[1] **Liqin Wang**

Analysis of frequent logistics routes with wireless RFID recognition using datamining

Journal of Electrical Systems

Abstract: - The issue of finding the best path in very frequent through application diagnostics has been efficiently resolved by the technology used in data mining simulation of logistics routes based on wireless RFID identification. This problem is not amenable to other methods for separable eigenbeamformers, including instantaneous indirect trade-offs. In order to enhance logistics job efficiency, it is necessary to construct data mining simulations of common logistics routes that use wireless RFID identification. This would eliminate the issue of logistics route selection.

Keywords: RFID recognition, Frequent paths, Sequential patterns, Data mining

1. Introduction:

Logistics generates a great deal of information, and RFID data is just one kind of that data. In order to extract useful information from logistics route data, this article develops the frequent sequence algorithm FSPMA [1-3]. Managers may make better judgments and traffic planners and designers can benefit from data extracted from algorithm analyses, such as the current path state of vehicle flow and a pattern of future move. Regular pattern mining using RFID route data and its critical importance Looking at the big picture, we can see the distribution of frequent vehicle movements at a macro level, and we can extract key details like the road's distribution at a micro level [4-6]; At the micro level, we can tell if this node is often used by cars, and at the deeper level, we can see which routes are most often used to enter and exit this node. A single node's examination is insufficient to reveal these It is necessary to examine the data from a macro viewpoint.

Constructing a data model for logistics routes to facilitate effective mining

1.1: Common Pattern Mining Ideas

Several related ideas of frequent pattern mining are introduced in this article first for the convenience of description. An itemset is a grouping of things. K itemsets is the name given to a set of k itemsets. As an example, the notation ABC is used to represent a 3-item set that consists of the objects A, B, and C3. Arranging the items in a specific order forms a sequences; the K itemsets is the organization of K things in the itemset. As an illustration, the notation BCA represents a three-sequence that includes three elements from A, B, and C.Backing: The percentage or decimal representation of the number of times a given set of items (sequence) appears in the dataset. Support for itemsets (sequences) must meet a minimal standard. Itemsets (sequences) with support similar or even more than the lowest level are deemed common. Each item in the sequence stands for a different location in the path. As an illustration, three places are denoted by A, B, and C. The path from point B to point C to point A is represented as BCA. A series of paths that satisfy the minimal support condition is called a frequent path. Itemsets (sequences) that have the potential to develop into recurring patterns are called candidate sets. Traditional frequent pattern mining uses a matrix-based data model (1.2). Scanning the dataset is where the frequent pattern mining algorithm spends the most time. The 0-1 matrix representation is

[1] Guangdong Engineering Polytechnic, Guangzhou, China, 510520. E-Mail: 1725692852@qq.com
Copyright © JES 2024 on-line : journal.esrgroups.org

commonly utilized in related studies to enhance the data set's scanning performance. There is a mapping from the data set's items to the matrix's columns, and from the rows to the item sets. A value of 0 here means that the item in question is not now in the list of items being considered, whereas a value of 1 means that it is. In Table 1, the two and three rows correspond to the item sets ACE and BCD, respectively.Each column in items A, B, C, D, and E are mapped in Table 1. By switching to this randomly searchable data structure, the candidate set scan's text search becomes a straightforward logic judgment.Items A and E should exist if the logical AND operation of the column data corresponding to them yields 1; otherwise, they do not; To ascertain whether record ACE in table 1's second row contains elements A and E, all that is needed is this information. Searching the string ACE item by item is not necessary in this case.

Table 1: A specific instance of a data set's 0-1 matrix.

A	B	C	D	E
1	0	1	0	1
0	1	1	1	0

This data model is no longer useful for mining common sequence patterns since sequence information is not merely 0s and 1s; as a result, inefficient string search is typically employed. The logistics path sequence has its own distinct spatial and temporal characteristics, therefore the identical things won't show up in it. In this study, we build a model for logistics data based on this feature. We take traditional logistics data, process it using the 0-1 matrix to make necessary adjustments, and then convert the traditional Radio Frequency Identification path data details on the route sequence data. In addition, it helps with mining datasets of frequent pathways efficiently. The next step is to create a data model that is appropriate for frequent path mining in the logistics industry. After that, you may implement the FSPMA algorithm for route mining.

