Trend Miner

Large-scale Cross-lingual Trend Mining Summarization of Real-time Media Streams

TrendMiner: Large-Scale Analysis of Political Attitudes in Public Facebook Messages

Márton Miháltz, Tamás Váradi

RESEARCH INSTITUTE FOR LINGUISTICS
HUNGARIAN ACADEMY OF SCIENCES

CogInfoCom 2015

Introduction

- Sentiment Analysis & Opinion Mining in NLP
  - Objective/subjective, positive/negative assessments, opinions etc.
  - Applications: market research, business analytics, politics, social sciences etc.
  - CogInfoCom: **affective computing** in intra-cognitive communication in textual modality
- Social Media content analysis to assess political attitudes
  - Current approaches: mainly only sentiment analysis
  - Introduce **new, social psychology-motivated measures**
- Use Facebook: dominant SM in Hungary
  - 4.25M users, 59.2% penetration (of people with internet access)
  - Analyze **public comments** for public posts on political pages
- Examine **correlation with traditional opinion poll** results
- **Cooperation** with *Institute of Cognitive Neuroscience And Psychology, MTA* (T. Pólya, É. Fülöp, I. Csertő, P. Kővágó)
- **TrendMiner Project** (FP7)

Data

- **1.9M public comments** (46M words), 141K posts
  - 01.10.2013 – 22.09.2014
  - Facebook Graph API
- **1344 fb** pages:
  - Organizations: political parties, subbranches
  - Persons: members of parliement, election candidates
  - Official and fan pan pages
- 3 categories
  - Hungarian Parliament 2010-2014
  - Hungarian Parliament 2014-2018 + candidates in April 2014
  - Hungarian members of EP 2014-2019 + candidates in 2014
- Sources: valasztas.hu, wikipedia.hu
- TrendMiner Multilingual Political OWL ontology

Processing Pipeline

- **Download comments** from Fb
- **Preprocessing**
  - Tokenization, sentence segmentation
  - Morphological analysis and part-of-speech tagging
  - Lemmatization
- **Named entity recognition**
  - Political actors (persons, organizations)
- **Social psychological content analysis**
  - **Sentiment analysis** + **4 new indicators**: *agency, communion, optimism, individualism*
  - Custom lexicons and grammars (*FSAs with NooJ*)
  - Scores based on counts

Trend Miner

Development of Lexicons

- Corpus: 176K comments, 5.45M words
  - 3500 most frequent words (f>=100)
  - Coding for categories by 6 annotators
  - Compiled into NooJ lexicons and grammars

Trend Miner

Issues in Social Media Text

- **Missing spaces**
  ... end of sentence.Beginning of another ...
- **Multiplicated** punctuation
  first part……. Second part
- **Contracted** words (slang)
  asszem = azt hiszem (“I think”)
- **Consonant** multiplication
  e.g. *pfffffffff, uffffff, ejjjjjjjj (pff(f*), uff(f*), ej(j*))*
- **Emoticons**
  : D
- Frequently **misspelled** words
  **dúrva-durva, *má-már, *enyi-ennyi, *korupt-korrupt, ...*
- **New** words
  traffipax, chipsadó, E-útdíj, MVM, nyugger, lájkol, ...

NLP for Social Media

- Existing NLP tools: different domain
  - Standard languages (newswire)
  - Lower performance on SM text
- Investigation corpus
  - 1.2M comments, 29M words, processing with vanilla NLP tools
  - Unknown tokens f>=15: 14,000 types
  - Manual analysis: common problems, lists of frequent unknown words
- 2-fold approach:
  - **Normalize input** to standard language (pre- & postprocessing)
  - **Adapt tools** to SM language (extend lexicons etc.)

Trend Miner

Identification of Political Actors

- *HunTag* NER tool (Maximum Entropy): low performance on SM text
  - Trained on newswire documents (standard language)
  - Category errors, entity boundary errors
- NooJ lexicon and grammar
  - **Person names**
    family_name (+ given_name), nickname
  - **Names of organizations**
    Official name, abbreviated and acronym form, nicknames
  - **Party affiliations**
  - Based on political **ontology**

Trend Miner

Sentiment Analysis

- Emotional polarity (valency)
  - Positive, negative, neutral
- Lexicon: 500 positive, 420 negative entries
  - content words, multi-word expressions, emoticons
- Negation context rules
- Score: (n_positive – n_negative) / n_tokens

