High Utility Frequent Itemset (HUFI) Mining: Key Concepts and Strategies

High Utility Frequent Itemset (HUFI) mining focuses on identifying patterns in transaction datasets that are both frequent and have high utility, such as profit, importance, or relevance. This approach goes beyond traditional frequent itemset mining by incorporating utility measures into the analysis. Below are the key factors and strategies for mining high utility and periodic itemsets.

Key Factors in High Utility Periodical Itemset Mining

1. **Unit Utility**
   - Represents the utility (e.g., price, profit, or significance) of a single item in a transaction.
   - This value can be positive (e.g., revenue) or negative (e.g., loss or cost).
2. **Frequency**
   - Indicates how often an itemset appears in transactions.

Criteria for High Utility and Frequent Itemsets

Table 1
Database.

Tid	Item	Quantity
T_1	a b c d	5 2 1 2
T_2	a c d g	1 1 1 3
T_3	a c f	1 1 1
T_4	a f g	1 4 2
T_5	a g	1 2
T_6	b c d e	3 2 3 1
T_7	c e	6 4
T_8	e f	1 3

Table 2
Unit utility.

Item	a	b	c	d	e	f	g
Utility	3	6	−3	12	−5	−2	−1

High Utility

- The total utility of an itemset across transactions must meet or exceed a predefined threshold (*minutil*).
- Example: An itemset like *rice + olive oil* may have high utility if it generates significant profit.

1. **Frequency (Periodic Constraints)**
   - **Max Periodic**: The itemset must appear at least once within a specified maximum number of transactions (*maxper*).
     - Ensures the pattern occurs regularly and is not overly rare.
   - **Min Periodic**: The itemset must not appear more frequently than a given minimum number of transactions (*minper*).
     - Prevents overly common items from forming irrelevant or misleading patterns.

Example: Rice as a High Utility Periodic Item

Rice is a staple food in many Asian households, appearing in nearly every transaction, such as:

- *Rice + olive oil*
- *Rice + apples + bananas*
- *Rice + fish + tank + aquarium décor*
- *Rice + hair dye + comb*

While rice has high utility, its ubiquity makes it unsuitable for certain recommendations (e.g., pairing *rice* with *hair dye* or *comb*).

- **Min Periodic** avoids recommending such trivial combinations.
- **Max Periodic** ensures rice-related patterns are still identified when relevant.

The Role of Average Periodic

Average periodic ensures itemsets are neither rare anomalies nor overly frequent, maintaining relevance in recommendations.

- **Reflects Business Cycles**: Aligns with real-world cycles, such as weekly restocking or seasonal trends.
- **Maintains Relevance**: Filters out outliers, ensuring itemsets represent consistent patterns rather than anomalies.

Example: If a merchant restocks every seven days, valid itemsets should align with this schedule to provide actionable insights.

Itemset scanning strategy

Efficient itemset scanning is essential to handle large datasets and identify meaningful patterns. Advanced algorithms are employed to:

- Prune irrelevant or low-utility patterns early.
- Optimize computations by leveraging periodic constraints and utility thresholds.