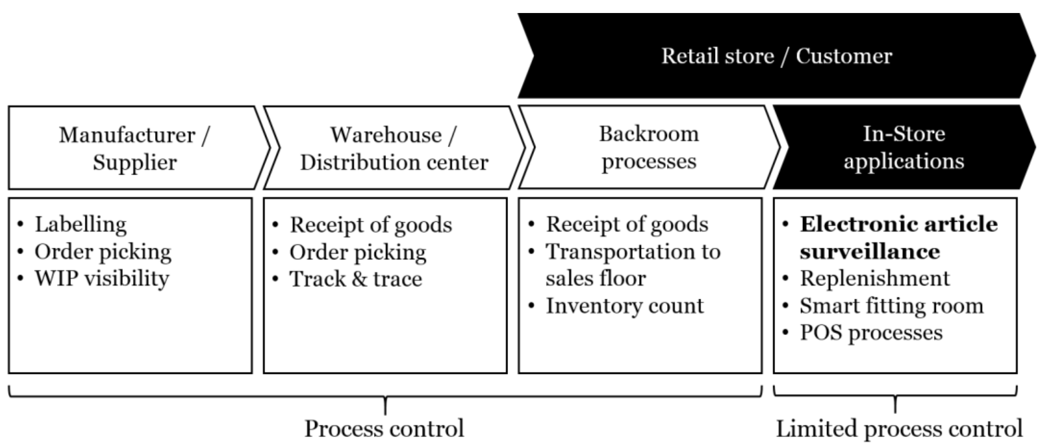

Figure 1. RFID-based processes along the retail supply chain

EAS portals are located at the entrances and exits of retail stores and trigger an alarm if a customer leaves the store with unpaid items. A viable EAS system necessitates the lowest possible number of false alarms. Otherwise, customer experience is at risk because a substantial number of customers frequently cause alarms. A typical countermeasure to avoid false positives is to reduce the power of the antennas, as this leads to tags not being read until they are very close to the gate (Bottani et al. 2012). However, this approach not only reduces the number of false positive tag reads but the read rate in general.

In a pre-study, we considered two test scenarios. In the first scenario, tagged garments were held in front of the carrier's body; in the second scenario, we put the tagged garments in a shopping bag, more precisely, a booster bag typically used for frozen food. For each of the two test scenarios, we walked 100 times in a straight line through the EAS portal. In the first setting, the antennas with and without reduced signal strength detected all RFID tags. In the second scenario, however, we found that the antennas with reduced signal strength detected only 77 out of 100 RFID tags (i.e., 33% false-negative tag reads and thus potential undetected thefts) whereas antennas with full signal strength detected all garments.

These results indicate that a reduction of signal strength is not an acceptable solution to reduce false positives. Consequently, we retain the antenna power levels at 100% and investigate the applicability of machine-learning techniques to overcome the before-mentioned trade-off between false-negative and false-positive reads.

Test Setting

Figure 2 illustrates our experimental setup. It shows the EAS portal with four far-field antennas and the paths we want to consider to simulate customer movements through and by the gate. For each of the two test scenarios described in the previous section, we want to analyze various distinct movement patterns, combinations of movement paths (e.g., straight movement through the gate or by the gate) and movement speeds (running or walking). Initially we only consider tests with one person and one tagged garment in the antennas' reading field during an individual trial. Therefore, we have to perform 28 single tests to sample data for all seven movement paths (i.e., three paths through the gate and four paths by the gate) in walking and running speeds for the two test scenarios. An error source we identified is shielding of RFID tags by two (or more) customers. In a second step, we therefore want to expand our analysis to more complex tests with multiple persons and tagged garments in the antennas' reading field during an individual trial (e.g., one person walking by the gate with tagged garments while a second person walks through the gate). Following Bottani et al. (2012) who describe 14 different stealing patterns, in a third step we want to analyze more test scenarios (e.g., a tagged garment held close to the body or a tagged item placed at the center of a full bag of untagged items). In addition, we intend to test different types of passive RFID tags.

