Association Rule Mining

Jiaze Li (23266049)

The process of association rule mining

1. Obtain two related views in SSMS.

After importing the data into SQL Server Management Studio, I create two groups of views for association rules mining. (The code of creating views was in the file of *CreateViews.sql*).

View name	Primary/Foreign Key	Other Column
vAssoCase02	WNKey	SubCrimeType
vAssoNested02	WNKey	Date, beat
vAssoCase03	DBKey	SubCrimeType
vAssoNested03	DBKey	Date, SubBuildingType

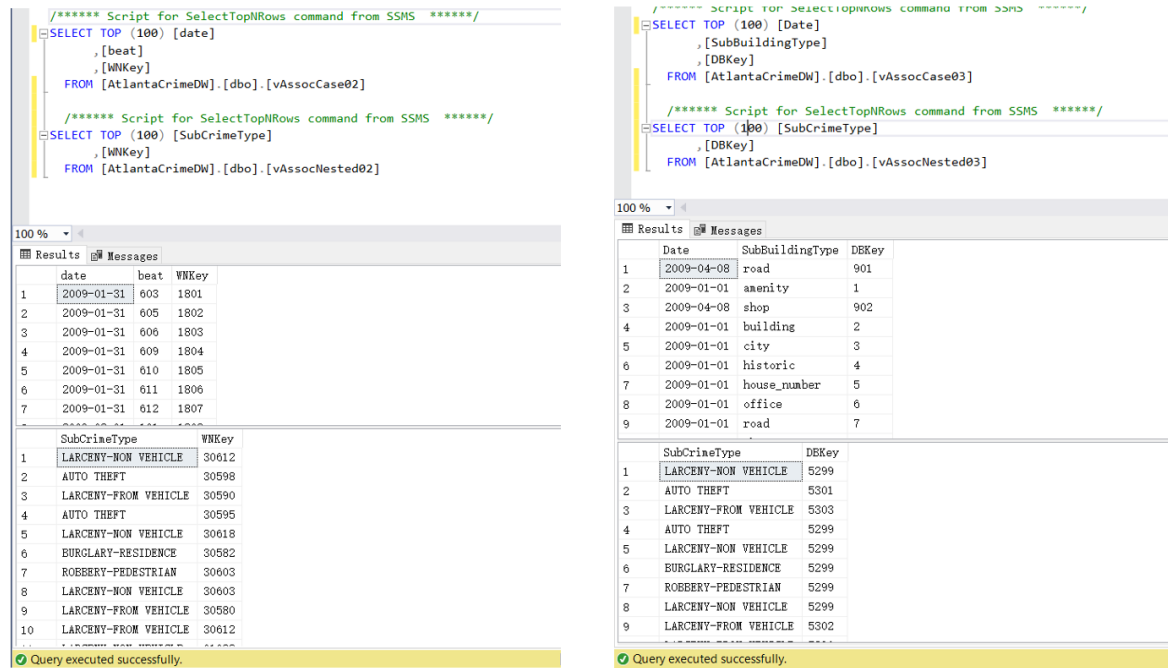*Table 1. Two groups of views for association rule mining*

Figure 1. Select top 100 rows to look at views results

For the first group of views (figure 1, left), we can see that in "dbo.vAssocNested02", the same order (identified by "WNKey") appear in multiple rows, as each SubCrimeType is associated with multiple crime types, and in "dob.vAssocCase02", each WNKey occupies a single row, which records the crime occur information (time and location). The objective of this association rule mining is to find what crime types occur together.

