



Semantic meta mining

2021.03.08

CONTENTS



From meta-learning to meta-mining



Data Mining Ontologies



An Ontology for Data Mining Optimization



PART 01

From meta-learning to meta-mining

From meta-learning to meta-mining

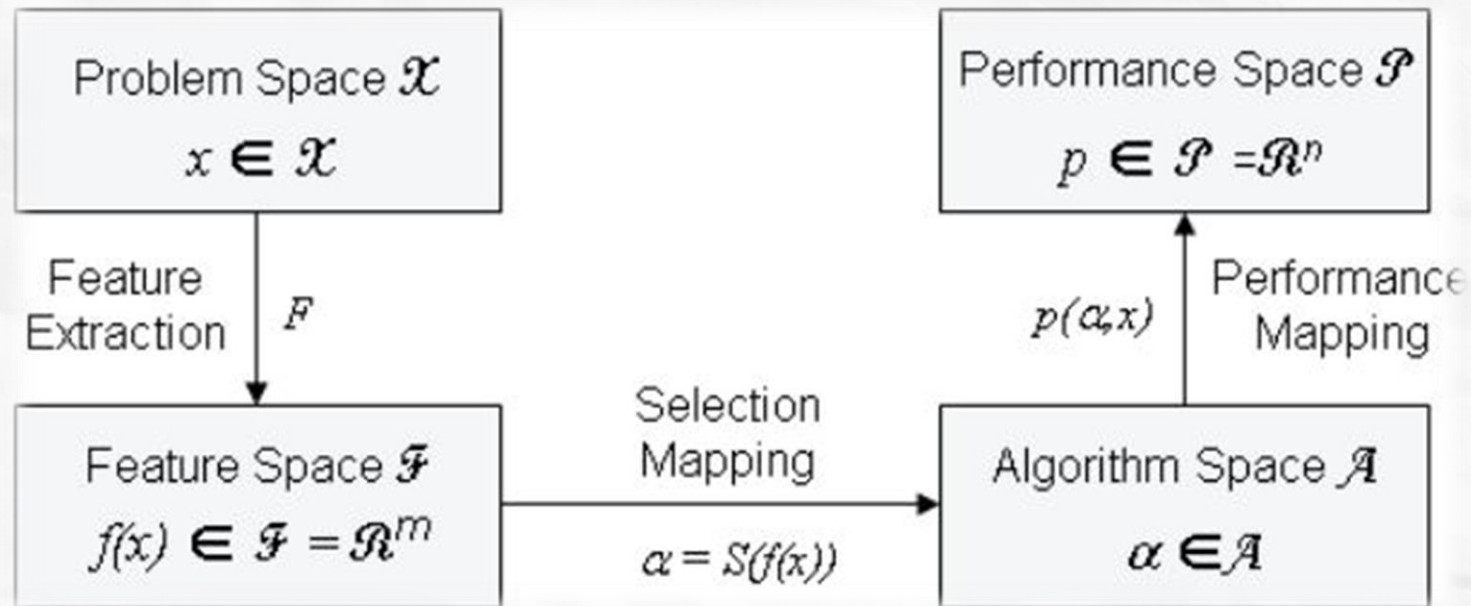
meta-learning

Learning to learn: use machine learning methods to improve learning

	Base-level learning	Meta-level learning
Application domain	any	machine learning
Ex. learning tasks	diagnose disease, predict stocks prices	select learning algorithm, parameters
Training data	domain-specific observations	meta-data from learning experiments

From meta-learning to meta-mining

Rice model



Abdulrahman, S. M., Adamu, A., Ibrahim, Y. A., & Muhammad, A. R. (2017). An Overview of the Algorithm Selection Problem. *International Journal of Computer (IJC)*, 26(1), 71-98.

