

UNDERSTANDING MULTI-TASK LEARNING WITH TRANSFER LEARNING

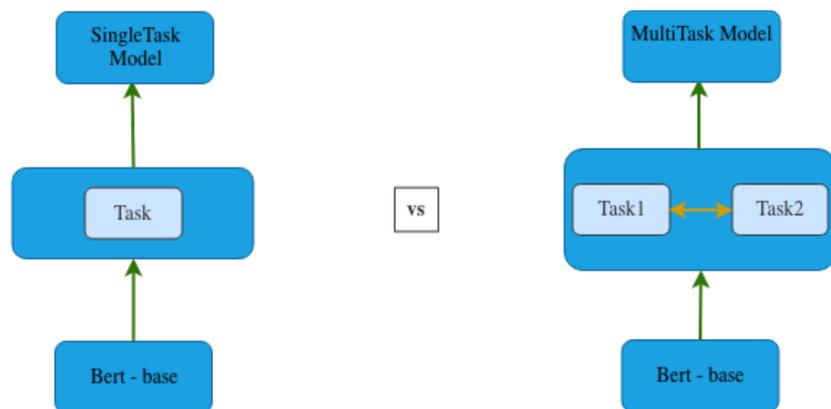
Abhishek Sharma
sharma21@stanford.edu

Introduction

Multitask Learning aims at improving generalization by learning multiple tasks at the same time. Idea behind it is that similar related tasks, aids in the learning of each other. This study focuses on the hypothesis whether training related tasks jointly, improves the performance of the model when exposed to transfer learning .

In particular, study focused on Question Answering (QA), Natural Language Inference (NLI) and Sentiment Analysis (SA) tasks and tried to understand the relationships between each of these by analyzing:

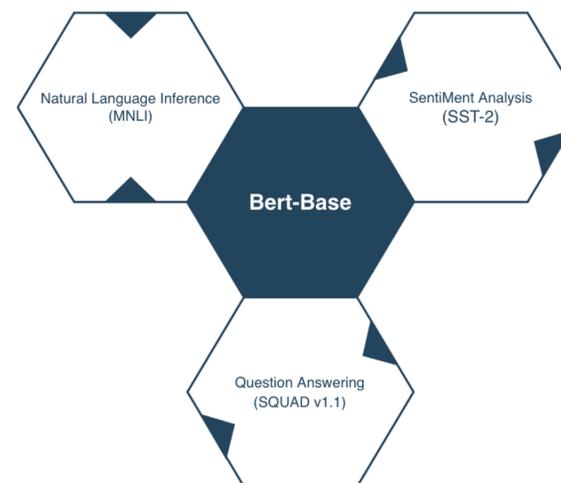
- i) the performance of the jointly trained model for the tasks in the same domain (for instance, SA and NLI both are classification tasks) on a pretrained model.
- ii) performance of the jointly trained model for the tasks in different domain (QA is span extraction problem which was trained with two different classification tasks) on a pretrained model.



Approach

- Same domain tasks:
 - joint model (MNLI, SST2) was trained by fine tuning Bert base uncased [1] model simultaneously by using two different classifier layers at the end for each of the task.
 - performance was then compared with the single task models: (MNLI, SST2) vs MNLI vs SST2
- Cross domain tasks:
 - joint models (QA + SST2) and (MNLI + QA), were trained alternately for each of the tasks on 50% of data on a pretrained Bert base uncased [1] model.
 - performance was then compared against single task models:
 - (QA + SST2) vs QA vs SST2 and (MNLI + QA) vs MNLI vs QA

Data/ Tasks



- Question Answering:
 - Model is given a question and a context and it returns the span (start and end) of the answer in the context.
 - Dataset: SQUAD v1.1 [2]
- Sentiment Analysis:
 - Find the polarity of the sentence.
 - Dataset: SST-2 [3] , binary classification
- Natural Language Inference:
 - Given a premise and a hypothesis, predict whether premise is contradicts, entails or is neutral wrt hypothesis.
 - Dataset: Multi-Genre Natural Language Inference

Results

Model	EM Score
SST-2	90.825
SST-2 + MNLI	93.577
SST-2 + SQUAD	92.087

Model	F1 Score
SQUAD	87.398
SQUAD + SST-2	88.135

Model	EM Score
MNLI	83.529
MNLI + SST-2	84.264
MNLI + SQUAD	83.311

Sentiment Analysis showed improved results when trained jointly with other tasks

Question Answering shows improvement when trained with SST-2

MNLI didn't show improvement when with QA but the score improved when trained with SST-2

Conclusion

- Multitask models for tasks in different domain showed some improvement (two out of three) whereas all multitask models for the same domain tasks showed great improvement that proving our hypothesis that training similar tasks jointly does aids to learning.

References

- [1] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- [2] Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P. (2016). Squad: 100,000+ questions for machine comprehension of text. *arXiv preprint arXiv:1606.05250*.
- [3] Nangia, N., Williams, A., Lazaridou, A., & Bowman, S. R. (2017). The repeval 2017 shared task: Multi-genre natural language inference with sentence representations. *arXiv preprint arXiv:1707.08172*.
- [4] Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A., & Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment