IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 26, 2020, accepted March 8, 2020, date of publication March 16, 2020, date of current version March 25, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2980952

Association Rule Mining Method Based on the Similarity Metric of Tuple-Relation in Indoor Environment

NAIXIA MOU[1,2], HONGEN WANG[1,2], HENGCAI ZHANG[2], AND XIN FU[3]

[1]College of Geomatics, Shandong University of Science and Technology, Qingdao 266590, China
[2]State Key Laboratory of Resources and Environmental Information System, IGSNRR, CAS, Beijing 100101, China
[3]School of Water Conservancy and Environment, University of Jinan, Jinan 250022, China

Corresponding authors: Naixia Mou (mounaixia@163.com) and Xin Fu (stu_fux@126.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFB0502104, and in part by the National Natural Science Foundation of China under Grant 41701521 and Grant 41771476.

ABSTRACT Association rules can detect the association pattern between POIs (point of interest) and serve the application of indoor location. In this paper, a new index, tuple-relation, is defined, which reflects the association strength between POI sets in indoor environment. This index considers the potential association information such as spatial and semantic information between indoor POI sets. On this basis, a new R-FP-growth (tuple-relation frequent pattern growth) algorithm for mining association rules in indoor environment is proposed, which makes comprehensive use of the co-occurrence probability, conditional probability, and multiple potential association information among POI sets, to form a new support-confidence-relation constraint framework and to improve the quality and application value of mining results. Experiments are performed, using real Wi-Fi positioning trajectory data from a shopping mall. Experimental results show that the tuple-relation calculation method based on cosine similarity has the best effect, with an accuracy of 87%, and 19% higher than that of the traditional FP-growth algorithm.

INDEX TERMS Association rule, data mining, indoor trajectory, network embedding.

I. INTRODUCTION

As one of the core tasks in the field of data mining, association rules [1] can get the association patterns between different data itemsets in the data set, so as to find out the internal commonness of user behavior habits [2]. At present, association rules have been widely used in recommendation system [3], [4], auxiliary business decision-making [5], [6] and other scenes. The core of association rules is to detect the interdependency and association between one or more than one item and other items. In the indoor environment, occupying more than 70% of human activities [7]–[9], users move and consume between indoor POIs(point of interest), reflecting the behavior and purchase habits of indoor users. The application of association rule algorithm in indoor environment can capture the interdependence of indoor POI, which plays an important role in indoor personalized information recommendation, indoor marketing strategy formulation and indoor user behavior pattern mining. For example, the association rule "Nike→Adidas" indicates that the shop POI "Nike" and "Adidas" in the indoor shopping mall has a strong association, and also reflects a shopping tendency of the user group, i.e., the users who have visited Nike tend to visit Adidas. Based on such association rules, shop information recommendation can be realized, together with user shopping pattern mining, shopping mall marketing strategy formulation and shop address selection as well as other indoor location-based services, so as to better user's shopping experience.

Faced with the massive indoor trajectory data that has exploded with the continuous development of wireless sensors and indoor positioning technologies [10]–[12], the indoor original trajectory is transformed into a semantic trajectory [13] containing the indoor POI semantic name that the user has visited. The association rule algorithm can detect the association patterns among different POI sets based on semantic trajectories, and find out the association rules among different POI sets that reflect the behavior habits of users, such as shopping. Therefore, it has a wide application prospect in indoor location-based services.

The associate editor coordinating the review of this manuscript and approving it for publication was Jing Bi.

The strength of association between indoor POIs is determined by many factors. To measure the strength of association between indoor POIs, multiple information should also be taken into consideration. Traditional association rule algorithms based on support-confidence constraint framework [14], such as the classic Apriori algorithm [5], DHP algorithm [15] and FP-growth algorithm [16], only use co-occurrence frequency and conditional probability between POI sets to measure the association strength of itemsets when mining the association rules between different POI sets in indoor semantic trajectory set, while ignoring the potential association information of spatial proximity trend and semantic trend between POIs in indoor environment. This leads to the fact that many weak association rules are involved in the mining results, which have low application value in indoor location-based service.

In view of this, this paper, by adding a network embedding algorithm [17] to its calculation process, puts forward a new constraint condition of tuple-relation, node2vec [18], considering the semantic information such as category attribute, price interval and spatial distance information between indoor POIs, synthesizing multiple potential association between indoor POIs, so as to better measure the association strength between indoor POI sets in association rules. By adding tuple-relation constraint to traditional association rule algorithm FP-growth, a new R-FP-growth (tuple-relation frequent pattern growth) association rule mining algorithm based on support-confidence-relation constraint framework is formed. When measuring the association strength of shop POI in rules such as ''Nike → Adidas'', we first consider the co-occurrence probability and conditional probability of users visiting two shops, and then take the semantic and spatial information such as whether two shops belong to the same clothing category or sports category, whether the price ranges of two shops are similar, and the distance between two shops as the side information. This comprehensive measurement of the association strength between shops from various aspects greatly improves the quality of association rule mining results and their application value in indoor location-based services.

The contributions of this paper are as follows.

(1) A new concept 'tuple-relation' is proposed to enhance the traditional support-confidence constraint framework of association rules. It represents a new constraint condition that incorporates multiple potential association information between indoor POIs.

(2) Based on this new measure, a novel indoor association rule mining algorithm R-FP-growth is designed to improve the quality and application value of the mining results of association rules in indoor environment, through considering both co-occurrence probability and conditional probability.

(3) The proposed approach is evaluated by using a real, continuous indoor trajectory, which includes 1,713,130 indoor trajectories and 25,794,539 location points over a period of two days. Results reveal the advantages of our approach compared to traditional FP-growth algorithm.

The rest of this paper is arranged as follows: Section 2 summarizes the related work; Section 3 illustrates the concept of tuple-relation and proposes R-FP-growth algorithm; Section 4 reports the experimental results and data analysis. Finally, Section 5 concludes the proposed approach, and presents several aspects of future work.

II. RELATED WORK

Given a data set, the problem of mining association rules can be transformed into the process of finding out the association rules that meet the minimum support and confidence threshold, which is mainly divided into two parts: (1) generating frequent itemsets, i.e., finding out the itemsets that meet the minimum support threshold; (2) generating association rules, i.e., finding out the rules that meet the minimum confidence threshold in frequent itemsets [19]. The former part focuses on how to improve the efficiency of generating frequent itemsets, while the latter part focuses on how to improve the quality and application value of generating association rules.

In order to improve the efficiency of generating frequent itemsets in association rule algorithm, Park *et al.* [15] propose an algorithm to generate frequent itemsets based on hash technology. The algorithm uses the cost of generating hash table and storage space of data set in exchange for the improvement of algorithm performance. Savasere [20] propose a partition-based algorithm that scans the database only twice during execution, significantly improving the performance of the algorithm; Han *et al.* [16] propose a FP-growth mining method that finds frequent itemsets without generating candidate sets. In the process of generating frequent itemsets, a divide-and-conquer strategy is used to generate suffix itemsets from the bottom up and construct a conditional FP-tree, which greatly improves Efficiency of generating frequent itemsets. Zaki and Gouda [21] propose an algorithm based on diffset data structure, dECLAT, which tracks only the difference of candidate patterns in producing frequent patterns, which greatly reduces the memory needed to store intermediate results. Uno *et al* [22] propose an algorithm LCM based on enumeration of reserved extensions, which is efficient and one of the most important ones so far [23].

