Robust Domain Adaptation by Adversarial Training and Classification

Nicolas Aagnes
naagnes@stanford.edu
Elliot Epstein
epsteine@stanford.edu

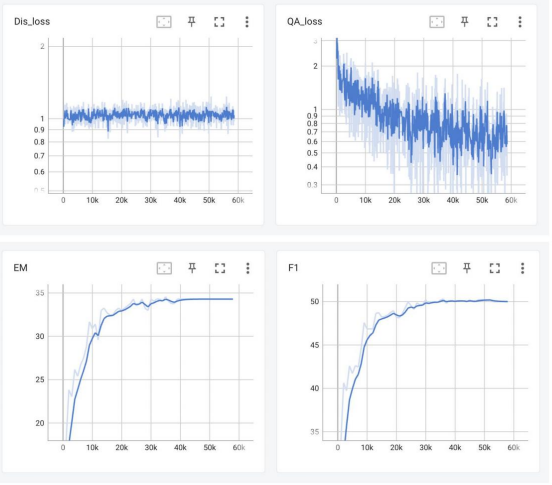
Introduction

- Common NLP benchmarks often neglect performance on out-of-distribution data, which can become problematic as inference data can often be different than training data.
- We extend a method proposed by Lee et al. whose novel idea is to jointly train a question-answering (QA) model together with a discriminative model that forces the encoder to learn domain-agnostic embeddings
- We include an additional classifier module to handle datasets that include questions without any answer.
- We also use enrich the the training procedure with data augmentation techniques such as backtranslation.

Experiments

Dataset	Train	Dev	Test
in-domain dataset			
SQuAD	50000	10507	-
NewsQA	50000	4212	-
Natural Questions	50000	12836	-
out-domain datasets			
DuoRC	127	126	1248
DuoRC BT x 10	127x11	126	-
RACE	127	126	419
RACE BT x 10	127x11	126	-
RelationExtraction	127	128	2693
RelationExtraction BT x 10	127x11	128	-

Results



Model	EM	F1
Baseline	31.41	47.57
Baseline+FT	34.555	49.881
Baseline+Train Longer	31.41	47.57
Baseline+FT+BTx 10	37.435	51.558
Dom. Adv. Training	34.50	50.48
Dom. Adv. Training +FT	32.41	48.97

Method

QA model loss:

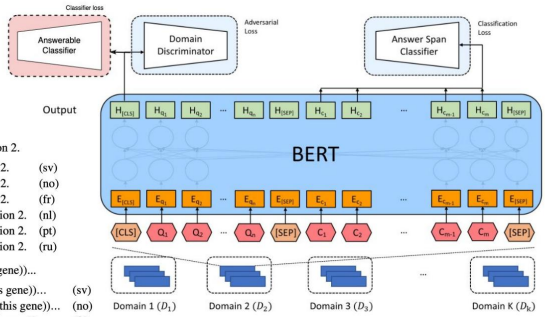
$$\mathcal{L} = \mathcal{L}_{QA} + \lambda \mathcal{L}_{Adv}$$

$$\mathcal{L}_{QA} = -\frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{N_k}[\log P_\theta(\mathbf{y}_{i,s}^{(k)}|\mathbf{x}_i^{(k)}, \mathbf{q}_i^{(k)}) + \log P_\theta(\mathbf{y}_{i,e}^{(k)}|\mathbf{x}_i^{(k)}, \mathbf{q}_i^{(k)})]$$

$$\mathcal{L}_{Adv} = \frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{N_k} KL(\mathcal{U}(l)\|P_\phi(l_i^{(k)}|\mathbf{h}_i^{(k)})),$$

Discriminator loss:

$$\mathcal{L}_D = -\frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{N_k}\log P_\phi((l_i^{(k)})|\mathbf{h}_i^{(k)})$$



Backtranslation:

(en) ... published by **Sony Computer Entertainment** of America, released on the PlayStation 2.

→
- ... published by **Sony Computer Entertainment** of America, released on PlayStation 2. (sv)
- ... published by **Sony Computer Entertainment** of America, released on PlayStation 2. (no)
- ... published by **Sony Computer Entertainment** of America, released on PlayStation 2. (fr)
- ... published by **Sony Computer Entertainment** of America, released on the PlayStation 2. (nl)
- ... published by **Sony Computer Entertainment** of America, released on the PlayStation 2. (pt)
- ... published by **Sony Computer Entertainment** of America, released on the PlayStation 2. (ru)

(en) Two genes located near each other on **chromosome 15** (CKMT1A and CKMT1B (this gene))...

→
- Two genes that are close to each other **chromosome 15** (CKMT1A and CKMT1B (this gene))... (sv)
- Two genes located close to each other on **chromosome 15** (CKMT1A and CKMT1B (this gene))... (no)
- Two genes that are close to each other **chromosome 15** (CKMT1A and CKMT1B (this gene))... (fr)
- Two genes located close to each other on **chromosome 15** (CKMT1A and CKMT1B (this gene))... (nl)
- Two genes located next to each other in **chromosome 15** (CKMT1A and CKMT1B (this gene))... (pt)
- Two genes located next to each other on **chromosome 15** (CKMT1A and CKMT1B (this gene))... (ru)

References

[1] Seanie Lee, Donggyu Kim, and Jangwon Park. Domain-agnostic question-answering with adversarial training, 2019.
[2] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473, 2014.
[3] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014.