Transportation Service Quality Improvement through Closed Sequential Pattern Mining Approach

Haisong Huang[1], Liguo Yao[1], Chieh-Yuan Tsai[2,3]

[1]*Key Laboratory of Advanced Manufacturing Technology, Ministry of Education, Guizhou University, Guiyang 550025, China*
[2]*Department of Industrial Engineering and Management, Yuan-Ze University, Chungli 32003, Taiwan*
[3]*Innovation Center for Big Data and Digital Convergence, Yuan-Ze University, Chungli 32003, Taiwan*
Emails: huang_h_s@126.com　　yaoliguo1990@163.com　　cytsai@saturn.yzu.edu.tw

Abstract: *With the improvement of people's living quality, more attention has been paid in food safety and quality. This is especially true for perishable agricultural and dairy products. It is quite often that customers receive poor or broken products due to mistakes or wrong ways in transportation. This leads customers the unsatisfied for companies' products are relatively low. To solve the above problem, this paper proposes a new approach of using frequent closed sequential mining technology to analysis logistics data for helping companies to track the possible transportation problems. The approach consists of several important steps: RFID-enabled raw data collection, frequent sequential patterns mining, and patterns analysis. The experiment shows the proposed analysis method can discover many inside transportation service causes.*

Keywords: *Customer satisfaction; Transportation service; Data mining; Sequential patterns.*

1. Introduction

With the rapid development of economics and Internet, shopping online has become a new trend. It is especially true for China. For example, in Single's Day, the company Alibaba raked in 5 billion dollars during the first 90 minutes of sale, totalling 14.3 billion dollars in just 24 hours through its online shopping platforms [1]. Numerous online merchandises delivery leads transportation an important issue since it might directly influent satisfaction of customers to merchandises. Third Party Logistics providers must possess the capability for continuous change and innovation, be able to develop new, and improve existing processes and services [2].

RFID (Radio Frequency IDentification) is an important communication technology in IOT (Internet Of Things). RFID, identifying specific targets without the need for a specific target recognition system between the mechanical or optical contact to establish, was widely used for logistics monitoring and managing in supply chain management. RFID is a famous auto-identification technology that uses radio frequency waves to identify, track and locate individual physical items [3]. Some researchers used RFID technology in logistics for tracing and tracking cargo transportation information. While reading through RFID will lead to information redundancy, it can generate large data including different kinds of aspect. The results obtained with the analysis and forecasts of the production process can help managers identify deficiencies and lead to improvements.

This paper proposes a method of mining transportation data in logistics and tries to improve the transportation service quality through the mining result. The main components in the proposed method include data collection of RFID-enabled merchandises transportation, frequent sequential pattern mining method for data processing, and decision making analyses.

The rest of this paper is organized as follows. Section 2 reviews previous works related to RFID-enabled transportation and frequent sequential pattern mining. Section 3 introduces the framework of proposed method. Section 4 illustrates a case to verify the feasibility of the proposed system. Lastly, Section 5 summarizes major conclusions and points out future directions of this paper.

2. Literature review

Logistics has become an important part in the global supply chain. Many logistics service providers have realized the importance of adoption of technologies that can help manufacturers, warehouses, and retailers to communicate with each other more efficiently. Among many logistics technologies, RFID has been identified as an important technology to improve logistics operations and supply chain management, and thus is increasingly gaining both practitioners' and researchers' attention [3]. Prasanna and Hemalatha [4] addressed the problems in cargo loaded vehicles such as the delayed delivery of goods, overloading on vehicles, and identification of vehicle location and misplacement of goods. Their proposed system used RFID tag to identify the goods to avoid misplacement. Weight sensor is used to avoid the overloads of goods in vehicle. The GPS and GSM are used to track the vehicle and also send the details of goods to source periodically. Mejjaouli and Babiceanu [5] presented a logistics RFID-WSN model to detect the condition of perishable products as they are moved downstream the supply chain before undesired totally loss of products occurs. The implementation of the proposed model carries important cost reduction prospects. Zhong et al. [6] proposed a holistic Big Data approach to excavate frequent trajectory from massive RFID-enabled shop floor logistics data, which are able to guide end-users to carry out associated decisions. Ting et al. [7] proposed a supply chain Quality Sustainability Decision Support System (QSDSS), adopting association rule mining and Dempster's rule of combination techniques. The aim of QSDSS is to support

managers in food manufacturing firms to define good logistics plans in order to maintain the quality and safety of food products. They conduct a case study of a Hong Kong red wine company in order to illustrate the applicability and effectiveness of QSDSS.