In order to improve the quality and application value of association rule mining results, researchers have done a lot in this area. Agrawal and Srikan [24] introduce the concept of interest index to find rules of interest to users; Klemettinen *et al.* [25] use the form of rule template to prune redundant and unintentional rules; Zhong and Liao *et al.* [26] introduce the concept of weighted dual confidence to reduce a large number of unintentional results and find interesting negative association rules; Wei *et al.* [27] propose a new matching method to replace the confidence constraints of traditional association rules algorithm, and the rules generated by the improved method are highly correlated; Zaki [28] proposes a new constraint framework based on the concept of closed itemsets, and finally proves the new framework can reduce association rules with low value; Wei *et al.* [29]

add corresponding weight to frequent itemsets in the process of generating association rules, putting forward the method of weighted association rules for the first time with better results. Although the above methods have achieved good results, yet they are not specifically for the improvement of special indoor environment.

At present, there are few researches on association rules in indoor environment. This paper focuses on how to improve the quality and application value of the results generated by association rules algorithm in indoor environment. Firstly, a constraint condition of tuple-relation is proposed to fuse the potential association information such as semantics and space between indoor POI sets. Then, a new R-FP-growth algorithm is proposed by combining tuple-relation with FP-growth algorithm, which greatly improves the association and application value of association rule mining results in indoor environment.

III. METHODOLOGY

Trajectory data has the characteristics of large data volume, fast update frequency, and low value density, resulting in low mining efficiency and unsatisfactory results based on original trajectory. Therefore, Alvares *et al.* [13] propose the concept of semantic trajectory, smaller in volume, higher in quality, and better in reflecting the user behavior than the original trajectory. In this paper, the association rules between indoor POIs are explored based on indoor semantic trajectory. For easier understanding, the method proposed in this article will be explained in an indoor shopping mall environment. POI in shopping mall mainly refers to one shop. The relevant definitions of semantic trajectory and association rules are given as follows:

Definition 1 (Trajectory): A trajectory $traj = \{pt_i\}_{i=1}^{n}$ is a sequence of point $pt_i = \left(id, x_i, y_i, t_i, f_i\right)$ in chronological order, where id, x_i, y_i, t_i, f_i respectively represent the unique identification of the user, longitude, latitude, time of acquisition and floor identification of the point.

Definition 2 (Stay Point): The stay point $sp = (s, f, t_{in}, t_{out}, x, y)$, as shown by the black point in Fig. 1, represents the POI that the user stays for more than a certain time, where $s$and f respectively represent the semantic name and floor identification of POI, t_{in} and t_{out} respectively indicate the time when the user enters and leaves POI, and $x = \sum_{i=1}^{n} x_i/n$ and $y = \sum_{i=1}^{n} y_i/n$, represent the mean of longitude and latitude of trajectory points in POI within a certain time range.

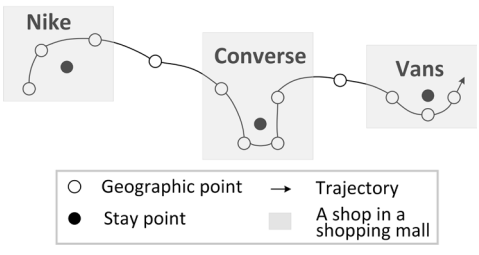


FIGURE 1. An example of semantic trajectory.

Definition 3 (Semantic Trajectory): The semantic trajectory $st = \{s_i\}_{i=1}^{n}$ is a sequence of POI semantic names with n stay points in chronological order. The user's original trajectory in Fig. 1 is represented by POI semantic name sequence <Nike, Converse, Vans>. Semantic trajectory set $ST = \left\{traj_i\right\}_{i=1}^{n}$ is a set consisting of all user semantic trajectories.

Definition 4 (Itemset): The total itemset $TI = \left\{s_j\right\}_{j=1}^{n}$ is a set of n unique POI semantic names. Itemset $I = \left\{s_j\right\}_{j=1}^{m}$ is a subset of the total itemset TI, where $m<=n$. Frequent itemsets are itemsets whose support meets the minimum support S_{min}.

Definition 5 (Association Rule): Set the itemset I_X, I_Y, where $I_X \subseteq TI$, $I_Y \subseteq TI$, and $I_X \cap I_Y = \emptyset$, the association rule is a logical implication of the form $I_X \rightarrow I_Y$ that satisfies the minimum confidence threshold C_{min}, i.e., I_X occurs first and I_Y occurs later. Among them, I_X and I_Y are also called the pre-itemset and post-itemset of association rules. A rule that has not yet been determined to satisfy C_{min} is a candidate rule.

Definition 6 (Support): The support of the association rule $S(I_X \rightarrow I_Y) = N(I_X \cup I_Y)/N(All)$ represents the co-occurrence probability that the user has been to the pre-itemset and post-itemset in the association rule, where $N(I_X \cup I_Y)$ represents the number of users who simultaneously contain I_X and I_Y in ST, and $N(All)$ represents the total number of users.

Definition 7 (Confidence): The confidence of the association rule $C(I_X \rightarrow I_Y) = N(I_X \cup I_Y)/N(I_X)$ is the conditional probability that the user first goes to the set I_X and then goes to the set I_Y, where $N(I_X \rightarrow I_Y)$ is the number of users including I_X and I_Y in ST, and $N(I_X)$ is the number of users including I_X in ST.

Definition 8 (Indoor POI Network): Indoor POI network $G = (V, E)$ is a directed weighted network constructed on user's semantic trajectory. V is the set of nodes in the network, each representing a POI. E is the set of directed edges between nodes in the network, each connecting two orderly nodes, from the first to the second, indicating the order in which user accesses POI. W_{ij}, the weight of the directed edge of node v_i pointing to node v_j, represents the total number of times the user first goes to the POI represented by node v_i and then goes to the POI represented by node v_j. Fig. 2 is a simple example of an indoor shop network generated by users' semantic trajectories in an indoor shopping mall. A, B, C, D, E, F represent six different shops, among which the weight of the directed edge between D and A in Fig. 2 (b) is 2, because both User 1 and User 3 first visit D and then visit A.

A. TUPLE - RELATION

Trajectory is generated by a user's special behavior activities between indoor POIs. In the mall, the user performs shopping behavior between different shops for consumption purposes, the semantic or spatial features of shops such as spatial distance, shop type, and prices of shop items affect

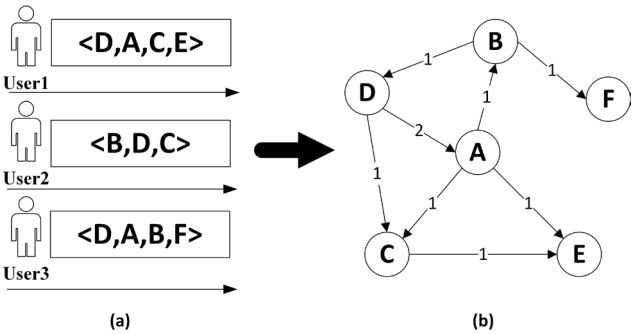


FIGURE 2. An example of shop network generated by users' semantic trajectories of shopping mall. (a) User semantic trajectories of shopping mall; (b) Shop network.

