Nowcasting User Behaviour with Social Media and Smart Devices

Adam Tsakalidis

The Alan Turing Institute

29/11/2018

Invited Talk

Overview

- **Introduction**
- **Part 1: Nowcasting Political Indices Using Social Media**
  - Predicting Election Results with Social Media
  - Nowcasting the Political Stance of Social Media Users
- **Part 2: Nowcasting Mental Health Using Heterogeneous Data**
  - Using Heterogeneous User Generated Content to Sense Well-being
  - Challenges in Assessing Mental Health using User Generated Content
- **Conclusion & Future Directions**

Overview

- **Introduction**
- **Part 1: Nowcasting Political Indices Using Social Media**
  - Predicting Election Results with Social Media
  - Nowcasting the Political Stance of Social Media Users
- **Part 2: Nowcasting Mental Health Using Heterogeneous Data**
  - Using Heterogeneous User Generated Content to Sense Well-being
  - Challenges in Assessing Mental Health using User Generated Content
- **Conclusion & Future Directions**

Introduction

Urban state monitoring

- Political preferences, health state of population
- Opinion polls, visits to doctors…
- Time consuming, manual effort, sparse…

Introduction

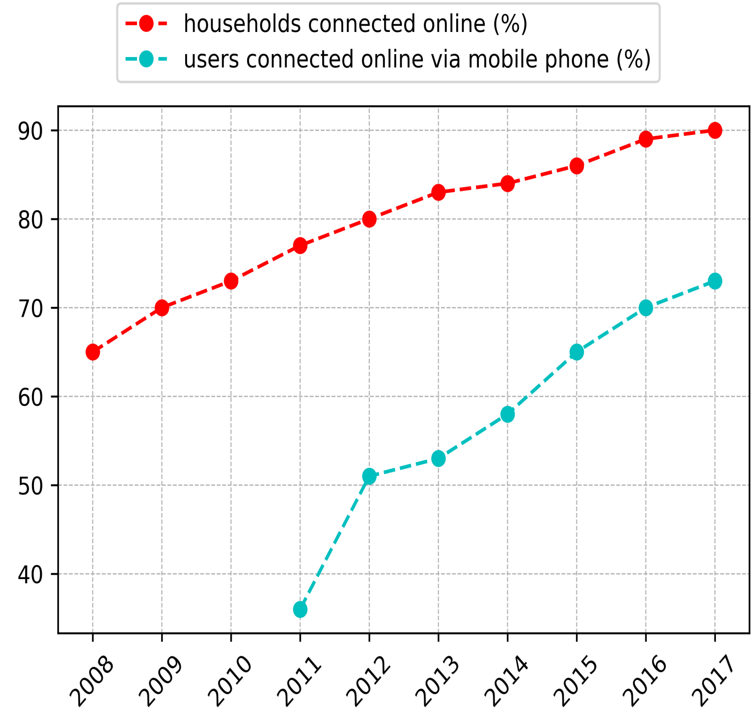
Urban state monitoring

- Political preferences, health state of population
- Opinion polls, visits to doctors…
- Time consuming, manual effort, sparse…

User-Generated Content (UGC)

- Large volumes in real-time
- Social media, smart devices

	Facebook	Twitter	Instagram	Sina Weibo	LinkedIn
Users	2.2B	0.3B	1.0B	0.4B	0.3B



➢ UGC: Real time "sensors"

Goal & Challenges

Aim: "nowcast" real-world and user-specific indices using UGC

- Macro-level (population)
- Micro-level (individual)

Longitudinal, asynchronous, heterogeneous data sources

Goal & Challenges

Aim: "nowcast" real-world and user-specific indices using UGC

- Macro-level (population)
- Micro-level (individual)

Longitudinal, asynchronous, heterogeneous data sources

Goal & Challenges

Aim: "nowcast" real-world and user-specific indices using UGC

- Macro-level (population)
- Micro-level (individual)

Longitudinal, asynchronous, heterogeneous data sources

Challenges

Generalise to new domains

Work under a real-world setting

Overview

- Introduction
- **Part 1: Nowcasting Political Indices Using Social Media**
  - **Predicting Election Results with Social Media**
  - Nowcasting the Political Stance of Social Media Users
- **Part 2: Nowcasting Mental Health Using Heterogeneous Data**
  - Using Heterogeneous User Generated Content to Sense Well-being
  - Challenges in Assessing Mental Health using User Generated Content
- **Conclusion & Future Directions**

Predicting Elections Using Social Media

Task: Predict the 2014 EU election results using UGC

Predicting Elections Using Social Media

Task: Predict the 2014 EU election results using UGC

Challenges

- Ability to generalise: *multiple* (three) electoral cases
- Real-world setting: predict *before* the election day

Predicting Elections Using Social Media

Task: Predict the 2014 EU election results using UGC

Challenges

- Ability to generalise: *multiple* (three) electoral cases
- Real-world setting: predict *before* the election day

Methodology

- Datasets: keyword-based (Twitter)
- Algorithms: LR, GP, SMO (average)
- Ground Truth: Opinion Polls
- Features: *(a)* past opinion polls *(b)* count-based *(c)* sentiment-based

Predicting Elections Using Social Media

Task: Predict the 2014 EU election results using UGC

Challenges

- Ability to generalise: *multiple* (three) electoral cases
- Real-world setting: predict *before* the election day

Methodology

- Datasets: keyword-based (Twitter)
- Algorithms: LR, GP, SMO (average)
- Ground Truth: Opinion Polls
- $\boxed{\text{Features: } \textit{(a)} \text{ past opinion polls} \quad \textit{(b)} \text{ count-based} \quad \textit{(c)} \text{ sentiment-based}}$

Real-world

Twitter-based

Results

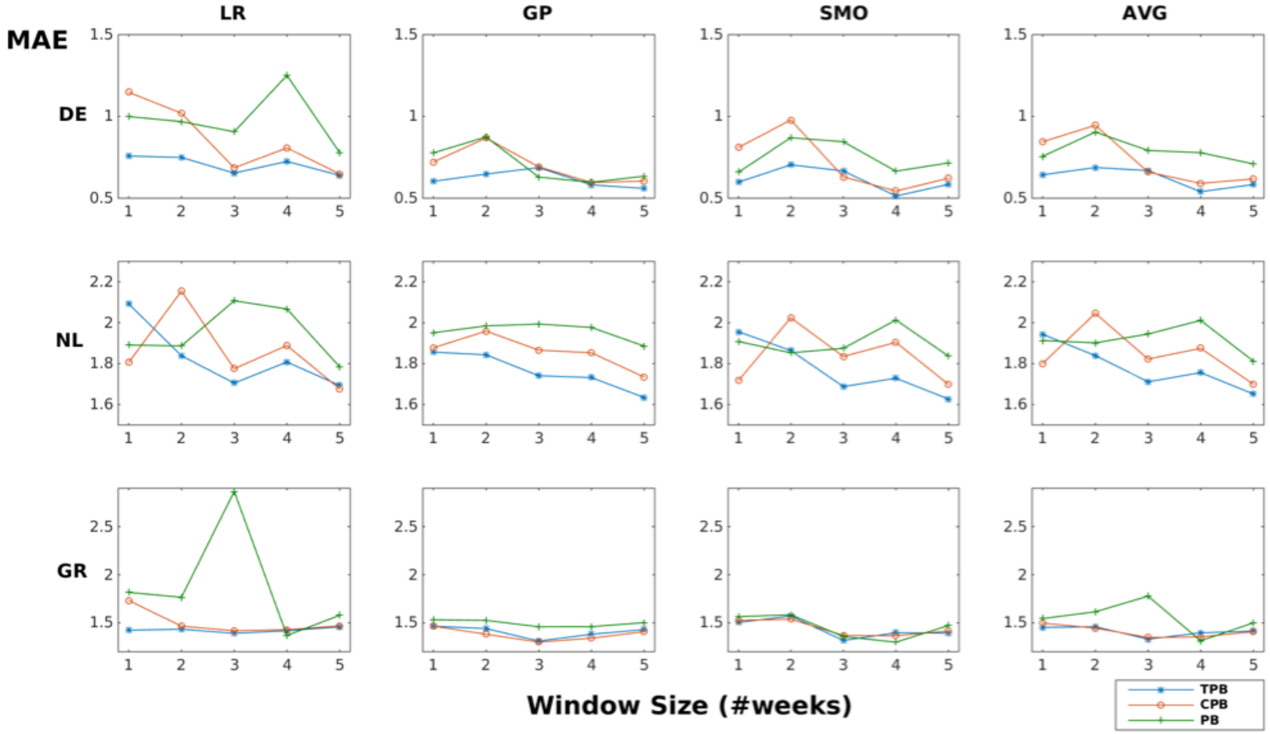
	Past Work			Polls	Websites		Time Series Models		
	CB1	**CB2**	**SB**	**Polls**	**MP**	**PW**	**PB**	**CPB**	**TPB**
MAE	4.87	5.19	2.59	1.57	1.39	1.73	1.40	1.38	**1.35**

PB: No Twitter Features
CPB: No Sentiment Features
TPB: All Features

Results

	Past Work			Polls	Websites		Time Series Models		
	CB1	CB2	SB	Polls	MP	PW	PB	CPB	TPB
MAE	4.87	5.19	2.59	1.57	1.39	1.73	1.40	1.38	**1.35**

PB: No Twitter Features
CPB: No Sentiment Features
TPB: All Features



TPB vs PB:
significant differences (.05)

TPB vs CPB:
significant differences (.05)
except for SMO

Issues

Reference: Tsakalidis, Adam, et al. "Predicting elections for multiple countries using Twitter and polls." *IEEE Intelligent Systems* 30.2 (2015): 10-17.