S h i m, C h o i and S u h [8] proposed CRM strategies for small-sized online shopping mall based on association rules and sequential patterns obtained by analysing the transaction data of the shop. H a n, K i m and S o h n [9] applied sequential association rules to extract the failure patterns and forecast failure sequences of ROKAF aircrafts according to various combinations of aircraft types, location, mission and season. W r i g h t et al. [10] applied the CSPADE algorithm to mine sequential patterns of diabetes medication prescriptions both at the drug class and generic drug level and ranked them by the support statistic. S h a w and G o p a l a n [11] concerned finding the frequent trajectories of moving objects in spatial-temporal data by a novel method adopting the concepts of clustering and sequential pattern mining. This approach may be of relevance in finding the frequent paths of migratory birds, or even to predict the next level of pattern characteristics in case of time series data with minor alterations and finding the frequent path of balls in certain games. T s a i and L a i [12] defined a Location-Item-Time (LIT) sequence to describe visitor's spatial and temporal behaviour, developing the Location-Item-Time PrefixSpan (LIT-PrefixSpan) mining algorithm to discover frequent LIT sequential patterns, which route suggestion procedure is proposed to retrieve suitable LIT sequential patterns for visitors under the constraints of their intended-visiting time, favourite regions, and favourite recreation facilities.

Sequential pattern mining is a data mining method for obtaining frequent sequential patterns in a sequential database. Although above researchers studied RFID-enabled logistics and mining logistics data, however, they did not combine two technologies to find transportation problems. This paper presents use of sequential patterns from RFID-enabled food transportation to mining logistics based on the customer's satisfaction, finding the reasons lead lower customer's satisfaction and analyzing the reasons.

3. Research methods

3.1. Framework

Fig. 1 shows the proposed research framework in this paper. With logistics information and transportation sequence dataset, sequential pattern mining algorithm can be applied to obtain frequent sequential patterns. Next step is to conduct decision analysis for find out the reasons of unsatisfied transportation service. If the decision result is not accepted, users can modify parameters in the sequential pattern mining algorithm, and repeat the whole process until satisfied.
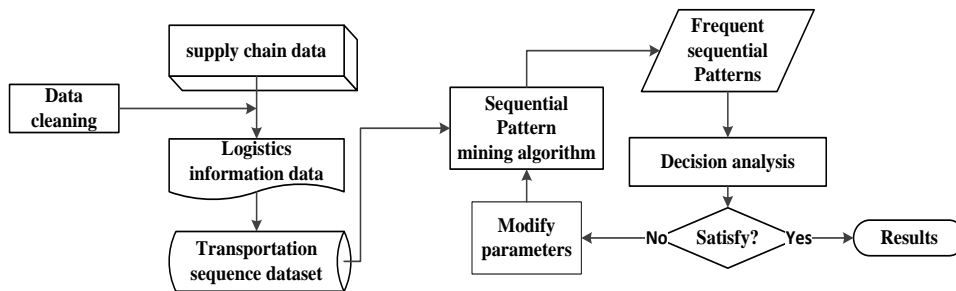
Fig. 1. Research framework

