

Cross-domain Soft Patterns for Sentiment Analysis

Anonymous NAACL submission

Abstract

This report describes experiments

1 Introduction

2 Related Work

3 Domain Adaptation with Soft Patterns

explain training proc here

4 Experimental Setup

4.1 Dataset

We use the provided dataset, a balanced subset of the reviews data extracted by McAuley et al. (2015). The data consists of users reviews on two domains –movies and TV, and games–, extracted from Amazon. We use Movies & TV category as source domain and Games as target domain. We extract a development subset from the source domain and further divide the target domain’s data into unlabeled, development, and test splits. Table 1 presents the sizes of each split considered in the experiments.

4.2 Implementation Details

We build upon the implementation of SoPa introduced by Schwartz et al. (2018).¹ All models are implemented in PyTorch². We use pre-trained 300-dimensional GloVe 840B embeddings Pennington et al. (2014) normalized to unit length. Training was performed using Adam (Kingma and Ba, 2014) as optimizer.

For hyper-parameter tuning, we resort to a subset of the training and development source data, consisting of 10,000 and 5,000 instances, respectively. These subsets were sampled following a

uniform distribution without replacement. We use a Tree-structured Parzen Estimator (TPE) optimization model over 30 iterations³. Table 2 shows the range of hyper-parameter values explored and their optimal values.

5 Results and Discussion

References

- Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Julian McAuley, Christopher Targett, Qinfeng Shi, and Anton Van Den Hengel. 2015. Image-based recommendations on styles and substitutes. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 43–52. ACM.
- Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.
- Roy Schwartz, Sam Thomson, and Noah A Smith. 2018. Sopa: Bridging cnns, rnns, and weighted finite-state machines. In *Proceedings of ACL*.

¹https://github.com/Noahs-ARK/soft_patterns

²<https://pytorch.org/>

³We use HyperOpt library (<http://hyperopt.github.io/hyperopt/>)

Domain	Train	Dev	Test	Unlabeled
Movies & TV (src)	89,998	17,999	10,000	-
Games (tgt)	-	5,000	11,142	5,000

Table 1: Size of data splits in source (src) and target (tgt) domains.

Hyper-parameter	Range	Optimal
Patterns	{6:10, 5:10, 4:10, 3:10, 2:10}, {6:10, 5:10, 4:10}	{6:10, 5:10, 4:10}
Learning rate	10^{-9} – 10^{-2}	0.00015
Dropout	0–0.2	0.0017
MLP hid. dim.	100–300	100
Batch size	10–64	20

Table 2: Range and optimal values of hyper-parameters tuned.