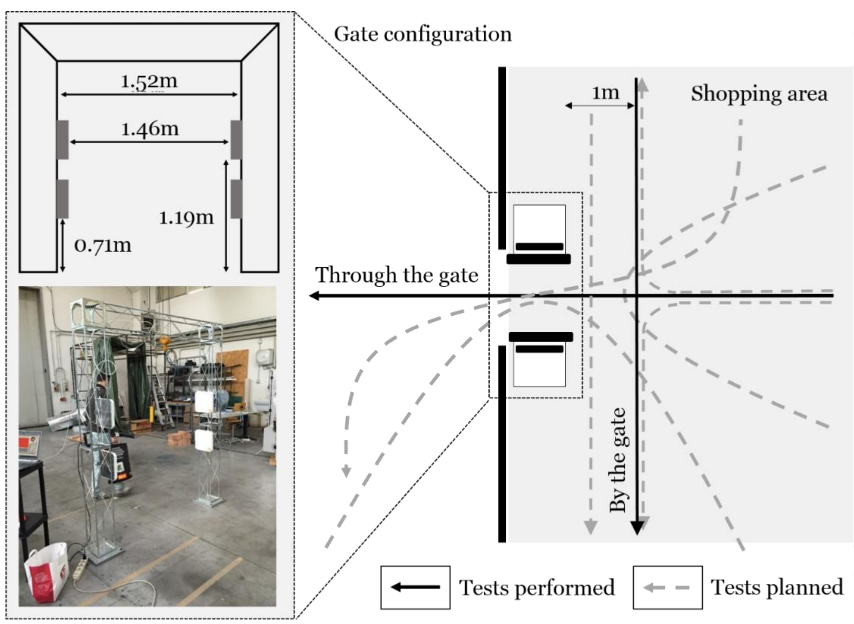

Figure 2. Test setting with typical movement paths in retail stores

For the initial analysis presented in this paper, we considered a subset of the tests with one person and one tagged garment during a trial. We sampled data for both test scenarios, walking speed and two movement paths, walking straight through the gate and walking straight by the gate with 1m distance to the gates. We repeated each of the four resulting tests 150 times. The experimental design yields 600 individual trials resulting in 63,230 tag read events. With the same number of repetitions for the other 24 tests, we will end up with 4,200 trials and more than 400,000 expected tag read events. Beyond that, consideration of more complex scenarios will significantly increase the amount of data.

Figure 3 shows antenna traces from our test data. It illustrates the central challenge of implementing RFID processes in non-standardized environments. In controlled settings with non-concealed tags “by the gate” and “through the gate” readings can reliably be distinguished by means of signal strength. In case the tagged garment is held in front of the carrier’s body the signal strength increases steeply when the person with the garment is approaching the gate. Also the maximum signal strength is considerably higher than in the “by the gate” setting and is reached when the person with the tag is closest to the antennas, that is, when the person is standing right in the middle of the gate. On the other hand, in realistic in-store situations with concealed tags (e.g., tag in a shopping bag), antenna readings become so strongly distorted that this separation becomes impossible.

Figure 3 also highlights the shortcomings of reducing the antenna power to reduce false-positives. For example, if the antenna power is reduced by 40%, tags are not read until they are very close to the gate, which means that for reduced antenna signal strength only the tag reads above the cut-off value in Figure 3 are recorded by the RFID hardware. In the scenario with garments held in front of the carrier’s body (comparable to the setting in upstream RFID processes) the “through the gate” tags are read, the others are not. However, in the setting with garments in a shopping bag, the antennas would still read some of the “by the gate” tags and at the same time miss many of the “through the gate” tags.