3.2. RFID-enabled logistics data

With RFID tag transportation sequence from suppliers to customers can be traced for each cargo. The transportation route might be consisted of a start city, middle cities, and a destination city, where middle cities might be many. The transportation from the start city to the destination city might result in have different sequences (routes). When cargos move around those cities, RFID reader will receive EPC code, entering time, and location. This tracing process lasts until cargo arrive destination. For example, as shown in Fig. 2, a cargo with RFID tag moves to middle city of A at time T1, then moves to F at time T6, and finally the route end in H at time T5. This route can be represented as sequence $\langle$(cargo1, A, T1), (cargo1, B, T2), (cargo1, C, T3), (cargo1, D, T4), (cargo1, E, T5), (cargo1, F, T6)$\rangle$. This route sequence will be stored into sequence dataset.
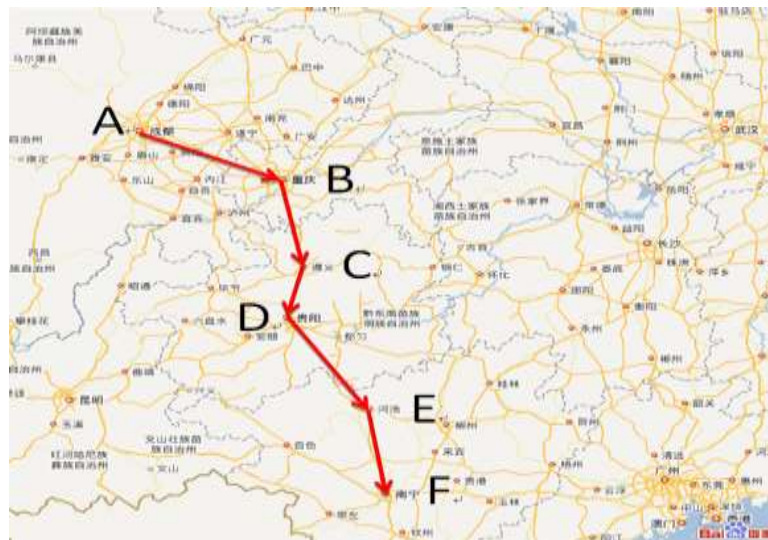

Fig. 2. Example routes of cargo movement

3.3. Closed sequential patterns mining

In this section, we first introduce some preliminary concepts, and then formalize the closed sequential pattern mining problem.

188

Definition 1 (Sequence support). The support of a sequence Y in a sequence database D is the number of sequences in D which contain Y. That is,

$$\text{Sup}(Y) = \big| \{Y | Y \in D\} \big|.$$

Definition 2 (Frequent Sequence). Given the minimum support threshold (min_sup), if the support of sequence Y in sequence database D is not less than min_sup, sequence Y is called a Frequent Sequence, referred to as FS. That is,

$$\text{FS} = \{Y | \text{Sup}(Y) \geq \min\_\text{sup}, Y \subseteq S, S \in D\}.$$

Definition 3 (Frequent Closed Sequence). Let sequences $a = (a_1, a_2, \cdots, a_i)$ and $b = (b_1, b_2, \cdots, b_j)$. If positive integer $m_1, m_2, \cdots, m_i$ where $1 \leq m_1 < m_2 < \cdots < m_i \leq j$ exist and satisfy $a_1 \subseteq b_{m1}$, $a_2 \subseteq b_{m2}, \cdots, a_i \subseteq b_{mi}$, b is called a super-sequence to a. If the support of a sequence is not identical with the support of super-sequence, and is a frequent sequence, the sequence is a Frequent Closed Sequence, referred to as FCS. That is,

$$\text{FCS} = \{Y | Y \in \text{FS}, \ X \notin \text{FS}, \ Y \subset X, \ \text{Sup}(Y) = \text{Sup}(X)\}.$$

There are many algorithms for mining sequential patterns, such as AprioriAll, GSP, FreeSpan, PrefixSpan and others. These algorithms exhibit good performances in databases containing short frequent sequences or the not very low support threshold. However, when long sequences are mined, or when a very low support threshold is used, the performance of such algorithms decreases dramatically and the number of frequent patterns increases sharply, resulting in too many meaningless and redundant patterns. Even worse, sometimes it is impossible to complete the algorithm execution due to a memory overflow.

Closed Sequences mining proposed by X i f e n g, J i a w e i and R a m i n [13] aim to decrease number of sequences beyond threshold and pick out long sequences, to reduce the amount of computing and calculating time. Due to operational performance better than the sequential pattern mining algorithms, closed sequential pattern mining algorithm was adopted by more and more people. Closed Sequential Patterns algorithm (ClaSP) is a new algorithm cited by Antonio G o m a r i z, M a r í n and G o e t h a l s [14], the motivation of ClaSP algorithm is mining frequent closed sequential patterns in temporal transaction data, which has more efficient than well-known algorithm CloSpan. The feature of ClaSP is combining vertical database format strategy and heuristic to prune non-closed sequences to mining closed frequent sequential patterns.

