

MAPPING ACROSS RELATIONAL DOMAINS FOR TRANSFER LEARNING WITH WORD EMBEDDINGS-BASED SIMILARITY

Thais Luca, Aline Paes, Gerson Zaverucha

INTRODUCTION

1. Most of the real-world data is relational.
2. Machine learning models assume training and test data to have the same distribution.

BACKGROUND KNOWLEDGE

Statistical Relational Learning

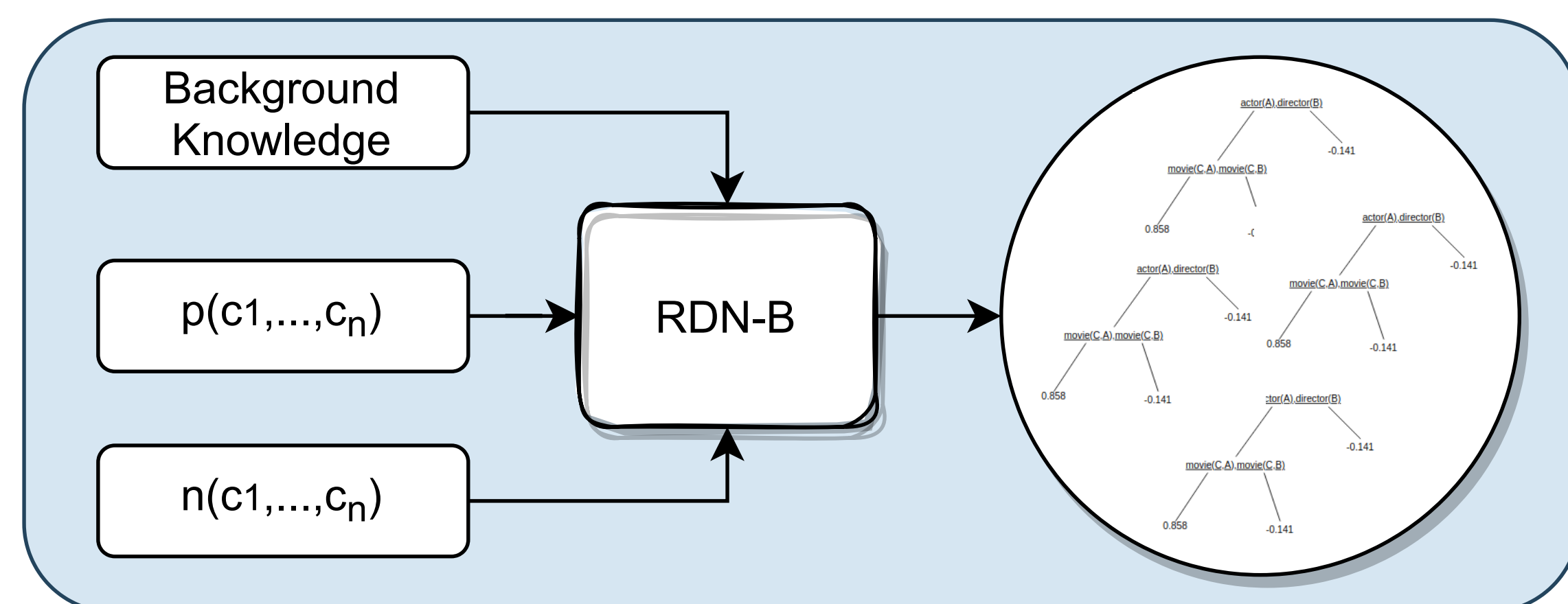


Figure 1: Conceptual Architecture of RDN-B

Transfer Learning

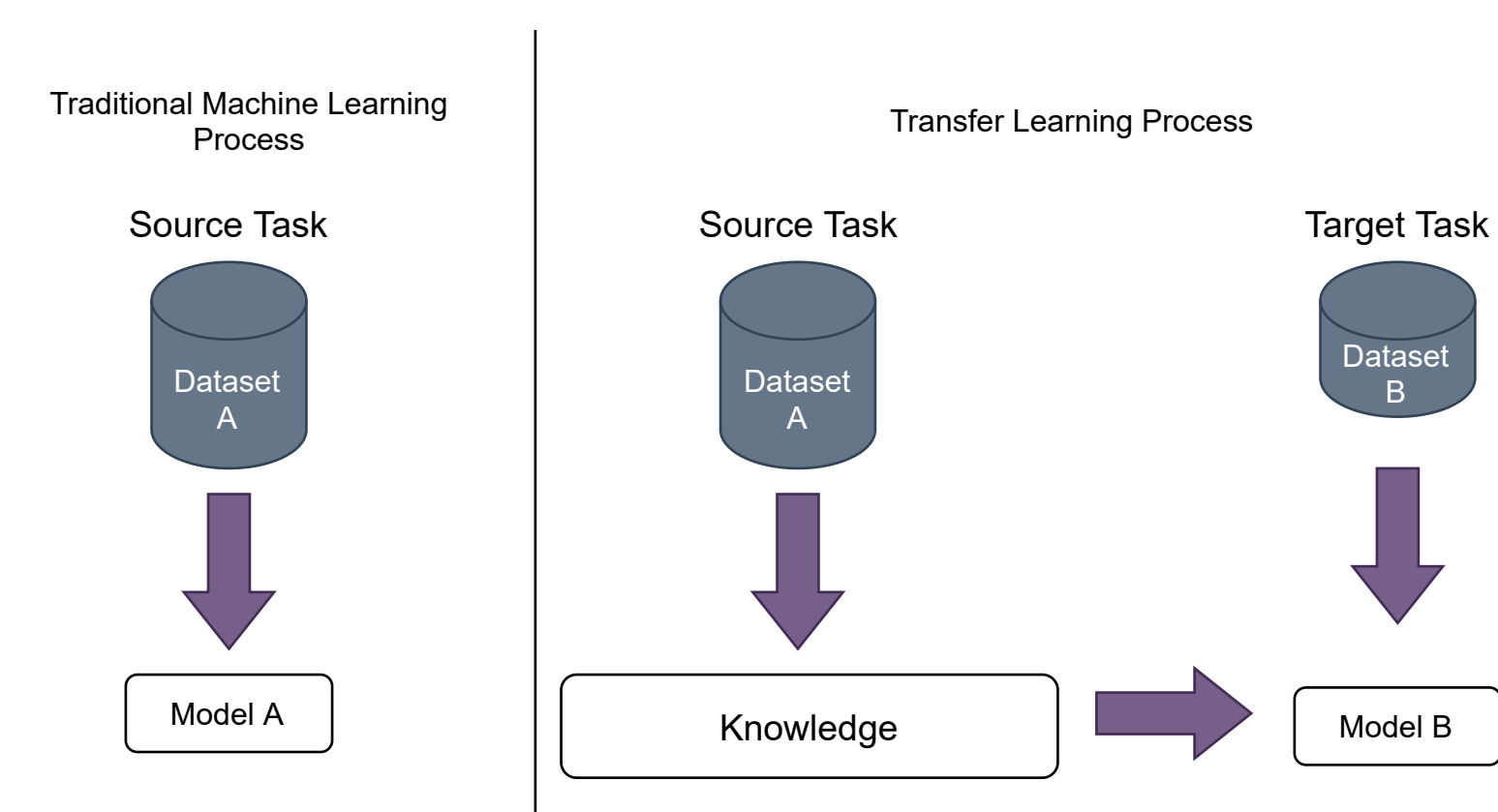


Figure 2: Difference between traditional machine learning processes (left) and transfer learning processes (right).