From meta-learning to meta-mining

meta-learning

Learning to learn: use machine learning methods to improve learning

Implicitly bound to the [Rice model](#) for algorithm selection

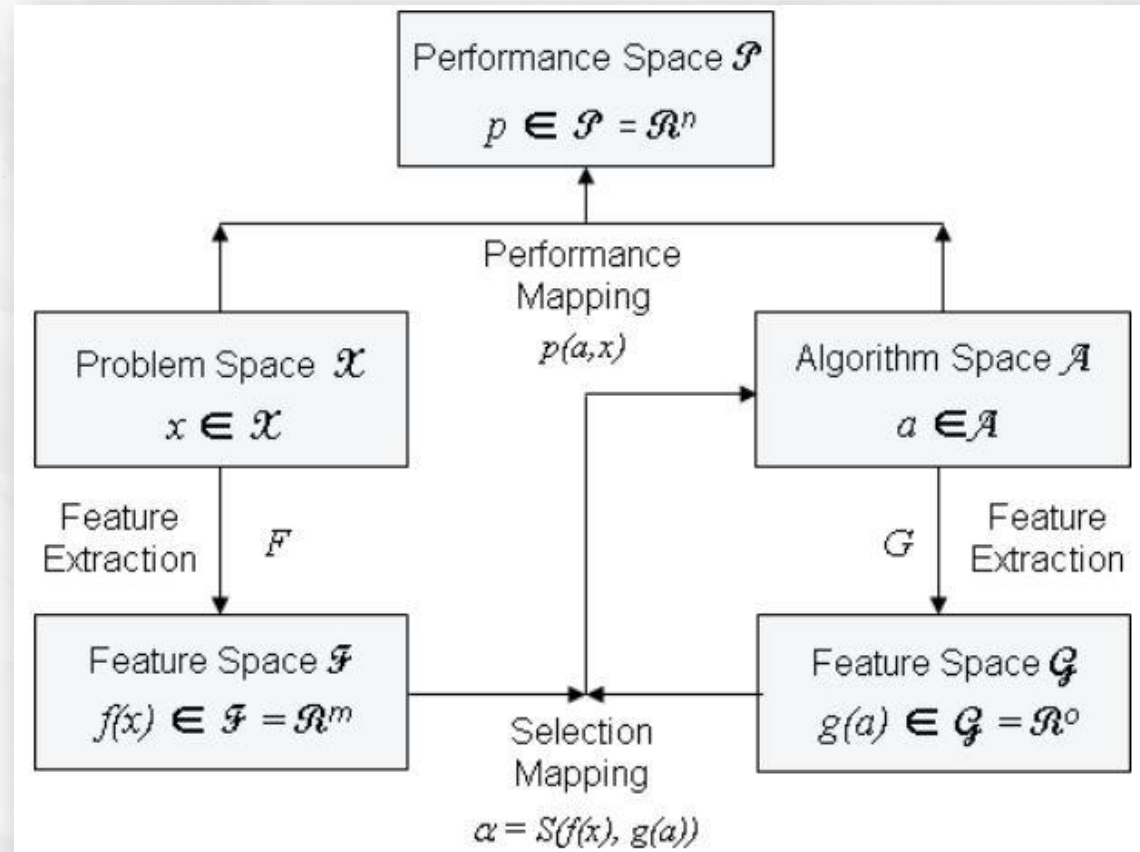
- Based solely on data characteristics.
- Algorithms treated as black boxes.

[Greedy](#): Restricted to the current (usually inductive) step of the DM process

[Purely data-driven](#): No integration of explicit DM knowledge into meta-learning

From meta-learning to meta-mining

Revised Rice model



Break the algorithmic black box and use both dataset and algorithm characteristics to meta-learn

From meta-learning to meta-mining

Semantic meta mining

- A semantic meta-learning problem would associate Algorithms Descriptors with Dataset Characteristics based on performance measures
- given some collection of datasets to which some algorithms were applied relying on DCs, the Algorithms Descriptors, and the algorithms performance measures
- A semantic meta-learning model can in principle make performance predictions for algorithms other than the ones on which it was created as long as the former are described in the DM ontologies.
- Very similar in nature to collaborative/content based filtering problems