The main steps of ClaSP algorithm are described as below:

(1) Generates a subset of Frequent Sequence (FS) and superset of Frequent Closed Sequence (FCS) called Frequent Closed Candidates (FCC) that is kept in main memory.

(2) Executes a post-pruning phase to eliminate from FCC all non-closed sequences to finally obtain exactly FCS (Frequent Closed Sequences).

Table 1 shows an example database with four sequences and Table 2 shows different algorithms to mining sequences under minimum support 0.6 for the example database.

Table 1. Sample sequences

SID	Sequences
1	(A, B)(A, B)(B, C)
2	(A)(A, B, C)
3	(C)(A)(A, B)(B, D)
4	(D)(A, D)(A, B)

Table 2. Sequential patterns resulting from different algorithms

Frequent sequential patterns	Support count	Frequent closed sequential patterns	Support count
(A)	4	(C)	3
(B)	4	(A)(A, B)	4
(C)	3		
(A)(A)	4		
(A)(B)	4		
(A)(A, B)	4		

3.4. Decision support analysis

The goal of this study is to use RFID-enabled technology to collect logistics data and conduct closed sequential pattern mining algorithms to find significant sequential patterns to identify potential logistics problems in supply chain.

The benefit of using closed sequential pattern mining algorithm is that the algorithm can reduce the amount of frequent patterns and improve the efficiency of data processing. However, choosing an appropriate minimum support value is important when conducting mining. Next, with all sequences, we will compare and observe the difference between sequences with high satisfaction and sequences with low satisfaction. This can be done by comparing different types of sequences and finding the reason why satisfaction of customers decrease.

4. Implementation and experiment results

The environment of following experiments is implemented using PC with Intel Core i7 3.40GHz system with 8G RAM and the analysis software is C++.

4.1. Closed sequential patterns mining

X is an agricultural transaction company in Guiyang, Guizhou Province, western of China. It produces four kinds of agricultural food and dairy products including fruits, vegetables, dairy and meat. The company delivers goods to almost all cities in Guizhou Province and other cities in different provinces. As shown in Fig. 3,

cargos move from start city Chengdu to destination city Nanning and go through 19 middle section cities.

Fig. 3. Different cargo movement routes

City names and city abbreviated codes are listed as follows. For example, Guiyang is city name and GA is city abbreviated code; Guiyang-GA, Liupanshui-GB, Zunyi – GC, Bijie – GF, Anshun – GG, Tongren – GD, Qiannan – GJ, Xingyi – GE, Guilin – KC, Luzhou – CE, Suining – CJ, Hechi – KM, Baise – KL, Chengdu – CA, Chongqing – YA, Guang'an – CX, Nanning – KA, Liuzhou – KB, Neijiang – CK. According to X company's trade and transport situation, 375 sequences were collected as shown in Table 3. 78 sequences belong to low satisfaction and the rest belongs to high satisfaction.

Table 3. 375 Route sequences

Sid	Route sequences	Satisfaction
1	CA-CJ-CX-YA-GD-KB-KA	High satisfaction
2	CA-CK-CE-GB-GJ-KM-KA	High satisfaction
–	–	–
195	CA-CK-YA-GB-GF-GB-GE-KL-KA	Low satisfaction
196	CA-CJ-CK-CE-GF-GA-GE-KL-KM-KA	High satisfaction
–	–	–
374	CA-CJ-CK-CE-GF-GB-GE-KL-KM-KA	Low satisfaction
375	CA-CK-CE-GC-GD-KB-KM-KA	High satisfaction

