Fuzzy Sentiment Analysis using Spanish Tweets

Session 20
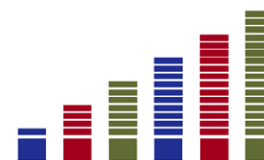
Pilar Rey del Castillo

Q2016
European Conference on
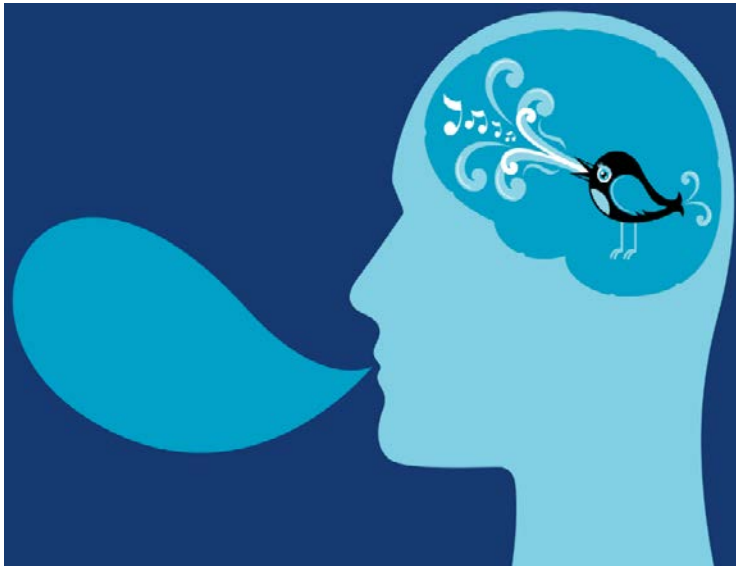**Quality** in **Official Statistics**
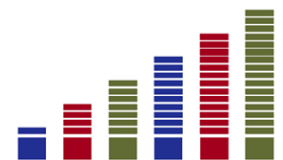
Madrid, May 31 – June 3

Outline

- Microblogs for sentiment analysis

- ANEW Spanish lexicon

- Model to perform sentiment analysis from tweets: results

- Final remarks

Microblogs for sentiment analysis

- Microblogs as wealthy sources for knowledge about opinions and sentiments

- Twitter 140-character limit forces users to extract content

Sentiment analysis by keyword spotting

- Naïve approach
- How to identify positive, negative, and neutral polarities in subjective sentences? => Classify texts based on the presence of unambiguous sentiment related words
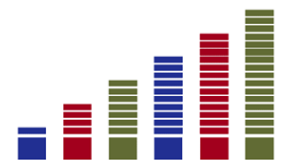- Sentiment polarity for words?

ANEW Spanish lexicon

- Translation of *Affective Norms for English Words*
- Developed at the **University of Santiago de Compostela** (2007)
- 1034 words
- Ratings based on 720 psychology students' assessments
- Conceptualizes emotion as having 3 dimensions:
  - ❑ Valence (from *unpleasant* to *pleasant*)
  - ❑ Arousal (from *calm* to *excited*)
  - ❑ Dominance (from *in control* to *out of control*)
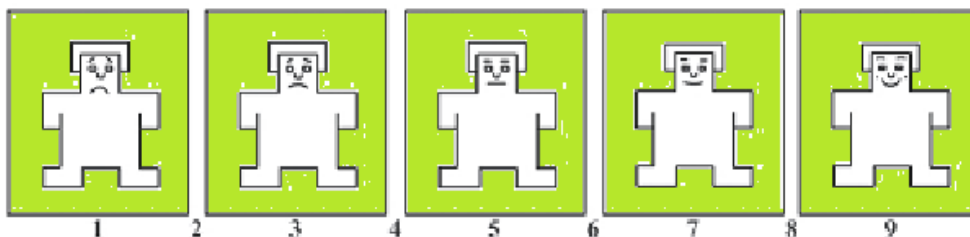
Model to perform sentiment analysis

