

Homework 8: PSL

DSCI 558 BUILDING KNOWLEDGE GRAPH

DUE DATE: Sunday, 04/04/2021 @ 11:59pm on Blackboard

Ground Rules

This homework must be done individually. You can ask others for help with the tools, but the submitted homework needs to be your own work.

Summary

In this homework, you will use Probabilistic Soft Logic

Task 1 – PSL tutorials (3pts)

Complete the PSL tutorial: <https://psl.lings.org/blog/2018/07/15/getting-started-with-psl.html>.

Complete the PSL entity resolution: <https://github.com/lings/psl-examples/blob/master/entity-resolution/>.

- Hint: Please check the PSL model and all the downloaded data files carefully so that you can understand the meaning of all predicates and rules. You *may* need these for Task 2.
- Hint: You should run the CLI version of the tutorial since it is the easiest way to run PSL. You are free to run/use other versions of PSL such as Python or Java in this homework if you want.

You need to submit the “*inferred-predicates*” folder which includes the target results files for each tutorial.

Task 2 – PSL relational entity resolution (7 pts)

Check the “Dataset Assignment.csv” file for your assigned dataset in this task. Download your dataset from <https://sites.google.com/site/anhaidgroup/useful-stuff/data>

For each dataset, you need to solve a **relational ER** problem with two different types of entities as shown in the “Dataset Assignment.csv” file.

The main idea of **relational ER** is that performing ER in one type of the entities can help solve the ER problem on the other type of entities. For example, ER for authors can help resolve ER for papers as shown in the ER tutorial. In this task, you need to:

- Identify another type of entities in your dataset: movies should be the first one.
State your choice in the report.

- Design a PSL model and generate the necessary data files (observations and targets).
List and explain the predicates and rules in your report.
- Run PSL inference to perform ER. Verify your performance using the labeled file L in your dataset. Include folder “inferred-predicates” in your submission. Report your precision, recall and F1-score in your report.
 - o Note: There will be two types of entities but you only need to report the performance of the entity type shown in L.
- Hint:
 - o You can modify the PSL model from the ER tutorial to fit in your dataset or you can create your own PSL model if you want.
 - o You need to make sure that your PSL model has collective ER rules (similar to rules shown in the lecture slides below):

Collective KG-based ER rules

$$\begin{array}{ll}
 \text{Relational equivalence:} & \text{CANDSAME}(E_1, E_2) \wedge \text{CANDSAME}(O_1, O_2) \\
 \text{Matching entities} & \wedge \text{SIM}(E_1, E_2) \wedge \text{SAME}(O_1, O_2) \\
 \text{should relate to objects} & \wedge \text{REL}(E_1, O_1, R) \wedge \text{REL}(E_2, O_2, R) \\
 \text{in the same way} & \Rightarrow \text{SAME}(E_1, E_2) \quad (12)
 \end{array}$$

$$\begin{array}{ll}
 \text{Relational parity:} & \text{CANDSAME}(E_1, E_2) \wedge \text{CANDSAME}(O_1, O_2) \\
 \text{If entities do not relate} & \wedge \text{SAME}(O_1, O_2) \wedge \neg \text{REL}(E_1, O_1, R) \\
 \text{to objects in the same} & \wedge \neg \text{NEWENTITY}(E_1) \wedge \neg \text{NEWENTITY}(O_1) \\
 \text{way they should not} & \wedge \text{REL}(E_2, O_2, R) \\
 \text{match} & \Rightarrow \neg \text{SAME}(E_1, E_2) \quad (13)
 \end{array}$$

Submission Instructions

You must submit the following files a single .zip archive named Firstname_Lastname_hw6.zip and submit it via Blackboard:

- Firstname_Lastname_hw8.pdf: A pdf file that contains your answers in Task 2
- Firstname_Lastname_hw8_psl.zip: A zip file that contains all the files needed for your PSL run (“.psl”, “.data” and “.txt” files).
- Firstname_Lastname_hw8_result.zip: A zip file that contains all “inferred-predicates” folders from Task 1 and Task 2
- Firstname_Lastname_hw8_src.zip: A zip file that contains all the source code you wrote for this homework.