Agency & Communion

- 2 dimensions in social value judgments
  - **Agency**: describes an individual in terms of the efficiency of their behavior oriented to their personal goals: *motivation, competence, control*
  - **Communion**: describes the moral and emotional aspects of an individual's relations to other group members, individuals or groups: *cooperation, social benefit, honesty, self-sacrifice, affection, friendship, respect, love etc.*
- Both have **positive-negative** range
- Lexicon: 650 words and multi-words
- Scores: (n_positive – n_negative) / n tokens

Optimism-pessimism

- Time of events plays a role in individual thinking
  - dominated by **past**: view the world unchangeable
  - dominated by **present**: importance of realistically attainable goals
  - dominated by **future**: sees open possibilities
- Based on PoS and morphology annotations, time expressions
- 2 scores for degree of optimism:
  - Present_verbs / past_verbs
  - Future_verbs / (present_verbs + past_verbs)

Trend Miner

Individualism-collectivism

- Individualism: importance of the category of the **self** when thinking about the world
  - *Individualistic*: focus on actions of self
  - *Collectivist*: focus on actions of groups
- Correlation between usage/omission of personal pronouns (**pronoun drop**) and levels of individualism in societies
- Extended to measure individualism/collectivism in groups
- Score:

  Pers_pronouns / (verbs_with_inflection + nouns_with_inflection)

Evaluation of Annotations

Gold standard corpus

- 1008 comments, 3 annotators
- Stratified random sampling: FIDESZ-KDNP 25.2%, EGYÜTT-2014 19.3%, JOBBIK 19.2%, MSZP 16.6%, DK 12.5%, PM 4.2%, LMP 2.9%

	Precision	Recall	F1
Named entity recognition	98.36%	57.14%	72.29%
Positive sentiment	82.56%	74.50%	77.38%
Negative sentiment	67.03%	53.68%	59.62%
Positive agency	70.59%	69.43%	52.83%
Negative agency	65.79%	25.51%	36.76%
Positive communion	65.75%	82.05%	22.43%
Negative communion	96.39%	13.80%	24.13%

Evaluation of Sentiment Polarity

- For every sentence:

Sentiment := (n_positive – n_negative) / n_tokens

Polarity := -1 if sentiment < 0
0 if sentiment = 0
1 if sentiment > 0

All sentences	**1295**
Baseline: gold standard most freq. polarity (neutral)	920 (71.04%)
Polarity correct	1096 (**84.63%**)

Comparison With Poll Data

- How well do **new measures** indicate **changes in political attitude** during **Hungarian parliamentary elections** in April 2014?
- Traditional public opinion poll data from **Tárki** (party preferences)
- Facebook data:
  - 1.9M comments
  - 6+6 months before/after elections
  - Scores aggregated monthly for each party
  - Assumption: commenters on a party's page are supporters

Individualism, Optimism

- **Individualism** correlates with party popularity over 12 months (r=.22, p=.052)

  Higher individualism - higher responsibility for party choices - higher party popularity

- **Individualism** increased after elections

  Decline of significance of cooperation and unity

- **Optimism** increased after elections only for winning parties

  Different experiences of success and failure

Agency, Communion

- Expected: **negative correlation** between *positive agency*, *negative communion* and party popularity
  - **Intergroup bias***:* overrate in-group & underrate outgroup in intergroup competence or conflict
  - Judge **in-group** through **agency**, **out-group** through **communion**

Agency, Communion Correlations

- 6 M. before elections, correlation w/ party popularity:
  - Negative correlation: pos. agency (r=-.429, p=.05)
  - Negative correlation: agency score (r=-.677, p=.05)
- 6 M. after elections, correlation w/ party popularity:
  - Negative correlation: neg. communion (r=-.574, p=.01)
  - Negative correlation: communion score (r=-454, p=.05)
- Averages for all parties:
  - Pos. agency > neg. agency for 12 M. (p=.001)
  - Neg. communion > pos. communion for 12 M. (p=.001)
  - Pos. agency decreased after elections (p=.01)

Summary

- Collected & analyzed public Fb comments about Hungarian politics
- Introduced 4 novel measures from social psychology
- Adaptated NLP tools to SM language
- Agency and communion measures may be valid for detecting changes in attitudes on social media sites of political groups.
- More information (code, data, papers):
  http://corpus.nytud.hu/trendminer

Trend Miner

Thank You!

http://corpus.nytud.hu/trendminer