CS 6350 Machine Learning Project Final Report

Predict Closed Questions on Stack Overflow.

1   Team Members

- Aravind Senguttuvan (u0877727)

- Anirudh Narasimhamurthy (u0941400)

2   Description of the problem

The problem which we worked on involves building a classifier which predicts whether or not a question posted on StackOverflow will be closed, given the question as submitted. Broadly speaking our problem is a multi-class classification problem in that the prediction labels given by the classifier are

1. Open

2. Off Topic(OT)

3. Not Constructive (NC)

4. Not a Real Question(NRQ)

5. Too Localized(TL)

After a question is submitted in StackOverflow, the proposed algorithm classifies whether a question will be closed or not. Classifier will output a label based on four features namely Off topic (OT), not constructive (NC), not a real question (NRQ), too lo calized (TL) or exact duplicate.

3   Interesting features

If the model works accurately, then there wouldn't be a need of manual intervention to close a question posted on StackOverflow.

4   Techniques Used

The paper `http://ceur-ws.org/Vol-1031/paper2.pdf` uses Random Forest, Support Vector Machine(SVM) and Vowpal Wabbit techniques to determine if a question should be closed. We are trying to prove the paper's evaluation.