Estimation-Based Search Space Traversal in PILP Environments

Joana Côrte-Real, Inês Dutra, and Ricardo Rocha
{*jcr, ines, ricroc*}*@dcc.fc.up.pt*
CRACS & INESC TEC, Faculty of Sciences, University of Porto

Introduction

Probabilistic Inductive Logic Programming (PILP) extends ILP by:

- representing knowledge using **probabilistic facts and rules**.
- learning probabilistic theories that can be used for **prediction**.

PILP theory search space uses algorithms similar to ILP to generate the logical part of the theories. It suffers from the same search space traversal efficiency issues as ILP. This adds a level of complexity w.r.t. ILP because every theory must be probabilistically evaluated.

Estimation Pruning

PILP pruning strategies can **reduce the number of candidate** theories based on their probabilistic values, which results in a **shorter execution time**.

Estimation pruning:

- calculates the **interval** where the predictions of a combination of two theories may lie.
- **estimates predictions** for the combination of theories (using a given estimator).
- **excludes** estimated theories that are too specific (AND operation) or too general (OR operation).

Estimators can be used to **determine the estimated predictions** of a theory within the possible interval of values.

	AND	OR
minimum	$\boldsymbol{max(0, A+B-1)}$	$\boldsymbol{max(A, B)}$
maximum	$\boldsymbol{min(A, B)}$	$\boldsymbol{min(A+B, 1)}$
center	$\frac{1}{2}(\boldsymbol{min(A, B)} + \boldsymbol{max(0, A+B-1)})$	$\frac{1}{2}(\boldsymbol{max(A, B)} + \boldsymbol{min(A+B, 1)})$
independence	$\boldsymbol{A \times B}$	$\boldsymbol{A+B-A \times B}$
exclusion	$\boldsymbol{max(0, A+B-1)}$	$\boldsymbol{min(A+B, 1)}$

Pruning criteria

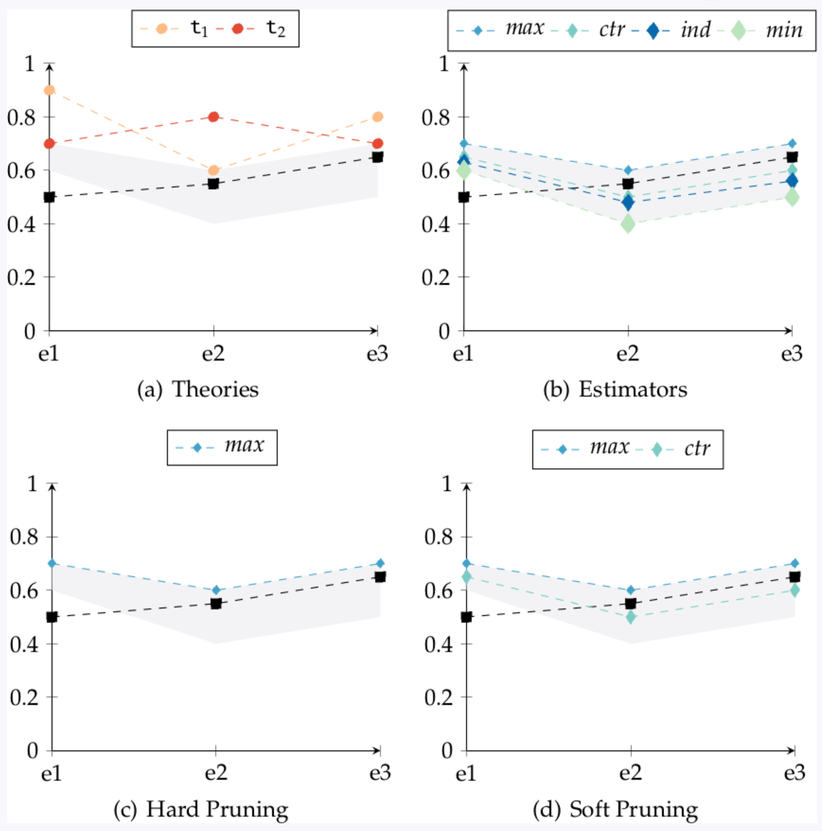
Pruning criteria are used to **decide if a theory should be evaluated** exactly based on its estimations.

The **hard** pruning criterion excludes theories if they too specific (AND operation)/general (OR operation) for **any estimation**.

The **soft** pruning criterion excludes theories if **all estimations** are overall too specific (AND operation)/general (OR operation).

