What MDL can bring to Pattern Mining
Tatiana Makhalova, Sergei Kuznetsov, Amedeo Napoli

To cite this version:
Tatiana Makhalova, Sergei Kuznetsov, Amedeo Napoli. What MDL can bring to Pattern Mining, ISWS 2018 - International Semantic Web Research Summer School, Jul 2018, Bertinoro, Italy. hal-01889792

HAL Id: hal-01889792
https://hal.archives-ouvertes.fr/hal-01889792
Submitted on 8 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Pattern Mining. What kind of patterns we should compute?

Types of patterns (defined for concept (A,B):

- MDL: is there a place for background knowledge?

Pattern Mining

Introduction

Patterns are subsets of attributes that describe an object.

Pattern Mining: Objective: find a small set of patterns that are well interpretable by experts.

Input data: binary table G * M, where G is a set of objects, M is a set of attributes, and I is a relation between them.

Interpretation of glm: object g ∈ G has attribute m ∈ M.

Examples of interestingness measures for concept (A,B)

- Typicality (representativeness)
  - It is measured by the usage of patterns, i.e. the frequency of the occurrence of patterns in the greedy covering, so the usage does not exceed the frequency.
  - It is not obvious which values are better. The high values of usage correspond to a subset of common patterns, while low values indicates that a subset contains less typical, but still interesting (w.r.t. interestingness measures) patterns.

- MDL-optimal (blue) vs top-n (green) closed itemsets

- MDL: is there a place for background knowledge?

- MDL in practice: greedy algorithm (Krimp)

- Used measures for ordering candidate sets.
  - The ordered list of candidates is used for greedy covering of data in Krimp

References