Issues

Further experiments

- 2015 General Elections (Greece): better estimates than *all* opinions polls

Reference: Tsakalidis, Adam, et al. "Predicting elections for multiple countries using Twitter and polls." *IEEE Intelligent Systems* 30.2 (2015): 10-17.

Issues

Further experiments

- 2015 General Elections (Greece): better estimates than *all* opinions polls
- 2015 General Elections (UK): LAB: 34% (29%)

  CON: 33% (36%)

Reference: Tsakalidis, Adam, et al. "Predicting elections for multiple countries using Twitter and polls." *IEEE Intelligent Systems* 30.2 (2015): 10-17.

Issues

Further experiments

- 2015 General Elections (Greece): better estimates than *all* opinions polls
- 2015 General Elections (UK): LAB: 34% (29%)
  CON: 33% (36%)

Issue

- Small-scale evaluation

Reference: Tsakalidis, Adam, et al. "Predicting elections for multiple countries using Twitter and polls." *IEEE Intelligent Systems* 30.2 (2015): 10-17.

Issues

Further experiments

- 2015 General Elections (Greece): better estimates than *all* opinions polls
- 2015 General Elections (UK): LAB: 34% (29%)
  CON: 33% (36%)

Issue

- Small-scale evaluation

→ **Predict the users' stance**

Reference: Tsakalidis, Adam, et al. "Predicting elections for multiple countries using Twitter and polls." *IEEE Intelligent Systems* 30.2 (2015): 10-17.

Issues

Further experiments

- 2015 General Elections (Greece): better estimates than *all* opinions polls
- 2015 General Elections (UK): LAB: 34% (29%)

  CON: 33% (36%)

Issue

- Small-scale evaluation

→ **Predict the users' stance**

Challenges

- Sudden/major political debates?
- No opinion polls?

Reference: Tsakalidis, Adam, et al. "Predicting elections for multiple countries using Twitter and polls." *IEEE Intelligent Systems* 30.2 (2015): 10-17.

Issues

Further experiments

- 2015 General Elections (Greece): better estimates than *all* opinions polls
- 2015 General Elections (UK): LAB: 34% (29%)
  CON: 33% (36%)

Issue

- Small-scale evaluation

→ **Predict the users' stance**

Challenges

- Sudden/major political debates?
- No opinion polls?

→ **Greek Bailout Referendum (2015)**

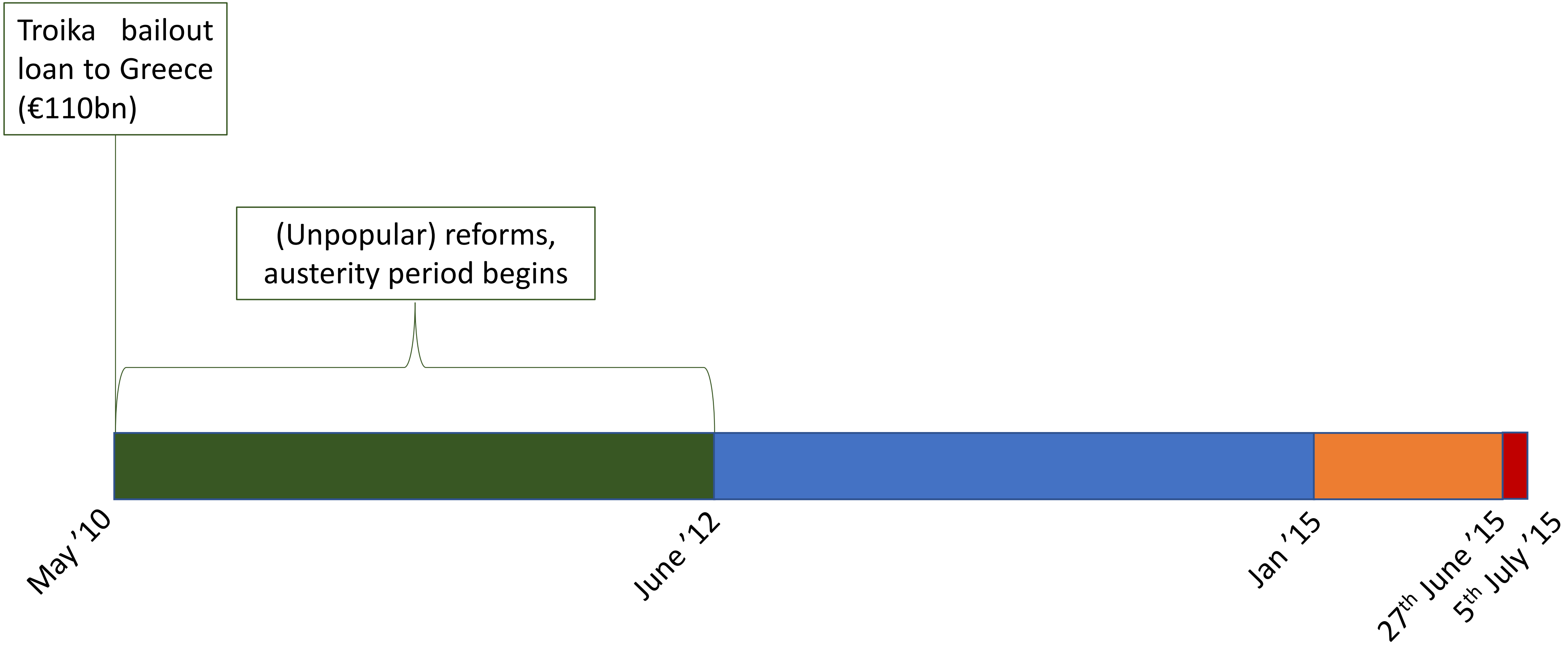
Reference: Tsakalidis, Adam, et al. "Predicting elections for multiple countries using Twitter and polls." *IEEE Intelligent Systems* 30.2 (2015): 10-17.

Overview

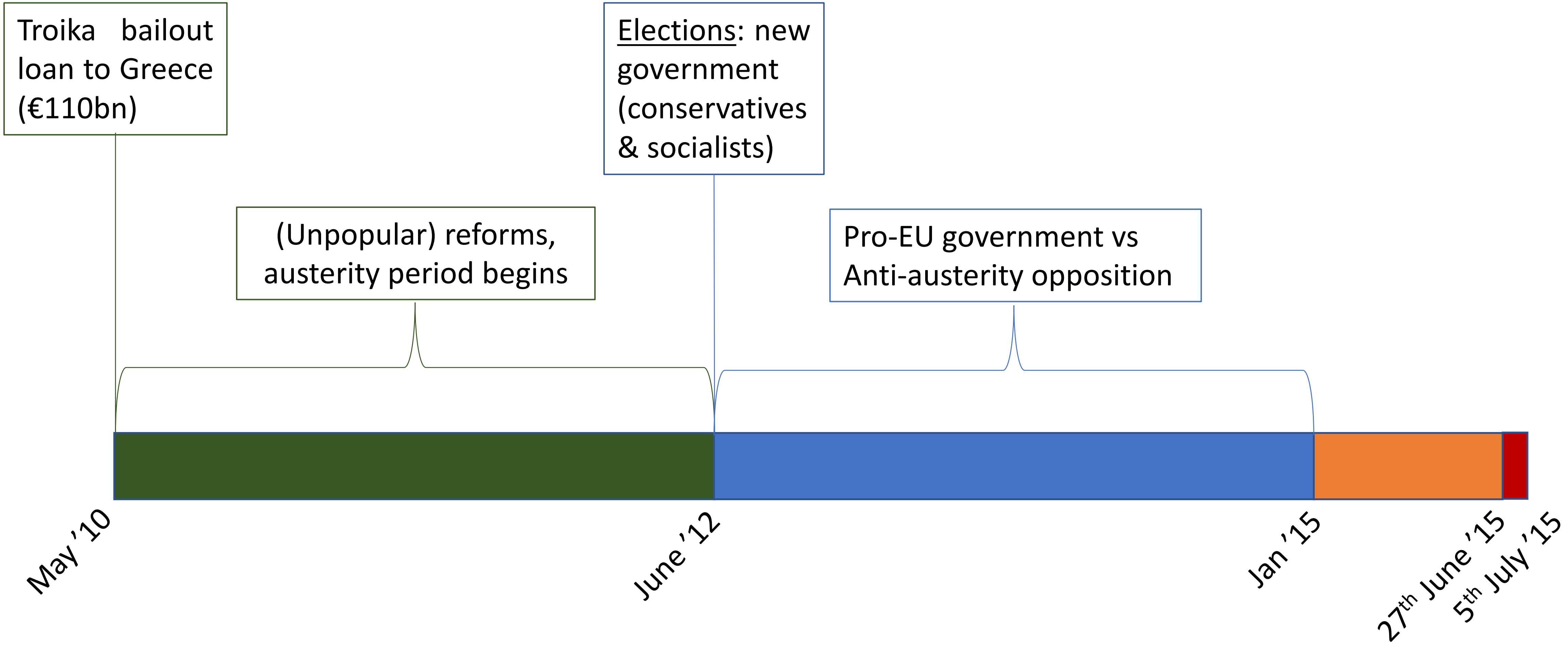
- Introduction
- **Part 1: Nowcasting Political Indices Using Social Media**
  - Predicting Election Results with Social Media
  - **Nowcasting the Political Stance of Social Media Users**
- Part 2: Nowcasting Mental Health Using Heterogeneous Data
  - Using Heterogeneous User Generated Content to Sense Well-being
  - Challenges in Assessing Mental Health using User Generated Content
- Conclusion & Future Directions