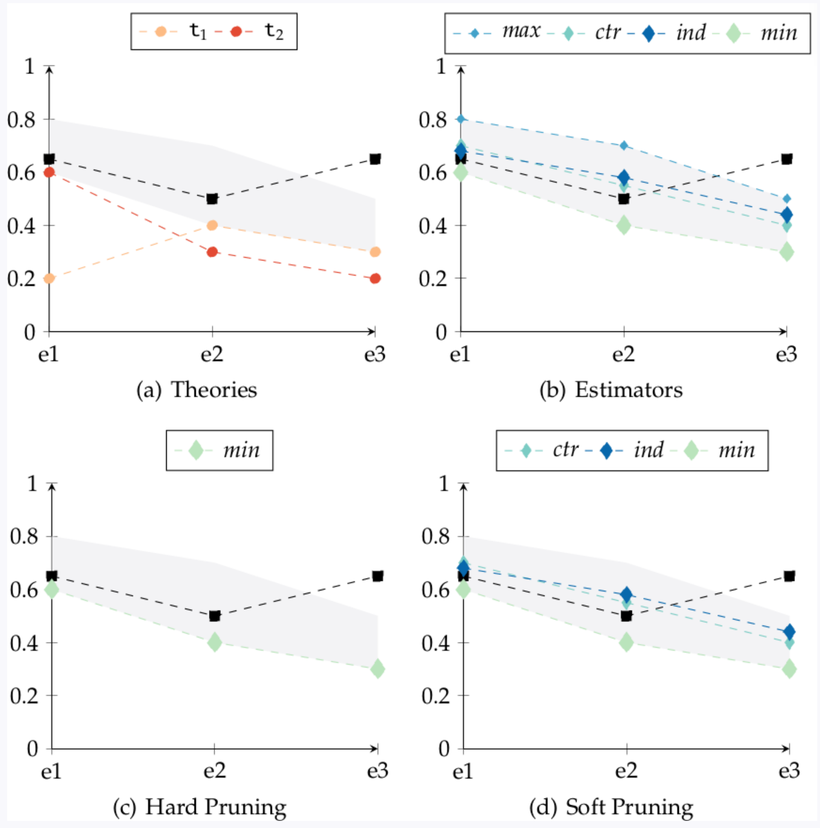
AND operation

Estimation pruning in the AND operation excludes combinations that are **too specific**, i.e. below the example values in Figures (c) and (d).



OR operation

Estimation pruning in the OR operation excludes combinations that are **too general**, i.e. above the example values in Figures (c) and (d).



Experiments

Three datasets were used in the experimental section:

metabolism adaptation of the dataset originally from the 2001 KDD Cup Challenge.

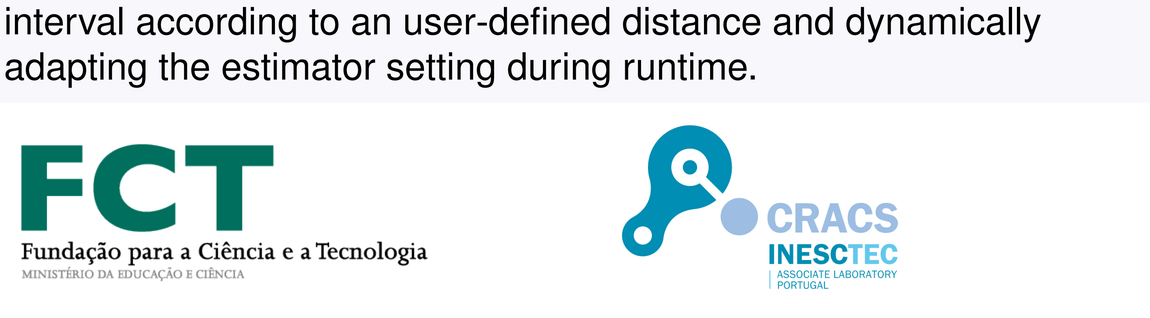
breast cancer data from 130 biopsies dating from January 2006 to December 2011.

athletes subset of facts regarding athletes and the sports they play collected by the never-ending language learner NELL.

Dataset	Examples	PBK	Folds	Size train	Size test
metabolism	230	7000 (46%)	30 n	160 (70%)	70 (30%)
athletes	721	4294 (100%)	30 n	505 (70%)	216 (30%)
breast cancer	130	13400 (3%)	130 lo	129 (99%)	1 (1%)

Results show that:

- All estimators **maintain predictive quality** and **reduce execution time**.
- The **HH pruning setting** showed the greatest speedups and also the greatest reduction in the number of probabilistic evaluations performed.
- Estimators *maximum* and *exclusion* are overall faster.



Conclusion

In this work:

- We proposed **five PILP estimators** to reduce the overhead caused by evaluation of candidate probabilistic theories.
- The estimators were implemented in the estimation pruning stage of the SkILL system, but can be **generalized to any PILP engine**.
- Candidate theories can be selected using **two pruning criteria**: soft and hard.
- Experiments using different pruning combinations were performed on **three real-world datasets**.

Future work includes adding an estimator that divides the estimation interval according to an user-defined distance and dynamically adapting the estimator setting during runtime.

Acknowledgements

FCT grant Joana Côrte-Real is funded by the FCT grant SFRH/BD/52235/2013.

ERDF COMPETE 2020 Programme within project POCI-01-0145-FEDER-006961

National Funds part of project UID/EEA/50014/2013 and North Portugal Regional Operational Programme, under the PORTUGAL 2020 Partnership Agreement

NanoSTIMA European Regional Development Fund as part of project NORTE-01-0145-FEDER-000016

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

CRACS INESCTEC
ASSOCIATE LABORATORY PORTUGAL