1. Assignment of sentiment scores to the words in the lexicon

2. Assignment of fuzzy sentiment polarities to tweets

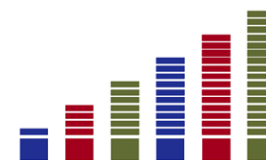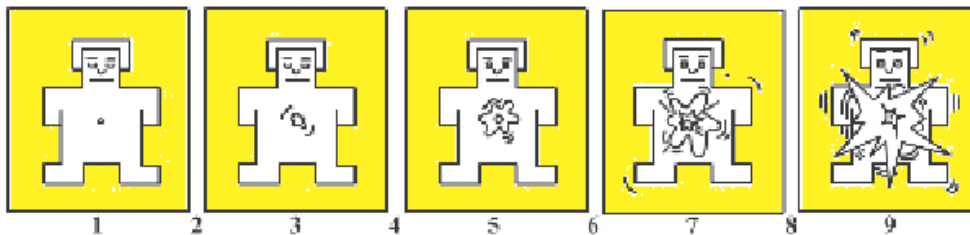3. Computation of a summary sentiment indicator

- **Valence**: level of pleasantness ≅ direction or quality of the sentiment (from *unpleasant:1* to *pleasant:9*)

- **Arousal**: ≅ amount and intensity of the sentiment (from *calm:1* to *excited:9*)

- **X = (Arousal – 1) / 8** => translates to [0, 1] to reflect intensity (from 0 to complete)
- **Y = Valence – 5** => translates to [-4, 4] to reflect opposition negative / positive
- **Score** for each word

$$ s = \frac{X.Y + 4}{8} $$

❑ Varies in [0, 1] interval
❑ Words expressing positive sentiments are assigned higher values
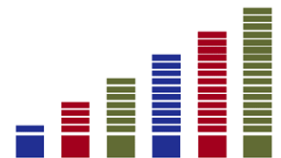❑ Words with negative connotation are assigned lower values
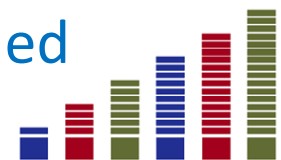
2. Assignment of fuzzy sentiment polarities to tweets

- Aim: score for each tweet reflecting *degree of membership* in the fuzzy category of sentiment
- Firstly pre-processed by removing special characters
- Second: detection of the presence of ANEW words (tweets not including any are discarded)
- Fuzzy sentiment polarity:

$$f(m) = \frac{\sum_{i=1}^{r} s(w_i)}{r}$$

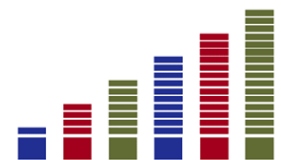(w_i are the words from ANEW detected in m)

- Aim: to obtain information about the trend of the sentiment for the whole Spanish population
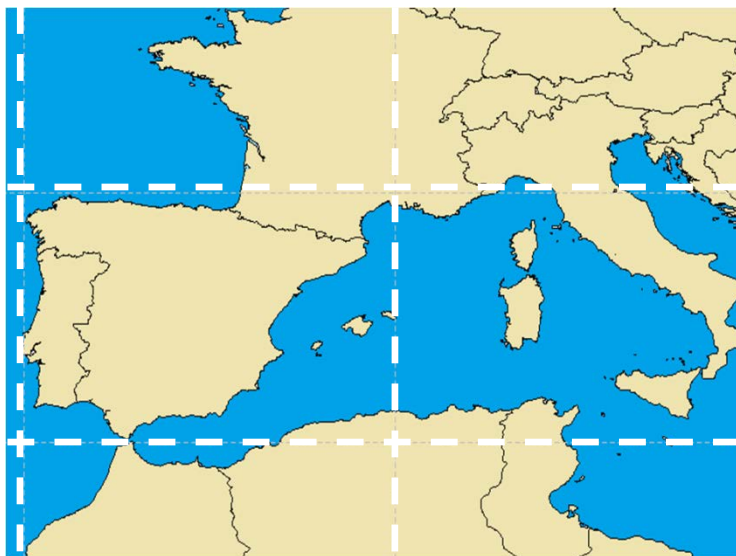
$$I_t = \frac{\sum_{k=1}^{n} f(m_k)}{n}$$

