

FP-Growth (Frequent Pattern Growth) algorithm

The FP-Growth (Frequent Pattern Growth) algorithm is a popular method for frequent itemset mining and association rule learning over transaction databases. It is more efficient than the Apriori algorithm, especially when dealing with large datasets. The FP-Growth algorithm avoids candidate generation and instead uses a data structure called an FP-tree (Frequent Pattern Tree) to represent the database.

Steps of FP-Growth Algorithm

1. Construct the FP-Tree:

- **Scan the database** to find the support count of each item. Discard infrequent items (those below a minimum support threshold).
- **Sort items in each transaction** in descending order of their frequency.
- **Build the FP-tree** by inserting transactions. Each transaction shares a common prefix, represented by the path in the tree.

2. Mine the FP-Tree:

- **Start from the frequent items** (those with minimum support) as the suffix.
- **Construct conditional FP-trees** for each frequent item.
- **Recursively mine each conditional FP-tree.**

Example

Let's walk through an example to illustrate the FP-Growth algorithm.

Sample Transactions

Consider the following transactions:

TID	Items
1	A, B, D, E
2	B, C, E
3	A, B, C, E
4	B, E
5	A, B, C, E

Let's set the minimum support threshold to 3.

Step 1: Construct the FP-Tree

1. Count the frequency of items:

- A: 3, B: 5, C: 3, D: 1, E: 4

Discard infrequent items (D in this case, as its support is less than 3).

- **Sort items in transactions by frequency:**

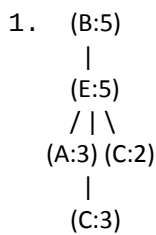
T1: B, E, A

T2: B, E, C
 T3: B, E, A, C
 T4: B, E
 T5: B, E, A, C

• **Build the FP-tree:**

- Insert T1: B -> E -> A
- Insert T2: B -> E -> C
- Insert T3: B -> E -> A -> C
- Insert T4: B -> E
- Insert T5: B -> E -> A -> C

The FP-tree will look like this:



Step 2: Mine the FP-Tree

1. **Start with the least frequent item:**

- Start with C:
 - Conditional pattern base: {{B,E,A}:1, {B,E}:1, {B,E,A}:1}
 - Conditional FP-tree: B:3, E:3, A:2
 - Frequent itemsets: {C}, {C,B}, {C,E}, {C,A}, {C,B,E}, {C,B,A}, {C,E,A}, {C,B,E,A}

2. **Next frequent item:**

- Move to A:
 - Conditional pattern base: {{B,E}:3}
 - Conditional FP-tree: B:3, E:3
 - Frequent itemsets: {A}, {A,B}, {A,E}, {A,B,E}

3. **Next frequent item:**

- Move to E:
 - Conditional pattern base: {{B}:5}
 - Conditional FP-tree: B:5
 - Frequent itemsets: {E}, {E,B}

4. **Most frequent item:**

- Finally, B:
 - Frequent itemsets: {B}

Summary of Frequent Itemsets

- {C}, {C,B}, {C,E}, {C,A}, {C,B,E}, {C,B,A}, {C,E,A}, {C,B,E,A}
- {A}, {A,B}, {A,E}, {A,B,E}

- $\{E\}, \{E,B\}$
- $\{B\}$