Multi-Label Based Learning for Better Multi-Criteria Ranking of Ontology Reasoners

Nourhène ALAYA (†,*), Myriam LAMOLLE (†) and Sadok BEN YAHIA (*)

(†) University of Paris 8, IUT of Montreuil, LIASD, France
(*) University of Tunis EL Manar, Faculty of Sciences de Tunis, LIPAH Tunisia

25 October 2017

16th International Semantic Web Conference (ISWC 2017)
I. Introduction

II. Proposed approach

III. Experimental evaluation

IV. Conclusion and future works
Inferring knowledge
Consistency

Reasoners
key components to work with OWL ontologies

Introduction
Proposed approach
Evaluations
Conclusion and future work

Context

In practice, reasoning is feasible

Theoretical complexity
Empirical complexity
## Motivation

### Disparity of reasoners’ computing times: the efficiency problem

<table>
<thead>
<tr>
<th>Ontologie</th>
<th>Profil OWL</th>
<th>LD</th>
<th>Classes</th>
<th>Axioms</th>
<th>Konclude</th>
<th>MORe</th>
<th>Hermit</th>
<th>FaCT++</th>
<th>TrOWL</th>
</tr>
</thead>
<tbody>
<tr>
<td>cas A</td>
<td>SHIQ(D)</td>
<td>DL</td>
<td>6967</td>
<td>17698</td>
<td>979</td>
<td>6594</td>
<td>6289</td>
<td>3643</td>
<td>9224</td>
</tr>
<tr>
<td>cas B</td>
<td>SHIQ(D)</td>
<td>DL</td>
<td>6193</td>
<td>14383</td>
<td>TIMEOUT</td>
<td>TIMEOUT</td>
<td>TIMEOUT</td>
<td>TIMEOUT</td>
<td>20360</td>
</tr>
</tbody>
</table>

Same size and expressivity class

Different computational times

### Disparity of reasoners’ computed results: the correctness problem

<table>
<thead>
<tr>
<th>Ontologie</th>
<th>Profil OWL</th>
<th>LD</th>
<th>Classes</th>
<th>Axiomes</th>
<th>Agreement level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cas C</td>
<td>DL</td>
<td>SHOIN(D)</td>
<td>106</td>
<td>723</td>
<td>60%</td>
</tr>
<tr>
<td>Cas D</td>
<td>DL</td>
<td>SHOIN(D)</td>
<td>145</td>
<td>362</td>
<td>100%</td>
</tr>
</tbody>
</table>

Same size and expressivity class

Disagreement

Various explanations

No tools to cope with these phenomena

[Gardiner et al., 2006] [Wang et al., 2007] [Gonçalves et al., 2012] [Lee et al., 2015]
Research question

**Growing number of reasoners**

**Various optimization techniques**

- JFacT
- MORe
- FaCT++
- Racer

**Variability of empirical performances**

- PelleT
- ELK
- TrOWL
- HermiT
- Konclude

**Lack of Knowledge**

How to assist a user to figure out the “best” reasoner to handle its ontology?

**OWL Profile**  **Efficiency**  **Correctness**
Related works

**Use machine learning techniques**
learn reasoner future behaviors from its past running

**Predict single reasoner performances given an input ontology**

- Predict the reasoner runtime [Kang et al., 2012, 2014] [Sazonau et al., 2014]
- Predict the reasoner robustness [Alaya et al., 2015]

**Predict the ranking of a set reasoners given an input ontology**

<table>
<thead>
<tr>
<th>Meta-Reasoner R2O2 [Kang et al., 2015]</th>
<th>RakSOR [Alaya et al., 2016]</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ Reasoning optimization technique</td>
<td>+ User assistance</td>
</tr>
<tr>
<td>- Only the runtime as criteria</td>
<td>+ Runtime and correctness</td>
</tr>
<tr>
<td>- No user assistance</td>
<td>- Only DL ontologies</td>
</tr>
<tr>
<td>- Cost of the prediction steps</td>
<td>- Complex process</td>
</tr>
<tr>
<td>- Only DL ontologies</td>
<td>- Time consuming</td>
</tr>
</tbody>
</table>
Outline

I. Introduction

II. Proposed approach

III. Experimental evaluation

IV. Conclusion and future works
**Contributions**

- **Multi-RakSOR**: assistance system to help a user select “best” reasoner for an ontology based application
  - **Multi-Criteria**: OWL profile, Correctness [Gardiner et al., 2006], Efficiency
  - **Informative**: Relevant/ Irrelevant reasoners are included in the ranking
  - **Multi-label learning paradigm**

- **Meta-RakSOR**: the evolvement of the ranking solution into a meta-reasoner

---

**Given an ontology within an OWL profile, the most preferred reasoner is the one capable to successfully achieve the reasoning task within the shortest time span**

- **Correct and in time**
  - **Unexpected and in time**
  - **Timeout**
  - **Halt**

**success**

**Failure**
Bucket order rules

\( \mathcal{R} = \{R_1, \ldots, R_m\} \) alternative reasoners

\( \mathcal{B} = \{B_1, \ldots, B_k\} \) buckets stand as a partition over \( \mathcal{R} \)

- \(<\) preferred to
- \((\mathcal{B}, <)\) strict total order across the buckets over \( \mathcal{R} \).
- Reasoners within one bucket are equally ranked (tied).

Ranks

|   | 4 | 2 | 3 | 1 | 4 | 6 | 6 |

Relevance

|   | 0 | 1 | 0 | 1 | 0 | 0 | 0 |
System Architecture: Multi-RakSOR

- **Ontologies**
- **Reasoner empirical evaluations**
- **Ranking component**
- **Ontology Profiler**

### Learning Dataset

<table>
<thead>
<tr>
<th>Features</th>
<th>Ranking</th>
<th>Relevance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ontos X0, X1, ..., Xd</td>
<td>Y11, Y12, ..., Ym1</td>
<td>Y12, Ym2</td>
</tr>
<tr>
<td>O1, x1, x1(1), ..., x1(d)</td>
<td>21, 3, ..., m</td>
<td>110, 101, ..., 00</td>
</tr>
<tr>
<td>On, x1(n), ..., x1(n)</td>
<td>13, 5, ..., m</td>
<td>101, 01, ..., 01</td>
</tr>
</tbody>
</table>

### Learning Component
- Repeat the learning for each OWL profile

### Multi-RakSOR Predictive Models
- **Proposed approach**
- **Evaluations**
- **Conclusion and future work**

- **Online**
- **Offline**

- **New Ontology**
- **Ontology Profiler**
- **Prediction Component**

- \((x_1, x_2, ..., x_d)^T\)

- **Ranked list of reasoners**
- **Predicted relevance (Success/Failure) per Reasoner**
Multi-label learning paradigm

\[ X = (x_1, x_2, \ldots, x_d) \xrightarrow{h(X)} \hat{Y} = (\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_m), \quad X \in \mathcal{X} \text{ and } \hat{Y} \in \mathcal{Y} \]

Ranks

\[ R_{s1} < \ldots < R_{sk} \]

Relevance

\( \mathcal{P}_\chi \) (relevant reasoners)
\( \mathcal{N}_\chi \) (irrelevant reasoners)

Learning to predict the ranks, \( \mathcal{Y} = \mathbb{R}^m \)

Learning to predict the relevance, \( \mathcal{Y} = \{0,1\}^m \)

Multi-Target Regression Algorithm, \( \text{ERC} \) [Xioufis et al., 2016]

Multi-label Classification Algorithm, \( \text{BR} \) [Ioannou et al., 2010]

Multi-RakSOR Learner
Ranks of relevant reasoners must respect a strict total order relation.

No irrelevant reasoner should be ranked lower than a relevant one.

\[ \forall R_i \in P_x \text{ and } \forall R_j \in N_x \text{ then } R_i < R_j \Rightarrow \sigma(R_i) < \sigma(R_j) \]
Example: 6 alternative reasoners

<table>
<thead>
<tr>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>R5</th>
<th>R6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Predicted ranks

Predicted relevance

Ranks of relevant reasoners

Ranks of irrelevant reasoners

Final Ranks

Introduction

Proposed approach

Evaluations

Conclusion and future work

Correction

2nd rule is not respected

Correction

Correction

not strict linear order
I. Introduction

II. Proposed approach

III. Experimental evaluation

IV. Conclusion and future works
Data describing the past running of a set of reasoners over a large set of ontologies.

Experimental Settings

- ORE Evaluation Framework (efficiency and correctness) [Gardiner et al., 2006]
- Ontologies [ORE Corpus, 2015]
  - 1191 DL
  - 763 EL
- Tasks: 2 classification challenges (DL, EL)
- 10 Reasoners
  - DL challenge: Konclude, FaCT++, MORe, JFacT, TrOWL, HermiT, Racer, Pellet
  - EL challenge: DL challenge reasoners + ELK + ELepHanT
- Time limit: 3 minutes,
- Memory: 10 Go Ram
- Machine: Intel Core I7, CPU running at 3.4GHz
Reasoner evaluation results

Results of DL classification challenge

Per Reasoner Success/Failure Rates

<table>
<thead>
<tr>
<th>Reasoner</th>
<th>Success</th>
<th>Unexpected</th>
<th>Timeout</th>
<th>Halt</th>
</tr>
</thead>
<tbody>
<tr>
<td>JFact</td>
<td>605</td>
<td>731</td>
<td>823</td>
<td>858</td>
</tr>
<tr>
<td>Racer</td>
<td>887</td>
<td>858</td>
<td>1005</td>
<td>1124</td>
</tr>
<tr>
<td>TrOWL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pellet</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FaCT++</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MORe</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hermit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Konclude</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Average reasoner runtime over the successfully classified ontologies

