Studying the feasibility of Domain Adaptation for Emerging Yelp Business Domains

Rafi Trad - Ali Hashaam - Imad Hajjar

Supervised by: M.Sc. Marcus Thiel

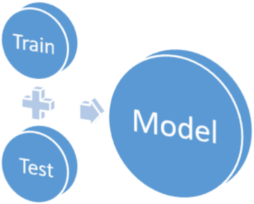
Scientific Project: Data and Knowledge Engineering

April 12, 2018

Table of Contents

Motivation

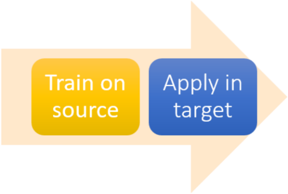
For a domain of interest $\mathcal{D}$:



- Scarce labelled data in $\mathcal{D}$ $\Rightarrow$ *Domain Adaptation* (transfer learning)

Domain Adaptation Intuition

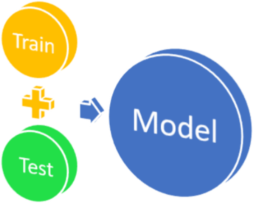
Resort to using $\mathcal{D}_S$ labelled data (abundant) to address the lack of data in $\mathcal{D}_T$



- *Domain adaptation* can be defined as a special case of the *transductive transfer learning* where the feature spaces between domains $\mathcal{X}$ are similar; viz.: $\mathcal{T}_S = \mathcal{T}_T \wedge \mathcal{D}_S \neq \mathcal{D}_T \wedge \mathcal{X}_S = \mathcal{X}_T$

BUT..

The *Discriminative Learning Methods' assumption* does not hold any more..



- Train data and Test data no longer conform to the same distribution!
- Acute effects on models' performance → A critical challenge

Methodology

Domain Adaptation can be performed in many ways

Feature Space Transformation

Prior Based Adaptation

Instance Selection and Weighting

Feature space transformation - generalisable feature selection:

- Set of $\mathcal{D}_S$ $\{d_k\}_{k=1}^{K}$ $\rightarrow$ train a model for a $\mathcal{D}_T$
- $\mathcal{X}_S \cap \mathcal{X}_T \neq \phi$

In our setting: one $\mathcal{D}_S \rightarrow$ two $\mathcal{D}_T$, one of which $\mathcal{X}_S \cap \mathcal{X}_T \neq \phi$, and the other $\mathcal{X}_S \cap \mathcal{X}_T \approx \phi$

General Outline

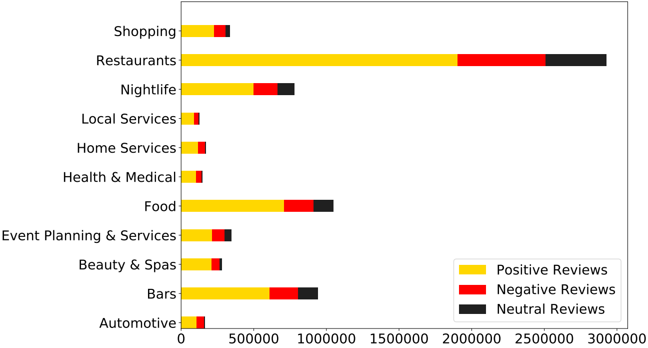
$\mathcal{T}$ = Sentiment Polarity Detection: Affection (negative or positive) towards the discussed aspects in textual inputs.

- **Domain Selection**
  - Select $\mathcal{D}_S$
  - Select the two $\mathcal{D}_T$
- **Domain Adaptation**
  - Train ξ_{ST} with labelled $\mathcal{X}_S$
  - Apply ξ_{ST} in $\mathcal{D}_T$
  - Evaluate

To what extent is the domain adaptation for the sake of sentiment polarity detection profitable for new emerging domains?

Dataset

- Yelp dataset (round 10).
- 7.27M reviews
- 11 businesses



Selecting $\mathcal{D}_S$ and $\mathcal{D}_T$

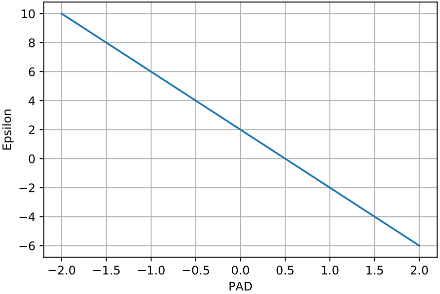
$\mathcal{D}_S$ was selected so that $|\mathcal{X}_S|$ is maximal $\Rightarrow$ Restaurants.