4.2. Discussion of minimum support

The minimum support in ClaSP mining algorithm is a threshold that will affect mining result frequent sequential patterns accuracy. To know how the support affects the result, the experiment sets the minimum support from 0.1 up to 0.6. Fig. 4 shows the numbers of two kinds of sequential patterns under different minimum supports.
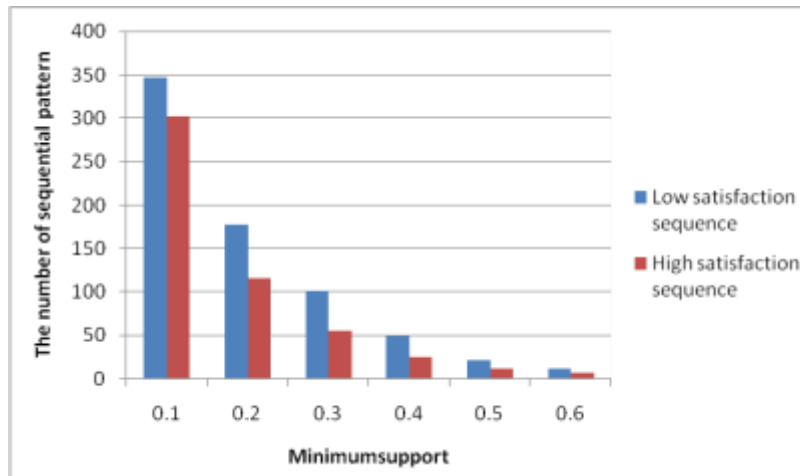

Fig. 4. The number of sequential patterns under different minimum supports

It is clear that, when minimum support increases, the number of closed sequential patterns decrease. If the minimum support value is set as 0.1, it will have 347 high satisfaction frequent sequential patterns and 302 low satisfaction frequent sequential patterns. However, when the minimum support value is set as 0.6, only 12 high satisfaction frequent sequential patterns and 7 low satisfaction frequent sequential patterns can be found. Therefore, based on the observation from Fig. 4, the minimum support value is set as 0.4 in the following experiment.

4.3. Experiment results

There are 49 frequent sequential patterns with high satisfaction and 25 frequent sequential patterns with low satisfaction when minimum support value of ClaSP algorithm is set as 0.4. The statistics of visited count for all middle cities are shown in Table 4. Sup-counts for sequences with high and low satisfaction were shown in Table 5.

Table 4. Visited count for all middle cities

Visited count	City code						
	CJ	CK	YA	CE	GC	KM	KL
In high satisfaction sequences	509	603	588	606	372	362	90
In low satisfaction sequences	134	594	564	103	394	407	88

192

Table 5. Sequences sup-count with high and low satisfaction

Sequences with high satisfaction	Sup_count	Sequences with low satisfaction	Sup_count
CA-CJ-YA-KA	110	CA-CK-YA-KA	98
CA-CJ-YA-CK-CE-KA	93	CA-CK-GC-KM-KL-KA	96
CA-CK-CE-GC-KA	88	CA-CK-YA-GC-KM-KA	83
CA-CJ-YA-CK-CE-GC-KA	84	CA-YA-KM-KA	78
CA-CJ-YA-CK-CE-KM-KA	81	CA-CJ-CK-GC-KA	74
CA-CJ-CE-KL-KM-KA	80	CA-GC-KM-KL-KA	73

With processing and analysing the mining results, three findings can be concluded. As shown in Table 4, visited counts of each city in sequences show the visited counts of CK, YA, GC, KM, KL are almost equal. However, the visited count for CJ and CE cities are very different for the cases of satisfaction and unsatisfaction. That is, CK, YA, GC, KM, KL are frequently visited cities, but these cities did not lead to different satisfaction level. Next, as shown in Table 5, sequences with high satisfaction and sequences with low satisfaction are compared. It is clear that satisfaction of customers will higher when goods go through path YA to CK and KL to KM. In contrast, goods go through path CK to YA and KM to KL will cause low satisfaction for customer. After observing the transportation process, we can find out that storage conditions from path CK to YA may lead goods metamorphism easily. In addition, the road condition from path KM to KL is bad so that merchandises collision during transport is quite often.

