Novel Algorithm of Spatiotemporal Association Rules Mining Based on Event-coverage

Gang Fang [1, 2], Yue Wu[1]
[1] School of Computer Science and Engineering
University of Electronic Science and Technology of China
Chengdu, Sichuan, 611731, P. R. China
[2] School of Computer Science and Engineering
Chongqing Three Gorges University, Wanzhou
Chongqing, 404000, P. R. China
gangfang07@sohu.com, ywu@uestc.edu.cn

ABSTRACT: *In order to eliminate data redundancy of spatiotemporal database, and flexibly create spatiotemporal association patterns, and fast discover spatiotemporal association rules, firstly, this paper adopts event-coverage to create spatiotemporal mining database; the method can divide the spatiotemporal domain into some spatiotemporal transaction cells, where each cell is made of attribute values and spatiotemporal predicate values created by the concept generalization method. Then we propose a novel algorithm of spatiotemporal association rules mining based on event-coverage, which can make each spatiotemporal association pattern be mapped to a mixed radix numeral, and uses power set to compute the support. The algorithm adopts simple data structure to discover frequent spatiotemporal association patterns, it only needs to read the database once, and need not generate candidate for mining spatiotemporal association patterns. Because of them, the algorithm overcomes these disadvantages of traditional classical algorithms for discovering frequent patterns. Finally, we discuss the optimal application environments of the algorithm to mine spatiotemporal association rules. For discovering frequent spatiotemporal association patterns on the application environments, these experimental results indicate that the algorithm is better than these traditional classical mining frameworks, particularly, the Apriori framework and the FP-growth framework.*

Keywords: Spatiotemporal Association Patterns, Event-coverage, Power set, Association Rules, Data mining.

Received: 12 January 2016, Revised 18 February 2016, Accepted 25 February 2016

1. Introduction

For spatial data mining and knowledge and discovery, mining spatial association rules from spatial database is one of important tools, and it also is a research hotspot of spatial data mining. Nowadays, spatial association rules have been applied to some domains of social life, such as Urban Traffic[1], Bioscience[2], Social Security[3], Climate forecasting[4], and Demographic survey[5]. Ref. 6 focuses on this specificity of spatial data mining by showing the suitability of join indices to this context. It describes the join index structure and shows how it could be used as a tool for spatial data mining. Ref. 7 discusses multiple level association rules mining, and further indicates spatiotemporal association rule mining should address issues of data integration, data classification, the representation and calculation of spatial relationships, and strategies for finding 'interesting' rules. In scientific data sets, Ref. 8 proposes a generalized framework to effectively discover different types of spatial and spatio temporal patterns, which can be used to capture a variety of interactions among objects of interest and the evolutionary behavior of interactions.

For the present research work, we can divide them into two groups as follows:

One is mining region association rules for numeric geographic elements with point, line and plane; i.e. firstly, we turn these numeric attributes into Boolean attributes with geographic elements, and then use transaction association rules mining methods to mine spatial association rules. The group is suitable for mining spatial association rules based on spatial location. Ref. 7 uses association rules to mine spatiotemporal relations among a set of variables that characterize socioeconomic and land cover change, but it only refer to the region. Ref. 9 proposes a robust geospatial multivariate association rules mining framework, where the attributes for geographic elements with point, line and plane could be turn into the region.

Ref. 10 proposes a novel framework to mine regional association rules based on a given class structure.

The other is mining spatial association rules for discrete geographic elements with spatial objects and layout relationship; i.e. firstly, we turn these discrete geographic elements into the category set, and then use transaction association rules mining methods to extract spatial association rules. Ref. 8 and 11 discuss the star association patterns, the sequence association patterns and the clique association patterns based on spatial distance for spatial objects relationship and layout relationship.

However, in the research work, there are some disadvantages as follows, firstly, the objects for mining spatial association rules are mainly from spatial database, where these algorithms do not fully regard temporal relationship with spatial association patterns; secondly, for mining spatial association patterns, their geographic elements are most one-dimensional. Finally, for these traditional mining algorithms, such as the algorithm Apriori, the algorithm FP-growth, and their improved algorithms also have some disadvantages as follows:

One is the mining framework based on Apriori, i.e. the mining algorithms discover frequent patterns via the ideas of Apriori, called the Apriori framework. The framework needs to generate candidate; its advantages are using simple data structure and costing less memory, and it is easy to program and maintain the algorithm; its disadvantages to read the database repeatedly and generate redundant candidate itemsets [12].

The other is the mining framework based on FP-growth, i.e. the mining algorithms discover frequent patterns via the data structure FP-tree [13], called the FP-growth framework. The framework need not generate candidate for discovering frequent patterns; though it uses a kinds of complex data structure to save reading the database, it needs to cost too much memory for traversing FP-tree. And its programming idea is complex and difficult to implement.

In this paper, our main contributions include the following two aspects:

Firstly, we propose the novel concept, eventcoverage, which can create the spatiotemporal mining database. The method can create flexible multidimensional spatiotemporal patterns, and make the mining database become very simple.

Secondly, we propose a novel algorithm of spatio temporal association rules mining based on event-coverage. The algorithm need not generate candidate, and uses power set to compute the support via reading the database once. The remainder parts are organized as follows:

In Section 2, we introduce how to create spatiotemporal mining database via event-coverage; In Section 3, we propose a novel algorithm of spatio temporal association rules mining based on event-coverage; In Section 4, we use some tests to show the performance of the algorithm and discuss it's the optimal application environments; In Section 5, we summary research results and discuss future work.

2. Creating Spatiotemporal Mining Database

2.1 Relational Definitions

Definition 2.1 Spatiotemporal event, denoted by E, is an object that affects a certain spatio temporal scopes, where the appearance, the disappearance or change of spatiotemporal object can be a spatiotemporal event, which can affect the other objects nearby.

For example, there are six spatiotemporal events:

(1) A taxi (No.t01) with passenger fast drove to the airport (No.a02) along the riverside (No.r03) eastward at 5 pm.
(2) A taxi (No.t02) with passenger slow left the airport (No.a01) across the Golden Gate Bridge (No.b01) westward at 8 am.
(3)A taxi (No.t03) without passenger slow left the school (No.s01) along the riverside (No.r01) southwest at 7 night.
(4) A taxi (No.t04) without passenger fast left the bank (No.b01) across the Rail Bridge (No.b02) northward at 3 pm.
(5) A taxi (No.t05) with passenger fast drove into the school (No.s02), and stop at the gate of bank (No.b03) southeast at 11 am.
(6) A taxi (No.t06) without passenger fast left the bank (No.b02) across the Rail Bridge (No.b01) northward at 4 pm.

Definition 2.2 Event-coverage, denoted by C_E, is a set of factors affected by an event, which consists of attribute factor, time factor, orientation factor, and object factor; where the orientation factor only includes an element; the others at least include an element.

For example, the event-coverage of the first event is described as follows:

The elements of attribute factor are$\{passenger, speed\}$;

Time factor is a given time division, such as $\{dawn, morning, afternoon, night\}$;

The elements of object factor are $\{taxi(No.t01), airport(No.a02), river(No.r03)\}$;

So, the event-coverage is $C_E = \{$ $passenger\ speed\ dawn, morning, afternoon, night, orientation, taxi(No.t01), airport(No.a02), river(No.r03)\}$.

Definition 2.3 Core element of a spatiotemporal event, denoted by e_c, is a spatial object in the object factor of C_E, where the core element is sole, and it is related with the attribute factor, the time factor, and the orientation factor of C_E.

Definition 2.4 Non-core element of the spatiotemporal event, denoted by e_n , is a spatial object in the object factor of C_E, where each non-core element is unrelated with the attribute factor, the time factor, the orientation factor of C_E, and it is only related with the core element e_c.

For the first event, the core element is $e_c = \{taxi\ (No.t01)\}$, the non-core element is $n_e = \{airport\ (No.a02), river(No.r03)\}$.

Definition 2.5 Spatiotemporal transaction based on event-coverage is a transaction, denoted by ST^{C_E}, called a spatiotemporal transaction cell, which is made of attribute values and spatio temporal predicate values, where the attribute values show the features of e_c, the predicate values include the time predicate values, the orientation predicate values with the e_c, and the topology predicate values between the e_c and n_e .

Here, we define these predicates as follows:

The time predicate set is $P_{time}(x, y) = \{before(x, y),\ after(x, y), equal(x, y)\}$;

The orientation predicate set is
$P_{orientation}(x, y) = \{east(x), south(x), west(x), north(x), southeast(x), southwest(x), northwest(x), northeast(x)\}$;

The topology predicate set is

$P_{topology}(x, y) = \{disjoint(x, y), contain(x, y), touch(x, y), cover(x, y), inside(x, y), coveredby(x, y), overlap(x, y)\}$.

2.2 The method of Creating Mining Database based on Event-Coverage

The method of creating spatiotemporal mining database is formally stated as follows:

$CreateDatabase(SPD_E, S_d, P_{time}, P_{orientation}, P_{topology})$

SPD_E is an original database of spatiotemporal events;

S_d , is a time division;

P_{time} , is a time predicate set;

$P_{orientation}$, is an orientation predicate set;

$P_{topology}$, is a topology predicate set;

Input: $SPD_E, S_d, P_{time}, P_{orientation}, P_{topology}$.

Output: *STMD* (a spatiotemporal mining database based on event-coverage);

(1) $S_a = \emptyset$ = ; //Saving the attribute factor for the C_E

(2) $S_o = \emptyset$ = ; //Saving the generalization categories of spatial objects

(3) Reading SPD_E;

(4) For $(\forall\ E \in SPD_E)$ do begin

(5) $T = \emptyset$; //Saving the values of transaction in the spatiotemporal mining database

(6) $S_a = S_a \cup \pi_{attribute}(C_E)$; // $\pi_a(C_E)$ is a set of elements for the factor *a* in C_E

(7) $S_o = \bigcup_{(e \in C_E \wedge e = e_n)} \{g(e)\} \cup S_o$; // $g(e)$ is a function of the concept generalization, its value is the category value of spatial object e ;

(8) For $(\forall\ e \in C_E \wedge e \neq e_c)$ do begin

(9) If $(e \in \pi_{attribute}(C_E))$ then

(10) $T = T \cup \{value(g(e_c), e)\}$; // $value(o,a)$ is the value of the spatial object *o* for attribute element *a* ;

(11) Else

(12) If $(e \in \pi_{object}(C_E))$ then

(13) $T = T \cup \{P_{topology}(g(e_c), g(e))\}$;

(14) Else

(15) $T = T \bigcup \{P_f(g(e_c), e)\}$; // $(e \in \pi_{time}(C_E) \wedge f = time) \vee (e \notin \pi_{time}(C_E) \wedge f = orientation)$

(16) End

(17) If $(\forall\ t \in D \wedge T \neq t.transaction)$ then begin

(18) $x.transaction = T$;

(19) $x.count = 1$;

(20) $D = D \bigcup \{x\}$;

(21) End

(22) Else

(23) $t.count = t.count + 1$;

(24) End

(25) Outputting $STMD = < S, D >$; // where $S = S_a \bigcup S_d \bigcup \{orientation\} \bigcup S_o$

Here S is the relational schema of spatiotemporal mining database $STMD$, D is a set of instances in the spatiotemporal mining database $STMD$.

For example, there are six spatiotemporal events described as definition 2.1, and we let the time division be $S_d = \{dawn,\ morning,\ afternoon,\ night\}$, and then we can get the following $STMD$.