The Greek Referendum: Background

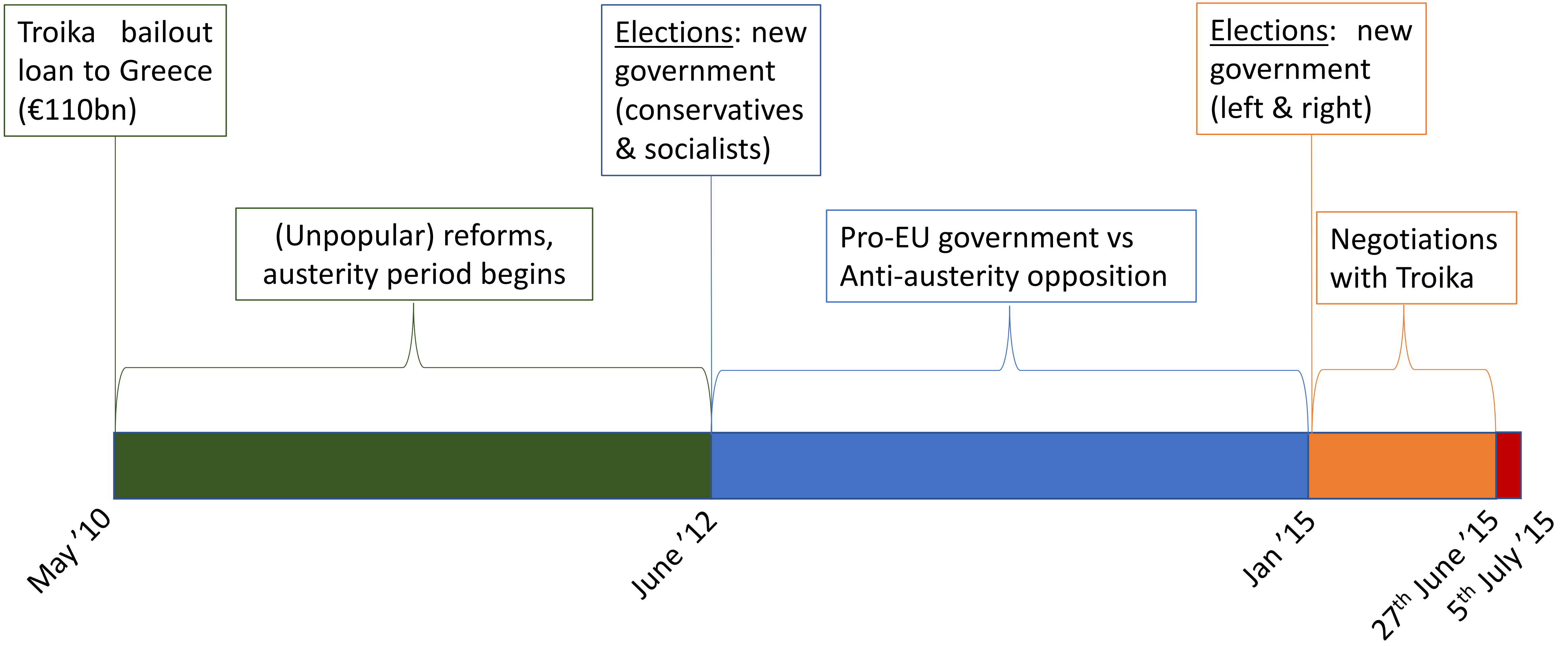
The Greek Referendum: Background



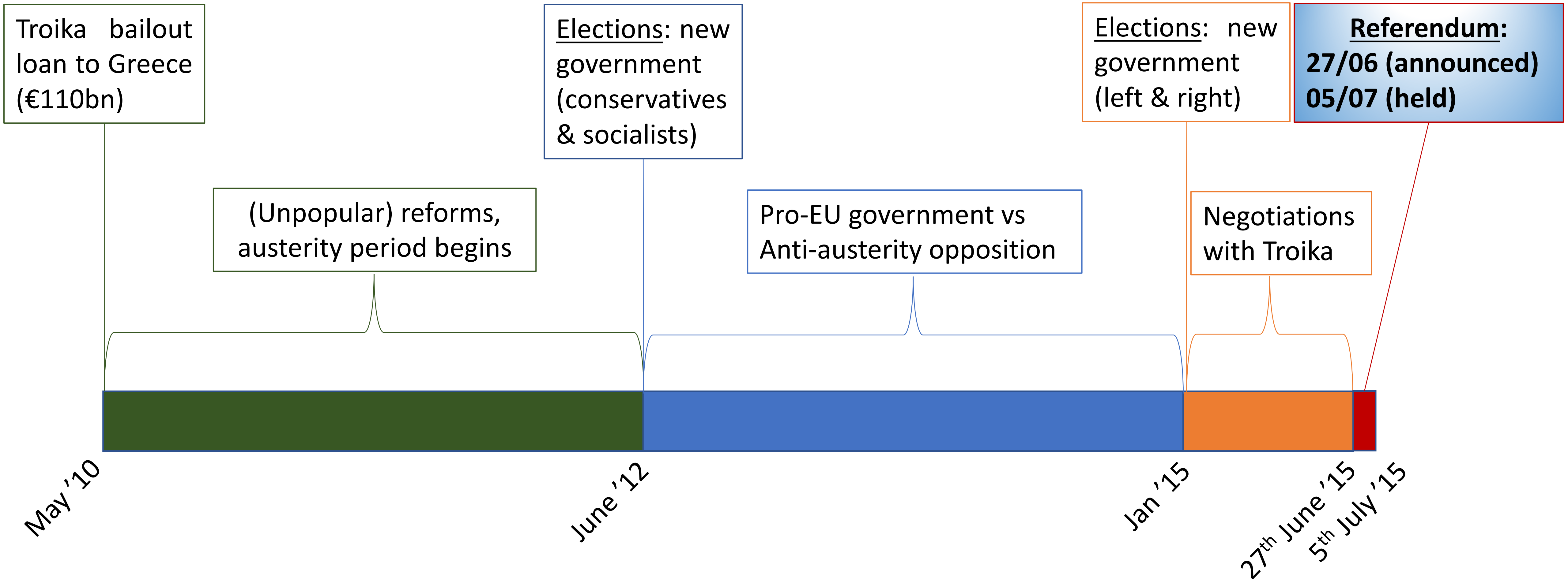
The Greek Referendum: Background



The Greek Referendum: Background



The Greek Referendum: Background



27th June-5th July extreme polarisation, demonstrations, capital controls…

Task Description

Aim: *nowcast the stance (YES/NO) of social media users* (binary task over time)

Task Description

Aim: *nowcast the stance (YES/NO) of social media users* (binary task over time)

For *t* in [0, 1, …, 8]: *#[26/06-04/07]*

1. Assume training set of *n* users at *t*:
2. Learn a function at *t*:
3. Evaluate performance on test users at *t*

$$D_t = \{(x_t^{(1)}, y^{(1)}), \ldots, (x_t^{(n)}, y^{(n)})\}$$
$$\hat{y} = f_t(x_t^{(j)})$$

Task Description

Aim: *nowcast the stance (YES/NO) of social media users* (binary task over time)

For t in [0, 1, …, 8]: *#[26/06-04/07]*

1. Assume training set of n users at t: $D_t = \{(x_t^{(1)}, y^{(1)}), \ldots, (x_t^{(n)}, y^{(n)})\}$
2. Learn a function at t: $\hat{y} = f_t(x_t^{(j)})$
3. Evaluate performance on test users at t

Challenges:

- real-world application; no manual effort – training set?
- combine *asynchronous* and *time sensitive* information sources?

Dataset

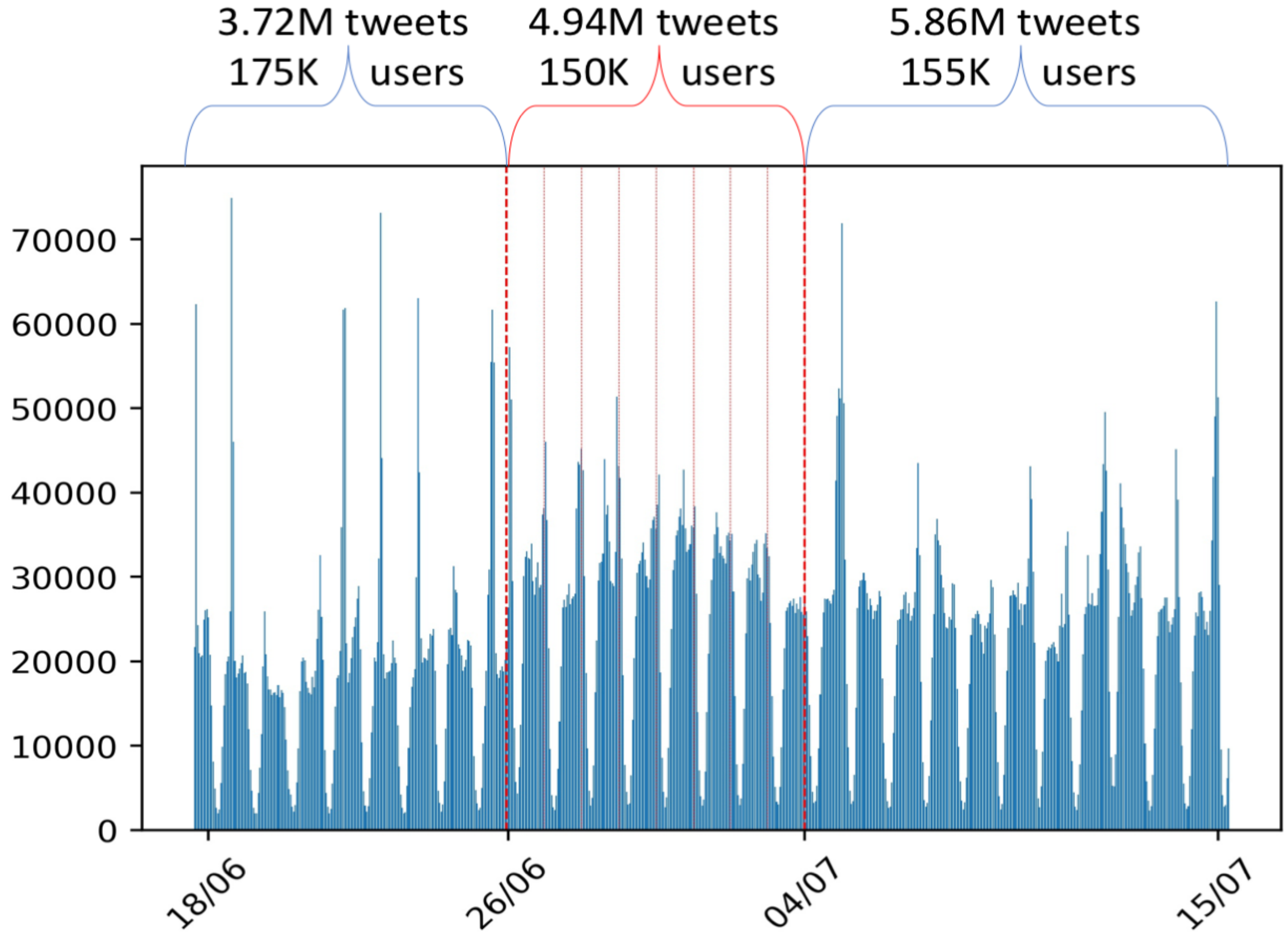
Twitter Streaming API (Greek stop-words)

