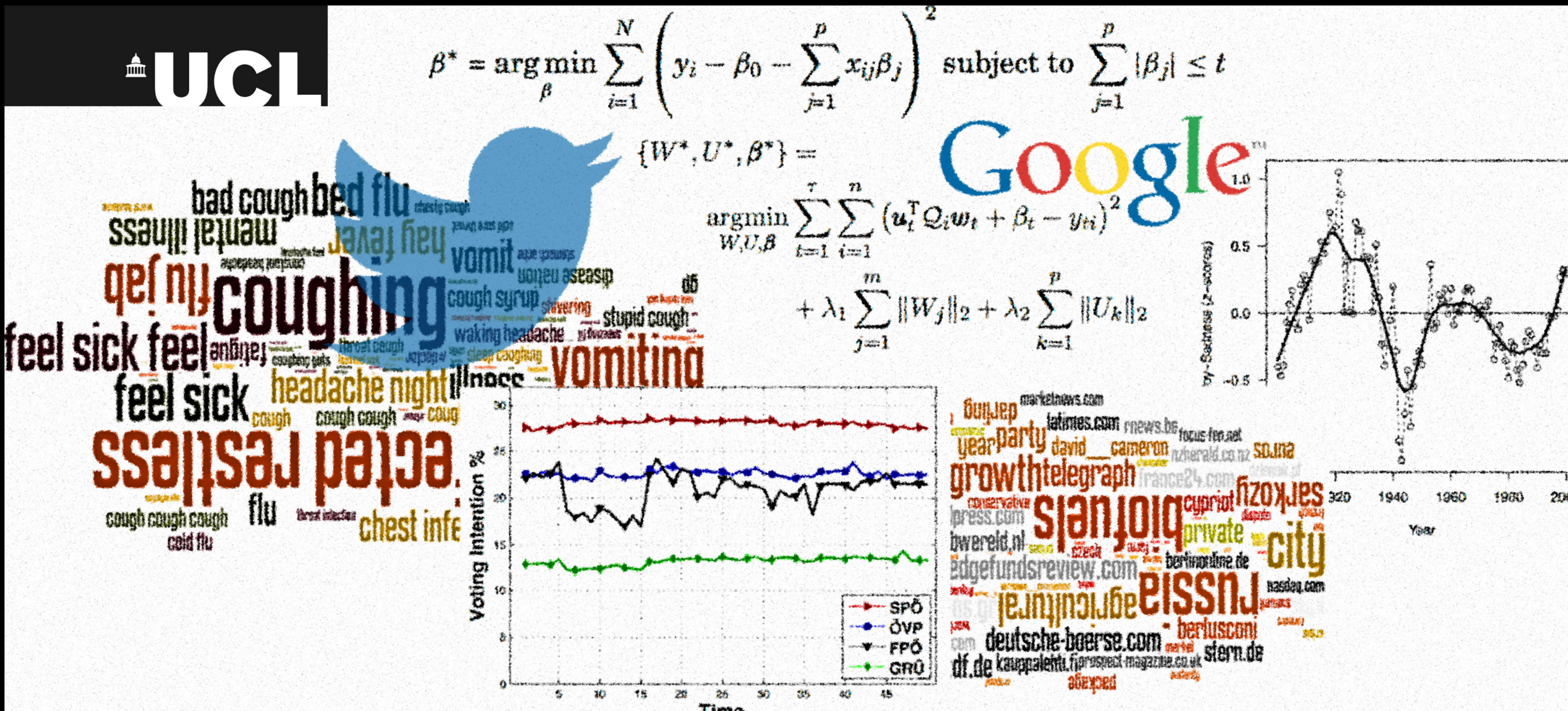
Extracting interesting concepts from large-scale textual data



Vasileios Lampos, University College London

New Frontiers of Automated Content Analysis in the Social Sciences (ACA) — July 1-3, 2015

Minimising the usual introduction

+ the Internet 'revolution'
+ successful web products feeding from user activity (search engines, social networks)
+ large volumes of digitised data ('***Big Data***')
+ lots of ***user-generated text*** & ***activity logs***

Can we arrive to better understandings of our 'world' from this data?

Minimising the usual introduction

+ the Internet ‘revolution’
+ successful web products feeding from user activity (search engines, social networks)
+ large volumes of digitised data (‘***Big Data***’)
+ lots of ***user-generated text*** & ***activity logs***

Can we arrive to better understandings of our ‘world’ from this data?

Overview

A. Prefixed keyword-based mining
B. Automating feature selection
C. User-centric (bilinear) modelling
D. Inferring user characteristics

5% 100%

Extracting interesting concepts from large-scale textual data

Word taxonomies for emotion

WordNet Affect

- builds on WordNet — automated word selection
- anger, disgust, fear, joy, sadness, surprise

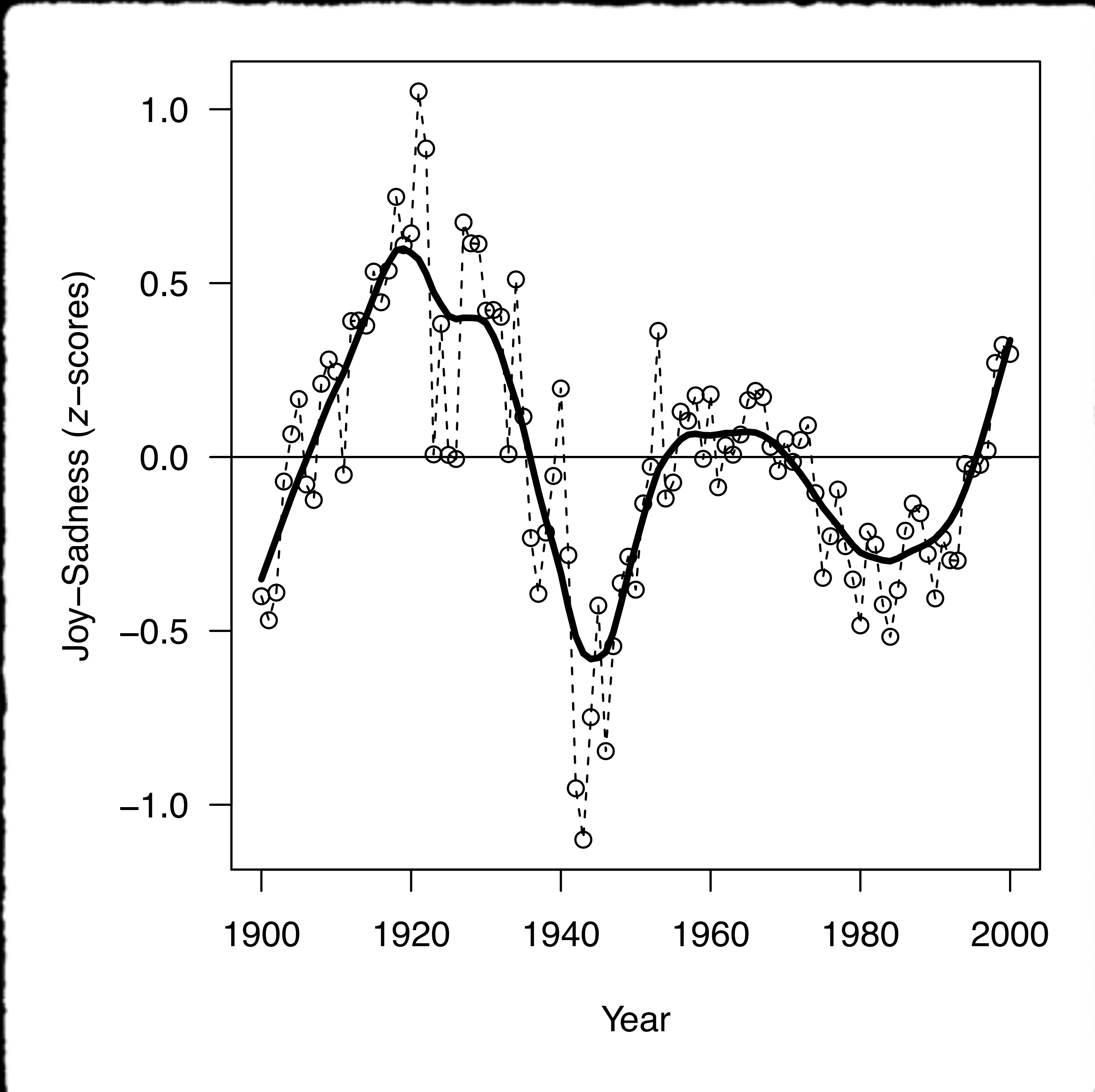
(*Strapparava & Valitutti, 2004*)

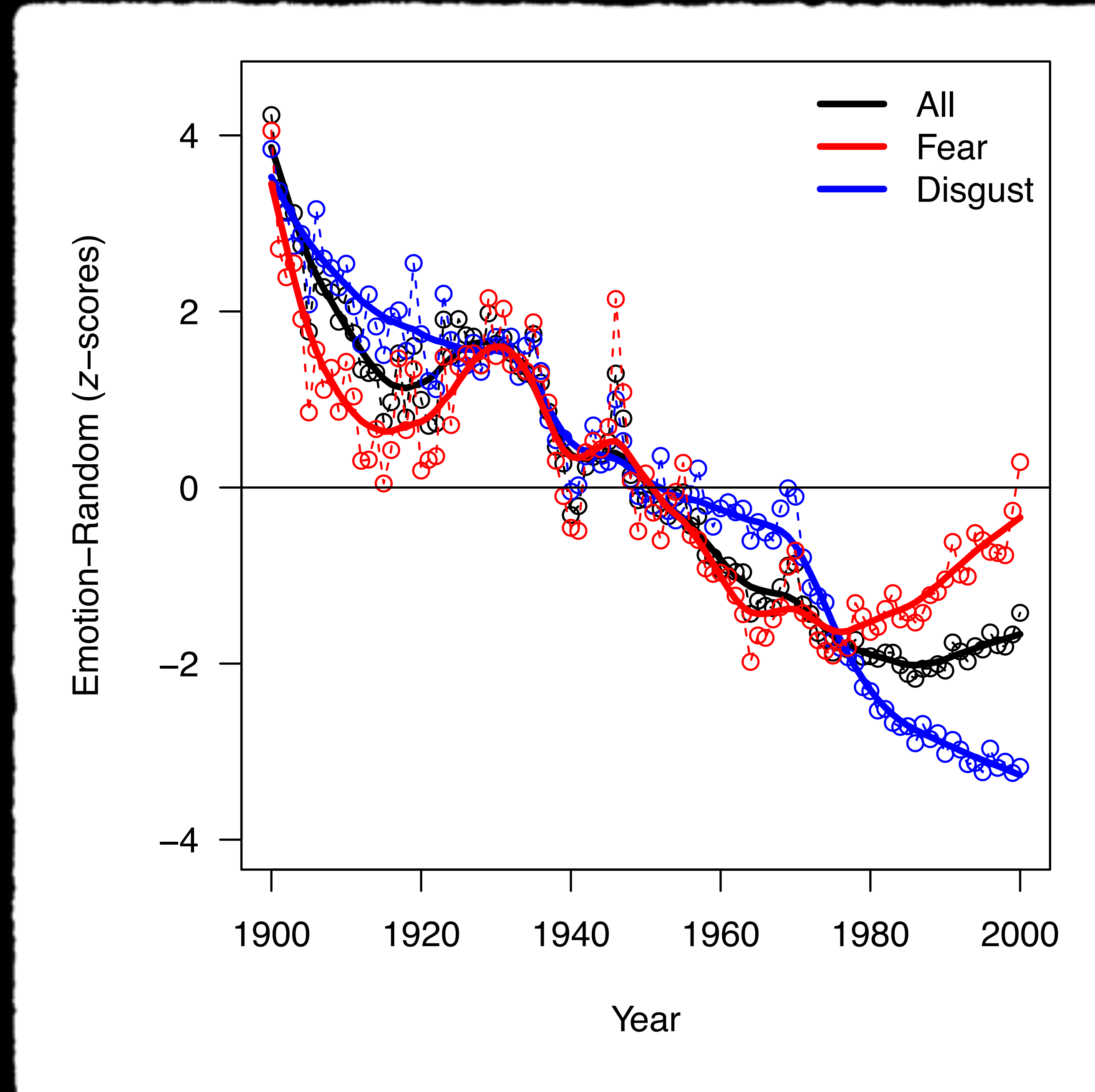
Linguistic Inquiry and Word Count (LIWC)

- taxonomies have been evaluated by human judges
- affect, anger, anxiety, sadness, negative or positive emotions

(*Pennebaker et al., 2007*)

Applying emotion taxonomies on ***Google Books***

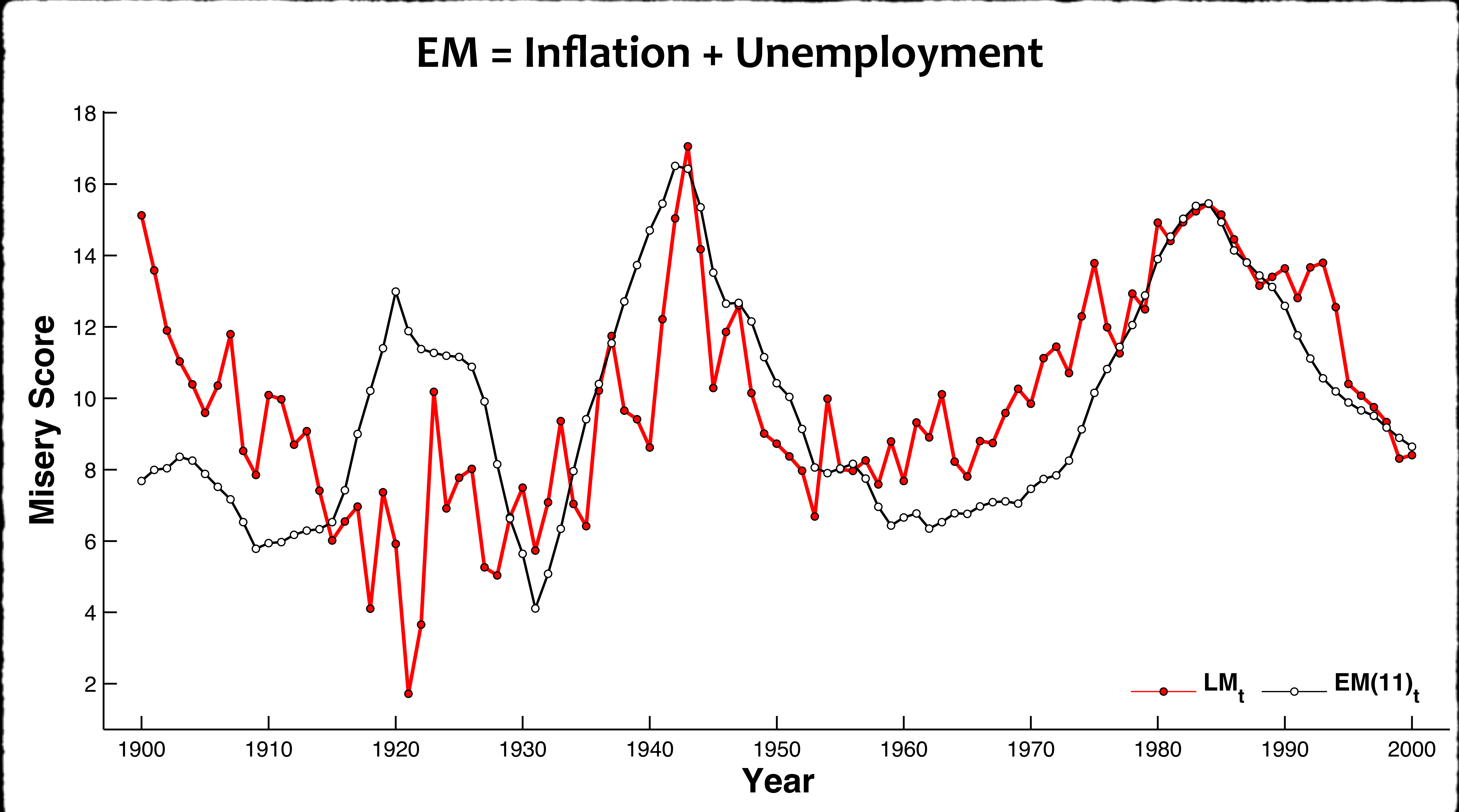




Emotion Score = *Mean of normalised emotion term frequencies*

Left: Joy minus Sadness — WWII, Baby Boom, Great Depression
Right: Emotional expression in English books decreases over the years