2. Establish the correct relationship in Visual Studio

In the operation interface of Visual Studio, I imported the relevant database from the data engine (SSMS) first, and create a data source view, then establish the relationship between two views.
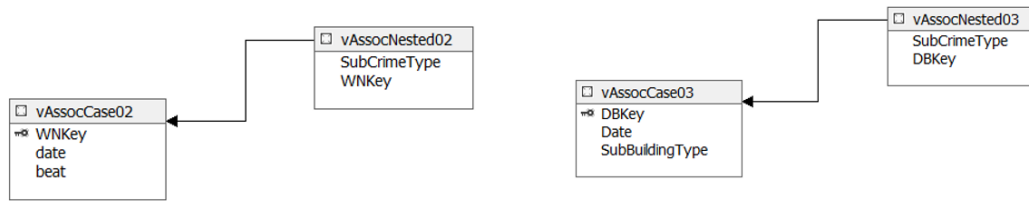

Figure 2. The relationship between two views

3. Estabilish Data Mining

Select the correct table for case table and nested table (Figure4). Then choose the single key, predictable column, and input columns (Figure5).
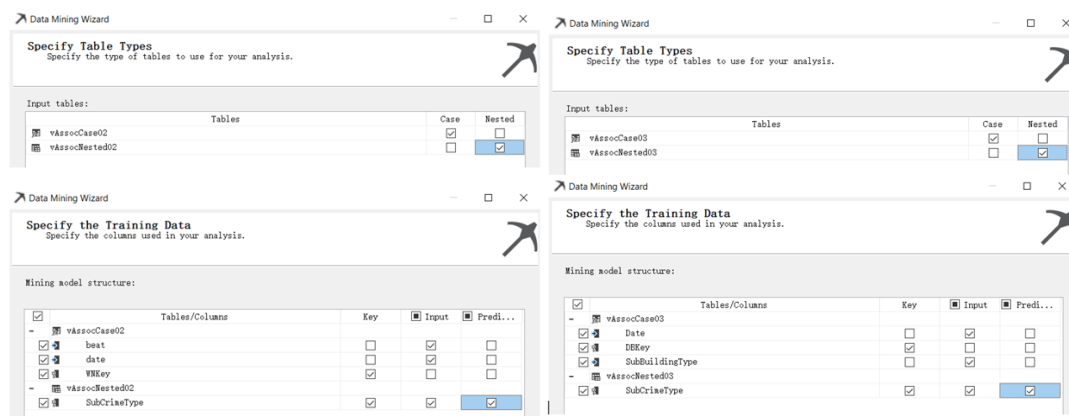

Figure 3. Case tables and Nested tables setting

After all, we can see the relationship plots as follow, then star processing the Data Mining and got result.
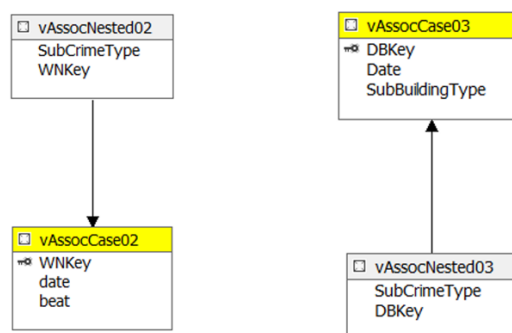

Figure 4. Relationship plots of two cases

The results of association rule mining

Rule1:

