



Twitter Sentiment Analysis

INF 298 I

Silvano Nogueira Buback



Agenda

- Motivação
- Estado da arte
- Entendimento
- Dataset - Twitter
- Resultado
- Conclusão



Motivação

- Avaliar o sentimento das pessoas acerca de um assunto
- Avaliar produtos e serviços
- Reputação de pessoas famosas/políticos



Estado da Arte

- Não existe uma comparação para o Twitter em portugues
- Naive Bayes: 78.7% (81.5% unig+POS)
- Hybrid SVM: 89%



Entendimento

- Entendimento do problema: moview review
 - treino: 1800 (ing) 1472 (port)
 - teste: 200 (ing) 162 (port)
- Resultado (ingles): 81.5%
- Resultado (portugues): 79%



Dataset - Twitter

- Inexistência de um bom dataset em português para treino
- Muitos erros de ortografia
 - “*aah gente o justin ja passou mais agora que eu onseguir ver*”
 - “*shorei horrores em toy story 3 como eu ri bastante tambem mas e bem trsite e nao tem musicas no 3*”



Resultado (I)

Positivos: 1.818 / Negativos: 862 / Treino: 500

	Positivo	Negativo	Acurácia
Básico	65.7%	72.9%	67.3%
Stemmer	65.2%	74.0%	67.1%
Stemmer+s/ acentos	65.3%	75.1%	67.4%
St+s/ac palavras chaves	69.7%	53.8%	66.3%



Resultado (2)

Positivos: 99.404 / Negativos: 32.012 / Teste: 500

	Positivo	Negativo	Acurácia
Básico (inv)	64.0%	59.4%	62.9%
Básico	86.6%	51.8%	69.2%
Stemmer+s/ acentos	85.0%	53.4%	69.2%



Conclusão

- Despreocupação com gramática/ortografia tornam o texto difícil de ser entendido até por humanos
- Mesma palavra com diversas grafias (haha)
- Enorme quantidade de erros de grafia



Trabalhos futuros

- Combinar técnicas diferentes (SVM)
- Admitir comentários NEUTRO
- Correção ortográfica do texto /
normalização da grafia de algumas palavras
(outro algoritmo de ml?)
- Relacionar tweets do mesmo usuário/
Replies