Ashwin Somasundaram

Reading ACL Paper Assignment

Paper: Simulating Bandit Learning from User Feedback for Extractive Question Answering
Authors: Ge Gao (Cornell PhD student) , Eunsel Choi (Professor at UT Austin), Yoav Artzi (Professor at Cornell University)

Problem Summary:

This paper explores the problem of the current dominant paradigm in NLP models. The model training is usually separate from its deployment, thus creating a divide. This prohibits models from being able to learn and improve themselves through user feedback. In the paper, the researchers study the potential of extracting feedback from users and then utilizing it to improve the accuracy of the model itself. The problem in this scenario is categorized as contextual bandit learning. This method is defined by the learner independently observing a single context during each time step and making decisions based off of that. There is a singular reward which corresponds solely with the action they have taken. The main aim of the learner is to minimize the cumulative regret: in this case this would be the backlash suffered by the learner due to previous policy actions. In simpler terms, this is a modified version of a reinforcement learning algorithm.

Prior work summary:

There has been plenty of prior work done in the past in the field of contextual bandit learning. It has also been applied to a variety of NLP applications including neural machine translation, structured prediction, semantic parsing and summarization. Building upon this, the paper's research shows the effectiveness of bandit feedback in terms of Question and Answering tasks. There is however a key difference in the focus of this research. The paper aims to reduce annotative data required during training and restrict all data flow to during the deployment of the model (in that particular domain itself). While implicit feedback has been studied in the past as well, in this experiment the researchers focus on the advantages of obtaining explicit feedback from users regarding the model's performance. Another research paper in the past has also mentioned work about improving QA systems by utilizing feedback-weighted learning which incorporates the simulated binary feedback as a part of training. The main contrast here is that the potential improvements of the QA systems are being studied through the lens of a bandit learning system.

Unique Contributions:

In order to explain the contributions of this paper, let us take a quick look at the approach and details entaling the research experiment at hand. The researchers study a situation where the users provide information/feedback which a QA model can learn from. The model is initialized at firstvia the regular supervised data training method. The next step is where the research differs from the norm. The team simulates bandit feedback utilizing some of the supervised data annotations which gives the impression of a ‘user’. The reward metric is also a lot more stringent, 1 for positive prediction/action and -1 for a negative prediction. An important point about this model is that the supervised data given at the beginning as a part of the same domain is minimized to gauge how well this model performs in terms of user feedback. Thus the main contribution of this paper is explaining how much merit there is in using a modified reinforcement learning approach to build Question and Answer type NLP models.

Evaluation of Work/Experiments:

The Pretrained model used for the purpose of this paper is the SpanBERT model. Empirical results from online learning show that there is an overall substantial performance gain when a weak initial model is utilized. Specifically, there are 6 QA datasets being used: SQuAD, NewsQA, SearchQA, TriviaQA, HotPotQA and NaturalQuestions. The model preformas pretty well in each of these. In certain cases however, bandit learning is not as effective with weaker base models. The researchers hypothesize that this may be due to a lack of quality annotated supervised data in the set. Another metric utilized by the researchers to conduct the same experiments offline and gauge how the performances are in comparison. The offline experiments show mixed results in comparison with the online results. In particular, ‘Table 3’ in the paper, signifies the drastic improvement in performance of some datasets, when compared with in-domain initial models and SQuAD initialized models including the source domain. In the research paper, they provide multiple tables and graphical diagrams to portray the performance metrics of each model with a dataset. Apart from the pictorial representations, they also make sure to explain unwarranted behavior or observations. This clearly emphasizes the approach and evaluation methods of the experiments conducted for this research paper.

Author information and Conclusion:

The primary author of this research paper is Ge Gao. She is currently a graduate student at Cornell University pursuing a PhD in Natural language processing and computational linguistics. She has 110 citations on her google scholar. Dr. Eunsel Choi is the secondary author and has 4936 citations on her google scholar page. Dr. Yoav Artzi is the faculty advisor (also affiliated with Cornell University) of Ge Gao and he has 7234 citations on his google scholar page.

I think that this paper was really innovative in terms of the way they crafted the learning model and executed the idea to improve it based on user feedback. Utilizing two types of learning methods was a smart idea, as it shows the real world performance of the model and how it would fare against existing technologies. A simple yet powerful example of NLP taking feedback from users, would be ChatGPT which utilizes information presented to it and learns off of that to make smarter and more informed decisions. I definitely believe that this paper is important to the field of NLP as having models which can be trained on the fly in production environments would be a massive improvement to the technology we currently have.