



程序代写代做 CS编程辅导



COMP90073 Data Mining: Methods and Applications

WeChat: cstutorcs

Assignment Project Exam Help

COMP90073

Email: tutorcs@163.com

QQ: 749389476 Sarah Erfani, CIS

<https://tutorcs.com> Semester 2, 2021

Outline

程序代写代做 CS编程辅导

- Introduction to Contrast Data Mining



- Apriori

- FP-Growth

WeChat: cstutorcs

- Applications of contrast mining in network traffic analysis and anomaly detection

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

程序代写代做 CS编程辅导

Contrast – “To compare or appraise in respect to differences” (Merriam Webster Dictionary)



Contrast data mining - The mining of patterns and models contrasting two or more datasets/conditions.

WeChat: cstutorcs

Assignment Project Exam Help

“Sometimes it's good to contrast what you like with something else. It makes you appreciate it even more”

Email: tutorcs@163.com

Darby Conley, Get Fuzzy QQ: 200749389476

<https://tutorcs.com>

What can be Contrasted?

- Objects at different *time* periods
— “Compare traffic patterns yesterday with today’s”
 - Objects for different *space* locations
— “Find the differences of location x for human DNA, versus location x for mouse DNA”
 - Object positions in a *ranking*
— “Find the differences between high- and low-income earners”
 - Objects across different *classes*
— “Find the differences between people with brown hair, versus those with blonde hair”
 - Objects *within* a class
— “Within the academic profession, there are no rich people”
— “Within computer science, most scientific articles come from USA or Europe”
- 程序代写代做 CS编程辅导
- WeChat: cstutorcs
- Assignment Project Exam Help
- Email: tutorcs@163.com
- QQ: 749389476
- <https://tutorcs.com>



- Applied to multivariate data
- Objects may be relational, graphs, models, classifiers, combinations of these
- Representation of contrasts is important. Needs to be
 - Interpretable, non redundant, potentially actionable
 - Tractable to compute
- Quality of contrasts is also important. Need
 - Statistical significance, which can be measured in multiple ways
 - Ability to rank contrasts is desirable, especially for classification



程序代写代做 CS编程辅导

- Reporting significant changes/differences
 - “Young children w
admission, compared to the rest of the population”
- Alerting, notification and warning
 - “Tell me when the dissimilarity index falls below 0.3”
- Building one/multi-class classifiers
 - Many different techniques
 - Also used for weighting and ranking instances
- Constructing synthetic instances
 - Good for rare classes

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

Example – Network Traffic Analysis

程序代写代做 CS编程辅导

- Extracting knowledge from the massive volumes of network traffic is an important task in network security management
- Network flows that are analyzed by anomaly detection systems often contain thousands of records. Analysts often check only the first few pages
- Having a concise and meaningful report of network traffic is more desirable
- An appropriate report can help managers to reduce the time and cost of security analysis and make smart decisions

WeChat: cstutorcs
Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>



Example – Network Traffic Analysis

程序代写代做 CS编程辅导

- **Summarization:** A good summarization is trade-off between two metrics:
Compaction gain and *information loss*



Table 1: Summary of network flows

	src IP	sPort	dst IP	Port	proto	flags	packets	bytes
T1	12.190.84.122	32178	100.10.20.4	80	tcp	APRS	[2,20]	[504,1200]
T2	88.34.224.2	51989	100.10.20.4	80	tcp	APRS	[2,20]	[220,500]
T3	12.190.19.23	2234	100.10.20.4	80	tcp	APRS	[2,20]	[220,500]
T4	98.198.66.23	27643	100.10.20.4	80	tcp	APRS	[2,20]	[42,200]
T5	192.168.22.4	5002	100.10.20.3	21	tcp	A-RSF	[2,20]	[42,200]
T6	192.168.22.4	5001	100.10.20.3	21	tcp	A-RSF	[40,68]	[220,500]
T7	67.118.25.23	44532	100.10.20.3	21	tcp	A-RS	[40,68]	[42,200]
T8	192.168.22.4	2765	100.10.20.4	113	tcp	APRS	[2,20]	[504,1200]
T9	98.198.66.23	5003	100.10.20.5	21	tcp	A-RSF	[2,20]	[220,500]

QQ: 749389476

<https://tutorcs.com>

Example – Network Traffic Analysis

程序代写代做 CS编程辅导

- **Summarization:** A good summarization is trade-off between two metrics:
Compaction gain and *information loss*



Table 1: Summary of network flows

	src IP	sPort	des IP	dPort	proto	flags	packets	bytes
T1	12.190.84.122	32178	100.10.20.4	80	tcp	APRS	[2,20]	[504,1200]
T2	88.34.224.2	51989	100.10.20.4	80	tcp	APRS	[2,20]	[220,500]
T3	12.190.19.23	2234	100.10.20.4	80	tcp	APRS	[2,20]	[220,500]
T4	98.198.66.23	27643	100.10.20.4	80	tcp	APRS	[2,20]	[42,200]
T5	192.168.22.4	5002	100.10.20.3	21	tcp	A-RSF	[2,20]	[42,200]
T6	192.168.22.4	5001	100.10.20.3	21	tcp	A-RSF	[40,68]	[220,500]
T7	67.118.25.23	44532	100.10.20.3	21	tcp	A-RS	[40,68]	[42,200]
T8	192.168.22.4	2765	100.10.20.4	113	tcp	APRS	[2,20]	[504,1200]
T9	98.198.66.23	5003	100.10.20.5	21	tcp	A-RSF	[2,20]	[220,500]

WeChat: estutorcs
Assignment Project Exam Help

Email: tutorcs@163.com

Table 2: Summarization by clustering

	size	src IP	sPort	des IP	dPort	proto	flags	packets	bytes
S1	5	***	*	100.10.20.4	**	tcp	APRS	[2,20]	***
S2	3	***	***	100.10.20.3	21	tcp	***	***	***

<https://tutorcs.com>

Example – Reporting Significant Differences Between Multiple Datasets

- Day1:

程序代写代做 CS编程辅导

	src IP	sPort	des IP	dPort	pro	flags	packets	bytes
T1	12.190.84.122	32	100.10.20.4	80	tcp	APRS	[2,20]	[504,1200]
T2	88.34.224.2	51989	100.10.20.4	80	tcp	APRS	[2,20]	[220,500]
T3	12.190.19.23	2234	100.10.20.4	80	tcp	APRS	[2,20]	[220,500]
T4	98.198.66.23	27643	100.10.20.4	80	tcp	APRS	[2,20]	[42,200]
T5	192.168.22.4	5002	100.10.20.3	21	tcp	A-RSF	[2,20]	[42,200]
T6	192.168.22.4	5001	100.10.20.3	21	tcp	A-RSF	[40,68]	[220,500]
T7	67.118.25.23	44532	100.10.20.3	21	tcp	A-RSF	[40,68]	[42,200]
T8	98.198.66.23	5003	100.10.20.5	21	tcp	A-RSF	[2,20]	[220,500]

- Day2:

WeChat: cstutorcs

Assignment Project Exam Help

Differences

	src IP	sPort	des IP	dPort	pro	flags	packets	bytes
T1	12.190.84.122	32178	100.10.20.4	80	tcp	APRS	[2,20]	[504,1200]
T2	88.34.224.2	51989	100.10.20.4	80	tcp	APRS	[2,20]	[220,500]
T3	12.190.19.23	2234	100.10.20.4	80	tcp	APRS	[2,20]	[220,500]
T4	98.198.66.23	27643	100.10.20.10	90	udp	---	[2,20]	[42,200]
T5	192.168.22.4	5002	100.10.20.10	90	udp	---	[2,20]	[42,200]
T6	192.168.22.4	5001	100.10.20.3	21	tcp	A-RSF	[40,68]	[220,500]
T7	67.118.25.23	44532	100.10.20.3	21	tcp	A-RSF	[40,68]	[42,200]
T8	98.198.99.23	5003	100.10.20.20	21	tcp	APRS	[40,68]	[1200,1500]

Example – Reporting Significant Differences Between Multiple Datasets

程序代写代做 CS编程辅导

- Output:

	src IP	sPort	QR Code	dPort	pro	flags	packets	bytes
C1	98.198.66.23	27643	1 	90	udp	---	[2,20]	[42,200]
C2	192.168.22.4	5002	1 	90	udp	---	[2,20]	[42,200]
C3	98.198.99.23	5003	1 	21	tcp	APRS	[40,68]	[1200,1500]

WeChat: cstutorcs

- We can use **contrast pattern mining** for finding important changes.
Assignment Project Exam Help

- **Contrast pattern mining** finds patterns whose support differs significantly from one dataset to another.

*Email: tutorcs@163.com
QQ: 749389476*

<https://tutorcs.com>

Definitions

程序代写代做 CS编程辅导

- **Itemset:** A collection of one or more items



- **k-itemset:** An itemset that contains k items

- **Count (X, D):** The number of transactions in dataset D containing pattern X

WeChat: cstutorcs

- **Support (X, D):** The percentage of transactions in dataset D containing pattern X

Assignment Project Exam Help

$$support(X, D) = \frac{Count(X, D)}{|D|}$$

Email: tutorcs@163.com

- **Frequent Itemset:** An itemset whose support is greater than or equal to a *minsup* threshold

QQ: 749389476
<https://tutorcs.com>

$$support(X, D) \geq minsup$$

程序代写代做 CS编程辅导

Method:

- Let $k=1$
- Generate frequent itemsets of length 1
- Repeat until no new frequent itemsets are identified
 - Prune candidate itemsets containing subsets of length k that are infrequent
 - Count the support of each candidate by scanning the database
 - Eliminate candidates that are infrequent, leaving only those that are frequent
 - Generate length $(k+1)$ candidate itemsets from length k frequent itemsets



WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

Apriori Principle:

- If an itemset is infrequent, then all of its superset must also be infrequent

Illustrating Apriori Principle – Revision

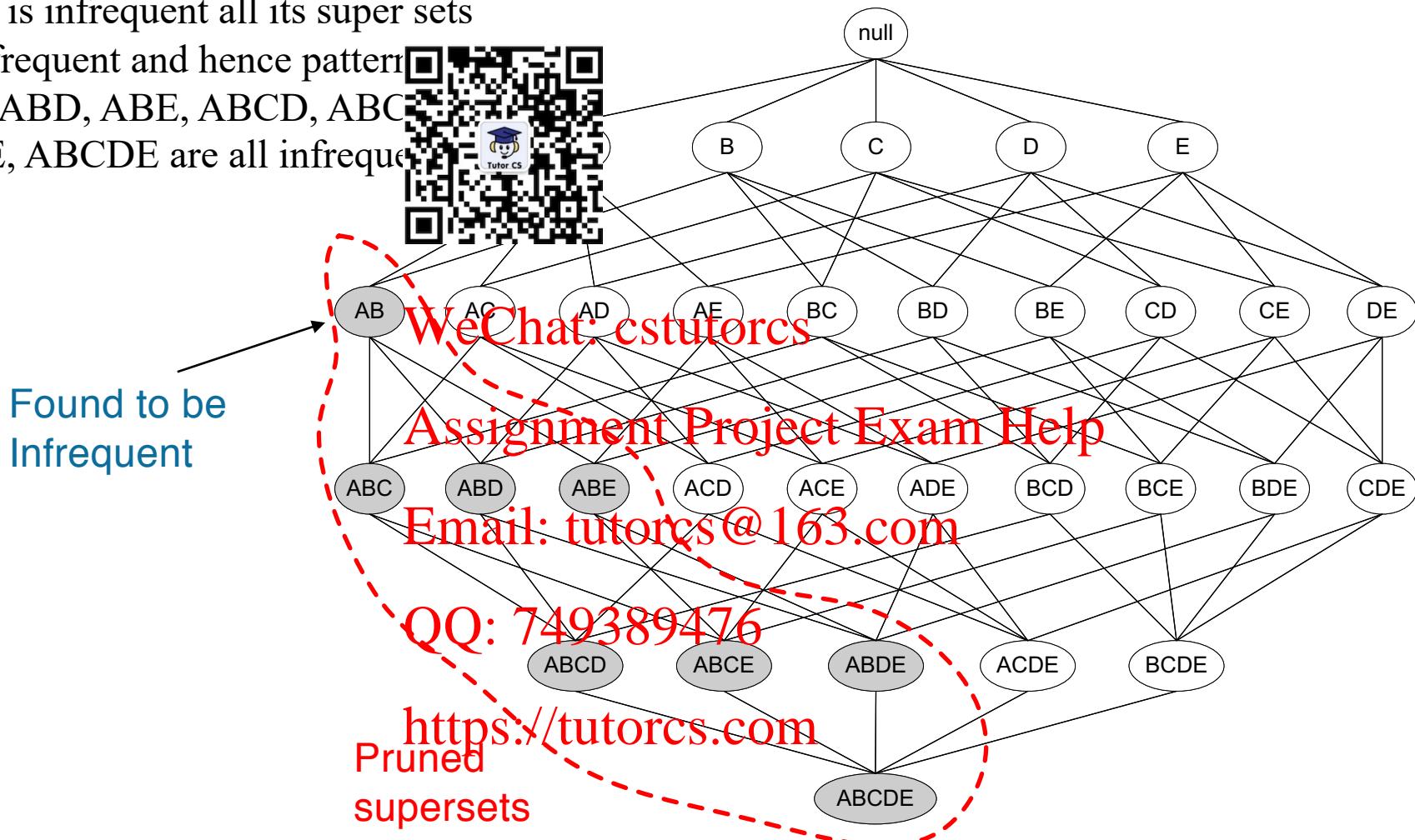
程序代写代做 CS 编程辅导

If AB is infrequent all its super sets

are infrequent and hence patterning

ABC, ABD, ABE, ABCD, ABC

ABDE, ABCDE are all infrequent



程序代写代做 CS编程辅导

- It is costly to handle a huge number of candidate sets
- If there are 10^4 frequent items, the Apriori algorithm will need to generate more than 10^{12} candidates and test their frequencies.
- Mining long patterns needs many passes of scanning and generates lots of candidates.

WeChat: cstutorcs

- It may need to repeatedly scan the whole database and check a large set of candidates by pattern matching.
- Bottleneck: candidate-generation-and-test@163.com
- Can we avoid candidate generation? [QQ: 749389476](#)

<https://tutorcs.com>



Overview of FP-Growth [2]

程序代写代做 CS编程辅导

- Find frequent single items, and partition the database based on each such item
- Recursively grow frequent patterns by doing the above for each portioned database
- To facilitate efficient processing, compress a large database into a compact, *Frequent-Pattern tree* (FP-tree) structure

WeChat: cstutors

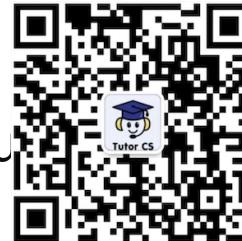
Assignment Project Exam Help

- Highly compacted, but complete for frequent pattern mining
- Avoid candidate generation
- Avoid costly repeated database scans

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

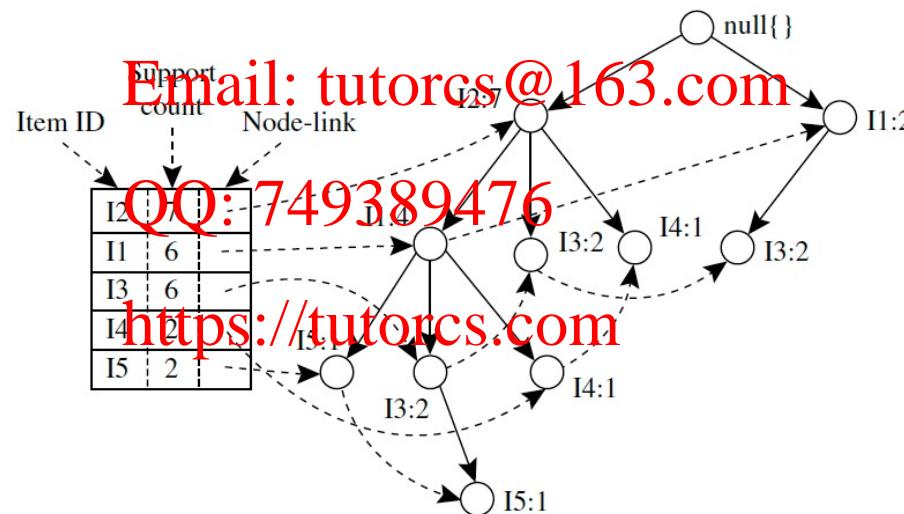


FP-Tree Definition

- FP-tree is a *frequent pattern tree*, and defined as below:


- One root labeled as “null{}”, a set of item prefix sub-trees as the children of the root, and a frequent item counter table.
- Each node in the item prefix sub-trees has three fields:
 - Item-name: Register the item this node represents,
 - Count: The number of transactions represented by the portion of the path reaching this node,

 - Node-link: Links to the next node in the FP-tree carrying the same item-name, or null if there is none.

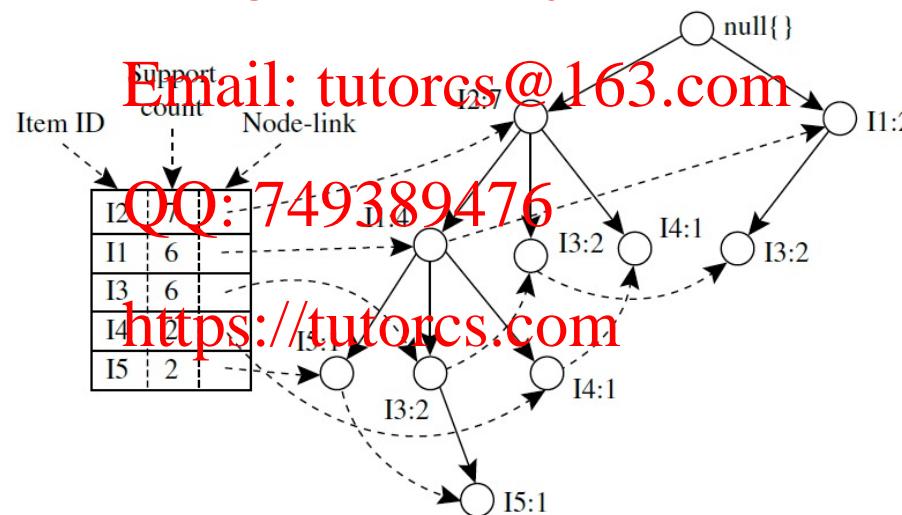
FP-Tree Definition

- 程序代写代做 CS编程辅导
- Each entry in the *frequent-item header table* has two fields,
 - Item-name,
 - Item support count
 - Head of node-link: points to the first node in the FP-tree carrying the item-name.



WeChat: cstutorcs

Assignment Project Exam Help



FP-tree: Construction and Design

STEP 1: Scan the transaction database for the first time, find frequent items (single item patterns) and order them into a list in frequency descending order.



WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

TID	List of item _IDs
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3, I6
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

FP-tree: Construction and Design

STEP 1: Scan the transaction database for the first time, find frequent items (single item patterns) and order them into a list in frequency descending order.



程序代写代做 CS编程辅导
 WeChat: cstutorcs
 L
 Assignment Project Exam Help

TID	List of item _IDs
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3, I6
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

FP-tree: Construction and Design

STEP 1: Scan the transaction database for the first time, find frequent items (single item patterns) and order them into a list in frequency descending order.



