Exploiting Human-Generated Text for Trend Mining

Vasileios Lampos

July, 2013

Outline

- $\perp$ **Motivation, Aims** [Facts, Questions]
- $\perp$ **Data**
- $\dashv$ **Nowcasting Events**
- $\dashv$ **Extracting Mood Patterns**
- $\dashv$ **Inferring Voting Intention**
- $\models$ **Conclusions**

Facts

We started to work on those ideas *back* in 2008, when...

- **Web** contained **1 trillion** unique pages (Google)
- **Social Networks** were rising, *e.g.*
  - *Facebook*: 100m (2008) → >**1.11 billion** active users (March, 2013)
  - *Twitter*: 6m (2008) → **554m** active users (July, 2013)
- New technologies to handle **Big Data** (*e.g.*, Map-Reduce)
- **User behaviour** was changing
  - Socialising via the Web
  - Giving up privacy (Debatin *et al.*, 2009)

General questions/aims

- Does human-generated text posted on web platforms (or elsewhere) contain **useful information**?

- How can we **extract** this information...

  ... **automatically**? Therefore, not we, but a **machine**.

- Practical / real-life **applications**?

- Can those large samples of human input **assist studies in other scientific fields**?
  Social Sciences, *Psychology*, *Epidemiology*

The Data (1/3) – Why Twitter?

Twitter...

- has a lot of content that is **publicly accessible**
- provides a well-documented **API** for several forms of data collection
- contains **opinions** and **personal statements** on various domains
- is ‘connected’ with current affairs (usually in **real-time**)
- includes **geo-located** content
- offers the option for personalised, **per-user modelling**

What does a @tweet look like?

Figure 1: Some biased and anonymised examples of tweets (limit of **140 characters**/tweet, # denotes a **topic**)

Why do I feel so happy today hihi. Bedtimeeee, good night. Yey thank You Lord for everything. Answered prayer ♥

Reply Retweet Favorite

(a) (user will remain anonymous)

RT if you love Justin Bieber. Delete ur account if you don't.

Reply Retweet Favorite

50 RETWEETS 1 FAVORITE

(b) they live around us

another demo covered by citizens today in Thessaloniki int'l fair. Citizen journalism on a speed rise in #Greece. check #deth and #rbnews

Reply Retweet Favorite

(c) citizen journalism

i think i have the flu but i still look fabulous

Reply Retweet Favorite

(d) flu attitude

The Data (3/3)

Data Collection & Preprocessing

- The easiest part of the process...
  - **not true**! → Storage space, crawler implementation, parallel data processing, adapt to new technologies
- Data collected via **Twitter's Search API**:
  - **collective sampling**
  - tweets geo-located in 54 urban centres in the UK
  - periodical crawling (every 3 or 5 minutes per urban centre)
- Data collected via **Twitter's REST API**:
  - **user-centric sampling**
  - preprocessing to approximate user's location (city & country)
  - ... or manual user selection from domain experts
  - get their latest tweets (3,000 or more)
- Several forms of **ground truth** (flu/rainfall rates, polls)

Nowcasting Events from the Social Web

‘Nowcasting’?

We do not predict the future, but **infer the present** $-\ \delta$

i.e. the very recent past

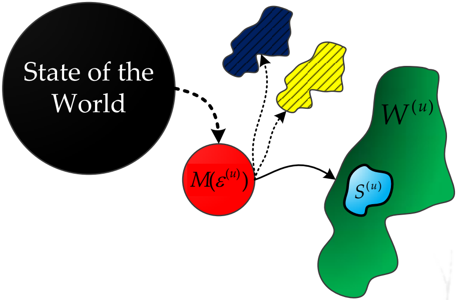


Figure 2: Nowcasting the magnitude of an event (ε) emerging in the real world from Web information

Our case studies: nowcasting (a) **flu rates** & (b) **rainfall rates (?!)**

What do we get in the end?

This is a **regression** problem

i.e. $\forall$ time interval i we aim to infer $y_i \in \mathbb{R}$ using text input $\boldsymbol{x}_i \in \mathbb{R}^n$

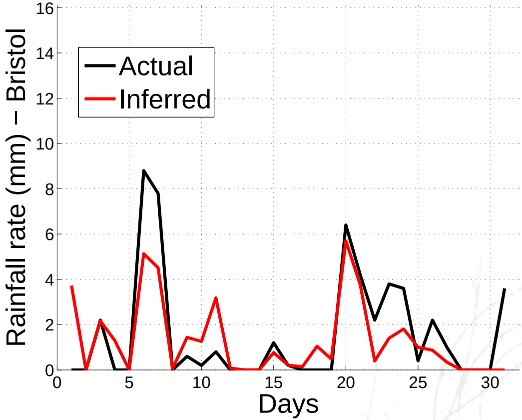


Figure 3: Inferred **rainfall rates** for Bristol, UK (October, 2009)

