Team Info

Yu Zhu (u1211862)

Submission 1

- **Submission ID?**
  File name: perceptron-average.csv
  Submission date: 2018-11-15 ~9:50pm
- **What algorithm did I use? How did I implement it?**
  Perceptron (average). I run average perceptron with epoch T = 20, margin = 0.01 and learning rate = 0.1.
- **Performance?**
  This version shoots an accuracy of 0.8801 on the test dataset and **0.87776 on the Kaggle** board.
- **Any feature engineering?**
  No.

Submission 2

- **Submission ID?**
  File name: submission-adaboost-w2v.csv
  Submission date: 2018-11-15 10:23pm
- **What algorithm did I use? How did I implement it?**
  AdaBoost + Perceptron (average). Epoch = 10. Each week classifier is an average perceptron with epoch T = 1, margin = 0.1, learning rate r = 1.
- **Performance?**
  This version gets an accuracy of 0.8526 on the test dataset and **0.8544 on the Kaggle** board.
- **Any feature engineering?**
  Instead of the original bag of words, I use *word embedding* to represent sentences. Specifically, each word is represented by a 300-dimension word vector, and each sentence is the average (centroid) of the words it contains. However, this feature engineering didn't beat the original bag of words.

What else have I tried?

- Other perceptron algorithm, like the simple, aggressive, margin, etc.
- Change epoch number from 1 to 20, margin from 0 to 1, and learning rate from 0.01 to 1

Roadmap

- **Feature engineering**
  - Train word embeddings based on the given data, instead of using an open source collection.
  - Add stop words.
  - Add syntactic parsing into the features.
- **Other algorithms**
  - SVM
  - Decision stump
  - Logistic
  - Neural Network