the shopping behavior of users and are reflected in user trajectory. The features of shops can reflect the association between shops. In order to extract multiple features of shops, and to comprehensively measure the association strength between shop sets in the candidate rules, this paper first converts user's original trajectory into user's semantic trajectory, and then builds a shop network based on user's semantic trajectory. Here, the network embedding algorithm node2vec is introduced, which defines the concepts of network homogeneity and structural equivalence. Shops with the same homogeneity or structural equivalence have semantic or spatial potential associations. Node2vec can automatically learn multiple features of a shop through a neural network and embed the shop into a feature vector matrix in the form of feature vectors, and then calculate the feature vector corresponding to the shop set based on the feature vector matrix. And finally, the tuple-relation between the feature vectors is calculated based on the cosine similarity. The greater the tuple-relation, the stronger the association strength. Given an association rule of $I_X \rightarrow I_Y$, the calculation equations of the tuple-relation are shown in (1)-(4):

$$\begin{cases} I_X = \{I_{x\varepsilon}\}_{\varepsilon=1}^{v}, & v \in N^* \\ I_Y = \{I_{y\varepsilon}\}_{\varepsilon=1}^{u}, & u \in N^* \end{cases} \tag{1}$$

$$\begin{cases} \overrightarrow{V_{x\varepsilon}} = M\left(I_{x\varepsilon}\right) \\ \overrightarrow{V_{y\varepsilon}} = M(I_{y\varepsilon}) \end{cases} \tag{2}$$

$$\begin{cases} \overrightarrow{V_X} = \{\dfrac{\sum_{\varepsilon=1}^{v} \overrightarrow{V_{x\varepsilon}}(i)}{v}\}_{i=1}^{m}, & m \in N^* \\ \overrightarrow{V_Y} = \{\dfrac{\sum_{\varepsilon=1}^{u} \overrightarrow{V_{x\varepsilon}}(i)}{u}\}_{i=1}^{m}, & m \in N^* \end{cases} \tag{3}$$

$$\mathrm{R} = \frac{\sum_{k=1}^{m} \overrightarrow{V_X}(k) \times \overrightarrow{V_Y}(k)}{\sqrt{\sum_{k=1}^{m} \overrightarrow{V_X}(k)^2} \times \sqrt{\sum_{k=1}^{m} \overrightarrow{V_Y}(k)^2}} \tag{4}$$

In Equation (1): $I_{x\varepsilon}$ represents the shop name, I_X and I_Y represent the set of multiple shop names; v and u represent the number of shops in the set. In Equation (2): $\overrightarrow{V_{x\varepsilon}}$ and $\overrightarrow{V_{y\varepsilon}}$ represent the feature vector corresponding to each shop name $I_{x\varepsilon}$; $M\left(I_{x\varepsilon}\right)$ represents the feature vector corresponding to the shop name from the shop feature vector matrix M. In Equation (3): $\overrightarrow{V_X}$ and $\overrightarrow{V_Y}$ represent feature vectors corresponding to I_X and I_Y shop sets, which is obtained by summing and averaging the corresponding vectors of the shops in the shop set; m represents the vector dimension. In Equation (4): the tuple-relation R is obtained by calculating the cosine value between the two vectors $\overrightarrow{V_X}$ and $\overrightarrow{V_Y}$; $\overrightarrow{V_X}(k)$ and $\overrightarrow{V_Y}(k)$ respectively represent the k-th value of the vector. The value range of R is $[-1, 1]$: when $\mathrm{R} = -1$, the angle between the two vectors is 180 degrees, which is opposite in the vector space, indicating that the association between the two vectors is the weakest; when $\mathrm{R} = 0$, the angle between the two vectors is 90 degrees, which is orthogonal in the vector space, indicating that the association between the two vectors is weak; when $\mathrm{R} = 1$, the angle between the two vectors is 0 degree, and the directions are the same in the vector space, when the association between the shop sets represented by $\overrightarrow{V_X}$ and $\overrightarrow{V_Y}$ is the strongest.

The feature vector matrix M of the shop in Equation (2) is an important part in the calculation of the tuple-relation. In order to obtain the matrix M shown in Fig. 3 (a), the constructed shop network $G = (V, E)$ is taken as the input of node2vec algorithm. The second-order random walk method is used to sample the adjacent nodes of a node, then the random walk sequence training set of the shop nodes shown in Fig. 3 (b) is obtained. In the process of representing shop nodes as feature vectors, network homogeneity and structural equivalence of shop nodes are preserved. Finally, the skip-gram neural network model shown in Fig. 3 (c) is used to train the training set to obtain the matrix M containing the feature vectors of all shop nodes in Fig. 3 (d).

In the process of using skip-gram neural network model to extract shop features, the problem is defined as the maximum likelihood optimization problem. Given the network $G = (V, E)$, let $f: V \rightarrow R^d$ be the mapping function of node to feature representation, d the dimension of network embedding, and f the matrix of size $|\mathrm{V}| \times \mathrm{d}$. For each node $u \in V$, $N_{s(u)} \in V$ is defined as the neighborhood node generated by the second-order random walk of node u.

The optimization goal of feature learning is to maximize the probability that nodes in the neighborhood will appear on the condition that the node is u. F (u) represents the mapping function of node u mapped to the feature vector.

$$\max_{f} \sum\nolimits_{u \in V} logPr(N_S(u)|f(u)) \tag{5}$$

To make the optimization problem simpler, two hypotheses are made:

(1) Hypothesis of conditional independence. Assuming that the neighboring nodes of node u appear independently of each other, the optimization goal will become:

$$\Pr\left(N_S\left(u\right)|f\left(u\right)\right) = \prod\nolimits_{n_i \in N_S(u)} \Pr(n_i|f(u)) \tag{6}$$

(2) Hypothesis of feature space symmetry. It is assumed that nodes are symmetrical to each other, and all nodes are under the same space. Since it is under the same

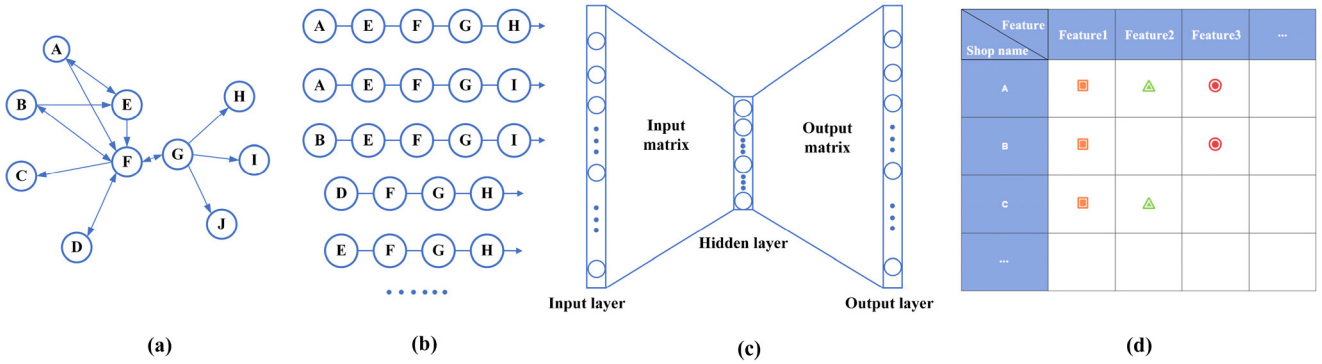


FIGURE 3. Calculation process of shop feature vector matrix M. (a) Shop network; (b) Random walk sequences; (c) Skip-gram neural network model; (d) Shop feature vector matrix M.

feature space, the conditional likelihood of a node and its neighboring nodes will be transformed into the following form:

$$\Pr(n_i|f(u) = \frac{\exp(f(n_i)\cdot f(u))}{\sum_{v\in V}\exp(f(v)\cdot f(u))}) \tag{7}$$

Based on these two hypotheses, the final objective function f can be optimized as follows:

$$\max_{f} \sum_{u\in V}[-logZ_u + \sum_{n_i\in N_S(u)} f(n_i)\cdot f(u)] \tag{8}$$

$Z_u = \sum_{v\in V}\exp(f(u)\cdot f(v))$ in Equation (8) is a normalization factor, which is optimized by using a negative sampling technique. According to the final objective function f, a shop feature vector matrix M is generated. In fact, for the same shop network, the shop feature vector matrix M can be given in advance, and the process of using the Skip-gram neural network model to extract the shop feature vector matrix M can be carried out just once.

