Neural Unsupervised Domain Adaptation: A Survey

Alan Ramponi[1,2] and Barbara Plank[3]

[1] Department of Inf. Eng. and Computer Science, University of Trento, Italy
[2] Fondazione The Microsoft Research – University of Trento COSBI, Italy
[3] Department of Computer Science, ITU Copenhagen, Denmark

IT UNIVERSITY OF COPENHAGEN

Problem

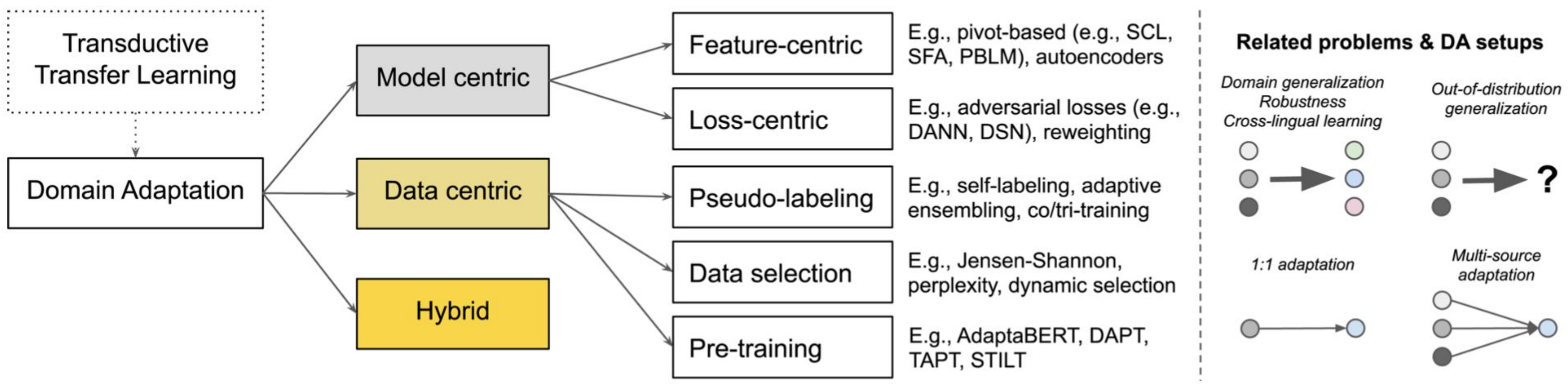
Learning under domain shift is a main challenge in NLP. We here review neural *unsupervised* domain adaptation (UDA) techniques which do not require labeled target domain data, a challenging, widely applicable setup in real-world scenarios.

From *domain* to the *variety space*

NLP is pervasively facing heterogeneity in data along many underlying (often unknown) dimensions, narrowly referred as *domains*. We suggest to use the more general term *variety*, rather than domain, which pinpoints better to the underlying linguistic differences and their implications rather than the technical assumptions.

Taxonomy

UDA is a special case of transductive transfer learning. We categorize methods into model-, data-centric and hybrid. Related problems are also discussed in the paper.



An overview of neural UDA in NLP (methods and tasks) is provided on the right.

Work	Method	SA (classif./inference)	LI	TC	NLI	POS (struct. prediction)	DEP	NER	RE
Model-centric:									
(Ziser and Reichart 2017; 2018a; 2018b; 2019)	Neural SCL	✓							
(Miller, 2019)	Neural SCL (Joint AE-SCL)	✓							
(Glorot et al., 2011)	SDA	✓							
(Chen et al., 2012)	MSDA	✓							
(Yang and Eisenstein, 2014)	MSDA					✓			
(Clinchant et al., 2016)	MSDA	✓		✓					
(Ganin et al., 2016)	DANN	✓							
(Li et al., 2017)	DANN+SCL MemNet	✓							
(Kim et al., 2017)	DANN/DSN			✓				✓	
(Sato et al., 2017)	DANN						✓		
(Wu et al., 2017)	DANN								✓
(Yasunaga et al., 2018)	DANN					✓			
(Shen et al., 2018)	DANN	✓							
(Li et al., 2018a)	DANN	✓	✓						
(Alam et al., 2018a)	DANN			✓					
(Wang et al., 2019)	DANN			✓					
(Shah et al., 2018)	DANN+Wasserstein			✓					
(Fu et al., 2017)	DANN								✓
(Rios et al., 2018)	DANN								✓
(Xu et al., 2019)	DANN			✓					
(Shi et al., 2018)	DSN (GSN)								✓
(Rocha and Lopes Cardoso, 2019)*	DANN, Shared encoders	✓			✓				
(Ghosal et al., 2020)	DANN (concept embeddings)	✓							
(Naik and Rose, 2020)	DANN (context embeddings)							✓	
Data-centric:									
(Ruder and Plank, 2018)	SSL, Multitask tri-training	✓				✓			
(Lim et al., 2020)	SSL					✓			
(Rotman and Reichart, 2019)	Deep self-training						✓		
(Han and Eisenstein, 2019)	AdaptaBERT◇					✓		✓	
(Li et al., 2019)	Adaptive pre-training						✓		
(Gururangan et al., 2020)	Adaptive pre-training (incl. multi-phase)	✓		✓	✓				✓
Hybrid:									
(Saito et al., 2017)	Asymmetric tri-training	✓							
(Desai et al., 2019)	Adaptive (temporal) ensembling			✓					
(Jia et al., 2019)	Cross-domain LM							✓	
(Cui and Bollegala, 2019)	SelfAdapt (pivots+co-training)	✓							
(Peng and Dredze, 2017)	Multi-task-DA◇					✓		✓	
(Guo et al., 2020)	DistanceNet-Bandit	✓							
(Ben-David et al., 2020)	PERL (pivots+context embeddings)	✓							

***Methods**: SCL = structural correspondence learning; AE = autoencoder; SDA = stacked denoising AE; MSDA = marginalized SDA; DANN = domain-adversarial neural network; DSN = domain separation network; GSN = genre separation network; SSL = semi-supervised learning; LM = language modeling. **Tasks**: SA = sentiment analysis; LI = language identification; TC = binary text classification (incl. machine reading, stance detection, intent classification, political data identification, etc.); NLI = natural language inference; POS = part-of-speech (incl. Chinese word segmentation); DEP = dependency parsing; NER = named entity recognition (incl. slot tagging, event trigger identification, etc.); RE = relation extraction. *with cross-lingual adaptation. ◇ applicable to UDA but main focus is supervised DA.*

Challenges and Directions

- Comprehensive UDA benchmarks (e.g., multiple tasks of increasing complexity);
- Back to the roots (e.g., revisiting classics, cf. SCL) and how knowledge transfers;
- Tackle the *X* scarcity (when unlabeled data is scarce or model training data is absent).