Methodology (1/5) — Text in Vector Space

Candidate features (n-grams): $\mathcal{C} = \{c_i\}$

Set of **Twitter posts** for a time interval u: $\mathcal{P}^{(u)} = \{p_j\}$

Frequency of c_i in p_j:

$$g(c_i, p_j) = \begin{cases} \varphi & \text{if } c_i \in p_j, \\ 0 & \text{otherwise.} \end{cases}$$

– g Boolean, maximum value for φ is 1 –

Score of c_i in $\mathcal{P}^{(u)}$:

$$s\left(c_i, \mathcal{P}^{(u)}\right) = \frac{\sum\limits_{j=1}^{|\mathcal{P}^{(u)}|} g(c_i, p_j)}{\left|\mathcal{P}^{(u)}\right|}$$

Methodology (2/5)

Set of **time intervals**: $\mathcal{U} = \{u_k\} \sim$ 1 hour, 1 day, ...

Time series of candidate features **scores**:

$$X^{(\mathcal{U})} = \left[\boldsymbol{x}^{(u_1)} \; \ldots \; \boldsymbol{x}^{(u_{|\mathcal{U}|})}\right]^{\mathsf{T}},$$

where

$$\boldsymbol{x}^{(u_i)} = \left[s\left(c_1, \mathcal{P}^{(u_i)}\right) \; \ldots \; s\left(c_{|\mathcal{C}|}, \mathcal{P}^{(u_i)}\right)\right]^{\mathsf{T}}$$

Target variable (event):

$$\boldsymbol{y}^{(\mathcal{U})} = \left[y_1 \; \ldots \; y_{|\mathcal{U}|}\right]^{\mathsf{T}}$$

Methodology (3/5) — Feature selection

Solve the following **optimisation problem**:

$$\min_{w} \quad \|X^{(\mathcal{U})}\boldsymbol{w} - \boldsymbol{y}^{(\mathcal{U})}\|_{\ell_2}^2$$

$$\text{s.t.} \quad \|\boldsymbol{w}\|_{\ell_1} \leq t,$$

$$t = \alpha \cdot \|\boldsymbol{w}_{\text{OLS}}\|_{\ell_1}, \;\; \alpha \in (0, 1].$$

- Least Absolute Shrinkage and Selection Operator (**LASSO**)

$$\underset{\boldsymbol{w}}{\operatorname{argmin}} \|X^{(\mathcal{U})}\boldsymbol{w} - \boldsymbol{y}^{(\mathcal{U})}\|_{\ell_2}^2 + \lambda\|\boldsymbol{w}\|_{\ell_1}$$

(Tibshirani, 1996)

- Expect a **sparse** $\boldsymbol{w}$ (feature selection)
- Least Angle Regression (**LARS**) – computes entire regularisation path ($\boldsymbol{w}$'s for different values of λ) (Efron *et al.*, 2004)

Methodology (4/5)

LASSO is **model-inconsistent**:

- inferred sparsity pattern may deviate from the true model, *e.g.*, when predictors are highly correlated (Zhao and Yu, 2006)

- bootstrap [**?**] LASSO (**Bolasso**) performs a more robust feature selection (Bach, 2008)
  ?:
  - in each bootstrap, input space is sampled with replacement
  - apply LASSO (LARS) to select features
  - select features with nonzero weights in all bootstraps

- better alternative — **soft-Bolasso**:
  - a less strict feature selection
  - select features with nonzero weights in p% of bootstraps
  - (learn p using a separate validation set)

- **weights** of selected features determined via OLS regression

Methodology (5/5) — Simplified summary

Observations: $X \in \mathbb{R}^{m \times n}$ (m time intervals, n features)

Response variable: $\boldsymbol{y} \in \mathbb{R}^{m}$

For $i = 1$ to *number of bootstraps*
    Form $X_i \subset X$ by sampling X with replacement
    Solve LASSO for X_i and $\boldsymbol{y}$, *i.e.* learn $\boldsymbol{w}_i \in \mathbb{R}^{n}$
    Get the $k \leq n$ features with nonzero weights
End_For

Select the $v \leq n$ features with nonzero weight in $p\%$ of the bootstraps

Learn their weights with OLS regression on $X^{(v)} \in \mathbb{R}^{m \times v}$ and $\boldsymbol{y}$

How do we form candidate features?

- Commonly formed by indexing the **entire corpus** (Manning, Raghavan and Schütze, 2008)
- We extract them from Wikipedia, Google Search results, Public Authority websites (*e.g.*, NHS)

  Why?