- 14.6M tweets
- 304K users

Dataset

Twitter Streaming API (Greek stop-words)

- 14.6M tweets
- 304K users



Dataset

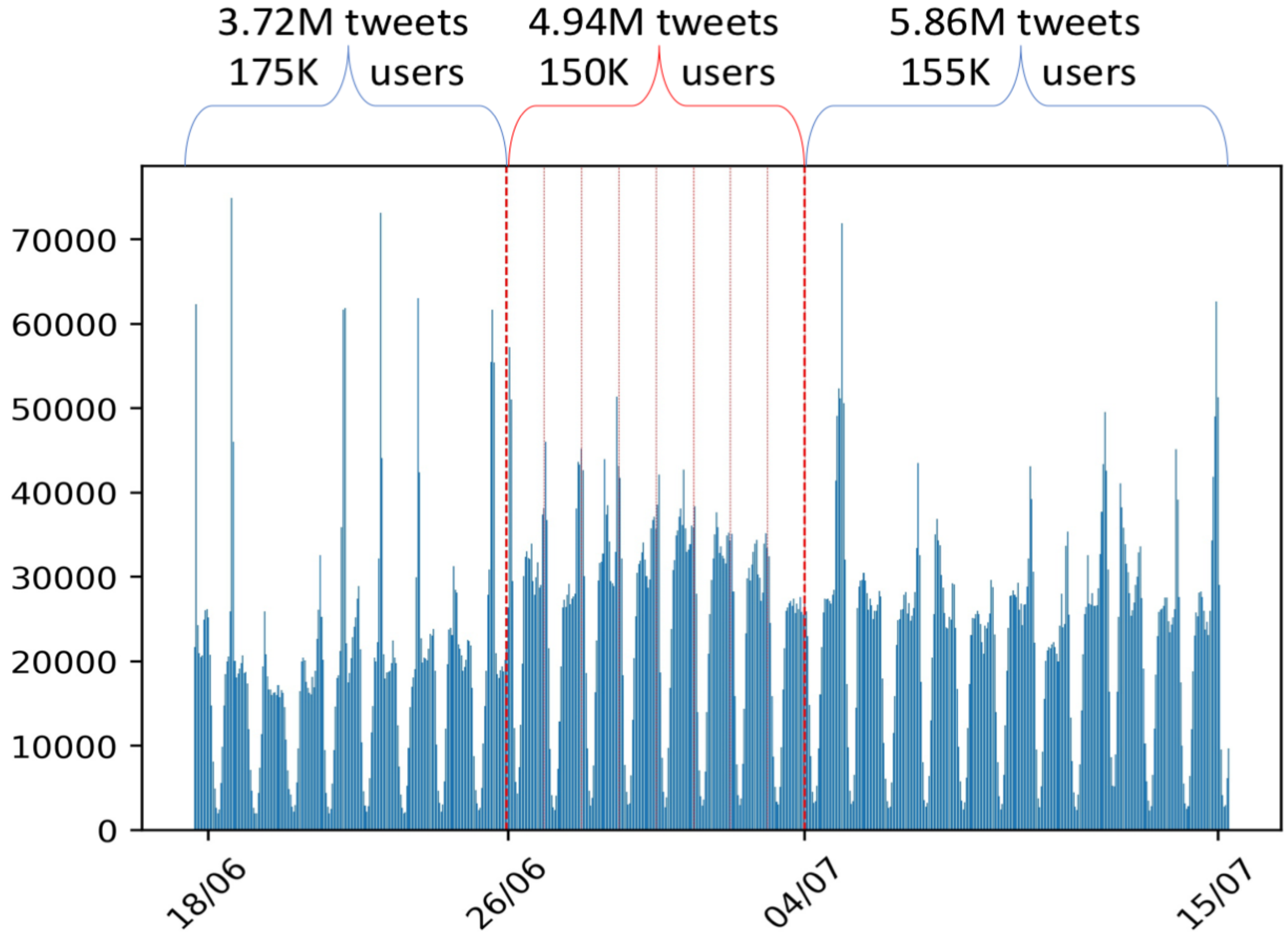
Twitter Streaming API (Greek stop-words)

- 14.6M tweets
- 304K users

Test set: annotate 2.7K users

- >10 tweets; >3 polarised hashtags
- Annotation: YES/NO (or N/A)
- Cohen's κ: .98 (.74)

- Final set: 2,197 users
- 77/23 NO/YES (result: 61/39)



Training Set

Seed accounts with "known" labels:

- political parties (pro-EU vs anti-austerity)
- MPs
- keyword-based profile search

148/119
YES/NO

Training Set

Seed accounts with “known” labels:

- political parties (pro-EU vs anti-austerity)
- MPs
- keyword-based profile search

148/119
YES/NO

Assumption: user RTs pro-EU accounts => vote for “YES”

Training Set

Seed accounts with “known” labels:

- political parties (pro-EU vs anti-austerity)
- MPs
- keyword-based profile search

148/119
YES/NO

Assumption: user RTs pro-EU accounts => vote for “YES”

$$scores = [PMI(u, YES) - PMI(u, NO), \forall u]$$

Training Set

Seed accounts with “known” labels:

- political parties (pro-EU vs anti-austerity)
- MPs
- keyword-based profile search

148/119
YES/NO

Assumption: user RTs pro-EU accounts => vote for “YES”

$$scores = [PMI(u, YES) - PMI(u, NO), \forall u]$$

$$scores[u] > t \Rightarrow YES$$

965 YES

$$scores[u] < -t \Rightarrow NO$$

1,165 NO

Data (summary)

	#users	#tweets	YES (%)	NO (%)
Train	2,121	867K	45%	55%
Test	2,197	768K	23%	77%

Methodology

User

Convolution kernels

text

network

K_w, K_{wt}

K_n, K_{nt}

Kernel combination

$$w_w K_w + w_{wt} K_{wt} + w_n K_n + w_{nt} K_{nt}$$

Convolution Kernels

Z_i: user representations (based on text or network structure) over times T_i:

$$Z_a = \{z_a^1, \ldots, z_a^N\} \qquad T_a = \{t_a^1, \ldots, t_a^N\}$$

$$Z_b = \{z_b^1, \ldots, z_b^M\} \qquad T_b = \{t_b^1, \ldots, t_b^M\}$$

Convolution Kernels

Z_i: user representations (based on text or network structure) over times T_i:

$$Z_a = \{z_a^1, \ldots, z_a^N\} \qquad T_a = \{t_a^1, \ldots, t_a^N\}$$

$$Z_b = \{z_b^1, \ldots, z_b^M\} \qquad T_b = \{t_b^1, \ldots, t_b^M\}$$

Convolution kernel(s):

$$K_z(a,b) = \frac{1}{|Z_a|\,|Z_b|} \sum_{i,j} k_z(z_a^i, z_b^j)$$

K_w (TEXT)
K_n (NET)

Convolution Kernels

Z_i: user representations (based on text or network structure) over times T_i:

$$Z_a = \{z_a^1, \dots, z_a^N\} \qquad T_a = \{t_a^1, \dots, t_a^N\}$$

$$Z_b = \{z_b^1, \dots, z_b^M\} \qquad T_b = \{t_b^1, \dots, t_b^M\}$$

Convolution kernel(s):

$$K_z(a,b) = \frac{1}{|Z_a|\,|Z_b|} \sum_{i,j} k_z(z_a^i, z_b^j)$$

K_w (TEXT)
K_n (NET)

Account for time (Lukasik & Cohn, 2016):

$$K_{zt}(a,b) = \frac{1}{|Z_a|\,|Z_b|} \sum_{i,j} k_z(z_a^i, z_b^j) k_t(t_a^i, t_b^j)$$

K_{wt} (TEXT)
K_{nt} (NET)