For $\mathcal{D}_T$: according to **similarity** to $\mathcal{D}_S$
Similar domains $\rightarrow$ to distinguish between them is difficult (proxy $\mathcal{A}$-distance, -PAD or $\hat{d}_A$-):

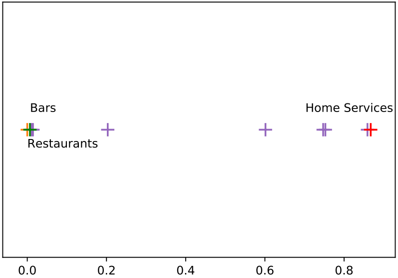
$$\hat{d}_A = 2(1 - 2\epsilon) \ .$$

- Linear bag-of-words SVM classifier was used (its error is ϵ)
- The most similar domain to Restaurants was Bars
- The least similar was Home Services

Selecting $\mathcal{D}_T$



(a)



(b)

Preprocessing

Two preprocessing pipelines:

- Preprocessing Pipeline for Baseline
- Preprocessing Pipeline for Gold Standard

Preprocessing - Baseline

For the baseline, we have applied standard preprocessing steps in following sequence:

- Dropping records with null values
- Replacing new line characters, slashes (/) and punctuation
- Case-folding
- Stopword Removal and Tokanisation
- Stemming

Preprocessing - Gold Standard (1)

Preprocessing steps, adopted in following sequence:

- Dropping records with null values (2 records)
- Replacing foreign accents with most likely letters by using Python's Unicode library, which provides the ASCII of transliterations Unicode words
- Case-folding
- Removing newlines, tabs, replacing " with '

Preprocessing - Gold Standard (2)

- Expanding the abbreviations before removing the punctuation, in order to pay attention to the linguistic abbreviations.
- Removing punctuations
- Tokenisation
- Adding PoS tags to tokens
- Lemmatising PoS tagged tokens

Preprocessing - Gold Standard (Feature Engineering)

- Detecting multi-word phrases inside a sentence (converting the common bi-grams into a single word by appending underscore between them).
- Feature Selection using ANalysis Of VAriance (ANOVA)

Modelling

- ξ_b: Baseline Model
- ξ: Advanced Model
- S: Source Domain
- T: Target Domain
- 3 or 5: The number of labels

Table 1: The Models required for our Experiments.

	Used In-Domain	Used Cross-Domain
Baseline	ξ_{bS3}, ξ_{bS5}, ξ_{bT}	ξ_{bST}
Advanced	ξ_T	ξ_{ST}

Models Used

- Multinomial Naive Bayes (MNB)

$$s_{map} = \arg\max_{s \in S} P(s|r) = \arg\max_{s \in S} P(s) \prod_{t \in r} P(t|s) \quad .$$

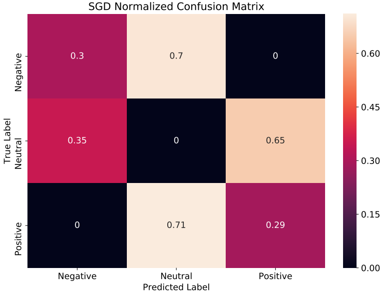
- Stochastic Gradient Descend (SGD)

$$E(w, b) = \frac{1}{n} \sum_{i=1}^{n} L(y_i, f(x_i)) + \alpha R(w) \quad .$$

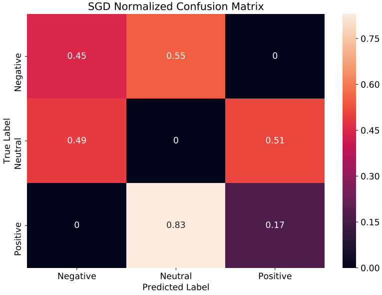
Modelling Steps

- Training set 70%, Test set 30%
- Parameter Optimization: Grid Search with Cross-validation
- 5-fold Cross-validation during training
- Determining neutral reviews:
    - MNB: when probability $\in]20\%, 80\%[$
    - SGD: utilizing Hinge Loss function

Confusion Matrices of ξ_{ST}



(a) Bars



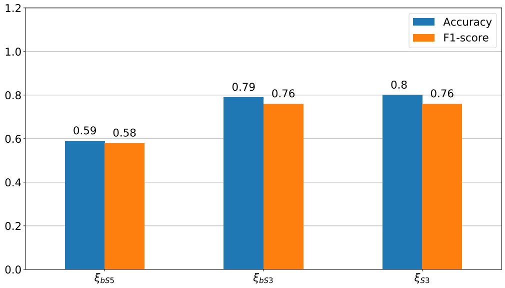
(b) Home Services

In-domain Evaluation Criteria

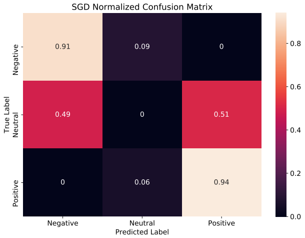
- Accuracy
- F1 score
- Cohen's Kappa

$$\kappa = \frac{p_o - p_e}{1 - p_e} \quad .$$

In-domain Evaluation Results



(a)