$S_a = \{passenger, speed\}$;
$S_o = \{airport, river, bridge, school, bank\}$;
$S = \{passenger,\ speed, dawn, morning, afternoon,$
$night, orientation, airport, river, bridge, school, bank\}$

By the concept generalization method, the event (1) is expressed as follows:

$$T_1 = \{value(taxi, passenger) = load, value(taxi, speed) = fast, before(taxi, night), after(taxi, morning), equal(taxi, afternoon), disjoint(taxi, airport), touch(taxi, river), east\ (taxi)\}.$$

We use the concept generalization method to change the others events, here, the event (4) and (6) can be turned into the same instance. And then, the spatiotemporal mining database $STMD$ can be expressed as table 1.

3. A Novel Spatiotemporal Association Rules Mining Algorithm

In this section, we propose a novel algorithm of frequent patterns mining without candidate generation, which can map the spatiotemporal mining database to a mixed radix notation system, and uses power set theory for the spatiotemporal event to compute the support.

3.1. Mapping from a Database to a Mixed Radix Notation System

Let $STMD = < S, D >$ be a spatiotemporal mining database, where

S, called attribute set, is the relational schema of spatiotemporal mining database;

D , is a set of instances in the spatiotemporal mining database, each instance includes two domains, namely, *transaction* and *count*.

And then, the $STMD = < S, D >$ can be mapped to the mixed radix notation system $M_{|S|}$ as follows:

$M_{|S|} = \begin{pmatrix} 1 & 2 & \ldots & |S| \\ |s_1|+1 & |s_2|+1 & \ldots & |s_{|S|}|+1 \end{pmatrix}$, where $|M_{|S|}| = \prod_{i=1}^{|S|}(|s_i|+1) - 1$, $|s_i|$ is the number of discrete values for the range of the attribute $s_i (i = 1, 2, \ldots, |S|)$. If the range of attribute factor is consecutive, then it needs to be turned into discrete category values.

So, each attribute value $v_j^{s_i} (i = 1, 2, \ldots, |S|, j = 1, 2, \ldots, |s_i|)$ can be mapped to a mixed radix numeral, denoted by a decimal integer $n_{<i,j>}$; each pattern P can be also mapped to a mixed radix numeral, denoted by a decimal integer n_P; and each transaction T can be also mapped to a mixed radix numeral, denoted by a decimal integer n_T ;

where $n_{<i,j>} = \begin{cases} j & i = 1 \\ j \cdot \prod_{\tau=1}^{i-1}(|s_\tau|+1) & i \neq 1 \end{cases}$; $n_P = \sum_{(v_j^{s_i} \in P)} n_{<i,j>}$; $n_T = \sum_{(v_j^{s_i} \in T)} n_{<i,j>}$

According to table 1, the spatiotemporal mining database can be mapped to the following mixed radix notation system:

$$M_{12} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ 3 & 3 & 3 & 3 & 4 & 3 & 6 & 2 & 2 & 2 & 3 & 3 \end{pmatrix};$$

TID / Attribute	T_1	T_2	T_3	T_4 (T_6)	T_5
Passenger	load	load	empty	empty	load
Speed	fast	slow	slow	fast	fast
Dawn	---	after	before	---	after
Morning	after	equal	---	after	equal
Afternoon	equal	before	after	equal	before
Night	before	---	equal	before	---
Orientation	east	west	south-west	north	south-east
Airport	disjoint	disjoint	---	---	---
River	touch	---	touch	---	---
Bridge	---	overlap	---	overlap	---
School	---	---	disjoint	---	inside
Bank	---	---	---	disjoint	touch
Count	1	1	1	2	1

Table 1. The spatiotemporal mining database

Let $P = \{value\ (taxi, passenger) = load, equal\ (taxi,\ night), disjoint\ (taxi, airport)\}$ be a spatiotemporal association pattern, for the M_{12}, P is mapped to a mixed radix numeral $(000010200001)_{12}$, namely, a decimal integer $n_P = 1 \cdot 1 + 2 \cdot 324 + 1 \cdot 5832 = 6481$, and then it can be also turned into a set $S_p = \{1, 648, 5832\}$, where $n_P = \sum^{s \in S_P} s = 1 + 648 + 5832$.

For the M_{12}, the spatiotemporal event (1) can be obviously mapped to $(000111122011)_{12}$, namely, a decimal integer $n_{e(1)} = 1 + 1 \cdot 3 + 2 \cdot 27 + 2 \cdot 81 + 1 \cdot 324 + 1 \cdot 972 + 1 \cdot 5832 + 1 \cdot 11664 = 19012$, and then it can be also turned into a set $S_{e(1)} = \{1,3,54,162,324,972, 5832,11664\}$, namely $n_{e(1)} = \sum^{s \in S_{e(1)}} s$.

3.2 Power Set Theory for a Spatiotemporal Event

In this section, according to the traditional power set, we introduce a power set theory for a spatiotemporal event, and use it to compute the support.

Let $STMD =< S, D >$ be a spatiotemporal mining database, which is mapped to the mixed radix notation system $M_{|S|}$; and let a spatiotemporal event E be turned into a set S_E, the power set theory for the spatiotemporal event can be defined as follows:

$P(S_E) = \{S \mid S \subseteq S_E \wedge S \neq \emptyset\}$, where

$\forall s(s \in S_E)$, is a decimal integer, which is mapped to the sole attribute value of the spatiotemporal event *E;*

$\forall p(p \in P(S_E))$, is a set of decimal integer, which is mapped to the sole spatiotemporal association pattern P .

And so, we have $\forall\ p \in P(S_E), p \subseteq S_E$, namely, the power set of S_E can be mapped to all the patterns supported by E. The method is used to compute the support.

Next, we discuss how to generate the power set for the spatiotemporal event E. The method can be expressed as the algorithm $PowerSet(S_E, S_P)$:

Input: S_E (a set of integer mapping a spatio temporal event E);

Output: S_P (a set of integer mapping many spatio temporal patterns);

(1) $L = |S_E|$; // the number of the elements in S_E

(2) For $(\forall\ i \in [1, 2^L - 1])$ do begin

(3) $B[1, ..., L] = IntegerToBinary(i)$; //an integer is turned into a binary number $B[1, ..., L]$

(4) $s = B[1,...,L] \cdot (S_E)$; // this is an inner product between B and S_E

(5) $S_P = S_P \cup \{s\}$;

(6) End

(7) Outputting S_P;

Let $\{value(taxi, passenger) = load, equal(taxi, night), disjoint\ (taxi, airport)\}$ be a spatiotemporal event, for the M_{12}, it is turned into a set S = {1, 648, 5832} ; and then generating these patterns supported by the event is described as follows:

$$i = 1, B = \{0,0,1\}, S_P[1] = (0,0,1) \cdot (1,648,5832) = 5832$$
$$\rightarrow disjoint(taxi, airport);$$
$$i = 2, B = \{0,1,0\}, S_P[2] = (0,1,0) \cdot (1,648,5832) = 648$$
$$\rightarrow equal(taxi, night);$$
$$i = 3, B = \{0,1,1\}, S_P[3] = (0,1,1) \cdot (1,648,5832) = 6480$$
$$\rightarrow disjoint(taxi, airport), equal(taxi, night);$$
$$i = 4, B = \{1,0,0\}, S_P[4] = (1,0,0) \cdot (1,648,5832) = 1$$
$$\rightarrow value(taxi, passenger) = load;$$
$$i = 5, B = \{1,0,1\}, S_P[5] = (1,0,1) \cdot (1,648,5832) = 5833$$
$$\rightarrow value(taxi,passenger) = load, disjoint(taxi,airport$$
$$i = 6, B = \{1,1,0\}, S_P[6] = (1,1,0) \cdot (1,648,5832) = 649$$
$$\rightarrow value(taxi, passenger) = load, equal(taxi, night);$$
$$i = 7, B = \{1,1,1\}, S_P[7] = (1,1,1) \cdot (1,648, 5832) = 6481$$
$$\rightarrow value(taxi, passenger) = load, equal(taxi, night),$$
$$disjoint(taxi, airport).$$

3.3 Spatiotemporal Association Patterns Mining Algorithm

In this section, based on Section 3.1 and 3.2, we mainly discuss the course of discovering frequent spatiotemporal association patterns, which is expressed as the algorithm *Mixed-radix*.

Input: SPD_E, S_d, P_{time}, $P_{orientation}$, $P_{topology}$ z

Output: F (the maximal frequent spatiotemporal association patterns);

(1) $FreCount = \emptyset$;

(2) $STMD = CreateDatabase(SPD_E, S_d, P_{time}, P_{orientation}, P_{orientation}, P_{topology})$; //see Section 2.2, namely, creating a spatiotemporal mining database

(3) Reading $STMD$;

(4) Mapping from $STMD$ to $M_{|S|}$;

(5) For $(\forall\ E \in STMD)$ do begin

(6) $S_E = TurnEventToSet(E.transaction, M_{|S|})$ // the event is turned into a set

(7) $PowerSet(S_E, S_P)$ //see Section 3.2

(8) For$(\forall\ s \in S_P)$ do begin

(9) If ($FreCount[s].count = 0$) then begin

(10) $FreCount[s].transaction = E.transaction$;

(11) $FreCount[s].count = E.count$;

(12) End

(13) Else

(14) $FreCount[s].count+ = E.count$;

(15) End

(16) End

(17) For $(i \in [1, |FreCount|])$ do begin

(18) If (*FreCount*[*i*].*count* ≥ *minimalSupport*) then

(19) Deleting the subsets of *FreCount*[*i*].*transaction* in *F*, and writing it to *F* ;

(20) End

(21) Outputting *F* ;

In this paper, we only discuss the main task of spatiotemporal association rules mining, namely, discovering frequent spatiotemporal association patterns by the algorithm *Mixed-radix*.

3.4 Performance Features

Here, we theoretically discuss the features of the *Mixed-radix*, and compare it with these classical mining frameworks, particularly, the Apriori and FP-growth framework. The algorithm has these performance features as table 2.

Comparing Items	Apriori framework	FP-growth framework	Mixed-radix
Data Structure	Simple	Complex	Simple
Read Database	Multiple	Twice	Once
Program	Simple	Complex	Simple
Complexity	High	High	Low
Memory usage	Less	More	Less
Candidate	Yes	No	No

Table 2. The comparisons of performance features

Based on the comparing results of table 2, we can draw the following conclusions:

(1)The *Mixed-radix* has these advantages of the Apriori framework, namely, it can turn each pattern into an integer, and so it can use simple array as data structure to reduce memory usage. In addition, its programming idea is simple and easy to implement it.

(2)The *Mixed-radix* also has these advantages of the FP-growth framework, namely, it need not generate candidate, and only needs to read the database once to compute the support by power set theory for spatiotemporal event.

(3)The *Mixed-radix* avoids these disadvantages of the two frameworks, namely, it is different from FP-growth to avoid using complex data structure; and it is different from Apriori that it need not read the database repeatedly.

4. Experimental Results

In this section, we discuss two problems, one is testing the performance of the *Mixed-radix* on the specified feature datasets; the other is assessing the trends of performance for the *Mixed-radix* on these different datasets. We mainly compare it with the Apriori framework and the FP-growth framework in the following tests. In general, as the FP-growth framework is better than the Apriori framework, so we do not compare them directly on these experiments.