程序代写代做 CS编程辅导
 WeChat: cstutorcs
 L
 Assignment Project Exam Help

Itemset	Count
{I1}	6
{I2}	7
{I3}	6
{I4}	2
{I5}	2
{I6}	1

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

TID	List of item _IDs
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3, I6
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

FP-tree: Construction and Design

STEP 1: Scan the transaction database for the first time, find frequent items (single item patterns) and order them into a list in frequency descending order.



WeChat: cstutorcs

L

Assignment Project Exam Help

L

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

Itemset	Count
{I1}	6
{I2}	7
{I3}	6
{I4}	2
{I5}	2
{I6}	1

Minsup = 2

Itemset	Count
{I2}	7
{I1}	6
{I3}	6
{I4}	2
{I5}	2

TID	List of item _IDs
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3, I6
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

FP-tree: Construction and Design

STEP 1: Scan the transaction database for the first time, find frequent items (single item patterns) and order them into a list in frequency descending order. 

STEP 2: For each transaction, sort its frequent items according to the order; Scan database a second time, construct FP-tree by putting each **frequency ordered transaction** onto it.

WeChat: cstutorcs

TID	List of Item _IDs	Ordered Frequent Items
T100	I1, I2, I5	I2, I1, I5
T200	I2, I4	I2, I4
T300	I2, I3	I2, I3
T400	I1, I2, I4	I2, I1, I4
T500	I1, I3	I4, I3
T600	I2, I3, I6	I2, I3
T700	I1, I3	I1, I3
T800	I1, I2, I3, I5	I2, I1, I3, I5
T900	I1, I2, I3	I2, I1, I3

