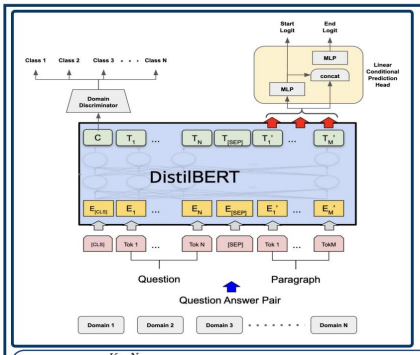
QAGAN: Adversarial Approach To Learning Domain Invariant Language Features

Shubham Shrivastava, Kaiyue Wang
Stanford University, CS224n Final Project

Introduction

- Large pre-trained language models like *BERT* perform extremely well on downstream QA tasks, but fails to generalize on *out-of-domain* dataset.
- *t-SNE* plots show that the **domain-gap truly exists** within datasets, and is further accentuated by downstream task training.
- We explore *adversarial approach* towards learning domain invariant features for better generalization.
- Many existing approaches predicts *start* and *end* logits independently; we also explore conditioning *end logits* prediction on *start logits*.



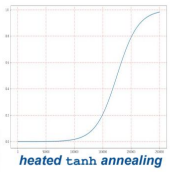
$$L_{QA} = -\frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{N_k}[logP_\theta(\hat{y}_{i,s}^{(k)}|x_i^{(k)}, q_i^{(k)}) + logP_\theta(\hat{y}_{i,e}^{(k)}|x_i^{(k)}, q_i^{(k)})]$$

$$L_{adv} = -\frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{N_k} y_{i,rand}^{(k)} logP_\theta(\hat{y}_{i,c}^{(k)}|h_{cls}^{(k)})$$

$$L_D = -\lambda_2\frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{N_k} y_{i,c}^{(k)} logP_\theta(\hat{y}_{i,c}^{(k)}|h_{cls}^{(k)})$$

$$L_{qagan} = L_{QA} + \lambda_1 L_{adv}$$

$$f_{anneal}(z) = \frac{tanh(2 * \frac{2z - n_{max} - n_{ws}}{n_{max} - n_{ws}}) + 1}{2}$$

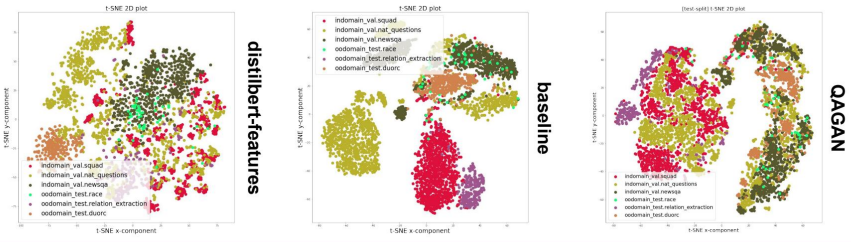


Technical Approach

- Given a ***question*** and a ***context*** paragraph, the objective of our model is to predict a probability of being *start* and *end* word of the answer span for each word in the ***context***.
- Our Question-Answering (QA) model is trained to minimize the *negative-log-likelihood* between ground-truth and prediction.
- We train a ***domain discriminator*** alongside our QA model to correctly classify the dataset domain class for each input sample, while the QA model works towards confusing it into outputting random classes..
- We designed *heated tanh annealing* function to aid discriminator training.
- To reduce overfitting we conducted data augmentation techniques using back translation and random word swap and observed improvements.

- **QAGAN minimizes** the domain-gap across various datasets through adversarial training.
- Conditional prediction head along with data augmentation techniques help the model perform better on *out-of-domain* dataset.

Domain-Gap Analysis



Experimental Results

Method	P_{head}	finetune	anneal	h_kld	ood_train	aug	D_{input}	D_{obj}	ind_val F1	ind_val EM	ood_val F1	ood_val EM
baseline	mlp	✗	-	-	✗	✗	-	-	70.49	54.48	48.29	30.89
baseline	mlp	✓	-	-	✗	✗	-	-	-	-	49.68	34.03
qagan	mlp	✗	✓	✗	✗	✗	[CLS]	KLD	70.10	54.24	46.56	31.15
qagan	mlp	✓	✓	✗	✗	✗	[CLS]	KLD	-	-	47.38	33.25
qagan	mlp	✗	✓	✗	✗	✗	[hidn]	NLL	68.88	52.69	46.95	30.89
qagan	mlp	✓	✓	✗	✗	✗	[hidn]	NLL	-	-	48.46	34.03
qagan	mlp	✗	✓	✗	✗	✗	[CLS]	NLL	69.85	53.84	46.92	31.68
qagan	mlp	✓	✓	✗	✗	✗	[CLS]	NLL	-	-	49.16	34.03
qagan	csat	✗	✓	✗	✗	✗	[CLS]	NLL	69.79	53.67	47.32	31.15
qagan	csat	✓	✓	✗	✗	✗	[CLS]	NLL	-	-	48.87	34.55
qagan	cmlp	✗	✗	✗	✗	✗	[CLS]	NLL	70.01	54.06	49.30	32.98
qagan	cmlp	✗	✓	✓	✗	✗	[CLS]	NLL	69.85	54.09	47.88	30.89
qagan	cmlp	✓	✓	✓	✗	✗	[CLS]	NLL	-	-	49.05	32.20
qagan	cmlp	✗	✓	✗	✗	✗	[CLS]	NLL	70.00	53.84	49.38	34.29
qagan	cmlp	✗	✓	✗	✓	✗	[CLS]	NLL	69.56	53.80	50.25	35.08
qagan	cmlp	✗	✓	✗	✓	✓	[CLS]	NLL	73.21	55.13	50.49	**35.90**
qagan	cmlp	✓	✓	✗	✗	✗	[CLS]	NLL	-	-	**51.00**	35.60

validation set

Method	P_{head}	finetune	ood_train	aug	D_{input}	D_{obj}	oodomain_test F1	oodomain_test EM
qagan	cmlp	✗	✗	✗	[CLS]	NLL	58.193	40.665
qagan	cmlp	✓	✗	✗	[CLS]	NLL	58.671	41.009
qagan	cmlp	✓	✓	✗	[CLS]	NLL	**58.898**	**42.362**

test set

QAGAN improves **F1** score by *5.6%* and **EM** score by *15.2%* over *baseline*.