Based on the mapping from a database to a mixed radix notation system in section 3.1, we can divide the real database into two types of datasets as follows:

(1) One is called a sparse patterns dataset, namely $\rho = log_{|D|}^{|M_{(S)}|} < 1$.

(2) The other is called a dense patterns dataset, namely $\rho = log_{|D|}^{|M_{(S)}|} \geq 1$.

Where $|M_{|S|}|$ is the number of nonempty combination for attribute values (its definition in section 3.1), and $|D|$ is the number of spatiotemporal event from the spatiotemporal database.

Here, we choose the following two original data sets, which are the specified sparse patterns data sets for testing the performance.

(1)The first original dataset is from the city GPS data of taxi, where each spatiotemporal event is consists of speed, loading passenger, time orientation, and space layout relationship. There are 323080 spatiotemporal events in the spatiotemporal database via data filtering; after the database is turned into the mining database based on event-coverage, it can be mapped to the following mixed radix notation system.

$$M_7 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 4 & 4 & 4 & 3 & 5 & 9 \end{pmatrix}, \text{ where } \rho = log_{|D|}{}^{|M_7|} = log_{323080}{}^{34559} = 0.824 < 1.$$

(2) Another original dataset is from the city GPS data of bus, where each spatiotemporal event consists of grade of service, speed, time, orientation, and space layout. According to the dataset, we create the mining database based on event-coverage consists of 40600 events, which can be mapped to the following mixed radix notation system.

$$M_6 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 5 & 4 & 6 & 7 & 8 \end{pmatrix}, \text{ where } \rho = log_{|D|}{}^{|M_6|} = log_{40600}{}^{20159} = 0.934 < 1.$$

All the experiments are run on Microsoft Window XP Professional with Intel (R) Core (TM) 2 Duo CPU (T6570@)2.10 GHz 1.19GHz) and 1.99 GB memory. The software development environment is based on C# with Microsoft Visual Studio 2008.

4.1. Testing the Performance

4.1.1 Testing on the GPS Data of Taxi

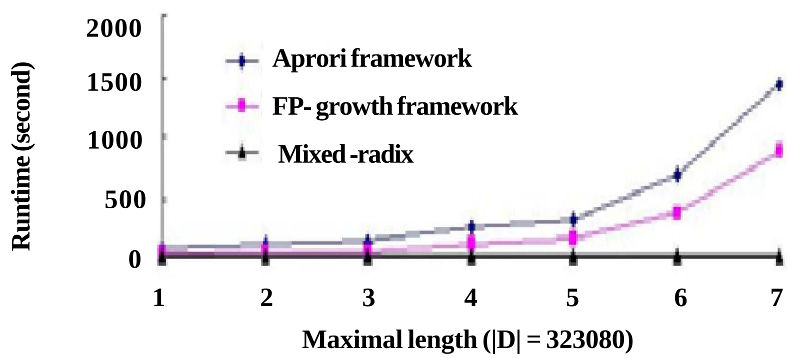


Figure 1. The varying maximal length for the taxi

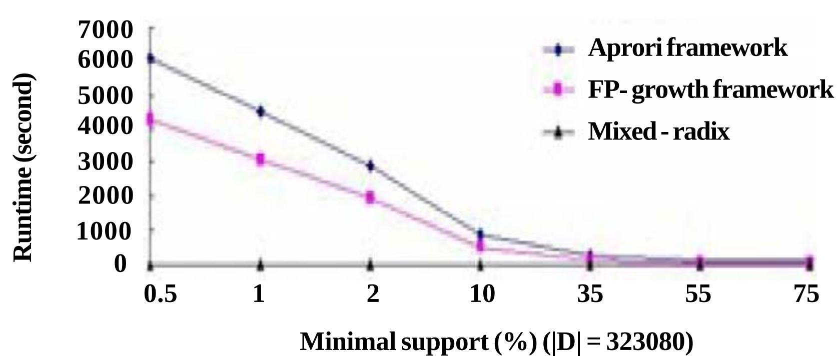


Figure 2. The varying minimal support for the taxi

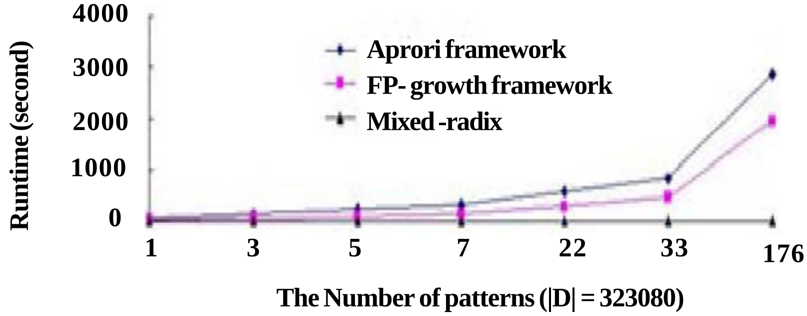


Figure 3. The varying number of patterns for the taxi

On the dataset, we compare the performance of *Mixed-radix* with the Apriori framework and the FP-growth framework as the maximal length varies, the comparing results are expressed as figure 1. For the varying minimal support, the comparing results are expressed as figure 2, and figure 3 shows the comparing results as the number of frequent patterns varies.

Based on these comparisons from figure 1 to 3, we know that the algorithm *Mixed-radix* is better than the Apriori framework and the FP-growth framework on the city GPS data of taxi.

4.1.2 Testing on the GPS data of bus

On the dataset, we compare the performance by the same way. The comparisons are expressed as figure 4, figure 5 and 6. For the comparisons, we also draw the same conclusion as above.

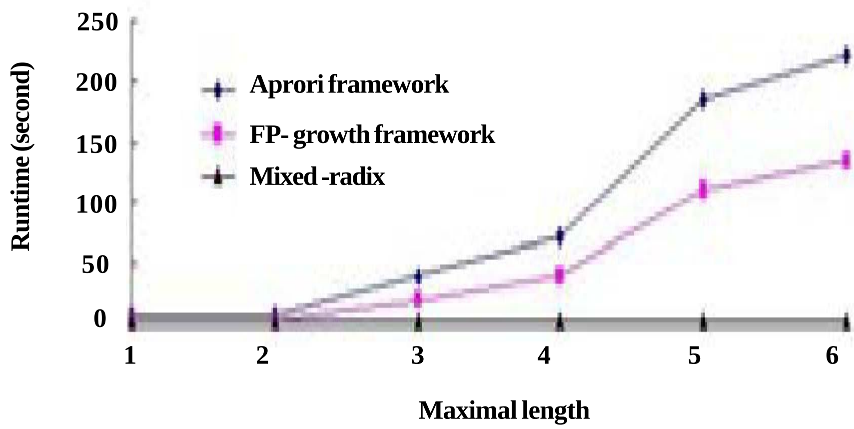


Figure 4. The varying maximal length for the bus

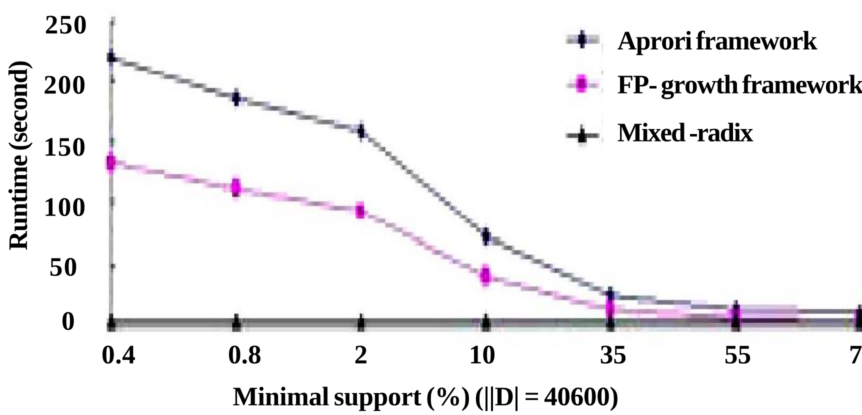


Figure 5. The varying minimal support for the bus

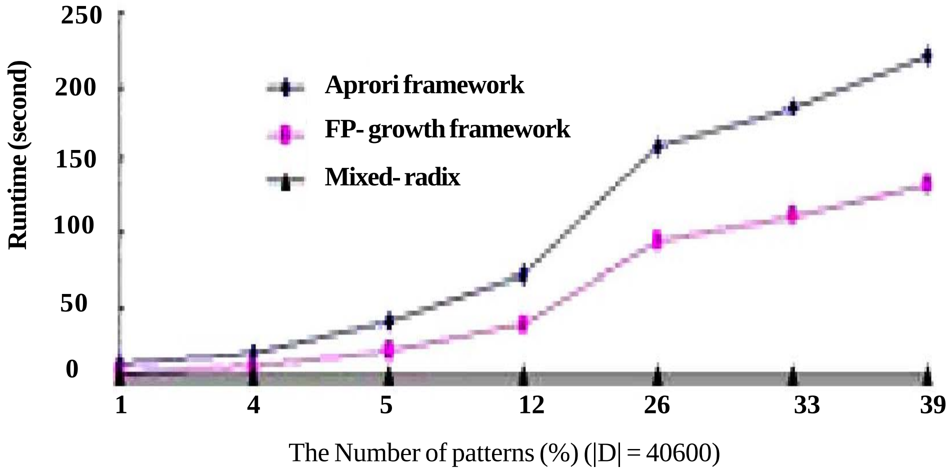


Figure 6. The varying number of patterns for the bus

By these experiments on the sparse patterns datasets (ρ <1) , we draw a conclusion as follows: The algorithm *Mixed-radix* is better than the Apriori framework and the FP-growth framework, the runtime of which is not affected by these parameters as the maximal length, the minimal support, and the number of frequent patterns.

4.2 Assessing the Trends of Performance

In this part, we mainly discuss the relationships between the performance and these parameters, such as $\rho = log_{|D|}^{|M_{|S|}|}$, $|M_{|S|}|$, $|D|$.

4.2.1 Testing on these Datasets Created by the First Original Dataset

For the first original dataset, by altering the number of events or data structure, we can create eight new datasets as table 3.

Name	Mixed radix notation system	ρ
Taxi data 1	(4, 4, 4, 5, 9)	$log_{403850}^{2879} = 0.617$
Taxi data 2	(4, 4, 4, 5, 9)	$log_{323080}^{2879} = 0.628$
Taxi data 3	(4, 4, 4, 4, 3, 5, 9)	$log_{403850}^{34559} = 0.810$
Original data	(4, 4, 4, 4, 3, 5, 9)	$log_{323080}^{34559} = 0.824$
Taxi data 4	(4, 4, 4, 4, 3, 5, 9, 3)	$log_{403850}^{103679} = 0.895$
Taxi data 5	(4, 4, 4, 4, 3, 5, 9, 3)	$log_{323080}^{103679} = 0.910$
Taxi data 6	(4, 4, 4, 5, 9)	$log_{3231}^{2879} = 0.986$
Taxi data 7	(4, 4, 4, 4, 3, 5, 9)	$log_{3231}^{34559} = 1.293$
Taxi data 8	(4, 4, 4, 4, 3, 5, 9, 3)	$log_{3231}^{103679} = 1.429$

Table 3. The new altered eight datasets for the first original dataset

(1)Assessing the trends with the varying ρ

On these datasets, we compare the *Mixed-radix* with the Apriori and the FP-growth framework; the results are expressed as figure 7 to 14.