  - reduce **dimensionality** to bound the error of LASSO

    $$\mathcal{L}(\boldsymbol{w}) \leq \mathcal{L}(\hat{\boldsymbol{w}}) + \mathcal{Q}, \text{ with } \mathcal{Q} \sim \min\left\{\frac{W_1^2}{N} + \frac{p}{N}, \frac{W_1^2}{N} + \frac{W_1}{\sqrt{N}}\right\}$$

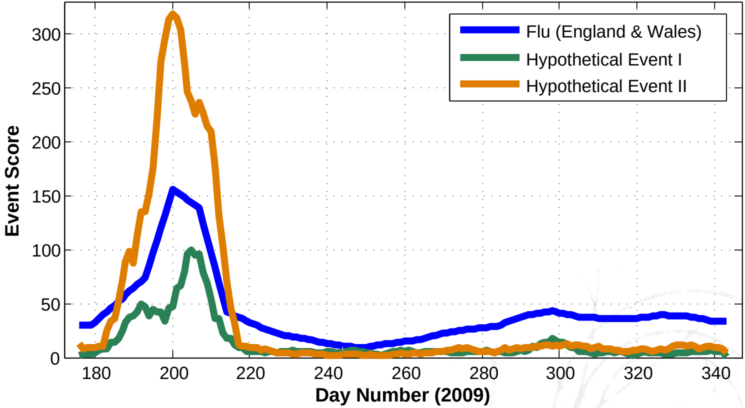
    p candidate features, N samples, empirical loss $\mathcal{L}(\hat{\boldsymbol{w}})$ and

    $\|\hat{\boldsymbol{w}}\|_{\ell_1} \leq W_1$ (Bartlett, Mendelson and Neeman, 2011)

  - **Harry Potter Effect!**

The 'Harry Potter' effect (1/2)

Figure 4: Events co-occurring (*correlated*) with the inference target may affect feature selection, especially when the sample size is small.



(Lampos, 2012a)

The 'Harry Potter' effect (2/2)

Table 1: Top 1-grams correlated with flu rates in England/Wales (06–12/2009)

1-gram	Event	Corr. Coef.
latitud	Latitude Festival	0.9367
flu	Flu epidemic	0.9344
swine	▲	0.9212
harri	Harry Potter Movie	0.9112
slytherin	▲	0.9094
potter	▲	0.8972
benicassim	Benicàssim Festival	0.8966
graduat	Graduation (?)	0.8965
dumbledor	Harry Potter Movie	0.8870
hogwart	▲	0.8852
quarantin	Flu epidemic	0.8822
gryffindor	Harry Potter Movie	0.8813
ravenclaw	▲	0.8738
princ	▲	0.8635
swineflu	Flu epidemic	0.8633
ginni	Harry Potter Movie	0.8620
weaslei	▲	0.8581
hermion	▲	0.8540
draco	▲	0.8533

Solution: ground truth with some degree of variability

(Lampos, 2012a)

About n-grams

1-grams

- decent (dense) representation in the Twitter corpus
- unclear semantic interpretation
  Example: "*I am not sick. But I don't feel great either!*"

2-grams

- very sparse representation in tweets
- sometimes clearer semantic interpretation

Experimental process indicated that...

a **hybrid combination*** of **1**-grams and **2**-grams
delivers the best inference performance

* refer to (Lampos, 2012a)

Flu rates – Example of selected features

Figure 5: Font size is proportional to the weight of each feature; flipped n-grams are negatively weighted. All words are stemmed (Porter, 1980).

(Lampos and Cristianini, 2012)

Rainfall rates – Example of selected features

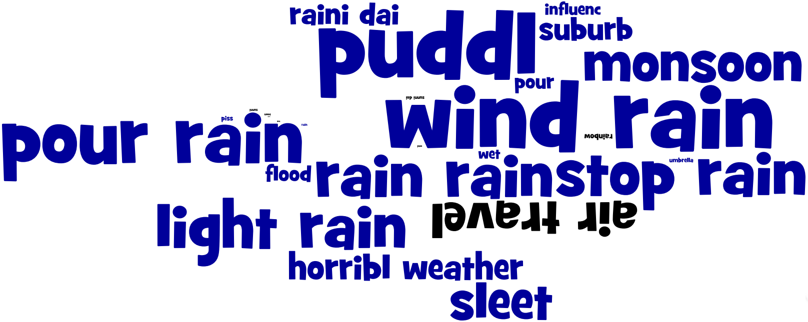
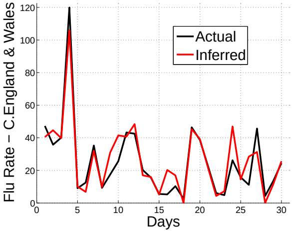


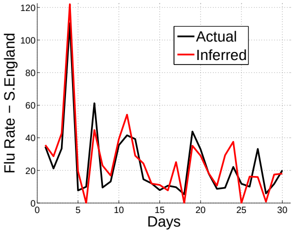
Figure 6: Font size is proportional to the weight of each feature; flipped n-grams are negatively weighted. All words are stemmed (Porter, 1980).