Results of EL classification challenge

Per Reasoner Success/Failure Rates

<table>
<thead>
<tr>
<th>Reasoner</th>
<th>Success</th>
<th>Unexpected</th>
<th>Timeout</th>
<th>Halt</th>
</tr>
</thead>
<tbody>
<tr>
<td>JFact</td>
<td>537</td>
<td>617</td>
<td>665</td>
<td>695</td>
</tr>
<tr>
<td>Racer</td>
<td>717</td>
<td>741</td>
<td>750</td>
<td>752</td>
</tr>
<tr>
<td>FaCT++</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pellet</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TrOWL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hermit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELephant</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Konclude</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MORe</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELK</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Average reasoner runtime over the successfully classified ontologies
Multi-RakSOR Evaluations

Data: Creating train and test datasets

<table>
<thead>
<tr>
<th></th>
<th>Train dataset</th>
<th>Test dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>OWL DL profile</td>
<td>800 (ontologies)</td>
<td>391</td>
</tr>
<tr>
<td>OWL EL profile</td>
<td>500</td>
<td>263</td>
</tr>
</tbody>
</table>

Models: Learning Relevance and Rank prediction models of Multi-RakSOR (4 models)

Task 1: Assessment of Multi-RakSOR relevance prediction model

![Graph for OWL DL Dataset](image)

![Graph for OWL EL Dataset](image)
Multi-RakSOR Evaluations

**Task 2: Quality assessment of Multi-RakSOR predicted rankings**

Assessing the correlation between the real and the predicted reasoner rankings

**Results over DL Dataset**

<table>
<thead>
<tr>
<th>Reasoner</th>
<th>SPearmanRHO</th>
<th>KendallTauX</th>
</tr>
</thead>
<tbody>
<tr>
<td>MultiRakSOR</td>
<td>0.86</td>
<td>0.80</td>
</tr>
<tr>
<td>ERC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR-kNN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LRT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCTR</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Results over EL Dataset**

<table>
<thead>
<tr>
<th>Reasoner</th>
<th>SPearmanRHO</th>
<th>KendallTauX</th>
</tr>
</thead>
<tbody>
<tr>
<td>MultiRakSOR</td>
<td>0.94</td>
<td>0.90</td>
</tr>
<tr>
<td>ERC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR-kNN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LRT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCTR</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Assessing the ranking quality at the top of the ranked lists

**Results over DL Dataset**

<table>
<thead>
<tr>
<th>Reasoner</th>
<th>MAP@1</th>
<th>MAP@3</th>
</tr>
</thead>
<tbody>
<tr>
<td>MultiRakSOR</td>
<td>0.89</td>
<td>0.80</td>
</tr>
<tr>
<td>LRT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR-kNN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCTR</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Results over EL Dataset**

<table>
<thead>
<tr>
<th>Reasoner</th>
<th>MAP@1</th>
<th>MAP@3</th>
</tr>
</thead>
<tbody>
<tr>
<td>MultiRakSOR</td>
<td>0.88</td>
<td>0.84</td>
</tr>
<tr>
<td>LRT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR-kNN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCTR</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Meta-RakSor Evaluations

**Meta-Reasoner**: Collection of reasoners + Intelligent selector [Kang et al., 2015], [SatZilla, 2008]

**Meta-RakSor**: Multi-RakSor Predictor + Calling the best ranked reasoner

**Experiments**: 2 ontology classification challenges (391 DL, 291 EL)

- **Count of Correctly Classified DL Ontologies**
- **Count of Correctly Classified EL Ontologies**

- **Average Runtime Over Correctly Classified DL Ontologies**
- **Average Runtime over Correctly Classified EL Ontologies**
Outline

I. Introduction
II. Proposed approach
III. Experimental evaluation
IV. Conclusion and future works
Our contributions are at the cross roads of two fields: the Semantic Web and the Machine Learning

- Multi-RakSOR can provide highly accurate ranking of reasoners based on correctness and efficiency while considering their capabilities over particular OWL profiles.
- Meta-RaksOR can outperform all of the studied reasoners in terms of result correctness.
- Meta-RakSOR shows good efficiency level over DL ontologies, but it is less efficient over EL ontologies.
Future works

Improve
- Meta-RakSOR efficiency over EL ontologies by setting up a default reasoner for the lightweight ontologies.

Extend
- Considering more reasoning tasks, different memory and machine configuration.

Embed
- our reasoner ranking solution in the meta-reasoner Chainsaw or in ontology publishing solutions like WIDOCO

Upcoming ..... the Web application
Thank you for your attention ...

Demo and Datasets are available on GitHub: https://github.com/Alaya2016/Multi-RakSORDemo/

Contact:
Nourhène ALAYA
n.alaya@iut.univ-paris8.fr
nourhen.alaya@inria.fr

Any questions? ...
References

Ontology Features

Ontology Profiler

\[ X = (x_1, x_2, \ldots, x_d)^T \subseteq \mathcal{F}^d \]

Vector of ontology feature Values

[Alaya et al., WIMS 2015]