TRANSBOOSTLER

Mapping Component

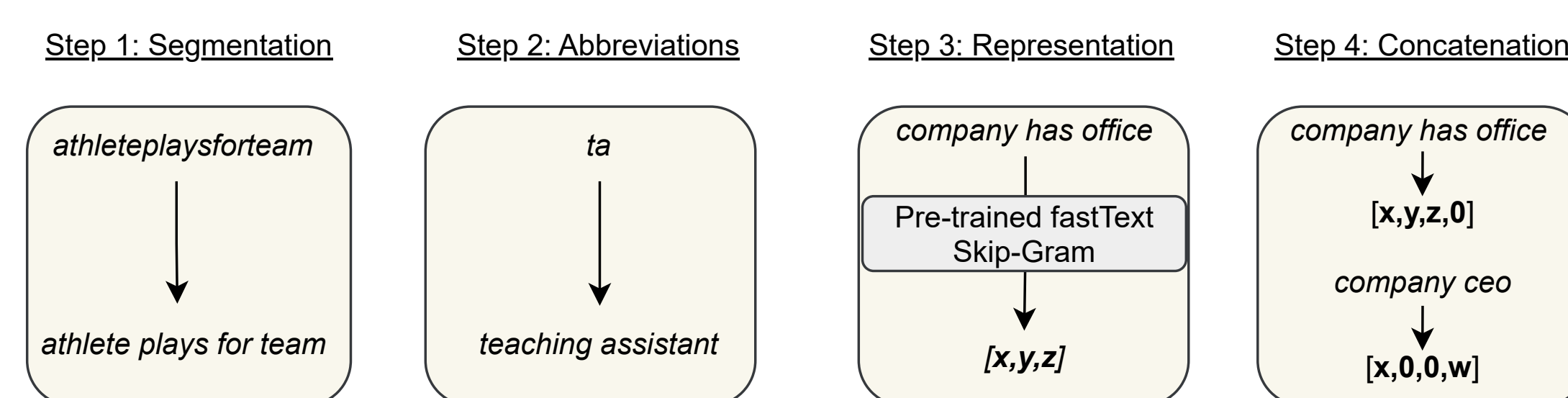


Figure 3: Pre-processing steps.

Similarity Metrics

Euclidean Distance $d(p, q) = \|p - q\|_2$

Soft Cosine $soft\ cosine(p, q) = \frac{\sum_{i,j} s_{i,j} p_i q_j}{\sqrt{\sum_{i,j} s_{i,j} p_i p_j} \sqrt{\sum_{i,j} s_{i,j} q_i q_j}}$

Word Mover's Distance (WMD)

Rank by Similarity

Similarity	Similarity
actor(person), student(person) 5.26	director(person), professor(person) 4.07
actor(person), professor(person) 5.36	director(person), student(person) 4.69

Table 1: Example of similarity ranking for *actor* and *director* using the Euclidean Distance.

Transferring and Revising the Structure

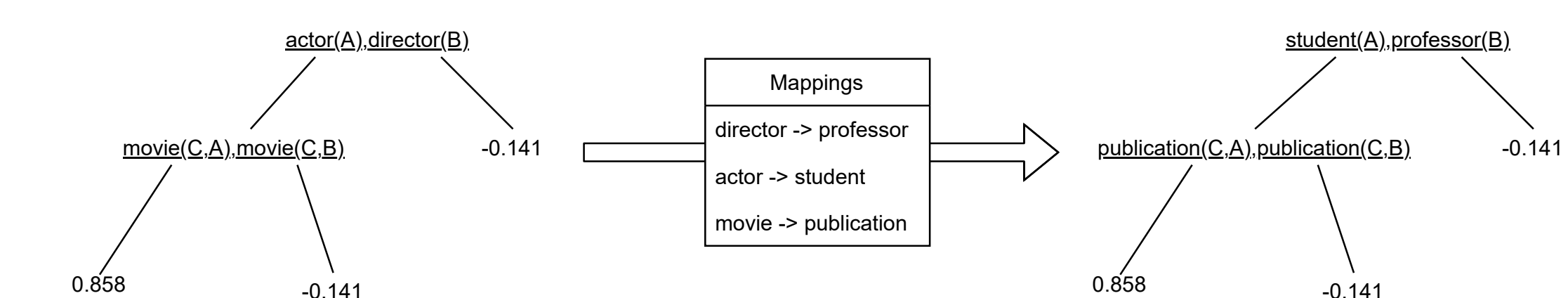


Figure 4: Transferring the structure by mapping similar predicates.