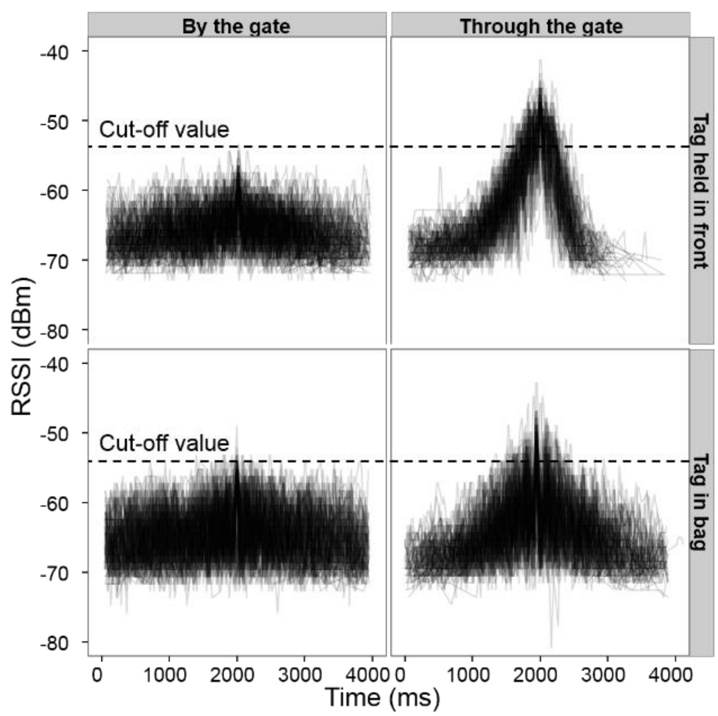

Figure 3. RFID tag signal time series

Conceptual Approach

Correctly classifying tag detections for bagged items is a central challenge of our research. This is a non-trivial task requiring deeper understanding of the structure of the various RFID read events. Table 1 provides an exemplary data excerpt from the raw data in our test data. Each row reflects a single tag read event triggered by one of the antennas. Here, *EPC* is the unique identifier of the RFID tag, *RSSI* is the radio signal's power measured in dBm and *Antenna* is the unique ID of the antenna that read the tag.

Table 1. Exemplary data excerpt					
EPC	Timestamp	RSSI	Frequency	Power	Antenna
0x3032d58e4700000000000002	14162082797	-71	865.7	31.5	2
0x3032d58e4700000000000002	14162082890	-70	865.7	31.5	2
0x3032d58e4700000000000002	14162082979	-72	865.7	31.5	2
0x3032d58e4700000000000002	14162083681	-69	865.7	31.5	2
0x3032d58e4700000000000002	14162084246	-70	865.7	31.5	1
0x3032d58e4700000000000002	14162084418	-70	865.7	31.5	1

In contrast to Figure 3, which does not incorporate information about the antenna ID of single read events, Figure 4 zooms in on the antenna level of two individual test runs. The left panel shows a trial where the tag was moved by the gate with one meter distance, whereas the right panel illustrates a trial where the tag was moved through the gate. In both cases, the tagged garment was held in front of the body. The lines reflect the read events of the two antennas on the left and right side of the gate respectively. We can identify a tempo-spatial characteristic on the antenna level: in the example on the right, the RSSI values measured by all antennas increase steeply until the maximum is met. All antennas exhibit their maximum roughly at the same time – this is the point when the tag is in the middle of the gate. After passing the antennas, the RSSI values decrease even more steeply, given that the body shields the tag while moving away from the gate. Conversely, in the left-hand panel the two sides' RSSI maxima are shifted in time. This can be explained by the fact that the tag moves by the gate in a straight line. Hence, the tag is first closer to one side of the gate and then moves towards the other side.

