A Million Tweets Are Worth a Few Points: Tuning Transformers for Customer Service Tasks

A. Hadifar, S. Labat, V. Hoste, C. Develder, T. Demeester
amir.hadifar@ugent.be
https://ugentt2k.github.io

Introduction

Customer service: all interactions of a company with its current and potential clients

How NLP can Help?

i. Insights to optimize customer interactions, e.g., customer satisfaction or sentiment analysis
ii. Facilitating real time decisions, e.g., categorizing or prioritizing customer tickets

Idea

Challenges: Small and medium size companies struggle with applying recent NLP techniques due to the limited size, noise and imbalance in their data

Contributions:

i. Multilingual conversational Twitter corpus
ii. Multilingual progressive tuning
iii. Demonstration of benefit of domain-specific representation from (i) + (ii) over generic LM models in various tasks

Transfer strategies

Transfer learning: Leveraging knowledge from a pre-trained model

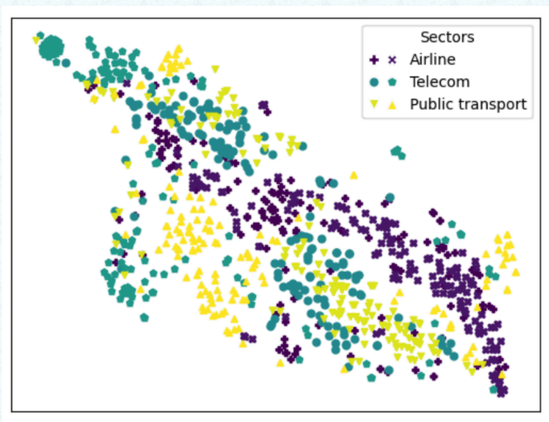
i. Freezing (extracting features): Use representations learned by a pre-trained model to extract meaningful features
ii. Finetuning: reuse a pretrained model on another task by adjusting its parameters.

Progressive tuning[1]: Further training of a generic pretrained model on a smaller and less diverse domain

Twitter Collection

Language	|convs|	|tweets|
English	135.1k	406.3k
French	60.9k	212.9k
Dutch	45.6k	141.0k
German	33.0k	104.5k
All	274.6k	864.7k

Number of collected conversations and tweets for each language



2D visualization of randomly sampled tweets for six companies, in different sectors and languages

Results

Language transferability: We evaluate progressive tuning on 6 tasks, spanning 4 languages

i. Generally, finetuning performs better than freezing (e.g., XLM→∅ vs. XLM→φ)
ii. Finetuned, domain specific model outperforms other models (e.g., BERTweet→φ vs. RoBERTa →φ)
iii. Classical models are strong competitors for the frozen transformers (e.g., SVM vs. BERTweet→∅)
iv. Further pretraining, followed by finetuning is beneficial (e.g., XLM→φ vs. XLM→π→φ)

	Complaint-2 (English)		Complaint-R (French)		Churn (English)		Subjectivity (Dutch)		Relevance (German)		Polarity (German)	
Model	**ACC**	**F1**	**ACC**	**F1**	**ACC**	**F1**	**ACC**	**F1**	**F1$_{syn.}$**	**F1$_{dia.}$**	**F1$_{syn.}$**	**F1$_{dia.}$**
Majority-class	72.4	42.0	56.0	35.8	78.4	43.9	55.0	35.5	81.6	83.9	65.6	67.2
LR (tf-idf)	83.5	77.0	57.5	57.4	85.1	71.7	71.6	70.9	88.4	87.7	71.1	70.4
SVM (tf-idf)	84.4	80.2	59.0	58.8	87.3	80.1	71.7	71.0	90.4	88.8	74.8	72.8
Reference	82.0[1]	62.7[1]	-	-	-	78.3[2]	-	-	85.2[3]	86.8[3]	66.7[3]	69.4[3]
BERTweet → ∅	80.5	71.6	-	-	79.3	55.2	-	-	-	-	-	-
BERTweet → φ	**90.0**	**86.1**	-	-	**93.0**	**90.0**	-	-	-	-	-	-
RoBERTa → ∅	77.9	74.5	-	-	78.3	59.7	-	-	-	-	-	-
RoBERTa → φ	87.5	85.1	-	-	88.4	84.8	-	-	-	-	-	-
XLM → ∅	76.2	61.6	44.0	30.5	61.6	55.7	63.8	62.4	83.1	84.7	64.5	66.8
XLM → φ	85.4	83.4	54.0	46.2	84.1	75.3	73.4	72.9	91.6	91.7	76.1	73.5
XLM → π→ ∅	81.8	76.8	56.5	54.1	79.7	66.0	71.6	71.1	84.4	85.3	65.1	68.0
XLM → π→ φ	86.9	82.7	**62.0**	**61.9**	87.8	83.7	**74.6**	**74.2**	**92.7**	**92.5**	**78.7**	**76.1**

XLM→π→φ means that the XLM model is further pretrained on the our twitter corpus, followed by finetuning on the end task

Task transferability: We evaluate progressive tuning on single task, spanning 2 related tasks

i. Further tuning on related tasks requires precise calibration (XLM→ π→φ vs. SVM)
ii. Progressive tuning slightly improves the performance (XLM→ π→φ vs. XLM→φ)

Test Dataset	**Complaint-9**		
Train Dataset	**C-9**	**C-2 & C-9**	**S & C-9**
Majority-class	39.1	-	-
SVM (tf-idf)	78.6	-	-
Preotiuc-Pietro et al.	79.0	-	-
XLM → φ	78.6	79.3	80.1
XLM → π→ φ	82.4	80.0	82.8

Conclusions

i. Further training on a moderately sized multilingual in-domain corpus is an effective way to improve generic multilingual language models
ii. The performance gain versus operational cost of frozen transformers should be taken into account

References

[1] Gururangan, Suchin, et al. 2020. *Don't Stop Pretraining: Adapt Language Models to Domains and Tasks*. In Proceedings of ACL.
[2] Code and dataset available at: https://github.com/hadifar/customerservicetasks