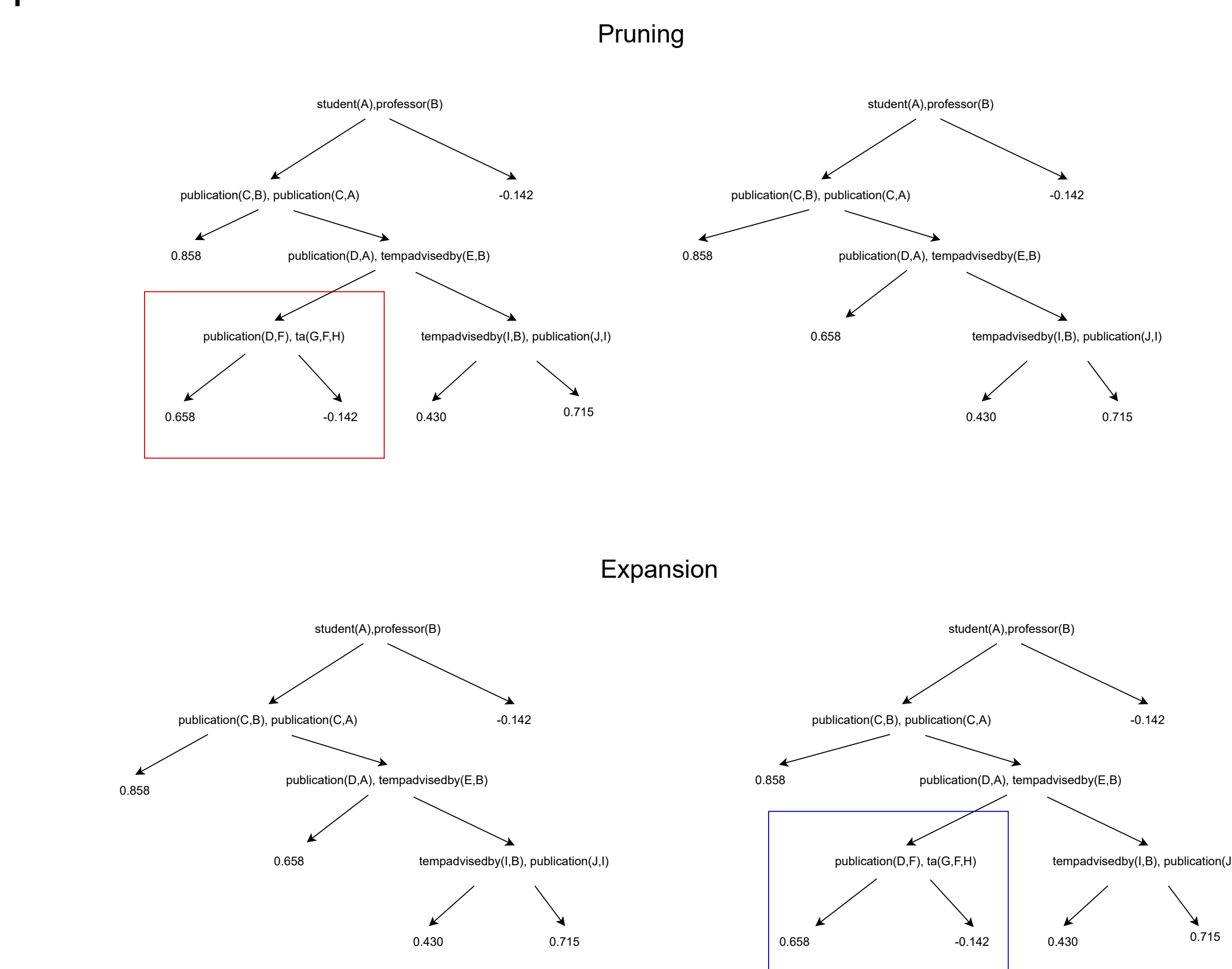


Figure 5: Example of pruning (top) and expansion (bottom) of nodes in a relational tree during revision.

Experimental Results

- indicates results only considering predicate mapping and parameter learning.
- ★ indicates when TransBoostler is significantly better than TreeBoostler.
- ◇ indicates that the difference between TransBoostler against RDN-B results is significantly better.

IMDB and Cora

	IMDB → Cora				Cora → IMDB			
	CLL	AUC	AUC	Run-	CLL	AUC	AUC	Run-
		ROC	PR	time(s)		ROC	PR	time(s)
RDN-B	-0.693	0.558	0.426	76.97	-0.075	1.000	1.000	2.89
TreeBoostler	-0.659	0.606	0.530	45.74	-0.075	0.999	0.954	4.29
TransBoostler Soft Cosine	-0.675★	0.599	0.464	51.18	-0.074	1.000★	1.000	4.36
TransBoostler Euclidean	-0.677	0.589	0.453	52.61	-0.076	0.999	0.927	4.42
TransBoostler WMD	-0.668	0.600	0.463	54.44	-0.076	0.999★	0.948	4.43
TreeBoostler•	-0.659	0.574	0.518	1.63	-0.115	0.982	0.888	0.95
TransBoostler• Soft Cosine	-0.699★	0.500	0.379	2.20	-0.306★◇	0.868	0.092	1.94
TransBoostler• Euclidean	-0.699★	0.500	0.379	2.15	-0.304★◇	0.868	0.092	1.90
TransBoostler• WMD	-0.699★	0.500	0.379	2.23	-0.308★◇	0.868	0.092	1.92

Table 2: Transfer experiment results for IMDB and Cora.

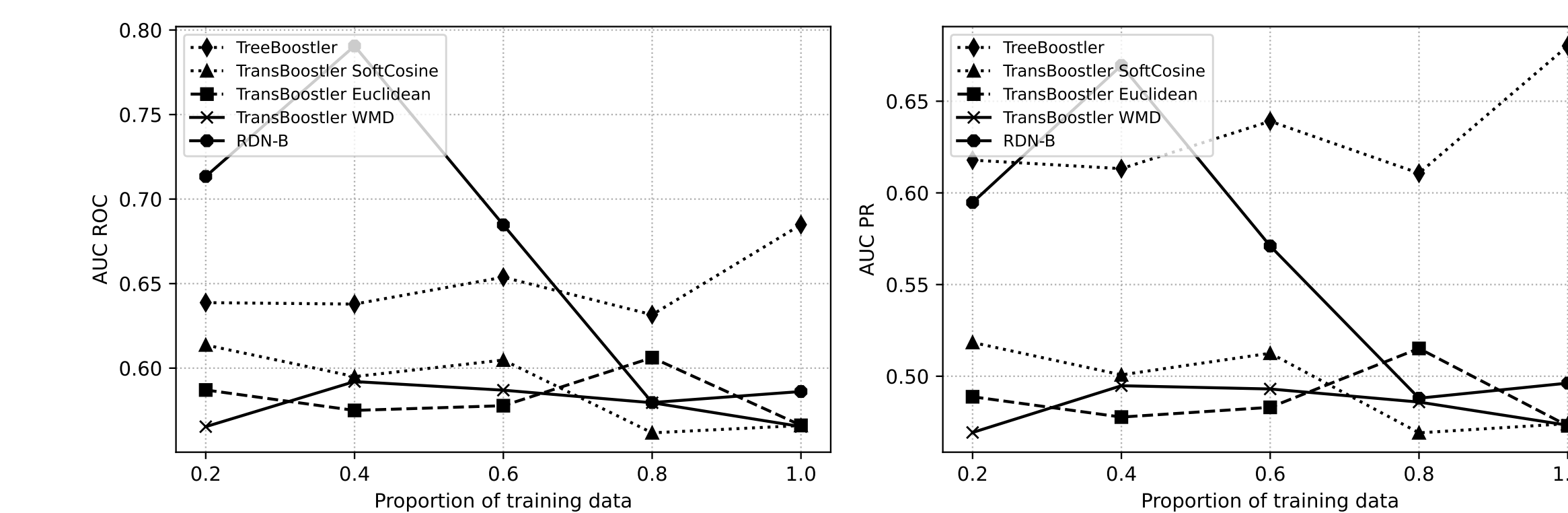


Figure 6: Learning curves for IMDB → Cora.

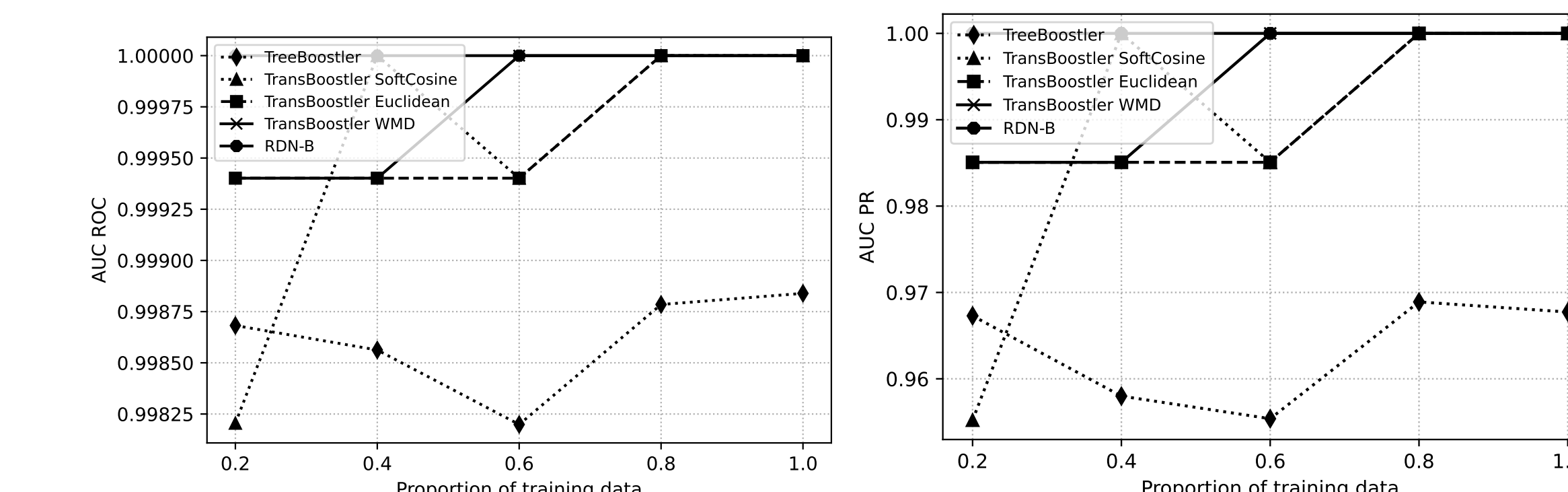


Figure 7: Learning curves for Cora → IMDB.