The conventional data used in logistics is Radio Frequency Identification data, which can be written as {EPC, Location, Time}. Here, EPC stands for Electronic Product Codes, which can be distinct identifier for the item, and the location is the precise location at a given time. There is no representation of time by the standard time here, but rather by numbers. From this set of records, we extract only those whose Locations indicate typical beginning points or transportation nodes. Figure 1 shows a logistics network with these nodes as its representative points: A, B, C, D, and EF. In Figure 1, the edges represent the connections between the nodes. Table 2 provides examples of Radio Frequency Identification data that are relevant to this figure. Furthermore, the following table with a specific item attributes can be created based on mining requirements, and multi-dimensional mining of frequent paths can be executed with path sequence and attribute separation realized. Table 3 shows the item type (KindID) and logistics mass (Weight) as attributes. The path frequency is measured by the logistics mass;a higher weight indicates that the associated route occurs more frequently. The mining process for obtaining the frequent pathways of a specific item type involves combining these two qualities. In an effort to keep things simple, the following tests compared logistical weight alone, ignoring additional factors like item categories, when calculating efficiency.

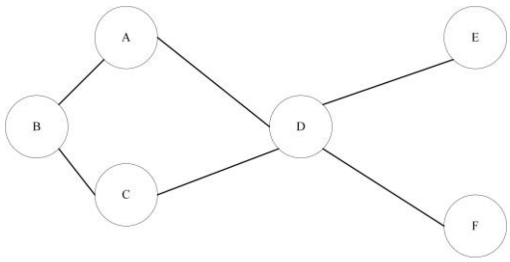

Figure 1: An example of a logistics network topology using RFID data

According to the extent of Time (i.e., the order of time), The values in the Location column are sorted into {epc_x, location_l, location_2,..., location_i} among the records that have the same EPC, where location_l to location_i are tables. The RFID data is classed by EPC. The Location of epc_x is the value of EPC in 2, and it is arranged chronologically. This is the RFID route sequence made up of the locations passed in chronological order and the objects with epe_x as the EPC. This is the RFID route sequence made up of the locations passed in chronological order and the objects with epe_x as the EPC.

2. **Applying data on the structure of the logistics network, the frequent path mining algorithm (FSPMA)**

Based on findings within the logistical network analysis, it is evident that A→B and B→A possess distinct logistical characteristics, and that A→B→C and A→C do not form an inclusion relationship. A⋈C does not have an offspring in A→B→C. Sequence is an additional irrelevant path; therefore, frequent logistics path mining differs from traditional association rule mining. This method is unique in sequence pattern extraction. Due to the dissimilarity between AB and BA, the number of candidates k sequence produced as A_n^k as compared to C_n^k (n is the total number of network nodes.) in the absence of pruning. Fortunately, within this particular network, the pruning method can also generate an algorithm corresponding to the Apriori feature, taking into account the attributes of the logistics network.

The conventional Apriori-like algorithm for frequent sequence mining consists primarily of two alternating steps: 1) Obtain candidate K sequences, which will be reduced in number by pruning algorithms pertinent to the candidate K sequences; 2) Scan In order to ascertain the frequency of the candidate sequence, the data set is utilised to identify the frequent K sequence. Subsequently, the process proceeds to step 1) to acquire the candidate K+1 sequence, and so forth, until no further sequences remain. Among them, the outcome of step 2) is utilised from step 1) onwards. The candidate K+1 sequence is pruned in accordance with the criterion that any sub-pattern of the frequent pattern is frequent; it is not feasible to eliminate the K+1 sequence from the frequent sequence. Clearly, step 2) is the more time-consuming of these two; however, step 1) has a direct bearing on the extent to which this step consumes time. Step 2 is less time-consuming when the sequences produced in step 1 are less prevalent. The logistics network possesses distinct attributes that are primarily manifested in two facets when frequent routes are mined.

