Framework for evaluating relative importance of tasks for cascaded Multi-task learning

Anhad Mohananey
anhad@nyu.edu

Pramit Mallick
pm2758@nyu.edu

Srinidhi Goud
sgm400@nyu.edu

Hypothesis

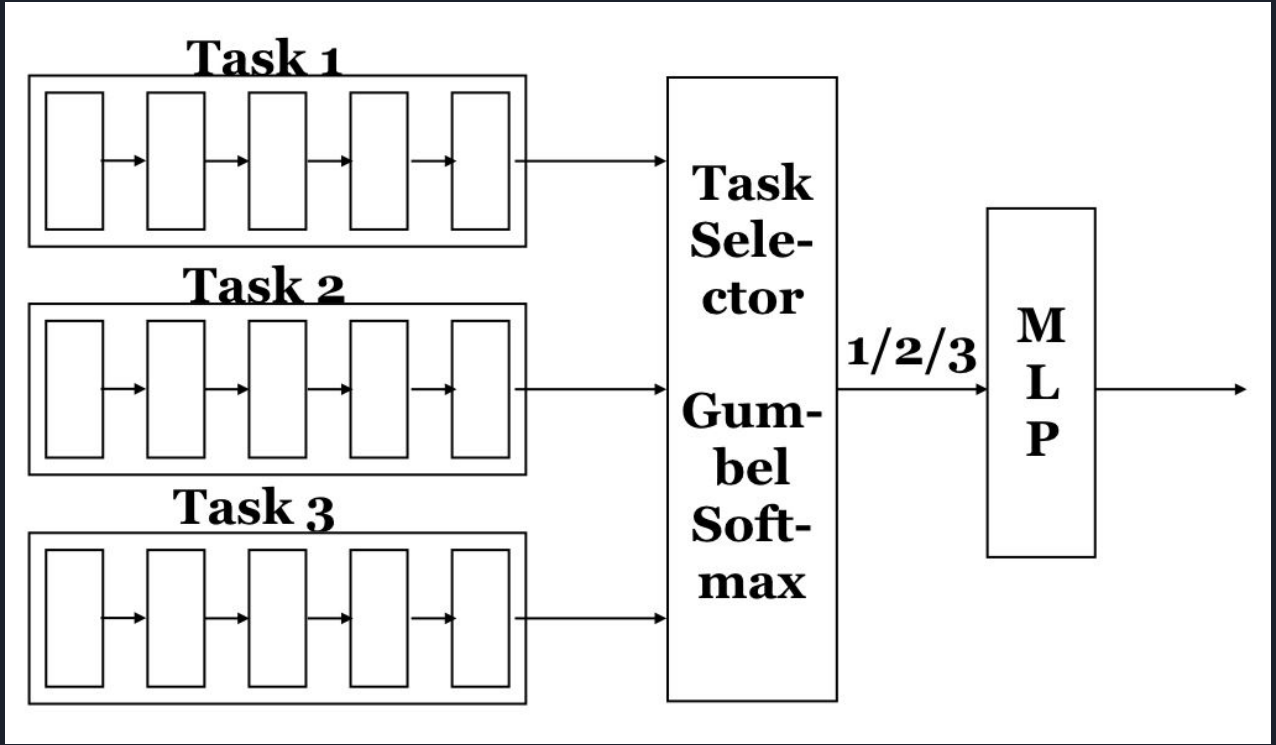
This project is about identifying potential hierarchy in the training of a sentence representation using already trained encoders.

Cascaded MTL generalizes the model while choosing certain hypothesis over others, we suspect there is an order in which the inductive bias can be introduced that might cause the model to converge better.

Our proposed framework, decides the order of importance in tasks, by calculating the number of times a task is chosen over other tasks while training to achieve the new objective.

This requires multiple pre-trained encoders connected to a Task Selector which trains the subsequent MLP layers over a specific chosen encoder for that point of time. Task Selector is also a trainable layer.

Model Architecture



Gumbel Softmax

Discrete decision made for task selection: normal softmax NOT differentiable.

A version of softmax that makes a continuous approximation of a one-hot vector: Gumbel Softmax.

The continuous approximation makes it differentiable, so now backpropagation works!

Gumbel Softmax: some details?

$$y^i = \frac{exp((log(k^i)+g^i)/temp)}{\sum exp((log(k^i)+g^i)/temp)}$$

$$g^i = -log(-log(u^i))$$

$$u^i = Uniform(0,1)$$

g(i): gumbel noise, k(i): unnormalized log probs, temp: temperature parameter

A *temp* value closer to zero: Similar to a one hot encoding.

Based on y(i):argmax for discrete decision on forward, but use the y(i) to calculate gradients for backpropagation!

Experiments and Results (so far..)

Gumbel				
Target	Accuracy	Quora (source)	NLI (source)	SST (source)
SST	69.2	40.4	59.6	
NLI	65.1	95.71		4.29
Quora	81.6		94.875	5.125

Experiments and Results (so far..)

Naive Task Selection			
Source\Target	Quora	SST	NLI
Quora	86.4	49.5	53.1
SST	61.5	79.9	61.5
NLI	83.9	64.1	75.7

Observations

From the completed experiments, amongst the three tasks, it seems like the NLI task is the best task to pre-train sentence encodings for MTL.

References

[1] Chen, Qian, et al. "Enhanced lstm for natural language inference." Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Vol. 1. 2017.
[2] Bingel, Joachim, and Anders Søgaard. "Identifying beneficial task relations for multi-task learning in deep neural networks." arXiv preprint arXiv:1702.08303 (2017).
[3] Conneau, Alexis, et al. "Supervised learning of universal sentence representations from natural language inference data." arXiv preprint arXiv:1705.02364 (2017).
[4] Ruder, Sebastian. "An overview of multi-task learning in deep neural networks." arXiv preprint arXiv:1706.05098 (2017).
[5] Meyerson, Elliot, and Risto Miikkulainen. "Beyond Shared Hierarchies: Deep Multitask Learning through Soft Layer Ordering." arXiv preprint arXiv:1711.00108 (2017).
[6] Han, Lei, and Yu Zhang. "Learning Multi-Level Task Groups in Multi-Task Learning." AAAI. 2015.