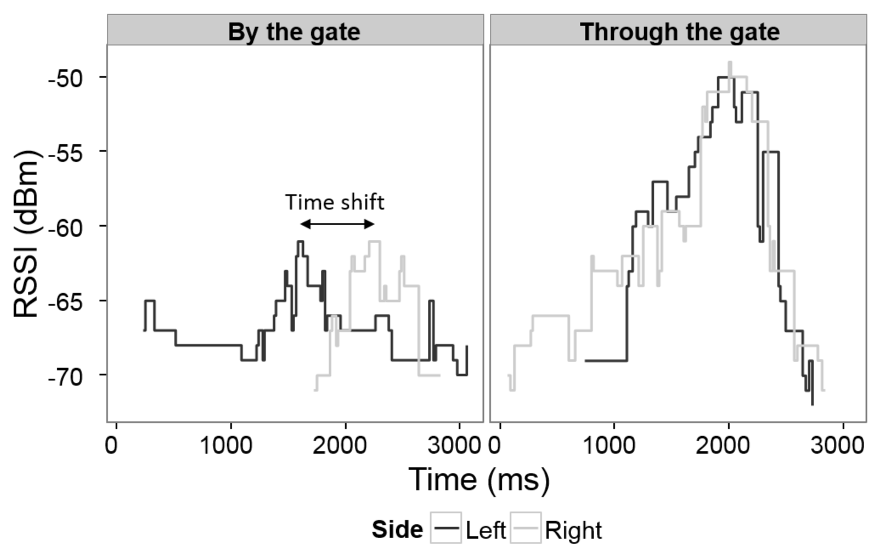

Figure 4. Side-by-side comparison of two RSSI plots

We seek to identify classifiers with high predictive power to achieve robust classification for all scenarios, including the scenario where the tag is located in a shopping bag. To this end, we split, aggregate and recombine the raw test data. This way, we derive additional attributes (predictors) to be considered when training the classifiers. The predictors are not just aggregations of the signal strength. Rather, individual reads are put into temporal relation to one another and are augmented with antenna information.

Table 2 provides a selection of over 40 predictors considered in our classification models. One example is $RSSI_{Max}$. We expect this predictor to be higher for test runs in which the tag moved through the gate, as it is very close to the antennas when it passes the gate, leading to high signal strength measured. Additionally, we measure the maximum RSSI values for the two antennas on the top and bottom respectively. $Antenna_{MaxTimeDispersion}$ is a predictor that captures the temporal relation of individual tag read events. It measures the time between the first antenna maximum and the last during a test run. In consideration of Figure 4, we expect this predictor to have higher values for tags that move by the gate, as the antenna maxima are often shifted in time. On the other hand, the predictor takes small values for tags that were moved through the gate, since the maxima will typically occur simultaneously.

Table 2. Selection of predictors

Predictor	Description	Stump Accuracy
$RSSI_{Sd}$	Standard deviation of RSSI measurements	89.80%
$RSSI_{Diff}$	Difference between the highest and lowest signal strength	86.73%
$RSSI_{Max}$	Maximum RSSI measurement	85.71%
$RSSI_{MaxBottom}$	Maximum RSSI value measured by the antennas on the bottom during the test run	84.69%
ReadCount	Number of tag reads	78.57%
$Antenna_{MaxTimeDispersion}$	Temporal shift between maximum RSSI value of antenna with earliest maximum value and the one with the latest	74.49%
$Antenna_{MaxTimeSd}$	Standard deviation of the time stamps of all four antennas' maximum RSSI measurements	73.47%
$RSSI_{MaxDiffTopBottom}$	Difference between the maximum RSSI values of the antennas on the top and on the bottom	69.39%
$RSSI_{Mean}$	Average signal strength of all reads	68.37%

We generate so-called “decision stumps” for each predictor, measuring classification performance based on only the specific predictor in question. Decision stumps are decision trees with only a single level and a single input feature (Caruana and Niculescu-Mizil 2005). The decision stumps provide us with insights about the relative value of individual attributes in terms of classification accuracy. Some predictors achieve accuracies of over 85%, yet no individual predictor's accuracy reaches 90%. This suggests that we cannot achieve satisfactory results with a single predictor. Yet, the combination of multiple predictors allows us to uncover hidden information within the data and consistently and correctly classify the test instances. Effectively, the combination of different predictors allows us to augment pure signal strength readings with spatial (antenna location information) as well as temporal (timing between sequential read events) information.