Kernel Combination

Kernel summation:

$$K(a,b) = \sum_{k=1}^{|K|} K_k(a,b)$$

Kernel Combination

Kernel summation:

$$K(a,b) = \sum_{k=1}^{|K|} K_k(a,b)$$

Multiple (convolution) kernel learning (Sonnenburg et al., 2006):

$$K(a,b) = \sum_{k=1}^{|K|} w_k K_k(a,b)$$

$$f(x) = sign\left(\sum_{i=1}^{N} a_i K(x, x_i) + b\right)$$

Experiments

Features

- TEXT: word2vec (Mikolov et al., 2013); 50 dim; trained on 15M tweets
- NETWORK: LINE (Tang et al., 2015); 50 dim; trained every 12 hours

Experiments

Features

- TEXT: word2vec (Mikolov et al., 2013); 50 dim; trained on 15M tweets
- NETWORK: LINE (Tang et al., 2015); 50 dim; trained every 12 hours

Baselines

- LR, FF, RF, SVM on feature aggregates

Experiments

Features

- TEXT: word2vec (Mikolov et al., 2013); 50 dim; trained on 15M tweets
- NETWORK: LINE (Tang et al., 2015); 50 dim; trained every 12 hours

Baselines

- LR, FF, RF, SVM on feature aggregates

Our models

- Single modality: SVM_w , SVM_{wt} , SVM_n , SVM_{nt}
- Combinations: SVM_+ , MCKL

Experiments

Features

- TEXT: word2vec (Mikolov et al., 2013); 50 dim; trained on 15M tweets
- NETWORK: LINE (Tang et al., 2015); 50 dim; trained every 12 hours

Baselines

- LR, FF, RF, SVM on feature aggregates

Our models

- Single modality: SVM_w , SVM_{wt} , SVM_n , SVM_{nt}
- Combinations: SVM_+ , MCKL

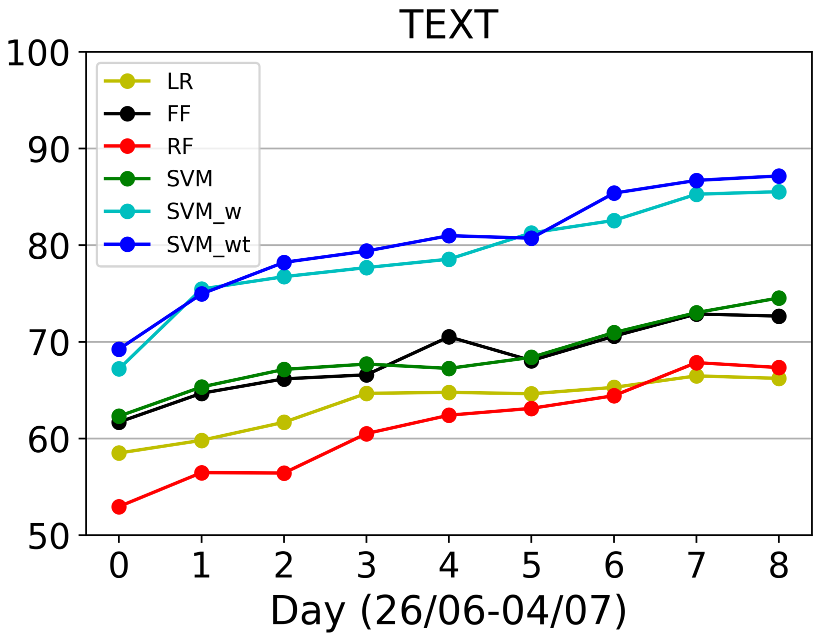
Evaluation

- macro-average F-score
- *nine* evaluation time points (every midnight)

Results

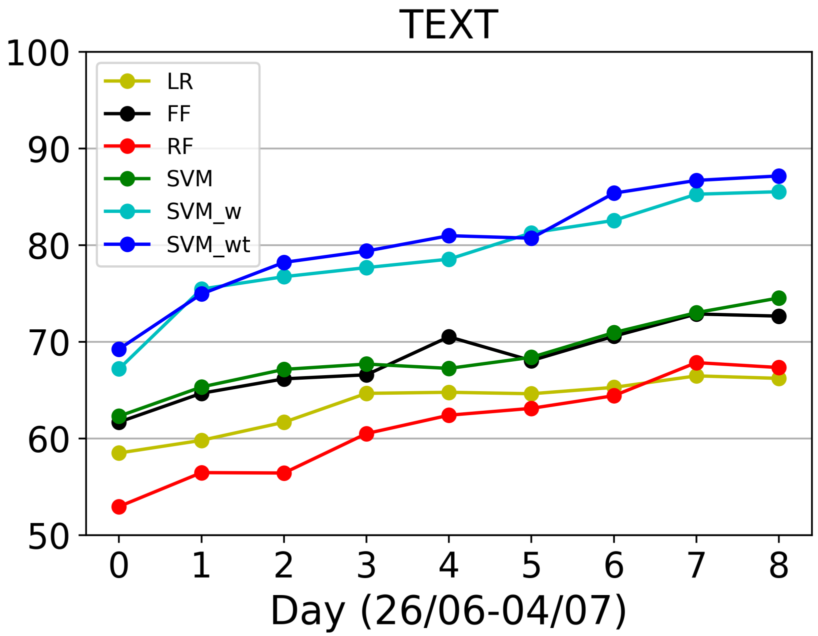
	LR	FF	RF	SVM	SVM_n	SVM_{nt}	SVM_+	MCKL
TEXT								
NET								
BOTH								

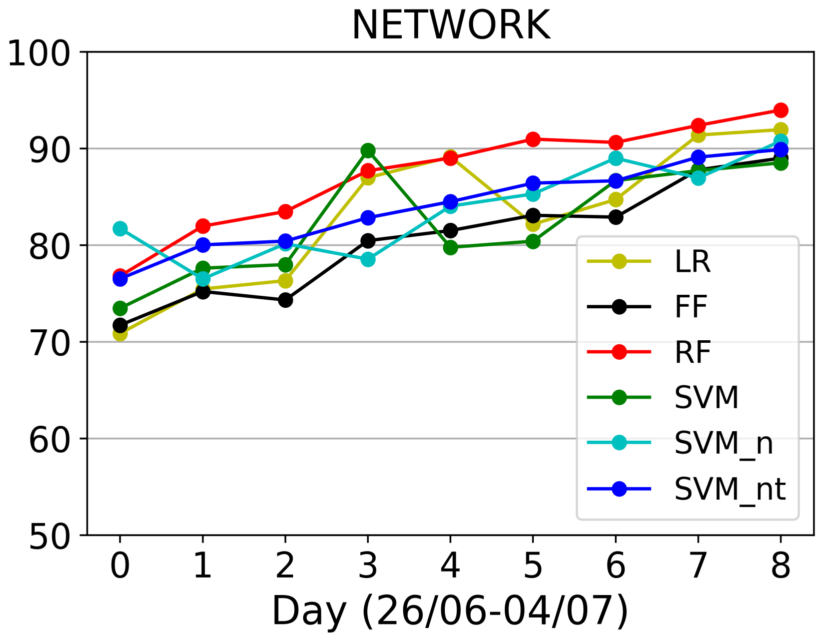
Results



	LR	FF	RF	SVM	SVM_z	SVM_{zt}	SVM_+	MCKL
TEXT	63.55	68.19	61.27	68.51	78.91	**80.30**	--	--
NET								
BOTH								

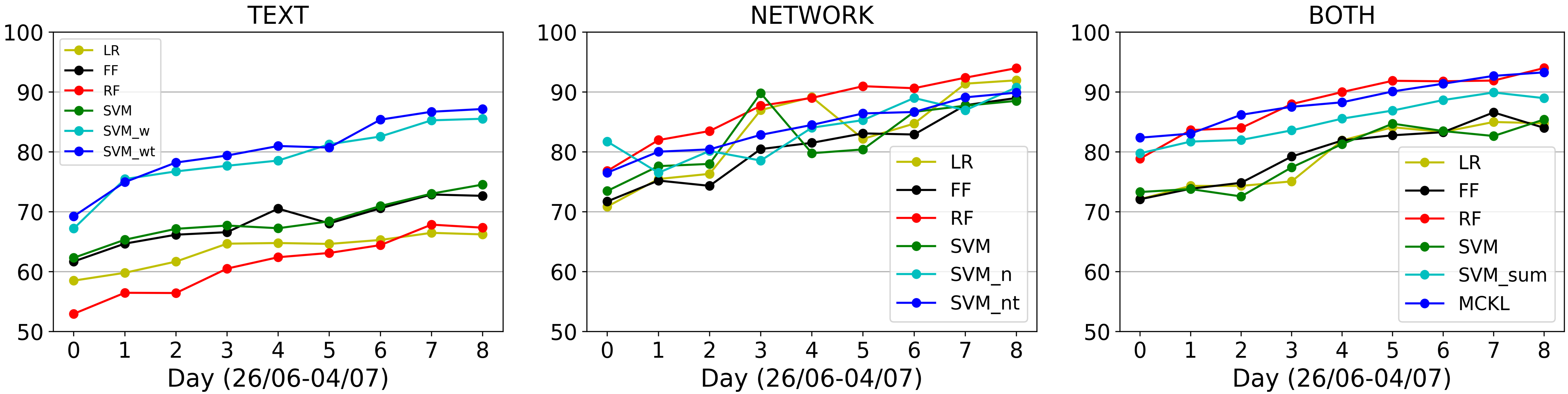
Results





	LR	FF	RF	SVM	SVM_z	SVM_{zt}	SVM_+	MCKL
TEXT	63.55	68.19	61.27	68.51	78.91	**80.30**	--	--
NET	83.21	80.66	**87.43**	82.43	83.65	84.03	--	--
BOTH								

