Cross-domain Soft Patterns for Sentiment Analysis

Anonymous NAACL submission

Abstract

This report describes experiments

1 Introduction

The objective of domain adaptation techniques is to adapt a hypothesis trained on a source data distribution so that it can perform well on a related target distribution. These techniques have been applied to a variety of NLP tasks such as sentiment analysis (Blitzer et al., 2007; McAuley and Leskovec, 2013; McAuley et al., 2015; Ruder and Plank, 2018), style transfer in text generation (Fu et al., 2018; Yang et al., 2018; Peng et al., 2018), textual and visual question answering (Chao et al., 2018; Zhao and Liu, 2018), and machine translation (Etchegoyhen et al., 2018; Britz et al., 2017), to name a few.

In the case of sentiment analysis of online user reviews, previous work has sought effective ways of transfer learning between product categories (Blitzer et al., 2007; Ruder and Plank, 2018). However, the task has been proven to be challenging since sentiment is expressed differently in different domains. For instance, Blitzer et al. (2007) identifies three types of feature behaviour across domains: (a) features that are highly predictive in the source domain but not in the target domain, (b) features that are highly predictive in the target domain but not in the source domain, and (c) features that positively predictive in the source domain but negatively predictive in the target domain (or viceversa).

In this report, we focus on unsupervised domain adaptation for the task of sentiment analysis, transfering from a single source domain into a single target domain. We build upon the recently proposed SoPa (Schwartz et al., 2018), a neural architecture that mimic the behaviour of a Weighted Finite State Machine. SoPa is able to learn soft lexical patterns, i.e. word patterns that might include a (possibly empty) wild card. We investigate the performance of SoPa under a self-training setup following calibration procedures proposed by Ruder and Plank (2018). Experiments on Amazon online reviews of two product categires show promising results.

2 Related Work

Early work on domain adaptation for sentiment analysis, namely non-neural approaches, reports that transfering from a source domain closer to the target domain yields better performance than combining several significantly varied domains (Blitzer et al., 2007; Aue and Gamon, 2005). One identified reason is the vocabulary mismatch between domains, leading to scenarios where features drawn from one domain are not present in the other or contradict each other, as reported by Blitzer et al. (2007). In the advent of neural networks, this problem is partially addressed with continuous representation of words. A more direct approach is taken by Barnes et al. (2018) who projects embeddings from both source and target domains into a common space in an adversarial setup. Furthermore, most neural architectures proposed so far rely on pretrained word embeddings that could be considered domain-independent given the datasets these embededdings trained on (Pennington et al., 2014; Peters et al., 2018). These huge benmark datasets are meant to be as varied as possible in terms of domains, e.g. Wikipedia, CommonCrawl.

However, highly specific domains will present word types that are likely not represented in these pretrained representations. In this case, a model will rely on the embedding module's robustness to represent OOV types. In this scenario, (Schwartz et al., 2018) proposes SoPa, a model that mimics the behaviour of a Weighted Finite State Machine. The model itself can be regarded as a restricted case of a one-layer CNN that consumes the input one token at a time, like an RNN. The architecture shifts the representation robustness from the token level to the phrase level by modelling a soft version of traditional lexical patterns. The model learns to represent fixed-length patterns of words with possibly empty components. For example, a pattern can match the sequence *A B C* as well as *A * C*.

The performance of SoPa is tested by Schwartz et al. (2018) for the task of sentiment analysis in single domain scenarios. In this report, we investigate the performance of SoPa under a transfer learning scenario from one source domain (Movies & TV) to one target domain (Games).