1) In the topological network, Since they are both nearby nodes in the sequence, they must also be similar. As an illustration, the AC sequence in Figure 1 cannot occur; only ABC or ADC can. Therefore, unlike traditional Apriori or other sequence pattern mining techniques, it is possible to generate frequent K+1 sequences from frequent K sequences. The equivalent K+1 sequences can be joined to produce the corresponding K+1 sequences, provided that two K sequences that satisfy the conditions exist. The K-1 sequence made up of the initial K-1 nodes of one sequence is precisely the subsequence of the other sequence, with the first node excluded. Hence, these two sequences are satisfied. For instance, you can connect ABC and BCD if they occur frequently. Obtain the ABCD candidate sequence;

2) The logistics network's topology data can be utilised to exclude candidates for frequently occurring sequences. In theory if a path is frequent between two points, then all paths between the two points having a cost that is equal to or less than the frequent path cost are likewise regular because logistical prefers to choose the path with the least cost. There are one or more minimal cost paths connecting two places in the logistics network. Whether the path is frequent or not, the undesirable path won't be frequent since the cost is too high compared to the minimum cost. The particular excess is deemed excessive and calls for a particular examination of a particular set of problems.

In this study, the "cost tolerance" parameter (TD) for the path sequence Rij in the logistics network from node i to node j is defined. The proportion of the price of Rij to the least expensive path from nodes i to nodes j is represented by the symbol TD. It is evident that the ratio is higher than 1 to 1. The highest limit of the path's cost tolerance that satisfies the requirements can be established during the candidate path sequence generation process every sequence of paths whose cost limitation beyond the higher bounds are eliminated. For instance, any path sequences whose expenses is more than 20% of the minimum expense will be trimmed if the upper limit is set to 1.2.

The fundamental topology data of the network of logistics, which includes the matrix of adjacency and node count. N×N is the matching matrix of adjacency

neib, assuming that there are N nodes in the network. Neib is a N×N two-dimensional array used in the algorithm's implementation, where node i and node j are represented as neib[i][j]. The cost or adjacency distance of j. It indicates that nodes i and j are not nearby if neib[i][j] = 0. Any two locations in the adjacency matrix can have their distance measured using the shortest path algorithm. The computation output is a matrix that is called the matrix of minimal costs

in this instance, and it is comparable to the adjacency matrix. The N×N two-dimensional array least is used in the algorithm implementation to represent the smallest cost matrix. The cost of the shortest road between nodes i and j is represented by least[i][j], and if zeas[i][j] equals zero, It suggests that there is not a route connecting nodes i and j. The cost tolerance pruning technique is the name given to the candidate sequence pruning technique that is presented in this paper's introduction to FSPMA. Firstly, ascertain the cost of the path that the candidates sequences represents. Secondly, compute the tolerance of cost parameter (which is the product of TD) and the lowest cost of the two locations that are connected to the path's ultimate destination. This is pruning procedure. Assume that the route is the one that goes from node I to node J, as indicated by equation (1):

$$\text{Ultimate} = \text{least}[i][j] \times TD \tag{1}$$

The final step involves comparing the cost to the ultimate, which is the highest allowable cost of path from node i to node j. The related candidate sequence must be eliminated from the candidate sequence if cost > ultimate since it does not match the standards.

Here's the way a PMWTI algorithm is explained:

Input: Basic topology details in the logistics network, minimum support SUP, cost tolerance parameter TD, and logistics path data collection.

Output: A list of all commonly used pathways satisfying the SUP support degree.

Step 1 reads the data set, creates a matrix using the formulas shown in table 5, and its stores the result in the memory.

Step 2 is to read the logistics network's basic topology information, get the matrix of adjacency (neib) and the number nodes (N), and then compute the less cost matrix (LCM) based on the neib.

Step 3: Acquire potential K sequences. The mining process stops if the The set of potential K sequences is null.; if not, step 4 is carried out. The set of all nodes in the logistical network serves as the sequence of canditate for the first execution of this step; if it is not executed for the first time, obtain the preliminary candidate k sequence using the previously discussed cost tolerance pruning method after obtaining the frequent K-1 connection of sequence in step 4.