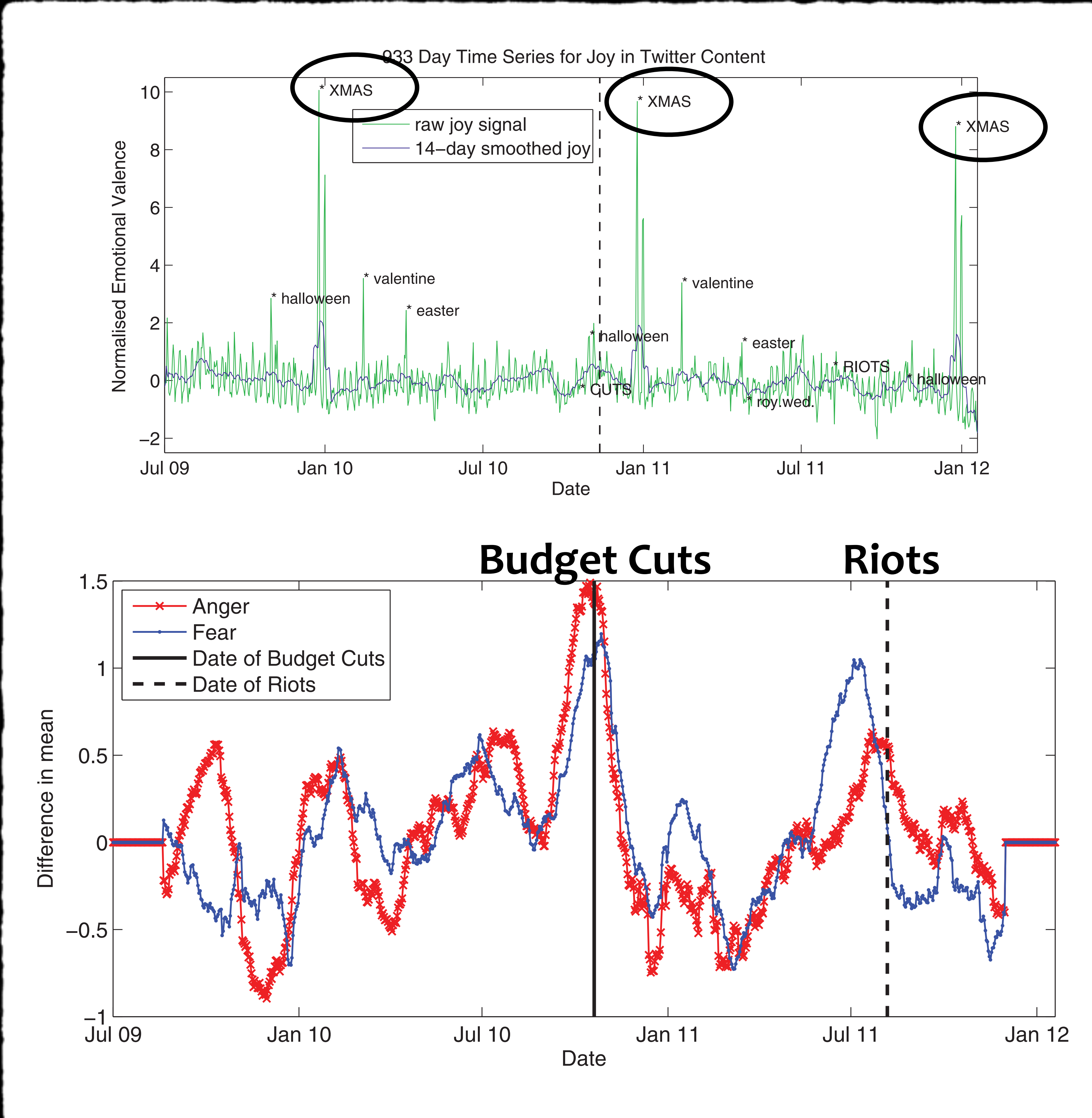
(*Acerbi, Lampos, Garnett & Bentley, 2013*)



Literary Misery (**LM = Sadness - Joy**) ***vs.***
Economic Misery (**EM**, 10-year past-moving average)
for books written in American English and US financial indices

(*Bentley, Acerbi, Ormerod & Lampos, 2014*)

… and now let's apply keyword based sentiment extraction tools on ***Twitter*** *content*



Collective mood patterns (UK)

Top:
'joy' time series across 3 years

Bottom:
rate of mood change for 'anger' and 'fear' (50-day window); **peaks** indicate **increase** in *mood change*

(*Lansdall-Welfare, Lampos & Cristianini, 2012*)

Overview

A. Prefixed keyword-based mining

B. Automating feature selection

C. User-centric (bilinear) modelling

D. Inferring user characteristics

15% 100%

Extracting interesting concepts from large-scale textual data

The case of influenza-like illness (ILI)

- existence of ‘**ground truth**’ enables optimisation of keyword selection
- supervised learning task (f: $\boldsymbol{X}$ —> $\boldsymbol{y}$)

Case study: ***nowcasting* ILI rates**

- infer ILI rates based on user data
- ‘ground truth’ provided via traditional health surveillance schemes
- complementary disease indicator
- earlier-warning
- applicable to parts of the world with less comprehensive healthcare systems
- noisy, biased demographics, media bias

(*Lampos & Cristianini, 2010*
Lampos, De Bie & Cristianini, 2010

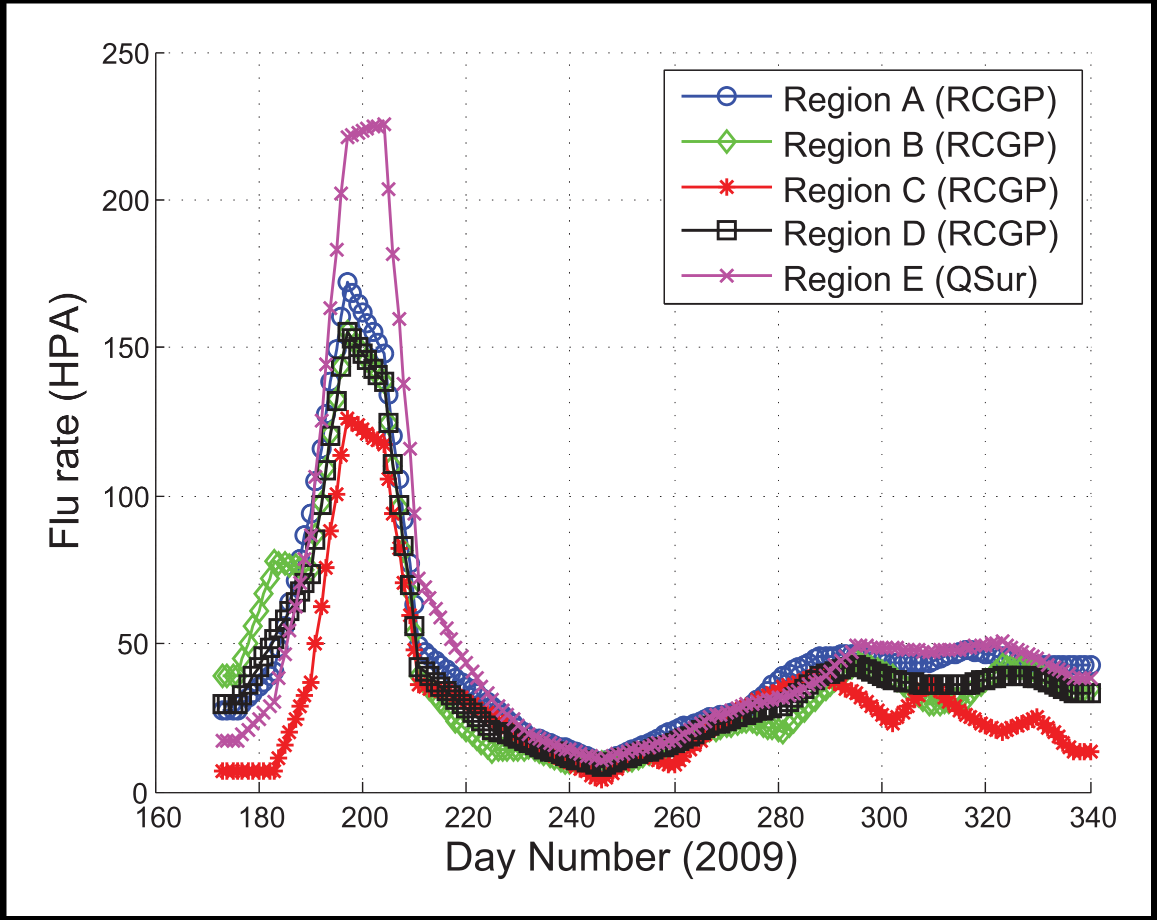
The case of influenza-like illness (ILI)

+ existence of '**ground truth**' enables optimisation of keyword selection
+ supervised learning task (f: $\boldsymbol{X}$ —> $\boldsymbol{y}$)

Case study: ***nowcasting* ILI rates**

+ infer ILI rates based on Twitter data
+ 'ground truth' provided via traditional health surveillance schemes
+ complementary disease indicator
+ earlier-warning
+ applicable to parts of the world with less comprehensive healthcare systems
+ noisy, biased demographics, media bias

Official ILI regional rates



(*Lampos & Cristianini, 2010* and *Lampos, De Bie & Cristianini, 2010*)

The case of influenza-like illness (ILI)

Twitter data
+ 27 million tweets from 54 UK urban centres
+ June 22 to December 6, 2009

Health surveillance data
+ ILI rates expressing GP consultations per 100,000 people, where the diagnosis was ILI

Feature extraction
+ a few handcrafted terms, *and*
+ all unigrams from related websites (Wikipedia, NHS, etc.)
+ = 1560 stemmed unigrams (*most of which unrelated*)

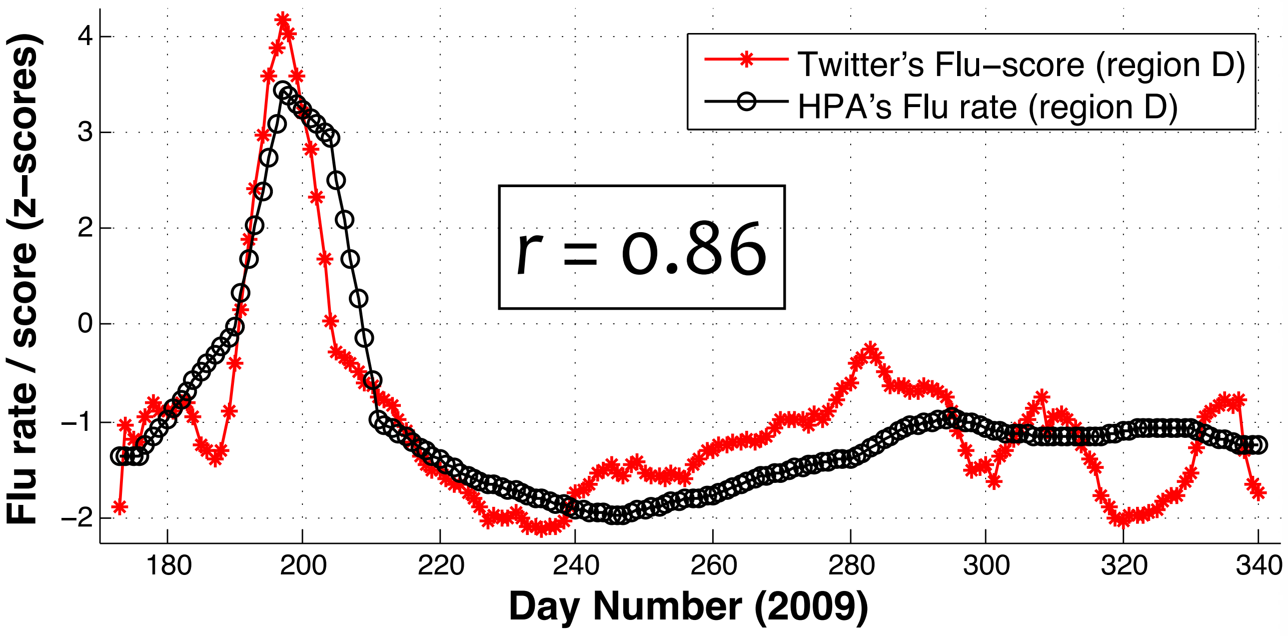
(*Lampos & Cristianini, 2010* and *Lampos, De Bie & Cristianini, 2010*)

Regularised text regression

- observations $\boldsymbol{x}_i \in \mathbb{R}^m$, $i \in \{1, ..., n\}$ — $\boldsymbol{X}$
- responses $y_i \in \mathbb{R}$, $i \in \{1, ..., n\}$ — $\boldsymbol{y}$
- weights, bias $w_j, \beta \in \mathbb{R}$, $j \in \{1, ..., m\}$ — $\boldsymbol{w}_* = [\boldsymbol{w}; \beta]$

