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?

The significance of association rule mining in the context of market is because of how it allows retailers to discover patterns of products frequently bought together. This can lead to insights for product placements, promotions, or even new product bundles.

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.

Association rule: Support >= 0.6 and Confidence >= 0.6 (threshold)

1) List all items and calculate their support
   - Support(Bread) = 3/5 = 0.6
   - Support(Milk) = 3/5 = 0.6
   - Support(Diapers) = 3/5 0.6
   - Support(Beer) = 1/5
2) Generate Size-2 Candidate Itemsets:
   - Since all bread, milk and diapers meet the support threshold, the pair should be: {Bread, Milk. Diapers}
   - The pairs involving 'Beer' should not be generated.
3) Compute Support for Size-2 Itemsets {Bread, Milk) – ikut soalan
   - Support(Bread, Milk) = 2/5 = 0.4
   - Since the support is below the threshold of 0.6, we don't consider this itemset as 'frequent'
   - Confidence = Support(X,Y)/Support(X)
   - Confidence = (Support(Bread, Milk)/Support(Bread)
   - Confidence = 0.4/0.6 = 0.67
   - With the updated threshold, 0.67 => 0.6, Hence this rule is accepted
   - Now new rule: If Milk, then Bread
   - Confidence = Support(Bread, Milk)/Suport(Milk) = 0.4/0.6 = 0.67
   - We accept this rule

   Thus, the conclusion is if someone buys Bread, there's 67% chance they'll also buy Milk. If someone buys milk, there's a 67% chance they'll also buy Bread.

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.

- Association rule mining is a data mining technique used to discover interesting relationships, patterns, and associations among a set of items in large datasets.
- The process involves identifying rules that highlight the relationship between different variables or items in a dataset.

General steps involved in association rule mnining with a focus on how the Apriori algorithm fits the process:

1. **Define the Problem:**
    - Clearly define the problem and the dataset for association rule mining.
    - Specify the minimum support and minimum confidence thresholds to control the rule generation process.
2. **Data Preprocessing:**
    - Clean and preprocess the data to handle missing values, outliers, and irrelevant information.
    - Transform the data into a suitable format for association rule mining.
3. **Transaction Identification:**
    - Represent the dataset as transactions, where each transaction is a set of items associated with a unique identifier.
4. **Generate Candidate Itemsets:**
    - Create the initial set of candidate itemsets by identifying frequent items in the dataset. An itemset is considered frequent if its support (frequency) is above a predefined minimum support threshold.
5. **Generate Association Rules:**
    - Create association rules from the frequent itemsets. An association rule typically has an antecedent (premise) and a consequent (outcome).
    - Rules are generated based on the minimum confidence threshold, which indicates the likelihood of the consequent occurring given the antecedent.
6. **Evaluate Rules:**
    - Evaluate the generated rules using metrics such as support, confidence, and lift to assess their significance and usefulness.

Now, let's focus on the Apriori algorithm:

- **Apriori Algorithm:**
    - The Apriori algorithm is a classic and widely used algorithm for association rule mining.

  - It operates in an iterative manner, generating frequent itemsets of increasing size.
  - The key idea is to use the "apriori property," which states that if an itemset is frequent, then all of its subsets must also be frequent.
  - The algorithm prunes candidate itemsets that do not meet the minimum support threshold, thus reducing the search space.
  - Apriori uses a breadth-first search strategy to discover frequent itemsets efficiently.

Key Characteristics of Apriori:

- **Candidate Generation:** It generates candidate itemsets of higher size from frequent itemsets of lower size.
- **Support Counting:** It scans the dataset multiple times to count the support of candidate itemsets.
- **Pruning:** It eliminates candidate itemsets that do not meet the minimum support threshold.

Advantages of Apriori Algorithm:

- It is straightforward and easy to understand.
- The apriori property reduces the search space, making it computationally efficient.
- It can handle large datasets with a large number of transactions and items.

The Apriori algorithm's efficiency and simplicity make it a commonly used and well-established technique for association rule mining.

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.

Queston 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).

- 1. Early Days of Data Mining (Late 1980s to Early 1990s)
  - During the late 1980s and early 1990s, the field of data mining began to emerge as databases grew in size and complexity. Researchers sought methods to extract meaningful patterns and insights from vast amounts of data.
- 2. Introduction of the Apriori Algorithm (1993)
  - The seminal work on association rule mining is the paper titled "Mining Association Rules between Sets of Items in Large Databases" by Rakesh

    Agrawal, Tomasz Imieliński, and Arun Swami in 1993. This was presented at the ACM SIGMOD conference.

    - In this work, the authors introduced the concept of mining association rules from transaction data and proposed the Apriori algorithm. The algorithm was designed to efficiently discover the most frequent combinations of items in transactional databases.
    - The Apriori algorithm's introduction spurred interest in the field, leading to numerous research initiatives aimed at optimizing and extending the algorithm.

