UMassAmherst
College of Information & Computer Sciences
Center for Intelligent Information Retrieval

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

Alibaba Group 阿里巴巴集团

Chen Qu[1], Feng Ji[2], Minghui Qiu[2‡], Liu Yang[1], Zhiyu Min[3], Haiqing Chen[2], Jun Huang[2], W. Bruce Croft[1]
[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?
	What **order** should the Matrix movies be watched in Is there any particular **order** in which I should watch the movies
Target (E-Com)	How can i get an **order** receipt or invoice? How do I get an invoice to pay?
	I need to understand why my **orders** have been cancelled Why my **order** have been closed?

Task Definition

Text Matching: Given two sentences and predict a binary label to show whether they are semantically related.

Transfer Learning: The Source and target tasks are the same while their domains are different.

Data Selection: Under a DNN based transfer learning setting. The data selection module intervenes before each source batch update and make selections.

Our Approach: Reinforced Transfer Learning

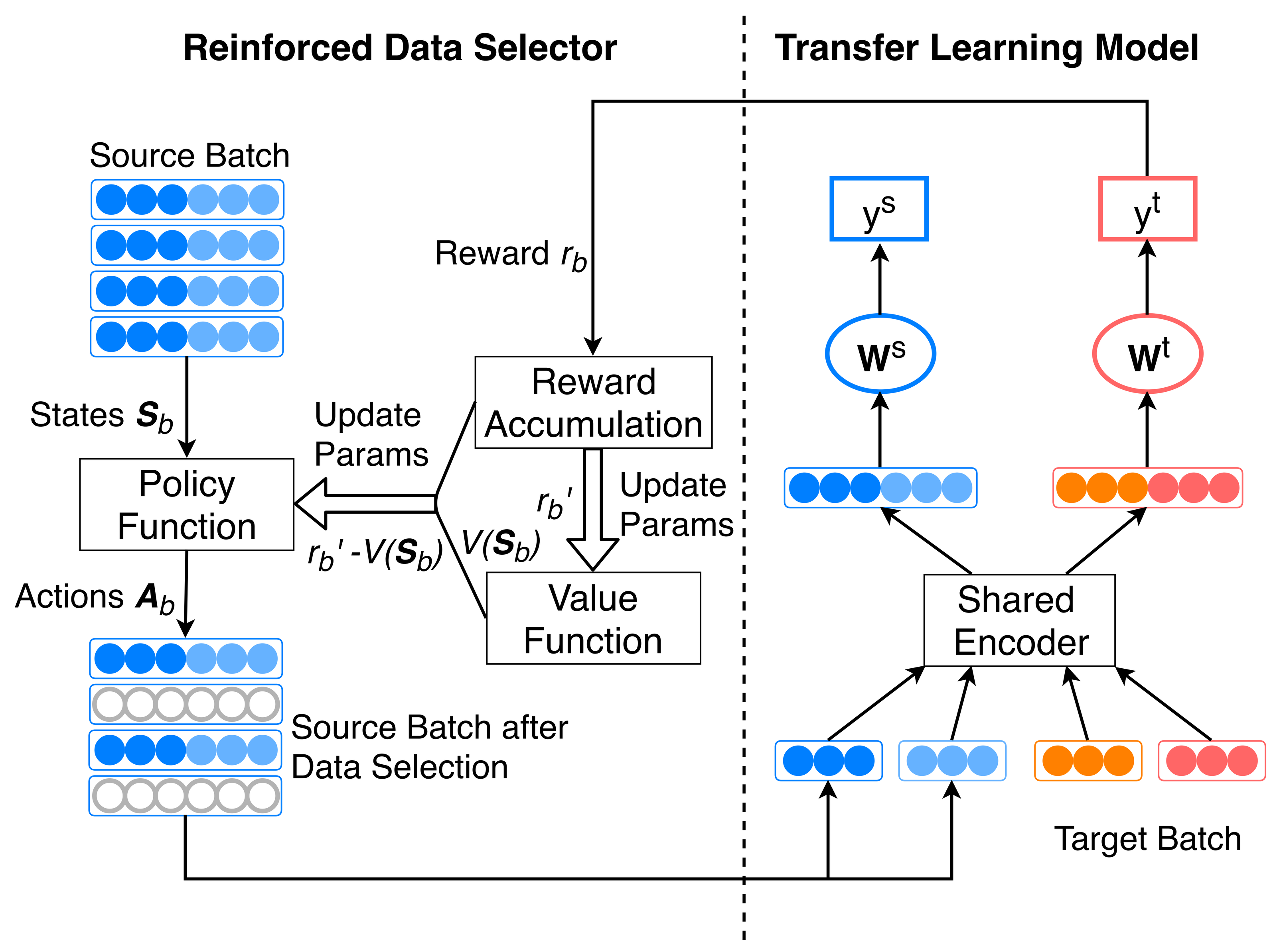

Figure: Architecture of the proposed RTL framework, which consists of two major parts: a reinforced data selector and a TL model. The "Shared Encoder" refers to the base model embedded in the TL model. The reinforced data selector selects a part of the source batch (blue) and feeds them into the TL model at each iteration. The TL model generates a reward on the target domain validation data for the data selector. Target batches (orange/pink) are fed into the TL model without data selection.

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 effectives 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 $(\mathrm{X}_1(i), \mathrm{X}_2(i))$ from the source batch.

Reward: The selected src batch $\mathcal{X}_b^{s'}$ is used to update the src model and get a reward r_b (Acc on tgt val data). We consider the future discounted reward. $r_b' = \sum_{k=0}^{N-b} \gamma^k r_{b+k}$

Optimization:

Policy network: $\Theta \leftarrow \Theta + \alpha \frac{1}{n} \sum_{i=1}^{n} v_i \nabla_\Theta \log \pi_\Theta(\mathrm{S}_i)$