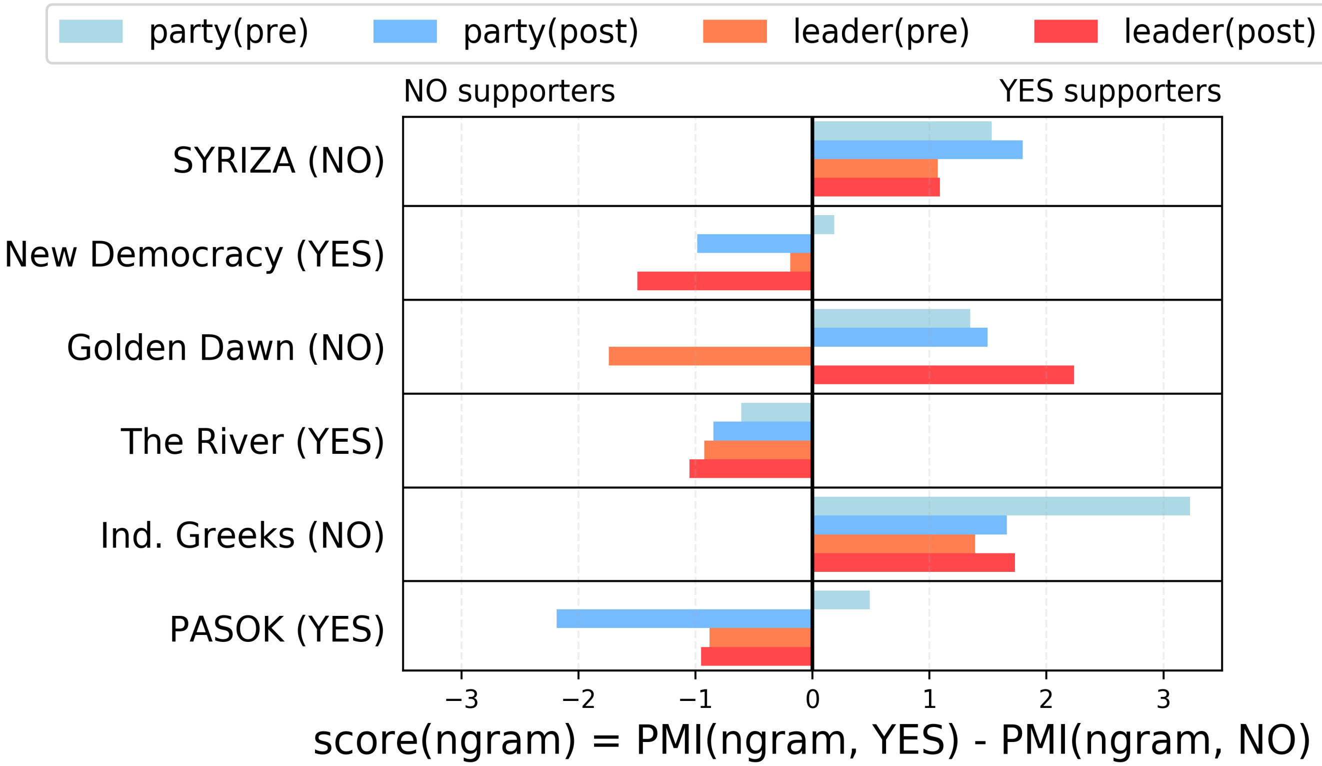
Results



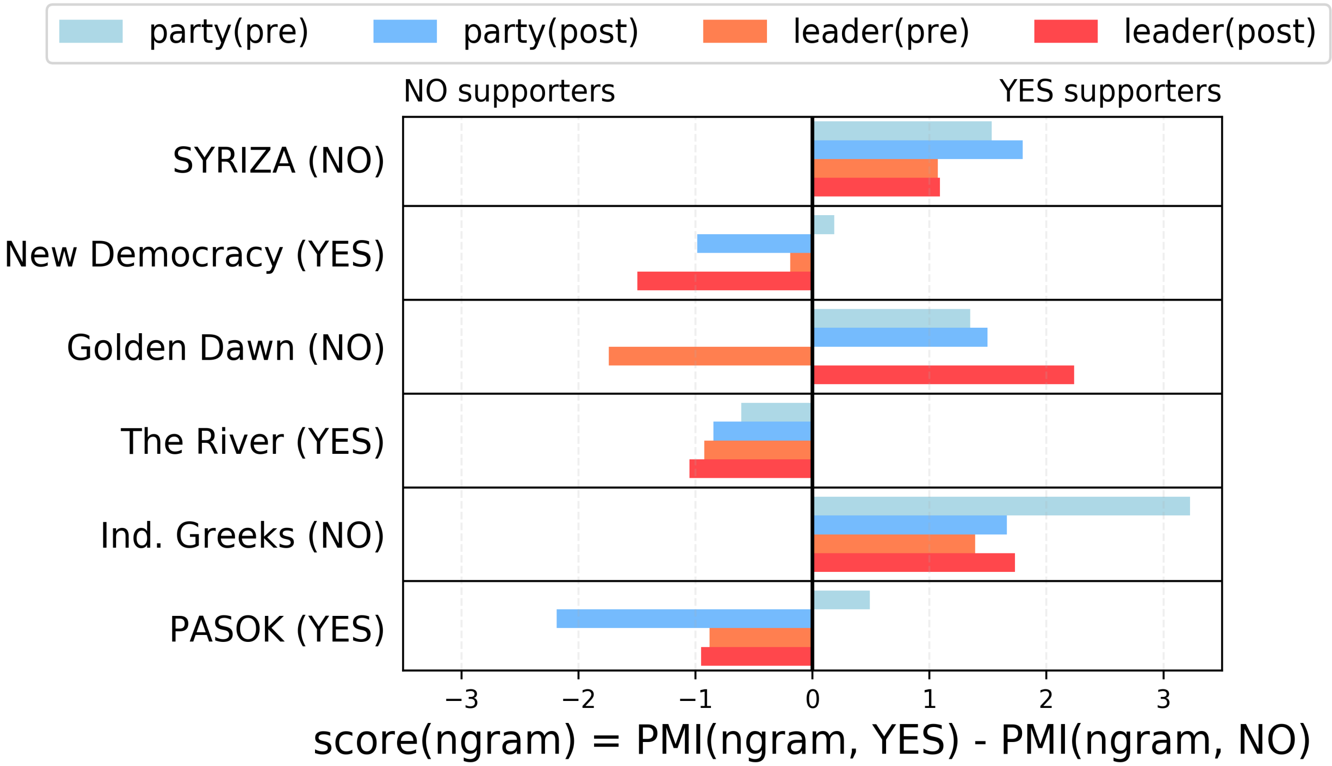
	LR	FF	RF	SVM	SVM_z	SVM_{zt}	SVM_+	MCKL
TEXT	63.55	68.19	61.27	68.51	78.91	**80.30**	--	--
NET	83.21	80.66	**87.43**	82.43	83.65	84.03	--	--
BOTH	79.43	79.83	88.22	79.39	--	--	85.22	**88.31**

Why temporal modelling? (text)

Why temporal modelling? (text)



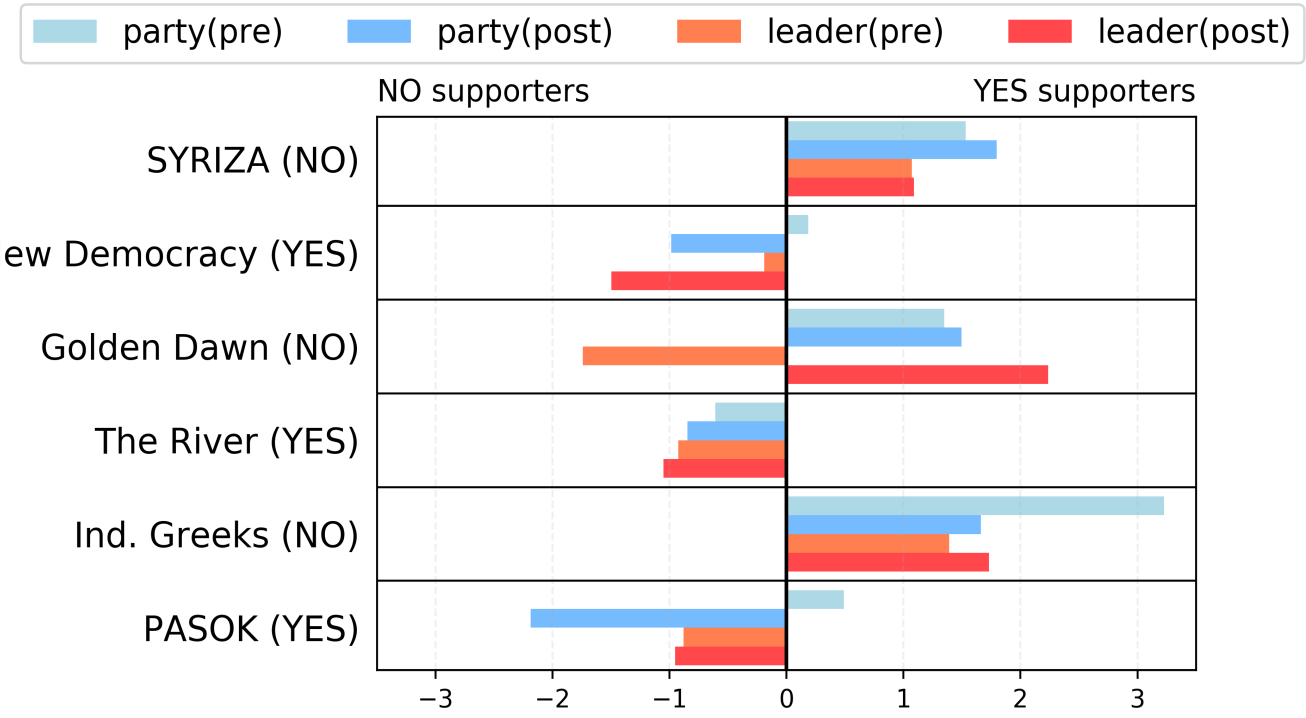
Why temporal modelling? (text)



