Deep Learning for Humanists @DHSI2023 – Day 3

Hoyeol Kim

These slides will be shared on GitHub

Mean Absolute Error = $\frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i |$

Mean Squared Error = $\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

Binary Cross Entropy (Log Loss) =

$$-\frac{1}{n}\sum_{i=1}^{n} \mathrm{y_i} * \log(\hat{y}_i) + (1 - y_i) * \log(1 - (\hat{y}_i))$$

L1 & L2 Loss Functions

$$L1LossFunction = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i |$$

$$L2LossFunction = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Loss

The image is from “https://openaccess.thecvf.com/content_cvpr_2018/papers/Chen_CartoonGAN_Generative_Adversarial_CVPR_2018_paper.pdf”

Pix2Pix

U-Net

- The encoder captures high-level features from the input image.
- The decoder generates a segmentation mask by combining the upsampled encoded features with features from earlier layers in the network.

Paper: https://arxiv.org/abs/1505.04597
Image credit: https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/

PatchGAN

Image credit: Geoffrey Hinton

- PatchGAN is used in a variety of GAN-based models such as conditional GAN and StarGAN.
- L1 & L2 loss can create blurry images in image generation tasks. Using PatchGAN, it is possible to generate detailed images with high frequency by limiting the size of patch for structures.
- PatchGAN decides whether or not outputs are true created by a generator.
- PatchGAN is a convnet, which is different from a regular GAN discriminator.
- PatchGAN can capture detailed information and provide precise feedback to the generator, which helps to ensure that the generated image maintains local consistency and captures fine-grained details.

GAN

(A)
Training set
Real sample
Discriminator
Latent space
(e.g., uniform)
Real
Fake
Generator
Noise vector
Fake sample
(B)
Training set
Bank
Real sample
Discriminator
Detective
Real
Fake
Generator
Counterfeiter
Noise vector
Inspiration
Fake sample

Transfer Learning

"Transfer learning is a machine learning method where a model developed for a task is reused as the starting point for a model on a second task." (Jason Brownlee)

Source: https://machinelearningmastery.com/transfer-learning-for-deep-learning/
Image: https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep-learning-212bf3b2f27a

Transfer Learning

1 - Semi-supervised training on large amounts of text (books, wikipedia..etc).

The model is trained on a certain task that enables it to grasp patterns in language. By the end of the training process, BERT has language-processing abilities capable of empowering many models we later need to build and train in a supervised way.

Semi-supervised Learning Step

Model:

Dataset:

Objective: Predict the masked word (langauge modeling)

2 - Supervised training on a specific task with a labeled dataset.

Supervised Learning Step

Model: (pre-trained in step #1)

Dataset:

Email message	Class
Buy these pills	Spam
Win cash prizes	Spam
Dear Mr. Atreides, please find attached…	Not Spam

Transfer Learning: T5

Image credit: https://ai.googleblog.com/2020/02/exploring-transfer-learning-with-t5.html

Transfer Learning: BERT

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

The Association for Computers and the Humanities

The Automatic Generation of Contextual Questions and Answers for English Learners

Minji Kang[1], Sungjin Ahn[2], Hoyeol Kim[3], Donghwan Kim[4], and Sung Won Han[1]

Korea University[1] Sungkyunkwan University[2] Texas A&M University[3] National University of Singapore[4]

Introduction

- We deployed the BERT, T5, and GPT-2 models for MC and Wh- question generation tasks.
- GPT-2 is a unidirectional language model, but it creates excellent results for text generation tasks.
- BERT is a bidirectional language model, the strength of which is predicting fill-in-the-blanks in context.
- T5 (Text-to-Text Transfer Transformer) is a pre-trained model with the C4 dataset. A T5 transformer can be fine-tuned for down-stream tasks.

Data

Description

- CoNLL 2014 and BEA 2019 are datasets containing essays written by ELs and native English speakers.
- Both datasets were originally designed for grammatical error correction tasks, so we fixed the errors on 170 of the essays to tailor them for our tasks.

Distribution Details

- To evaluate our generator, we randomly sampled 20 essays from the 170 corrected essays.

Datasets	Topics	Number of Essays
CoNLL 2014	Genetic risk	1
	Social media	3
BEA 2019	Write & Improve	10
	LOCNESS	6

- We created 20 MC and 20 Wh- questions, each with 4 answer choices.

Methodology

..... This instrument has brought other worlds right into the living room, making it possible to see life in other countries in a way that was never before possible. ..... The entertainment aspect of television has offered society an easy escape from its problems and difficulties. Therefore, since the invention of television, society has become more informed but too lazy to think about this new information, and has distanced itself from family life.

MC Question Generation Process

Summarize by Extracting (T5)

- We extracted the main sentences from the input text to make a summary. We used the T5 model developed by Google.

Paraphrase Sentences (T5)

- We paraphrased each of the main sentences with the T5 model.

Parse Sentences (Benepar)

- We parsed the paraphrased sentences to create inputs compatible with GPT-2. We used Berkeley Neural Parser by Berkeley NLP Group (ACL 2018).

Generate False Sentences (GPT-2)

- We generated false sentences to create distractors for the MC questions. We used the GPT-2 model developed by OpenAI.