(Lampos and Cristianini, 2012)

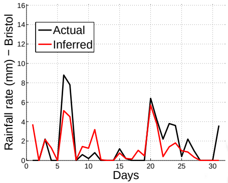
Examples of inferences



(a) Central England/Wales (flu)



(b) South England (flu)



(c) Bristol (rain)

Figure 7: Examples of flu and rainfall rates **inferences** from Twitter content (Lampos and Cristianini, 2012)

Performance figures

Table 2: RMSE for **flu rates** inference (5-fold cross validation), 50m tweets, 21/06/2009–19/04/2010

Method	**1-grams**	**2-grams**	**Hybrid**
Baseline*	12.44±2.37	13.81±3.29	11.62±1.58
Bolasso	**11.14±2.35**	**12.64±2.57**	**10.57±2.2**
CART ensemble**	**9.63**±5.21	13.13±4.72	**9.4**±4.21

Table 3: RMSE (in *mm*) for **rainfall rates** inference (6-fold cross validation), 8.5m tweets, 01/07/2009–30/06/2010

Method	**1-grams**	**2-grams**	**Hybrid**
Baseline*	2.91±0.6	3.1±0.57	4.39±2.99
Bolasso	**2.73±0.65**	2.95±0.55	**2.60**±0.68
CART ensemble**	**2.71**±0.69	**2.72±0.72**	**2.64±0.63**

* As implemented in (Ginsberg *et al.*, 2009)

** Classification and Regression Tree (Breiman *et al.*, 1984) & (Sutton, 2005)

Flu Detector

URL: `http://geopatterns.enm.bris.ac.uk/epidemics`

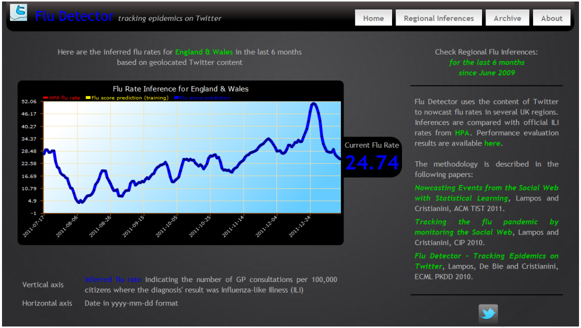


Figure 8: Flu Detector uses the content of Twitter to nowcast flu rates in several UK regions

(Lampos, De Bie and Cristianini, 2010)

Extracting Mood Patterns from Human-Generated Content

Computing a mood score

Table 4: **Mood terms** from **WordNet Affect**

Fear	Sadness	Joy	Anger
afraid	depressed	admire	angry
fearful	discouraged	cheerful	despise
frighten	disheartened	enjoy	enviously
horrible	dysphoria	enthousiastic	harassed
panic	gloomy	exciting	irritate
...	...	...	...
(*92 terms*)	(*115 terms*)	(*224 terms*)	(*146 terms*)

Mood score computation for a time interval d using n mood terms

$$\mathrm{ms}_d = \frac{1}{n} \sum_{i=1}^{n} \frac{c_i^{(t_d)}}{N(t_d)}$$

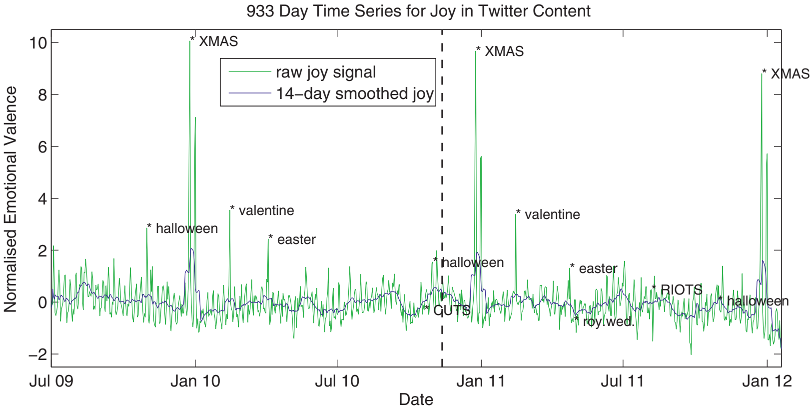
$c_i^{(t_d)}$: count of term i in the Twitter corpus of day d
$N(t_d)$: number of tweets for day d

Using the sample of d days, compute a standardised mood score:

$$\mathrm{ms}_d^{\mathrm{std}} = \frac{\mathrm{ms}_d - \mu_{\mathrm{ms}}}{\sigma_{\mathrm{ms}}}$$

The mood of the nation (1/5)

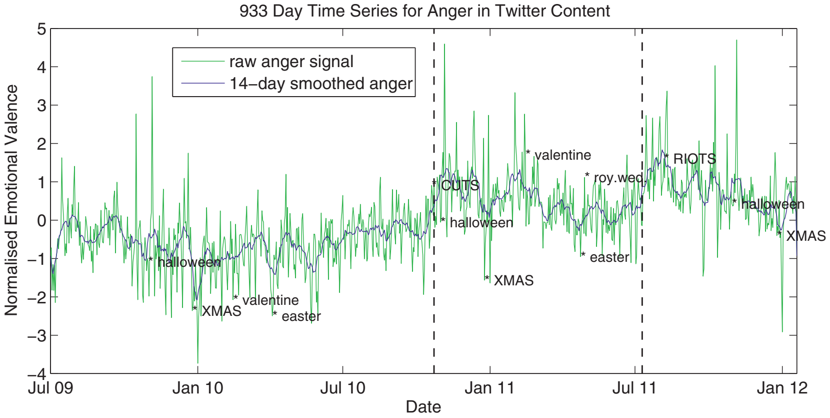
Figure 9: Daily time series (actual & their 14-point moving average) for the mood of **Joy** based on Twitter content geo-located in the **UK**



(Lansdall-Welfare, Lampos and Cristianini, 2012a&b)

The mood of the nation (2/5)

Figure 10: Daily time series (actual & their 14-point moving average) for the mood of **Anger** based on Twitter content geo-located in the **UK**



(Lansdall-Welfare, Lampos and Cristianini, 2012a&b)

The mood of the nation (3/5)

Window of **100 days**: 50 before & after the point of interest

$$\mathrm{ms}_i^{\mathrm{std}} = \mu\left(\mathbf{ms}_{i+1\to i+50}^{\mathrm{std}}\right) - \mu\left(\mathbf{ms}_{i-50\to i-1}^{\mathrm{std}}\right)$$

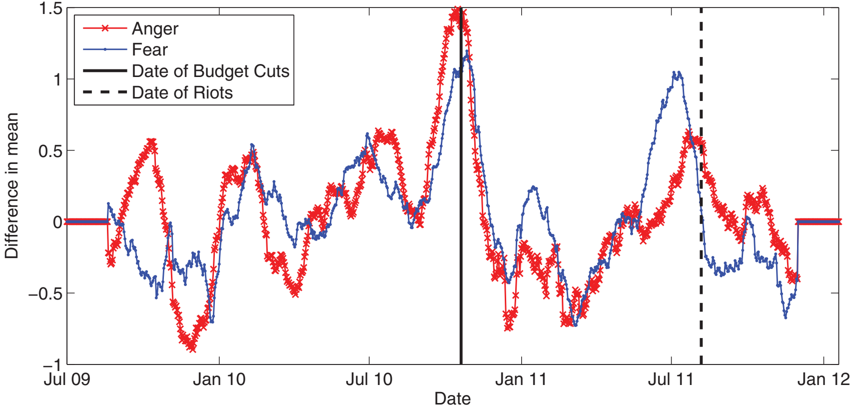
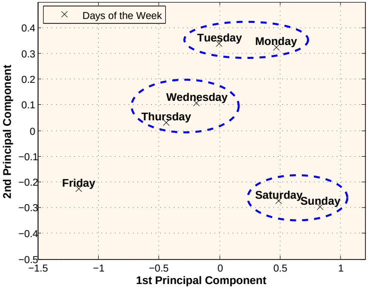


Figure 11: Change point detection using a 100-day moving window

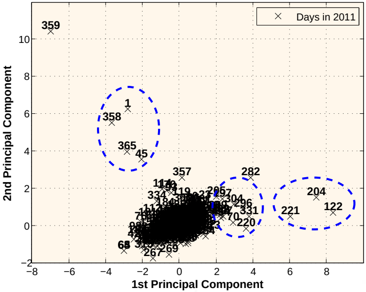
(Lansdall-Welfare, Lampos and Cristianini, 2012a)

The mood of the nation (4/5)

Figure 12: Projections of **4-dimensional mood score signals** (joy, sadness, anger and fear) on their **top-2 principal components** (PCA) – Twitter content from 2011



(a) Days of the week (2011)



(b) Days of the year (2011)

Cluster I

New Year (**1**), Valentine's (**45**), Christmas Eve (**358**), New Year's Eve (**365**)

Cluster II

O.B. Laden's death (**122**), Winehouse's death + Breivik (**204**), UK riots (**221**)

(Lampos, 2012a)

The mood of the nation (5/5)

URL: `http://geopatterns.enm.bris.ac.uk/mood`

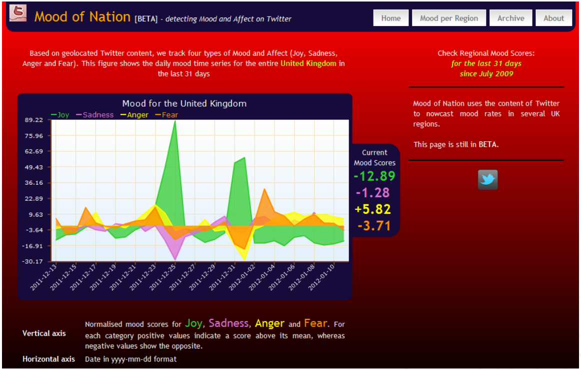


Figure 13: Mood of the Nation uses the content of Twitter to nowcast mood rates in several UK regions

(Lampos, 2012a)

Circadian mood patterns (1/3)

Compute **24-h** mood score patterns

Mood score computation for a **time interval** $u = 24$hours using n **mood terms** (WordNet) and a sample of D **days**:

$$\mathcal{M}_s(u) = \frac{1}{|D|} \sum_{j=1}^{|D|} \left(\frac{1}{n} \sum_{i=1}^{n} sf_i^{(t_{j,u})} \right)$$

$$sf_i^{(t_{d,u})} = \frac{f_i^{(t_{d,u})} - \bar{f}_i}{\sigma_{f_i}}, \;\; i \in \{1, ..., n\}.$$

$f_i^{(t_{d,u})}$: normalised frequency of a mood term i during time interval u in day $d \in D$

Circadian mood patterns (2/3)

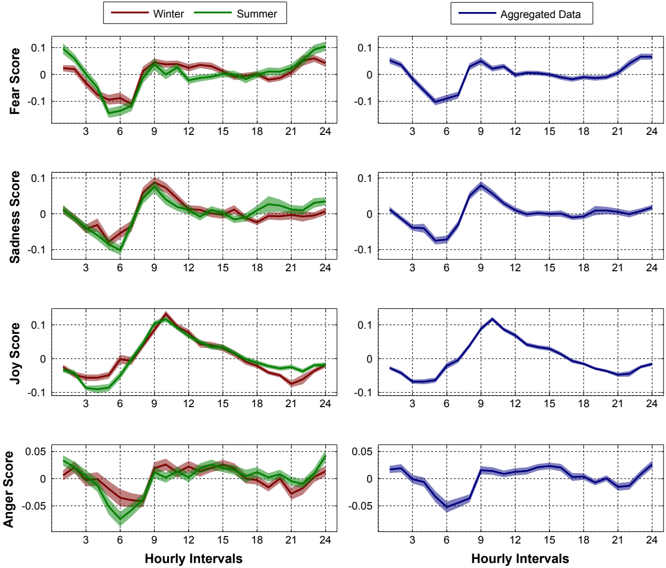
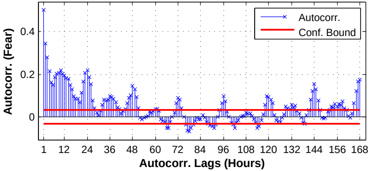


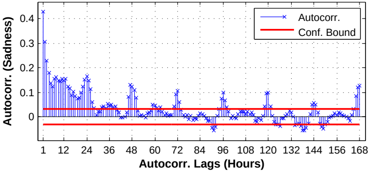
Figure 14: Circadian (**24-hour**) mood patterns based on UK Twitter content

Circadian mood patterns (3/3)

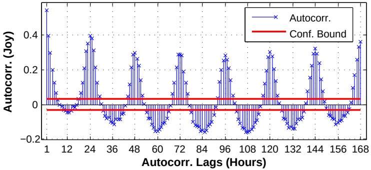
Figure 15: **Autocorrelation** of circadian mood patterns based on **hourly lags** revealing daily and weekly periodicities



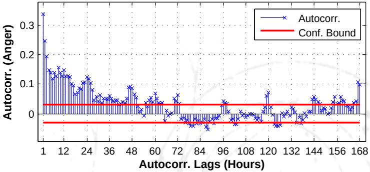
(a) Fear



(b) Sadness



(c) Joy

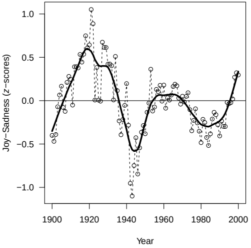


(d) Anger

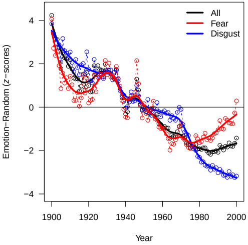
Further analysis available in (Lampos, Lansdall-Welfare, Araya and Cristianini, 2013)

Emotion in Books

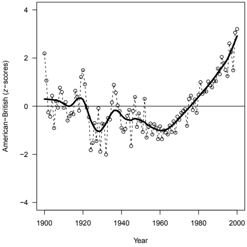
Input: **Google Ngram corpus** of $\sim$ 5m digitised books (Michel *et al.*, 2010)
Tool: WordNet Affect (Strapparava and Valitutti, 2004)



(a) Joy minus Sadness



(b) Use of emotion-related terms through time



(c) American versus British English

Figure 16: Emotion trends in 20th century books

(Acerbi, Lampos, Garnett and Bentley, 2013)

Inferring Voting Intention from Social Media Content

... and a new way for modelling text regression

Motivations and Aims

- Social Media contain a **vast amount of information** about various topics (health, politics, finance)
- This information (X) can be used to assist **predictions** (y)
- $f : X \rightarrow y$, f usually formulates a **linear** regression task
- X accounts only for word frequencies; can we incorporate **user information** as well?
- Could we also exploit the statistical information held in **multiple response variables**?

Data Sets

UK case study

- **60m tweets** by **42K users** from 30/04/2010 to 13/02/2012
- Random selection and distribution of geo-located users proportional to regional population figures
- Main language: **English**
- **240** unique voting intention **polls** from YouGov
  - percentages for Conservatives (**CON**), Labour Party (**LAB**) and Liberal Democrats (**LIB**)

Austrian case study

- **800K tweets** by **1.1K users** from 25/01 to 01/12/2012
- Users manually selected by Austrian political analysts
- Main language: **German**
- **98** unique voting intention **polls** from various pollsters
  - percentages for Social Democratic Party (**SPÖ**), People's Party (**ÖVP**), Freedom Party (**FPÖ**) and Green Alternative Party (**GRÜ**)

The Bilinear Model (1/2)

The **main idea** is simple:

$$f(X) = \boldsymbol{u}^{\mathsf{T}} X \boldsymbol{w} + \beta$$

$X \in \mathbb{R}^{m \times p}$: matrix of user-word frequencies
$\boldsymbol{u}$, $\boldsymbol{w}$: user and word weights

Our original **bilinear text regression model**:

$$\{\boldsymbol{w}^*, \boldsymbol{u}^*, \beta^*\} = \underset{\boldsymbol{w},\boldsymbol{u},\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \left(\boldsymbol{u}^{\mathsf{T}} \mathcal{Q}_i \boldsymbol{w} + \beta - y_i \right)^2 + \psi(\boldsymbol{w}, \rho_1) + \psi(\boldsymbol{u}, \rho_2)$$

$\mathcal{Q}_i$: X for time instance i, $\boldsymbol{y} \in \mathbb{R}^n$: response variable (voting intention)
$\boldsymbol{w} \in \mathbb{R}^m$, $\boldsymbol{u} \in \mathbb{R}^p$: word and user weights, $\beta \in \mathbb{R}$: bias
$\psi(\cdot)$: a regularisation function

Elastic Net (Zhou and Hastie, 2005) for $\psi(\cdot)$
$\rightarrow$ Bilinear Elastic Net (**BEN**) (Lampos, Preoţiuc-Pietro and Cohn, 2013)

The Bilinear Model – Multi-Task Learning (2/2)

Apply ℓ_1/ℓ_2 regulariser (Argyriou, Evgeniou and Pontil, 2008)
Extends the notion of **Group LASSO** (Yuan and Lin, 2006) for a τ-dimensional $\boldsymbol{y}$

Bilinear Group ℓ_1/ℓ_2 (**BGL**)

$$\{W^*, U^*, \boldsymbol{\beta}^*\} = \underset{W,U,\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{t=1}^{\tau} \sum_{i=1}^{n} \left(\boldsymbol{u}_t^{\mathrm{T}} \mathcal{Q}_i \boldsymbol{w}_t + \beta_t - y_{ti} \right)^2 + \lambda_1 \sum_{j=1}^{m} \|W_j\|_2 + \lambda_2 \sum_{k=1}^{p} \|U_k\|_2,$$

$W = [\boldsymbol{w}_1 \ \ldots \ \boldsymbol{w}_\tau]$: words weight matrix – $\boldsymbol{w}_t$ refers to t-th political entity
$U = [\boldsymbol{u}_1 \ \ldots \ \boldsymbol{u}_\tau]$: users weight matrix
W_j, U_j: j-th rows of weight matrices W and U respectively
$\boldsymbol{\beta} \in \mathbb{R}^\tau$: bias term per task

(Lampos, Preoțiuc-Pietro and Cohn, 2013)

Evaluation – Performance Tables (1/2)

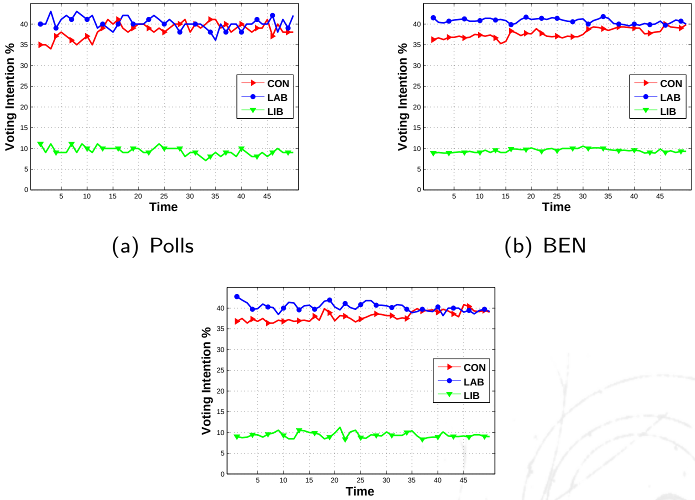
Table 5: UK case study — Average RMSEs representing the error of the inferred voting intention percentage for a 10-step validation process

	CON	LAB	LIB	μ
B_μ	2.272	1.663	1.136	1.69
B_{last}	2	2.074	1.095	1.723
LEN	3.845	2.912	2.445	3.067
BEN	1.939	1.644	1.136	1.573
BGL	**1.785**	**1.595**	**1.054**	**1.478**

Table 6: Austrian case study

	SPÖ	ÖVP	FPÖ	GRÜ	μ
B_μ	1.535	1.373	3.3	1.197	1.851
B_{last}	**1.148**	1.556	**1.639**	1.536	1.47
LEN	1.291	1.286	2.039	**1.152**	1.442
BEN	1.392	1.31	2.89	1.205	1.699
BGL	1.619	**1.005**	1.757	1.374	**1.439**

Evaluation (2/3)