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

Itemset	Count
{I2}	7
{I1}	6
{I3}	6
{I4}	2
{I5}	2

- STEP 2: Construct FP-tree

Ordered Frequent Items

I2, I1, I5

I2, I4

I2, I3

I2, I1, I4

I1, I3

I2, I3

I1, I3

I2, I1, I3, I5

I2, I1, I3

程序代写代做 CS编程辅导



Null



WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

• STEP 2: Construct FP-tree

Ordered Frequent
Items

I2, I1, I5

I2, I4

I2, I3

I2, I1, I4

I1, I3

I2, I3

I1, I3

I2, I1, I3, I5

I2, I1, I3

程序代写代做 CS编程辅导



Null

WeChat: I2:1

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

• STEP 2: Construct FP-tree

Ordered Frequent
Items

I2, I1, I5

I2, I4

I2, I3

I2, I1, I4

I1, I3

I2, I3

I1, I3

I2, I1, I3, I5

I2, I1, I3

程序代写代做 CS编程辅导

<https://tutorcs.com>

• STEP 2: Construct FP-tree

Ordered Frequent Items

I2, I1, I5

I2, I4

I2, I3

I2, I1, I4

I1, I3

I2, I3

I1, I3

I2, I1, I3, I5

I2, I1, I3

程序代写代做 CS编程辅导

<https://tutorcs.com>

• STEP 2: Construct FP-tree

Ordered Frequent Items

I2, I1, I5

I2, I4

I2, I3

I2, I1, I4

I1, I3

I2, I3

I1, I3

I2, I1, I3, I5

I2, I1, I3

程序代写代做 CS编程辅导



• STEP 2: Construct FP-tree

Ordered Frequent Items

I2, I1, I5

I2, I4

I2, I3

I2, I1, I4

I1, I3

I2, I3

I1, I3

I2, I1, I3, I5

I2, I1, I3

程序代写代做 CS编程辅导

<https://tutorcs.com>

• STEP 2: Construct FP-tree

Ordered Frequent Items

I2, I1, I5

I2, I4

I2, I3

I2, I1, I4

I1, I3

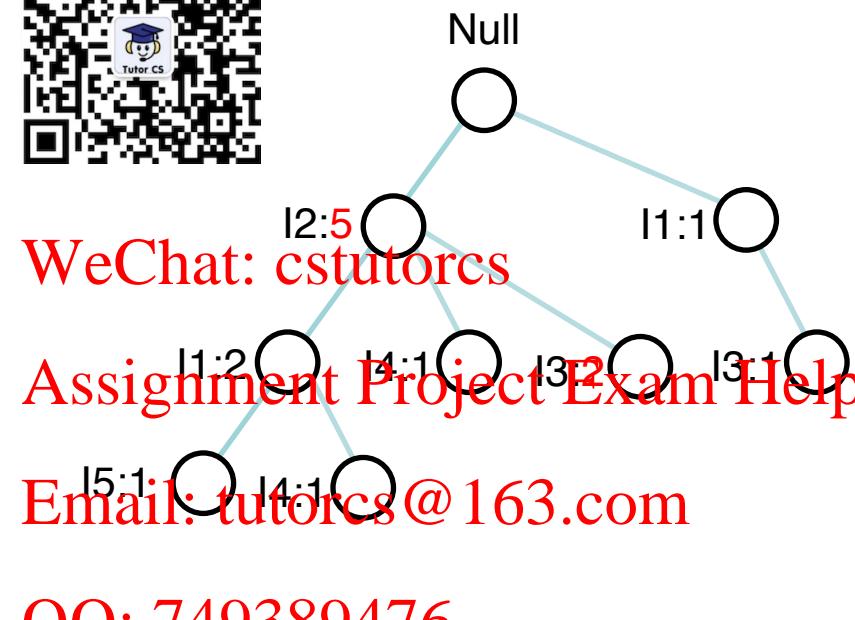
I2, I3

I1, I3

I2, I1, I3, I5

I2, I1, I3

程序代写代做 CS编程辅导



<https://tutorcs.com>

• STEP 2: Construct FP-tree

Ordered Frequent Items

I2, I1, I5

I2, I4

I2, I3

I2, I1, I4

I1, I3

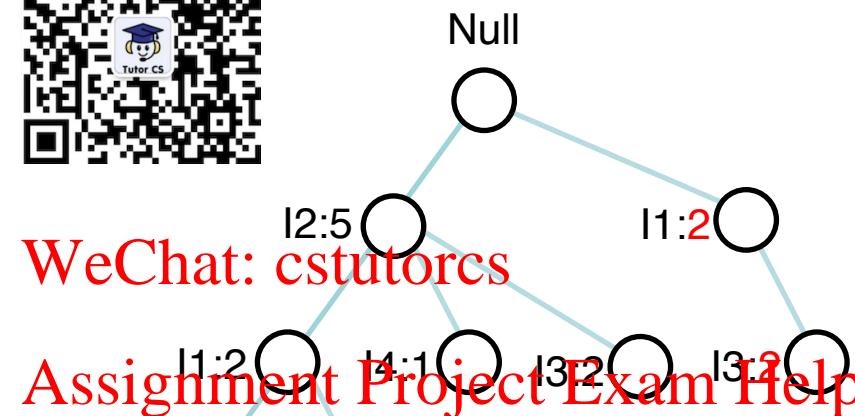
I2, I3

I1, I3

I2, I1, I3, I5

I2, I1, I3

程序代写代做 CS编程辅导



• STEP 2: Construct FP-tree

Ordered Frequent Items

I2, I1, I5

I2, I4

I2, I3

I2, I1, I4

I1, I3

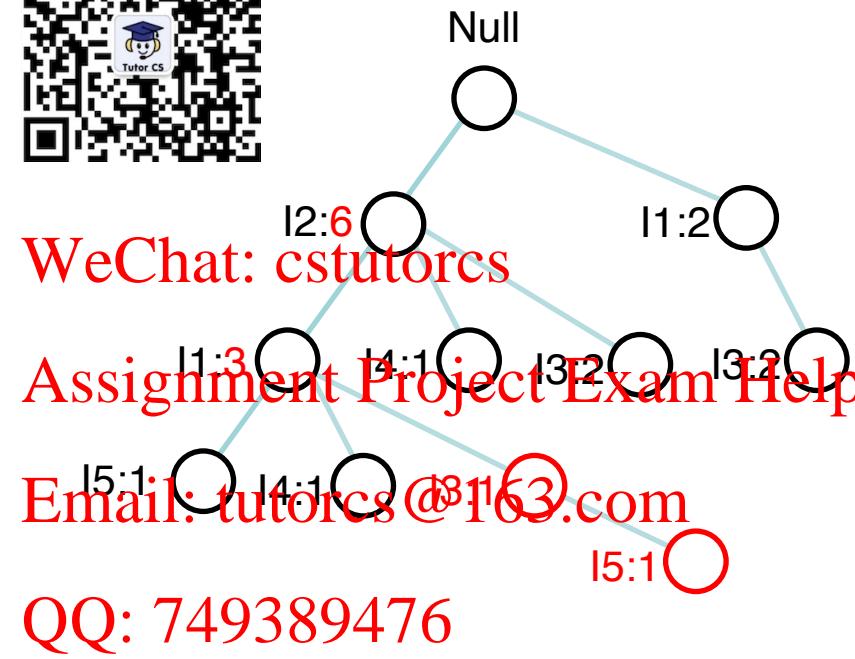
I2, I3

I1, I3

I2, I1, I3, I5

I2, I1, I3

程序代写代做 CS编程辅导



<https://tutorcs.com>

• STEP 2: Construct FP-tree

Ordered Frequent Items

I2, I1, I5

I2, I4

I2, I3

I2, I1, I4

I1, I3

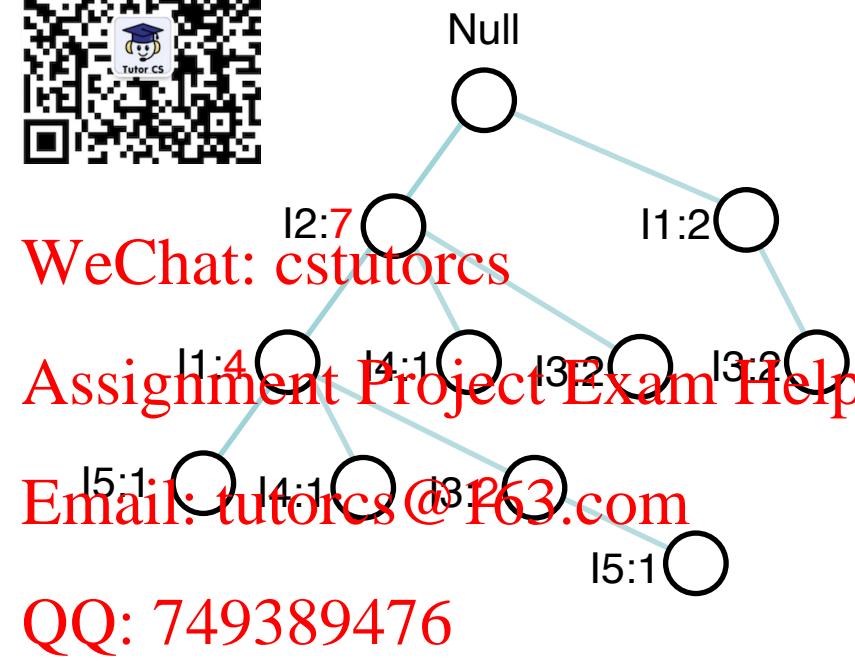
I2, I3

I1, I3

I2, I1, I3, I5

I2, I1, I3

程序代写代做 CS编程辅导



• STEP 2: Construct FP-tree

Ordered Frequent Items

I2, I1, I5

I2, I4

I2, I3

I2, I1, I4

I1, I3

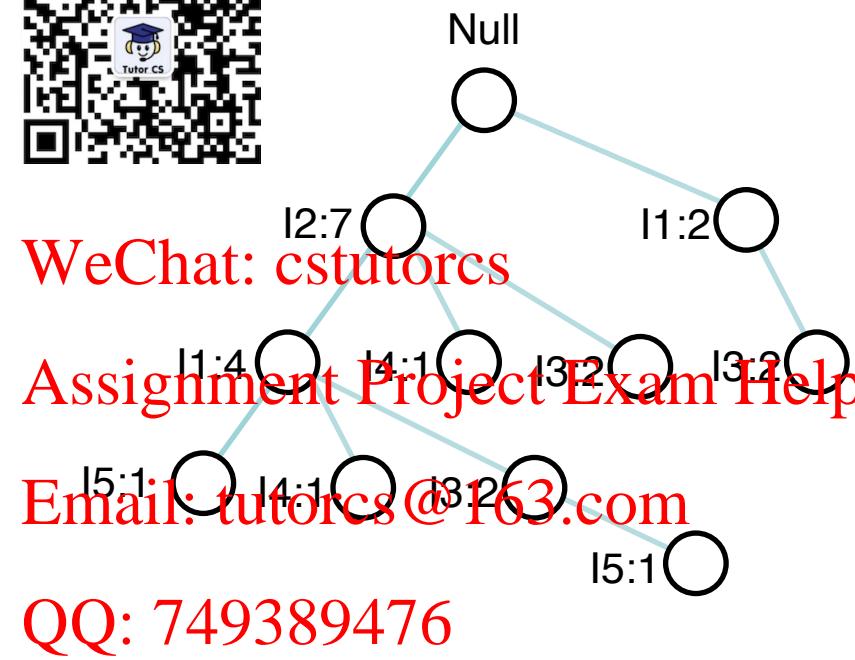
I2, I3

I1, I3

I2, I1, I3, I5

I2, I1, I3

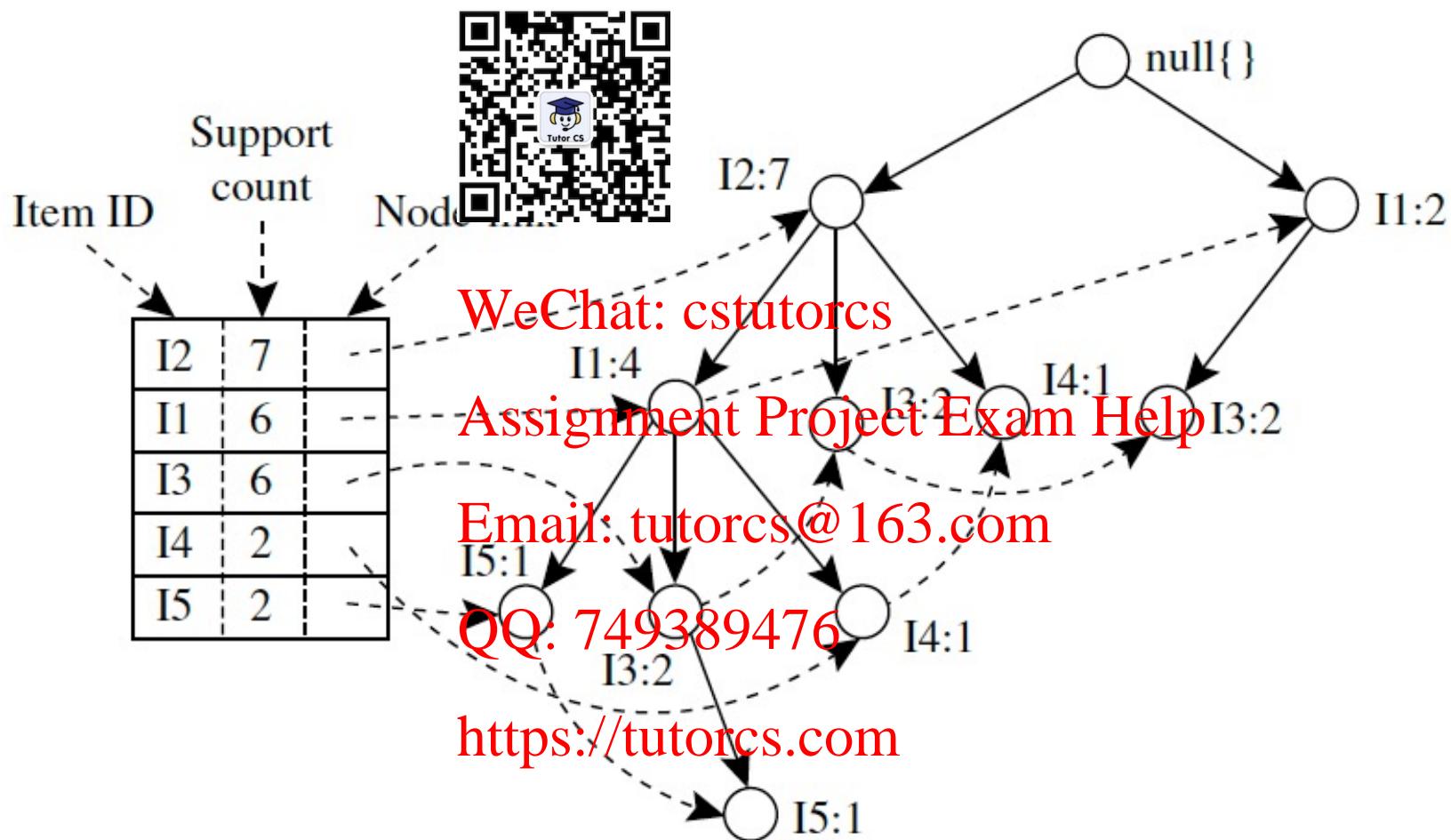
程序代写代做 CS编程辅导



<https://tutorcs.com>

FP-tree: Construction and Design

程序代写代做 CS编程辅导



程序代写代做 CS编程辅导

Starting the processing from the end of list L:

Step 1:



Construct **conditional pattern base** for each item in L

Step 2:

WeChat: cstutorcs

Construct **conditional FP-tree** from each conditional pattern base

Assignment Project Exam Help

Step 3:

Email: tutorcs@163.com

Recursively mine conditional FP-trees and grow frequent patterns obtained so far.

QQ: 749389476

<https://tutorcs.com>

- If the conditional FP-tree contains a single path, simply enumerate all the patterns

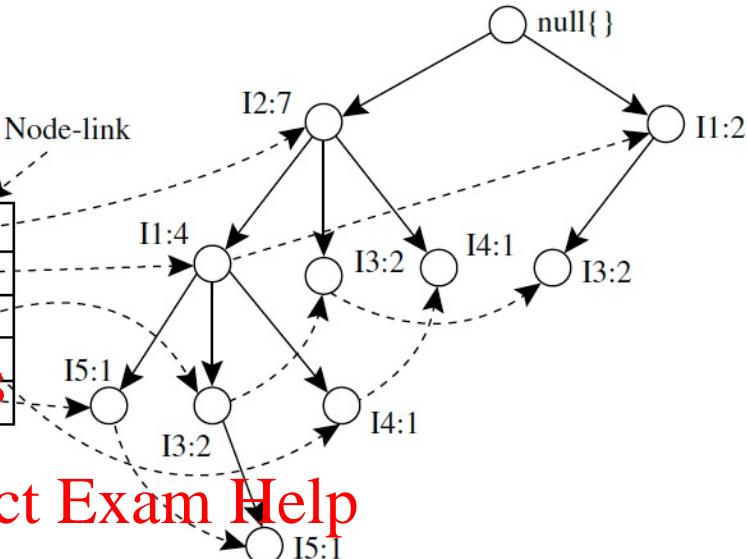
Step 1: Construct Conditional Pattern Base

程序代写代做 CS编程辅导

- Starting at the bottom of frequent-item header table of the FP-tree
- Traverse the FP-tree by following the link of each frequent item
- Accumulate all of transformed prefix paths of that item to form a conditional pattern base



