

Date ; 03/01

Topic: advanced pattern mining

Mining Diverse Patterns:

- Multiple-level association
- Rare pattern and negative pattern.

Q: What is a closed pattern?

A: "Closed frequent itemsets are defined as frequent patterns, no superset of which have the same frequency as that itemset. Closed patterns can be viewed as the maximal patterns from each group of equi-support patterns (i.e., patterns with the same support). All maximal patterns are, therefore, closed." (Frequent Pattern Mining p.5)

E.g.: {a, b, c} could be a closed pattern, which includes examples of {a,b}, {a,c}, etc.

For closed sequential patterns, an example could be {a,b,c,a,c,(a,b)} which contains {a,b} {b,a}

Q2: What is the difference between closed pattern and maximal pattern? Are they the same?

A2: currently, we think they are the same. (no super level patterns)

Q3: Here, sequential pattern doesn't require the consistent relationship, instead, only the relative sequence is concerned, which means, any insertions are welcomed. For example, in sequential pattern mining of words like "advertising", patterns of "-ing", "ad-", and also "-ding", "-vt-" might be caught.

A3: The consideration of the requirement of consistent relationship might be similar to the consideration between "precision" and "recall". The consistent relationship is a more restraint requirement; hence, it might identify more closely related pattern. On the other hand, without this restraint, we could find patterns which are loosely related. In summary, the consideration of whether the consistent relationship should be required or not depends on what kind of result you want to achieve. If you want to ensure that the found patterns are closely related, you might want to restrain the consistent relationship. In contrast, if you want to find as many related patterns as possible, you might want to loosen this restraint.

E.g1. In economics and some researches of recommender system, people emphasize more on the surprising effect of the recommendation, which refers to the concept of recall, rather than the precision. Customers are okay with bad recommendations, meanwhile, conservative recommendations are bad for the vitality of the system.

E.g2.: "斯德哥尔摩症候群"中, "斯德哥尔摩", "症候", "群"彼此不可被打散。
我们想要"斯德哥尔摩", 不想要"斯症候群"。

To generalize, sometimes, when inserted, the original local idea would be broken. We argue that there should be an unbreakable minimum unit in finding the sequential pattern.