Step4: To identify frequent sequences from candidate sequences, step 4 scans the matrix from step 1. The mining process concludes if the necessary frequently series set is not filled in; if it is,

move on to step 3.

PMWTI provides a least matrix of cost where n is the quantity of nodes. This matrix can be calculated and stored for O(n3) and O(n2), respectively, in terms of time and space. In general, regular use of route mining expenses the difference is minimal unless the bare minimum of support is set unnecessarily high, which would instantly stop the algorithm and mining patterns. This mining is, of course, worthless.

2. Experimental analysis

There is only one run of the experiment. There are 1,206,480 records in the data collection that was used for the experiment. The Every record has a mean mass of 3438.7, and the total logistics weight of all the records is 480800030. Using 0.001 as the minimal support as an example, that is A route must have a total logistical weight more than or equal to 480800030×0.001 in order to be considered frequent. First, run two algorithms on the data set. The primary is the FSPMA technique that this paper proposes; the second is the classic Apriori-like mining algorithm, which consists of the description of the FSPMA algorithm plus the TRADITIONAL component that employs topological information for pruning. The outcomes of TRADITIONAL mining with the same minimal support are contrasted with the mining results obtained with FSPMA. The cost tolerance numbers used by FSPMA are 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, and 2.0. A minimum of 0.001, 0.003, o.005, 0.007, 0.01, 0.03, 0.05, and 0.07 are the support degrees. In order to guarantee consistency between the mining results of the two methods, the outcome of the comparison is displayed in Figure 2, where the ordinate represents the minimal cost tolerance and the abscissa represents the minimal support. The mining results of FSPMA with a cost tolerance of 1.6 and above are consistent with the mining results of TRADITIONAL, whereas the mining results of 1.5 and below are inconsistent with the mining results of TRADITIONAL, for instance, when the minimum support is 0.001. From 0.001 to 0.07, the longest common path that was found from the shortest supportive grade is 9, 8, 7, 6, 5, 3, 1, and 1 in that sequence, according to an analysis of the mining data. Since practically all paths become common paths and the minimal support is too low in relation to the experiment's data set, the mining findings are worthless, which explains why the minimal The degree of price acceptance is as high as 1.6 when the minimal support is 0.001. The benefits of FSPMA cannot be fully utilised in this situation. The influence of FSPMA is strongly expressed when the minimum support is raised from 0.001 to 0.003, which also results in a decrease in the minimum cost tolerance from 1.6 to 1.3. In other words, when the minimum support is slightly raised and the excavated frequent pathways have a specific significance. The minimum cost tolerance steadily decreases to 1.0 as the minimum support increases. This implies that, at a given minimum support level, either the often dug paths are the lowest paths of cost or no path satisfies the minimal support. For instance, the mined nodes are all single nodes rather than sequence of path when minimal support is between 0.05 and 0.07.

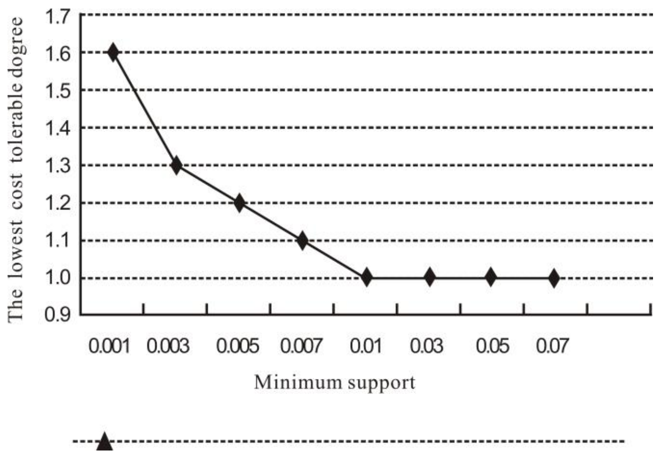

Figure 2: The range of the minimum cost tolerance under various minimum support levels.

