Explanation for Experiment Result

Hai Dang Tran

November 23, 2016

In this folder, there are two sub-folders for IMDB and YAGO. Each of them contain the following files:

- **patterns.txt.** Each line in this file is in the format: h p q f, that means a positive rule h(X, Y) :- p(X, Z), q(Z, Y) (see Setup part in Evaluation section of the paper) with f being its absolute support, that is, the number of triples (X, Y, Z) that satisfies h(X, Y), p(X, Z), q(Z, Y) (see Relational Association Rule Mining part in Preliminaries section of the paper). In patterns.txt file all rules are sorted by decreasing order of absolute support.

- **selected.patterns.txt.** This file has the same format as patterns.txt file. We only add exceptions to Horn rules in this file and the corresponding nonmonotonic rules are used to infer new facts. The quality of these rules are described in Table 2 of the paper.

- **ideal.data.txt.** Each line in this file indicates a triple with the format <subject> <predicate> <object> (see Motivation part in the section Introduction of the paper). This file is the approximated ideal graph used in the Evaluation section.

- **training.data.txt.** This file has the same format as ideal.data.txt file and it is the training data in the Evaluation section. We generate this file by removing 20% facts for every predicate from ideal.data.txt file (see Dataset part of Evaluation section).

- **revised-rules.naive.** This file is generated by Naive ranking using top 100 patterns in patterns.txt and sampled ones in selected.patterns.txt.

There are two parts in this file, one for Naive ranking and the other for chosen revised rules at the end. The first part is the list of Horn rules given in patterns.txt file with statistics such as conviction (Conv) and confidence (Conf) measure, followed by their top exceptions. The maximum number of top exceptions for each rule is 10 and they are sorted by decreasing order of positive-negative conviction (PosNegConv, please see formula 8 in Step 4, section Methodology of the paper). In case two exceptions have the same positive-negative conviction, the one which has higher standard conviction measure (Conv) will have higher rank. The second part is a list of best revisions (rank 1) of sampled Horn rules in selected.patterns.txt file.

- **revised-rules.pm.** This file is similar to revised-rules.naive, the only difference is that it is generated by PM ranking.

- **revised-rules.opm.** This file is similar to revised-rules.naive, the only difference is that it is generated by OPM ranking.

- **extension.pos.** This file is a list of new facts predicted from training data by sampled positive rules in selected.patterns.txt file. It has the same format as training file.

- **extension.naive.neg.** This file is similar to extension.pos file, the only difference is that it is generated by Naive revision of sampled Horn rules (the second part of revised-rules.naive).

- **extension.pm.neg.** This file is similar to extension.pos file, the only difference is that it is generated by PM revision of sampled Horn rules (the second part of revised-rules.pm).

- **extension.opm.neg.** This file is similar to extension.pos file, the only difference is that it is generated by OPM revision of sampled Horn rules (the second part of revised-rules.opm).