Target: $v_i = r_b' - V_\Omega(\mathrm{S}_i)$

Value network: $\Omega \leftarrow \Omega + \alpha \frac{1}{n} \sum_{i=1}^{n} \nabla_\Omega \mathrm{MSE}(r_b', V_\Omega(\mathrm{S}_i))$

Policy function: π parameterized by Θ. Value function: V parameterized by Ω. Learning rate: α. Batch size: n. Target: v_i. Estimated future total reward: $V_\Omega(\mathrm{S}_i)$. State: S_i

Experiments

Datasets:

Paraphrase Identification:
Quora Question Pairs (open domain) → AnalytiCup (E-commerce)

Natural Language Inference:
MultiNLI (open domain) → SciTail (science)

Results:

Table: Testing performance in the target domain for PI and NLI

Methods	PI		NLI	
	Acc	AUC	Acc	AUC
Base Model	0.8393	0.8548	0.7300	0.7663
Transfer Learning Model	0.8488	0.8706	0.7453	0.8044
Ruder and Plank	0.8458	0.8680	0.7521	0.8062
RTL	**0.8616**‡	**0.8829**	**0.7672**‡	**0.8163**

Performance Interpretation

Method: **Wasserstein distance** measures the distance between two probability distributions. We compute this metric between the **term distributions** of the target domain and the source domains.

Results:

Table: The Wasserstein distances between the term distributions of different domains.

Name	Domains in Comparison	PI	NLI
D_{origin}	Target ↔ Source	5.250E-06	3.256E-06
D_{select}	Target ↔ Source (Selected)	4.963E-06	3.190E-06
D_{drop}	Target ↔ Source (Dropped)	5.320E-06	3.290E-06
D_{rand}	Target ↔ Source (Random)	5.232E-06	3.243E-06

Observations:

- $D_{rand} \approx D_{origin}$: random selection only influences the term distribution slightly. This sets a baseline for other distances.
- $D_{select} < D_{origin}$: the source domain data selected by the reinforced data selector is closer to the target domain data.
- $D_{drop} > D_{origin}$: the source domain data dropped by the reinforced data selector is not very similar to the target domain data.

Chen Qu, Feng Ji, Minghui Qiu, Liu Yang, Zhiyu Min, Haiqing Chen, Jun Huang, W. Bruce Croft. Learning to Selectively Transfer: Reinforced Transfer Learning for Deep Text Matching. In Proceedings of 12th ACM International Conference on Web Search and Data Mining. Melbourne, Australia. February 11-15, 2019. ‡ Corresponding author.