B. R-FP-GROWTH

Mining association rules in a data set is actually a process of finding out association rules that satisfy constraints. Traditional FP-growth algorithm first mines frequent itemsets in the data set, then generates candidate rules based on frequent itemsets, and finally generates association rules based on confidence constraint. R-FP-growth algorithm introduces the network embedding algorithm node2vec to generate the shop feature matrix M, and then combines the tuple-relation generated by M with traditional support-confidence constraint framework to form a new support-confidence-relation constraint framework, This way can remove the rules that meet the support and confidence constraints but that have weak association, so as to improve the accuracy of the algorithm. The overall flow of the R-FP-growth algorithm is shown in Algorithm 1.

In the process of filtering candidate rules with the new constraint framework, as is shown in Equation (9), the confidence C of the candidate rule is first calculated, followed by determining whether or not the confidence C of the candidate rule is greater than the minimum confidence C_{min}. If not, it indicates that the rule's credibility is not high, and delete it directly; if yes, continue to calculate the tuple-relation R of the candidate rule, and then determine whether or not the R is greater than the minimum tuple-relation R_{min}. If not, it means that the rule is not strongly associated and delete it directly; if yes, it means that the rule has strong credibility and association, and has practical application value, keep the rule.

$$\begin{cases} C < C_{min} \rightarrow Delete \\ C \geq C_{min} \text{ and } R < R_{min} \rightarrow Delete \\ C \geq C_{min} \text{ and } R \geq R_{min} \rightarrow Save \end{cases} \tag{9}$$

Algorithm 1 Indoor Association Rule Mining Algorithm

Input: Semantic trajectory dataset: ST
Support threshold: S_{min}
Confidence threshold: C_{min}
Tuple-relation threshold: R_{min}
Hyper parameters of node2vec: network, d
Output: Shop association rules: AssociationRules
function R-FP-growth (ST, S_{min}, C_{min}, R_{min}):
// *Build fptree*
1: fptree = fptreeEstablish (ST, S_{min})
// *Mine shop frequent itemsets*
2: frequentItemsets = fpMine (fptree , S_{min})
// *Calculate the shop feature vector matrix M*
3: M = node2vec (network, d)
4: for itemset in frequentItemsets:
// *Generate candidate rules*
5: candidateRules = crulesMine (itemset)
6: for candidateRule in candidateRules:
// *Calculating candidate rule confidence C*
7: C = confidenceCalculation (candidateRule)
// *Filter candidate rules*
8: if $C >= C_{min}$:
// *Calculate candidate rule tuple-relation R*
9: R = tuple-relationCalculation (candidateRule, M)
10: if $R > R_{min}$:
11: AssociationRules.append (candidateRule)
12: return AssociationRules
13: end R-FP-growth

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. DATA AND PROCESSING

The experimental part of this paper uses real indoor trajectory point data set and vector data of a shopping mall in Jinan City, China. The trajectory point data set is collected through the Wi-Fi sensor in the shopping mall. Once the user's smartphone is connected to the Wi-Fi in the shopping mall, the Wi-Fi fingerprinting positioning method [30] will obtain the user's position. The mall shop vector data contains a total of 489 shops. The Wi-Fi positioning trajectory data covers eight floors of the mall from 23 December 2017 to 24 December 2017, with a total of 1,713,130 user trajectories and 25,794,539 trajectory points. The data format of Wi-Fi positioning trajectory points is shown in Table 1. Each column in the table represents the user's unique identification, sampling time, X coordinate, Y coordinate, and the floor where the user is located. The user trajectory points are connected in chronological order to form the user trajectory. An example of user trajectory points, trajectory, and shop vector data is shown in Fig. 4.

TABLE 1. Trajectory point data format details.

ID	Time	X (m)	Y (m)	FloorID
000004F***	20171223153941	13021***	43904***	1
000004F***	20171223153945	13021***	43904***	1
……	……	……	……	……
000004F***	20171223154008	13021***	43904***	2
000004F***	20171223154013	13021***	43904***	2

Note: In order to protect the privacy of the user, the user's ID and XY coordinates are represented by ***.

Because the indoor environment is very complicated, the positioning sensor signals are easily affected by reflection, occlusion, and attenuation, resulting in insufficient accuracy, incomplete recording, and positioning errors [31]. In order not to affect the experimental results, we clean the following types of data:

(1) Meaningless user trajectories, i.e., users with too few trajectory points and too short stay in the mall.
(2) Time anomaly points, such as abnormal trajectory points with a sampling interval of 0 and beyond normal business hours of the mall.
(3) Spatial anomaly points, i.e. the trajectory points that deviate from the normal space of the mall.
(4) Floor anomaly points, e.g. the floor of a point is the second floor in one second, but it reaches the fifth floor in the next second, contrary to common sense.

The total number of trajectories after data cleaning is 29008, and the total number of trajectory points is 171534. We use this as a basis to cite Indoor-STDBSCAN algorithm in [32], [33] to identify the user's stay points, and then perform semantic position matching to obtain 10521 user semantic trajectories, then remove the semantic trajectories with less than 5 sequences, and finally get 8063 User semantic trajectories, detailed information is shown in Table 2.

TABLE 2. Examples of user semantic trajectory.

User ID	User Semantic Trajectory
000AF5A***	Adidas, Dairy Fairy, Cabbeen Urban, Love Happy…
000AF5C***	Taro, Flower & Bake, Love In Colors, MUJI…
000C020***	MLB, C&A, YIERYI, Gozo…

Note: In order to protect the privacy of the user, the user's ID are represented by ***.

B. EVALUATION METRICS

The effectiveness of the association rule algorithm requires quantitative indicators for evaluation. As is shown in Equations (10) and (11), accuracy and lift [34]–[36] are

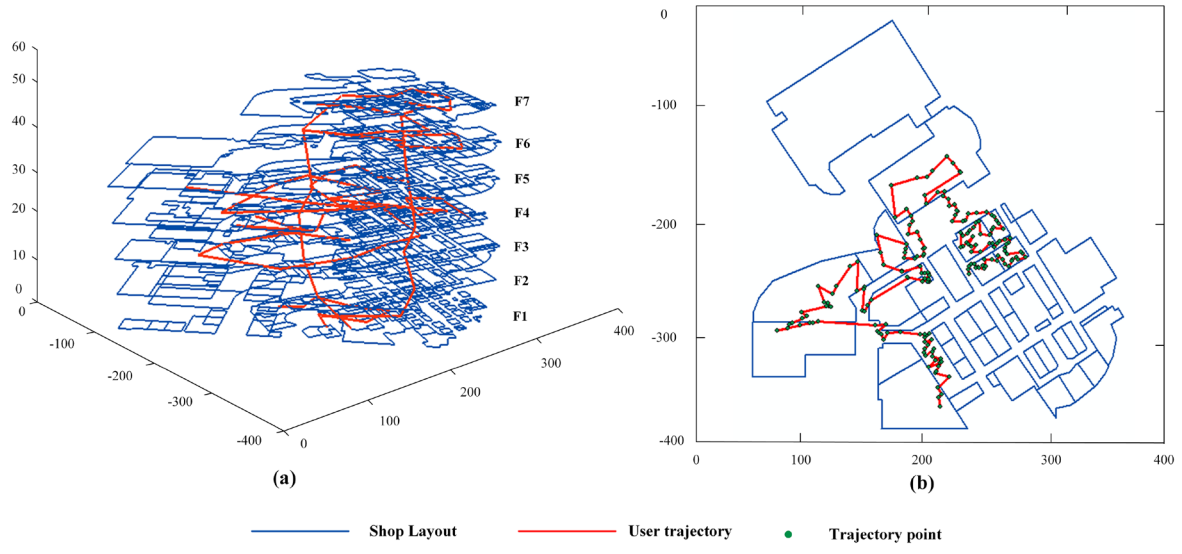


FIGURE 4. User trajectory examples. (a) An entire trajectory of a user; (b) A part of a user's trajectory on one floor.