Item ID	Support count	
I2	7	
I1	6	
I3	6	
I4	2	
I5	1	



WeChat: cstutorcs
Assignment Project Exam Help
Email: tutorcs@163.com

Item	Conditional Pattern Base
I5	{ {I2, I1: 1}, {I2, I1, I3: 1} }

QQ: 749389476

<https://tutorcs.com>

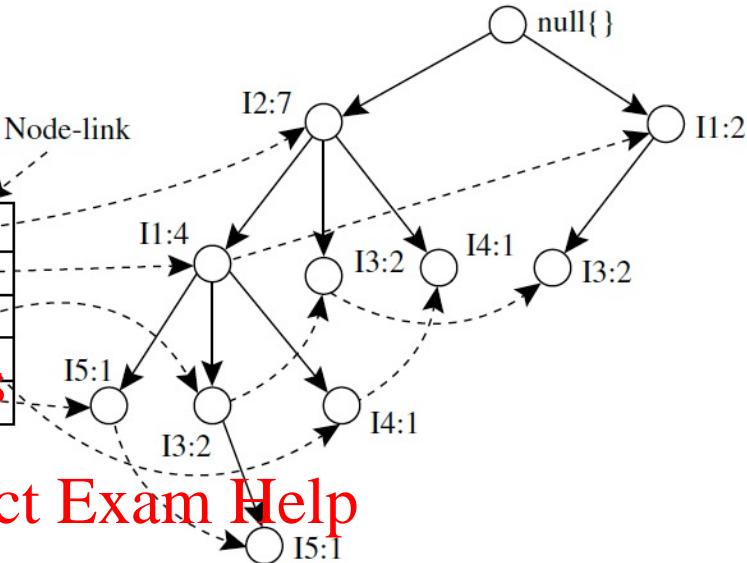
Step 1: Construct Conditional Pattern Base

程序代写代做 CS编程辅导

- Starting at the bottom of frequent-item header table, traverse the FP-tree
- Traverse the FP-tree by following the link of each frequent item
- Accumulate all of transformed prefix paths of that item to form a conditional pattern base



Item ID	Support count	
I2	7	
I1	6	
I3	6	
I4	2	
I5	1	



Item Conditional Pattern Base

I5 $\{\{I2, I1: 1\}, \{I2, I1, I3: 1\}\}$

I4 $\{\{I2, I1: 1\}, \{I2: 1\}\}$

QQ: 749389476

<https://tutorcs.com>

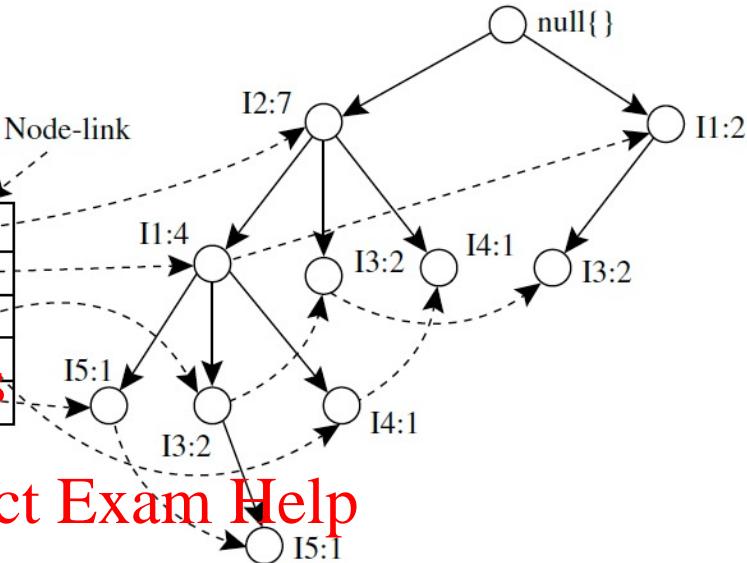
Step 1: Construct Conditional Pattern Base

程序代写代做 CS编程辅导

- Starting at the bottom of frequent-item header table, traverse the FP-tree
- Traverse the FP-tree by following the link of each frequent item
- Accumulate all of transformed prefix paths of that item to form a conditional pattern base



Item ID	Support count	
I2	7	
I1	6	
I3	6	
I4	2	
I5	1	



WeChat: cstutorcs
Assignment Project Exam Help
Email: tutorcs@163.com

Item Conditional Pattern Base

I5 $\{\{I2, I1: 1\}, \{I2, I1, I3: 1\}\}$

I4 $\{\{I2, I1: 1\}, \{I2: 1\}\}$

I3 $\{\{I2, I1: 2\}, \{I2: 2\}, \{I1: 2\}\}$

QQ: 749389476

<https://tutorcs.com>

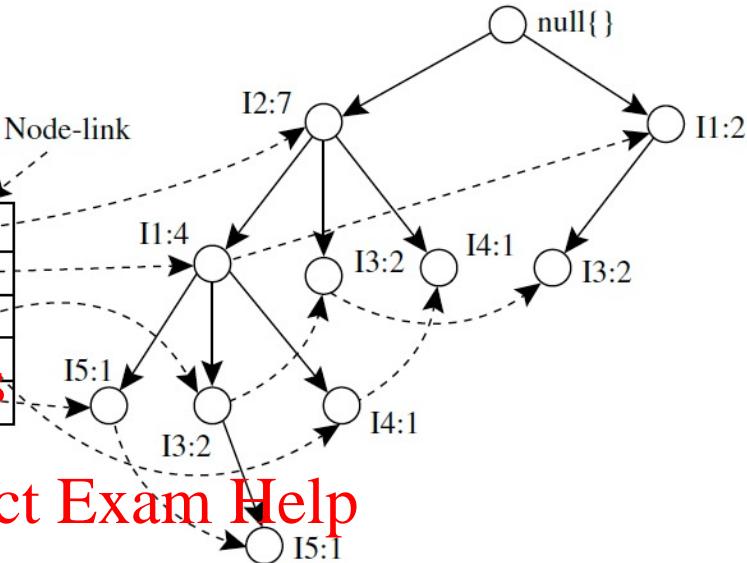
Step 1: Construct Conditional Pattern Base

程序代写代做 CS编程辅导

- Starting at the bottom of frequent-item header table, traverse the FP-tree
- Traverse the FP-tree by following the link of each frequent item
- Accumulate all of transformed prefix paths of that item to form a conditional pattern base



Item ID	Support count	
I2	7	
I1	6	
I3	6	
I4	2	
I5	1	



Item Conditional Pattern Base

I5 $\{\{I2, I1: 1\}, \{I2, I1, I3: 1\}\}$

I4 $\{\{I2, I1: 1\}, \{I2: 1\}\}$

I3 $\{\{I2, I1: 2\}, \{I2: 2\}, \{I1: 2\}\}$

I1 $\{\{I2: 4\}\}$

QQ: 749389476

<https://tutorcs.com>

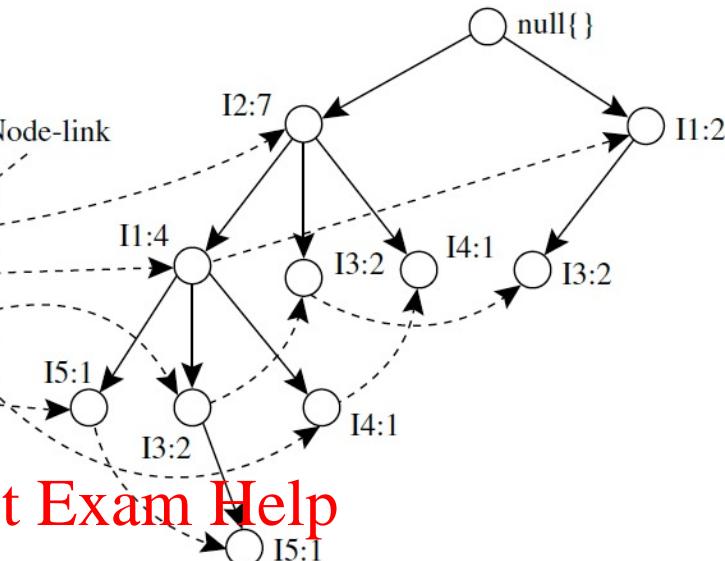
Step 2: Construct Conditional FP-tree

程序代写代做 CS 编程辅导

- For each pattern base
 - Accumulate the count for each item in the pattern base
 - Construct the conditional FP-tree for the frequent items in the pattern base
 - Minsup=2



ID	Support count
I2	7
I1	6
I3	6
I4	2
I5	2



Assignment Project Exam Help

Email: tutorcs@163.com

Item	Conditional Pattern Base	Conditional FP-tree
I5	$\{\{I2, I1: 1\}, \{I2, I1, I3: 1\}\}$	$\langle I2: 2, I1: 2 \rangle$
I4	$\{\{I2, I1: 1\}, \{I2: 1\}\}$	https://tutorcs.com
I3	$\{\{I2, I1: 2\}, \{I2: 2\}, \{I1: 2\}\}$	
I1	$\{\{I2: 4\}\}$	

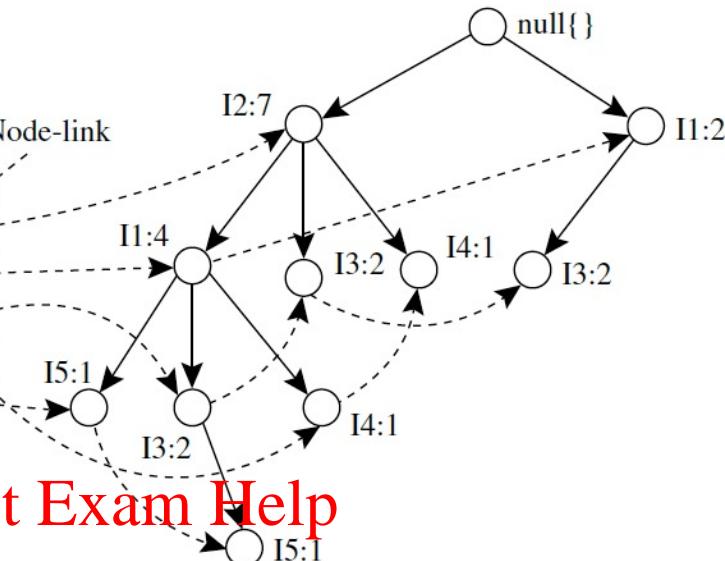
Step 2: Construct Conditional FP-tree

程序代写代做 CS 编程辅导

- For each pattern base
 - Accumulate the count for each item in the pattern base
 - Construct the conditional FP-tree for the frequent items in the pattern base
 - Minsup=2