Figure 3 compares the efficiency of FSPMA with TRADITIONAL methods. The abscissa represents the minimal support degree, the ordinate represents the excavation time (in seconds), and the tolerance of cost is FSPMA in 1.6. Without a doubt, FSPMA is much more efficient than TRADITIONAL with any level of assistance. When mining logistics data for regular routes, TRADITIONAL performs terribly in terms of efficiency since it ignores any logistics network topology information. The two cannot be compared. Following examination, it is discovered that the creation of the sequence of applicant 2 is where TRADITIONAL's primary issue resides. Given that the candidate 2 sequences in TRADITIONAL are derived from frequent 1 sequence by permutation, n (n-1) sequences of candidate 2 will be produced for every n frequent 1 sequences. However, because the sequence of path in the logistical network includes the attribute of path and any two adjacent sites in the path must be adjacent in the logistics network topology, some of the sequences of candidate 2 cannot appear in the sets of data (here A sequence made up of all non-adjacent to nodes cannot exist in the logistics data set; in this case, a sequence that is adjacent to inside the sequence but not adjacent within the logistical network is considered illegal). This sequences is referred to as a legal sequence. Basic topological knowledge can be introduced to fix this issue.This article presents the third approach, which involves adding the matrix of adjacency parameter neib to TRADITIONAL and using neib to judge when creating the sequence of candidate 2 ij if node i and node sequence of Candidate 2 ij should be added to sequence of candidate 2 if j is not contiguous. Because the candidate K sequence (K>2) is connected based on the legal sequence, and all sequences obtained through the connection of the legal sequence are legal sequences, this judgment is added only during the sequence of candidate 2 creation. During the generating process, the candidate sequence naturally acquires this pruning effect; hence, the procedure is called NATURAL. When compared to TRADITIONAL, NATURAL is more efficient and more in line with FSPMA. As a result, TRADITIONAL is no longer included in the comparative data provided below, which only contains the experimental data of NATURAL and FSPMA.

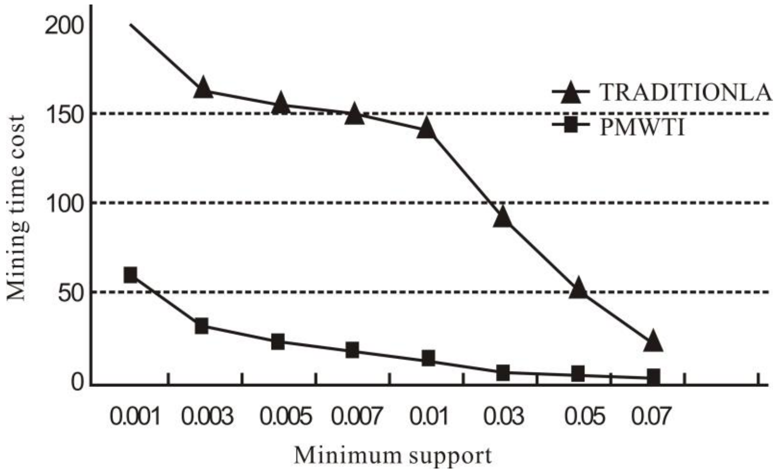

Figure 3: Evaluating the effectiveness of FSPMA and TRADITIONAL methods with different levels of minimal assistance.

Initally, as indicated in Figure 4, compare FSPMA with NATUAL when the minimal support is 0.003. The tolerance of cost is represented by the abscissa, while the ordinate shows the amount of time the method took in seconds. Because NATURAL lacks this parameter—the cost tolerance is only applicable to FSPMA—it takes a constant 37.328 seconds in Figure 4. Figure 4 shows how the time-consuming FSPMA gradually approaches NATURAL with continuous improvements in cost tolerance; in contrast, modern price tolerance is too low, despite the fact that FSPMA's mining efficiency is extremely high in contrast, and the mining output can be compare with those of conventional mining algorithms. Choosing an acceptable cost tolerance is crucial because the outcomes are not consistent.

