TRANSFER LEARNING FOR BOOSTED RELATIONAL DEPENDENCY NETWORKS THROUGH GENETIC ALGORITHM

Leticia Freire de Figueiredo, Aline Paes, Gerson Zaverucha

Engenharia de Sistemas e Computação PESC/Coppe

COPPE UFRJ
Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia

INTRODUCTION

1. RDN-Boost: statistical relational learning method

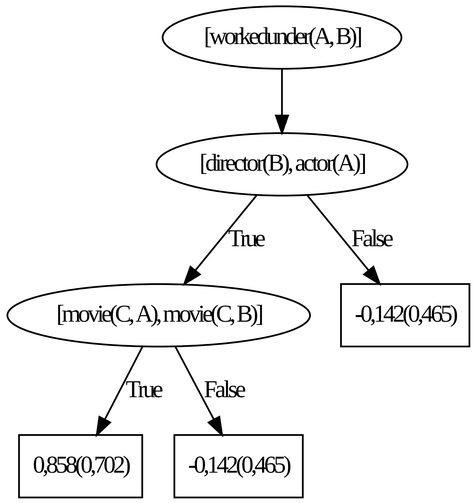

Figure 1: RDN-Boost with 1 tree.

2. When data is scarce and collecting more data is hard, it is necessary to use transfer learning technique.

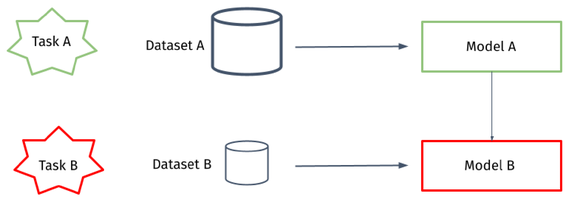

Figure 2: Transfer learning example between two tasks.

3. Transfer learning in relational data is a challenge. An idea is to use GENETIC ALGORITHM.

GROOT

1. Make transfer learning, finding the best mapping between the predicates, using genetic algorithm in a relational setting, with trees generated by the source dataset using the RDN-Boost

Creating the population

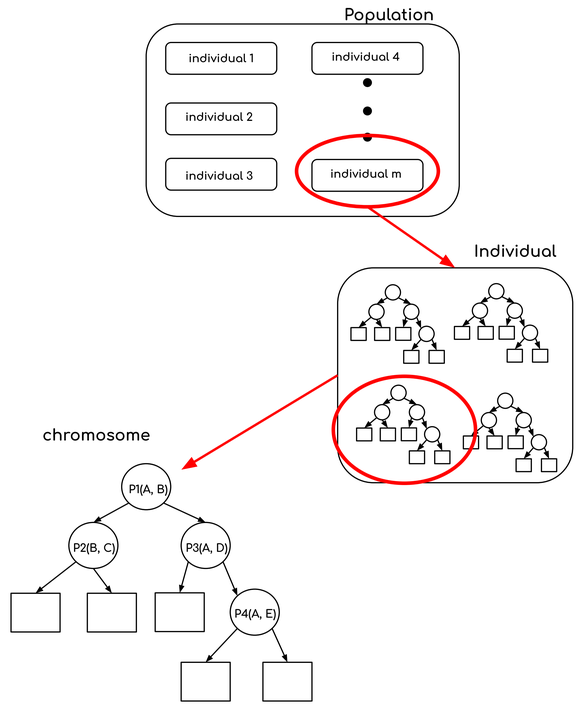

Figure 3: Population definition.

Genetic Operators

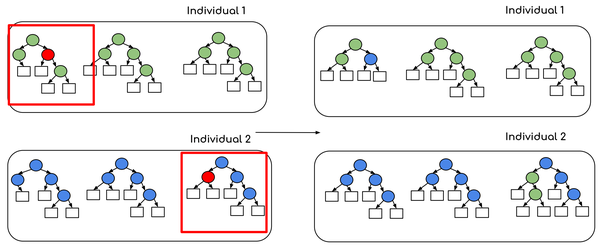

Figure 4: Example of crossover between two individuals with 3 chromosomes. The red nodes are selected to be exchanged.