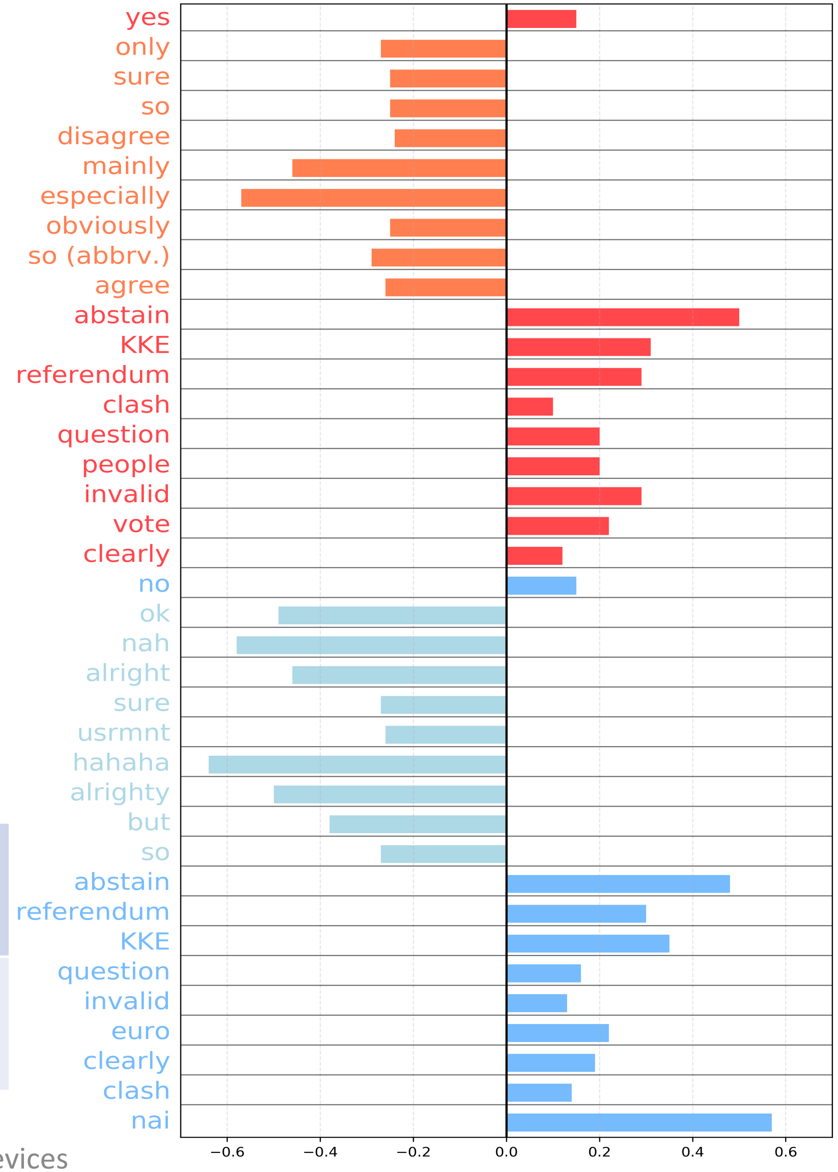
They say that there is a long queue of people in ATMs and they show 6 people waiting; this is not a queue, this is PASOK.

I want to write something funny regarding the statements made by Kammenos, but I cannot find something funnier than the statements made by Kammenos.

Why temporal modelling? (text)



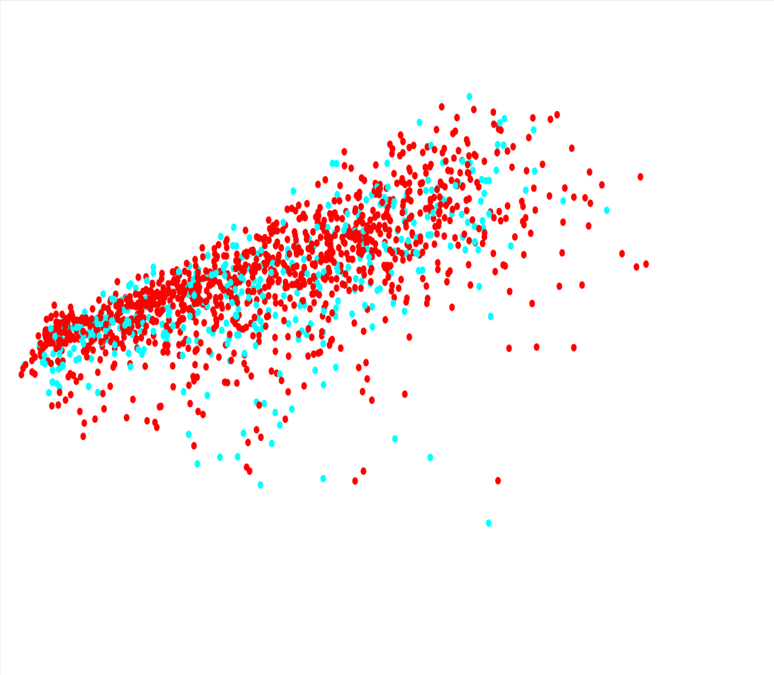
score(ngram) = PMI(ngram, YES) - PMI(ngram, NO)



They say that there is a long queue of people in ATMs and they show 6 people waiting; this is not a queue, this is PASOK.

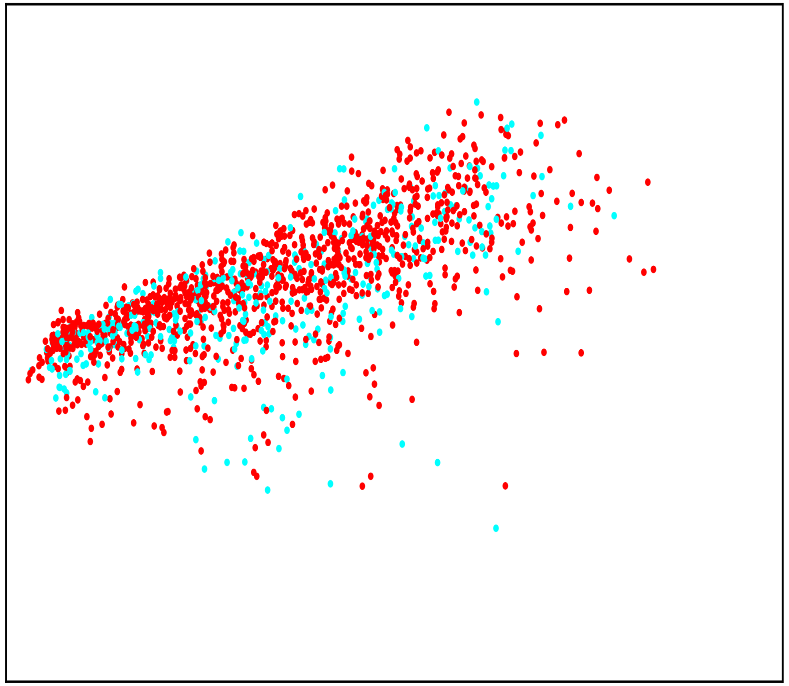
I want to write something funny regarding the statements made by Kammenos, but I cannot find something funnier than the statements made by Kammenos.

Why temporal modelling? (network)

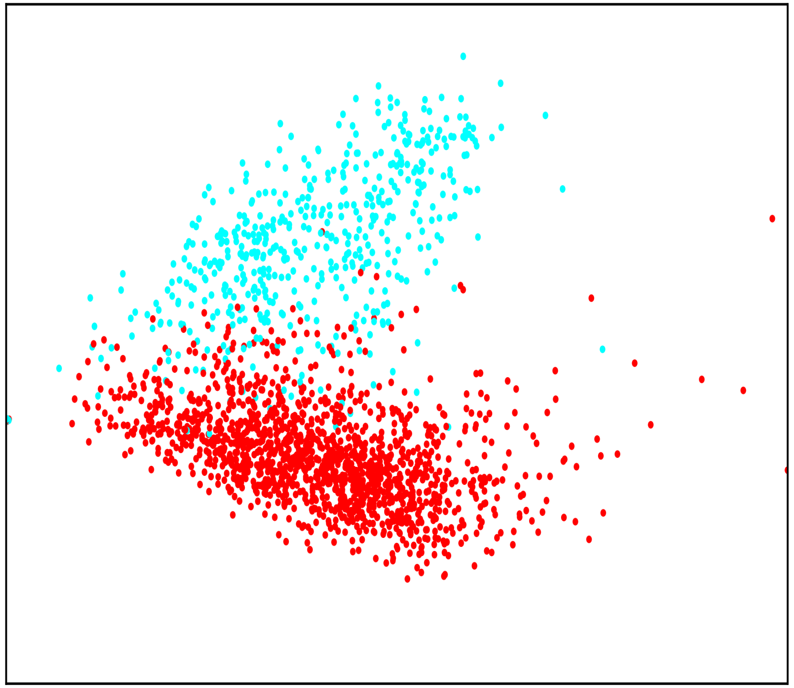


User representations **before** the announcement of the referendum

Why temporal modelling? (network)