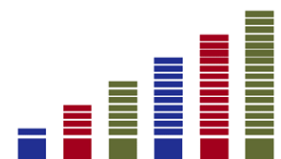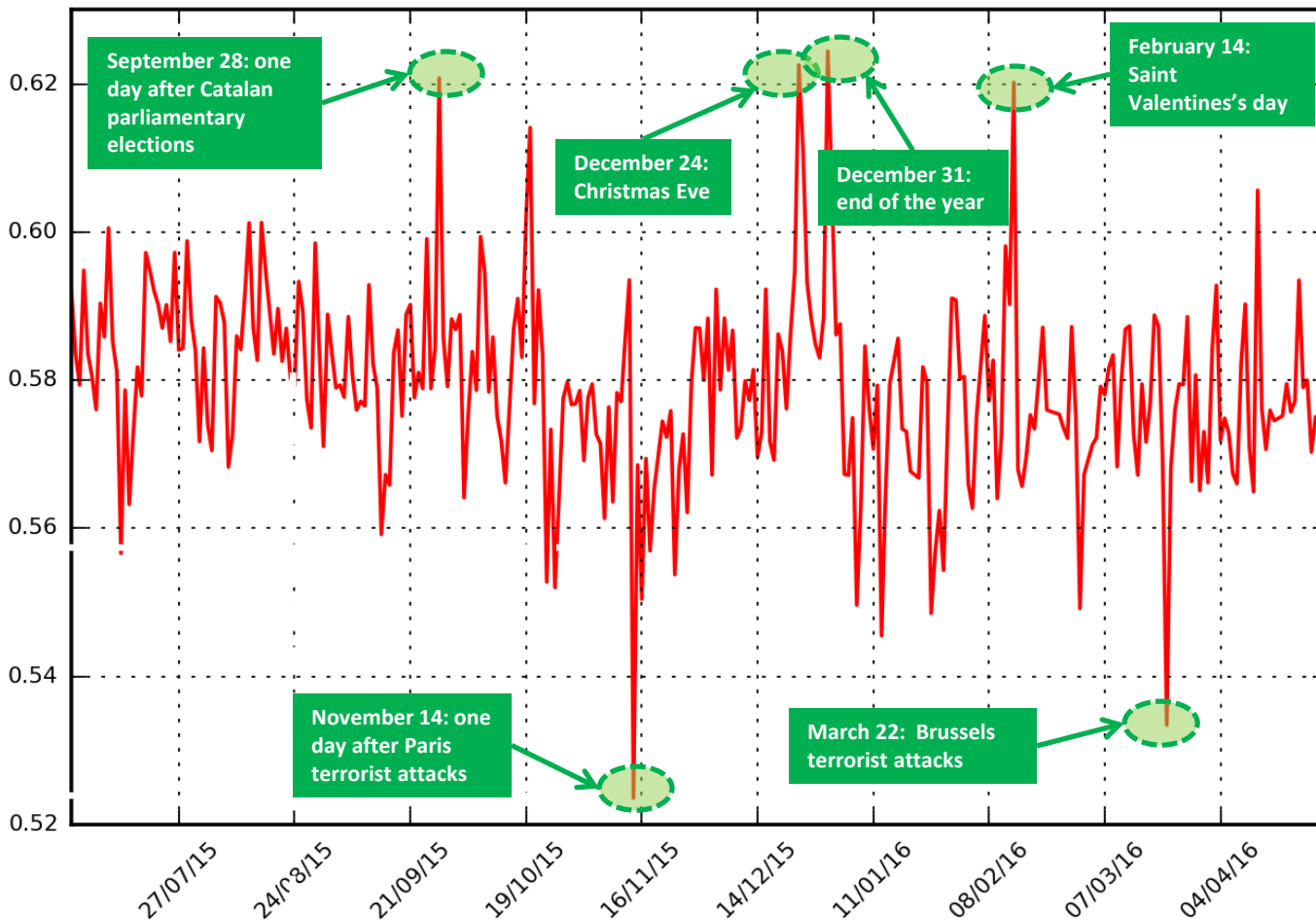where m_k are the tweets in period t

