Frequent Pattern Mining in Data Mining

Frequent pattern mining in data mining is the process of identifying patterns or associations within a dataset that occur frequently. This is typically done by analyzing large datasets to find items or sets of items that appear together frequently.

There are several different algorithms used for frequent pattern mining, including:

1. Apriori algorithm: This is one of the most commonly used algorithms for frequent pattern mining. It uses a "bottom-up" approach to identify frequent itemsets and then generates association rules from those itemsets.
2. ECLAT algorithm: This algorithm uses a "depth-first search" approach to identify frequent itemsets. It is particularly efficient for datasets with a large number of items.
3. FP-growth algorithm: This algorithm uses a "compression" technique to find frequent patterns efficiently. It is particularly efficient for datasets with a large number of transactions.
4. Frequent pattern mining has many applications, such as Market Basket Analysis, Recommender Systems, Fraud Detection, and many more.

Advantages:

1. It can find useful information which is not visible in simple data browsing
2. It can find interesting association and correlation among data items

Disadvantages:

1. It can generate a large number of patterns
2. With high dimensionality, the number of patterns can be very large, making it difficult to interpret the results.

The increasing power of computer technology creates a large amount of data and storage. Databases are increasing rapidly and in this computerized world everything is shifting online and data is increasing as a new currency. Data comes in different shapes and sizes and is collected in different ways. By using data mining there are many benefits it helps us to improve the particular process and in some cases, it costs saving or revenue generation. Data mining is commonly used to search a large amount of data for patterns and trends, and not only for searching it uses the data for further processes and develops actionable processes.

Data mining is the process of converting raw data into suitable patterns based on trends.

Data mining has different types of patterns and **frequent pattern mining** is one of them. This concept was introduced for mining transaction databases. Frequent patterns are patterns(such as items, subsequences, or substructures) that appear frequently in the database. It is an analytical process that finds frequent patterns, associations, or causal structures from databases in various databases. This process aims to find the frequently occurring item in a transaction. By frequent patterns, we can identify strongly correlated items together and we can identify similar characteristics and associations among them. By doing frequent data mining we can go further for clustering and association.

Frequent pattern mining is a major concern it plays a major role in associations and correlations and disclose an intrinsic and important property of dataset.

Frequent data mining can be done by using association rules with particular algorithms eclat and apriori algorithms. Frequent pattern mining searches for recurring relationships in a data set. It also helps to find the inheritance regularities. to make fast processing software with a user interface and used for a long time without any error.
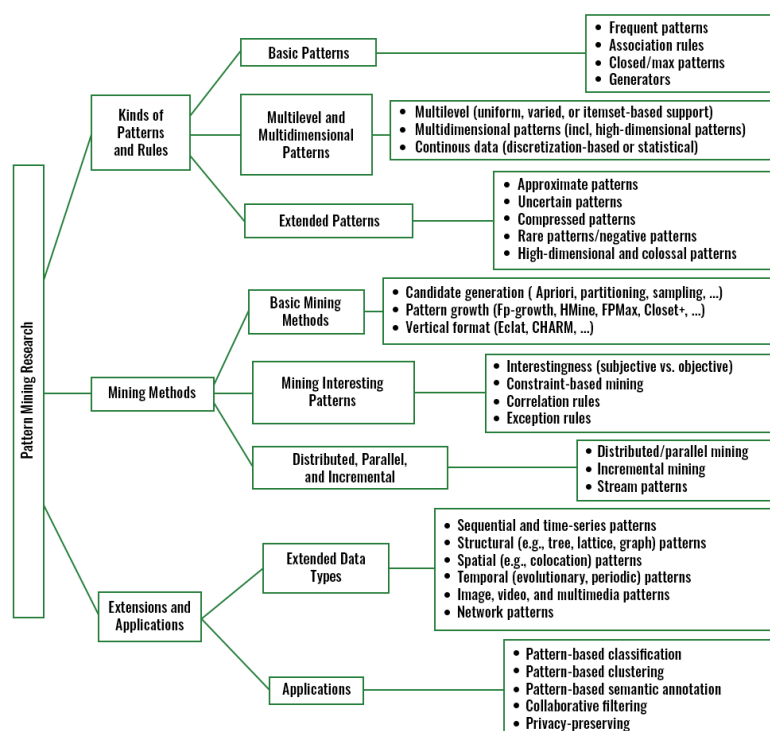
Association Rule Mining:

It is easy to find associations in frequent patterns:

- for each frequent pattern x for each subset y c x.
- calculate the support of y-> x – y.

if it is greater than the threshold, keep the rule. There are two algorithms that support this lattice

1. Apriori algorithm
2. eclat algorithm

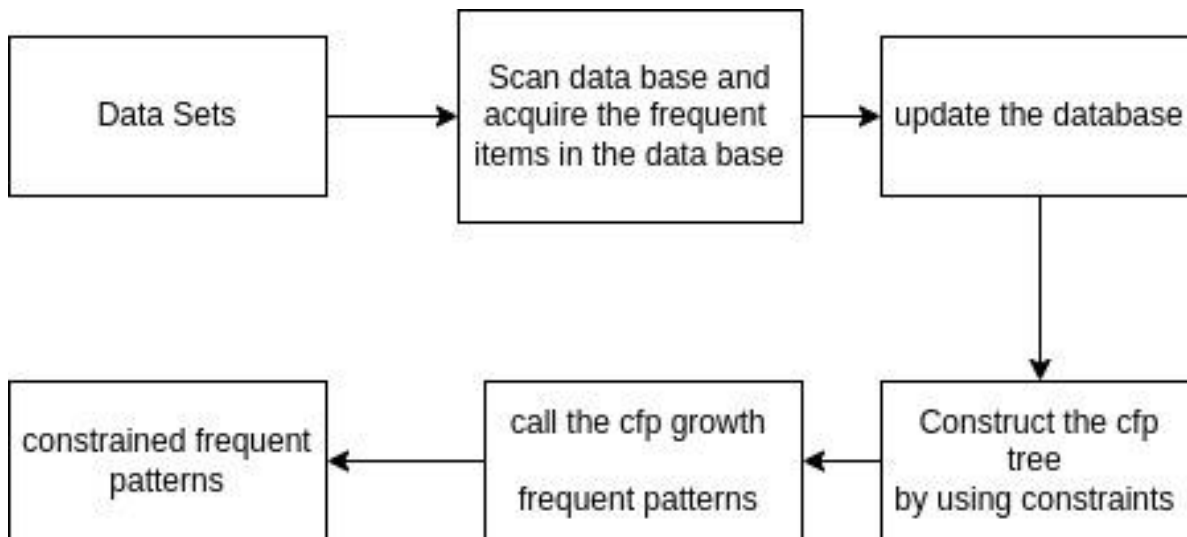| Apriori	Eclat
It performs "perfect" pruning of infrequent item sets.	It reduces memory requirements and is faster.
It requires a lot of memory(all frequent item sets are represented) and support counting takes very long for large transactions. But this is not efficient in practice.	Its storage of transaction list.

The words support and confidence support the association rule.

- **Support:** how often a given rule in a database is mined? support the transaction contains x U y
- **Confidence:** the number of times the given rule in a practice is true. The conditional probability is a transaction having x as well as y.

```
┌─────────────────┐      ┌─────────────────────┐      ┌─────────────────────┐
│                 │      │  Scan data base and │      │                     │
│   Data Sets     │─────▶│  acquire the frequent│────▶│ update the database │
│                 │      │  items in the data base│    │                     │
└─────────────────┘      └─────────────────────┘      └─────────────────────┘
                                                                  │
                                                                  │
                                                                  ▼
┌─────────────────┐      ┌─────────────────────┐      ┌─────────────────────┐
│                 │      │  call the cfp growth│      │  Construct the cfp  │
│ constrained frequent│◀─│                     │◀─────│       tree          │
│    patterns     │      │  frequent patterns  │      │ by using constraints│
└─────────────────┘      └─────────────────────┘      └─────────────────────┘
```

working principle (it is a simple point of scale application for any supermarket which has a good off-product scale)

- the product data will be entered into the database.
- the taxes and commissions are entered.
- the product will be purchased and it will be sent to the bill counter.
- the bill calculating operator will check the product with the bar code machine it will check and match the product in the database and then it will show the information of the product.
- the bill will be paid by the customer and he will receive the products.
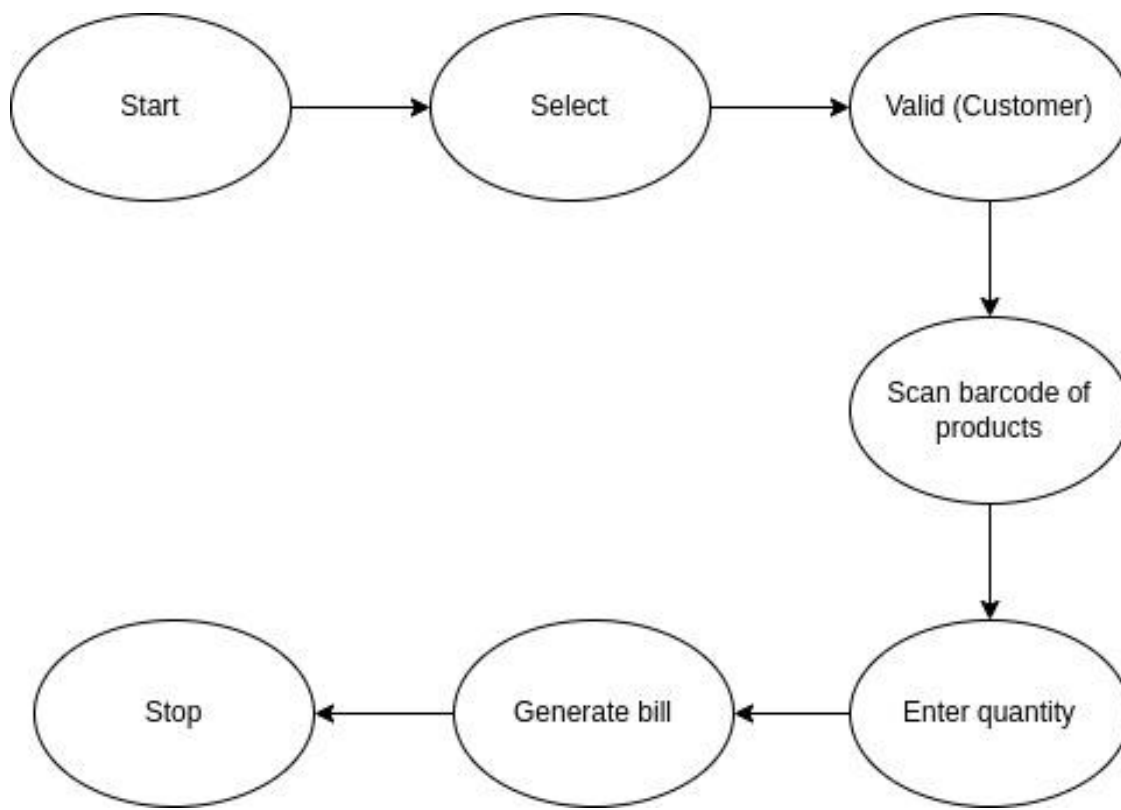
Tasks in the frequent pattern mining:

- Association
- **Cluster analysis:** frequent pattern-based clustering is well suited for high-dimensional data. by the extension of dimension the sub-space clustering occurs.
- **Data warehouse:** iceberg cube and cube gradient
- Broad applications

There are some to improve the efficiency of the tasks.

Closed Pattern:

A frequent pattern, it meets the minimum support criteria. All super patterns of a closed pattern are less frequent than the closed pattern.

Max Pattern:

It also meets the minimum support criteria(like a closed pattern). All super patterns of a max pattern are not frequent patterns. both patterns generate fewer numbers of patterns so therefore they increase the efficiency of the task.

