

Day 10: Text Mining from Social Media

Ken Benoit

Essex Summer School 2013

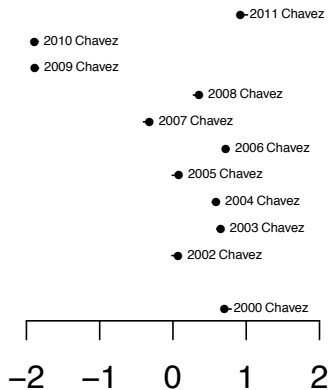
August 2, 2013

Comparability of the Results of Text Analysis

- ▶ Very common objective from analysis: to compare over *time* or *location*
- ▶ Inherently problematic for inductive approaches

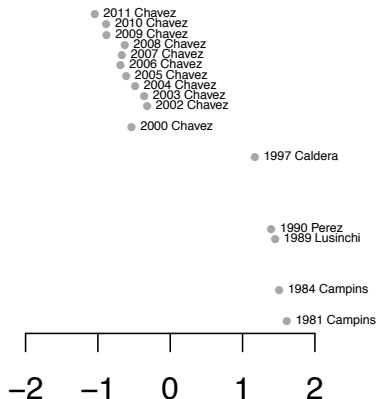
Example of inductive scaling instability

Venezuela



Left-Right Scale

Venezuela



Ideological Space

Solutions to this problem

- ▶ Use a scaling system with stable referents
 - ▶ Supervised learning approaches
 - ▶ Wordscores scaling
 - ▶ Dictionaries
- ▶ Limit your comparisons

Twitter uses: Sentiment analysis

- ▶ A huge industry in analyzing Tweets for business products
- ▶ Often uses dictionaries or machine learning, or a combination
- ▶ Challenge for dictionaries: abbreviations and slang
- ▶ Challenge for machine learning: very short numbers of tokens (in 160 characters)

Twitter uses: Scaling

- ▶ Supervised methods such as Wordscores
- ▶ Unsupervised methods such as CA

Twitter uses: Exploiting the meta-data (non-textual)

- ▶ location
- ▶ time
- ▶ username
- ▶ user descriptions
- ▶ networks of followers
- ▶ retweets of followers and texts

Twitter challenges

- ▶ Large amounts of data
 - ▶ storage problems
 - ▶ analysis problems
- ▶ Language is informal and often non-textual (emoticons, links, images)
- ▶ lots of fake users
- ▶ A lot of the content is non-message oriented e.g.
`http://twitter.com/search?q=%23JamesCallSam`
- ▶ A lot of the content is moronic

Example: Twittdiots



Michael Matthews
@YourBuddyBurns



I'm tired of this terrorist bullshit fucking w our country. Fuck it, just nuke Czechoslovakia

[← Reply](#) [↻ Retweet](#) [★ Favorite](#) [⋮ More](#)



InstrumentalStash
@HashHitz



I Can't believe that pair in the Boston bombing was NOT Towel heads!!! They are Czechoslovakian! Daamn!! FUCK Czechoslovakia!

[← Reply](#) [↻ Retweet](#) [★ Favorite](#) [⋮ More](#)



Kaitlynn Schuler
@KaitlynnSchuler



Some Czech mother fucker is about to get LITTTT up. #gethim

[← Reply](#) [↻ Retweet](#) [★ Favorite](#) [⋮ More](#)



s_elliott11

What did America ever do to the Czech Republic? Where even is the Czech Republic? Have fun with the devil terrorboy

[🕒 2 days ago](#) [← Reply](#) [↻ Retweet](#) [☆ Favorite](#)



Jafar El-Shabazz
@llcooljaff



The media fucked up! They was sayin the suspect was a dark skinned male..turned out to be a Czech republican. ???!?

[← Reply](#) [↻ Retweet](#) [★ Favorite](#) [⋮ More](#)