

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/21760> holds various files of this Leiden University dissertation.

Author: Duivesteijn, Wouter

Title: Exceptional model mining

Issue Date: 2013-09-17

Stellingen behorend bij het proefschrift “Exceptional Model Mining” van Wouter Duivesteijn

1. Paradoxically, under certain circumstances, an increase in the price of a commodity will lead to an increase in the demand for it. The Exceptional Model Mining instance with regression as model class can find a stronger occurrence of this effect than the occurrence found in the first field study whose results featured the effect (Chapter 7).
2. Using the Distribution of False Discoveries, we can compute a p-value for a found description. When the corresponding null hypothesis is refuted, we have reason to believe the the description is not a false discovery, but rather represents an underlying concept in the dataset (Chapter 8).
3. The Distribution of False Discoveries can be used to compare quality measures, by how well they distinguish true from false discoveries. For Subgroup Discovery, the χ^2 measure ranks best, and is significantly better than the measures Confidence, Laplace, Specificity, Purity, and Sensitivity (Chapter 8).
4. The part of Europe where the maximal temperature in March is below 7.97°C and the maximal temperature in September is below 17.65°C , consists of the union of two geographically coherent regions: Northern areas, and mountainous areas. In this union, the conditional dependence relations between mammals deviate from pan-European behavior (Chapter 6).
5. Employing descriptions, found with the Exceptional Model Mining instance of Chapter 6, as constructed binary features in multi-label Support Vector Machine classifiers, can improve predictive performance of the classifiers (Chapter 9).
6. The Beam Search algorithm for Top-q Exceptional Model Mining has a worst-case computational complexity of $\mathcal{O}(dwkn(c + M(N, m) + \log(wq)))$, if domain experts demand constraints that are reasonable (such as an upper bound on the description complexity), rather than conditioning them on, e.g., $P = NP$ (Chapter 3).
7. The primary goal of data mining is extracting meaningful information from large quantities of data. Therefore, “Because it tells us something about our dataset” should be enough motivation to pursue a line of data mining research.
8. The way in which academic success is evaluated in data mining causes underappreciation of negative results and of evaluating previously published results.
9. The essence of subjective data mining tasks, such as Subgroup Discovery and Exceptional Model Mining, cannot be properly captured by a formal Problem Statement, although such a statement is necessary for formal analysis.
10. Software development speed prohibits proper referencing.
11. “[...] there is an inherent weakness in trying to justify one’s concerns by saying that they are useful. Useful is a word of engineering.” (Yuri I. Manin)
(in “Good Proofs are Proofs that Make us Wiser”, interview by Martin Aigner and Vasco A. Schmidt, The Berlin Intelligencer, 1998, pp. 16–19, Springer-Verlag)