Yeast and Twitter

	Yeast → Twitter				Twitter → Yeast			
	CLL	AUC	AUC	Run-	CLL	AUC	AUC	Run-
		ROC	PR	time(s)		ROC	PR	time(s)
RDN-B	-0.122	0.990	0.347	23.45	-0.253	0.926	0.230	15.55
TreeBoostler	-0.096	0.994	0.395	86.63	-0.166	0.986	0.267	34.96
TransBoostler Soft Cosine	-0.127	0.994★	0.382	23.38	-0.280★◇	0.920	0.169	20.39
TransBoostler Euclidean	-0.107	0.994	0.389	26.09	-0.282★◇	0.894	0.325	13.05
TransBoostler WMD	-0.107	0.994	0.374	24.55	-0.240★	0.953	0.282	19.64
TreeBoostler•	-0.103	0.993	0.334	7.17	-0.166	0.986	0.267	2.17
TransBoostler• Soft Cosine	-0.154	0.993	0.339	4.51	-0.336★◇	0.820	0.299	3.35
TransBoostler• Euclidean	-0.110	0.994	0.405	5.65	-0.336★◇	0.820	0.307	2.44
TransBoostler• WMD	-0.110	0.994	0.391	4.45	-0.336★◇	0.820	0.304	2.51

Table 3: Transfer experiment results for Yeast and Twitter.

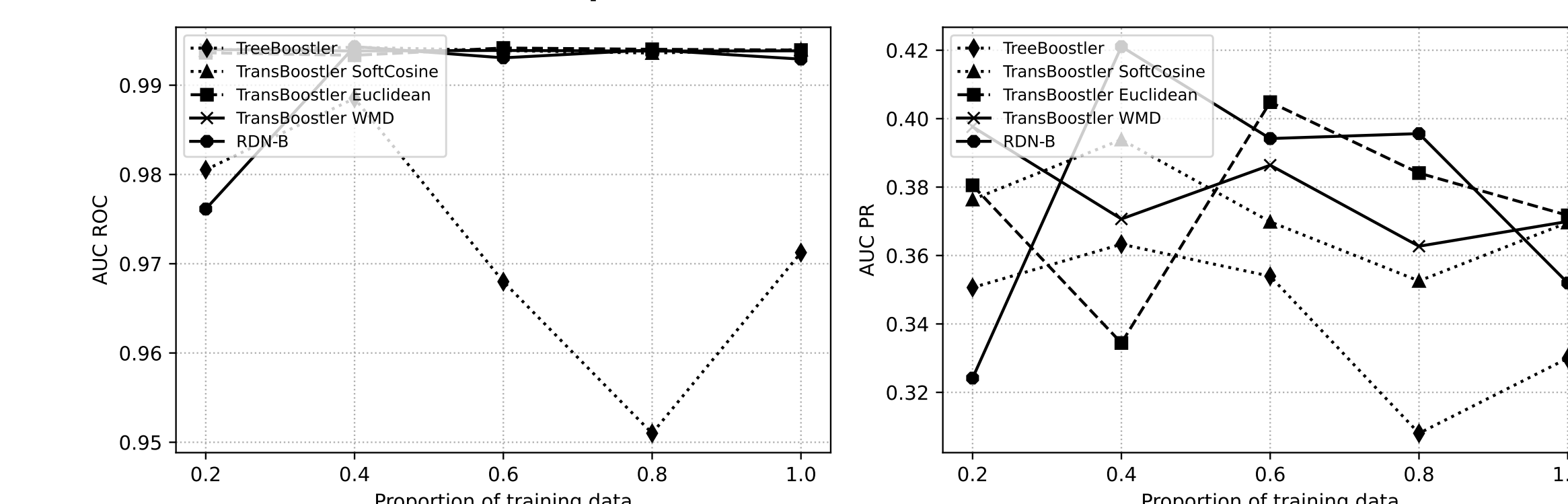


Figure 8: Learning curves for Yeast → Twitter.