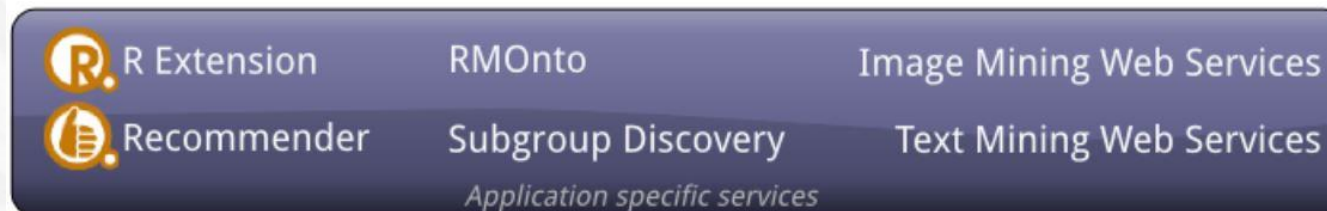
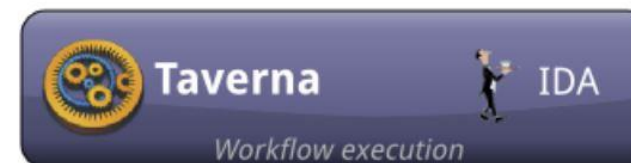
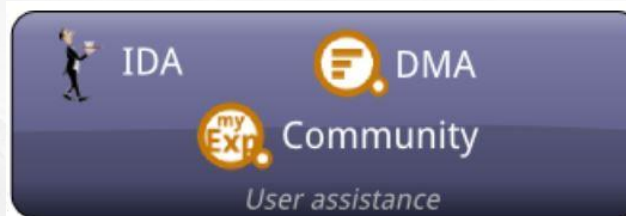
PART 02

Data Mining Ontologies

Data Mining Ontologies

E-LICO

An *e-Laboratory* for
Interdisciplinary
Collaborative Research in
Data Mining and Data-
Intensive Science



VideoLectures.Net data set
for video recommendation



KUPKB
Kidney and Urinary Pathways
Knowledge Base



Data Mining Ontologies

Ontologies in E-LICO

- The Data Mining Work Flow (DMWF) Ontology.
 - <http://www.e-lico.eu/ontologies/dmo/e-Lico-eProPlan-DMWF-HTN.owl>
- The Data Mining Optimization (DMOP) Ontology.
 - <http://www.e-lico.eu/ontologies/dmo/DMOP/DMKB.owl>.

Data Mining Ontologies

OntoDM series

- Ontology of Datatypes (OntoDT)
- Ontology of Core Data Mining Entities (OntoDM-core)
- Ontology of Data Mining Investigations (OntoDM-KDD)

Data Mining Ontologies

Others

- **The Knowledge Discovery (KD) Ontology**
- **The OntoDTA ontology**
- **The KDDONTO Ontology**
- **...**



