

Enhancing Inductive Programming by Function Ranking

A Machine Learning Application for Data Wrangling Automation

Lidia Contreras-Ochando

liconoc@upv.es

@liconoc

Joint work with: Cèsar Ferri, José Hernández-Orallo,
Susumu Katayama, Fernando Martínez-Plumed and
María José Ramírez-Quintana



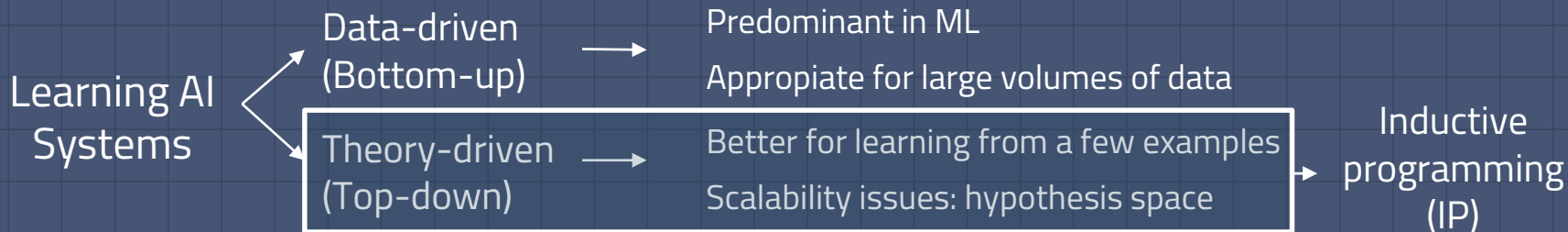
UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



宮崎大学
University of Miyazaki



Introduction



Example

Id	Input	Output
1	25-03-74	25
2	03/29/86	29
3	1998/12/25	25
4

IP for automating data wrangling problems

Domain: Dates

Appropriate Background Knowledge (BK)

Size (b)

Minimum (d)

Goal

Automation of data wrangling tasks, controlling d and b of the inductive inference by choosing the correct domain-specific background knowledge (DSBK) for the problem.

Experiments

Infer the subset of appropriate functions:

1. Take the first example from a dataset
2. Extract its metafeatures
3. Detect the domain
4. Predict & Ranking functions



- 6 Domains
- 124 datasets (91 for training & 33 for testing)
- 54 Metafeatures (descriptive characteristics)
- IP Learning System: MagicHaskeller
- 5 Strategies for selecting BK

Conclusions

We have a general IP system that:

1. Can work with different BK
2. Is able to solve many data wrangling problems
3. Works by using only one example of data



Results at the poster!