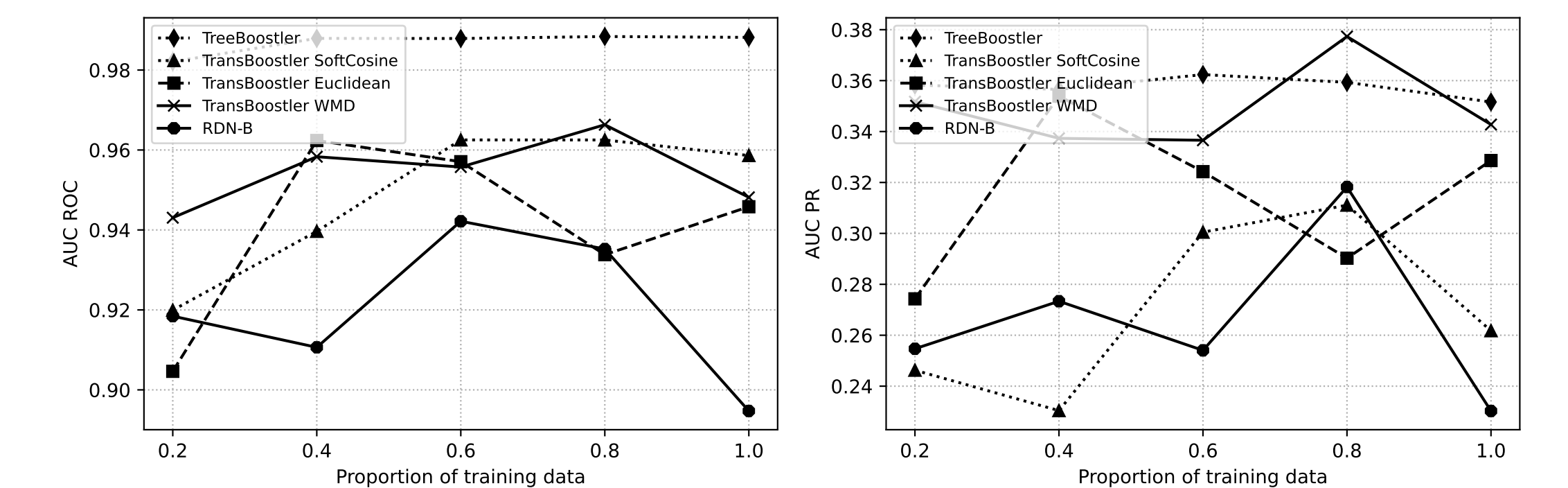


Figure 9: Learning curves for Twitter → Yeast.

IMDB → UW-CSE and NELL Sports → NELL Finances

	IMDB → UW-CSE				NELL Sports → NELL Finances			
	CLL	AUC	AUC	Run-	CLL	AUC	AUC	Run-
		ROC	PR	time(s)		ROC	PR	time(s)
RDN-B	-0.257	0.940	0.282	8.74	-0.323	0.692	0.062	24.86
TreeBoostler	-0.247	0.939	0.302	4.78	-0.165	0.980	0.071	124.59
TransBoostler Soft Cosine	-0.255	0.936	0.284	5.94	-0.321★	0.721	0.087	62.70
TransBoostler Euclidean	-0.254	0.936	0.275	6.44	-0.320★	0.750	0.069	54.04
TransBoostler WMD	-0.247	0.936	0.274	5.89	-0.324★◇	0.741◇	0.079	53.81
TreeBoostler•	-0.267	0.930	0.293	0.63	-0.315	0.979	0.068	8.85
TransBoostler• Soft Cosine	-0.385★◇	0.608	0.035	1.17	-0.366★◇	0.531	0.001	6.92
TransBoostler• Euclidean	-0.296★◇	0.906	0.131	1.53	-0.365★◇	0.558	0.002	5.88
TransBoostler• WMD	-0.288★◇	0.906	0.131	1.19	-0.365★◇	0.540	0.002	5.66

Table 4: Transfer experiment results for pairs of datasets IMDB and UW-CSE and NELL Sports and NELL Finances.