Applications of Frequent Pattern Mining:

basket data analysis, cross-marketing, catalog design, sale campaign analysis, web log analysis, and DNA sequence analysis.

Issues of frequent pattern mining

- flexibility and reusability for creating frequent patterns
- most of the algorithms used for mining frequent item sets do not offer flexibility for reusing
- much research is needed to reduce the size of the derived patterns
- Frequent pattern mining has several applications in different areas, including:
- Market Basket Analysis: This is the process of analyzing customer purchasing patterns in order to identify items that are frequently bought together. This information can be used to optimize product placement, create targeted marketing campaigns, and make other business decisions.
- Recommender Systems: Frequent pattern mining can be used to identify patterns in user behavior and preferences in order to make personalized recommendations.
- Fraud Detection: Frequent pattern mining can be used to identify abnormal patterns of behavior that may indicate fraudulent activity.
- Network Intrusion Detection: Network administrators can use frequent pattern mining to detect patterns of network activity that may indicate a security threat.
- Medical Analysis: Frequent pattern mining can be used to identify patterns in medical data that may indicate a particular disease or condition.
- Text Mining: Frequent pattern mining can be used to identify patterns in text data, such as keywords or phrases that appear frequently together in a document.
- Web usage mining: Frequent pattern mining can be used to analyze patterns of user behavior on a website, such as which pages are visited most frequently or which links are clicked on most often.
- Gene Expression: Frequent pattern mining can be used to analyze patterns of gene expression in order to identify potential biomarkers for different diseases.

These are a few examples of the application of frequent pattern mining. The list is not exhaustive and the technique can be applied in many other areas, as well.

Conclusion:

It is impossible to give complete coverage of this topic with the limited space and our limited knowledge. Frequent pattern mining has achieved tremendous progress and claimed a good set of applications. However in-depth research is required that the field may have a long-lasting and deep impact on data mining applications.

Article Tags : Data Mining Data Warehouse

Recommended Articles

1. Generalized Sequential Pattern (GSP) Mining in Data Mining
2. Classification Using Frequent Patterns in Data Mining
3. Text Mining in Data Mining
4. Mining Collective Outliers Data Mining
5. Pattern Evaluation Methods in Data Mining
6. Different Types of Data in Data Mining
7. Data Cube or OLAP approach in Data Mining
8. Difference between Data Profiling and Data Mining
9. Data Mining - Time-Series, Symbolic and Biological Sequences Data
10. Data Mining For Financial Data Analysis
11. Clustering High-Dimensional Data in Data Mining
12. Complex Data Types in Data Mining
13. Data Replication Methods in Data Mining
14. Outlier Detection in High-Dimensional Data in Data Mining
15. General Strategies for Data Cube computation in Data Mining
16. Data Mining Tutorial
17. Techniques To Evaluate Accuracy of Classifier in Data Mining
18. Associative Classification in Data Mining
19. Statistical Methods in Data Mining
20. Data Mining Techniques
21. Graph Clustering Methods in Data Mining
22. Discovery Driven Cube Space Exploration in Data Mining
23. Introduction to Data Mining
24. Advantages and Disadvantages of ANN in Data Mining
25. Determining the Number of Clusters in Data Mining

Read Full Article

A-143, 9th Floor, Sovereign Corporate Tower,
Sector- 136, Noida, Uttar Pradesh (201305)
feedback@geeksforgeeks.org

Company

About Us

Careers

In Media

Contact Us

Privacy Policy

Copyright Policy

Advertise with us

Learn

DSA

Algorithms

Data Structures

SDE Cheat Sheet

Machine Learning

CS Subjects

Video Tutorials

Courses

NEWS

Top News

Technology

Work & Career

Business

Finance

Lifestyle

Knowledge

Languages

Python

Java

CPP

Golang

C#

SQL

Kotlin

Web Development

Web Tutorials

Django Tutorial

HTML

JavaScript

Bootstrap

ReactJs

NodeJs

Contribute

Write an Article

Improve an Article

Pick Topics to Write

Write Interview Experience

Internships

Video Internship