ID	Support count
I2	7
I1	6
I3	6
I4	2
I5	2



Assignment Project Exam Help

Email: tutorcs@163.com

Item	Conditional Pattern Base	Conditional FP-tree
I5	$\{\{I2, I1: 1\}, \{I2, I1, I3: 1\}\}$	$\langle I2: 2, I1: 2 \rangle$
I4	$\{\{I2, I1: 1\}, \{I2: 1\}\}$	https://tutorcs.com
I3	$\{\{I2, I1: 2\}, \{I2: 2\}, \{I1: 2\}\}$	
I1	$\{\{I2: 4\}\}$	

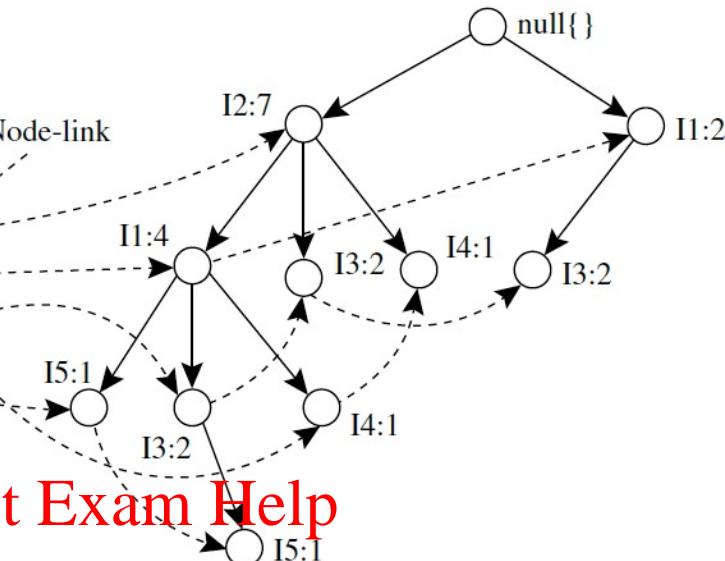
Step 2: Construct Conditional FP-tree

程序代写代做 CS 编程辅导

- For each pattern base
 - Accumulate the count for each item in the pattern base
 - Construct the conditional FP-tree for the frequent items in the pattern base
 - Minsup=2



ID	Support count
I2	7
I1	6
I3	6
I4	2
I5	2



Assignment Project Exam Help

Email: tutorcs@163.com

Item	Conditional Pattern Base	Conditional FP-tree
I5	$\{\{I2, I1: 1\}, \{I2, I1, I3: 1\}\}$	$\langle I2: 2, I1: 2 \rangle$
I4	$\{\{I2, I1: 1\}, \{I2: 1\}\}$	$\langle I2: 2 \rangle$ https://tutorcs.com
I3	$\{\{I2, I1: 2\}, \{I2: 2\}, \{I1: 2\}\}$	
I1	$\{\{I2: 4\}\}$	

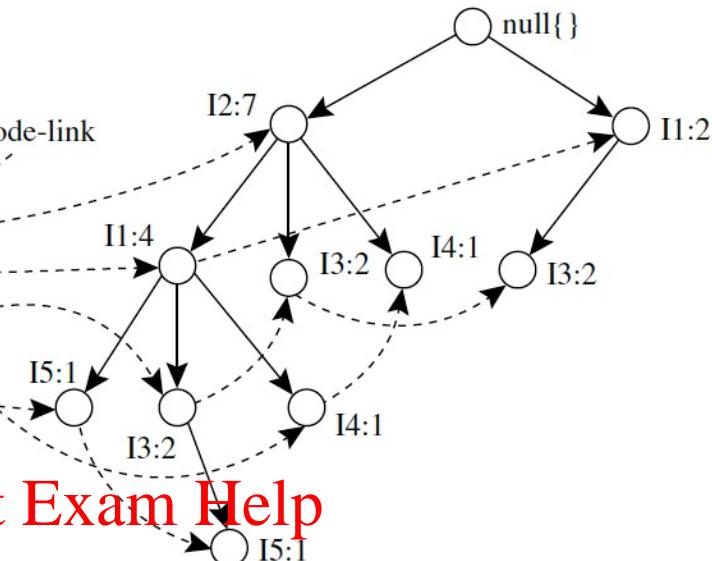
Step 2: Construct Conditional FP-tree

程序代写代做 CS 编程辅导

- For each pattern base
 - Accumulate the count for each item in the pattern base
 - Construct the conditional FP-tree for the frequent items in the pattern base
 - Minsup=2



ID	Support count
I2	7
I1	6
I3	6
I4	2
I5	2



Assignment Project Exam Help

Email: tutorcs@163.com

Item	Conditional Pattern Base	Conditional FP-tree
I5	$\{\{I2, I1: 1\}, \{I2, I1, I3: 1\}\}$	$\langle I2: 2, I1: 2 \rangle$
I4	$\{\{I2, I1: 1\}, \{I2: 1\}\}$	$\langle I2: 2 \rangle$
I3	$\{\{I2, I1: 2\}, \{I2: 2\}, \{I1: 2\}\}$	$\langle I2: 4, I1: 2 \rangle, \langle I1: 2 \rangle$
I1	$\{\{I2: 4\}\}$	

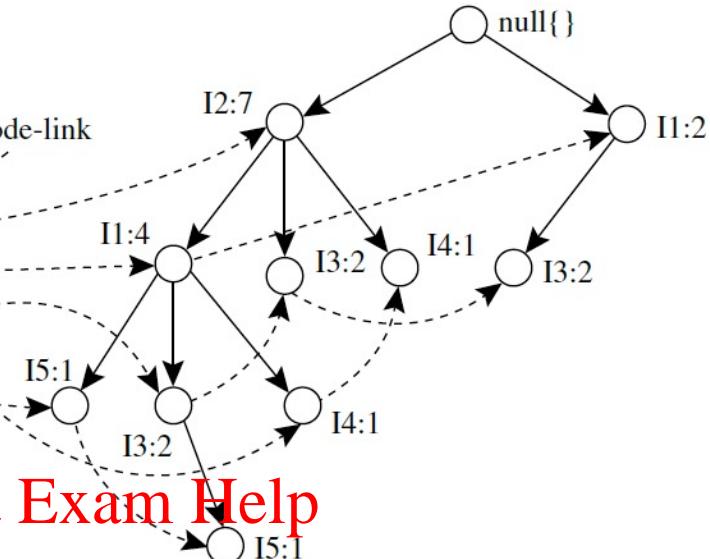
Step 2: Construct Conditional FP-tree

程序代写代做 CS编程辅导

- For each pattern base
 - Accumulate the count for each item in the base
 - Construct the conditional FP-tree for the frequent items of the pattern base
- Minsup=2



Item ID	Support count	Node-link
I2	7	
I1	6	
I3	6	
I4	2	
I5	1	



WeChat: cstutors

Assignment Project Exam Help

Email: tutorcs@163.com

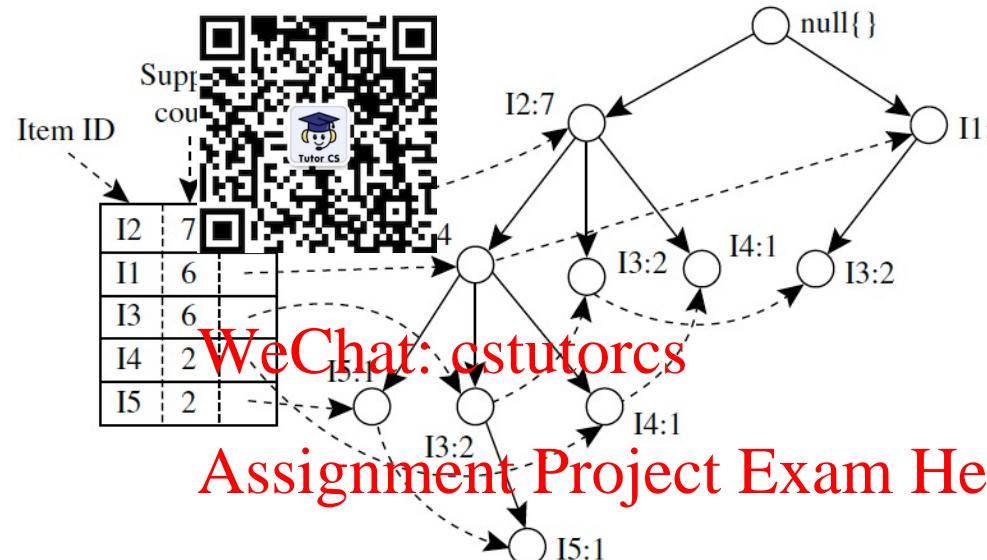
Item	Conditional Pattern Base	Conditional FP-tree
I5	$\{\{I2, I1: 1\}, \{I2, I1, I3: 1\}\}$	$\langle I2: 2, I1: 2 \rangle$
I4	$\{\{I2, I1: 1\}, \{I2: 1\}\}$	$\langle I2: 2 \rangle$
I3	$\{\{I2, I1: 2\}, \{I2: 2\}, \{I1: 2\}\}$	$\langle I2: 4, I1: 2 \rangle, \langle I1: 2 \rangle$
I1	$\{\{I2: 4\}\}$	$\langle I2: 4 \rangle$