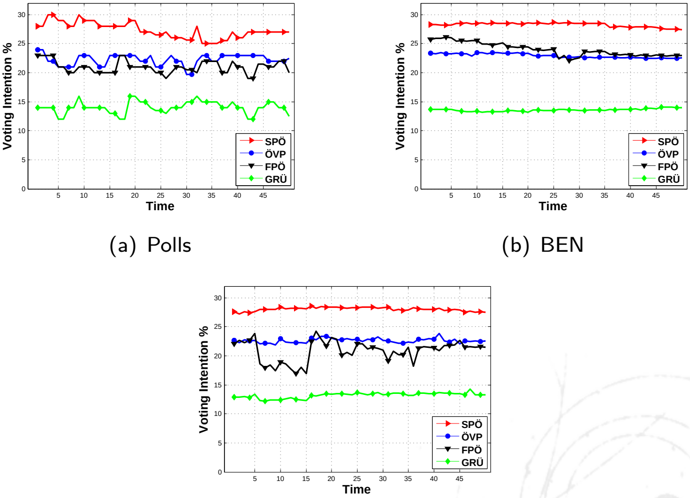
(a) Polls

(b) BEN

(c) BGL

Figure 17: UK case study — 50 consecutive poll predictions

Evaluation (3/3)



(a) Polls (b) BEN (c) BGL

Figure 18: Austrian case study — 50 consecutive poll predictions

Conclusions

- **Social Media** hold **valuable information**
- We can develop **methods** to extract portions of this information **automatically**
  - **detect, quantify, nowcast events** (examples of flu and rainfall rates)
  - extract **collective mood** patterns (we can do this for books too!)
  - model other domains (such as **politics**)
- Different types of information (word frequencies, user accounts) can be fused for improved inference performance
- Side effect: **user privacy**

Significant collaborators...

Prof. Nello Cristianini, University of Bristol (*Artificial Intelligence*)

Prof. Alexander Bentley, University of Bristol (*Anthropology*)

Dr. Trevor Cohn, University of Sheffield (*Natural Language Processing*)

Dr. Alberto Acerbi, University of Bristol (*Anthropology*)

Daniel Preoţiuc-Pietro, University of Sheffield (*Computer Science*)

The end.

Any questions?

Download the slides from
http://www.lampos.net/research/presentations-and-posters

References

- Acerbi, Lampos, Garnett and Bentley. **The Expression of Emotions in 20th Century Books**. PLoS ONE, 2013.
- Argyriou, Evgeniou and Pontil. **Convex multi-task feature learning**. Machine Learning, 2008.
- Bach. **Bolasso: Model Consistent Lasso Estimation through the Bootstrap**. ICML, 2008.
- Bartlett, Mendelson and Neeman. **L1-regularized linear regression: persistence and oracle inequalities**. PTRF, 2011.
- Debatin, Lovejoy, Horn and Hughes. **Facebook and Online Privacy: Attitudes, Behaviors, and Unintended Consequences**. JCMC, 2009.
- Efron *et al.*. **Least Angle Regression**. The Annals of Statistics, 2004.
- Ginsberg *et al.* **Detecting influenza epidemics using search engine query data**. Nature, 2009.
- Lampos and Cristianini. **Tracking the flu pandemic by monitoring the Social Web**. CIP, 2010.
- Lampos, De Bie and Cristianini. **Flu Detector – Tracking Epidemics on Twitter**. ECML PKDD, 2010.
- Lampos and Cristianini. **Nowcasting Events from the Social Web with Statistical Learning**. ACM TIST, 2012.
- Lampos. **Detecting Events and Patterns in Large-Scale User Generated Textual Streams with Statistical Learning Methods**. Ph.D. Thesis, University of Bristol, 2012.(a)
- Lampos. **On voting intentions inference from Twitter content: a case study on UK 2010 General Election**. CoRR, 2012.(b)
- Lampos, Preoţiuc-Pietro and Cohn. **A user-centric model of voting intention from Social Media**. ACL, 2013.
- Lampos, Lansdall-Welfare, Araya and Cristianini. **Analysing Mood Patterns in the United Kingdom through Twitter Content**. CoRR, 2013.
- Lansdall-Welfare, Lampos and Cristianini. **Effects of the Recession on Public Mood in the UK**. WWW, 2012.(a)
- Lansdall-Welfare, Lampos and Cristianini. **Nowcasting the mood of the nation**. Significance, 2012.(b)
- Manning, Raghavan and Schütze. **Introduction to Information Retrieval**, 2008.
- Michel *et al.* **Quantitative Analysis of Culture Using Millions of Digitized Books**. Nature, 2010.
- Porter. **An algorithm for suffix stripping**. Program, 1980.
- Strapparava and Valitutti. **WordNet-Affect: an affective extension of WordNet**. LREC, 2004.
- Tibshirani. **Regression Shrinkage and Selection via the LASSO**. JRSS, 1996.
- Yuan and Lin. **Model selection and estimation in regression with grouped variables**. JRSS, 2006.
- Zhao and Yu. **On model selection consistency of LASSO**. JMLR, 2006.