WIE3007 2023/2024 S1

Tutorial 9

Question 1: Explain the significance of association rule mining in the context of market basket analysis. How can businesses benefit from the insights derived from this technique?

Association rule mining is a data mining technique that focuses on discovering interesting relationships, patterns, or associations among a set of items in large datasets. In the context of market basket analysis, association rule mining is particularly significant. Market basket analysis involves analysing customer purchase patterns to identify relationships between products that are frequently bought together. The primary goal is to uncover associations that can help businesses make informed decisions regarding product placement, promotions, and inventory management.

Benefits:

- Cross-Selling Opportunities
- Inventory Management
- Marketing Strategies

Question 2: Define the terms "Support," "Confidence," and "Lift." Given a set of transactions:

1. Bread, Milk
2. Bread
3. Milk, Diapers
4. Bread, Milk, Diapers
5. Diapers, Beer

Calculate the support for the itemset {Bread, Milk} and the confidence for the rule Bread → Milk.

- **Support**: The support for an itemset is the proportion of transactions in the dataset that contain that itemset. For the itemset {Bread, Milk}:

$$Support(\{Bread, Milk\}) = \frac{Number\ of\ transactions\ containing\ Bread\ and\ Milk}{Total\ number\ of\ transactions}$$

$$\frac{2}{5} = 0.4$$

- **Confidence**: Confidence measures the likelihood that if item A is purchased, item B will also be purchased. For the rule Bread → Milk:

$$Confidence\ (Bread\ \rightarrow Milk) = \frac{Support\ (Bread, Milk)}{Support\ (Bread)}$$

$$\frac{2}{3} = 0.67$$

- **Lift**: Lift measures how much more likely item B is purchased when item A is purchased, compared to when item B is purchased without considering item A.

$$Lift\ (Bread\ \rightarrow Milk) = \frac{Confidence\ (Bread\ \rightarrow Milk)}{Support\ (Milk)}$$

$$\frac{2}{3} \div \frac{3}{5} = 1.11$$

Question 3: Outline the general steps involved in association rule mining. Highlight how the Apriori algorithm fits into this process and what makes it a commonly used algorithm for this purpose.

General Steps:

1. *Data collection*
2. *Data pre-processing*
3. *Support calculation*
4. *Rule generation*
5. *Rule evaluation*

The Apriori algorithm is a widely used algorithm in association rule mining. It works by iteratively discovering frequent itemsets and generating association rules. It uses a bottom-up approach and employs the "apriori" property, which states that any subset of a frequent itemset must also be frequent. This property allows the algorithm to prune the search space, making it more efficient.

The Apriori algorithm is commonly used due to its simplicity, ease of implementation, and ability to handle large datasets. It efficiently prunes the search space, reducing computational complexity.

Question 4: Trace the evolution of association rule mining from its inception to its current status. Discuss the main challenges that early algorithms faced and how they were addressed in subsequent versions.

1. ***Computational Complexity***
    - *Generating association rules from large datasets was computationally expensive.*
2. ***Memory Usage***
    - *Storing and processing large item sets required substantial memory.*
3. ***Multiple Scans of Data***

- *Early algorithms often required multiple passes over the dataset.*

Ways to address these challenges:

1. ***Improved Efficiency***
    - *New algorithms, like FP-Growth, addressed computational complexity by using a more compact data structure (FP-tree) and requiring fewer passes over the data.*
2. ***Memory Efficiency***
    - *Some algorithms reduced memory usage by employing tree-based structures or compact data representations.*
3. ***Parallelization***
    - *Distributed and parallel algorithms emerged to handle large-scale datasets more efficiently.*
4. ***Dynamic Itemset Counting***
    - *Some algorithms dynamically adjusted support thresholds to focus on interesting item sets, reducing the need for multiple scans.*

Question 5. Describe a scenario where association rule mining can significantly improve the performance of a deep learning model in human activity recognition (slide 30 – presentation).

In a scenario related to human activity recognition using deep learning, association rule mining can enhance performance by providing insights into sequential patterns of activities. For example, if association rule mining identifies frequent sequences of activities such as "walking -> jogging -> running," this information can be used to create more informed training datasets for deep learning models. By incorporating the knowledge of frequently occurring activity sequences, the deep learning model can better understand and predict the sequential nature of human activities, leading to improved accuracy and performance.

Queston 6. Discuss a challenge in using mobile and wearable sensors for human activity recognition (slide 30 – presentation) and propose a solution using a combination of association rule mining and deep learning.

Challenge: Variability and noise in sensor data, leading to inaccuracies in activity recognition.

Solutions:

1. Data Pre-processing
    - Use association rule mining to identify patterns and associations in the sensor data. This may include recognizing common sequences of sensor readings associated with specific activities.

2. Rule-Based Filtering

- Develop rules based on association rule mining results to filter out noise or irrelevant sensor data. This helps in reducing the impact of inconsistent readings on the overall recognition process.

3. Feature Engineering
    - Extract relevant features from the pre-processed data, incorporating both raw sensor readings and rule-based filtered data. This enhances the discriminative power of the features used by the deep learning model.

4. Hybrid Model
    - Train a deep learning model (e.g., recurrent neural network or LSTM) on the pre-processed and feature-engineered data. The model can leverage the learned representations from the data while being guided by the insights obtained from association rule mining.

5. Continuous Learning
    - Implement a system for continuous learning where the model adapts to new patterns and associations identified by association rule mining over time. This ensures that the model remains robust in dynamic environments.