The rules and dependency network displayed after using association rule mining are as follow (figure 5).
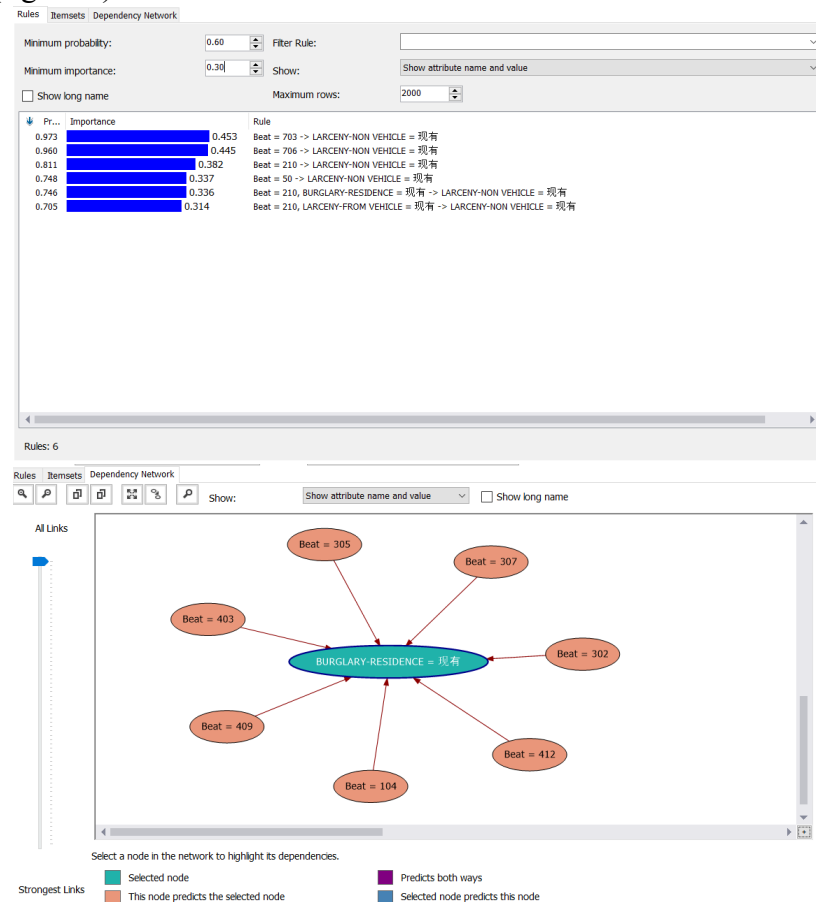

Figure 5. Rules and dependency network of case1[1]

This graph illustrates the following association rule:
When one or more events from {Beat104, Beat305, Beat307, Beat302, Beat403, Beat409, Beat412} occur => BURGRALY-RESIDENCE event is likely to occur.

From this association rule, we can infer a phenomenon: these specific areas (Beat104, Beat305, Beat307, Beat302, Beat403, Beat409, Beat412) may have certain security risks and are highly correlated with residential burglary events. Based on this phenomenon, relevant departments or communities can strengthen security measures in these areas to reduce the risk of residential burglaries.

---

[1] The computer I used has already changed the language, however, it still showed some Chinese word ('现有') above, which means 'existing'.
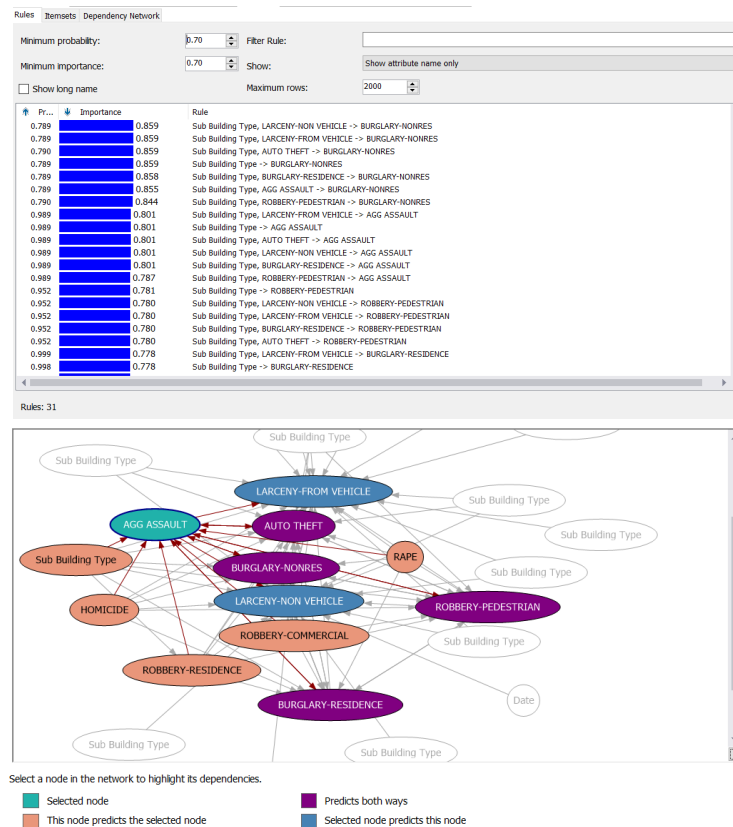
Rule2:

Figure 6. Rules and dependency network of case2

This graph illustrates the following association rules and phenomena:

1. Orange nodes predicting the green node:

When events from {Residential, HOMICIDE, ROBBERY-RESIDENCE, ROBBERY-COMMERCIAL, RAPE} occur => AGG ASSAULT event is likely to occur.