used to evaluate R-FP-growth algorithm proposed in this paper. Here, lift appear as one of the two most promising quality measures to be used as metrics in association rule mining problem [36], which has been widely used to evaluate association rule mining results [37]–[40].

$$\text{accuracy} = \frac{N_{true}}{N_{all}} \tag{10}$$

$$\text{lift}(I_X \rightarrow I_Y) = \frac{P(I_X I_Y)}{P(I_X)P(I_Y)} \tag{11}$$

In Equation (10): Accuracy indicates the accuracy of the association rule algorithm, i.e., the ratio of the number of valid rules N_{true} to the total number of rules N_{all} in the association rule mining result, which uses lift to determine whether a certain rule is valid. In Equation (11): I_X and I_Y are the pre-itemset and post-itemset of association rules. $P(I_X)$ represents the probability of I_X occurrence, $P(I_Y)$ represents the probability of I_Y occurrence, and $P(I_X I_Y)$ represents the probability of I_X, I_Y occurring simultaneously. Lift is actually the ratio of the probability of occurrence of I_Y when I_X occurs to the probability of occurrence of I_Y without considering any factors. It reflects the degree of interaction between I_X and I_Y. The threshold of lift is $[0, +\infty]$. When the lift of an association rule is 1, it indicates that I_X and I_Y of the association rule do not affect each other, hence such a rule is almost meaningless. When the lift is less than 1, it indicates that the occurrence of I_X will inhibit the occurrence of I_Y, and the application value is low. When the lift is greater than 1, it indicates that the occurrence of I_X will promote the occurrence of I_Y, which is the rule required by most application scenes. Therefore, this paper regards association rules with a lift less than or equal to 1 as invalid association rules, and those with greater than 1 as valid association rules.

C. VARIABLE ESTIMATION

The hyper-parameters of R-FP-growth algorithm mainly include the minimum support S_{min}, the minimum confidence C_{min}, and the minimum tuple-relation R_{min}. In order to determine the optimal hyperparameters of the algorithm, this paper uses the control variable method in order to obtain the optimal parameter value combination of the algorithm.

Let the R_{min} be 0 and observe the relationship between S_{min} and the accuracy of the algorithm under different C_{min} values, where $S_{min} \in \{0.002, 0.004, 0.006, 0.008, 0.01, 0.012\}$, and $C_{min} \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$. As is shown in Fig. 5, under different C_{min}, as S_{min} increases, the accuracy generally shows a downward trend; but when C_{min} is 0.4 or 0.5, the curve has a small upward trend in the S_{min} band of 0.004 $\sim$ 0.006. The accuracy of the algorithm is the highest when $S_{min} = 0.006$ and $C_{min} = 0.5$, so the R-FP-growth algorithm is set to $S_{min} = 0.006$ and $C_{min} = 0.5$.

After determining the minimum support S_{min} and the minimum confidence C_{min}, In order to select the appropriate value of the minimum tuple-relation R_{min}, we continue to observe the impact of the R_{min} on the accuracy of R-FP-growth algorithm. As is shown in Fig. 6, with the increase of R_{min},

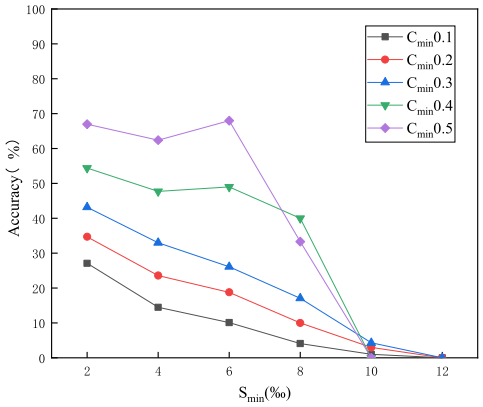


FIGURE 5. Accuracy variation of R-FP-growth under different C_{min}.

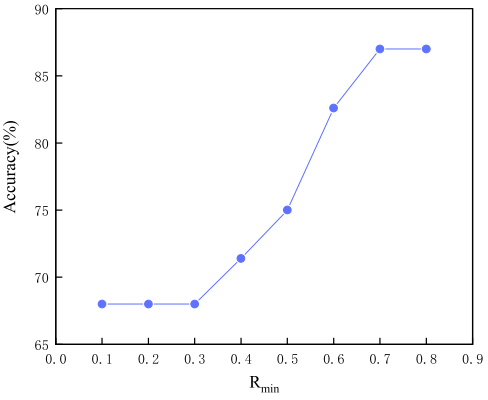


FIGURE 6. Accuracy variation of R-FP-growth.

the accuracy curve of the algorithm presents a trend of first stable, then rising, and finally stable change, and reaches the highest point when $R_{min} = 0.7$, so the minimum tuple-relation $R_{min} = 0.7$, the three optimal parameters of S_{min}, C_{min} and R_{min} of R-FP-growth algorithm are {0.006, 0.5, 0.7}.

D. COMPARISON OF TUPLE-RELATIONS BASED ON FOUR DIFFERENT SIMILARITIES

The cosine similarity calculation method used in this paper is compared with the three common vector similarity calculation methods of Euclidean distance, Manhattan distance, and Chebyshev distance, to verify the validity of the cosine similarity-based tuple-relation in this paper. The tuple-relations calculated based on four different similarities are represented by R_C, R_ED, R_MD, and R_CD, respectively.

Set the minimum support S_{min} to 0.006, and observe the influence of four different tuple-relation on the accuracy of the mining results of R-FP-growth algorithm under different minimum confidences C_{min}, where the $C_{min} \in \{0.3, 0.4, 0.5\}$. Fig. 7 (a) shows the effect of R_C on the accuracy of R-FP-growth algorithm. Among the different C_{min} values, changes in each curve are generally consistent. As R_{min} increases, each curve as a whole shows a trend of stability first, then rises, and then stabilizes, and rises to the highest