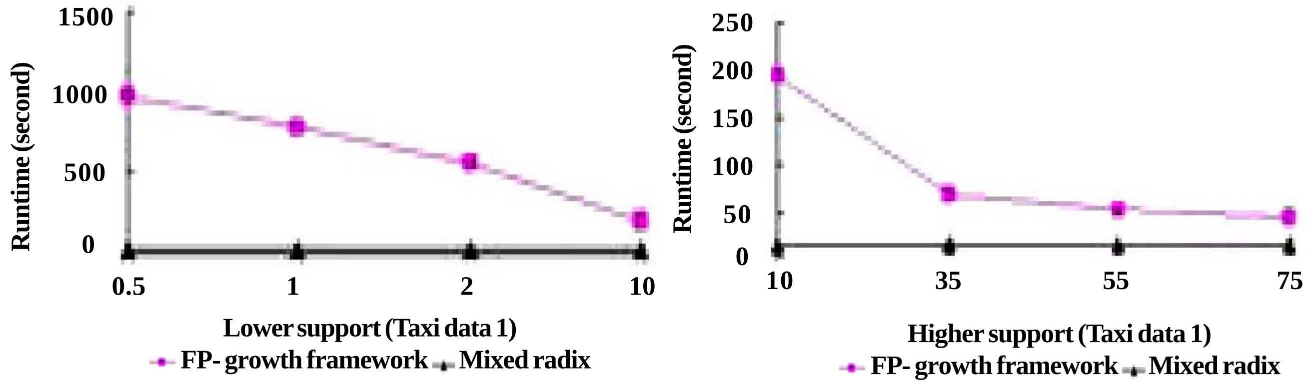


Figure 7. The comparisons on $\rho = 0.617$

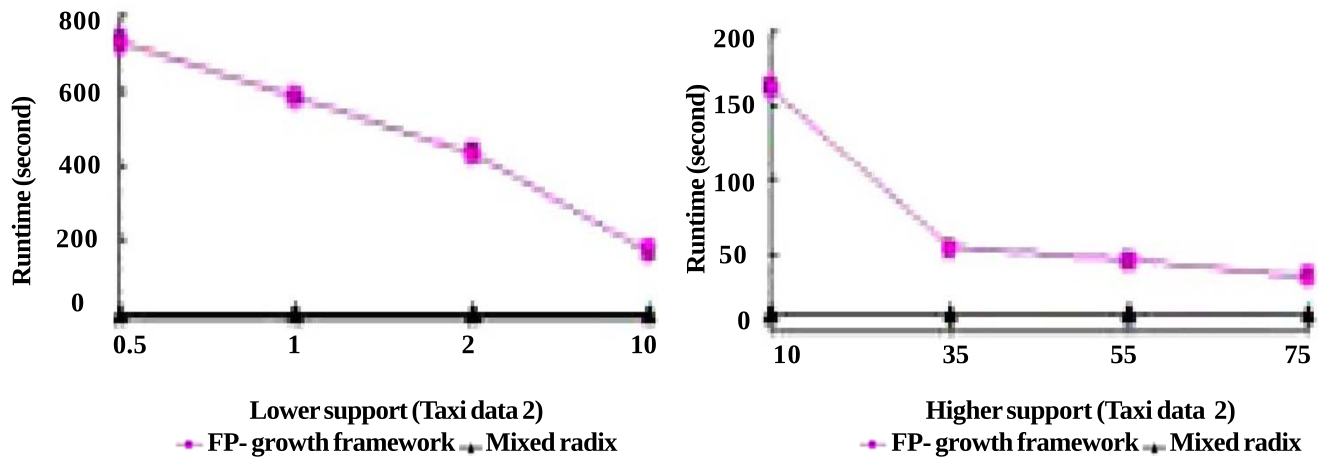


Figure 8. The comparisons on $\rho = 0.628$

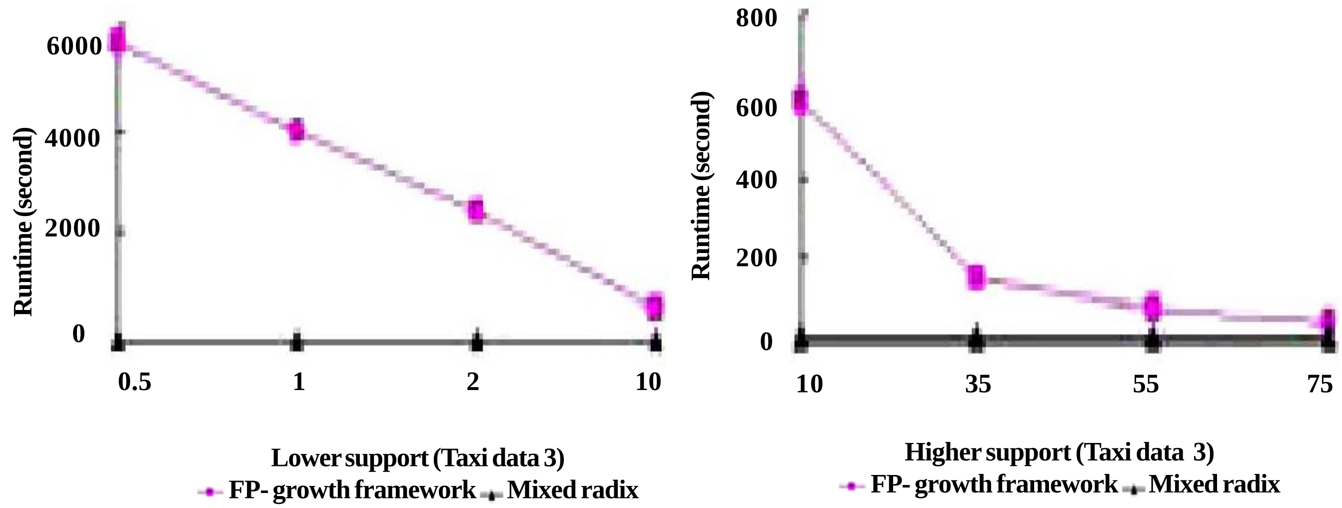


Figure 9. The comparisons on $\rho = 0.810$

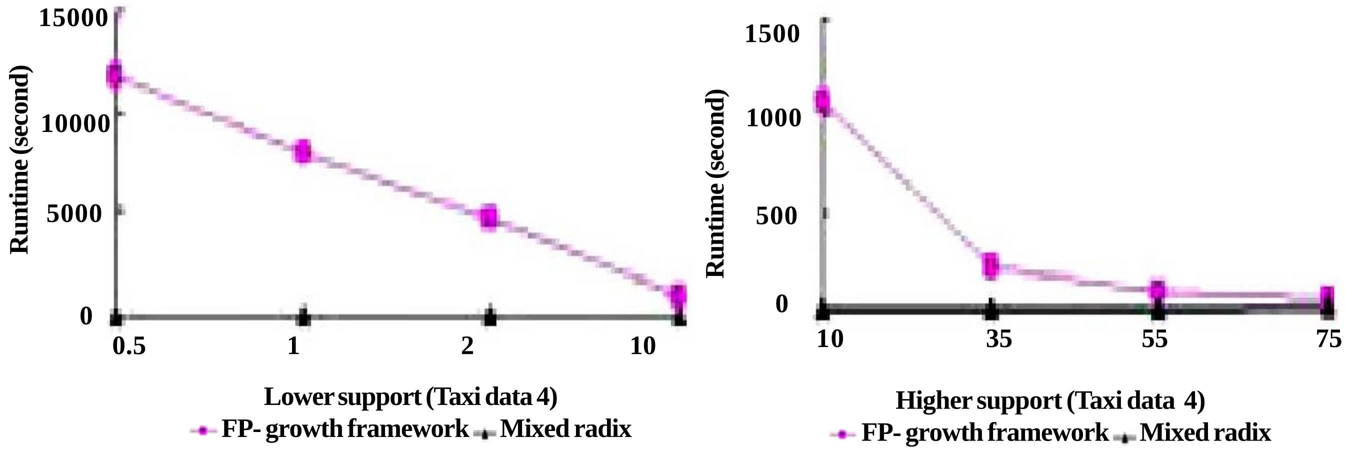
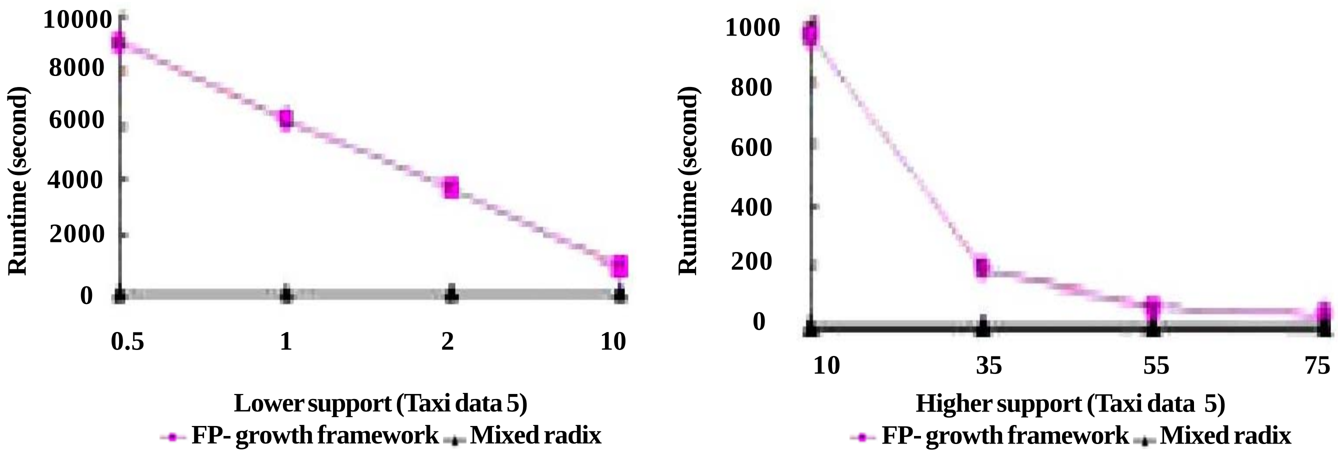


