Taxonomy of Data Interestingness

Both helps to understand the current trend and methods followed for interesting facts generation.

- Association rule mining
  - What — Barket basket analysis, Its a non supervised ML method. uses supporta nd confidence as default measure
  - Literature
    - Apriori - frequent items mining — Huge no. of rules and may not be strong rules.
    - FP Growth - Frequent pattern tree. Divide and conqur method — Increase in FP- Tree size exponentially increases time.
    - FP-based ARM for Ontology / Identifying : Known, Unexpected, Novel rules — No semantics in the generated rules..
    - Fuzzy association rule mining / context sensetive fuzzy clustering — No structural relationships.
    - Distance based association rule mining / CLARNS / Concept hierarchy / CBPNARM — Positive and negative rules - KB
    - AMIE / AMIE improved — Horn rules among predicates. — Only labelled data is used.
    - Ontology & hypergrapgs are used to discover latent AR. — Ontology helps to prune and filter the rules.
    - Generalized and hierarchical AM — User knowledge is required & Only precise knowledge is used.
    - WARM - Weighted ARM — Uses inductive logic programming
- Ontology
  - What
    - Explicit, formal specification of a shared conceptualization
    - O=(TBOX +ABOX,G)
  - Why
    - Semantci structuring
    - Same semantcis across data points
  - Ontology-Based Methods
    - Ontology-guided generalization — Information theory based measure
    - Annotation dataset
    - Cross ontology relationships — Distinct characteristics
    - Ontology - ARM — Integrated rule information content (IRIC)
    - SWARM. Common behavioral patterns — Only concentrate on learning TBOX from incomplete ABOX.
    - ODIS- ontology driven information system. — Learning based semantic search algorithm.
    - Others — Multi ontology / semantic support / annotation inconsistency / Post mining and filtering
    - SEGS & g-SEGS — Inductive logic programming (ILP) .
      - Only use labelled data
      - Hierarcgy schema is not used.
      - ILP enrich the ontology.
    - Partial completeness assumption (PCA). — PCA confidence measure.
  - Semantic similarity based
    - Least common ancestor (LCA).
    - Dept of ontology hierarchy.
    - Common and non-common ancestors.
    - Ontology
      - Crisp ontology (CO)
      - Fuzzy Ontology (FO) — Fuzzy FCA)FFCA) — Uses common feature of concepts.
- Interestingness metrics
  - What
    - Prune or rank the disocvered rule.
    - statstical, ML, data mining techniques..
    - Three categories [29]. — 1. Rules thara re both unexpected and actionable. 2. Rules thata are both unexpected and not actionable. 3. Rules thata are actionable but expected.
- Metrics in Data mining for interestingness
  - Objective measures — Support. Confidence, Based on probability and based on form of rules.
  - Subjective measures — Unexpectedness, Actionability, Surprisingness & Novelty
    - Interest factor
    - Gini Index — Treats both postive and negative facts as same.
  - Semantic measures — Utility and Actionability
- Evaluation methods
  - Manual
    - Indicating the expectations — Comparing obtained via manually extracted.
    - Domain experts evaluations
  - Statistical
    - Precession,Recall, F-measure — Relation learning recall & Relation learning precision
    - Chi-Square
    - Learning Acurracy.....
  - Multi ontology — Instance-based / Match concepts of multiple ontology
  - Other measures
    - Ground truth association
    - True rule set
]