QQ: 749389476

<https://tutorcs.com>

Step 3: Recursively Mine the Conditional FP-tree

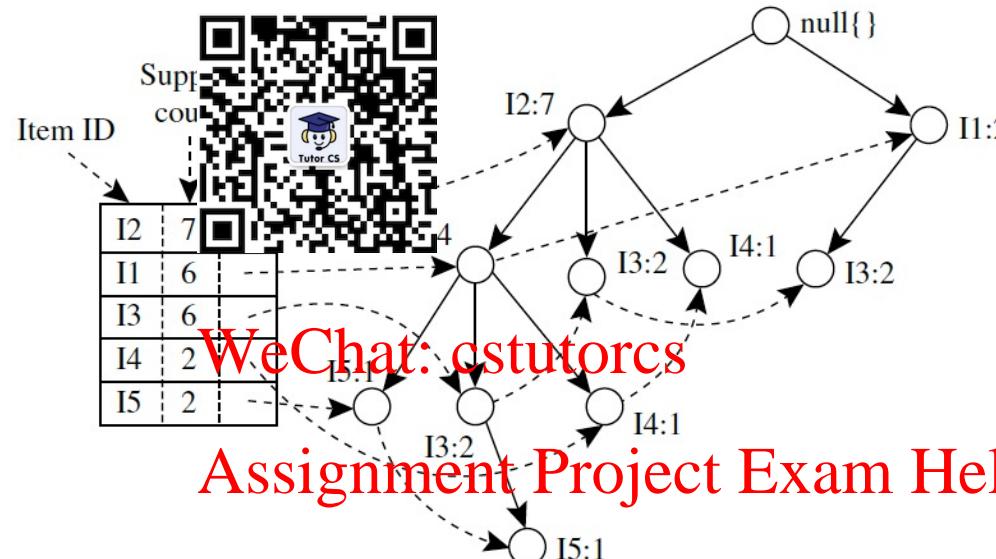
程序代写代做 CS编程辅导



Item	Conditional Pattern Base	Conditional FP-tree	Frequent Patterns Generated
I5	$\{\{I2, I1: 1\}, \{I2, I1, I3: 1\}\}$	$\langle I2: 2, I1: 2 \rangle$	$\{I2, I5: 2\}, \{I1, I5: 2\}, \{I2, I1, I5: 2\}$
I4	$\{\{I2, I1: 1\}, \{I2: 1\}\}$	$\langle I2: 2 \rangle$	https://tutorcs.com
I3	$\{\{I2, I1: 2\}, \{I2: 2\}, \{I1: 2\}\}$	$\langle I2: 4, I1: 2 \rangle, \langle I1: 2 \rangle$	
I1	$\{\{I2: 4\}\}$	$\langle I2: 4 \rangle$	

Step 3: Recursively Mine the Conditional FP-tree

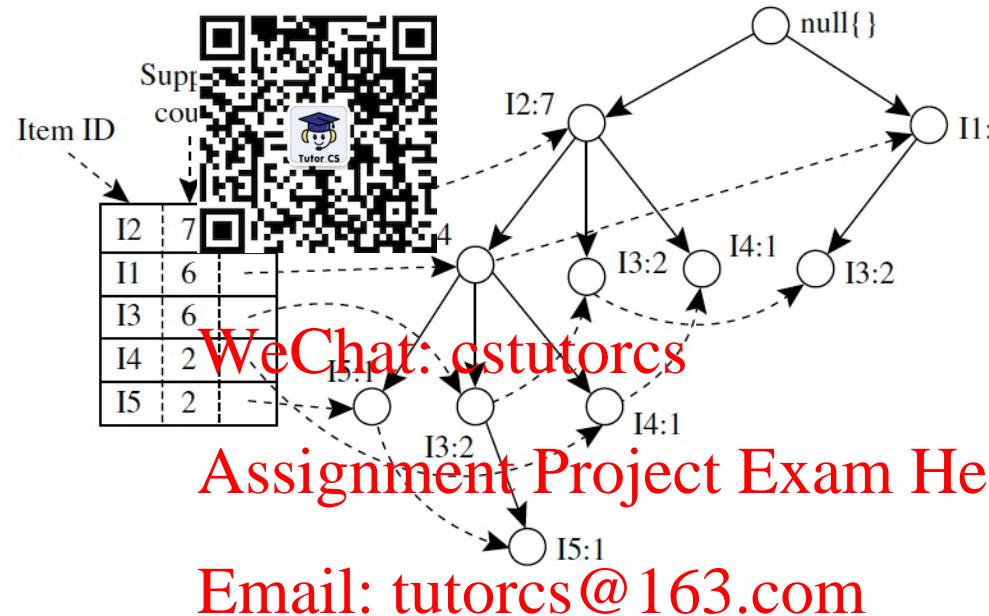
程序代写代做 CS编程辅导



Item	Conditional Pattern Base	Conditional FP-tree	Frequent Patterns Generated
I5	$\{\{I2, I1: 1\}, \{I2, I1, I3: 1\}\}$	$\langle I2: 2, I1: 2 \rangle$	$\{I2, I5: 2\}, \{I1, I5: 2\}, \{I2, I1, I5: 2\}$
I4	$\{\{I2, I1: 1\}, \{I2: 1\}\}$	$\langle I2: 2 \rangle$	$\{I2, I4: 2\}$
I3	$\{\{I2, I1: 2\}, \{I2: 2\}, \{I1: 2\}\}$	$\langle I2: 4, I1: 2 \rangle, \langle I1: 2 \rangle$	
I1	$\{\{I2: 4\}\}$	$\langle I2: 4 \rangle$	

Step 3: Recursively Mine the Conditional FP-tree

程序代写代做 CS编程辅导



Item	Conditional Pattern Base	Conditional FP-tree	Frequent Patterns Generated
I5	$\{\{I2, I1: 1\}, \{I2, I1, I3: 1\}\}$	$\langle I2: 2, I1: 2 \rangle$	$\{I2, I5: 2\}, \{I1, I5: 2\}, \{I2, I1, I5: 2\}$
I4	$\{\{I2, I1: 1\}, \{I2: 1\}\}$	$\langle I2: 2 \rangle$	$\{I2, I4: 2\}$
I3	$\{\{I2, I1: 2\}, \{I2: 2\}, \{I1: 2\}\}$	$\langle I2: 4, I1: 2 \rangle, \langle I1: 2 \rangle$	$\{I2, I3: 4\}, \{I1, I3: 4\}, \{I2, I1, I3: 2\}$
I1	$\{\{I2: 4\}\}$	$\langle I2: 4 \rangle$	

Step 3: Recursively Mine the Conditional FP-tree

程序代写代做 CS编程辅导



Item	Conditional Pattern Base	Conditional FP-tree	Frequent Patterns Generated
I5	$\{\{I2, I1: 1\}, \{I2, I1, I3: 1\}\}$	$\langle I2: 2, I1: 2 \rangle$	$\{I2, I5: 2\}, \{I1, I5: 2\}, \{I2, I1, I5: 2\}$
I4	$\{\{I2, I1: 1\}, \{I2: 1\}\}$	$\langle I2: 2 \rangle$	$\{I2, I4: 2\}$
I3	$\{\{I2, I1: 2\}, \{I2: 2\}, \{I1: 2\}\}$	$\langle I2: 4, I1: 2 \rangle, \langle I1: 2 \rangle$	$\{I2, I3: 4\}, \{I1, I3: 4\}, \{I2, I1, I3: 2\}$
I1	$\{\{I2: 4\}\}$	$\langle I2: 4 \rangle$	$\{I2, I1: 4\}$

程序代写代做 CS编程辅导

Advantages

- Only needs to read the database once as opposed to Apriori who reads it once for every iteration.
- Removes the need to calculate the pairs to be counted, which is very processing heavy, because it uses the FP-Tree. This makes it $O(n)$ (which is much faster than Apriori).

Assignment Project Exam Help

- Stores a compact version of the database in memory.

Email: tutorcs@163.com

Bottlenecks

QQ: 749389476

- The interdependency problem is that for the parallelization of the algorithm some shared data still needs to be shared, which creates a bottleneck in the shared memory.



WeChat: cstutorcs

Definitions

程序代写代做 CS编程辅导

- **Growth Rate:** Given a pair of dataset D_p (positive/target dataset) and D_n (negative/source dataset)



$$gr(X, D_p) = \frac{supp(X, D_p)}{supp(X, D_n)}$$

- **Emerging Patterns (EPs):** Patterns whose support is significantly different from one dataset to another. If $gr(X, D_p) \geq \rho$, pattern X is an emerging pattern for dataset D_p .
- *Emerging patterns* also known as *Contrast patterns (CP)*, and *Discriminative patterns*.
- **Jumping Emerging Pattern (JEP):** An emerging pattern whose support is non-zero in the positive dataset but zero in the negative dataset is called a, and $gr(X, D_p) = \infty$.

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

Discriminative patterns.

QQ: 749389476

<https://tutorcs.com>

Example – Emerging and Jumping Emerging Patterns

程序代写代做 CS 编程辅导

Positive Dataset

	Src IP	des IP	pro	packets
T1	68.22.1	10.10.10.1	udp	[2,20]
T2	68.55.2	10.10.10.4	udp	[40,68]
T3	68.22.1	10.10.10.1	tcp	[2,20]
T4	192.168.20.1	10.10.10.2	tcp	[2,20]

Negative Dataset

	WeChat	Src IP	des IP	pro	packets
T1	192.168.44.2	10.10.10.2	tcp	[40,68]	
T2	192.168.20.1	10.10.10.2	tcp	[2,20]	
T3	192.168.20.1	10.10.10.2	tcp	[2,20]	
T4	192.168.22.1	10.10.10.1	udp	[2,20]	

QQ: 749389476

- Find an EP and JEP given <https://tutorcs.com>

Example – Emerging and Jumping Emerging Patterns

程序代写代做 CS 编程辅导

Positive Dataset

	Src IP	des IP	pro	packets
T1	68.22.1	10.10.10.1	udp	[2,20]
T2	68.55.2	10.10.10.4	udp	[40,68]
T3	68.22.1	10.10.10.1	tcp	[2,20]
T4	192.168.20.1	10.10.10.2	tcp	[2,20]

Negative Dataset

	WeChat	Src IP	des IP	pro	packets
T1	Assignment Project Exam Help	192.168.44.2	10.10.10.2	tcp	[40,68]
T2	Email: tutors@163.com	192.168.20.1	10.10.10.2	tcp	[2,20]
T3	QQ: 749389476	192.168.20.1	10.10.10.2	tcp	[2,20]
T4		192.168.22.1	10.10.10.1	udp	[2,20]

Emerging and Jumping Emerging Patterns

Growth rate

C1 {srcIP=192.168.22.1, destIP=10.10.10.1, pro=[2,20]}

2

C2 {srcIP=192.168.55.2, destIP=10.10.10.4, pro=udp, Pkt=[2,20]}

∞

⋮

程序代写代做 CS编程辅导

- **Objective:** Provide a concise and meaningful report of significant changes in multiple data sets.



