

A user-centric model of voting intention from Social Media

Vasileios Lamos, Daniel Preoțiu-Pietro & Trevor Cohn

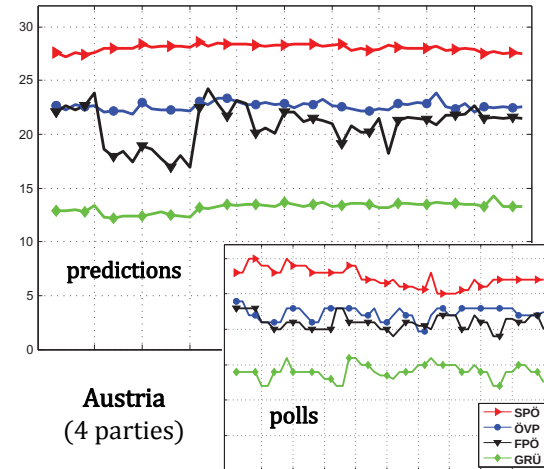
Computer Science Department, University of Sheffield, UK

Trend Miner

$$\begin{matrix} \begin{bmatrix} u_{p1} \\ u_{p2} \\ \vdots \\ u_{pk} \end{bmatrix}^T \downarrow \begin{matrix} k \text{ users} \end{matrix} \end{matrix} \begin{matrix} \xrightarrow{m \text{ words}} \end{matrix} \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{k1} & x_{k2} & \cdots & x_{km} \end{bmatrix} \begin{matrix} \begin{bmatrix} w_{p1} \\ w_{p2} \\ \vdots \\ w_{pm} \end{bmatrix} \end{matrix}$$



x_{ij} : frequency of word j for user i during time interval t



$$\min_{W, U, \beta} \sum_{p=1}^n \sum_{t=1}^{\tau} (\mathbf{u}_p^T \mathbf{Q}_t \mathbf{w}_p + \beta_p - y_{tp})^2 + \lambda_1 \sum_{j=1}^m \|\mathbf{w}_j\|_2 + \lambda_2 \sum_{j=1}^k \|\mathbf{u}_j\|_2$$

bi-linear

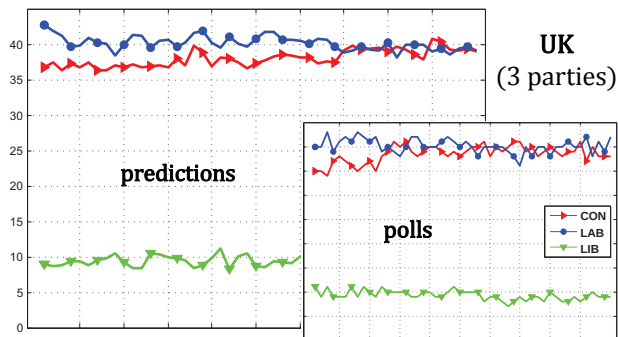
voting intention % for political party p during time interval t

regularisation parameter for word weights

filtering out words & users

$\ell_{2,1}$ -norm $\mathbf{u}_j$: j th row of $\mathbf{U}$

$$\begin{aligned} \mathbf{W} &= [\mathbf{w}_1 \dots \mathbf{w}_p \dots \mathbf{w}_n] \in \mathbb{R}^{m \times n} \\ \mathbf{U} &= [\mathbf{u}_1 \dots \mathbf{u}_p \dots \mathbf{u}_n] \in \mathbb{R}^{k \times n} \\ \beta &\in \mathbb{R}^n \quad \lambda_1, \lambda_2 \in \mathbb{R}^{>0} \\ \mathbf{Q}_t &\in \mathbb{R}^{k \times m} \quad \mathbf{y} \in \mathbb{R}^{\tau \times n} \end{aligned}$$



Bi-convex iterative learning

1. Solve $\min_{W, \beta}$
2. Fix $\mathbf{W}$ and solve $\min_{U, \beta}$
3. Fix $\mathbf{U}$ and solve $\min_{W, \beta}$
4. Validate ? Go to Step 2 : END

prediction performance

RMSE (%)	Method	Austria	UK
training set benchmark	mean(poll)	1.851	1.69
	Last poll	1.47	1.723
$\mathbf{w}$	Linear	1.442	3.067
$\mathbf{w}, \mathbf{u}$	Bilinear	1.699	1.573
$\mathbf{W}, \mathbf{U}$	Bilinear Multi-task	1.439	1.478

1.1K users
23K words

42K users
81K words