Data Augmentation with Adversarial Examples and Back-Translation

Yuzu Ido, Stephan Sharkov, Eunice Yang
{yuzu, stpshrkv, eunicey}@stanford.edu | CS224N Winter 2022

Background

- **Question answering** is a critical NLP task and long-standing challenge in AI
- Models are given a question and related context as input, then try to answer the question correctly.
- State-of-the-art NLP models tend to **have trouble generalizing deeply** beyond their given training distribution to unseen domains.
- Data augmentation allows us to encode additional examples with **label preserving invariances** to increase diversity of our training set.

Problem

- How do we develop a QA system that is more **robust to out-of-distribution data** using a data augmentation approach?

Datasets

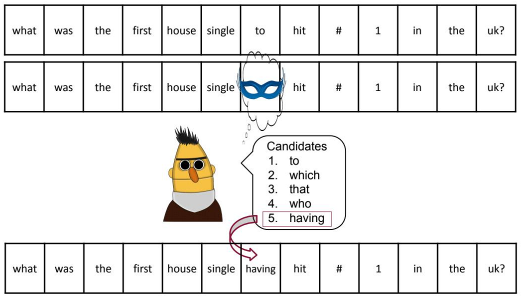
Dataset	Passage Source	Train	Dev	Test
In Domain				
SQuAD	Wikipedia	50000	10507	N/A
NewsQA	News articles	50000	4212	N/A
Natural Questions	Wikipedia	50000	12836	N/A
Out of Domain				
DuoRC	Movie reviews	127	126	1248
RACE	Examinations	127	128	419
RelationExtraction	Wikipedia	127	128	2693

Example (SQuAD)

Context: Southern California, often abbreviated SoCal, is a geographic and cultural region that generally comprises California's southernmost 10 counties…is a major economic center for the state of California and the United States.
Q: What is Southern California often abbreviated as?
A: SoCal

Methods

- **BAE**: BERT-based Adversarial Examples
  - Choose one word at random from question and mask
  - Predict top 5 choices of masked word with DistilBERT
  - Replace with lowest probability word (most adversarial)



what | was | the | first | house | single | to | hit | # | 1 | in | the | uk?

what | was | the | first | house | single | | hit | # | 1 | in | the | uk?

Candidates
1. to
2. which
3. that
4. who
5. having

what | was | the | first | house | single | having | hit | # | 1 | in | the | uk?

- **Back-Translation**
  - Translate question to Russian then back to English
- Baseline Model: **DistilBERT**
  - Use "distilled" version of original BERT transformer model pre-trained on SQuAD, NewsQA, and Natural Questions

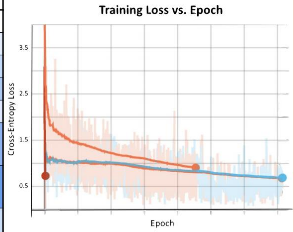
Discussion

- Data Augmentation **improved** DistilBERT performance on out-of-domain examples
- **Overfitting** when feeding model pure out-of-domain + BAE out-of-domain + back translated out-of-domain
  - Eliminating of pure out-of-domain improved model's performance
- **Randomizing** the BAE and backtranslated examples was **better** than our initial layering approach with BAE then backtranslation

>

Experiments

Finetuning on Baseline	F1
None (Baseline)	47.72
IN-BAE	47.99
IN-BT	48.26
IN-BAE < OUT-BAE	48.48
IN-BT < OUT-BT	48.05
IN-BAE < IN-BT < OUT-BAE + OUT-BT	48.76
IN-BAE + IN-BT < OUT-BAE + OUT-BT	49.07
IN-BAE + IN-BT + OUT-BAE + OUT-BT	48.2



Context: BPB Peptidoglycan, also known as murein, … Peptidoglycan serves a structural role in the bacterial cell wall, giving structural strength, as well as… binary fission during bacterial cell reproduction.
Q: bacterial cell walls are made rigid by the presence of
A: Peptidoglycan **Prediction:** Peptidoglycan

Context: (CNN) – Actor Gary Coleman is in critical condition in a Provo, Utah, hospital… the spokeswoman for Utah Valley Regional Medical Center, confirmed that… contributed to this report.
Q: What is the name of the hospital where Gary Coleman was admitted?
A: Utah Valley Regional Medical Center **Prediction:** Provo, Utah,

Future Work

- Address challenge of accidentally masking out a crucial word for question answering
  - Choosing word to **mask with importance** rather than random choice
- Use adversarial learning framework to conduct **domain adversarial training** and learn domain invariant features

References

1. Siddhant Garg and Goutham Ramakrishnan. Bae: Bert-based adversarial examples for text classification. In EMNLP, 2020.
2. Robert Östling, et al. The Helsinki Neural Machine Translation System. In EMNLP, 2017.