Figure 10. The comparisons on $\rho = 0.895$



Figurre 11. The comparisons on $\rho = 0.910$

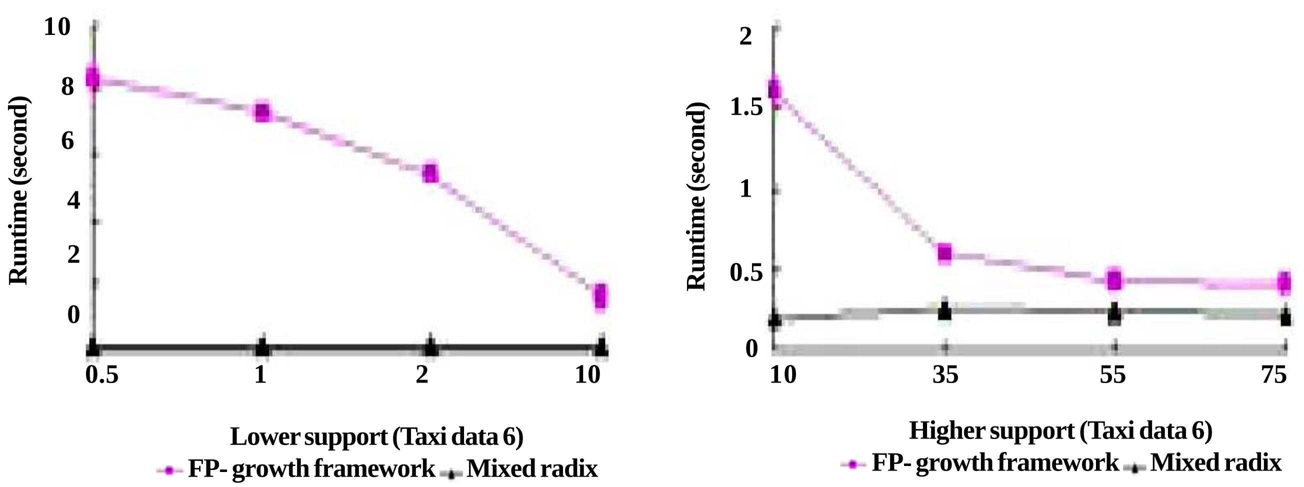


Figure 12. The comparisons on $\rho = 0.986$

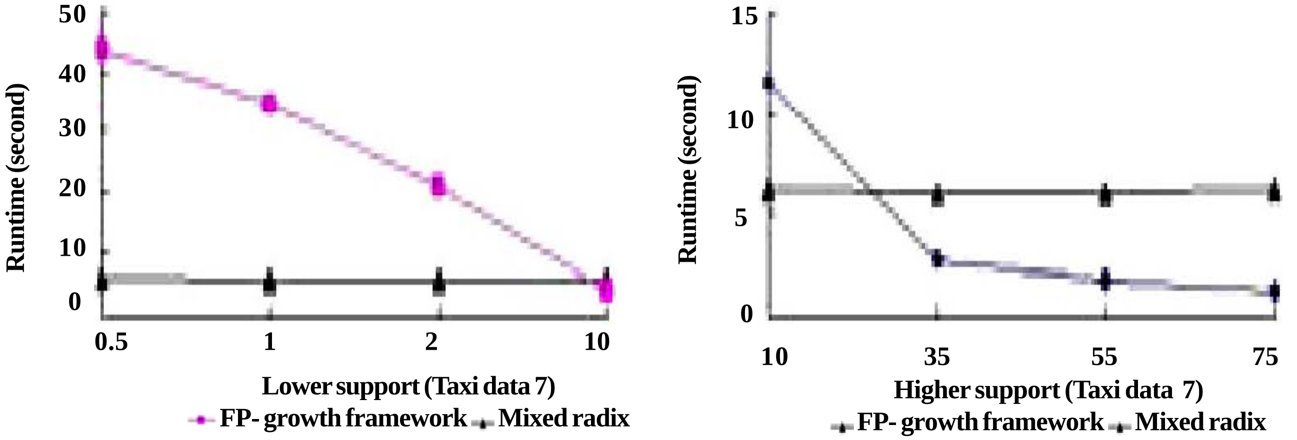


Figure 13. The comparisons on ρ =1.293

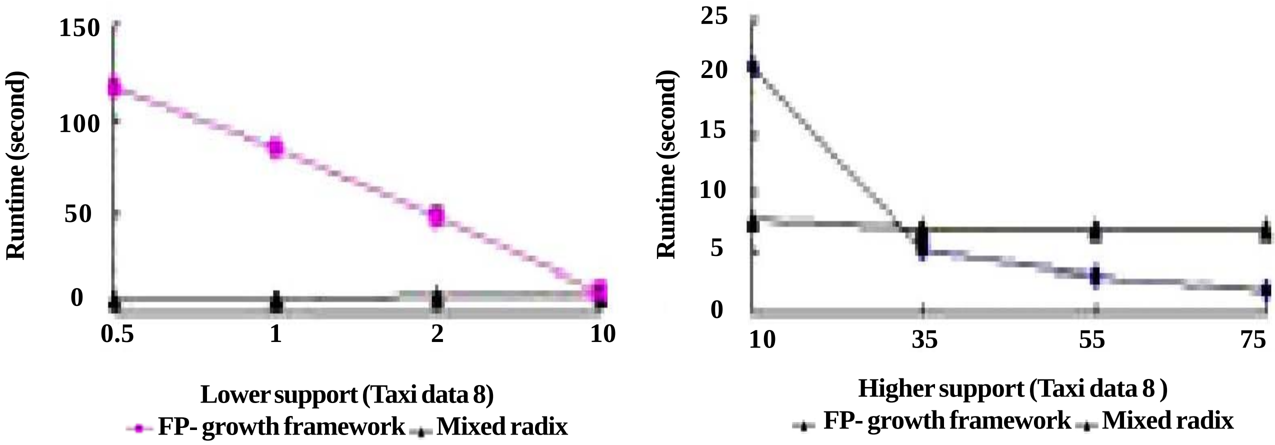


Figure 14. The comparisons on ρ=1.429

Based on the comparisons, we know that the *Mixed-radix* is better than the Apriori framework and the FP-growth framework on the sparse patterns datasets (ρ <1) ; but it is not better than them on the dense patterns datasets ($\rho \geq 1$) .

(2) Assessing the trends with the varying $|M_{|S|}|$ For the constant $|D|$, the tests with the varying $|M_{|S|}|$ are expressed as figure 15 to 17. Based on the comparisons, we know the performance of *Mixed-radix* becomes worse as $|M_{|S|}|$ increases.

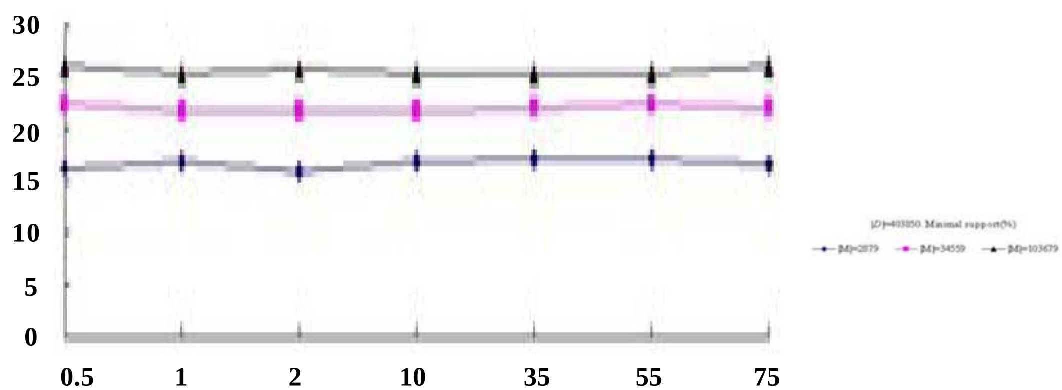


Figure 15. The varying $|M_{|S|}|$ on $|D|$ = 403850

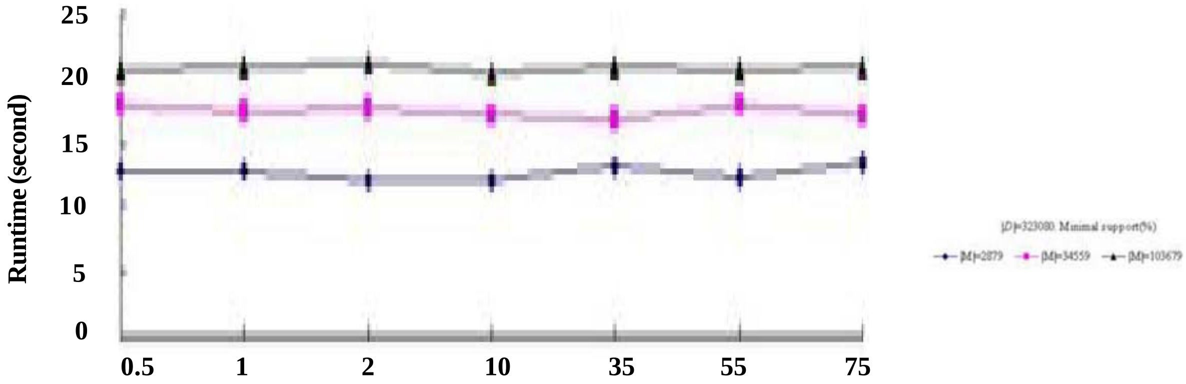


Figure 16. The varying $|M_{|S|}|$ on $|D|$= 323080

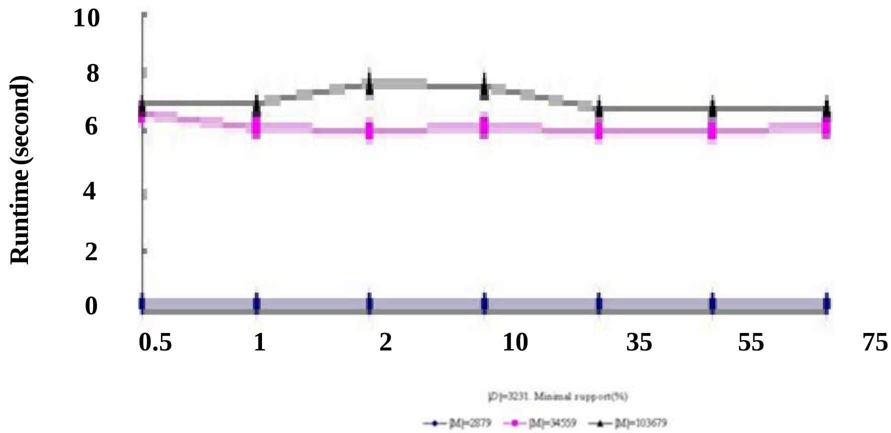


Figure 17. The varying $|M_{|S|}|$ on $|D|$ = 3231

(3) Assessing the trends with the varying $|D|$ For the constant $|M_{|S|}|$, the tests with the varying $|D|$ are expressed as figure 18 to 20. Based on the comparisons, we also find the performance of *Mixed-radix* becomes worse as increases $|D|$.