PART 03

An Ontology for Data Mining Optimization

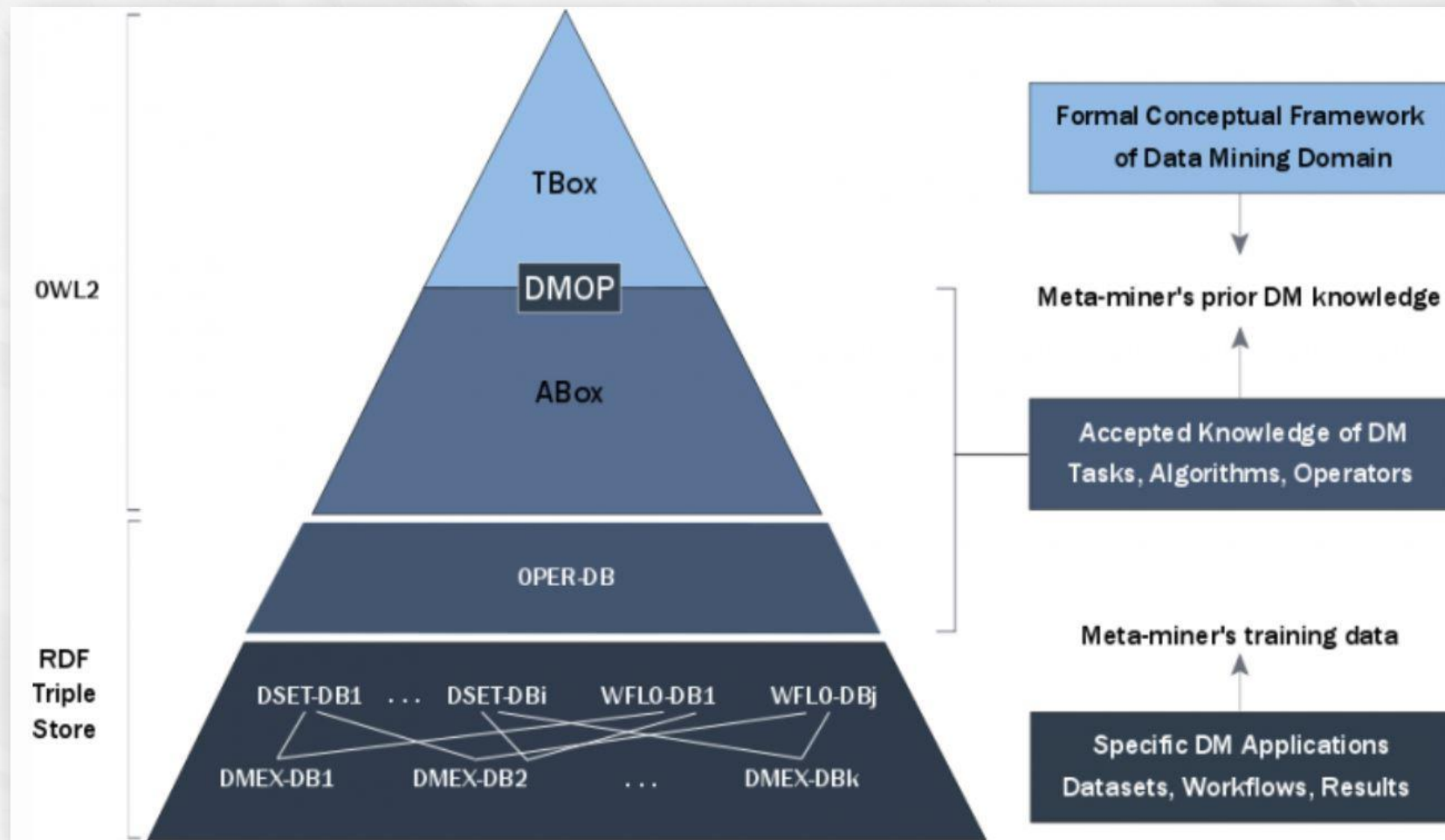
An Ontology for Data Mining Optimization

DMOP: Data Mining OPTimization ontology

- **Purpose:** structure the space of DM tasks, data, models, algorithms, operators and workflows
 - higher-order feature space in which meta-learning can take place
- **Approach:** model algorithms in terms of their underlying assumptions and other components of bias
 - allows for generalization over algorithms and hence over workflows
 - supports **semantic meta-mining**