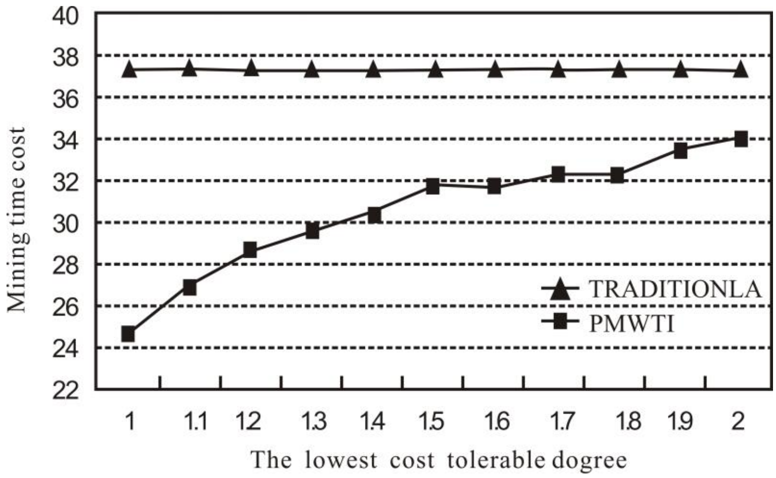

Figure 4: Analyzing the mining efficiency of FSPMA and NATURAL under various costlimits.

3. Conclusions

This study aims to create a relevant logistics data model and using the FSPMA mining method to effectively extract regular route using logistics data, taking into account the features of the data. The mining algorithm and the data model operate independently of one another. Space and time are exchanged at the foundation of the data model. The cost tolerance pruning technique, or FSPMA algorithm, designs overall potential route sequences pruning technique. When compared to the same technique without the pruning algorithm, this strategy decreases the amount of data that needs to be scanned and significantly increases mining efficiency.

Acknowledgement

2022 China Society of Logistics Research Project "Simulation Research on Frequent Route Data Mining in Logistics Relying on Wireless RFID Identification" (No. 2022CSLKT3-167).

References

[1] Lapkin A A , Heer P K , Jacob P M , et al. Automation of route identification and optimisation based on data-mining and chemical intuition[J]. Faraday Discussions, 2017, 202(26):1-10.

[2] Nezihe Yıldıran, Tacer E. Identification of photovoltaic cell single diode discrete model parameters based on datasheet values[J]. Solar Energy, 2016,4(6):101-111.

[3] Cho J S , Jeong Y S , Park S O . Consideration on the brute-force attack cost and retrieval cost: A hash-based radio-frequency identification (RFID) tag mutual authentication protocol[J]. Computers & Mathematics with Applications, 2015, 69(1):58-65.

[4] Chen Y , Zhou F , Liu X , et al. Online adaptive parameter identification of PMSM based on the dead-time compensation[J]. International Journal of Electronics, 2015, 102(7-9):1132-1150.

[5] Janosovsky J , Danko M , Labovsky J , et al. Software approach to simulation-based hazard identification of complex industrial processes[J]. Computers & Chemical Engineering, 2018, 122(4):66-79.

[6] Nguyen C T , Bui A T H , Nguyen V D , et al. Modified Tree-based Identification Protocols for Solving Hidden-Tag Problem in RFID Systems over Fading Channels[J]. Iet Communications, 2017, 11(7):1132-1142.

[7] Ganjar Alfian et al., Integration of RFID, wireless sensor networks, and data mining in an e-pedigree food traceability system. Journal of Food Engineering,November 2017, *https://doi.org/10.1016/j.jfoodeng.2017.05.008*

[8] Yahia Zare Mehrjerdi.,RFID-enabled supply chain systems with computer simulation,April 2009,Assembly Automation 29(2):174-183. DOI:10.1108/01445150910945624

[9] Abad, E., Palacio, F., Nuin, M., Zárate, A. González de, Juarros, A., Gómez, J.M and Marco, S., 2009. RFID smart tag for traceability and cold chain monitoring of foods: demonstration in an intercontinental fresh fish logistic chain, Journal of Food Engineering, 93 (4) , pp. 394–399.