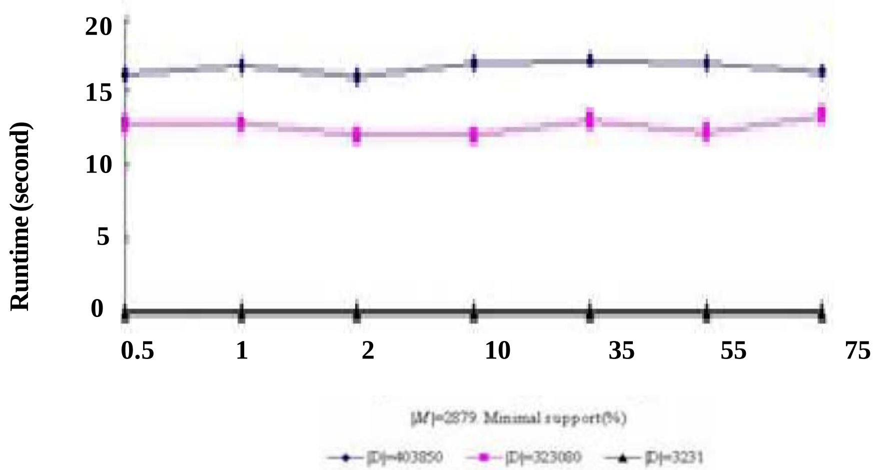


Figure 18. The varying $|D|$ on 2879

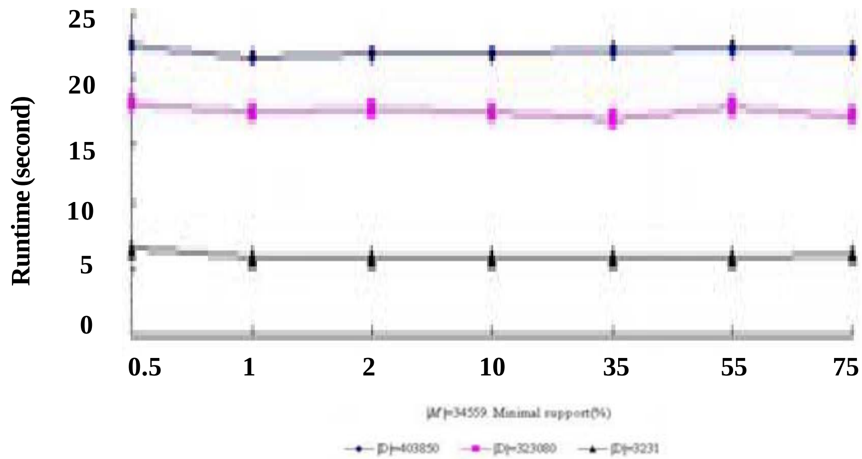


Figure 19. The varying $|D|$ on $|M_{|S|}|$=34559

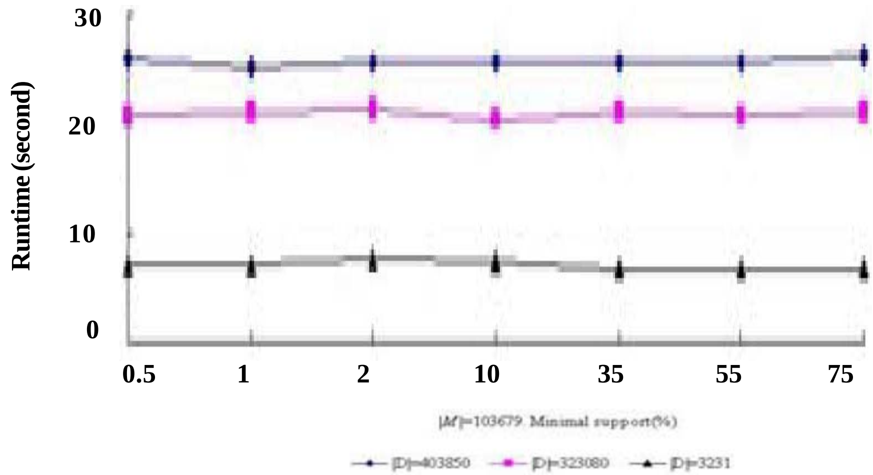


Figure 20. The varying $|D|$ on $|M_{|S|}|$= 103679

4.2.2 Testing on these Datasets Created by the Second Original Dataset

For the second dataset, we also use the same method to create eight datasets as table 4.

Name	Mixed radix notation system	ρ
Bus data 1	(3,5,6,7,8)	$log_{121800}{}^{5039} = 0.728$
Bus data 2	(3,5,6,7,8)	$log_{40600}{}^{5039} = 0.803$
Bus data 3	(3,5,4,6,7,8)	$log_{121800}{}^{20159} = 0.846$
Original data	(3,5,4,6,7,8)	$log_{40600}{}^{20159} = 0.934$
Bus data 4	(3,5,4,6,7,8,3,6)	$log_{121800}{}^{362879} = 1.093$
Bus data 5	(3,5,6,7,8)	$log_{2030}{}^{5039} = 1.119$
Bus data 6	(3,5,4,6,7,8,3,6)	$log_{40600}{}^{362879} = 1.206$
Bus data 7	(3,5,4,6,7,8)	$log_{2030}{}^{20159} = 1.301$
Bus data 8	(3,5,4,6,7,8,3,6)	$log_{2030}{}^{362879} = 1.954$

Table 4. The new altered eight datasets for the second original dataset

(1)Assessing the trends with the varying ρ On these datasets, we compare the *Mixed-radix* with the Apriori framework and the FP-growth framework; the results are expressed as figure 21 to 28. For the comparisons, we also know that the *Mixed-radix* is better than the Apriori framework and the FP-growth framework on the sparse patterns datasets (ρ <1) ; but it is not better than them on the dense patterns datasets ($\rho \geq 1$) .