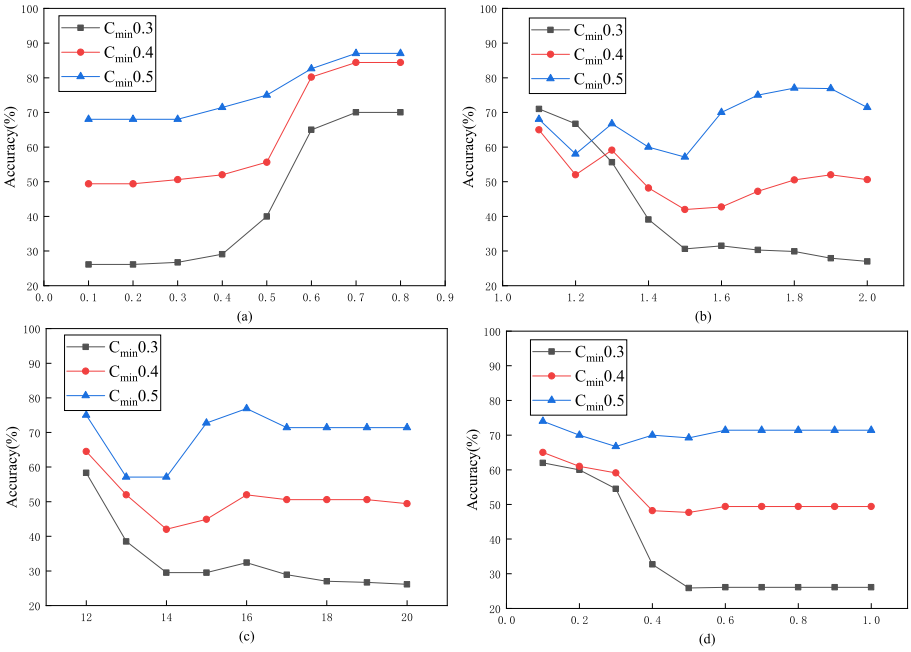


FIGURE 7. Effect of different tuple-relation on R-FP-growth accuracy. (a) Effect of R_C on algorithm accuracy; (b) Effect of R_ED on algorithm accuracy; (c) Effect of R_MD on algorithm accuracy; (d) Effect of R_CD on algorithm accuracy.

point when $C_{min} = 0.5$ and $R_{min} = 0.7$. Fig. 7 (b) and (c) show the effects of R_ED and R_MD on the accuracy of R-FP-growth algorithm. The curves in the two images fluctuate greatly, but the overall curves first decline, then rise, and then tend to be stable, both reaching the highest point under ($C_{min} = 0.5$, $R_{min} = 1.8$), ($C_{min} = 0.5$, $R_{min} = 16$) respectively. Fig. 7 (d) shows the effect of R_CD on the accuracy of R-FP-growth algorithm. As R_{min} increases, each curve in the image first decreases, then stabilizes and obtains the highest accuracy when $C_{min} = 0.5$ and $R_{min} = 0.1$.

After determining the optimal parameters of the algorithm under four different tuple-relations, the accuracy obtained by R-FP-growth algorithm based on four different tuple-relations is compared, as is shown in Fig. 8. the accuracy of R-FP-growth algorithm based on four kinds of tuple-relation is 87%, 77%, 76.9%, 74%, respectively.

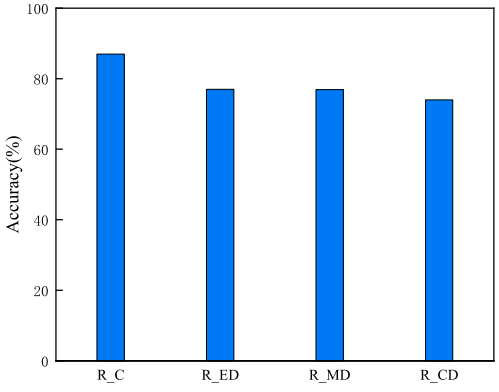


FIGURE 8. Comparison of the accuracy of four tuple-relation.

Experimental results show that the choice of different tuple-relation calculation methods will affect the accuracy of the algorithm. We choose the tuple-relation calculation method based on cosine similarity, because the relation is calculated based on the feature vector matrix generated by the network embedding algorithm node2vec, and the inner product of the vector is used in the process of generating the feature vector matrix. So, using cosine similarity of vectors to calculate the degree of correlation is theoretically the most reasonable way. When comparing the effects of four different tuple-relation calculation methods on cosine, European distance, Manhattan distance, and Chebyshev distance on the accuracy of the algorithm, the tuple-relation calculation method based on cosine similarity obtains an accuracy of 87%, significantly higher than that of the other three calculation methods, which also proves the rationality of the tuple-relation calculation method based on cosine similarity.

E. COMPARISON WITH FP-GROWTH

In order to verify R-FP-growth algorithm to improve the quality of association rule results, R-FP-growth algorithm is compared with traditional FP-growth algorithm to prove the superiority of the former.

Set the minimum support S_{min} of the two algorithms to 0.006, and then compare the accuracy of FP-growth algorithm when the minimum confidence C_{min} is 0.3, 0.4 and 0.5 with that of R-FP-growth algorithm. As is shown in Fig. 9, the accuracy of FP-growth algorithm when C_{min} is 0.3, 0.4 and 0.5 is 43.2%, 54.4% and 68%, respectively, which are lower than the accuracy of R-FP-growth algorithm by 87%. It is

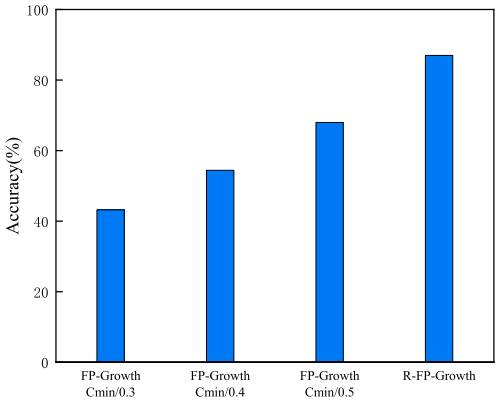


FIGURE 9. Comparison of accuracy between R-FP-growth and FP-growth.

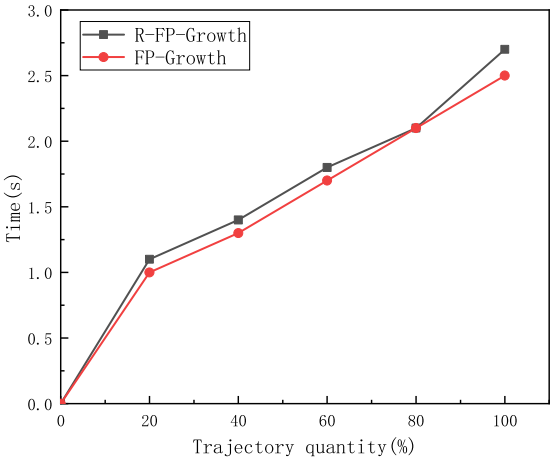


FIGURE 10. Comparison of running time between R-FP-growth and FP-growth.

proved that superiority of R-FP-growth algorithm. At the same time, it is found that the accuracy change of FP-growth algorithm is the same as that of R-FP-growth algorithm in Fig. 5 when the tuple-relation is 0, which explains why we set the minimum support S_{min} to 0.006.

FP-growth algorithm based on support-confidence constraint framework only considers the co-occurrence and conditional probabilities between the sets of shops, yet cannot well measure the association between the sets of shops, which causes the algorithm to have more weakly associated rules in the results of mining shop association rules, and that is why FP-growth algorithm gets lower accuracy in experiments. However, R-FP-growth algorithm is based on FP-growth algorithm, and considers multiple potential spatial and semantic association information between the set of shops, which can better measure the strength of association between association rules. Therefore, the invalid rules with low association strength are filtered out, and the accuracy of the algorithm is significantly improved.

In order to compare and analyze the running time of this method more fully, we divide the data set with 8063 semantic trajectories into 5 data sets with different volumes and different number of trajectories, with the number of trajectories occupying 20%, 40%, 60%, 80% and 100% respectively. As is shown in Figure 10, the ordinate represents the running time of the algorithm and the abscissa represents the number of tracks in the data set as a percentage of the number of tracks in the original data set. It can be seen from the figure that the running time of the R-FP-growth algorithm is slightly increased compared with the FP-growth algorithm, but the running time of the two algorithms on the whole is not much different.