We approach the classification problem by first building classification models based on white-box methods. White-box models are comprehensible to humans, while it is very difficult or even impossible for humans to understand the knowledge black-box models use to classify the instances. The white-box models we consider are decision trees (including decision stumps) and logistic regression. On the other hand, the black-box classification models we use are random forests, neural networks and support vector machines.

Preliminary Results

We base our evaluation on the criteria Accuracy, Precision, Recall and F-Measure. Accuracy is the share of correctly classified test runs. Precision is the share of instances classified as “moved through the gate” that actually were moved through the gate. In our application, if we erroneously classify tags that were not moved through the gate as "moved through the gate" (false alarms), precision is diminished. Recall measures the proportion of correctly classified “through the gate” instances. For very conservative classifiers that tend to classify instances as "Not through the gate" in uncertain cases, recall will be low. F-Measure considers both, precision and the recall, and is calculated as the harmonic mean of the two evaluation criteria.

In our tests, we fit our models with 80% of the 600 instances and test them with the remaining 20%. Our classification results are presented in Table 3. The table shows each classifier's accuracy, precision, recall and F-Measure. The best performances of each category are highlighted. While decision stump accuracy was below 90% for all predictors we evaluated, feeding them into machine learning algorithms yields accuracies of over 95% for all classification models considered. The results show that there is no single best classifier. Based on the results below, we consider the black box classifiers (i.e. the random forest, the neural network and the support vector machine) to be the most promising classifiers. With almost 99% accuracy and no false positive tag read, the SVM arguably achieves the best classification result.

Table 3. Classification Results

Classifier	Accuracy	Precision	Recall	F-Measure
Decision Tree	95.92%	95.45%	95.45%	95.45%
Logistic Regression	96.94%	95.56%	97.73%	96.63%
Random Forest	97.96%	97.73%	97.73%	97.73%
Neural Network	97.96%	95.65%	**100.00%**	97.78%
Support Vector Machine	**98.98%**	**100.00%**	97.73%	**98.85%**

Expected Contribution and Future Work

Our initial study considers two customer movement paths at a single movement speed. Going forward, we expand our test setting to more paths (e.g., parabolic movement through the gate) and different movement speeds (walking or running). We want to analyze data not only from trials with one person and one tagged garment in the antennas' reading field but more complex settings as well. Our initial analysis establishes the feasibility of our conceptual approach. Although the results are very promising, we expect that the additional tests will introduce new challenges. Additional predictors may be required to correctly classify trials in which a person runs out of the store or trials in which a person walks very close to the gate but not through the gate. Moreover, the sheer amount of data might require adaptions of the classification models. The test setting's complexity is necessary to ensure real world application. Our ultimate objective is to ensure feasibility of our approach under real-world conditions to facilitate a subsequent roll-out in a real retail environment. This would reflect the deployment step of the CRISP-DM process and effectively close the relevance cycle of the design science research cycle.

In light of our initial findings, we can identify opportunities for further research in various directions. First, real-time classification may increase the chances of catching a shoplifter red-handed. The time required to identify a shoplifter is critical. In our analysis, we derive aggregate attributes for single runs. By doing so, classification is done *after* a tag moved through the gate, which may be too late to catch a thief. Using non-aggregated data on the single read level (or aggregate data only for short-time intervals) may allow us to detect tags in the *very moment* (or shortly after) they pass the gate. Second, the integration of RFID-based EAS in business processes may improve gate performance as contextual information, e.g., if a specific item was sold at the check-out desk, can be taken into consideration when assessing whether or not a specific tag in the reading field of the antennas is stolen. Semantic modeling may be leveraged to extract the latent knowledge in these situations (Brock et al. 2005).

References

Angeles, R. 2005. “Rfid Technologies: Supply-Chain Applications and Implementation Issues,” *Information Systems Management* (22:1), pp. 51–65.

Bai, Y., Wang, F., and Liu, P. 2006. “Efficiently filtering rfid data streams,” in *1st International Very Large Data Base Workshop on Clean Databases (CleanDB)*, pp. 50–57.

Bamfield, J. 2011. *The Global Retail Theft Barometer 2011*, Boughton, Notts: Centre for Retail Research.