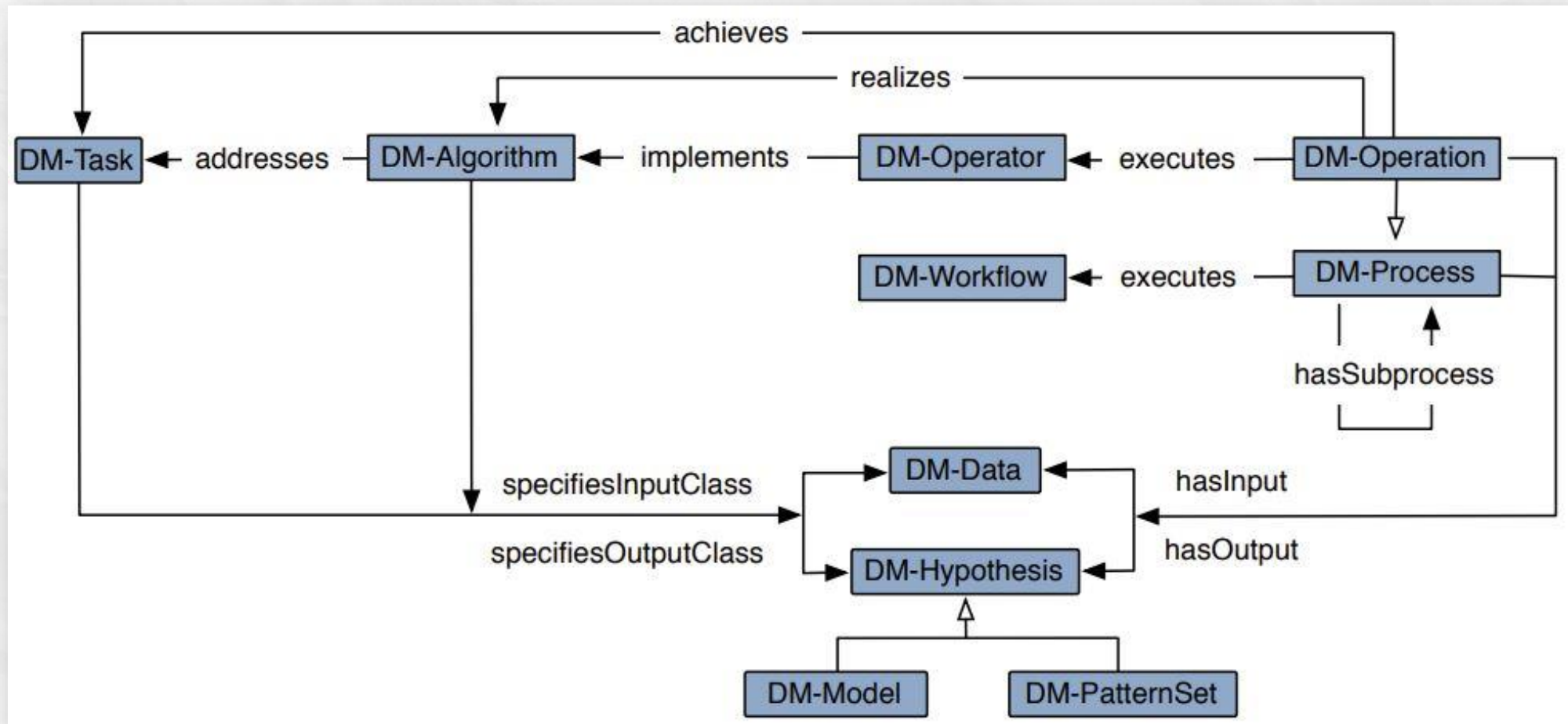
An Ontology for Data Mining Optimization