Because the R-FP-growth algorithm adds the tuple-relation calculation part on the basis of the FP-growth algorithm, it will increase the time complexity of the algorithm, which makes the operation time of the algorithm slightly increase compared with the FP-growth, and the operation efficiency is reduced, but the loss of this little running efficiency will result in a significant improvement in the quality of the mining results, which we think is worthwhile. It has to be noted that the efficiency of the algorithm does have room for improvement, which is also the need for further improvement in our future work.

F. DISCUSSION

In special indoor environment, the spatial and semantic association information between POIs will greatly affect the behavior of users. Traditional association rule algorithm based on support-confidence constraint framework only uses co-occurrence frequency and conditional probability between POI sets to mine the association rules. It cannot well measure the strength of the association between indoor POI sets and therefore it is not applicable to the mining of association rules in indoor environment. Meanwhile, there are few studies on improving the quality of association rule mining results in indoor environment. Therefore, this paper proposes R-FP-growth algorithm, which takes into account co-occurrence and conditional probabilities between indoor POI sets, and introduces tuple-relation constraint. In the process of calculating tuple-relation, the network embedding algorithm is introduced to extract multiple potential spatial and semantic association information between indoor POIs, so that the algorithm can better measure the strength of associations between indoor POI sets, significantly improving the accuracy of algorithm Rate and quality of mining results. Because the algorithm considers the spatial and semantic association information between POIs, it also greatly increases the application value of association rules in indoor location-based services.

This paper mainly describes R-FP-growth algorithm in indoor shopping mall scene, but the algorithm is also applicable in other indoor scenes such as hospitals or airports. In different indoor scenes, although the types of POIs are different and the user's purpose is different, the spatial and semantic information between POIs will affect user behavior and be reflected in user trajectories. The method in this paper can be used to extract the spatial and semantic features between POIs, to calculate the tuple-relation between POIs, and finally

to generate POI association rules in different indoor scenes. It should be noted that the optimal parameter combination of support, confidence and tuple-relation may be different in different indoor scenes, so parameters need to be adjusted in different indoor scenes. And because the method in this paper is mainly designed for indoor scenes, it may not be applicable to other outdoor scenes.

The tuple-relation constraint proposed in this paper can be combined not only with FP-growth algorithm, but also with other association rule mining methods. The reason why we chose the FP-growth algorithm is that it operates with relatively high efficiency, and that the calculation process of tuple-relation has low coupling with other modules of the algorithm. Therefore, tuple-relation can be combined with association rule algorithms such as Apriori and DHP to mine association rules in indoor environment to achieve equally good results, with difference in calculation efficiency.

The POI feature matrix is obtained through a skip-gram neural network in our article; therefore, for the same data set, although the vector matrix obtained each time through the neural network may be different, the distance between the vectors of each node is relatively stable because the objective function optimized by the neural network at each time the matrix is generated is the same. Our final goal is to compute the distance between nodes through the vector matrix, so the process of using the neural network to generate the node feature vector matrix for the same data set is only once, and the vector matrix can be fixed beforehand. In fact, for the indoor place A, with a relatively rich data set, we only need to get the feature vector matrix of the POI node of place A according to this data set in advance, and we only need to use this matrix when calculating the distance or similarity between the nodes in the subsequent calculation, and we do not have to extract the matrix again. But for indoor place B, it is necessary to re-extract its POI node feature vector matrix according to the data set of place B, and the matrix obtained from the data set of place A can no longer be used.

V. CONCLUSIONS AND FUTURE WORK

This paper proposes a new method for mining POI association rules based on tuple-relation in indoor environment. This method can mine indoor POIs with strong association based on the semantic trajectory of indoor users and serve indoor location applications. In order to better measure the association strength of indoor POIs, after considering potential association information such as spatial proximity trend and semantic trend between indoor POIs, the concept of tuple-relation is proposed for the first time by combining a network embedding algorithm with an association rule mining algorithm. Then, in order to improve the quality of association rule mining results and their application value in indoor location-based services, a new R-FP-growth association rule mining algorithm is proposed by combining tuple-relation with traditional association rule mining algorithm FP-growth. Finally, experiments are performed on a real indoor Wi-Fi positioning trajectory dataset. Experimental results prove the effectiveness of the proposed method and the rationality of the tuple-relation calculation method.

In future work, the following aspects will be further studied: (1) to further validate the proposed algorithm using more types of indoor data (such as airport trajectory data and hospital trajectory data); (2) to improve the operating efficiency of the algorithm.

ACKNOWLEDGMENT

The authors are grateful to Shanghai Palmap Science & Technology Company Limited for providing indoor trajectory data support, which has made this research possible.

REFERENCES

[1] Y. Zhang, L. Zhang, G. Nie, and Y. Shi, "A survey of interestingness measures for association rules," in *Proc. Int. Conf. Bus. Intell. Financial Eng.*, Jul. 2009, pp. 460–463.
[2] L. Zhou and S. Yau, "Efficient association rule mining among both frequent and infrequent items," *Comput. Math. with Appl.*, vol. 54, no. 6, pp. 737–749, Sep. 2007.
[3] J. Chen, C. Miller, and G. G. Dagher, "Product recommendation system for small online retailers using association rules mining," in *Proc. Int. Conf. Innov. Design Manuf. (ICIDM)*, Aug. 2014, pp. 71–77.
[4] Y. S. Cho, S. C. Moon, and K. H. Ryu, "Mining association rules using RFM scoring method for personalized U-commerce recommendation system in emerging data," in *Computer Applications for Modeling*. Berlin, Germany: Springer, 2012, pp. 190–198.
[5] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in *Proc. Int. Conf. Very Large Data Bases*, 1994, pp. 487–499.
[6] T. Brijs, G. Swinnen, K. Vanhoof, and G. Wets, "Using association rules for product assortment decisions: A case study," in *Proc. 5th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 1999, pp. 254–260.
[7] S. Guo, H. Xiong, X. Zheng, and Y. Zhou, "Activity recognition and semantic description for indoor mobile localization," *Sensors*, vol. 17, no. 3, p. 649, Mar. 2017, doi: 10.3390/s17030649.
[8] C. Koehler, N. Banovic, I. Oakley, J. Mankoff, and A. K. Dey, "Indoor-ALPS: An adaptive indoor location prediction system," in *Proc. ACM Int. Joint Conf. Pervas. Ubiquitous Comput. UbiComp Adjunct*, 2014, pp. 171–181.
[9] H. Li, H. Lu, L. Shou, G. Chen, and K. Chen, "In search of indoor dense regions: An approach using indoor positioning data," in *Proc. IEEE 35th Int. Conf. Data Eng. (ICDE)*, Apr. 2019, pp. 2127–2128.
[10] M. Fu, W. Zhu, Z. Le, D. Manko, I. Gorbov, and I. Beliak, "Weighted average indoor positioning algorithm that uses LEDs and image sensors," *Photonic Netw. Commun.*, vol. 34, no. 2, pp. 202–212, Oct. 2017, doi: 10.1007/s11107-016-0682-8.
[11] W. Guan, Y. Wu, S. Wen, H. Chen, C. Yang, Y. Chen, and Z. Zhang, "A novel three-dimensional indoor positioning algorithm design based on visible light communication," *Opt. Commun.*, vol. 392, pp. 282–293, Jun. 2017, doi: 10.1016/j.optcom.2017.02.015.
[12] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 37, no. 6, pp. 1067–1080, Nov. 2007, doi: 10.1109/tsmcc.2007.905750.
[13] L. O. Alvares, V. Bogorny, B. Kuijpers, A. Palma, B. Kuijpers, B. Moelans, and J. A F. de Macedo, "Towards semantic trajectory knowledge discovery," Hasselt University, Hasselt, Belgium, Tech. Rep., Oct. 2007.
[14] M. Muyeba, M. S. Khan, and F. Coenen, "Fuzzy weighted association rule mining with weighted support and confidence framework," in *Proc. Pacific-Asia Conf. Knowl. Discovery Data Mining*, Osaka, Japan, 2008, pp. 49–61.
[15] J. S. Park, M.-S. Chen, and P. S. Yu, "An effective hash-based algorithm for mining association rules," *ACM SIGMOD Rec.*, vol. 24, no. 2, pp. 175–186, May 1995.
[16] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," *ACM SIGMOD Rec.*, vol. 29, no. 2, pp. 1–12, 2000.
[17] P. Cui, X. Wang, J. Pei, and W. Zhu, "A survey on network embedding," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 5, pp. 833–852, May 2019.
[18] A. Grover and J. Leskovec, "Node2vec: Scalable feature learning for networks," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2016, pp. 855–864.