Bottani, E., Ferretti, G., Montanari, R., Rizzi, A., and Volpi, A. 2012. “Performances of RFID, acousto-magnetic and radio frequency technologies for electronic article surveillance in the apparel industry in Europe: A quantitative study,” *International Journal of RF Technologies: Research and Applications* (3:2), pp. 137–158.

Brock, D. L., Schuster, E. W., Allen, S. J., and Kar, P. 2005. “An Introduction to Semantic Modelling for Logistical Systems,” *Journal of Business Logistics* (26:2), pp. 97–117.

Brusey, J., Floerkemeier, C., Harrison, M., and Fletcher, M. 2003. “Reasoning about Uncertainty in Location Identification with RFID,” in *Workshop on Reasoning with Uncertainty in Robotics at IJCAI-2003*.

Caruana, R., and Niculescu-Mizil, A. 2005. “An Empirical Comparison of Supervised Learning Algorithms Using Different Performance Metrics,” in *Proc. 23rd Intl. Conf. Machine learning (ICML'06)*, pp. 161–168.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., and Wirth, R. 2000. "CRISP-DM 1.0 Step-by-step data mining guide"

DeHoratius, N., and Raman, A. 2008. “Inventory Record Inaccuracy: An Empirical Analysis,” *Management Science* (54:4), pp. 627–641.

Goller, M., and Brandner, M. 2011. “Improving classification performance of RFID gates using Hidden Markov models,” in *Instrumentation and Measurement Technology Conference (I2MTC)*, pp. 1–5.

GS1 Germany 2007. “Supply Chain Management in the European Textile Industry,” *Technical reports*.

Hevner, A., and Chatterjee, S. 2010. "Design Research in Information Systems" in Integrated Series in Information Systems (22), pp. 9-22.

Jiang, B., Fishkin, K. P., Roy, S., and Philipose, M. 2006. “Unobtrusive Long-Range Detection of Passive RFID Tag Motion,” *IEEE Transactions on Instrumentation and Measurement* (55:1), pp. 187–196.

Keller, T., Thiesse, F., and Fleisch, E. 2014. “Using Dynamic Time Warping to Identify RFID Tag Movement in a Logistics Scenario with and without Additional Process Knowledge,” in *Proceedings of the International Conference on Information Systems (ICIS) 2014*.

Keller, T., Thiesse, F., Kungl, J., and Fleisch, E. 2010. “Using low-level reader data to detect false-positive RFID tag reads,” *2010 Internet of Things (IOT)*, pp. 1–8.

Lee, H., and Özer, Ö. 2007. “Unlocking the Value of RFID,” *Production and Operations Management* (16:1), pp. 40–64.

Liao, W.-P., Lin, T. M., and Liao, S.-H. 2011. “Contributions to Radio Frequency Identification (RFID) research: An assessment of SCI-, SSCI-indexed papers from 2004 to 2008,” *Decision Support Systems* (50:2), pp. 548–556.

Ngai, E., Cheng, T., Au, S., and Lai, K.-h. 2007. “Mobile commerce integrated with RFID technology in a container depot,” *Decision Support Systems* (43:1), pp. 62–76.

Ni, L. M., Liu, Y., Lau, Y. C., and Patil, A. P. 2004. “LANDMARC: Indoor Location Sensing Using Active RFID,” *Wireless Networks* (10:6), pp. 701–710.

Roberts, C. M. 2006. “Radio frequency identification (RFID),” *Computers & Security* (25:1), pp. 18–26.

Thiesse, F., Staake, T., Schmitt, P., and Fleisch, E. 2011. “The rise of the “next-generation bar code”: an international RFID adoption study,” *Supply Chain Management: An International Journal* (16:5), pp. 328–345.

Tu, Y.-J., and Piramuthu, S. 2008. “Reducing False Reads in RFID-embedded Supply Chains,” *Journal of Theoretical and Applied Electronic Commerce Research* (3:2), pp. 60–70.