STAT479 Project Proposal

Richard Yang gyang79@wisc.edu
Junyao Wang jwang2286@wisc.edu
Qiaochu Yu qyu65@wisc.edu

1. Introduction

1.1. Dataset

In our project, we are going to use data from Quora to analyze the sincerity of questions.

According to the statement in Quora [1], an insincere question is defined as "a question intended to make a statement rather than look for helpful answers". Some characteristics that can signify that a question is insincere as listed in kaggle competition website[2]:

- " has an exaggerated tone to underscore a point about a group of people."
- "is rhetorical and meant to imply a statement about a group of people."
- "suggests a discriminatory idea against a protected class of people, or seeks confirmation of a stereotype."
- "makes disparaging attacks/insults against a specific person or group of people based on an outlandish premise about a group of people."
- "disparages against a characteristic that is not fixable and not measurable Isn't grounded in reality."

"The training data includes the question that was asked, and whether it was identified as insincere (target = 1). The ground-truth labels contain some amount of noise: they are not guaranteed to be perfect."

The dataset contains the question IDs, question text and its label (sincere or not), and the train set contains 1.3 million examples while the test set has 376K examples, which is quite large and suitable for deep learning models.

1.2. Our plan

Because it's a natural language processing problem, we may use specific type of deep neural network for modeling, typically RNN, such as LSTM or GRU[4], to finish the language information detection or implement transfer learning using some well-trained basic language models as the precursor of our model . And we will pick proper hyperparameters without trying too many combinations to gain high accuracy in our test set. After that if possible, we want to regenerate sincere or non-sincere questions using some generative model like a GAN[1] or an Autoencoder [2] to find what kind of questions are learned to be sincere in our model, and this generator can also be useful in real world application.

The whole procedure can be roughly shown as belows:

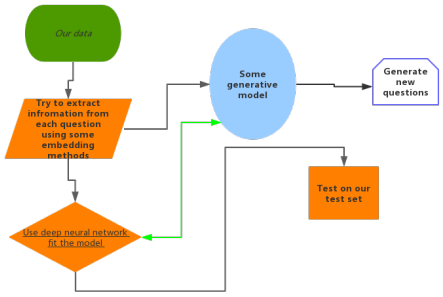

Figure 1. The flow chart of our thoughts

1.3. Related work

Some previous work both for classification and generating can be useful to us:

1. *Recurrent neural network for text classification with multi-task learning–arXiv*[3]
2. *Document modeling with gated recurrent neural network for sentiment classification*[5]
3. *Generating text via adversarial training–NIPS workshop on Adversarial Training*[6]

2. Motivation

Quora is a question-and-answer website where questions are asked, answered, edited and organized by its community of users, which was co-founded by former Facebook

[1] https://www.kaggle.com/c/quora-insincere-questions-classification/data
[2] https://www.kaggle.com/c/quora-insincere-questions-classification/data

employees Adam D′Angelo and Charlie Cheever in June 2009. When asked why they choose the name, Cheever stated, *" I associate it with quorum or public congregation."* Users post questions within Quora and other users can answer the questions. Also users are able to vote for or comment against answers posted by others and the answers are sorted by the results in its default state. These interactions naturally select the best answer for a question. As the origin of its name indicates, Quora has a social network backbone that nicely integrates its user base. The social aspects give uniqueness to Quora and deserves a further investigation. In the past years, there is an exponential growth in the community of users of Quora, and it also poses a problem necessary to be solved how to handle toxic and divisive contents[3].

The sincerity of questions and answers provides safety for its users when they are sharing their knowledge with the world. How to weed out insincere questionsthose founded upon false premises, or those intend to make a statement rather than look for helpful answers, is an important step to solve the problem. Otherwise there would be lots of such insincere questions making people hard to get unbiased information or solution in time[4].

3. Evaluation

Our goal is to fit a deep NN model in our training set what can achieve high accuracy in the test set, thus we can tell if a given question is sincere or not with high confidence level.

The success of our model is:

- our model can distill words or sentences information sufficiently.
- the neural network fitted using our training set can generalize well.
- our generative model can generate meaningful question to validate our model. To test meaningful, we can try to use generated data to see if it can fool our discriminator .

4. Resources

4.1. Dataset

For the datasets, we will use an existing data from Kaggle competition, with the theme insincere questions classification

[https://www.kaggle.com/c/quora-insincere-questions-classification/data] .

This site contains data sets for testing and training.

[3] https://www.kaggle.com/c/quora-insincere-questions-classification

[4] https://www.kaggle.com/c/quora-insincere-questions-classification

4.2. Word embedding

Also, the above website provides trained word embeddings:

- GoogleNews-vectors-negative300 [https://code.google.com/archive/p/word2vec/]
- glove.840B.300d [https://nlp.stanford.edu/projects/glove/]
- paragram_300_sl999 [https://cogcomp.org/page/resource_view/106]
- wiki-news-300d-1M [https://fasttext.cc/docs/en/english-vectors.html]

4.3. Coding and computation

For the computer hardware, we will use Google cloud server to handle the enormous computation. Meanwhile, we'll use PyTorch mainly to construct the model and do the computation.

5. Contributions

For the writing part, Richard Yang will be responsible for the Proposed Method and Experiments parts. Junyao Wang will be responsible for the Results and Discussion and Conclusions parts. Qiaochu Yu will be responsible for the rest parts.

For experiments, we will take part in model building and experiments together since that's the main part of the project and is of large work load. Also, by cooperating, we can try different models or hyperparameters simultaneously to be efficient.

References

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. pages 2672–2680, 2014.

[2] C.-Y. Liou, W.-C. Cheng, J.-W. Liou, and D.-R. Liou. Autoencoder for words. *Neurocomputing*, 139:84–96, 2014.

[3] P. Liu, X. Qiu, and X. Huang. Recurrent neural network for text classification with multi-task learning. *arXiv preprint arXiv:1605.05101*, 2016.

[4] T. Mikolov, M. Karafiát, L. Burget, J. Černockỳ, and S. Khudanpur. Recurrent neural network based language model. In *Eleventh annual conference of the international speech communication association*, 2010.

[5] D. Tang, B. Qin, and T. Liu. Document modeling with gated recurrent neural network for sentiment classification. In *Proceedings of the 2015 conference on empirical methods in natural language processing*, pages 1422–1432, 2015.

[6] Y. Zhang, Z. Gan, and L. Carin. Generating text via adversarial training. In *NIPS workshop on Adversarial Training*, volume 21, 2016.