Distractor Generation with Generative Adversarial Nets for Automatically Creating Fill-in-the-blank Questions

Chen Liang[1], Xiao Yang[2], Drew Wham[3], Bart Pursel[3], Rebecca Passonneau[2], C. Lee Giles[1]

[1]Information Sciences and Technology [2]Computer Science and Engineering [3]Teaching and Learning with Technology

PennState

Background

- Fill-in-the-blank (FITB) question

  Branch roots of the primary root of a flowering plant are initiated in the _________. Stem

  (a) cortex (b) pericycle (c) epidermis (d) endodermis

  Key Distractor

- Distractor generation (DG): generate distractor answers given the question sentence (stem) and the key to the question.

Motivation

- DG is a crucial step for fill-in-the-blank (FITB) question generation.
- Previous DG methods were mostly based on semantic similarities between the key and the candidate distractor
  - WordNet synonyms [1]
  - Embedding-based similarity [2, 3]
  - Co-occurrence likelihoods [3, 4]
- Existing DG methods have NOT fully explored how to utilize *context information* (stem) regarding the question.

Our Contribution

- Propose a machine learning-based approach for DG, which is fundamentally different from previous unsupervised similarity-based methods.
- Proposed method only uses stem information and it can be used in combination with existing key-based methods.
- A new dataset collected from college-level biology exams for evaluating DG or FITB generation.

Method – GAN

- Learn distractor distribution conditioned on the stem
- Adapt generative adversarial nets (GANs) [5] to tackle DG
- A generative model *G* that captures real data (key to the FITB question) distribution given a context (stem).
- A discriminative model *D* estimates the probability that a sample comes from the real training data rather than *G*.