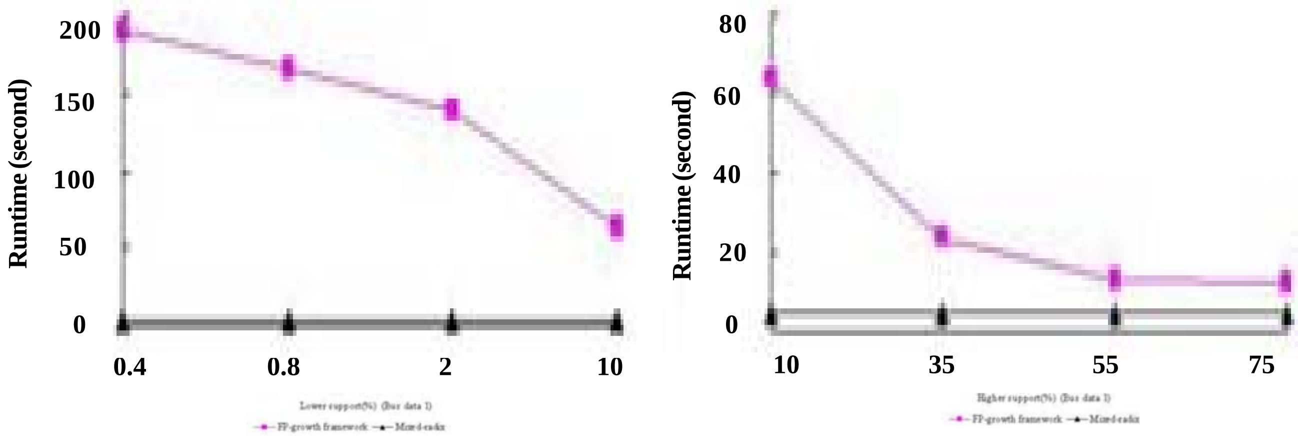


Figure 21. The comparisons on ρ = 0.728

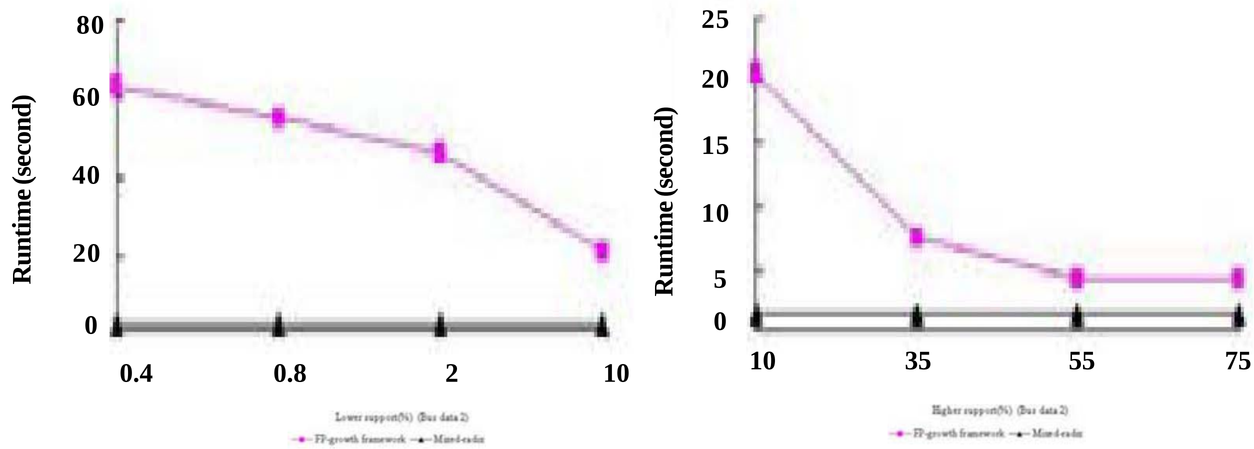


Figure 22. The comparisons on ρ = 0.803

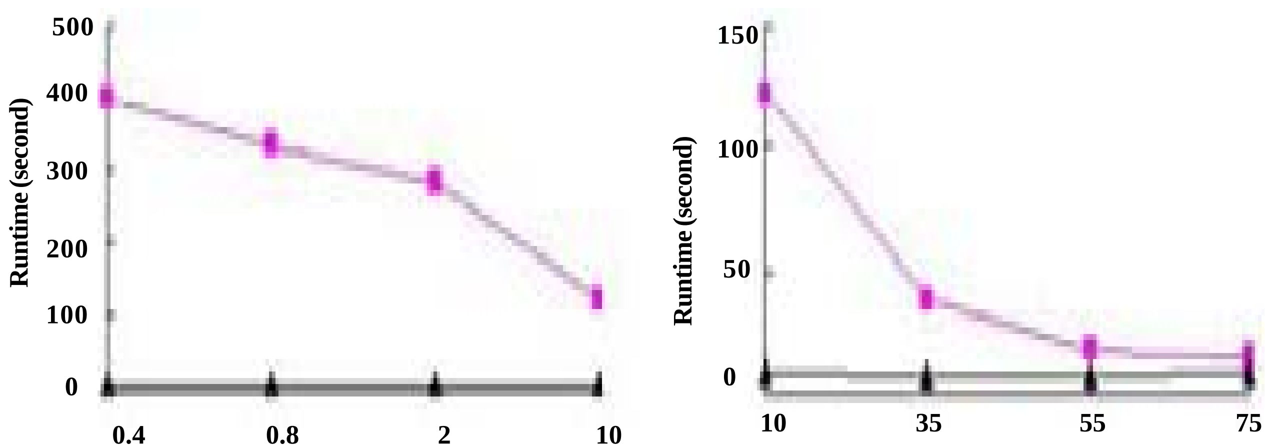


Figure 23. The comparisons on ρ = 0.846

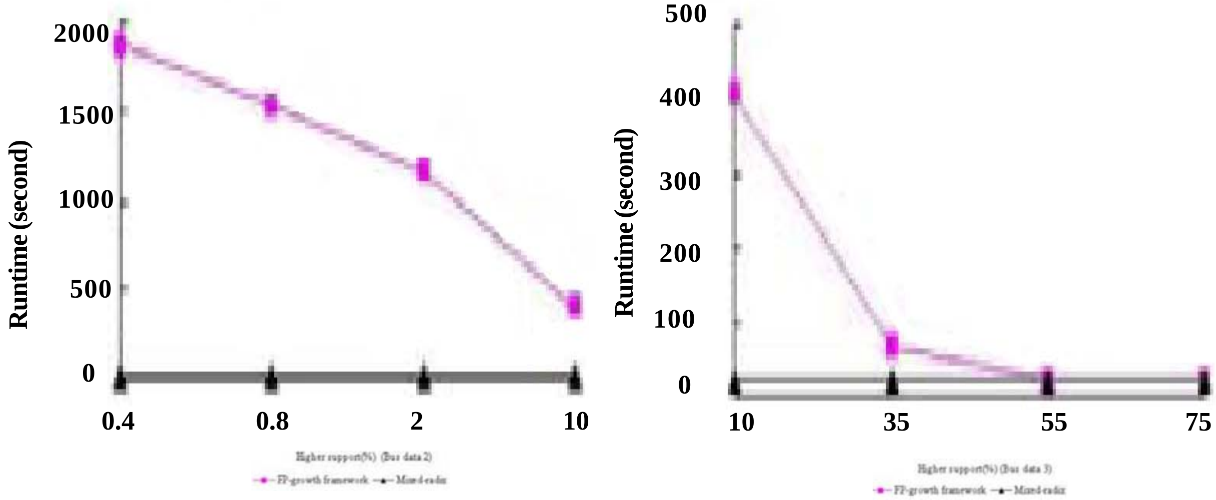


Figure 24. The comparisons on $\rho = 1.093$

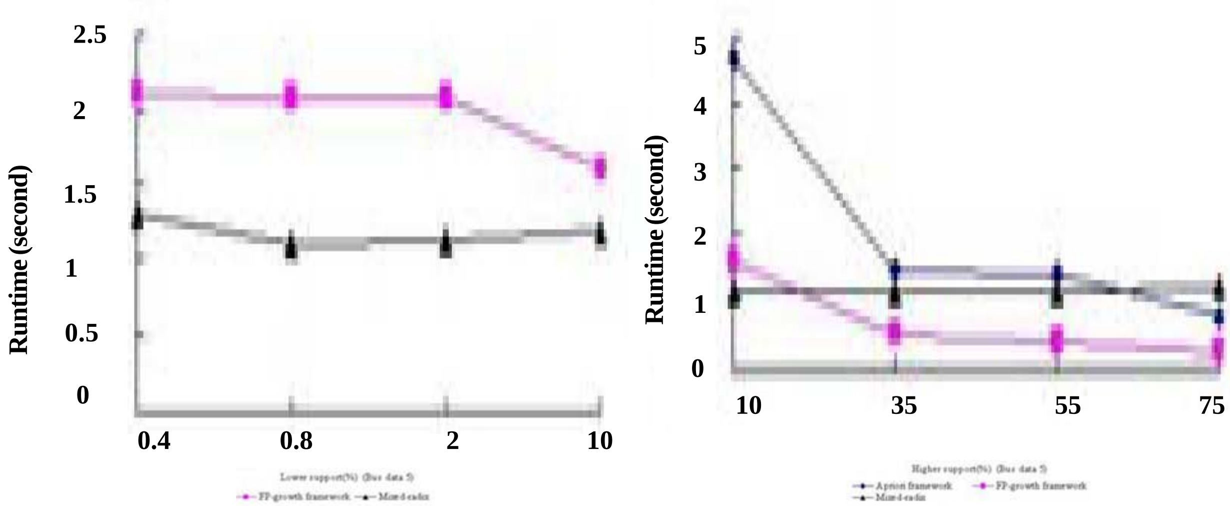


Figure 25. The comparisons on $\rho = 1.119$

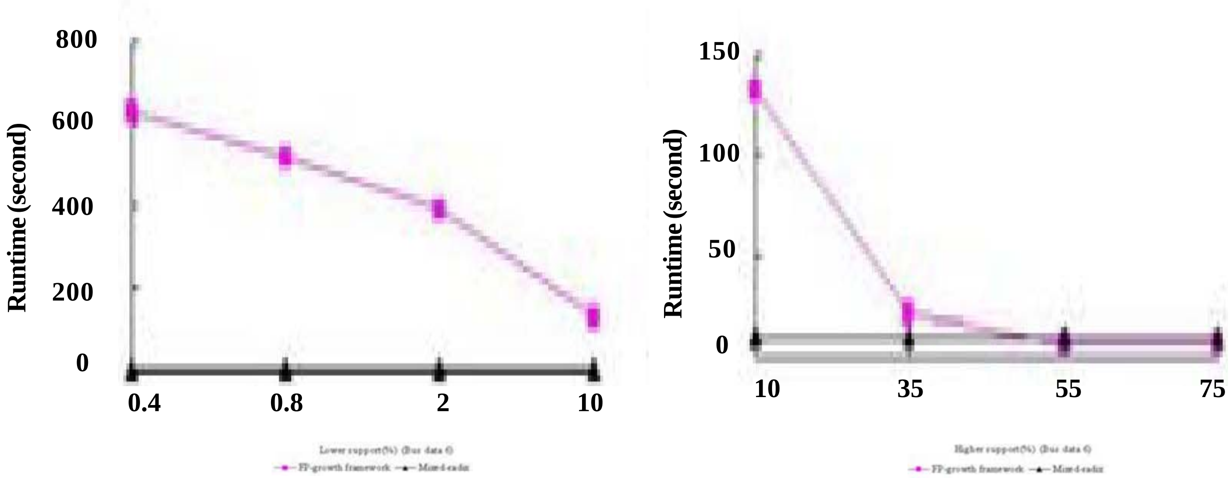


Figure 26. The comparisons on $\rho = 1.206$

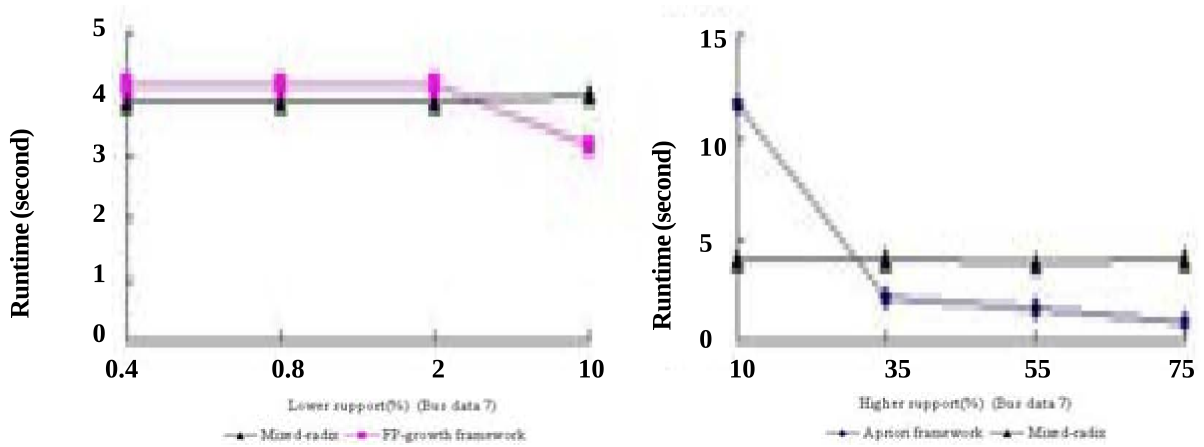


Figure 27. The comparisons on ρ =1.293

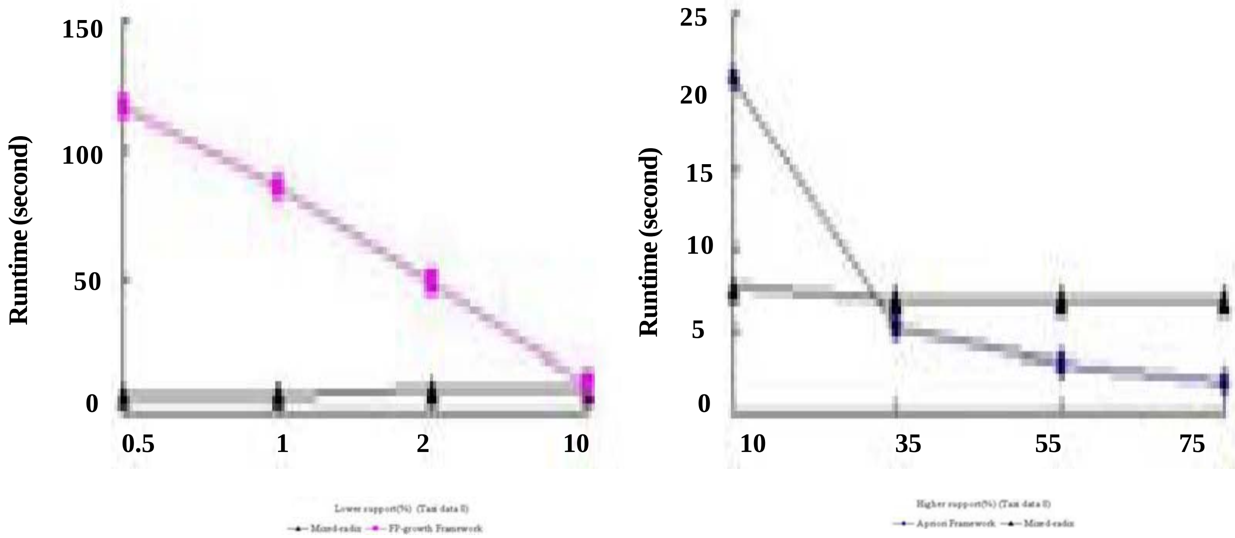


Figure 28. The comparisons on ρ = 1.954

(2)Assessing the trends with the varying $|M_{|S|}|$ For the constant$|D|$, the tests with the varying $|M_{|S|}|$ are expressed as figure 29 to 31. For these comparisons, we know that the performance of *Mixed-radix* becomes worse as $|M_{|S|}|$ increases.

