

Outlier Detection and Explanation

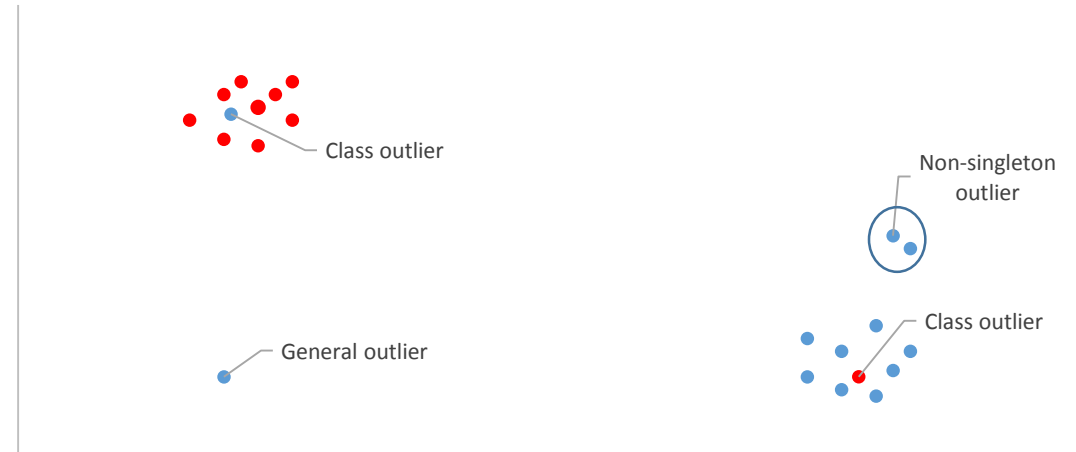
Bachelor's thesis defence

Václav Blahut, 395963

June 2015

Outliers

- Objects that do not comply with general behavior or model of the data
- Many methods of outlier detection
- Few methods for class outlier detection, outlier explanation or non-singleton outlier detection
- Many applications (fraud detection, system health issues...)



Examples of outliers

About this method

- Proposed by Angiulli and Fassetti in 2009
- Unsupervised, ILP-based
- Outlier is a set of examples
- Datasets in Aleph format
 - Strongly relational
 - Background knowledge – determinations, modes, facts
 - Positive and negative examples

The Algorithm

- Main idea – excluding outlier from dataset will simplify its model
- Compares models (Aleph theories) generated including and excluding each possible outlier
- Three types of abnormality
 - Irregularity
 - Anomaly
 - Outlier
- Implemented in Prolog
- Output – set of detected abnormalities with their explanations

Evaluation

- Four datasets
 - ZOO
 - Student Loan 100
 - Small and Medium Enterprises
 - House Votes 84 (Republicans vs. Democrats)
- Results compared to weka-peka and weka-CODB
- Explanations compared to RF-OEX

Thank you for your attention.