DMOP: Architecture



An Ontology for Data Mining Optimization

DMOP: content



An Ontology for Data Mining Optimization

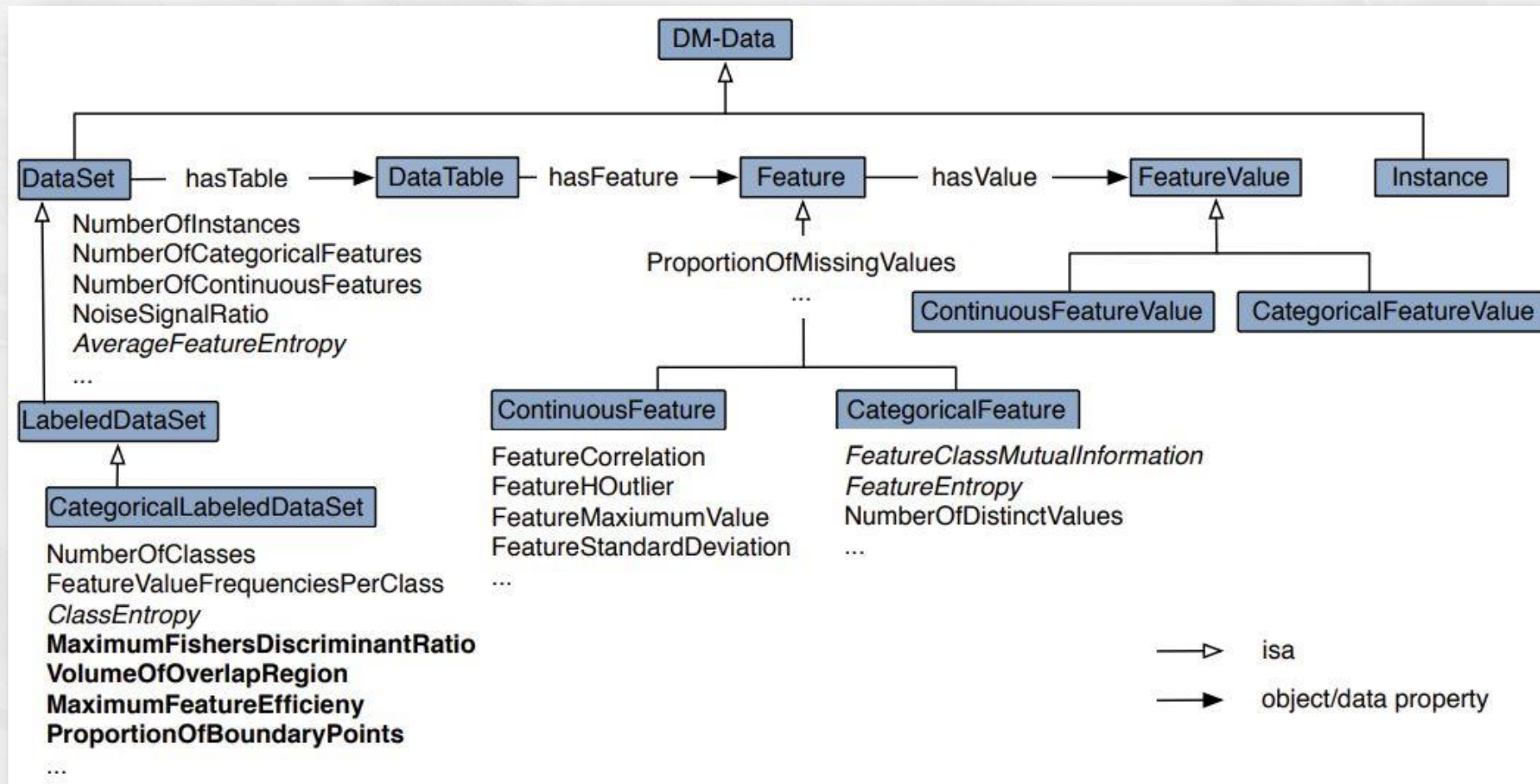
DMOP: DM-tasks

DM tasks:

