TECHNICAL DETAILS OF REPUTATION AGENT

TRAINING AND TESTING OF MACHINE LEARNING MODEL.

Given a set of labeled reviews (i.e., reviews that are labeled as to whether they are fair or unfair), Reputation Agent uses stratified sampling to split the labeled data into training, test, and validation sets under proportions of 80%, 10%, and 10% (the validation set helps to avoid overfitting). Using Python 3 and the Keras framework with Tensorflow, we trained eight models to learn to recognize reviews that evaluate workers based on mission-critical metrics and non-critical ones. Our goal was to identify the machine learning models which worked the best for different gig markets. For this purpose, we trained different machine learning models which used as feature vectors either word vectors or word embeddings:

Word ngram + LR: Logistic regression with word ngrams.

Char ngram + LR: Logistic regression with character ngrams.

(Word + Char ngram) + LR: Logistic regression with word and character ngrams.

RNN no embedding: Recurrent neural network (bidirectional GRU) without pre-trained embeddings.

RNN + GloVe embedding: Recurrent neural network (bidirectional GRU) with GloVe pre-trained embeddings.

CNN (multi-channel): Multi-channel Convolutional Neural Network.

RNN + CNN: Recurrent neural network (Bidirectional GRU) + Convolutional Neural Network.

Google BERT (Devlin et al., 2018): Bidirectional Encoder Representations from Transformers, is a new method of pre-training language representations which obtains state-of-the-art results on a wide range of Natural Language Processing (NLP) tasks.

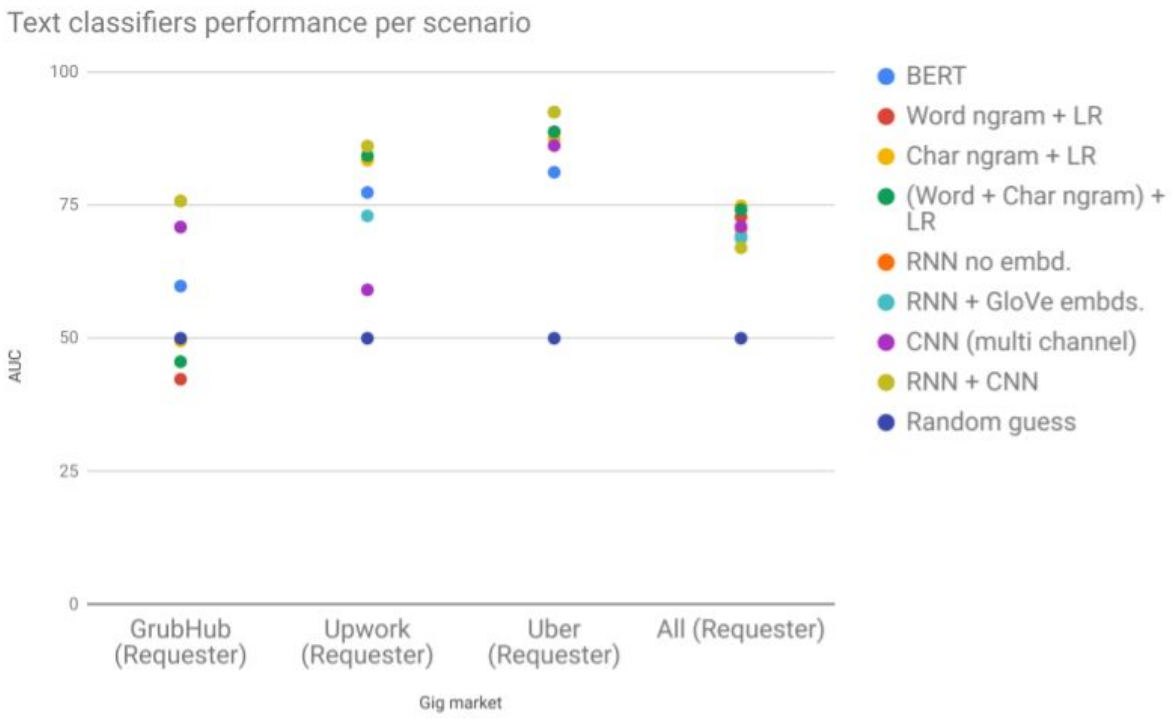

Figure X. Text classifier benchmark. Recurrent Neural Networks (RNN) + Convolutional Neural Network (CNN) approach performed better across conditions.

We implemented early stopping as a method to stop training once the model performance stops improving on a hold out validation dataset. For the deep learning models, we used a binary cross-entropy loss function, ADAM as an optimizer, and a learning rate of 0.001.

Fig. X presents an overview of the benchmark of the training models (i.e., the figure shows the performance metrics of each model). We note that different machine learning models performed better on different gig markets. However, RNN (Recursive Neural Network, a Deep Learning Algorithm) performed in general the best across all gig markets. This was the reason we eventually choose to utilize this model.

After the model has been trained, it is exposed as a REST web service via JSON requests to our front-end interface. The service is consumed directly by Reputation Agent's Accuracy Promoter and it displays the messages accordingly.