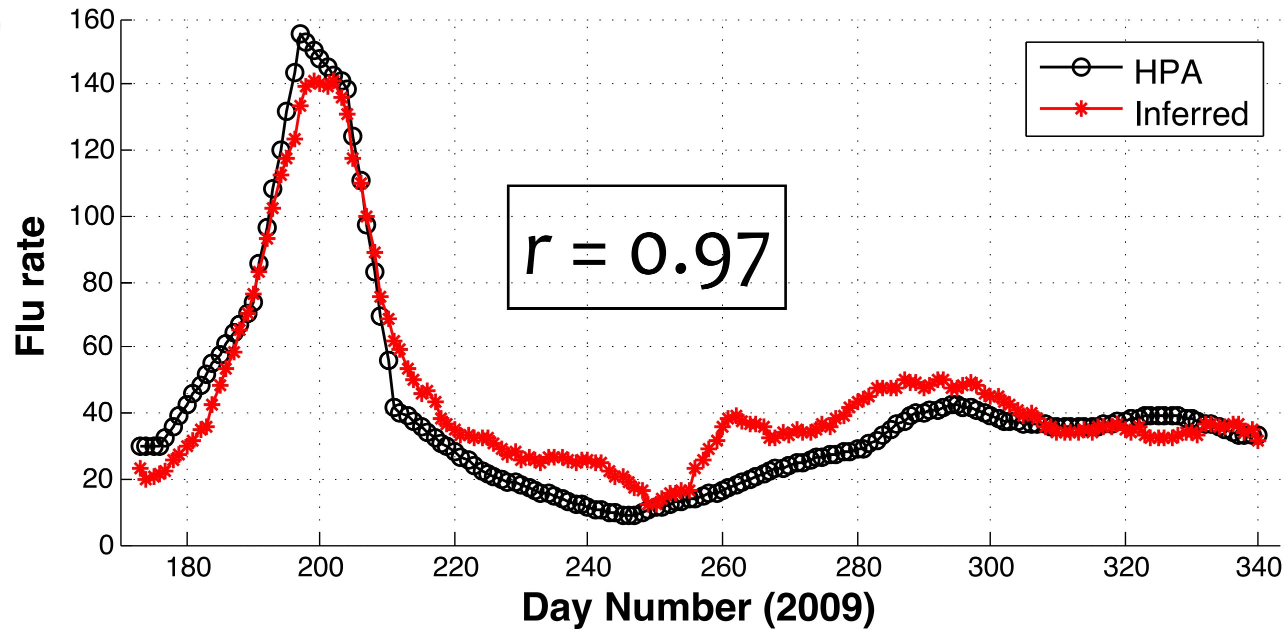
$$\underset{\boldsymbol{w}_*}{\operatorname{argmin}} \left\{ \|\boldsymbol{X}_* \boldsymbol{w}_* - \boldsymbol{y}\|_{\ell_2}^2 + \boxed{\lambda \|\boldsymbol{w}\|_{\ell_1}} \right\}$$

broadly known as the ‘***lasso***’ (*Tibshirani, 1996*)



41 handcrafted markers

blood, cold, cough, dizzy, feel sick, feeling unwell, fever, flu, headache, runny nose, shivers, sore throat, stomach ache (…)



Automatically selected unigrams

lung, unwel, temperatur, like, headach, season, unusu, chronic, child, dai, appetit, stai, symptom, spread, diarrhoea, start, muscl, weaken, immun, feel, liver (…)

Manual vs. automated feature selection

(*Lampos & Cristianini, 2010*)

Robustifying the previous algorithm

Lasso may not select the *true model* due to collinearities in the feature space (*Zhao & Yu, 2006*)

Bootstrap lasso ('*bolasso*') for feature selection (*Bach, 2008*)

- For a number (N) of bootstraps, i.e. iterations
  - Sample the feature space with replacement ($\boldsymbol{X}_i$)
  - Learn a new model ($\boldsymbol{w}_i$) by applying lasso on $\boldsymbol{X}_i$ and $\boldsymbol{y}$
  - Remember the n-grams with nonzero weights
- Select the n-grams with nonzero weights in p% of the N bootstraps
- p can be optimised using a held-out validation set

— ***Will all this generalise to a different case study?***

(*Lampos, De Bie & Cristianini, 2010* and *Lampos & Cristianini, 2012*)

Robustifying the previous algorithm

Lasso may not select the *true model* due to collinearities in the feature space (*Zhao & Yu, 2006*)

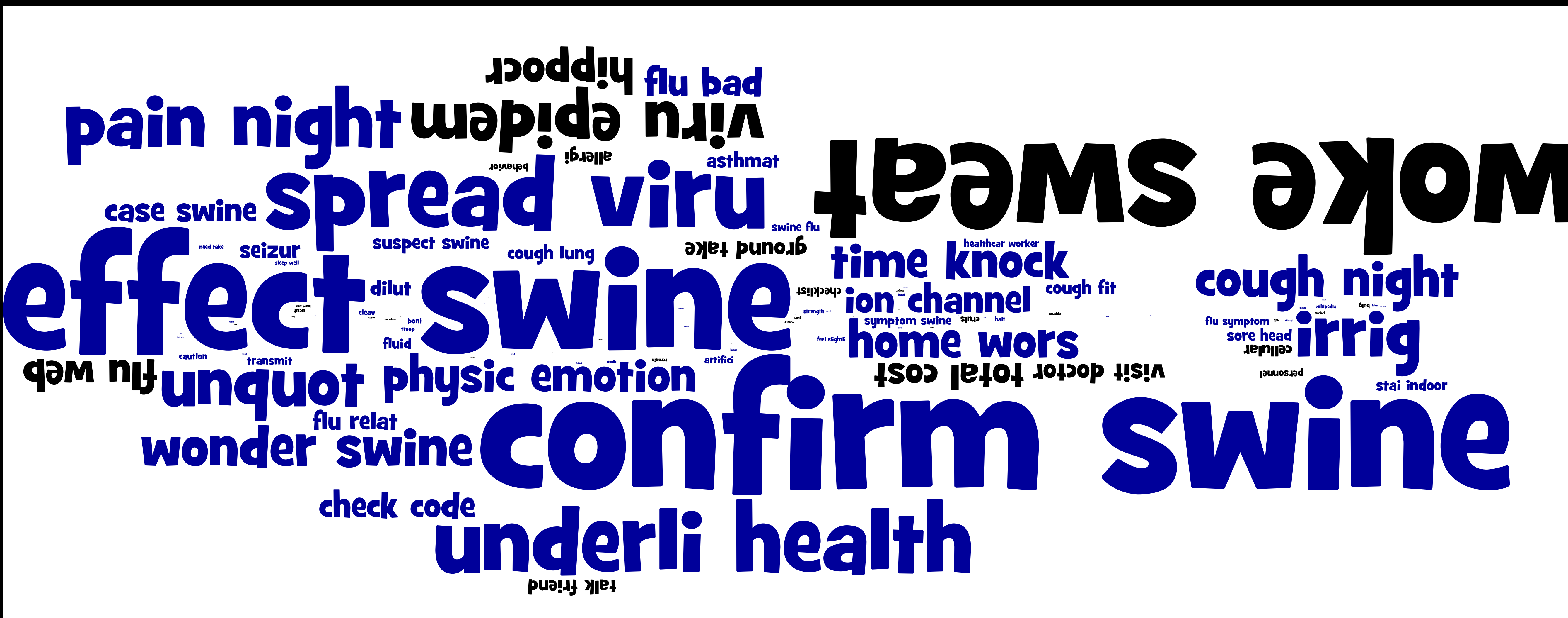
Bootstrap lasso ('*bolasso*') for feature selection (*Bach, 2008*)

- For a number (*N*) of bootstraps, i.e. iterations
  - Sample the feature space with replacement ($\boldsymbol{X}_i$)
  - Learn a new model ($\boldsymbol{w}_i$) by applying lasso on $\boldsymbol{X}_i$ and $\boldsymbol{y}$
  - Remember the n-grams with nonzero weights
- Select the n-grams with nonzero weights in *p*% of the *N* bootstraps
- *p* can be optimised using a held-out validation set

— ***Will all this generalise to a different case study?***

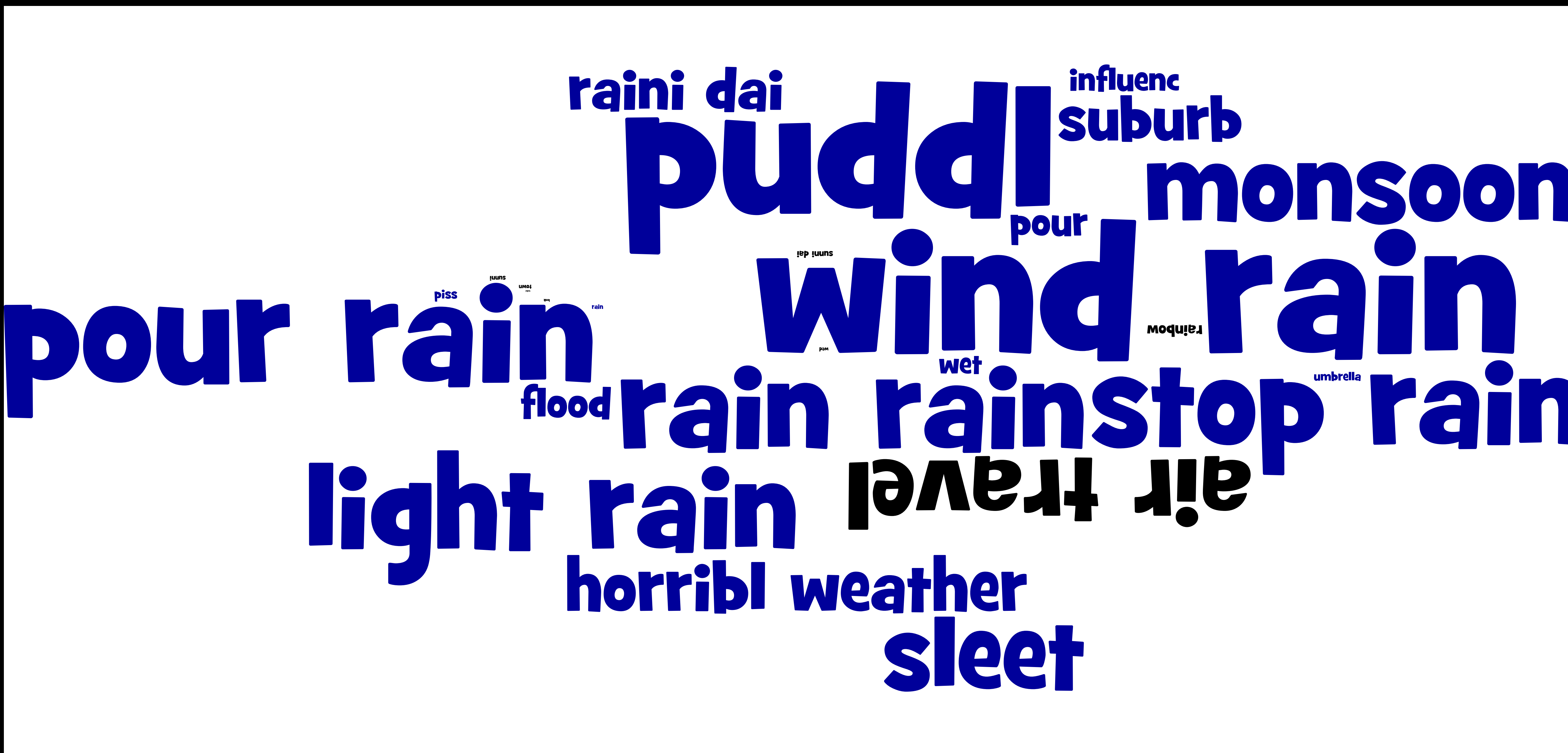
(*Lampos, De Bie & Cristianini, 2010* and *Lampos & Cristianini, 2012*)

Word cloud — Selected n-grams for ILI



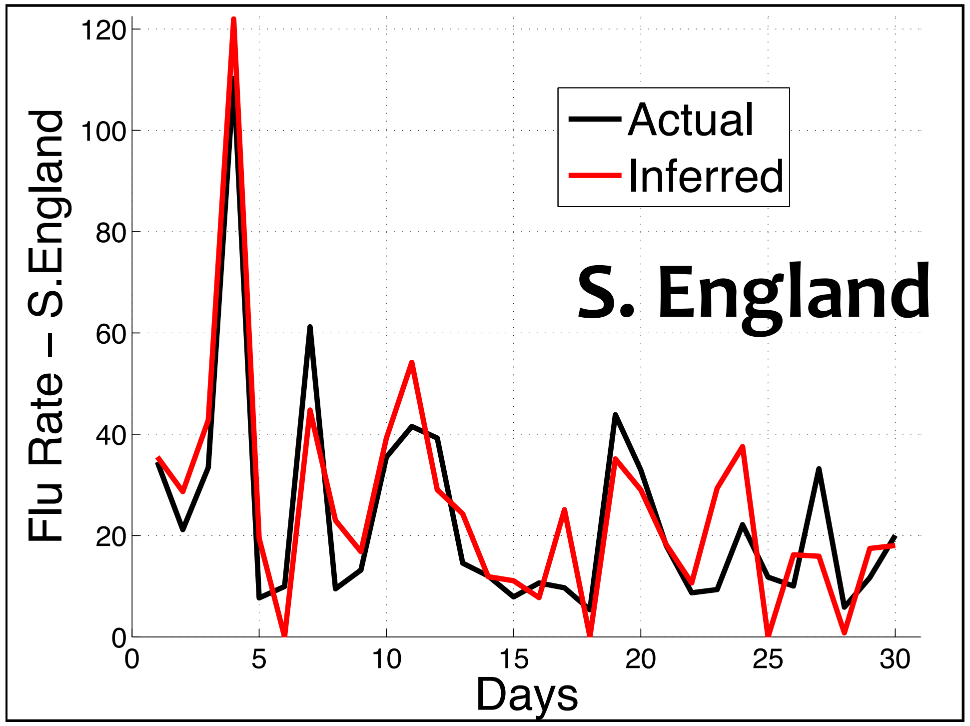
So, apart from flu,
we also tried to nowcast
rainfall rates.