- DataProcessingTask
- InductionTask
- HypothesisProcessingTask

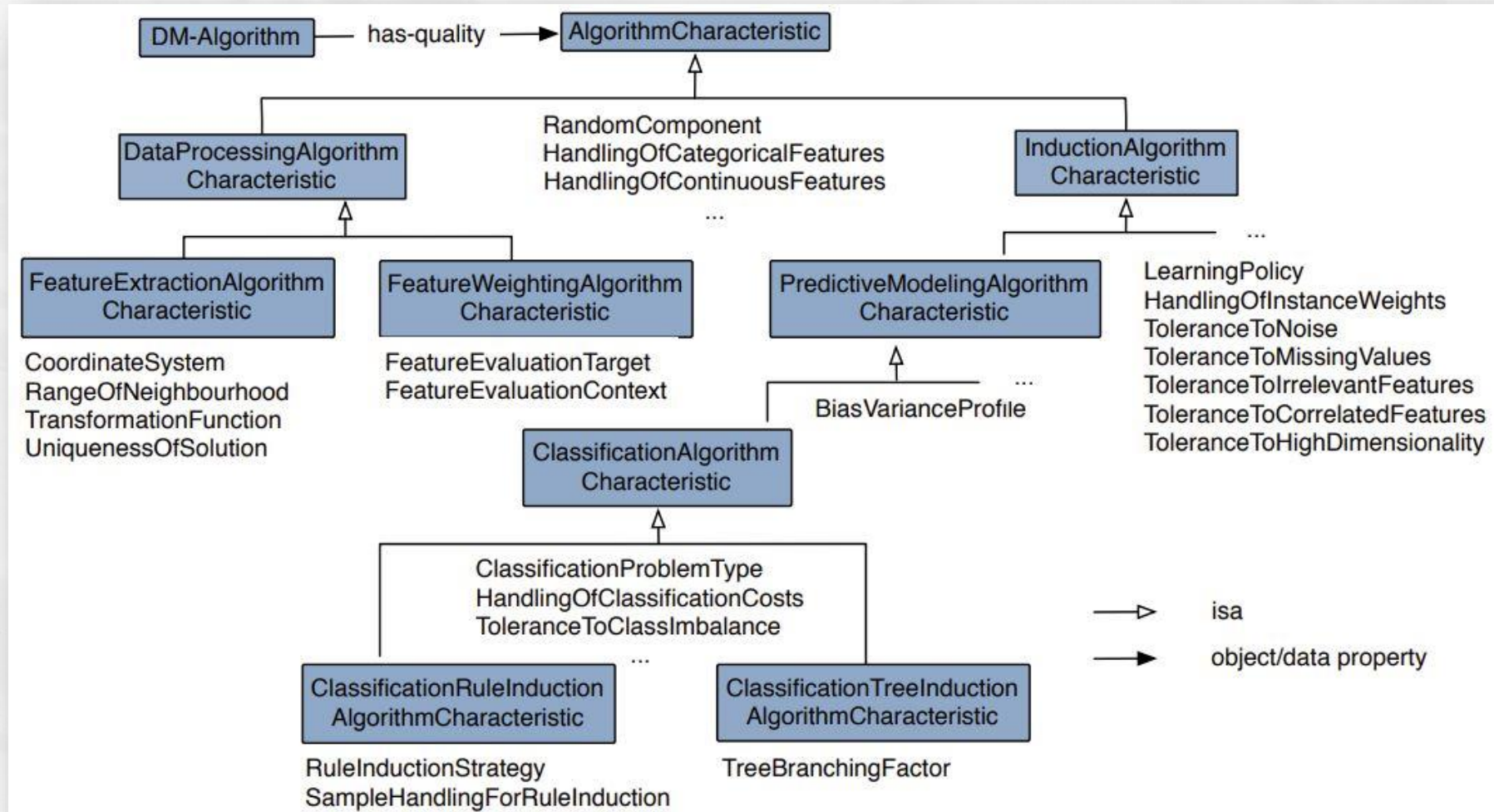
An Ontology for Data Mining Optimization