- Evaluate the quality of generated patterns.
- Select the best set of patterns, emerging patterns belong to either an attack class or a normal class.
- Emerging patterns can efficiently distinguish between attack and normal traffic.

Email: tutorcs@163.com

- **Extracting contrast patterns:**

QQ: 749389476

- GC-Growth algorithm

<https://tutorcs.com>

Case Study – Goodness of Contrast Data Mining for Network Traffic Analysis

程序代写代做 CS编程辅导

- Attack ratio: Measures the probability that a given contrast pattern X belongs to the attack class

$$Attack\ Ratio = \frac{Count(X, D_{pos})}{Count(X, D_{neg})}$$



- All contrast patterns with a high growth rate are attack patterns

WeChat: cstutorcs

Assignment Project Exam Help

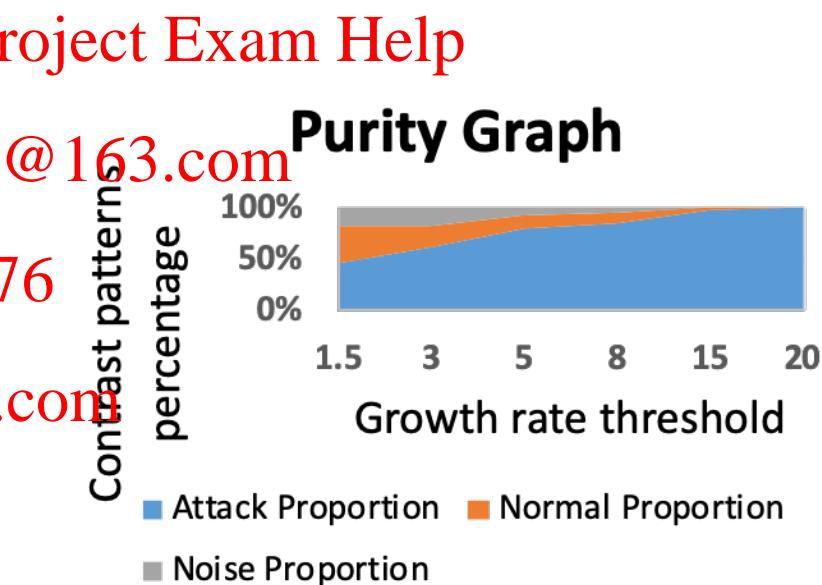
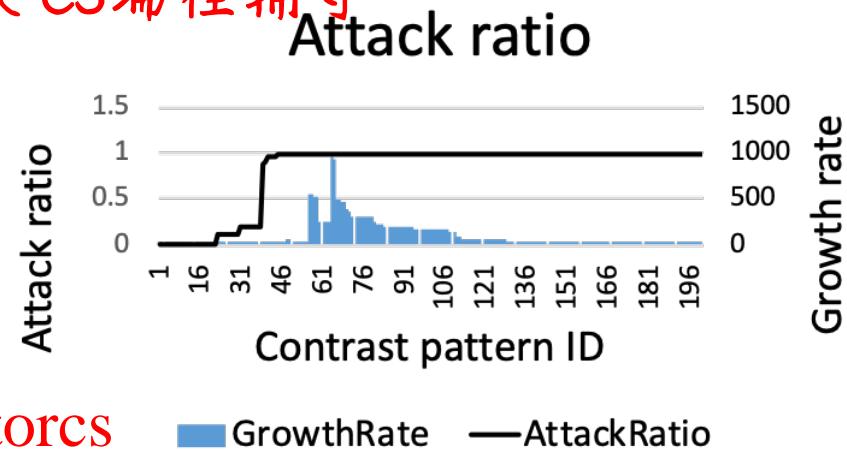
- Most of the pure patterns (i.e., patterns belong to either an attack class or a normal class) correspond to attacks

Email: tutorcs@163.com

QQ: 749389476

- The proportion of attack patterns increases significantly with an increase of the growth rate threshold

https://tutorcs.com



Application: One-class Classification [1]

程序代写代做 CS编程辅导

- **OCLEP:** Build some CP length statistics
 - Uses the training data to derive the length statistics
 - For each new test case compare the length statistics for the test case and the length statistics of the training data.
- **Property:** Provided that all transactions of T come from one class, length statistic tends to contain long EPs when test X' and train X come from the same class, and it tends to contain short EPs when test and train come from different classes.

WeChat: cstutorcs
Assignment Project Exam Help
Email: tutorcs@163.com
QQ: 749389476
<https://tutorcs.com>

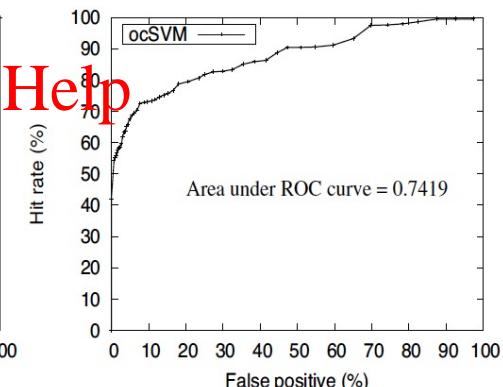
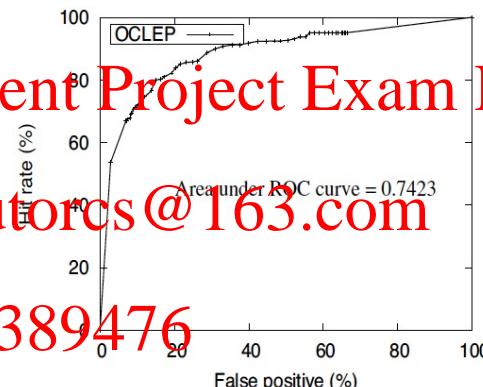


Figure: ROC curves for OCLEP and OCSVM.

Application: Comparing Anomaly Detection Models

程序代写代做 CS编程辅导

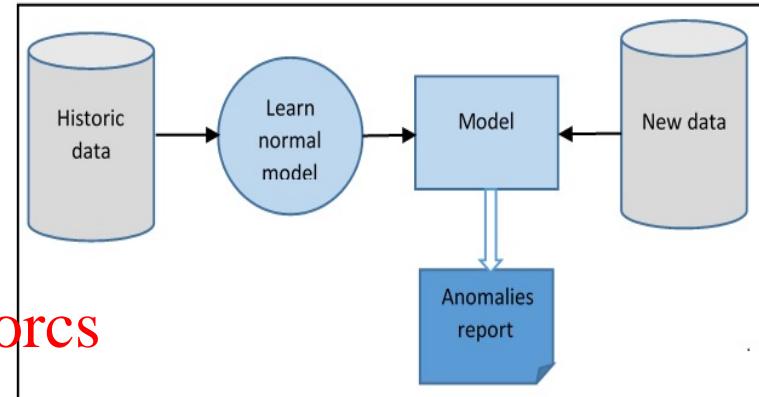
- **Anomaly detection model:**

- Learns a model of normal behaviours from historic dataset
- Applies the learned rules to current data
- Detects anomalies



WeChat: cstutorcs

Anomaly detection model



- **CPM technique:**

Assignment Project Exam Help

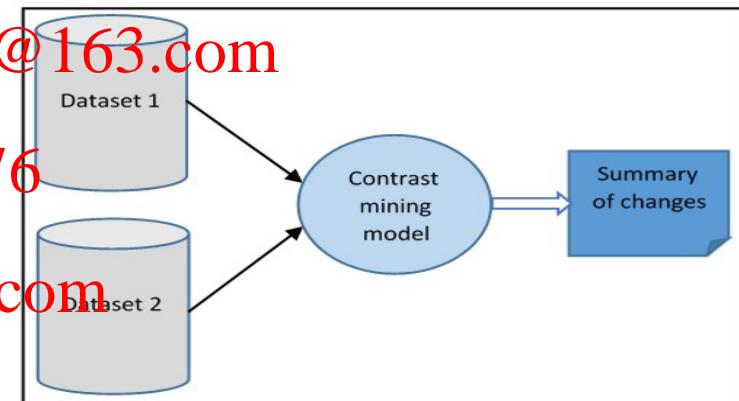
Contrast mining model

- Compares two current dataset and historic dataset
- Extracts significant changes
- Presents a succinct report

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>



Summary

程序代写代做 CS编程辅导

- Why contrast data mining is important and when it can be used?
- What algorithms can be used for contrast data mining?
- How it can be used for network traffic analysis and unsupervised learning?
WeChat: cstutorcs



Assignment Project Exam Help

Next: Adversarial Machine Learning

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

References

程序代写代做 CS编程辅导

1. Guozhu Dong and James Bailey. "Contrast data mining: concepts, algorithms, and applications". CRC Press, 2012.
2. Jiawei Han, Micheline Kamber, Jian Pei, "Data Mining: Concepts and Techniques ", 2011, China
3. Elaheh Alipour Chavary, Sarah Erfani, Christopher Leckie, "Summarizing Significant Changes in Network Traffic Using Contrast Pattern Mining", ACM International Conference on Information and Knowledge Management (CIKM), 2017.

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