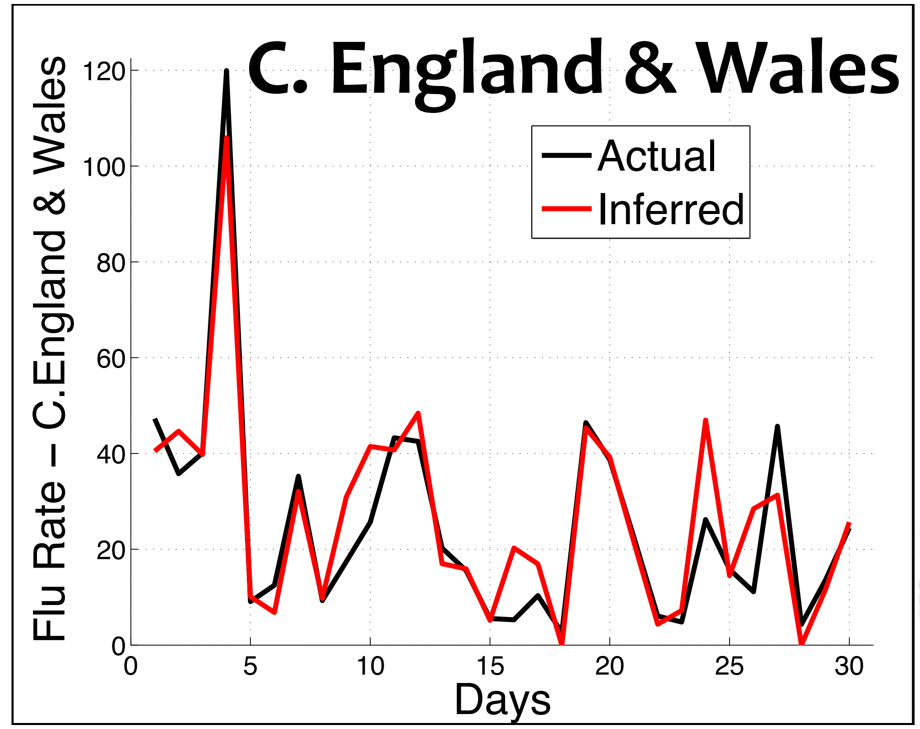
Word cloud — Selected n-grams for rain



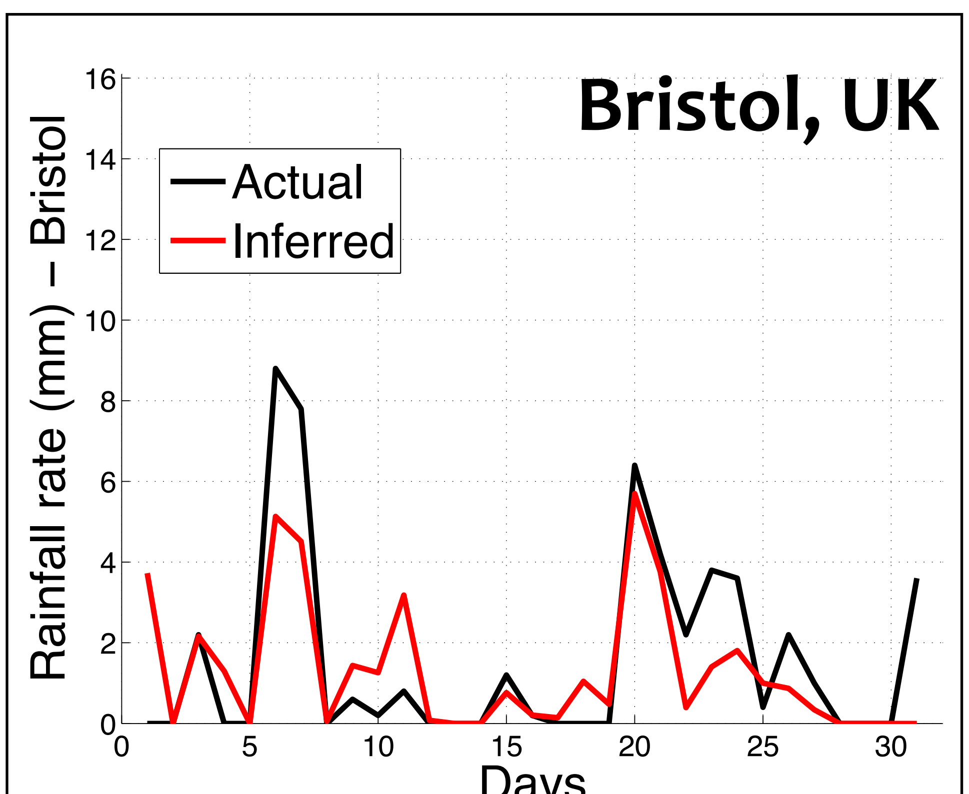
Inference examples

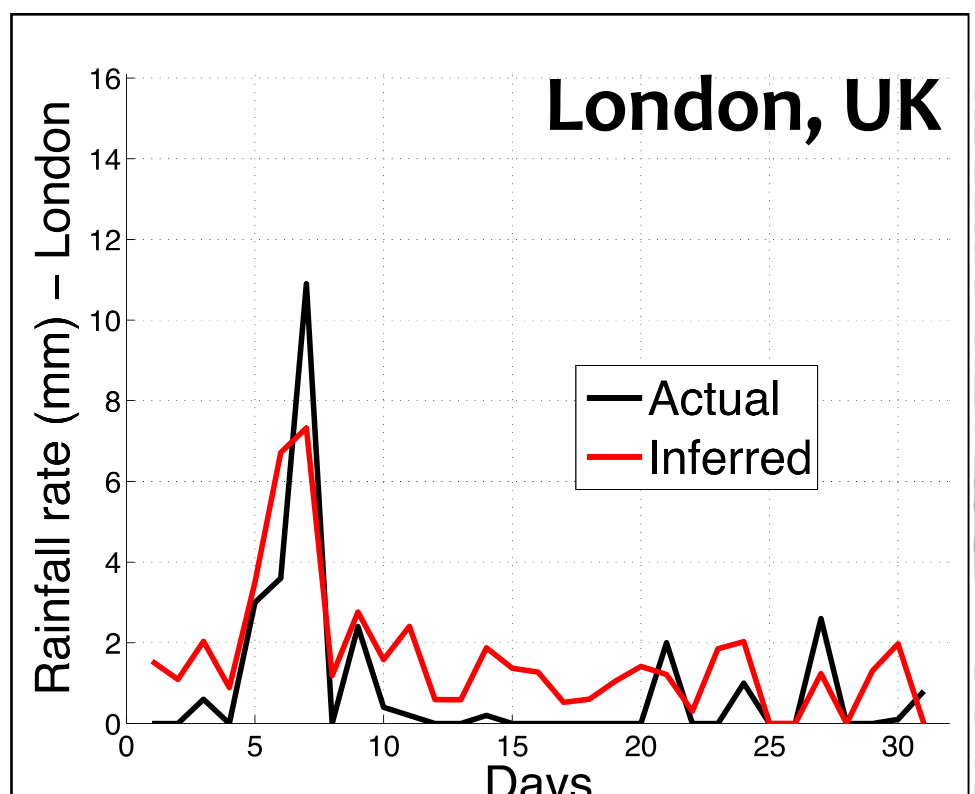
ILI rates





Rainfall rates





Top:
ILI rates

Bottom:
Rainfall rates

Overview

A. Prefixed keyword-based mining
B. Automating feature selection
C. User-centric (bilinear) modelling
D. Inferring user characteristics

User-centric modelling : why?

+ text regression models usually focus on the word space
+ **social media** context —> **words**, but also **users**
+ models may benefit by incorporating a form of user contribution in the current word modelling
+ in this way more relevant users contribute more, and irrelevant users may be filtered out

(*Lampos, Preotiuc-Pietro & Cohn, 2013*)

'bilinear' modelling : definition

Linear regression $f\left(\boldsymbol{x}_i\right) = \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{w} + \beta$

- observations $\boldsymbol{x}_i \in \mathbb{R}^m,$ $i \in \{1, ..., n\}$ — $\boldsymbol{X}$
- responses $y_i \in \mathbb{R},$ $i \in \{1, ..., n\}$ — $\boldsymbol{y}$
- weights, bias $w_j, \beta \in \mathbb{R},$ $j \in \{1, ..., m\}$ — $\boldsymbol{w}_* = [\boldsymbol{w}; \beta]$

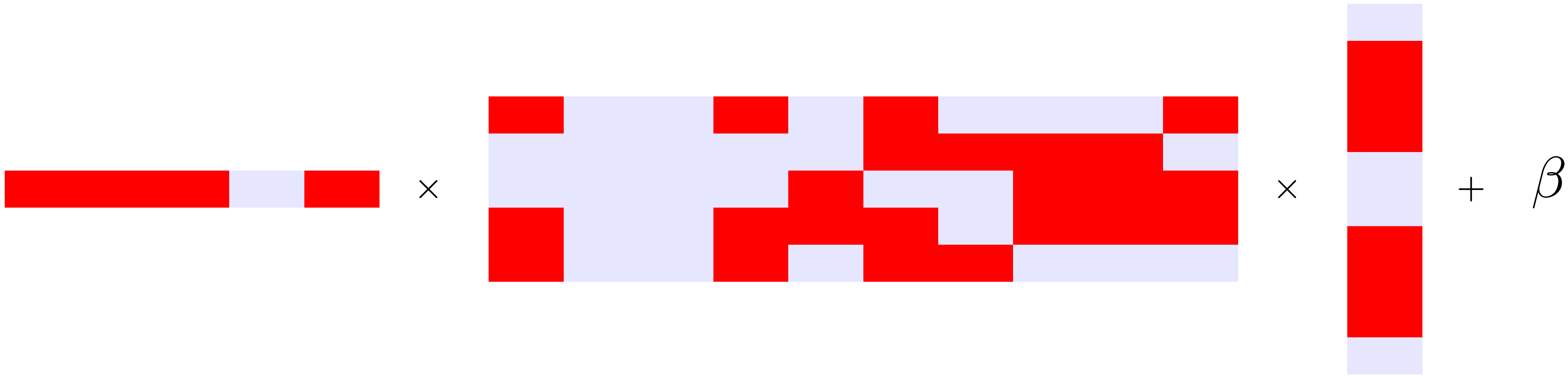
Bilinear regression $f\left(\boldsymbol{Q}_i\right) = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{Q}_i\boldsymbol{w} + \beta$

- users $p \in \mathbb{Z}^+$
- observations $\boldsymbol{Q}_i \in \mathbb{R}^{p \times m},$ $i \in \{1, ..., n\}$ — $\boldsymbol{\mathcal{X}}$
- responses $y_i \in \mathbb{R},$ $i \in \{1, ..., n\}$ — $\boldsymbol{y}$
- weights, bias $u_k, w_j, \beta \in \mathbb{R},$ $k \in \{1, ..., p\}$, $j \in \{1, ..., m\}$ — $\boldsymbol{u},\ \boldsymbol{w},\ \beta$

‘bilinear’ modelling : definition

- users $\quad p \in \mathbb{Z}^{+}$
- observations $\quad \boldsymbol{Q}_i \in \mathbb{R}^{p \times m}, \quad i \in \{1, \ldots, n\} \quad — \quad \boldsymbol{\mathcal{X}}$
- responses $\quad y_i \in \mathbb{R}, \quad i \in \{1, \ldots, n\} \quad — \quad \boldsymbol{y}$
- weights, bias $\quad u_k, w_j, \beta \in \mathbb{R}, \quad k \in \{1, \ldots, p\}, \; j \in \{1, \ldots, m\} \quad — \quad \boldsymbol{u}, \boldsymbol{w}, \beta$

$$f\left(\boldsymbol{Q}_i\right) = \boldsymbol{u}^{\mathrm{T}} \boldsymbol{Q}_i \boldsymbol{w} + \beta$$



$\boldsymbol{u}^{\mathrm{T}}$ $\quad \boldsymbol{Q}_i$ $\quad \boldsymbol{w}$

Bilinear regularised regression

• users	$p \in \mathbb{Z}^+$			
• observations	$\boldsymbol{Q}_i \in \mathbb{R}^{p \times m}$,	$i \in \{1, ..., n\}$	—	$\boldsymbol{\mathcal{X}}$
• responses	$y_i \in \mathbb{R}$,	$i \in \{1, ..., n\}$	—	$\boldsymbol{y}$
• weights, bias	$u_k, w_j, \beta \in \mathbb{R}$,	$k \in \{1, ..., p\}$ $j \in \{1, ..., m\}$	—	$\boldsymbol{u}$, $\boldsymbol{w}$, β

$$\underset{\boldsymbol{u},\boldsymbol{w},\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \left(\boldsymbol{u}^{\mathrm{T}} \boldsymbol{Q}_i \boldsymbol{w} + \beta - y_i \right)^2 + \psi(\boldsymbol{u}, \theta_u) + \psi(\boldsymbol{w}, \theta_w) \right\}$$

$\psi(\cdot)$: **regularisation function** with a set of hyper-parameters (θ)

- if $\psi(\boldsymbol{v}, \lambda) = \lambda \|\boldsymbol{v}\|_{\ell_1}$ — Bilinear Lasso
- if $\psi(\boldsymbol{v}, \lambda_1, \lambda_2) = \lambda_1 \|\boldsymbol{v}\|_{\ell_2}^2 + \lambda_2 \|\boldsymbol{v}\|_{\ell_1}$ — Bilinear Elastic Net (**BEN**)

(*Lampos, Preotiuc-Pietro & Cohn, 2013*)

An extension: bilinear & multi-task

+ optimise (*learn the model parameters for*) a number of tasks **jointly**

+ attempt to **improve generalisation** by exploiting domain specific information of related tasks

+ good choice for under-sampled distributions (*knowledge transfer*)

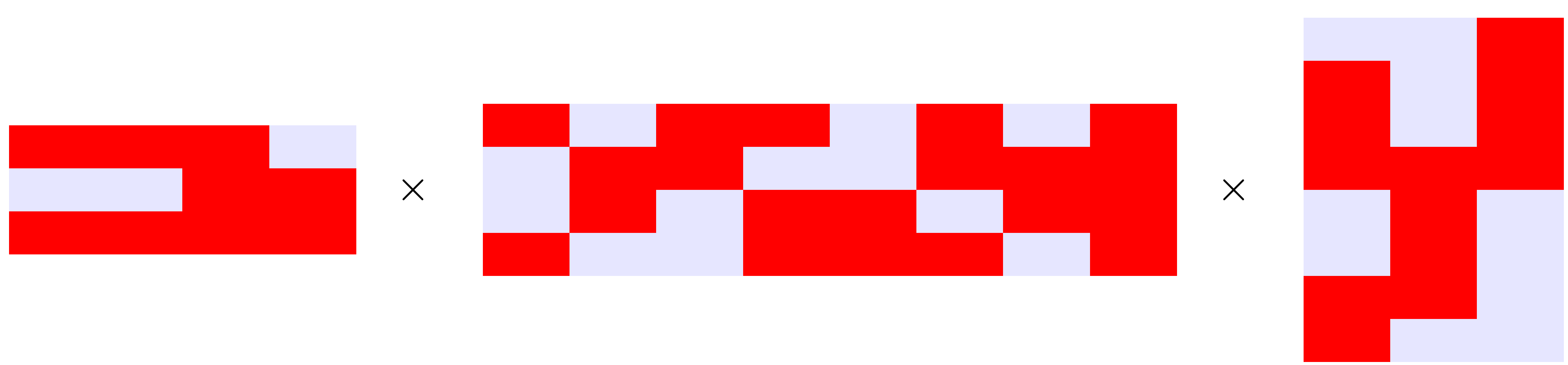
+ **application-driven reasons** (e.g. voting intention modelling)

(*Caruana, 1997; Lampos, Preotiuc-Pietro & Cohn, 2013*)

Bilinear *multi-task* text regression

- tasks $\tau \in \mathbb{Z}^+$
- users $p \in \mathbb{Z}^+$
- observations $\boldsymbol{Q}_i \in \mathbb{R}^{p \times m}$, $i \in \{1, ..., n\}$ — $\boldsymbol{\mathcal{X}}$
- responses $\boldsymbol{y}_i \in \mathbb{R}^{\tau}$, $i \in \{1, ..., n\}$ — $\boldsymbol{Y}$
- weights, bias $\boldsymbol{u}_k, \boldsymbol{w}_j, \boldsymbol{\beta} \in \mathbb{R}^{\tau}$, $k \in \{1, ..., p\}$, $j \in \{1, ..., m\}$ — $\boldsymbol{U}$, $\boldsymbol{W}$, $\boldsymbol{\beta}$

$$f\left(\boldsymbol{Q}_i\right) = \mathrm{tr}\left(\boldsymbol{U}^{\mathrm{T}} \boldsymbol{Q}_i \boldsymbol{W}\right) + \boldsymbol{\beta}$$



$\boldsymbol{U}^{\mathrm{T}}$ $\boldsymbol{Q}_i$ $\boldsymbol{W}$

Bilinear Group $\ell_{2,1}$ (BGL)

• tasks	$\tau \in \mathbb{Z}^+$			
• users	$p \in \mathbb{Z}^+$			
• observations	$\boldsymbol{Q}_i \in \mathbb{R}^{p \times m}$,	$i \in \{1, \dots, n\}$	—	$\boldsymbol{\mathcal{X}}$
• responses	$\boldsymbol{y}_i \in \mathbb{R}^{\tau}$,	$i \in \{1, \dots, n\}$	—	$\boldsymbol{Y}$
• weights, bias	$\boldsymbol{u}_k, \boldsymbol{w}_j, \boldsymbol{\beta} \in \mathbb{R}^{\tau}$,	$k \in \{1, \dots, p\}$ $j \in \{1, \dots, m\}$	—	$\boldsymbol{U}, \boldsymbol{W}, \boldsymbol{\beta}$