Results using Spanish tweets

- Raw idea of the average general sentiment in the country
- Twitter streaming API: collection of 15 mns tweets at 14:00h

Fig 1. Fuzzy sentiment indicator by day

September 28: one day after Catalan parliamentary elections

December 24: Christmas Eve

December 31: end of the year

February 14: Saint Valentines's day

November 14: one day after Paris terrorist attacks

March 22: Brussels terrorist attacks

Results using Spanish tweets (4)

Geographical area is split into points whose latitude is < and whose latitude is ≥ 40º North
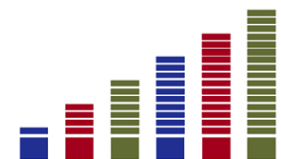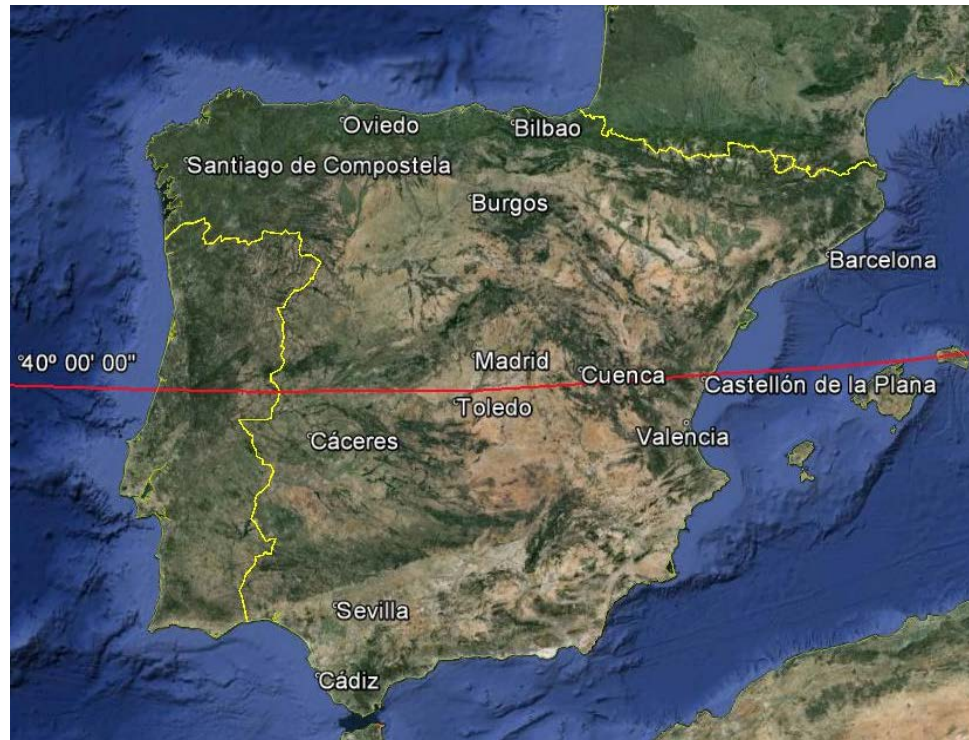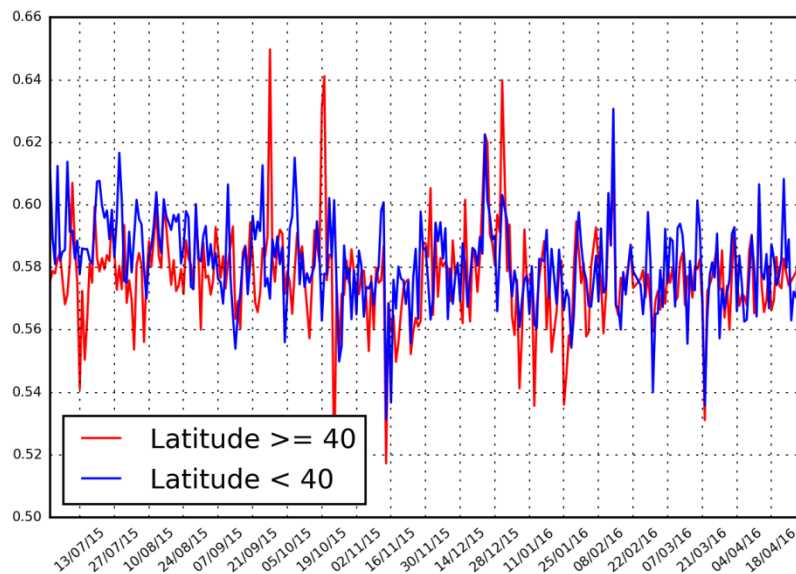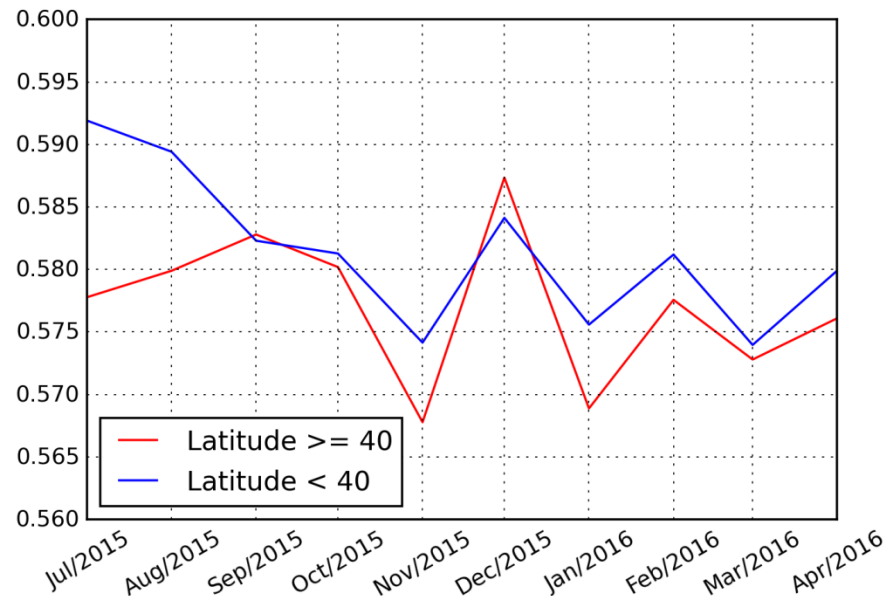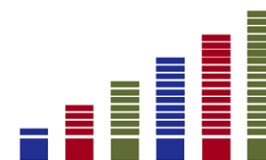
Fig 4. Fuzzy sentiment indicators by day

Fig 5. Fuzzy sentiment indicators by month

- Collective mood is more positive in the less than 40 degrees North area
- In accordance with general understanding: people from the South in Spain are more cheerful and have more positive feelings

14

Final remarks

- The ANEW Spanish lexicon offers a useful tool to perform sentiment analysis
- The procedure provides with indicators of the average mood in the country at an unknown until now time granularity level
- Noise component of the indicators could probably be reduced by increasing the number of collected tweets
- Use in assessing the impact of specific social, political or economic events