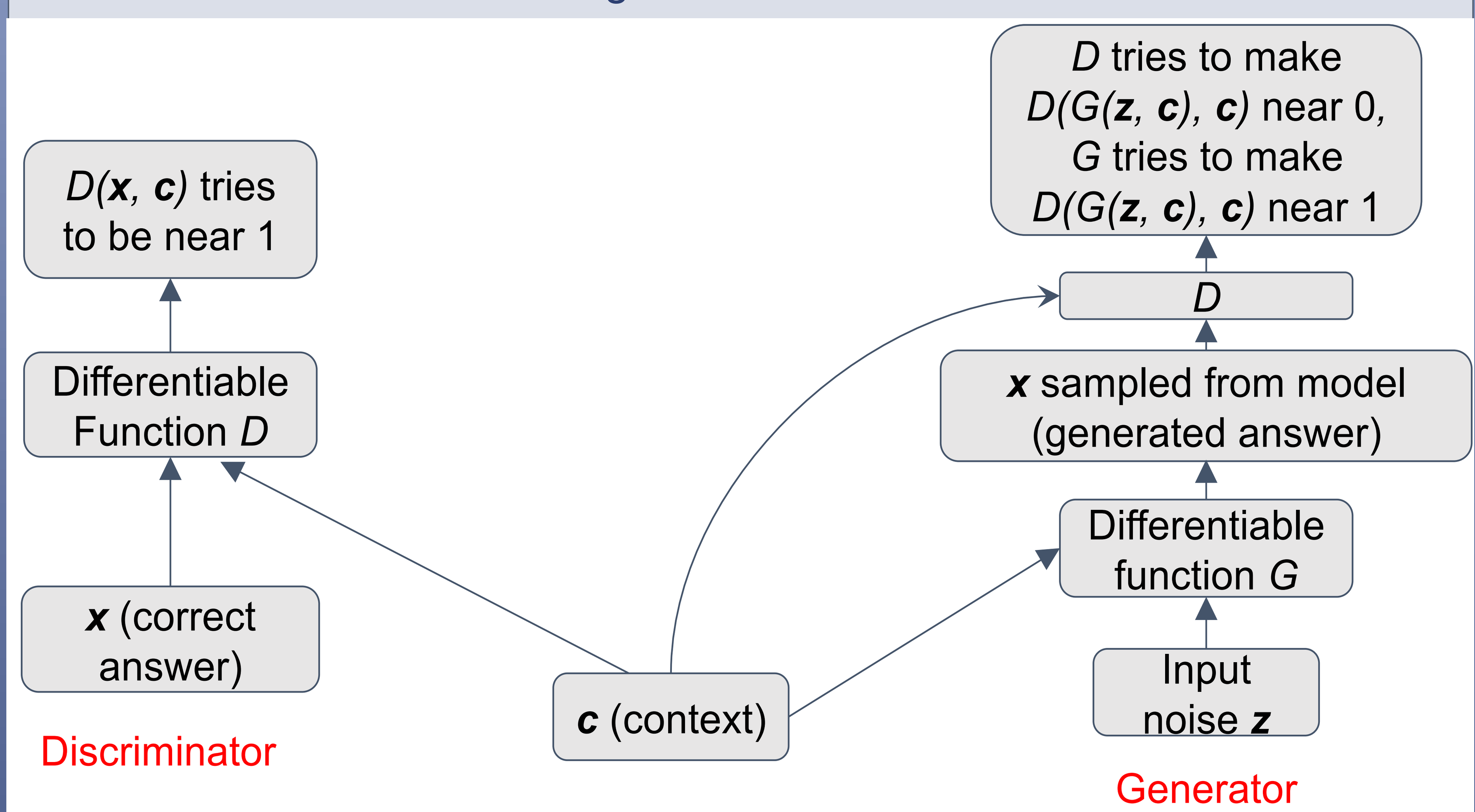

Figure 1: Conditional GANs for Distractor Generation

Challenge

- GAN requires that the composition of the generator and the discriminator are fully differentiable.
- Sampling a discrete token is not differentiable
- Use Gumbel softmax trick (Jang et al., ICLR'17) to enable back-propagating training signals from the discriminator to the generator.

Experiments

Data Preparation

- Utilize the Wikipedia corpus for creating the training set of (stem, key) pairs.
- Substitute the link in a Wikipedia sentence with a blank to get the stem and use the linked Wiki concept as the key.
- **Training set**: 1.62 million (stem, key) pairs, with a vocabulary of 8879 biology-related concepts
- **Test sets:**
  - **Wiki-FITB**: 30 FITB questions based on sentences in Wikipedia, selected by a domain expert
  - **Course-FITB**: 92 FITB questions from actual exams for two college-level biology courses and biology GRE 2016.

Experiment Settings

- For each (stem, key) pair, we apply the proposed DG method to generate a list of distractors
- Three domain experts were asked collaboratively to label each of the top-4 predictions as a *Good*, *Fair*, or *Bad* distractor
- **GAN**: the proposed GAN-based DG model
- **W2V**: a frequently used similarity-based method, which generates distractors based on the word2vec similarity between the candidate and the key
- **GAN+W2V**: Use the prediction score and the ranking of GAN and W2V as four features and train a logistic regression classifier to predict the probability of a distractor being good, fair, or bad.

Experiment Results

Table 1: Distractor generation results. Numbers are 95% confidence intervals of percentages of generated distractors being good, fair, or bad, calculated in a leave-one-out manner.

Methods	Good	Fair	Bad
GAN	28.4 (10.8)	10.8 (5.5)	60.8 (10.5)
W2V	35.8 (6.8)	10.0 (5.0)	54.2 (8.0)
GAN + W2V	40.0 (7.8)	11.7 (5.0)	48.3 (8.6)

(a) Wiki-FITB

Methods	Good	Fair	Bad
GAN	17.7 (5.0)	9.2 (3.5)	73.1 (5.9)
W2V	32.9 (5.3)	11.9 (3.5)	55.2 (5.7)
GAN + W2V	34.3 (5.7)	14.1 (3.9)	51.6 (6.0)

(b) Course-FITB

GAN	W2V	GAN + W2V
Speciation	Natural selection	Speciation
Transcription	Macroevolution	Natural selection
Inbreeding	Evolutionary biology	Microevolution
Genetic drift	Microevolution	Genetic drift

Table 2: Distractor generation examples for question "Changes in gene frequency over time describes the process of ______.", whose key is *Evolution*. (Legend: Good, Fair, Bad)

Acknowledgements

We gratefully acknowledge partial support from the Pennsylvania State University Center for Online Innovation in Learning.

References

[1] Mitkov, R., LE AN, H. A., & Karamanis, N. (2006). A computer-aided environment for generating multiple-choice test items. *Natural language engineering*, *12*(2), 177-194.

[2] Kumar, G., Banchs, R. E., & D'Haro Enriquez, L. F. (2015). Revup: Automatic gap-fill question generation from educational texts. Association for Computational Linguistics.

[3] Jiang, S., & Lee, J. (2017). Distractor Generation for Chinese Fill-in-the-blank Items. *BEA@EMNLP*.

[4] Hill, J., & Simha, R. (2016). Automatic Generation of Context-Based Fill-in-the-Blank Exercises Using Co-occurrence Likelihoods and Google n-grams. In *BEA@ NAACL-HLT*

[5] Goodfellow, Ian, et al. "Generative adversarial nets." *Advances in neural information processing systems*. 2014.