[19] T. Karthikeyan and N. Ravikumar, "A survey on association rule mining," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 3, no. 1, pp. 2278–1021, 2014.

[20] A. Savasere, "An efficient algorithm for mining association rules in large databases," in *Proc. 21st Int. Conf. Very Large Databases*, 1995, pp. 432–444.

[21] M. J. Zaki and K. Gouda, "Fast vertical mining using diffsets," in *Proc. 9th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, Washington, DC, USA, Aug. 2003, pp. 326–335.

[22] T. Uno, M. Kiyomi, and H. Arimura, "LCM ver. 2: Efficient mining algorithms for frequent/closed/maximal itemsets," in *Proc. IEEE ICDM Workshop Frequent Itemset Mining Implement. (FIMI)*, Brighton, U.K., 2004, pp. 1–11.

[23] J. M. Luna, P. Fournier-Viger, and S. Ventura, "Frequent itemset mining: A 25 years review," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 9, no. 6, Nov. 2019, Art. no. e1329.

[24] R. Srikant and R. Agrawal, "Mining quantitative association rules in large relational tables," *ACM SIGMOD Rec.*, vol. 25, no. 2, pp. 1–12, Jun. 1996.

[25] M. Klemettinen, H. Mannila, P. Ronkainen, H. Toivonen, and A. I. Verkamo, "Finding interesting rules from large sets of discovered association rules," in *Proc. 3rd Int. Conf. Inf. Knowl. Manage. (CIKM)*, vol. 10, 1994, pp. 401–407.

[26] Y. Zhong and Y. Liao, "Research of mining effective and weighted association rules based on dual confidence," in *Proc. 4th Int. Conf. Comput. Inf. Sci.*, Aug. 2012, pp. 1228–1231.

[27] J.-M. Wei, W.-G. Yi, and M.-Y. Wang, "Novel measurement for mining effective association rules," *Knowl.-Based Syst.*, vol. 19, no. 8, pp. 739–743, Dec. 2006.

[28] M. J. Zaki, "Mining non-redundant association rules," *Data Mining Knowl. Discovery*, vol. 9, no. 3, pp. 223–248, Nov. 2004.

[29] W. Wang, J. Yang, and P. S. Yu, "Efficient mining of weighted association rules (WAR)," in *Proc. 6th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2000, pp. 270–274.

[30] P. Bahl and V. N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," in *Proc. 19th Annu. Joint Conf. IEEE Comput. Commun. Soc.*, Mar. 2000, pp. 775–784.

[31] Y. Liu, "Inferring gender and age of customers in shopping malls via indoor positioning data," *Environ. Planning B, Urban Anal. Sci.*, vol. 2019, Apr. 2019, Art. no. 2399808319841910, doi: 10.1177/2399808319841910.

[32] P. Wang, H. Wang, H. Zhang, F. Lu, and S. Wu, "A hybrid Markov and LSTM model for indoor location prediction," *IEEE Access*, vol. 7, pp. 185928–185940, 2019, doi: 10.1109/ACCESS.2019.2961559.

[33] P. Wang, S. Wu, H. Zhang, and F. Lu, "Indoor location prediction method for shopping malls based on location sequence similarity," *ISPRS Int. J. Geo-Inf.*, vol. 8, no. 11, p. 517, 2019. [Online]. Available: https://www.mdpi.com/2220-9964/8/11/517

[34] S. Brin, R. Motwani, J. D. Ullman, and S. Tsur, "Dynamic itemset counting and implication rules for market basket data," *ACM SIGMOD Rec.*, vol. 26, no. 2, pp. 255–264, Jun. 1997.

[35] S. Lallich, O. Teytaud, and E. Prudhomme, "Association Rule Interestingness: Measure and Statistical Validation," in *Quality Measures in Data Mining*, F. J. Guillet and H. J. Hamilton Eds. Berlin, Germany: Springer, 2007, pp. 251–275.

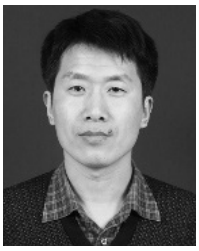
[36] J. M. Luna, M. Ondra, H. M. Fardoun, and S. Ventura, "Optimization of quality measures in association rule mining: An empirical study," *Int. J. Comput. Intell. Syst.*, vol. 12, no. 1, p. 59, 2018.

[37] S. Brin, R. Motwani, and C. Silverstein, "Beyond market baskets: Generalizing association rules to correlations," in *Proc. ACM SIGMOD Int. Conf. Manage. Data (SIGMOD)*, Tucson, AX, USA, 1997, pp. 265–276, doi: 10.1145/253260.253327.

[38] C. Silverstein, S. Brin, and R. Motwani, "Beyond market baskets: Generalizing association rules to dependence rules," *Data Mining Knowl. Discovery*, vol. 2, no. 1, pp. 39–68, 1998.

[39] T. Brijs, G. Swinnen, K. Vanhoof, and G. Wets, "Using association rules for product assortment decisions: A case study," in *Proc. 5th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, San Diego, CA, USA, 1999, pp. 254–260.

[40] C. Clifton, R. Cooley, and J. Rennie, "TopCat: Data mining for topic identification in a text corpus," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 8, pp. 949–964, 2004.



NAIXIA MOU is currently an Associate Professor with the College of Geomatics, Shandong University of Science and Technology, China. His main research interests are data mining, big data analysis, tourism, and transport geography.

HONGEN WANG is currently pursuing the M.S. degree with the College of Geomatics, Shandong University of Science and Technology. His research interests focus on indoor location services and spatial-temporal data mining.

HENGCAI ZHANG received the Ph.D. degree from the Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences. He is currently an Assistant Professor at the Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences. He is a member of the Theory and Methodology Committee of the Chinese Association of Geographic Information System, and of the Chinese Branch of ACM SIGSPATIAL. His interests focus on moving objects database and spatial-temporal data mining. During the past years, he has published over 26 refereed journal and conference papers.

XIN FU received the Ph.D. degree from the Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences. She is currently an Associate Professor with the School of Water Conservancy and Environment, University of Jinan. Her interests focus on trajectory data mining, pattern mining, human mobility, human behavior analysis, trajectory clustering, trajectory reconstruction, and trajectory prediction. During the past years, she has published over 18 refereed journal and conference papers.