Learning to Selectively Transfer: Reinforced Transfer Learning for Deep Text Matching

Chen Qu¹, Feng Ji², Minghui Qiu²¹, Liu Yang¹, Zhiyu Min³, Haiqing Chen², Jun Huang², W. Bruce Croft¹
¹ University of Massachusetts Amherst, ² Alibaba Group, ³ Carnegie Mellon University

Overview

Research Problem: Data selection for DNN based supervised transfer learning in a deep text matching setting.

Contributions:
- We propose a reinforcement learning based data selector to select high-quality source data to help the DNN based transfer learning model.
- In contrast to do data selection instance by instance, we propose a batch based strategy to sample the actions in order to improve the model training efficiency.
- We perform thorough experimental evaluation on PI and NLI tasks that involves four benchmark datasets. We find that the proposed reinforced data selector can effectively improve the performance of the TL model and outperform several existing baseline methods.
- We use Wasserstein distance to interpret the model performance.

An Example of Negative Transfer in PI

Domain Sentences
Source (Open) Which answers does Quora show first for each question? How does Quora decide the order of the answers to a question?
Target (E-Com) How can I get an order receipt or invoice? How do I get an invoice to pay?

Model Details

Base Model: Decomposable Attention Model.
Transfer Learning Model: leveraging a large amount of source domain data in a multi-task learning manner.

Reinforced Data Selector: handling source domain data selection and maximize the effectiveness of the TL model.

State:
- (1) A hidden representation by the shared encoder.
- (2) Train loss on src model. (3) Test loss on the tgt model.
- (4) Pred probs on src model. (5) Pred probs on tgt model.

Action: denoted as \( a_i \in \{0, 1\} \), which indicates whether to drop or keep \( (X_1(i), X_2(i)) \) from the source batch.

Reward: the selected src batch \( X_b^* \) is used to update the src model and get a reward \( r_b \) (Acc on tgt val data). We consider the future discounted reward: \( r_b' = \sum_{k=0}^{\infty} \gamma^k r_{b+k} \).