Automatic Detection of Irony and Humour in Twitter

Francesco Barbieri

Horacio Saggion

Universitat Pompeu Fabra Barcelona

talp research group

Presentation Outline

- Introduction
- DataSet / Resources
- Features for irony and humour classification
- Classification Experiments
  - Irony and Humour cross domain classification
  - Creative VS non-creative
- Conclusions and Current Work

Introduction

- Twitter
- Irony and humour
- Automatic detection
- Formal definition

Introduction

- **#Irony**
  - Bush sent more troops than Obama to create Peace in Afghanistan but Obama got the NOBEL!
  - I tell you a secret… I love Christmas!
- **#Humour**
  - Computers are like air conditioners. They work fine until you start opening windows.
  - What's the difference between government and the mafia? One is organized.

Introduction

- Being able to detect ironic and humorous statements can be useful in many human-computer interaction applications
- As in previous work (Reyes, Rosso and Veale (2013)) we cast the problem as binary classification and use a machine learning algorithm to discriminate ironic from non-ironic messages and Humorous

Dataset & Text Processing

- **Tweets:** Corpus of 40.000 tweets equally divided into four different topics (Reyes et al, 2013) #irony, #humour, #education, #politics
- **ANC Frequency Data (oral / written)** (Ide and Suderman, 2004)
- **TwitIE** (Bontcheva et al., 2013) tokeniser and Part of Speech Tagger.
- **WordNet** (Miller, 1995)
- **SentiWordNet** (Esuli and Sebastiani, 2006)
- **Potts' Intensity Scores** (Potts, 2011)

Features

1. **Frequency** *(gap between rare and common words)*
2. **Written-Spoken** *(written-spoken style uses)*
3. **Intensity** *(intensity of adverbs and adjectives)*
4. **Structure** *(length, punctuation, emoticons)*
5. **Sentiments** *(gap between positive and negative terms)*
6. **Synonyms** *(common vs. rare synonyms use)*
7. **Ambiguity** *(measure of possible ambiguities)*

Frequency

- Gap between rare and common words, i.e. register inconsistencies in the same tweet as a mark of unexpectedness (Lucariello 1994, Venour 2013)
- We compute the frequency imbalance between words (ANC)

1. **frequency mean**
2. **rarest frequency**
3. **frequency gap**

Written-Spoken

- Unexpectedness created by using spoken style words in a mainly written style tweet or vice versa using ANC spoken and written corpora

1. **written mean**
2. **spoken mean**
3. **written spoken gap**

Intensity

- Intensity scores of Potts (2011):
  horrible (-1.9) → *bad (-1.1)* → *good (0.2)* → *nice (0.3)* → *great (0.8)*

1. **adj (adv) tot**
2. **adj (adv) mean**
3. **adj (adv) max**
4. **adj (adv) gap**

Structure

- Length, punctuation, emoticons

1. **Length**
2. **N. words**
3. **Word length mean**
4. **N. verbs, N. nouns, N. adjectives, N. adverbs**
5. **Verb Ratio, noun R., adjective R., adverb R.**
6. **Punctuation (? ! , . … ; “ ” *)**
7. **Laughing (LOL, …)**
8. **Emoticon**
9. **Internet Links**

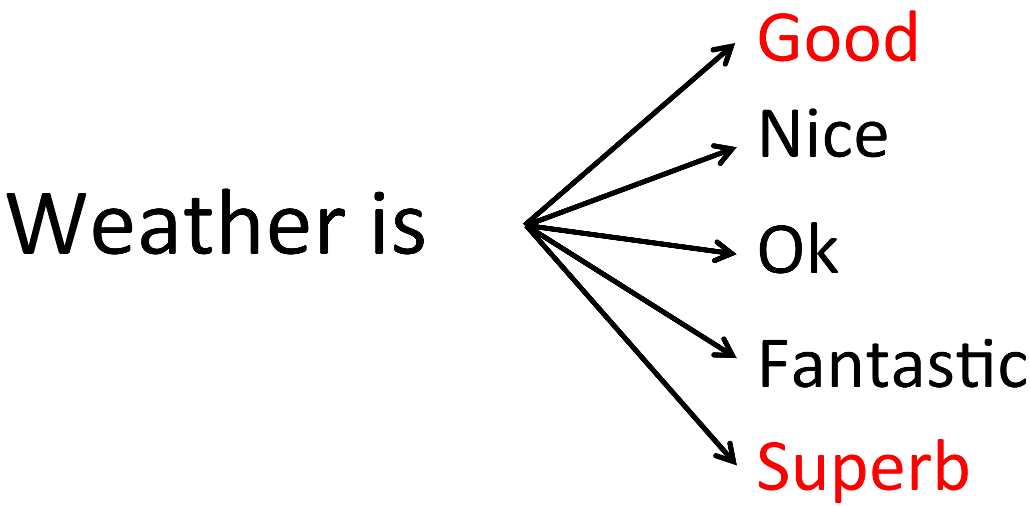
Sentiments

- Gap between positive and negative terms
- SentiWordnet (Esuli and Sebastiani, 2006)

1. **positive sum**
2. **negative sum**
3. **positive negative mean**
4. **positive-negative gap**
5. **positive single gap**
6. **negative single gap**

Synonyms

- Model choice of synonym using frequencies from ANC



Synonyms

- W1 … Wi … Wn

- W1 … S1 … S1 ← Most frequent
- S1 … S2 … Wn
- S2 … S3 … S2
- S3 … Wi … S3 ← Least frequent

Synonyms

- W1 … Wi … Wn

- W1 … S1 … S1
- S1 … S2 … Wn
- S2 … S3 … S2
- S3 … Wi … S3

1. **syno lower**
2. **syno lower mean**
3. **syno lower gap**
4. **syno greater**

Synonyms

- W1 … Wi … Wn

- W1 … S1 … S1
- S1 … S2 … Wn
- S2 … S3 … S2
- S3 … Wi … S3

1. syno lower = 3
2. syno lower mean
3. syno lower gap
4. syno greater

Synonyms

- W1 … Wi … Wn

- W1 … S1 … S1
- S1 … S2 … Wn
- S2 … S3 … S2
- S3 … Wi … S3

1. syno lower = 3
2. syno lower mean

   = (3+0+2)/3 = 1.66
3. syno lower gap
4. syno greater

Synonyms

- W1 … Wi … Wn

- W1 … S1 … S1
- S1 … S2 … Wn
- S2 … S3 … S2
- S3 … Wi … S3

1. syno lower = 3
2. syno lower mean

   = (3+0+2)/3 = 1.66
3. syno lower gap

   = (n1) - (n2) = 1.34
4. syno greater

Synonyms

- W1 … Wi … Wn

- W1 … S1 … S1
- S1 … S2 … Wn
- S2 … S3 … S2
- S3 … Wi … S3

1. syno lower = 3

2. syno lower mean

= (3+0+2)/3 = 1.66

3. syno lower gap

= (n1) - (n2) = 1.34

4. syno greater = 3

Ambiguity

- Using a word with many meanings allows us to *say something meaning something else*

1. **mean of number of synsets**
2. **max num of synsets**
3. **gap max - mean**

Experiments and Results

- **Irony** Cross-domain Classification
  - Training Set:
    - *7500 Irony vs 7500 Education*
    - *7500 Irony vs 7500 Politics*
    - *7500 Irony vs 7500 Humour*
  - Test Set:
    - *2500 Irony vs 2500 Education*
    - *2500 Irony vs 2500 Politics*
    - *2500 Irony vs 2500 Humour*

Experiments and Results

- **Irony** Cross-domain Classification
  - Training Set:
    - *7500 Irony vs 7500 Education*
    - *7500 Irony vs 7500 Politics*
    - *7500 Irony vs 7500 Humour*
  - Test Set:
    - *2500 Irony vs 2500 Education*
    - *2500 Irony vs 2500 Politics*
    - *2500 Irony vs 2500 Humour*

Experiments and Results

- **Humour** Cross-domain Classification
  - Training Set:
    - *7500 Humour vs 7500 Education*
    - *7500 Humour vs 7500 Politics*
    - *7500 Humour vs 7500 Irony*
  - Test Set:
    - *2500 Humour vs 2500 Education*
    - *2500 Humour vs 2500 Politics*
    - *2500 Humour vs 2500 Irony*

Experiments and Results

- Irony + Humour = **Creative**
- Education + Politics = **Non-creative**

- **Creative** VS **Non-creative** classification
  - Training Set:
    - *15000 (Irony + Humour) vs 15000 (Education + Politics)*
  - Test Set:
    - *5000 (Irony + Humour) vs 5000 (Education + Politics)*

Experiments and Results

- **Irony** Cross-domain Classification (F1)

Test	Training: Education	Training: Humour	Training: Politics
Education	**87**	86	86
Humour	77	**88**	76
Politics	82	82	**88**

- **Humour** Cross-domain Classification (F1)

Test	Training: Education	Training: Irony	Training: Politics
Education	**78**	46	71
Irony	71	**88**	69
Politics	73	51	**80**

Experiments and Results

- **Creative** VS **Non-creative** classification

  F1 = 0.80

Experiments and Results

- **Best 10 Features** (Information Gain)

Irony	Humour	Creative
links	syno. greater	syno. lower
questions	rarest frequency	rarest frequency
syno. greater	adv. max	word length mean
rarest frequency	link	noun ratio
fullstop	questions	adv. max
syno. lower	syno. lower	punctuation
punctuation	positive sum	syno. greater
n. words	adv. total	adv. mean
noun ratio	adv. mean	written spoken gap
tw.lenght	word mean	avgWritten

Conclusion and Current Work

- We presented a model and experiments on binary classification of “ironic” and “non-ironic” tweets and “humorous” and non-humorous.
- The proposed model does not rely on word-based information but on word characteristics
- We are creating our own dataset for sharing / comparing different approaches
- We are currently working on other aspects of language

Conclusion and Current Work

- We presented a model and experiments on binary classification of “ironic” and “non-ironic” tweets and “humorous” and non-humorous.
- **The proposed model does not rely on word-based information but on word characteristics**
- We are creating our own dataset for sharing / comparing different approaches
- We are currently working on other aspects of language

Experiments and Results

- Information gain Pearson **Correlation** of information gain values for each feature over different topics when training:

Irony

Model	Education	Humour	Politics
Education	1	0.76	**0.96**
Humour		1	0.76
Politics			1

Humour

Model	Education	Irony	Politics
Education	1	0.48	**0.89**
Irony		1	0.36
Politics			1