- 3. Extensions and Improvements (Mid to Late 1990s)
    - Following the introduction of the Apriori algorithm, several variations and improvements were proposed to make the algorithm faster and more memory-efficient.
    - Algorithms like FP-growth (Frequent Pattern growth) were developed, which eliminated the need for candidate generation, one of the bottlenecks in the Apriori approach.
    - Other research focused on mining association rules in more complex data types, such as time-series data, spatial data, and multi-relational databases.
- 4. Applications Beyond Market Basket Analysis (2000s)
    - As the concept matured, the application of association rule mining extended beyond retail and market basket analysis. It started to be applied in various domains like biology (for gene pattern discovery), web usage mining (to understand user 5.
- 5. Integration with Big Data Technologies (2010s and Beyond)
    - With the advent of big data technologies and platforms like Hadoop and Spark, association rule mining algorithms were scaled to handle even larger datasets.
    - Distributed versions of association rule mining algorithms were developed to leverage the parallel processing capabilities of these platforms.
- Current Trends
    - Today, association rule mining continues to be a valuable tool in the data scientist's toolkit. With the increasing availability of data, there's a renewed interest in extracting meaningful patterns using association rules.
    - Current research in the area focuses on real-time association rule mining, mining from streaming data, and integrating other machine learning techniques with association rules for more insightful analyses.
    - The history of association rule mining showcases its evolution from a novel concept for understanding transaction data to a versatile tool used across various domains. It also mirrors the broader trajectory of data mining and

analytics, from handling relatively small datasets to today's big data challenges.

- behavior), and healthcare (to find patterns in patient data).

Imagine a scenario where a deep learning model is employed for human activity recognition using sensor data from wearable devices, such as accelerometers and gyroscopes. The goal is to classify and understand various human activities, such as walking, running, sitting, and standing. While deep learning models are powerful in capturing complex patterns in data, association rule mining can complement these models by providing insights into temporal patterns and co-occurrences of activities. This combined approach can significantly improve the performance of the overall system.

Here's how association rule mining can enhance the performance of the deep learning model in this scenario:

1. **Data Preprocessing:**
    - Collect sensor data from wearable devices, ensuring a rich dataset that captures various activities and their transitions over time.
2. **Deep Learning Model:**
    - Train a deep learning model (e.g., a recurrent neural network or a long short-term memory network) on the sensor data to recognize human activities.
3. **Association Rule Mining:**
    - Apply association rule mining to discover temporal patterns and associations between different activities. For example, the rules might reveal that certain activities tend to follow or precede others in specific sequences.
4. **Rule Integration with Deep Learning Model:**
    - Integrate the discovered association rules into the deep learning model to enhance its predictions. The rules can serve as additional context and constraints, guiding the model to make more accurate predictions based on the observed patterns of human behavior.
5. **Improving Model Robustness:**
    - Use association rules to improve the robustness of the deep learning model. For instance, if the model predicts an activity that contradicts the discovered association rules (e.g., standing immediately after a prediction of running), it can be flagged as a less likely or erroneous prediction.
6. **Handling Noisy Data:**
    - Association rule mining can help identify and handle noisy or ambiguous instances in the sensor data. Rules can capture common variations or exceptions in activity patterns, allowing the deep learning model to adapt to real-world variability.

7. **Enhancing Explainability:**
    - Association rules provide a transparent way to interpret and explain the model's decisions. This can be crucial in applications where understanding why a certain prediction was made is important, such as in healthcare or assisted living scenarios.

By combining the strengths of deep learning for feature learning and association rule mining for pattern discovery, this integrated approach can lead to a more robust and context-aware human activity recognition system. It leverages the complementary nature of the two techniques, addressing challenges related to temporal dependencies, variability in activity sequences, and noisy sensor data.

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: Inconsistencies in Sensor Data Due to Device Placement and Orientation

- **Issue:** Mobile and wearable sensors can suffer from inconsistencies in data due to variations in device placement and orientation on the human body. For instance, a smartphone in a pocket may provide different accelerometer readings compared to the same device attached to the wrist or placed in a bag. These variations make it challenging for human activity recognition models to generalize across different users and device placements.
- **Solution using a Combination of Association Rule Mining and Deep Learning:**
    1. **Association Rule Mining for Device-Specific Patterns:**
        - Use association rule mining to identify device-specific patterns and relationships between sensor data and human activities based on the placement and orientation of the device.
        - Discover rules that capture how accelerometer and gyroscope readings correlate with activities for specific device placements (e.g., pocket, wrist, bag).
    2. **Adaptive Feature Learning with Deep Learning:**
        - Train a deep learning model (e.g., a recurrent neural network) to learn adaptive features from the sensor data. This model should be capable of capturing device-specific variations by considering the identified rules from association rule mining.
    3. **Context-Aware Fusion of Rules and Deep Features:**
        - Develop a mechanism to fuse the device-specific rules obtained from association rule mining with the deep features learned by the neural network.

- The fusion process should dynamically adapt to the device placement and orientation, allowing the model to leverage both the discovered patterns and the adaptive features for accurate activity recognition.

4. **Real-Time Calibration and Feedback:**
    - Implement a real-time calibration mechanism that continuously adjusts the model based on the incoming sensor data.
    - Provide feedback to users about the quality of data collection, suggesting optimal device placement for improved accuracy.
5. **User-Specific Fine-Tuning:**
    - Allow for user-specific fine-tuning of the model by incorporating individual preferences and adjustments.
    - Users can provide feedback on the accuracy of activity recognition, helping the system refine its understanding of device-specific patterns for each individual.

Benefits of the Proposed Solution:

- The combination of association rule mining and deep learning addresses the challenge of inconsistencies in sensor data by providing a more context-aware and adaptive approach to human activity recognition.
- The model becomes capable of learning and adapting to device-specific variations, making it more robust across different scenarios, users, and wearable placements.
- Real-time calibration and user-specific fine-tuning enhance the overall user experience and improve the accuracy of activity recognition in diverse settings.

By leveraging both the strengths of association rule mining for pattern discovery and deep learning for adaptive feature learning, this integrated solution addresses the challenges associated with variations in sensor data due to device placement and orientation in mobile and wearable devices.

Note: The presentation is derived from the following video:
https://www.youtube.com/watch?v=OGIiGg8qKFg&ab_channel=UniversitiMalaya