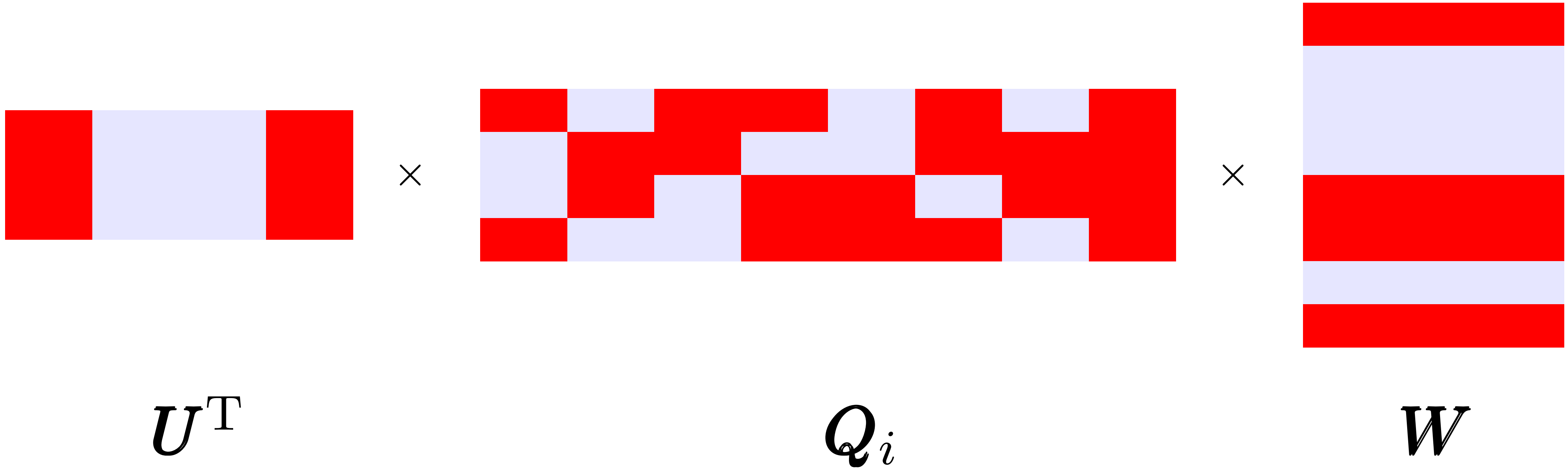
$$\underset{\boldsymbol{U},\boldsymbol{W},\boldsymbol{\beta}}{\operatorname{argmin}} \left\{ \sum_{t=1}^{\tau} \sum_{i=1}^{n} \left(\boldsymbol{u}_t^{\mathrm{T}} \boldsymbol{Q}_i \boldsymbol{w}_t + \beta_t - y_{ti} \right)^2 \right.$$

(*Argyriou et al., 2008*)

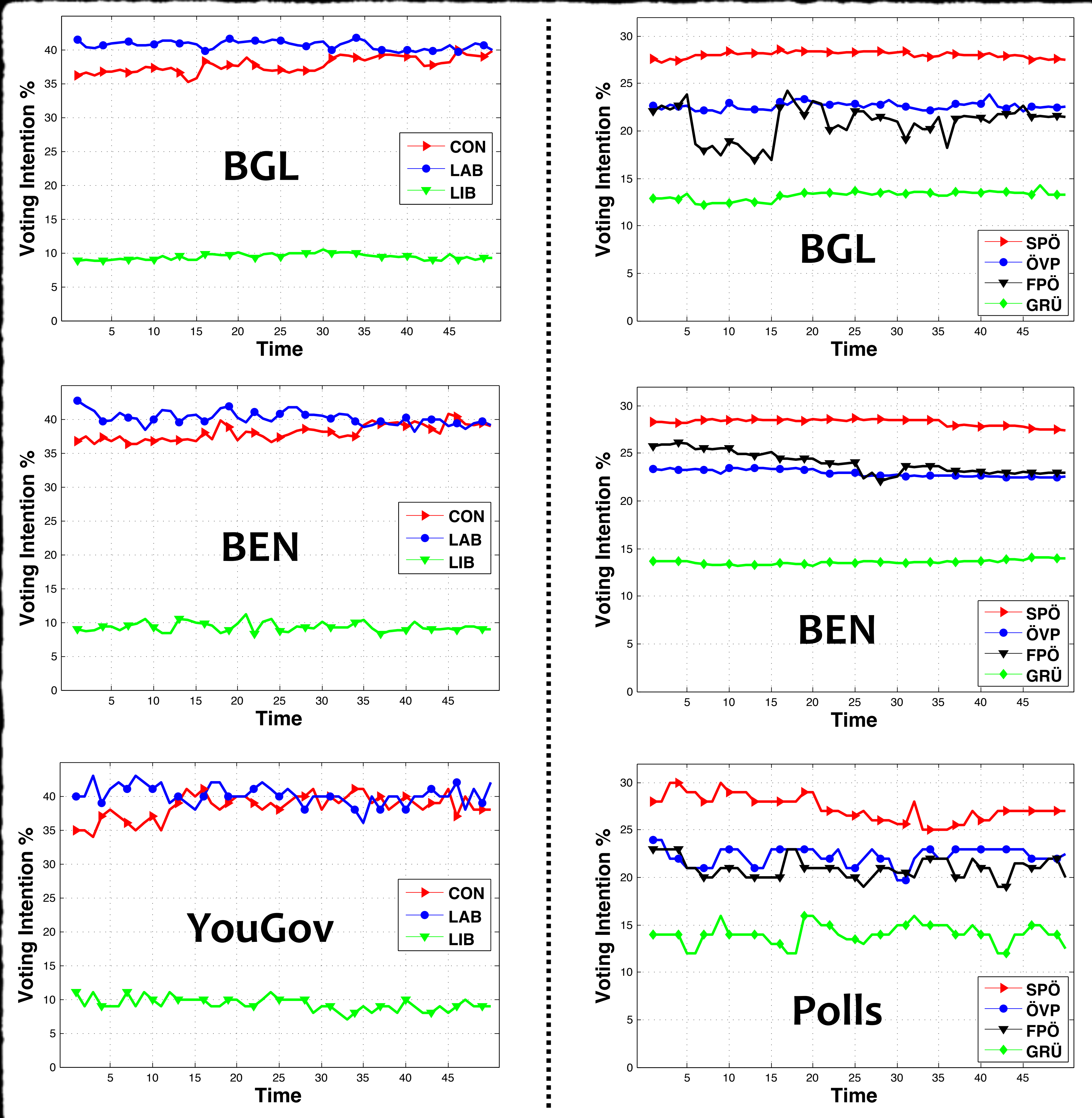
$$\left. + \lambda_u \sum_{k=1}^{p} \|\boldsymbol{U}_k\|_2 + \lambda_w \sum_{j=1}^{m} \|\boldsymbol{W}_j\|_2 \right\}$$

BGL's main property

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



- a feature (user or word) is usually **selected** (activated) **for all tasks,** but with different weights
- useful in the domain of ***political preference inference***

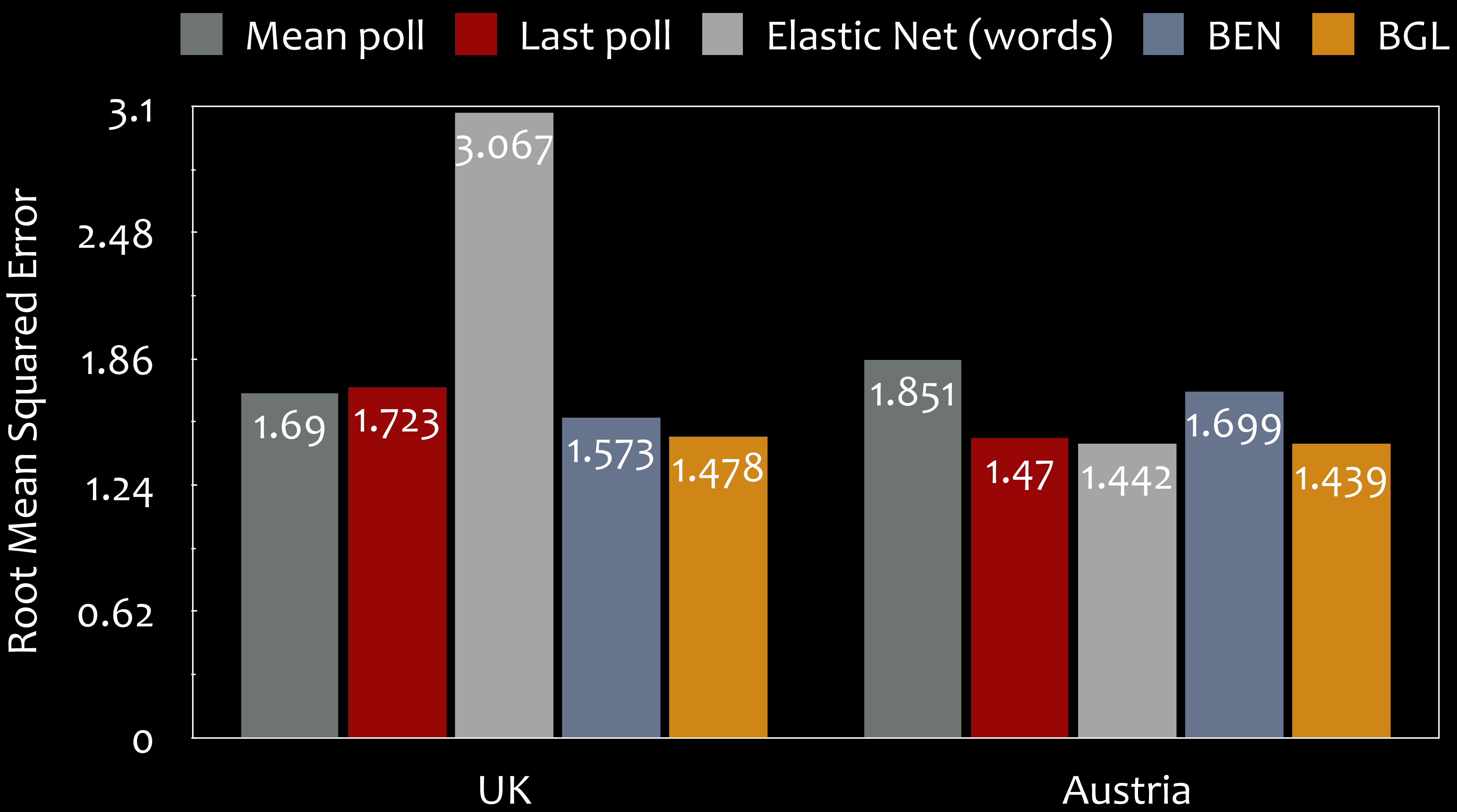


Inferring voting intention via Twitter

Left side:
UK — 3 parties, 42K users (~ to regional population), 81K unigrams, 240 polls, 2 years

Right side:
Austria — 4 parties, 1.1K manually selected users, 23K unigrams, 98 polls, 1 year

Performance figures — BGL prevails



BGL-scored tweet examples (Austria)

Party	Tweet	Score	User type
SPÖ	*Inflation rate in Austria slightly down in July from 2.2 to 2.1%. Accommodation, Water, Energy more expensive.*	0.745	Journalist
ÖVP	*Can really recommend the book "Res Publica" by Johannes #Voggenhuber! Food for thought and so on #Europe #Democracy*	-2.323	User
FPÖ	*Campaign of the Viennese SPO on "Living together" plays right into the hands of right-wing populists*	-3.44	Human rights
GRÜ	*Protest songs against the closing-down of the bachelor course of International Development: <link> #ID_remains #UniBurns #UniRage*	1.45	Student Union

Overview

A. Prefixed keyword-based mining
B. Automating feature selection
C. User-centric (bilinear) modelling
D. Inferring user characteristics

63% 100%

Predicting user impact on Twitter

- *Validate a hypothesis: "User behaviour on a social platform reflects on user impact"*

- *What parts of user behaviour are more relevant to a notion of user impact?*

- *In this regard, how informative are the text inputs from the users?*

(*Lampos, Aletras, Preotiuc-Pietro & Cohn, 2014*)

Defining an impact score (S)

$$\mathrm{S}(\phi_{\text{in}}, \phi_{\text{out}}, \phi_{\lambda}) = \ln\left(\frac{(\phi_{\lambda}+\theta)(\phi_{\text{in}}+\theta)^2}{\phi_{\text{out}}+\theta}\right)$$

$$(\phi_{\text{in}}^2/\phi_{\text{out}}) = (\phi_{\text{in}} - \phi_{\text{out}}) \times (\phi_{\text{in}}/\phi_{\text{out}}) + \phi_{\text{in}}$$