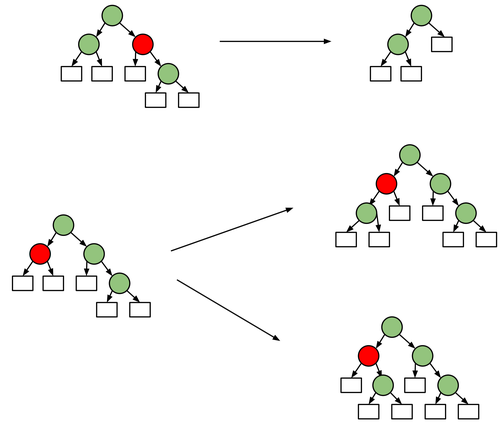

Figure 5: A mutation example. Top: pruning starts from the red node and erases all the nodes below it. Bottom: the expansion example shows the possibility to include a new node in one leaf.

EXPERIMENTAL RESULTS

Methodology

Table 1: Genetic algorithm hyperparameters.

	Parameters
Mutation rate	[0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4]
Crossover rate	[0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95]
Number of individuals	[10, 30, 50]

Figure 6: Schema showing how evaluation occurs.

1. Cross-validation: each fold trained 5 times and tested with the remaining folds
2. Internal cross-validation: each training fold splitted in 3 sub-folds to optimize the genetic hyperparameters
3. Results compared with:
   (a) Learning from scratch:
   RDN-Boost, with 10 trees (RDN-B)
   RDN-Boost, with 1 tree (RDN-B-1)
   (b) Transfer with another framework:
   TreeBoostler
4. When training, the amount of negative has been the ratio of two negatives for one positive; when testing, all the examples from the test dataset are used

Results

Table 2: Results for the experiment with IMDB and UW-CSE datasets.

	IMDB → UW-CSE			
	CLL	AUC ROC	AUC PR	Time
RDN-B-1	-0.239 ± 0.000	0.796 ± 0.000	0.085 ± 0.000	2.595 ± 0.124 s
RDN-B	**-0.814 ± 0.003**	0.801 ± 0.005	0.094 ± 0.011	7.582 ± 0.457 s
TreeBoostler	-0.368 ± 0.004	0.905 ± 0.004	0.168 ± 0.014	11.152 ± 0.700 s
GROOT	-0.262 ± 0.033	**0.939 ± 0.010**	**0.336 ± 0.018**	18.3 ± 6.5 min

Table 3: Results for the experiment with IMDB and Cora datasets.

	IMDB → Cora			
	CLL	AUC ROC	AUC PR	Time
RDN-B-1	-0.213 ± 0.004	0.534 ± 0.008	0.012 ± 0.000	5.9 ± 5.1 min
RDN-B	-0.500 ± 0.01	0.542 ± 0.006	0.013 ± 0.001	42.8 ± 7.6 min
TreeBoostler	-0.325 ± 0.008	0.729 ± 0.006	0.261 ± 0.022	194.0 ± 50.1 min
GROOT	-0.326 ± 0.006	0.582 ± 0.005	0.183 ± 0.010	41.0 ± 0.6 min

Table 4: Results for the experiment with Cora and IMDB datasets.

	Cora → IMDB			
	CLL	AUC ROC	AUC PR	Time
RDN-B-1	-0.224 ± 0.000	0.843 ± 0.000	0.487 ± 0.000	2.249 ± 0.067 s
RDN-B	-0.697 ± 0.000	0.843 ± 0.000	0.487 ± 0.000	4.100 ± 0.137 s
TreeBoostler	-0.236 ± 0.000	0.958 ± 0.001	0.541 ± 0.055	9.564 ± 0.140 s
GROOT	-0.208 ± 0.015	**0.965 ± 0.012**	0.326 ± 0.176	86.4 ± 30.8 min

Table 5: Results for the experiment with Nell datasets.

	Nell Sports → Nell Finances			
	CLL	AUC ROC	AUC PR	Time
RDN-B-1	-0.178 ± 0.005	0.601 ± 0.069	0.045 ± 0.025	10.948 ± 2.003 s
RDN-B	**-0.284 ± 0.019**	0.796 ± 0.032	0.114 ± 0.027	56.401 ± 11.519 s
TreeBoostler	-0.167 ± 0.006	**0.979 ± 0.003**	0.083 ± 0.026	2.6 ± 0.8 min
GROOT	-0.197 ± 0.028	0.976 ± 0.012	**0.167 ± 0.080**	534.8 ± 219.3 m

ACKNOWLEDGEMENTS

CNPq Conselho Nacional de Desenvolvimento Científico e Tecnológico

FAPERJ Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro