

Emotion classification using domain adaptation



**ASSIGN
BUSTER**

Emotion classification maps texts of different granularities to a set of pre-defined emotions. Emotion classification arouses interest in different areas as it performs fine-grained analysis according to the set of distinct emotions. It was applied in a variety of applications, e. g. determining popularity of products and governments (Mohammad, 2012a) and improving human-computer interaction (Velasquez, 1997).

With the increasing interest in emotion classification, requirements on different emotion models for the variety of applications rises also. There are various emotional models from psychology (Bostan and Klinger, 2018) to define a set of emotions, e. g., categorical or dimensional models. Categorical models assume a small number of distinct emotional classes, e. g., anger, disgust, fear, happiness, sadness, and surprise (Ekman, 1992). Dimensional models represent emotions in a dimensional form, such as valence, arousal and dominance (Russell and Mehrabian, 1977). The growing interest in emotion classification according to the various emotional models raises the need of labeled data in different domains. Thus, domain adaptation could be a possible solution.

Domain adaptation is the process of adapting one or more source domains providing labeled training data to the target domain on which the classifier will be tested. The training and testing data are built from different distributions. There are two main reasons why we want to apply domain adaptation: in order to cover the lack of labeled data; in order to create the robust classifier which can capture the different data distributions in and across domains. To annotate the data necessary for accurate classification for each new domain requires huge cost and effort. Emotions are very subjective, e. g., (from <https://assignbuster.com/emotion-classification-using-domain-adaptation/>

Schuff et al. (2017)) the tweet “ 2 pretty sisters are dancing with cancered kid ” was marked as fear and sadness by one annotator and with joy and sadness by another. There are no clear guides for annotation. It is well known that accuracy of the classifier will drop if the data distribution of the domain on which the classifier was trained is different than the domain on which the classifier is tested. The words used in different domains for expressing of the same emotion are quite different and vice versa the same word used in different domains may refer to different emotions.

For example, the sentence “ Thank you everyone!” (from corpus of blog posts, Aman and Szpakowicz (2007)) refers to happiness and the sentence “ Thank you @Macys! You are and will continue to be my favorite store! Bye bye #DoucheBag” (from SSEC corpus, Schuff et al. (2017)) refers to anger, anticipation, disgust, joy, surprise, trust. That is the reason why we can not apply a classifier trained on one domain directly to other domains. Thus, domain adaptation is necessary. While domain adaptation was successfully applied for different tasks, there are only few works in text-based emotion classification. Blitzer et al. (2006), Pan et al. (2010), Glorot et al. (2011) investigated the adaptation method for features, as it is more robust to domain shift. Blitzer et al. (2006), Blitzer et al. (2007) performed a set of experiments in domain adaptation for the sentiment analysis task. With the structural correspondence learning adaptation method (Blitzer et al., 2006; Blitzer et al., 2007) for automatically detection of the correspondence between features from different domain. An example from Blitzer et al. (2006): there are many words like “ excellent” or “ awful” which have the same meaning in each domain, but there are also many words which are

domain-specific like “ dual-core” for computer reviews or “ reception” for cell phone reviews. But if these domain-specific words have high correlation with the word “ excellent” and low correlation with word “ awful” than they can be aligned. The features like “ fast dual-core” or “ good-quality reception” are so-called non-pivot features and the words “ excellent” and “ awful” are pivot features. SCL first chooses a set of pivot features, e. g., bigrams which occur frequently in both domains. Then, SCL correlates between pivot and non-pivot features by training a linear predictor which is characterized by a weight vector. Positive entries in the weight vector mean that non-pivot feature is highly correlated with the corresponding pivot. Blitzer et al. (2007) applied additionally mutual information to pivot features and the source label for correction of structural correspondence misalignments. Misalignments can occur when a projection is discriminative in the source domain but not in the target. For example, the book domain is quite broad: the topics can vary between religion, politics and so on, and the kitchen domain is much narrower. They also introduced the A-distance metric to measure how good the source domain is for the adaptation. A-distance measures the divergence of two domains after the SCL projection.

We aim at contributing to this topic and want to perform feature-based domain adaptation for emotion classification task. We hypothesize that ... domain adaptation helps to improve the performance of a model trained on another domain than the target domain, as it helps to cover the lack of labeled data and to capture the different data distributions in and across domains. and arrived at the conclusion that: “ Adding more labeled data always helps, but diversifying training data does not”.

In addition, with model introspection tools (Ribeiro et al., 2016, Ribeiro et al. (2018)) based on features we will detect the similarities and dissimilarities between different data sets. These introspection results are helpful in identifying which domain would be a good proxy for all other domains and should be used for training. Also in determining improvement steps of applied adaptation method, this knowledge could be helpful, as we can understand the reasons behind predictions.