(b)

In-domain Evaluation Results

$$\kappa = \frac{p_o - p_e}{1 - p_e} \quad .$$

Most Frequent	Stratified
0	-0.02

Cross-domain Evaluation Criteria

- Transfer Loss:

$$Loss(S,T) = e(S,T) - e_b(T,T) = e(\xi_{ST}) - e(\xi_{bT}) \; .$$

- Adaptation Loss:

$$Adaptation\ Loss \; = acc(\xi_T) - acc(\xi_{ST}) \; .$$

- Relative Reduction of Error:

$$Relative\ Reduction\ of\ Error = \frac{acc(\xi_{ST}) - acc(\xi_{bST})}{acc(\xi_T) - acc(\xi_{bST})} \; .$$

Cross-domain Evaluation Criteria cont.

- McNemar's Test:

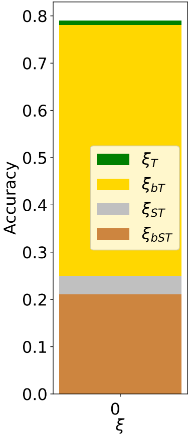
$$P = 2\sum_{m=n_{10}}^{k} \binom{k}{m} \left(\frac{1}{2}\right)^k : n_{10} > \frac{k}{2} \quad .$$

$$P = 2\sum_{m=0}^{n_{10}} \binom{k}{m} \left(\frac{1}{2}\right)^k : n_{10} < \frac{k}{2} \quad .$$

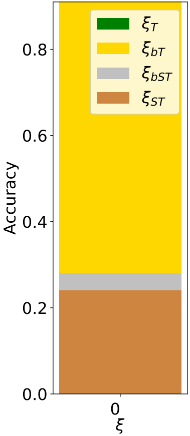
	Correct	Incorrect
Correct	N_{00}	N_{01}
Incorrect	N_{10}	N_{11}

$$k = N_{10} + N_{01}$$

Cross-domain Evaluation Results

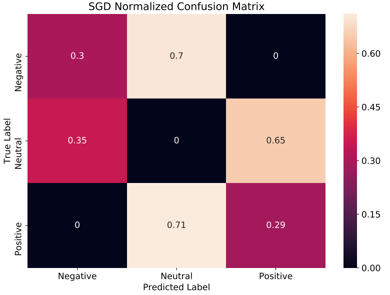


(c) Bars

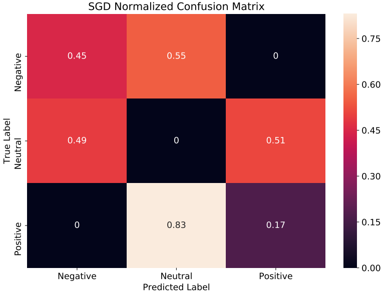


(d) Home Services

Cross-domain Evaluation Results cont.

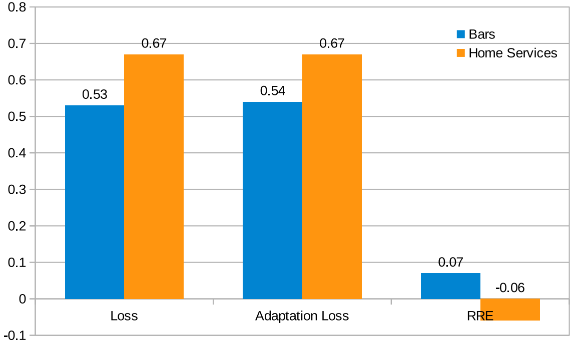


(a) Bars



(b) Home Services

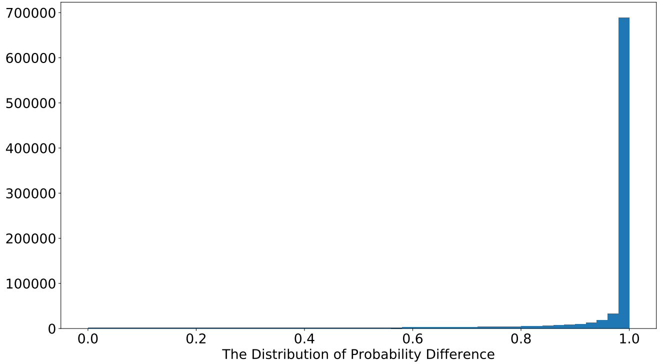
Cross-domain Evaluation Results cont.



Differences in classifiers' performance were *significant*, according to McNemar's test.

Thank You

Appendix - Determining MNB Neutral Reviews



Appendix - Determining SGD Neutral Reviews