This phenomenon suggests that there is a correlation between these crime types and aggravated assault. When events like homicide, residential robbery, commercial robbery, and rape occur, there is a higher probability of an aggravated assault taking place.

2. Green node predicting the blue nodes:

When AGG ASSAULT event occurs => events from {LARCENY-FROM VEHICLE, LARCENY-NON VEHICLE} are likely to occur.

This phenomenon suggests that when an aggravated assault occurs, it may be followed by or related to larceny events, including larceny from vehicles and other types of larceny.

3. Purple nodes predicting both ways with the green node:

- When AUTO THEFT event occurs <=> AGG ASSAULT event is likely to occur and vice versa.
- When BURGLARY-NONRES event occurs <=> AGG ASSAULT event is likely to occur and vice versa.
- When BURGLARY-RESIDENCE event occurs <=> AGG ASSAULT event is likely to occur and vice versa.
- When ROBBERY-PEDESTRIAN event occurs <=> AGG ASSAULT event is likely to occur and vice versa.

This phenomenon suggests that there is a strong correlation between aggravated assault and these crime types. They can be considered as mutually predictive events.

Based on these association rules and phenomena, relevant authorities or communities can focus on addressing these correlated crime types to effectively reduce the risk and occurrence of aggravated assaults. By understanding these relationships, we can better allocate resources and devise strategies to combat crime.

Top K rules with crime

Top k association rules are the k most frequent association rules in the database having a confidence that is higher or equal to minimum probability and minimum importance. Like when we set k=3, the minimum probability = 0.70 and minimum importance = 0.70, we would get the top 3 rules that the confidence higher than 0.70.

Probability	Importance	Rules
0.744	0.815	Residential, ROBBERY-RESIDENCE -> **BURGLARY-NONRES**
0.719	0.814	Residential, LARCENY-FROM VEHICLE -> **BURGLARY-NONRES**
0.718	0.810	Residential, AGG ASSAULT-> **BURGLARY-NONRES**

Table 2. Top3 rules with Atlanta crime data

In this example, the top 3 rules highlight specific associations between different crime types. These rules have been ranked based on their importance, which is a metric that indicates the strength and relevance of the relationships in the context of the dataset.

The meaning of the k rules

In plain English, "k rules" refer to a specific number (k) of rules, principles, or guidelines in each context. These rules help us understand patterns or relationships within a particular system or situation. In this project, we are focusing on Atlanta crime data and the three rules generated through association rule mining.

Given these top 3 rules, we can provide the following interpretation:
- Rule 1 suggests that when residential robbery occurs, there is a high probability (74.4%) of non-residential burglary occurring. The importance of this rule is 0.815, indicating that it has a high priority among the mined rules.
- Rule 2 suggests that when larceny from a vehicle occurs in a residential area, there is a high probability (71.9%) of non-residential burglary occurring. The importance of this rule is 0.814, also indicating that it has a high priority among the mined rules.
- Rule 3 suggests that when an aggravated assault occurs in a residential area, there is a high probability (71.8%) of non-residential burglary occurring. The importance of this rule is 0.810, again indicating that it has a high priority among the mined rules.

These rules suggest that specific crime types in residential areas (ROBBERY-RESIDENCE, LARCENY-FROM VEHICLE, AGG ASSAULT) have a strong relationship with the occurrence of BURGLARY-NONRES which means in Atlanta, if a crime involves robbery or vehicle theft or aggravated assault in a residential area, there is a high probability that a non-residential burglary will follow. Law enforcement could use this insight to be more proactive in preventing non-residential burglaries after incidents of aggravated assault in residential areas by increasing police presence and vigilance.