Frequent Pattern Mining

General

* The field of data mining has four main “super-problems” corresponding to clustering,

classification, outlier analysis, and frequent pattern mining.

* Anomaly detection = outlier analysis
* Frequent pattern mining != feature engineering (should be \in)

Frequent pattern mining: The problem of frequent pattern mining is that of finding relationships among the items in a database. The problem can be stated as follows.

*Given a database* D *with transactions* T1 . . . TN *, determine all patterns* P *that are present in at least a fraction s of the transactions*.

Sequential pattern mining: <https://en.wikipedia.org/wiki/Sequential_pattern_mining>

Feature engineering: <https://en.wikipedia.org/wiki/Feature_engineering>

[We don’t fully agree the definition from wikipedia, in our perspective, FE is the process of feature creation, recreation, optimization, evaluation, etc. for the purposes of better describing the data (entity?) In DB, the world is made of entities and relationship. FE is for polishing the attributes of entities.]

[Our understanding of FPM. We agree with the book definition. They are for detecting relationship (or the patterns) among (of) the transactions (not items), utilizing the items in all transaction space.]

Thus, in our understanding, the scope of FPM is a subset of that of FE.??

Or they are parallel, due to one for entity and one for relationship.??

(To be discussed.)

[Why do we need frequent pattern mining?]

* Decision making
* Distance (combination of items)
* Tree structure
* Conditional probability

Application of DAG (Directed Acyclic graph):

* Deep learning
* Bayes Network
* CRF / HMM
* Sequential relationship (chain)
* etc.

Non-sequential

Sequential