DMOP: DM-Data



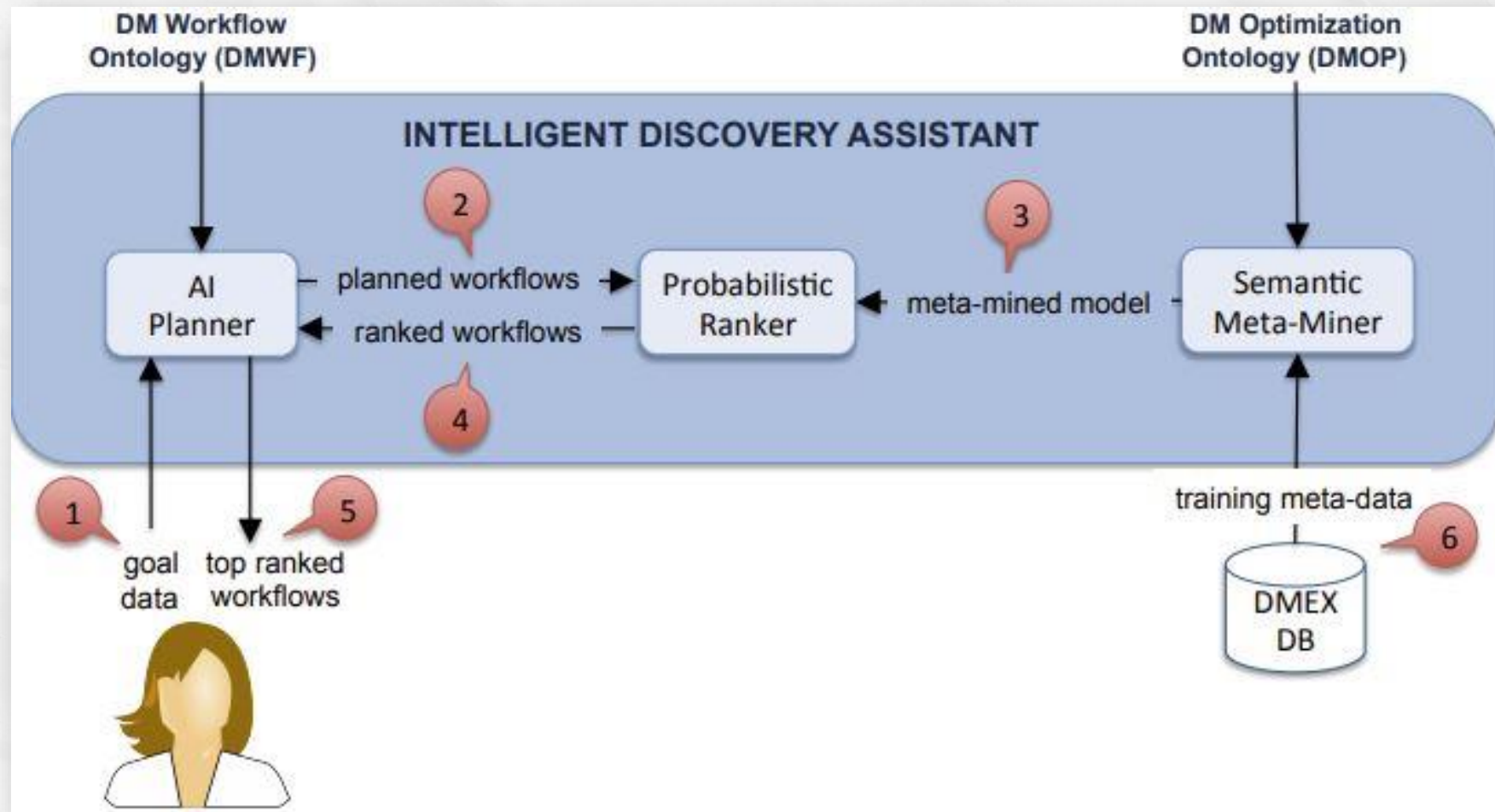
An Ontology for Data Mining Optimization

DMOP: DM-Algorithm



An Ontology for Data Mining Optimization

Usage of DMOP in semantic meta-mining





THANK YOU

2021.03.08

mantx626@gmail.com