5. Conclusions and future work

For past years, RFID technology has been used in transportation to monitoring, tracing and tracking merchandises, which takes advantage of helping companies to optimize supply chain management. Although RFID technology has been applied to inventory management, few researches discuss how to use it to improve the cargo transportation problems. To solve the above problem, this paper proposes a new approach of using frequent closed sequential mining technology to analysis logistics data for helping companies to track the possible transportation problems. An example of 19 cities traffic problems are given to illustrate the way of how to find out the routes leading low satisfaction. The proposed approach can help companies easily identify the causes of transportation service quality problems, such as location, route condition.

R e f e r e n c e s

1. How Alibaba Turned an Obscure, Made-Up Chinese Holiday into a 14.3 Billion Dolllors Shopping Extravaganza that's Bigger than Black Friday. **http://www.businessinsider.com/how-alibaba-made-143-billion-on-singles-day-2015-11**

2. W a g n e r, S. M., R. S u t t e r. A Qualitative Investigation of Innovation between Third-Party Logistics Providers and Customers. – International Journal of Production Economics, Vol. **140**, 2012, No 2, pp. 944-958.

3. R a m a n a t h a n, R., U. R a m a n a t h a n, L. W. L. K o. Adoption of RFID Technologies in UK Logistics: Moderating Roles of Size, Barcode Experience and Government Support. – Expert Systems with Applications, Vol. **41**, 2014, No 1, pp. 230-236.

4. P r a s a n n a, K. R., M. H e m a l a t h a. RFID GPS and GSM Based Logistics Vehicle Load Balancing and Tracking Mechanism. – Procedia Engineering, Vol. **30**, 2012, pp. 726-729.

5. M e j j a o u l i, S., R. F. B a b i c e a n u. RFID-Wireless Sensor Networks Integration: Decision Models and Optimization of Logistics Systems Operations. – Journal of Manufacturing Systems, Vol. **35**, 2015, pp. 234-245.

6. Z h o n g, R. Y., G. Q. H u a n g, S. L a n, Q. Y. D a i, X. C h e n, T. Z h a n g et al. A Big Data Approach for Logistics Trajectory Discovery from RFID-Enabled Production Data. – International Journal of Production Economics, Vol. **165**, 2015, pp. 260-272.

7. T i n g, S. L., Y. K. T s e, G. T. S. H o, S. H. C h u n g, G. P a n g. Mining Logistics Data to Assure the Quality in a Sustainable Food Supply Chain: A Case in the Red Wine Industry. – International Journal of Production Economics, Vol. **152**, 2014, pp. 200-209.

8. S h i m, B., K. C h o i, Y. S u h. CRM Strategies for a Small-Sized Online Shopping Mall Based on Association Rules and Sequential Patterns. – Expert Systems with Applications, Vol. **39**, 2012, No 9, pp. 7736-7742.

9. H a n, H. K., H. S. K i m, S. Y. S o h n. Sequential Association Rules for Forecasting Failure Patterns of Aircrafts in Korean Airforce. – Expert Systems with Applications, Vol. **36**, 2009, No 2, pp. 1129-1133.

10. W r i g h t, A. P., A. T. W r i g h t, A. B. M c C o y, D. F. S i t t i g. The Use of Sequential Pattern Mining to Predict Next Prescribed Medications. – Journal of Biomedical Informatics, Vol. **53**, 2015, pp. 73-80.

11. S h a w, A. A., N. P. G o p a l a n. Finding Frequent Trajectories by Clustering and Sequential Pattern Mining. – Journal of Traffic and Transportation Engineering (English Edition), Vol. **1**, 2014, No 6, pp. 393-403.

12. T s a i, C.-Y., B.-H. L a i. A Location-Item-Time Sequential Pattern Mining Algorithm for Route Recommendation. – Knowledge-Based Systems, Vol. **73**, 2015, pp. 97-110.

13. X i f e n g, Y., H. J i a w e i, A. R a m i n. CloSpan: Mining Closed Sequential Patterns in Large Datasets. – In: In SDM, 2003, pp. 166-177.

14. G o m a r i z, M. C., R. M a r í n, B. G o e t h a l s. CLASP: An Efficient Algorithm for Mining Frequent Closed Sequences. – Advances in Knowledge Discovery and Data Mining, 2013, pp. 50-61.