

CS 583 Project 2: Twitter Data Mining

Xuelong Wang & Shuangxi Zhu

University of Illinois at Chicago

April 20, 2017

Data processing

- ▶ Features extraction
 1. Stop word;
 2. Combine Unigram and N-grams models (3-5).
- ▶ Grid Search:
 1. Elastic Net;
 2. Hyperparameter of penalty for model complexity;
 3. Classes weights.
- ▶ Ensemble Classifiers:
 1. Majority vote among three classifiers:
 - ▶ SVM
 - ▶ MultinomialNB
 - ▶ Logistic Regression

Classification and Results

- blue represents Obama; red represents Romney

Algorithms	Results						
	Positive class			Negative class			Overall Acc
	p	r	F	p	r	F	
Ensemble	0.67	0.65	0.66	0.61	0.68	0.64	0.64
	0.58	0.49	0.53	0.66	0.77	0.71	0.62