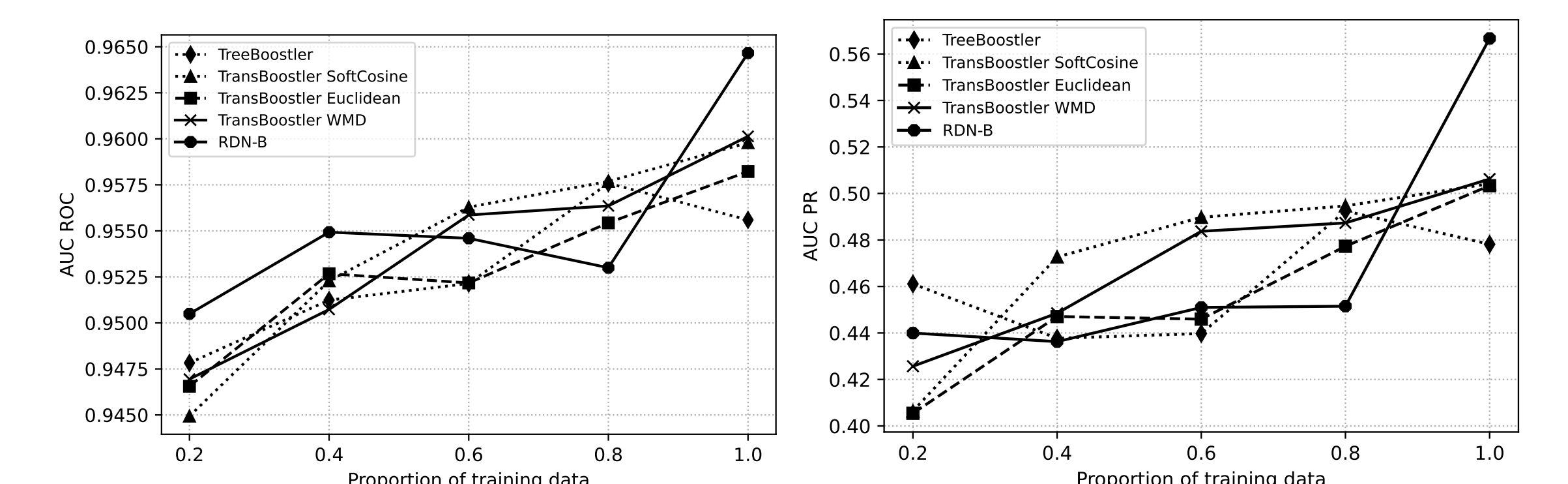


Figure 10: Learning curves for IMDB → UW-CSE.

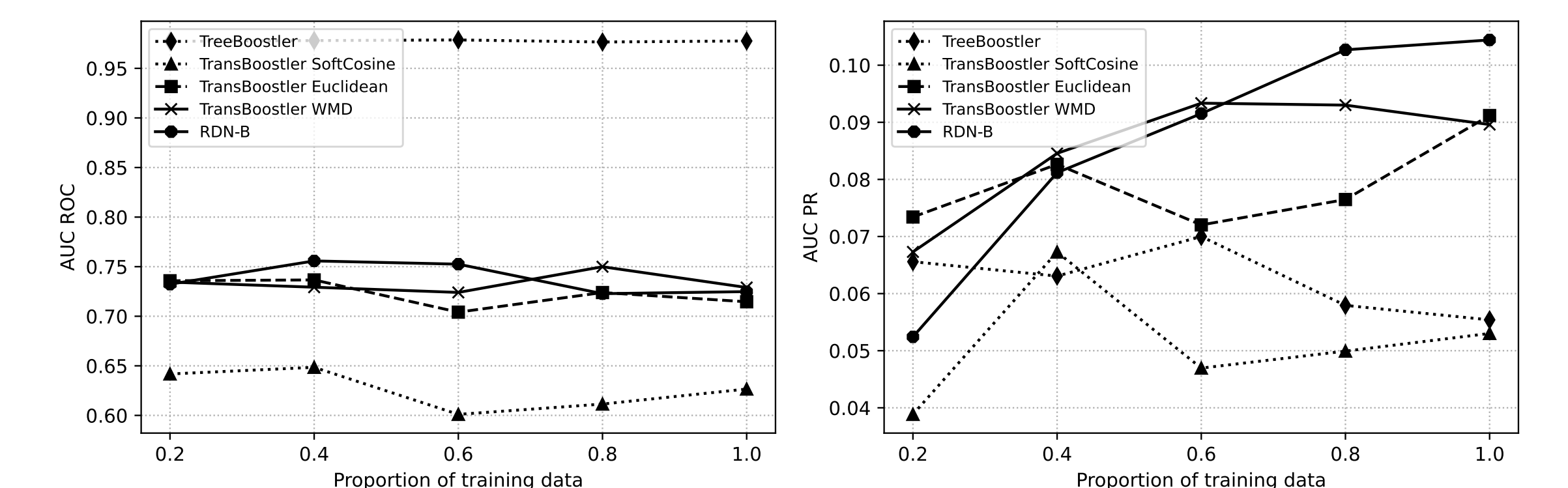


Figure 11: Learning curves for NELL Sports → NELL Finances.

Acknowledgements