ϕ_{in} —> number of followers

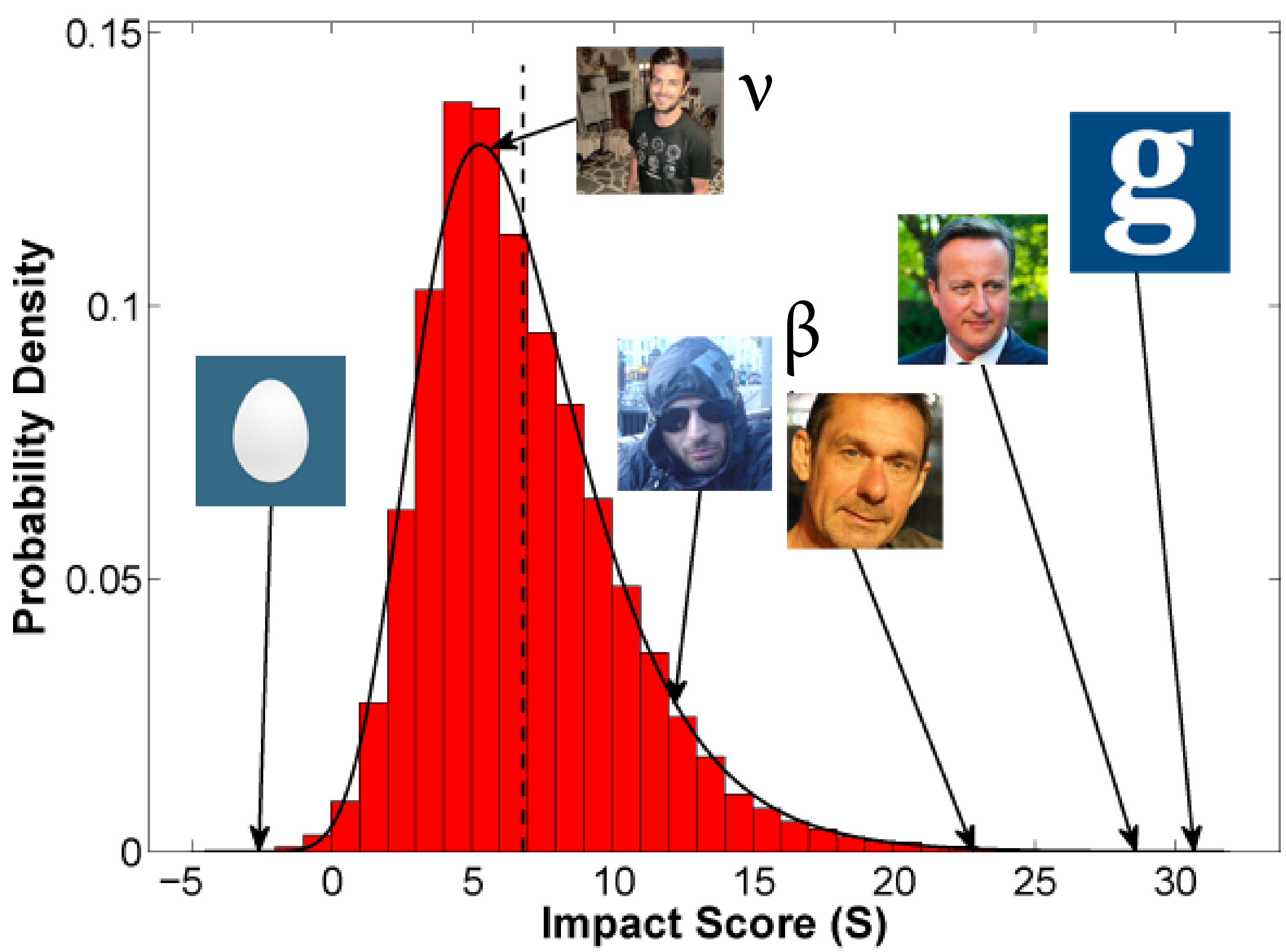
ϕ_{out} —> number of followees

ϕ_{λ} —> number of times listed

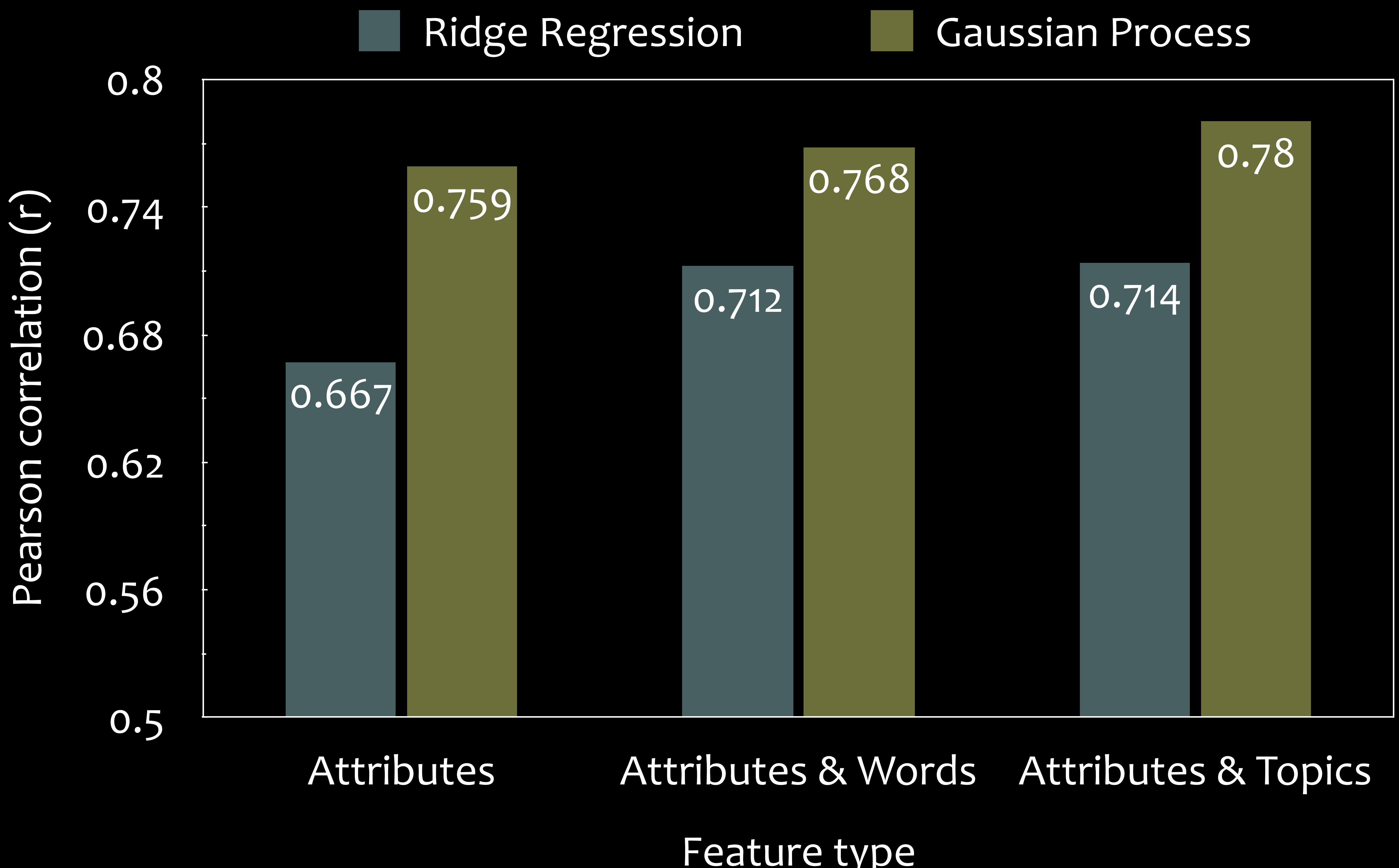
$\theta = 1$ —> logarithm is applied on a positive

β Vasileios Lampos ~ ***@lampos***
ν Nikolaos Aletras ~ ***@nikaletras***

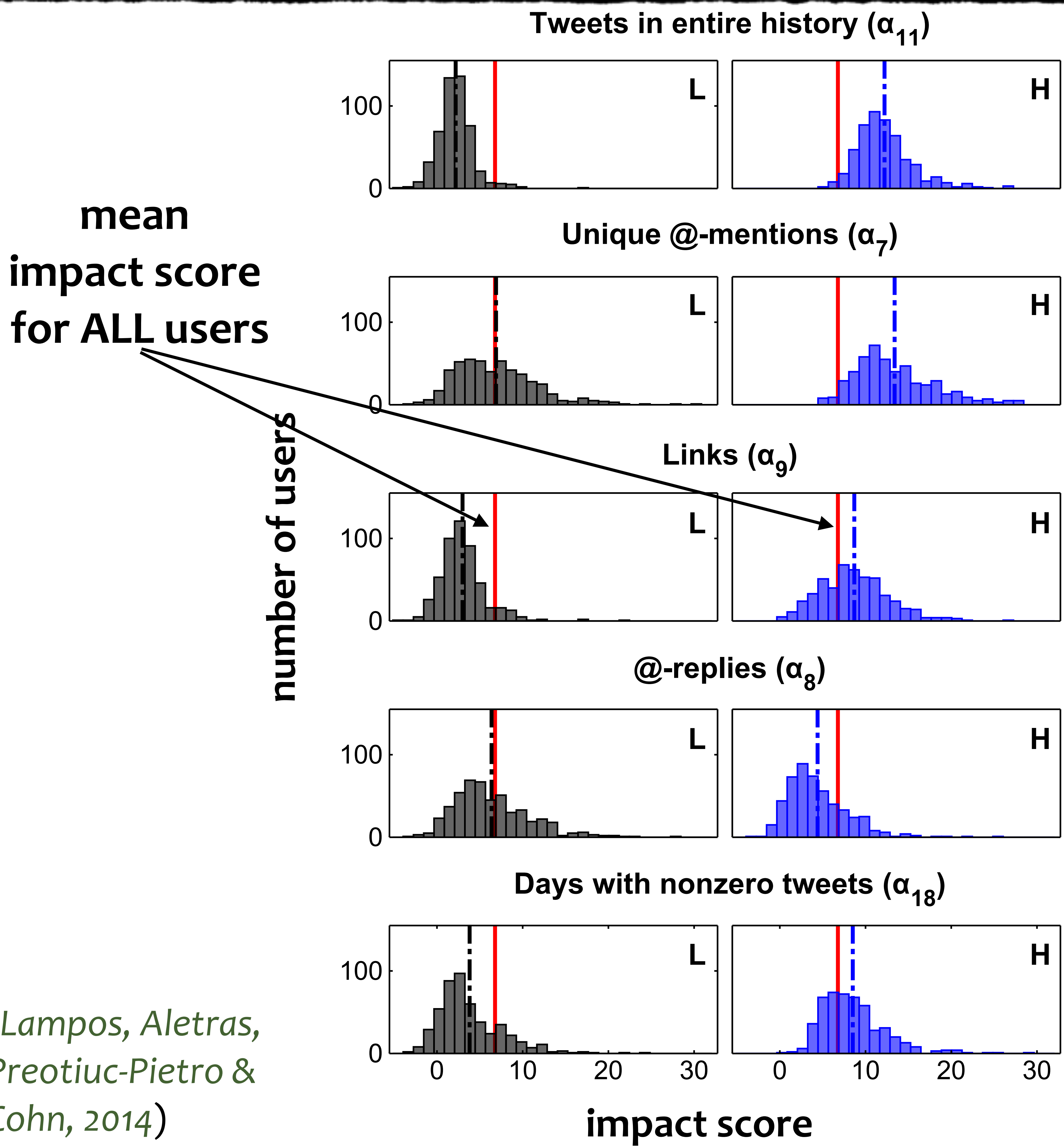
40K Twitter accounts (UK) considered



Impact prediction as a regression task



(*Lampos, Aletras, Preotiuc-Pietro & Cohn, 2014*)



(*Lampos, Aletras, Preotiuc-Pietro & Cohn, 2014*)

Some of the most important **user attributes** for **impact** (*excl. topics*)

500 accounts with the lower (**L**) and higher (**H**) impacts for an attribute

Impact ***plus***

- more tweets
- more @mentions
- more links
- less @replies
- less inactive days

Impact ***minus***

- less tweets
- less links
- more inactive days

We can guess the impact of user from user activity, but can we infer his / her ***occupation***?

Inferring the occupational class of a Twitter user

"Socioeconomic variables are influencing language use."
(*Bernstein, 1960*; *Labov, 1972/2006*)

+ Validate this hypothesis on a larger data set
+ Downstream applications
  + research (social science & other domains)
  + commercial
+ Proxy for income, socioeconomic class etc., i.e. further applications

(*Preotiuc-Pietro, Lampos & Aletras, 2015*)

Standard Occupational Classification (SOC, 2010)

C1 Managers, Directors & Senior Officials — *chief executive, bank manager*

C2 Professional Occupations — *mechanical engineer, pediatrist, **postdoc (!)***

C3 Associate Professional & Technical — *system administrator, dispensing optician*

C4 Administrative & Secretarial — *legal clerk, company secretary*

C5 Skilled Trades — *electrical fitter, tailor*

C6 Caring, Leisure, Other Service — *nursery assistant, hairdresser*

C7 Sales & Customer Service — *sales assistant, telephonist*

C8 Process, Plant and Machine Operatives — *factory worker, van driver*

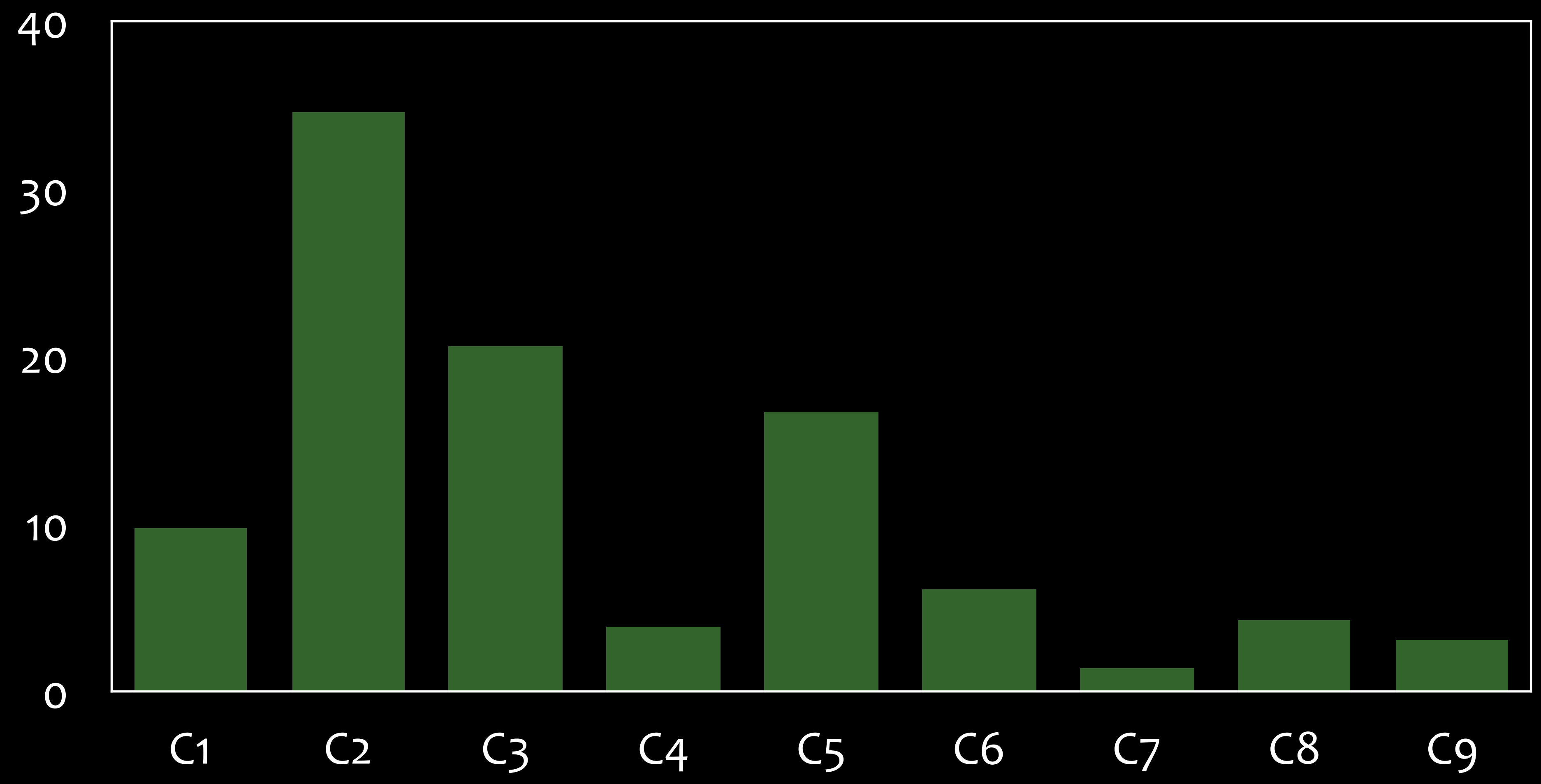
C9 Elementary — *shelf stacker, bartender*

Google "ONS" AND "SOC" for more information

Data

- 5,191 users mapped to their occupations, then mapped to one of the 9 SOC categories — *manual* (!) labelling
- 10 million tweets
- Get processed data: **http://www.sas.upenn.edu/~danielpr/jobs.tar.gz**

% of users per SOC category



Features

User attributes (18)