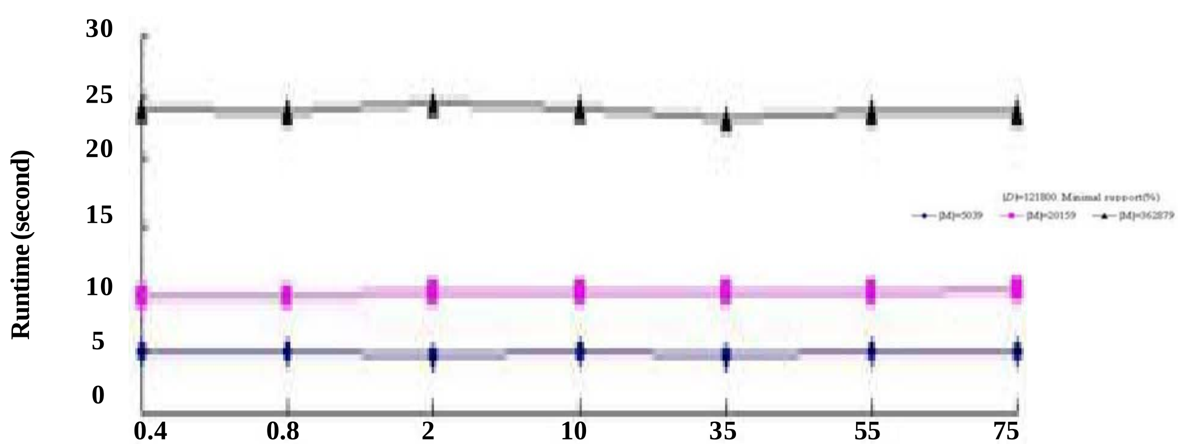


Figure 29. The varying $|M_{|S|}|$ on$|D|$ =121800

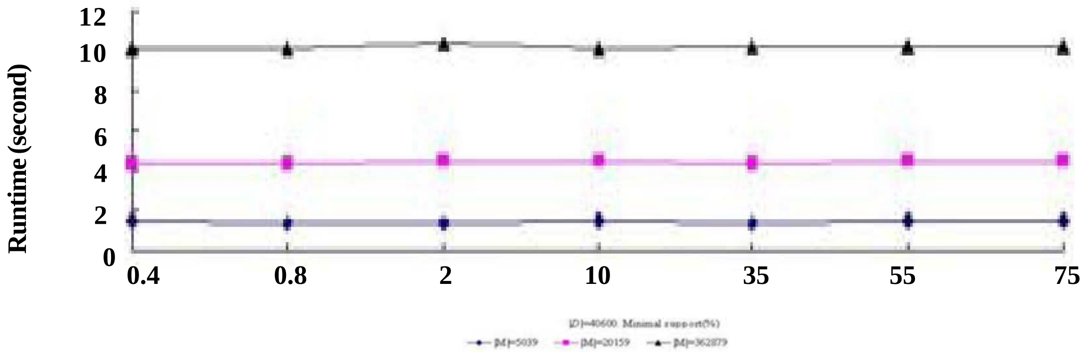


Figure 30. The varying $|M_{|S|}|$ on $|D|= 40600$

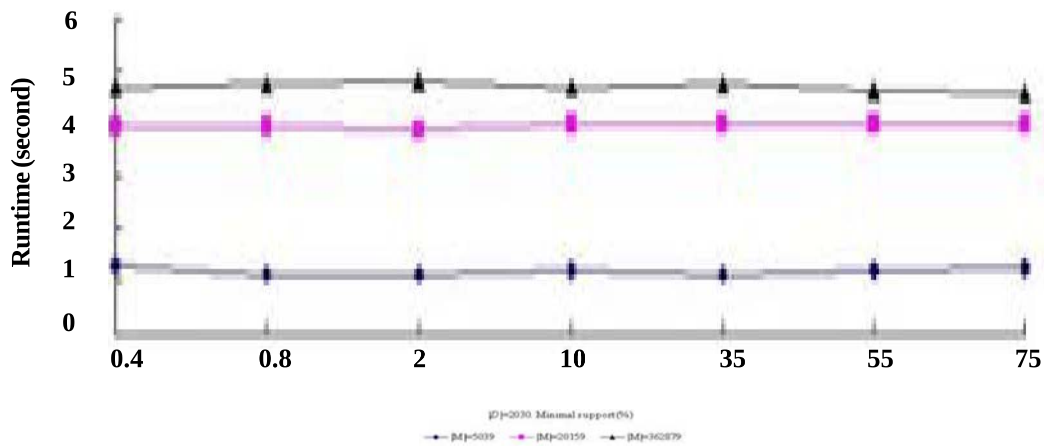


Figure 31. The varying $|D|$ on $|M_{|S|}| = 2030$

(3)Assessing the trends with the varying$|D|$ For the constant $|M_{|S|}|$, the tests with the varying $|D|$ are figure 32 to 34. We also clearly know the *Mixed-radix* becomes worse as$|D|$ increases.

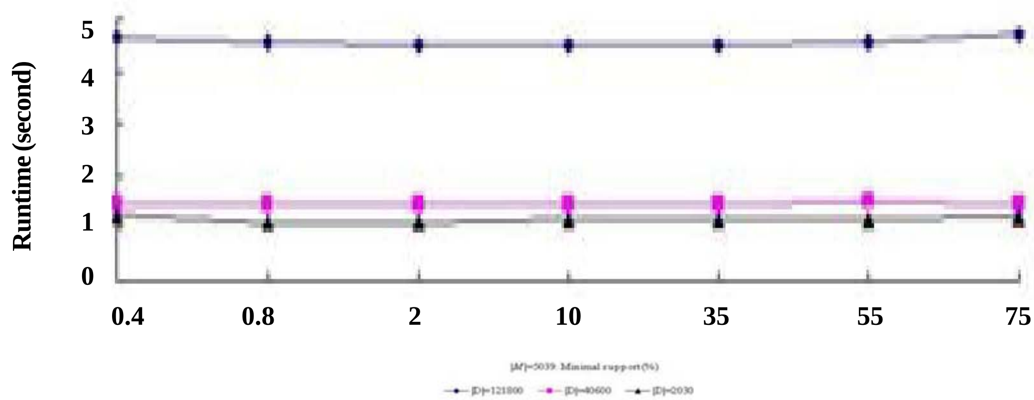


Figure 32. The varying$|D|$ on $|M_{|S|}| = 5039$ |

According to all these tests, we can draw the following conclusions with the optimal application environments for the *Mixed-radix*:

(1)The performance of the *Mixed-radix* becomes worse and worse as $|M_{|S|}|$ increases.

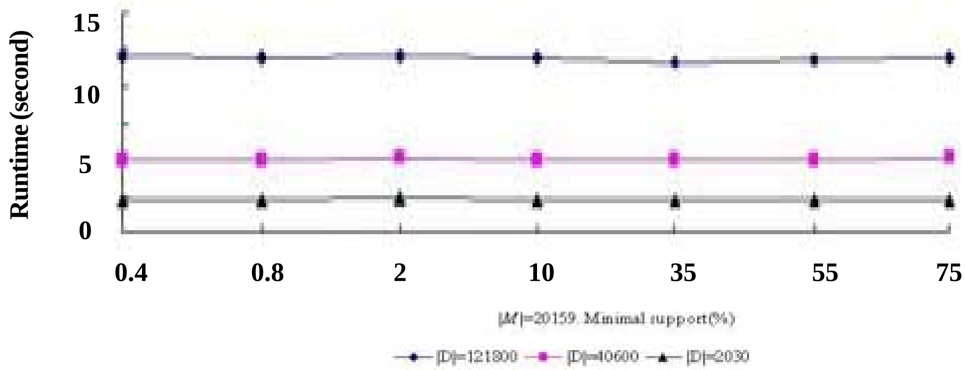


Figure 33. The varying| D| on $|M_{|S|}|$ = 20159

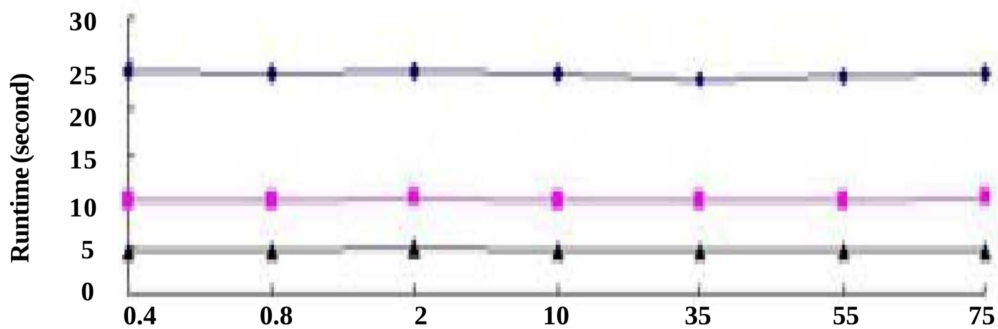


Figure 34. The varying| D| on $|M_{|S|}|$362879

(2)The performance of the *Mixed-radix* becomes worse and worse as| D| increases.

(3)The *Mixed-radix* is suitable for discovering frequent spatiotemporal association patterns on sparse patterns dataset (ρ <1) when $|M_{|S|}| \leq \mu$. (μ is a parameter with computing environments) Because the computing environments generally have the performance bottleneck, if $|M_{|S|}|$ is too large, the range of search frequent patterns will become bigger and bigger, and then the performance of *Mixed-radix* will become worse and worse. Hence, the optimal application environments need $|M_{|S|}| \leq \mu$ (for our computing environments in this paper, where $\mu = 2^{25}$).

5. Conclusions

In this paper, we introduce the new algorithm of spatiotemporal association rules mining based on event-coverage, which is suitable for discovering frequent spatiotemporal association patterns on sparse patterns dataset. Based on the feature of reading the dataset once, the algorithm is also suitable for discovering frequent patterns from dynamic datasets. But it is unsuitable for mining frequent spatiotemporal association patterns on the dense patterns datasets. Hence, we need to solve the disadvantage in the future.

Acknowledgments

This work was a project supported by Scientific and Technological Research Program of Chongqing Municipal Education Commission (Grant No.KJ121107) in China. And it was also supported by Scientific and Technological Research Program of Chongqing Three Gorges University (Grant No.13ZD20).

References

[1] Cucchiara R., Piccardi M., Mello P. (2000). Image analysis and rule-based reasoning for a traffic monitoring system. *IEEE Trans.*

on Intelligent Transportation Systems, 1 (2) 119–130.

[2] Pandey G., Atluri G., Steinbach M., et al. (2009). An association analysis approach to biclustering. *In*: Proc. of the 15th ACM SIGKDD Conf. on *Knowledge Discovery and Data Mining*, Paris, France, p. 677–686.

[3] Lee I., Phillips P. (2008). Urban crime analysis through areal categorized multivariate associations mining. *Applied Artificial Intelligence*, 22 (5) 483–499.

[4] Huang Y., Kao L., Sandnes F. (2007). Predicting ocean salinity and temperature variations using data mining and fuzzy inference. *Int. J. of Fuzzy Syst.* 9 (3) 143–151.

[5] Chang C., Shyue S. (2009). Association rules mining with GIS: An application to Taiwan census 2000. *In Proc. of the 6th int. conf. on Fuzzy systems and knowledge discovery*, Tianjin, China, p. 65–69.

[6] Zeitouni K., Yeh L., Aufaure M. (2000). Join indices as a tool for spatio data mining. *In:* Proc. of Int. Workshop on Temporal, Spatio and Spatiotemporal *Data Mining*, Berlin, Springer, p. 102–114.

[7] Mennis, J., Liu, J. (2005). Mining association rules in spatiotemporal data: An analysis of urban socioeconomic and land cover change, *Trans. in GIS* 9 (1) 5–17.

[8] Yang, H., Parthasarathy, S. (2006). Mining spatio and spatiotemporal patterns in scientific data. *In:* Proc. of 22nd Int. Conf. on Data Engineering Workshops, Atlanta, GA, USA, p x146.

[9] Lee, I. (2004). Mining multivariate associations within GIS environments. *In:* Proceedings of 17th Int. Conf. on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, Ottawa, Canada, p. 1062–1071.

[10] Ding, W., Eick, C., Wang, J., et al. (2006). A framework for regional association rule mining in spatio datasets. *In* Proc. of the Sixth IEEE Int. Conf. on Data Mining, IEEE Press, Hong Kong, p. 851–856.

[11] Yang, H., Parthasarathy S., Mehta S. (2005). Mining spatio object associations for scientific data. *In:* Proc. of the 19th Int. Joint Conf. on Artificial Intelligence, Edinburgh, UK, p. 902– 907.

[12] Boutsinas, B. (2013). A new biclustering algorithm based on association rule mining, *International Journal on Artificial Intelligence Tools*, 22 (3) 1350017-1-13.

[13] Shen, B., Yao M., Wu, Z. H. et al. (2010). Mining dynamic association rules with comments, *Knowledge and Information Systems*, 23 (01) 73– 98.

Author biographies

Gang Fang, Born in December 1978, China. He graduated from Chongqing Three Gorges University in 1999, and obtained the master degree from University of Electronic Science and Technology of China in 2009. He is working toward the PhD degree at the School of Computer Science and Engineering at the University of Electronic Science and Technology of China. Now he is an Associate Professor in Chongqing Three Gorges University. His primary research interests are in data mining, granular computing, and GIS.

Yue Wu, Born in May 1958, China. He graduated from Harbin Institute of Technology, China, in 1981, and obtained the master degree from Harbin Institute of Technology, in 1984. Now he is currently a professor in the School of Computer Science and Engineering at the University of Electronic Science and Technology of China. His primary research interests are in network computing, data mining, database system, and mobile agent technology