Sort Sentences (BERT)

- We sorted the generated sentences based on dissimilarity from the corresponding raw sentences using the BERT model developed by Google.

According to the passage, which of the following sentence is true?

1) **Television has brought other worlds right into the living room, making it possible to see life in other countries in a way that was never possible before (True)**
2) **Television has the power to bring the war to the home as well as artistic expressions of those involved in program production giving society the opportunity to see both sides. (False)**

Wh- Question Generation Process

Extract Answer Tokens (Benepar)

- We used Benepar Parser to set the answer for a given passage. After parsing every sentence, we extracted all of the noun phrases and used them as answer tokens.

Generate Wh- Questions (T5)

- We generated Wh-questions using the answer tokens and the passages. We used the sequence-to-sequence question generator model, pre-trained on several question-answer pair datasets, with T5.

Evaluate and Choose Questions (BERT)

- We evaluated the semantic relevance of the generated question and answer pairs to ultimately generate a questionnaire. We used the sequence classification model pre-trained with BERT.

Generate Distractors (BERT)

- We masked part of the correct answer tokens and generated new words to use as distractors using the BERT-Based model.

What is the most important aspect of television?

1) **The entertainment aspect (answer)**
2) **The technical aspect**
3) **The broadcasting aspect**
4) **The political aspect**

Results

- Four experts rated 40 pairs of generated question and answer pairs using the given criteria and the results are shown in the table below.
  - Q1: MC distractors acceptability (1~5)
  - Q2: MC distractors grammar correctness (1~5)
  - Q3: Wh-distractors acceptability (1~5)
  - Q4: Wh-distractors grammar correctness (1~5)
  - Q5: Wh-question acceptability (1~5)
  - Q6: Wh-question grammar correctness (0, 1)*

	Q1	Q2	Q3	Q4	Q5	Q6
Evaluator 1	3.85	4.45	2.95	4.45	3.40	0.95
Evaluator 2	4.15	3.80	2.30	3.95	3.90	0.80
Evaluator 3	4.10	4.40	2.60	4.25	3.60	0.95
Evaluator 4	3.85	4.70	2.80	4.90	4.00	1.00
Average	**3.99**	**4.34**	**2.66**	**4.39**	**3.73**	**0.93**

Conclusion

- T5, BERT, and GPT-2 generate grammatically correct questions and distractors.
- Although the datasets were created by ESL students, they need to be tailored for MC and Wh- question generation tasks.
- There is room for improvement for MC and Wh- questions generation tasks. For instance, the model sometimes creates out-of-context Wh- distractors. Fine-tuning the model, curating the datasets, or creating tailored datasets can improve the results.

References

- Bryant, C., Felice, et al. (2019). The BEA-2019 Shared Task on Grammatical Error Correction. *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications*, 52–75. https://doi.org/10.18653/v1/w19-4406
- Devlin, Jacob, et al. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding.
- Ng, H. T., et al. (2014). The CoNLL-2014 Shared Task on Grammatical Error Correction. *Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task*, 1–14. https://doi.org/10.3115/v1/W14-1701
- Radford, Alec, et al. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, *1*(8), 9.
- Raffel, Colin, et al. (2019). Exploring the limits of transfer learning with a unified text-to-text transformer.

- Tutorials: https://pseudo-lab.github.io/Tutorial-Book/chapters/NLP/intro.html
- Homepage: https://pseudo-lab.com/

PseudoLab

Playground

chatGPT: https://chat.openai.com/auth/login

Bard: https://bard.google.com/

RoBERTa: https://huggingface.co/roberta-base

T5: https://huggingface.co/t5-base

Multimodal Deep Learning

(a) Video-Only Deep Autoencoder

(b) Bimodal Deep Autoencoder

Image credit: “Multimodal Deep Learning” by Jiquan Ngiam et al.

Images credit: https://towardsdatascience.com/multimodal-deep-learning-ce7d1d994f4

CLIP

Sources: https://github.com/openai/CLIP
https://openai.com/blog/clip/

DALLE2

DALL·E 2 can create original, realistic images and art from a text description. It can combine concepts, attributes, and styles.

TEXT DESCRIPTION

An astronaut Teddy bears A bowl of soup

riding a horse lounging in a tropical resort in space playing basketball with cats in space

in a photorealistic style in the style of Andy Warhol as a pencil drawing

→

DALL·E 2

DALL·E2: https://openai.com/dall-e-2

Stable Diffusion

Easy to use

stablediffusionweb.com is an easy-to-use interface for creating images using the recently released Stable Diffusion image generation model.

High quality images

It can create high quality images of anything you can imagine in seconds–just type in a text prompt and hit Generate.

GPU enabled and fast generation

Perfect for running a quick sentence through the model and get results back rapidly.

Stable Diffusion: https://stablediffusionweb.com/

Playground

DALL·E2: https://openai.com/dall-e-2

CLIP: https://huggingface.co/spaces/vivien/clip

Stable Diffusion: https://stablediffusionweb.com/

Face Generation: https://huggingface.co/spaces/CVPR/ml-talking-face

Any Questions?