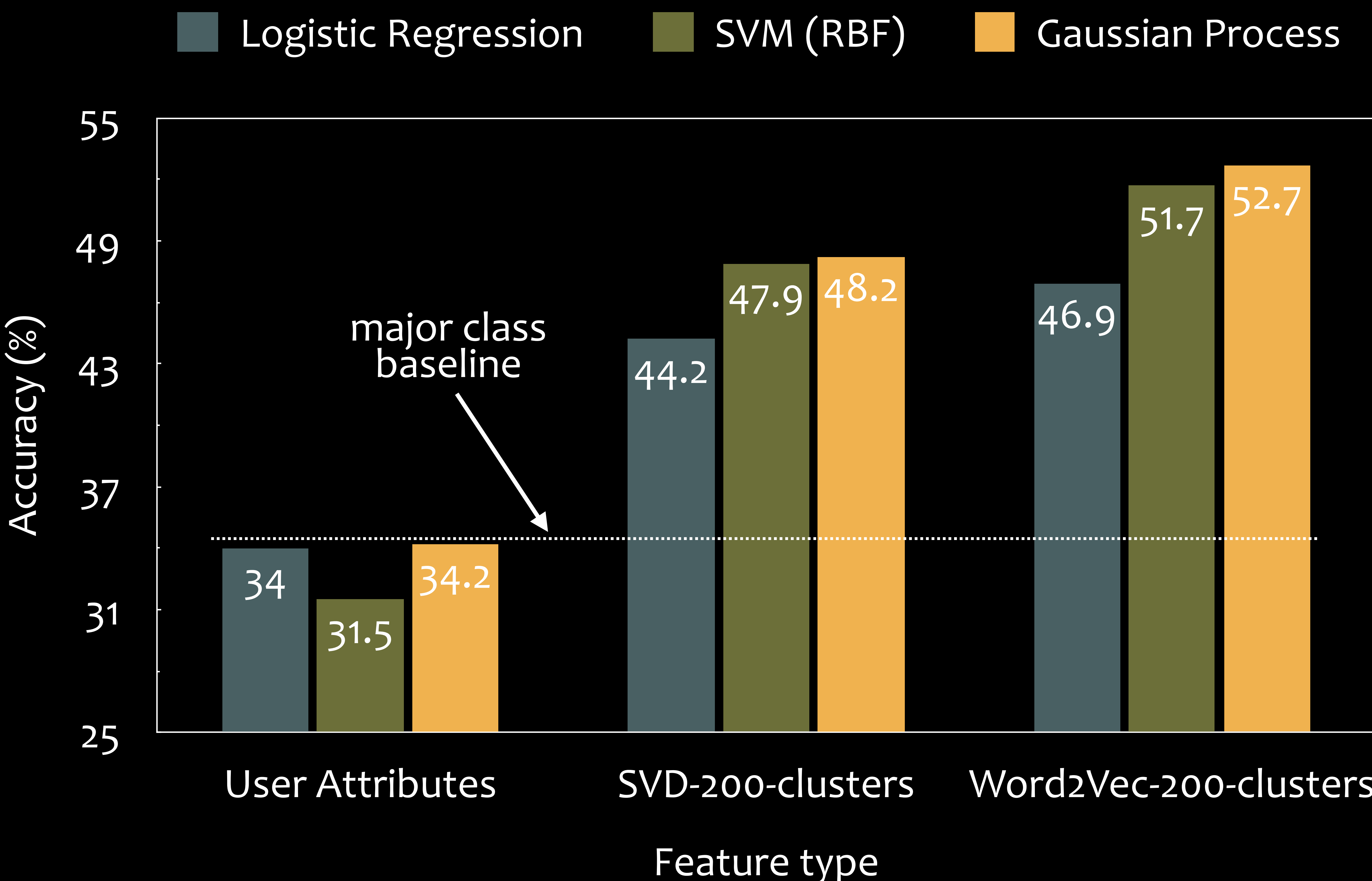
+ number of followers, friends, listings, follower/friend ratio, favourites, tweets, retweets, hashtags, @-mentions, @-replies, links and so on

Topics — Word clusters (**200**)

+ ***SVD*** on the graph laplacian of the word x word similarity matrix using normalised PMI, i.e. a form of spectral clustering (*Bouma, 2009*; *von Luxburg, 2007*)

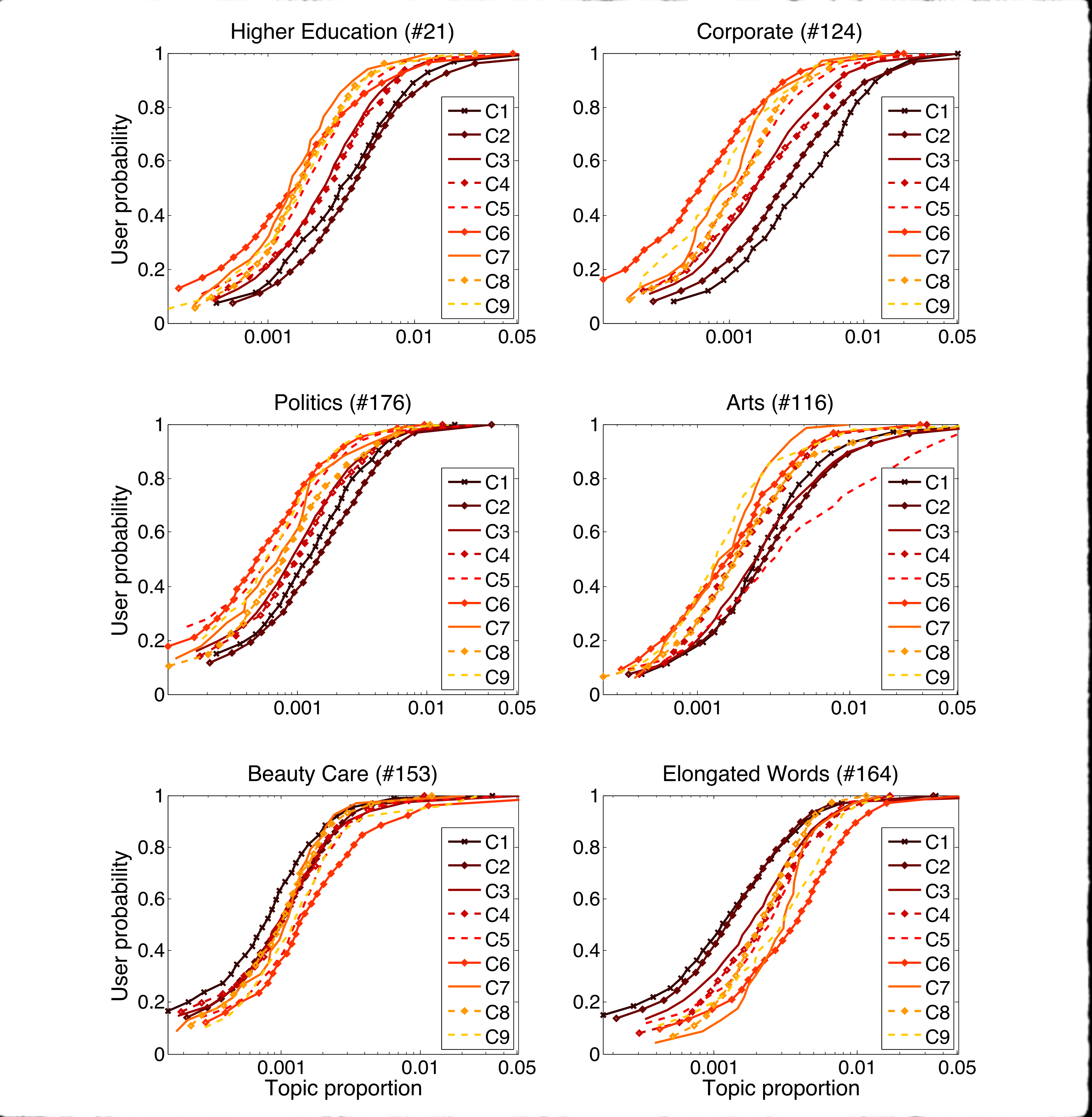
+ Skip-gram model with negative sampling to learn word embeddings (***Word2Vec***); pairwise cosine similarity on the embeddings to derive a word x word similarity matrix; then spectral clustering on the (*Mikolov et al., 2013*)

Occupational class (9-way) classification



Topics

Manual label	Most central words; Most frequent words	Rank
Arts	archival, stencil, canvas, minimalist; art, design, print	1
Health	chemotherapy, diagnosis, disease; risk, cancer, mental, stress	2
Beauty Care	exfoliating, cleanser, hydrating; beauty, natural, dry, skin	3
Higher Education	undergraduate, doctoral, academic, students, curriculum; students, research, board, student, college, education, library	4
Software Engineering	integrated, data, implementation, integration, enterprise; service, data, system, services, access, security	5
Football	bardsley, etherington, gallas; van, foster, cole, winger	7
Corporate	consortium, institutional, firm's; patent, industry, reports	8
Cooking	parmesan, curried, marinated, zucchini; recipe, meat, salad	9
Elongated Words	yaaayy, wooooo, woooo, yayyyyy, yaaaaay, yayayaya, yayy; wait, till, til, yay, ahhh, hoo, woo, woot, whoop, woohoo	12
Politics	religious, colonialism, christianity, judaism, persecution, fascism, marxism; human, culture, justice, religion, democracy	16



Discussion topics per occupational class — CDF plots

Plots explained

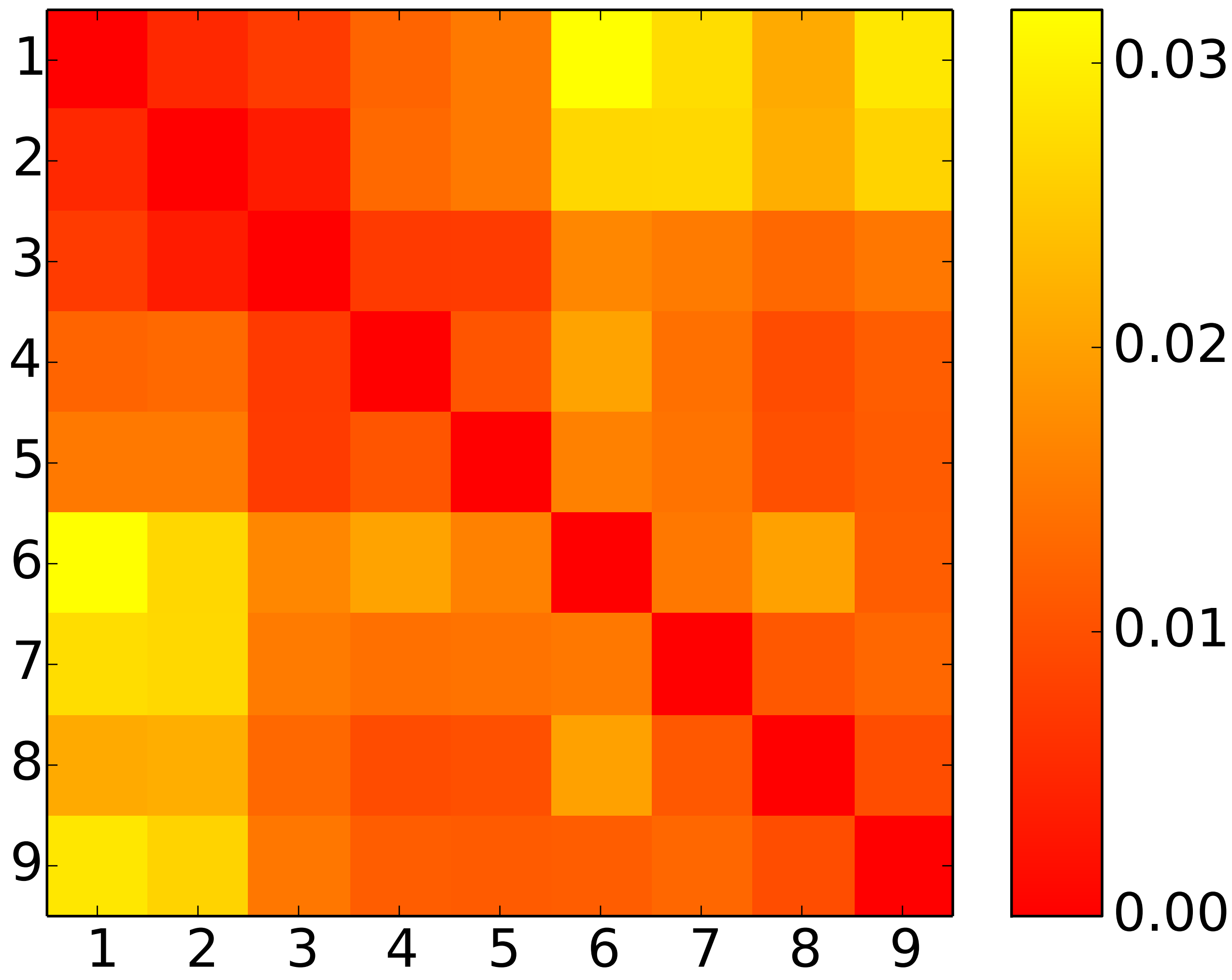
Topic more prevalent in a class (C1-C9), if the line leans closer to the bottom-right corner of the plot

Upper classes

- Higher Education
- Corporate
- Politics

Lower classes

- Beauty Care
- Elongated Words



Topics	C 1–2	C 6–9
Arts	4.95	2.79
Health	4.45	2.13
Beauty Care	1.40	2.24
Higher Education	6.04	2.56
Software Engineering	6.31	2.54
Football	0.54	0.52
Corporate	5.15	1.41
Cooking	2.81	2.49
Elongated Words	1.90	3.78
Politics	2.14	1.06

Left: Distance (Jensen-Shannon divergence) between topic distributions for the different occupational classes, depicted on a heatmap

Right: Comparison of mean topic usage between supersets of occupational classes (1-2 vs. 6-9)

concluding…

95%

Conclusions

Publicly available, **user-generated content** can be used to better understand:

+ ***collective emotion***
+ disease rates or the ***magnitude of some target events***
+ ***voting intentions***
+ ***user attributes*** (impact, occupation)

A number of studies (*too many to cite*) have attempted different — sometimes improved — approaches on the methods presented here.

Many studies have also explored different data mining scenarios (e.g. infer user gender, financial indices etc.).

Some of the challenges ahead

- Work closer with domain experts (social scientists to epidemiologists)
  - *e.g. in collaboration with Public Health England we proposed a method for assessing the impact of a health intervention through social media and search query data*
    (*Lampos, Yom-Tov, Pebody & Cox, 2015*)

- Understand better the biases of the online media (when it is desirable to conduct more generic conclusions)
  - *note that sometimes these biases may be a good thing*

- Attack more interesting (usually more complex) questions
  - *e.g. generalise the inference of offline from online behaviour*

- Improve on existing methods

Collaborators participating in the work presented today

(*in alphabetical order*)

Alberto Acerbi	Anthropology, Eindhoven University of Technology
Nikolaos Aletras	Natural Language Processing, University College London
Alex Bentley	Anthropology & Archaeology, University of Bristol
Trevor Cohn	Natural Language Processing, University of Melbourne
Nello Cristianini	Artificial Intelligence, University of Bristol
Tijl De Bie	Computational Pattern Analysis, University of Bristol
Philip Garnett	Complex Systems, University of York
Thomas Lansdall-Welfare	Computer Science, University of Bristol
Paul Ormerod	Decision Making and Uncertainty, University College London
Daniel Preotiuc-Pietro	Natural Language Processing, University of Pennsylvania

Extracting interesting concepts from large-scale textual data

Thank you!

slides available at

http://www.lampos.net/sites/default/files/slides/ACA2015.pdf

Bonus slides

100%

Training Bilinear Elastic Net (BEN)

BEN's ***objective function*** — — — — —>

$$\underset{\boldsymbol{u},\boldsymbol{w},\beta}{\operatorname{argmin}}\left\{\sum_{i=1}^{n}\left(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Q}_i\boldsymbol{w}+\beta-y_i\right)^2 + \lambda_{u_1}\|\boldsymbol{u}\|_{\ell_2}^2+\lambda_{u_2}\|\boldsymbol{u}\|_{\ell_1} + \lambda_{w_1}\|\boldsymbol{w}\|_{\ell_2}^2+\lambda_{w_2}\|\boldsymbol{w}\|_{\ell_1}\right\}$$

Biconvex problem

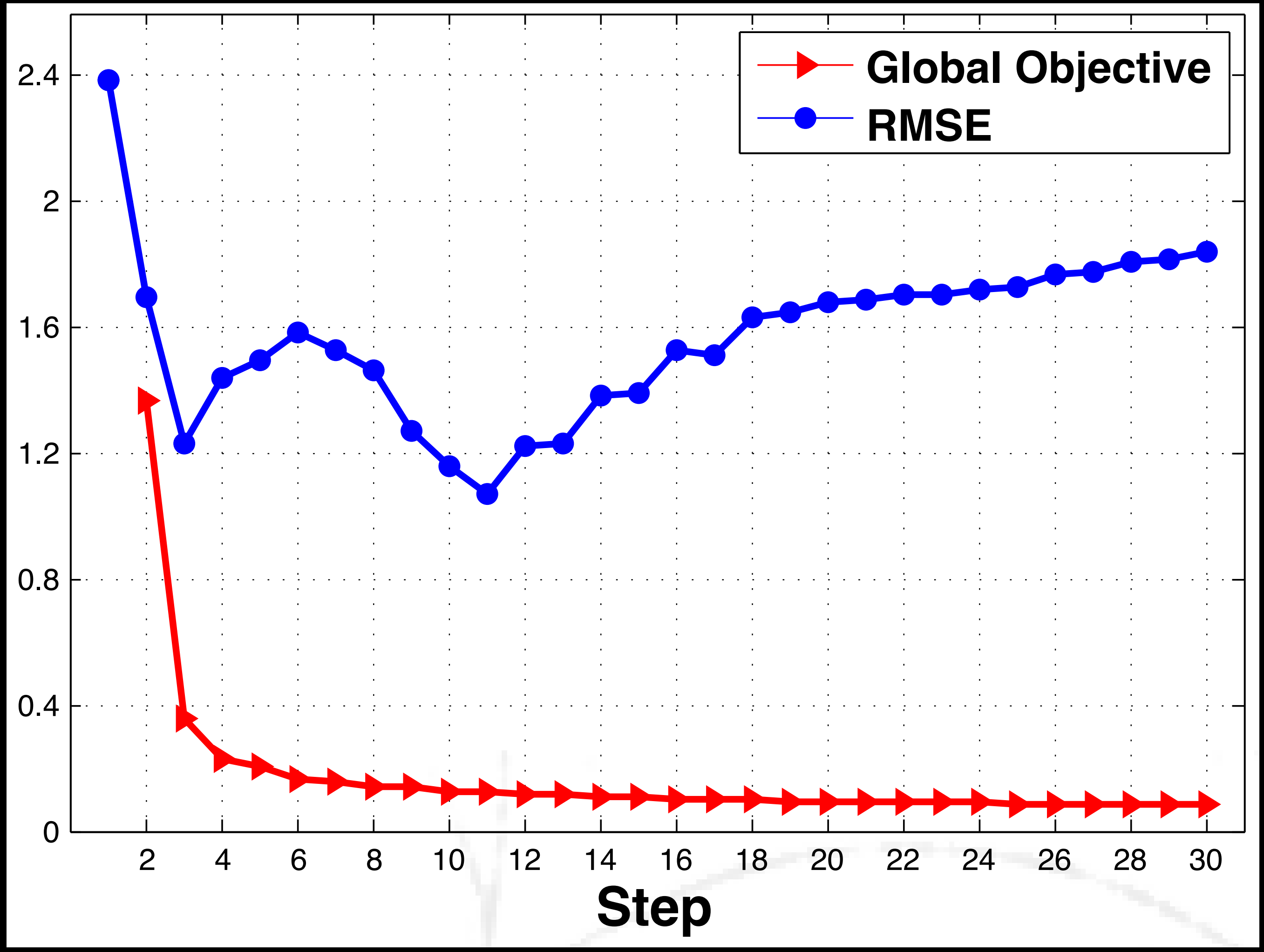
+ fix ***u*****,** learn ***w*** and vice versa
+ iterate through convex optimisation tasks

Large-scale solvers available

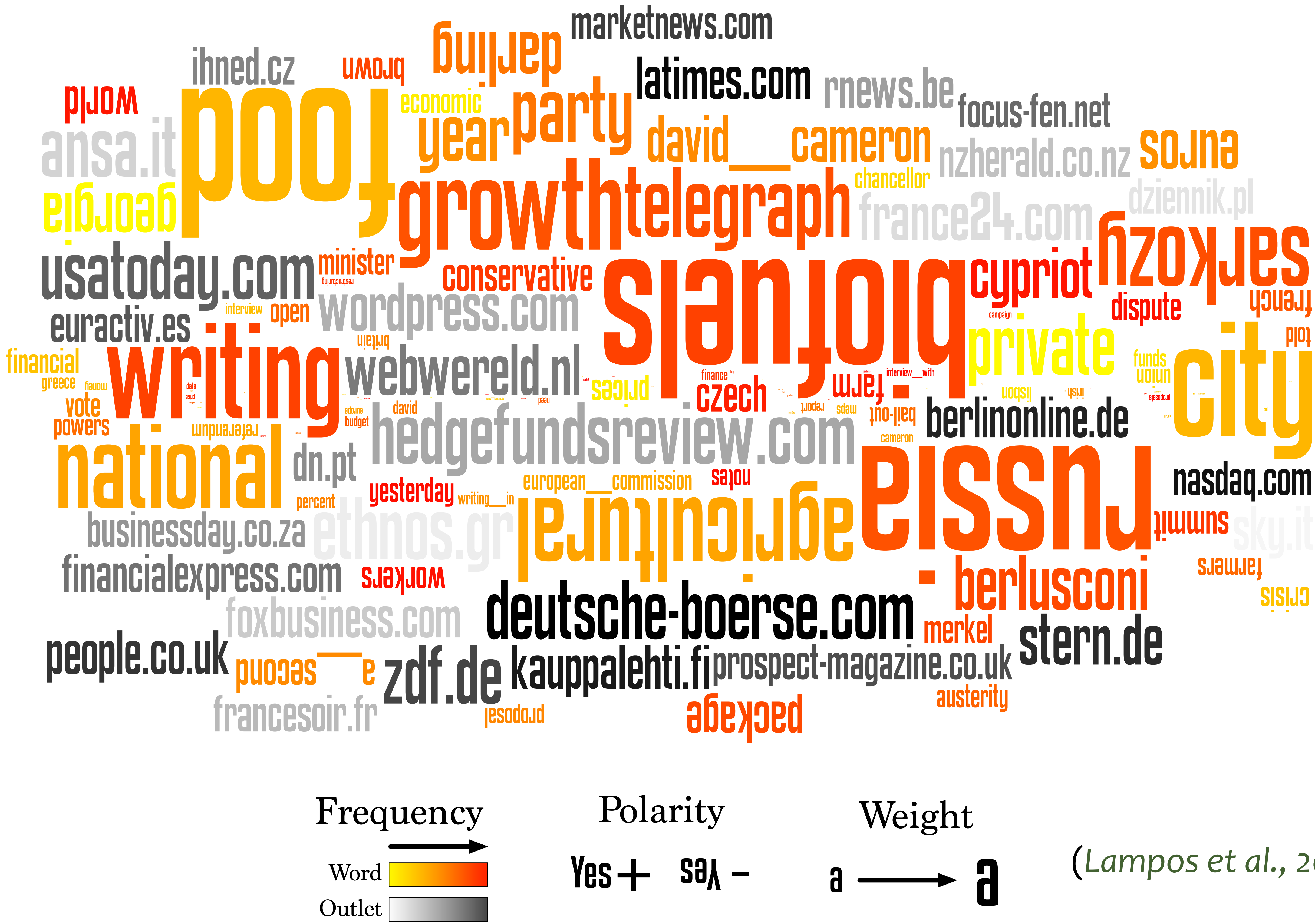
+ **FISTA** implemented in **SPAMS** library (*Beck & Teboulle, 2009*; *Mairal et al., 2010*)

Global objective function during training (***red***)

Corresponding prediction error on held out data (***blue***)



Bilinear modelling of EU unemployment via news summaries



(*Lampos et al., 2014*)

More information about Gaussian Processes

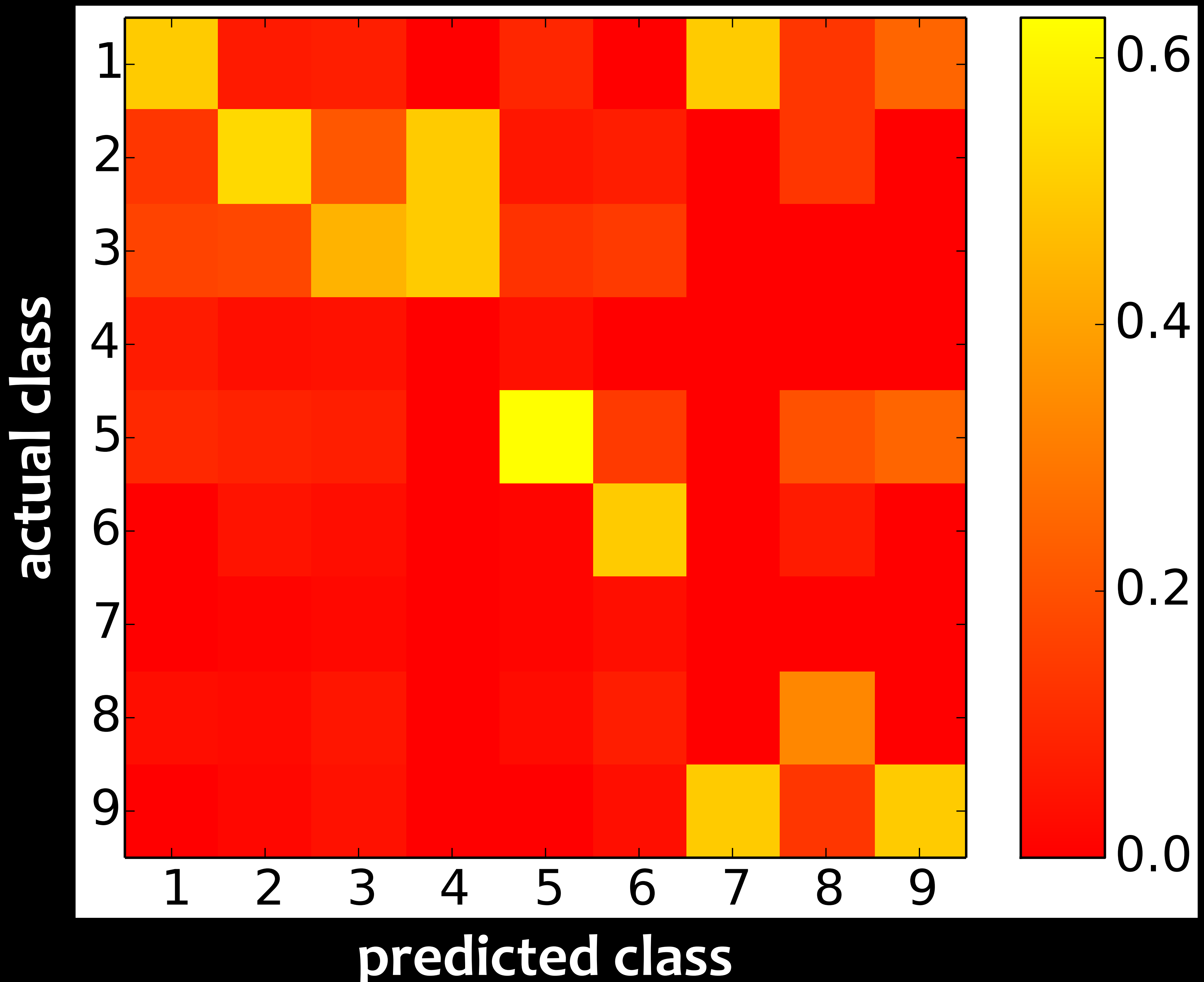
+ non-linear, kernelised, non-parametric, modular
+ applicable in both regression and classification scenarios
+ interpretable

Pointers

+ Book — "*Gaussian Processes for Machine Learning*"
  http://www.gaussianprocess.org/gpml/
+ Tutorial — "*Gaussian Processes for Natural Language Processing*"
  http://people.eng.unimelb.edu.au/tcohn/tutorial.html
+ Video-lecture — "*Gaussian Process Basics*"
  http://videolectures.net/gpip06_mackay_gpb/
+ Software I — *GPML* for Octave or MATLAB
  http://www.gaussianprocess.org/gpml/code
+ Software II — *GPy* for Python
  http://sheffieldml.github.io/GPy/

(*Rasmussen & Williams, 2006*)

Occupational class (9-way) classification confusion matrix



References

Acerbi, Lampos, Garnett & Bentley. The Expression of Emotions in 20th Century Books. PLoS ONE, 2013.

Bach. Bolasso: Model Consistent Lasso Estimation through the Bootstrap. ICML, 2008.

Beck & Teboulle. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. J. Imaging. Sci., 2009.

Bentley, Acerbi, Ormerod & Lampos. Books Average Previous Decade of Economic Misery. PLoS ONE, 2014.

Bouma. Normalized (pointwise) mutual information in collocation extraction. GSCL, 2009.

Caruana. Multi-task Learning. Machine Learning, 1997.

Labov. The Social Stratification of English in New York City, 1972; 2nd ed. 2006.

Lampos & Cristianini. Tracking the flu pandemic by monitoring the Social Web. CIP, 2010.

Lampos & Cristianini. Nowcasting Events from the Social Web with Statistical Learning. ACM TIST, 2012.

Lampos, De Bie & Cristianini. Flu Detector - Tracking Epidemics on Twitter. ECML PKDD, 2010.

Lampos, Preotiuc-Pietro & Aletras. Predicting and characterising user impact on Twitter. EACL, 2014.

Lampos, Preotiuc-Pietro & Cohn. A user-centric model of voting intention from Social Media. ACL, 2013.

Lampos, Yom-Tov, Pebody & Cox. Assessing the impact of a health intervention via user-generated Internet content. DMKD, 2015.

Lampos et al. Extracting Socioeconomic Patterns from the News: Modelling Text and Outlet Importance Jointly. ACL LACSS, 2014.

Mairal, Jenatton, Obozinski & Bach. Network Flow Algorithms for Structured Sparsity. NIPS, 2010.

Mikolov, Chen, Corrado and Dean. Efficient estimation of word representations in vector space. ICLR, 2013.

Pennebaker et al. Linguistic Inquiry and Word Count: LIWC2007, Tech. Rep., 2007.

Preotiuc-Pietro, Lampos & Aletras. An analysis of the user occupational class through Twitter content. ACL, 2015.

Rasmussen & Williams. Gaussian Processes for Machine Learning. MIT Press, 2006.

Strapparava & Valitutti. Wordnet-Affect: An affective extension of WordNet. LREC, 2004.

Tibshirani. Regression shrinkage and selection via the lasso. JRSS Series B (Method.), 1996.

von Luxburg. A tutorial on spectral clustering. Statistics and Computing, 2007.

Zhao & Yu. On model selection consistency of lasso. JMLR, 2006.