3 Soft Patterns

A soft pattern, as introduced by Davidov et al. (2010), is a pattern that supports partial matching on a given span of text by skipping some words of the pattern. Let WFSA-ϵ be a WFSA that support ϵ transitions (a transition that skips an input word) as well as self-loops (a transition that repeats the insertion of an input word). Let WFSA-ϵ be defined by the tuple $F = \langle S, V, \pi, T, \eta \rangle$ where S is the set of states of size d, V is the vocabulary, $\pi \in \mathbb{R}^d$ is an initial weight vector, $T : (V \cup \epsilon) \rightarrow \mathbb{R}^{d \times d}$

3.1 SoPa as a WFSA

Let the WSFA with ϵ transitions, i.e. a transition that does not consume a token, be defined by the tuple

what sopa defines and does for 1 pattern

how patterns are aggregated and scoring a document

3.2 Scoring a document

4 Domain Adaptation with SoPa

explain training proc here

5 Experimental Setup

We build upon the implementation of SoPa introduced by Schwartz et al. (2018).[1] All models are implemented in PyTorch[2].

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

[2] https://pytorch.org/

5.1 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.

5.2 Training of source domain

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 tunning, 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 iterations[3]. Table 2 shows the range of hyper-paramter values explored and their optimal values.

5.3 Self-training of target domain

6 Results and Discussion

References

Anthony Aue and Michael Gamon. 2005. Customizing sentiment classifiers to new domains: A case study. In *Proceedings of recent advances in natural language processing (RANLP)*, volume 1, pages 2–1. Citeseer.

Jeremy Barnes, Roman Klinger, and Sabine Schulte im Walde. 2018. Projecting embeddings for domain adaption: Joint modeling of sentiment analysis in diverse domains. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 818–830.

John Blitzer, Mark Dredze, and Fernando Pereira. 2007. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *Proceedings of the 45th annual meeting of the association of computational linguistics*, pages 440–447.

[3] We 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.

Denny Britz, Quoc Le, and Reid Pryzant. 2017. Effective domain mixing for neural machine translation. In *Proceedings of the Second Conference on Machine Translation*, pages 118–126.

Wei-Lun Chao, Hexiang Hu, and Fei Sha. 2018. Cross-dataset adaptation for visual question answering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5716–5725.

Dmitry Davidov, Oren Tsur, and Ari Rappoport. 2010. Enhanced sentiment learning using twitter hashtags and smileys. In *Proceedings of the 23rd international conference on computational linguistics: posters*, pages 241–249. Association for Computational Linguistics.

Thierry Etchegoyhen, Anna Fernández Torné, Andoni Azpeitia, Eva Martínez Garcia, and Anna Matamala. 2018. Evaluating domain adaptation for machine translation across scenarios. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*.

Zhenxin Fu, Xiaoye Tan, Nanyun Peng, Dongyan Zhao, and Rui Yan. 2018. Style transfer in text: Exploration and evaluation. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Julian McAuley and Jure Leskovec. 2013. Hidden factors and hidden topics: understanding rating dimensions with review text. In *Proceedings of the 7th ACM conference on Recommender systems*, pages 165–172. ACM.

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.

Nanyun Peng, Marjan Ghazvininejad, Jonathan May, and Kevin Knight. 2018. Towards controllable story generation. In *Proceedings of the First Workshop on Storytelling*, pages 43–49, New Orleans, Louisiana. Association for Computational Linguistics.

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.

Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *Proc. of NAACL*.

Sebastian Ruder and Barbara Plank. 2018. Strong baselines for neural semi-supervised learning under domain shift. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1044–1054.

Roy Schwartz, Sam Thomson, and Noah A Smith. 2018. Sopa: Bridging cnns, rnns, and weighted finite-state machines. In *Proceedings of ACL*.

Zichao Yang, Zhiting Hu, Chris Dyer, Eric P Xing, and Taylor Berg-Kirkpatrick. 2018. Unsupervised Text Style Transfer using Language Models as Discriminators. In S Bengio, H Wallach, H Larochelle, K Grauman, N Cesa-Bianchi, and R Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 7287–7298. Curran Associates, Inc.

Helen Jiahe Zhao and Jiamou Liu. 2018. Finding answers from the word of god: Domain adaptation for neural networks in biblical question answering. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.