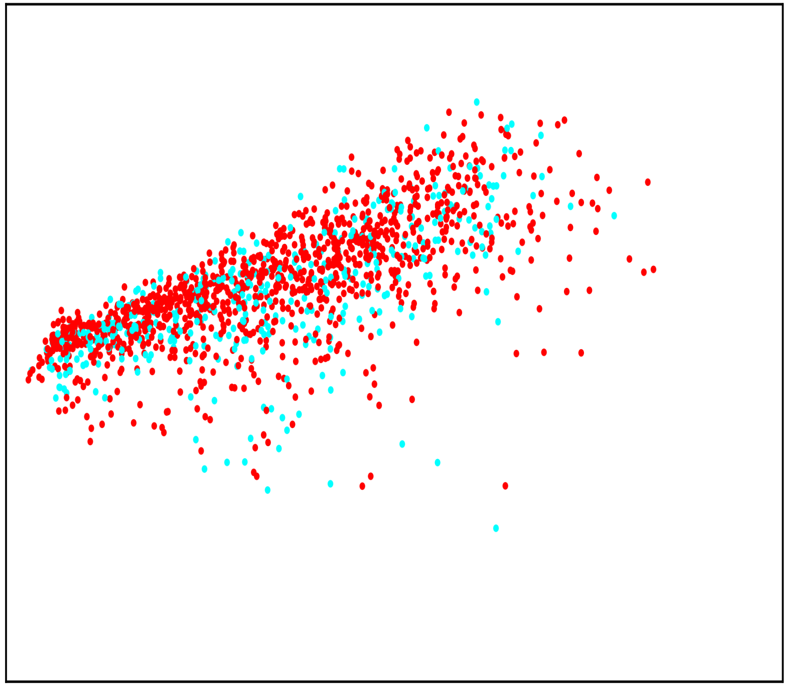


User representations **before** the announcement of the referendum

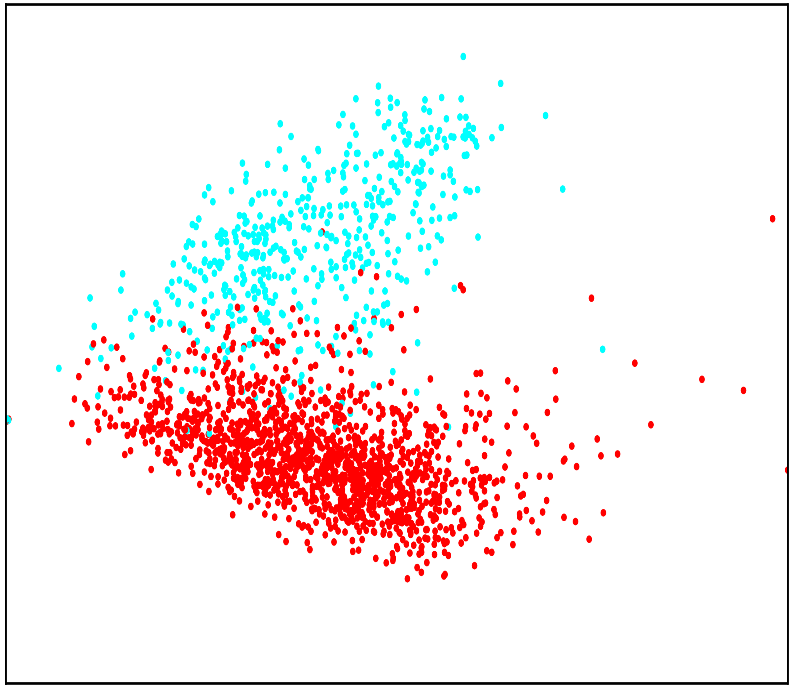


User representations **after** the announcement of the referendum

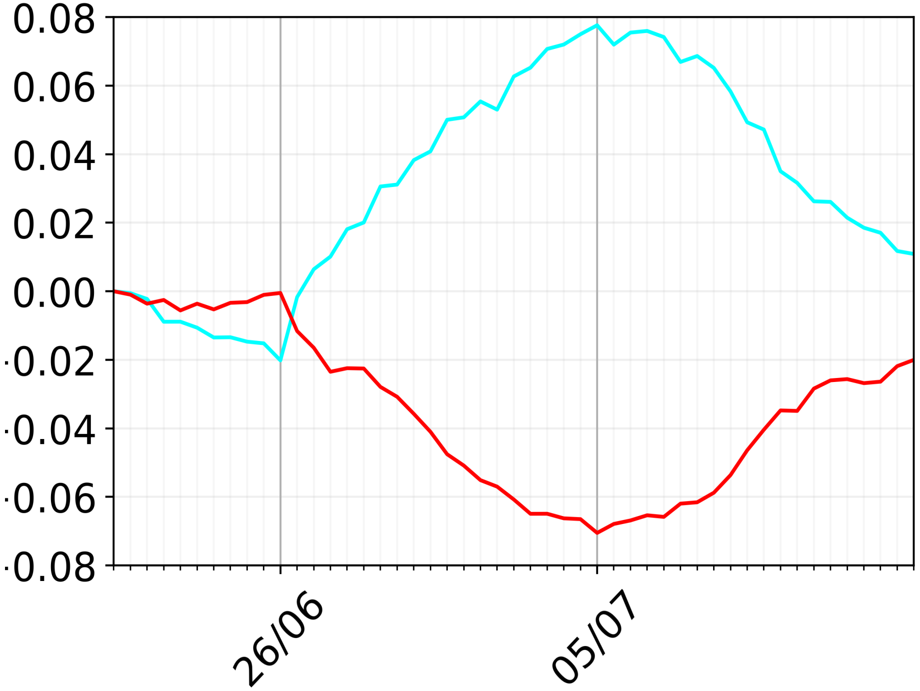
Why temporal modelling? (network)



User representations **before** the announcement of the referendum



User representations **after** the announcement of the referendum



User representations in a sliding window:
sim(YES, YES*)-sim(YES, NO*)
sim(NO, YES*)-sim(NO, NO*)

Summary

Contributions

- Effective/robust MCKL approach (asynchronous, longitudinal)
- Importance of temporal modelling
- Real-time simulation framework

Challenges

- **Time-varying target?**
- **Heterogeneous information?**

Reference: Tsakalidis, Adam, et al. "Nowcasting the Stance of Social Media Users in a Sudden Vote: The Case of the Greek Referendum." *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. ACM, 2018.

Overview

- **Introduction**
- **Part 1: Nowcasting Political Indices Using Social Media**
  - Predicting Election Results with Social Media
  - Nowcasting the Political Stance of Social Media Users
- **Part 2: Nowcasting Mental Health Using Heterogeneous Data**
  - **Using Heterogeneous User Generated Content to Sense Well-being**
  - Challenges in Assessing Mental Health using User Generated Content
- **Conclusion & Future Directions**

Continuous Sensing for Mental Heath Assessment

Mental Health Assessment

- Self-reports (time-consuming) → real-time?

Social Media & Smart Devices

- Real-time sensors: location, text, moving patterns…
- Relation to mental health?

Continuous Sensing for Mental Heath Assessment

Mental Health Assessment

- Self-reports (time-consuming) → real-time?

Social Media & Smart Devices

- Real-time sensors: location, text, moving patterns…
- Relation to mental health?

Longitudinal Task Description:

- Small group of subjects monitored over time
- Features: social media or smart devices
- Target: daily self-reported mental health scores

Continuous Sensing for Mental Heath Assessment

Mental Health Assessment

- Self-reports (time-consuming) → real-time?

Social Media & Smart Devices

- Real-time sensors: location, text, moving patterns…
- Relation to mental health?

Longitudinal Task Description:

- Small group of subjects monitored over time
- Features: social media or smart devices
- Target: daily self-reported mental health scores

Our goal: combine heterogeneous data

Task & Dataset

Task & Dataset

29 users, ~4 months: social media/SMS | DA data   mental health scores

Task & Dataset

29 users, ~4 months: social media/SMS | DA data

mental health scores

Target: daily self-assessed {positive, negative, wellbeing} scores (2,436)

Task & Dataset

29 users, ~4 months: social media/SMS | DA data

mental health scores

Target: daily self-assessed {positive, negative, wellbeing} scores (2,436)

TEXT (>110K)

- ngrams
- lexicons
- word embeddings
- topics
- others (count-based)

DA (>42GB)

- locations
- wifi
- calls
- others (headphones, charging, ...)

RQ3: Experiment & Results

5-fold CV | MKL vs {LASSO, SVR, RF}

		positive	negative	wellbeing
DA	LASSO			
	RF			
	SVR			
	MKL			
TEXT	LASSO			
	RF			
	SVR			
	MKL			
DA+TEXT	LASSO			
	RF			
	SVR			
	MKL			

RQ3: Experiment & Results

5-fold CV | MKL vs {LASSO, SVR, RF}

		positive	negative	wellbeing
DA	LASSO	.31	.11	.30
	RF	.69	.43	.75
	SVR	.58	.35	.62
	MKL	.61	.38	.65
TEXT	LASSO	.53	.23	.55
	RF	.70	.45	.74
	SVR	.60	.32	.62
	MKL	.62	.36	.65
DA+TEXT	LASSO	.49	.18	.54
	RF	.71	.46	.76
	SVR	.60	.34	.62
	MKL	.65	.41	.68

RQ3: Experiment & Results

5-fold CV | MKL vs {LASSO, SVR, RF}

		positive	negative	wellbeing
DA	**LASSO**	.31	.11	.30
	RF	.69	.43	.75
	SVR	.58	.35	.62
	MKL	.61	.38	.65
TEXT	**LASSO**	.53	.23	.55
	RF	.70	.45	.74
	SVR	.60	.32	.62
	MKL	.62	.36	.65
DA+TEXT	**LASSO**	.49	.18	.54
	RF	.71	.46	.76
	SVR	.60	.34	.62
	MKL	.65	.41	.68

RQ3: Experiment & Results

5-fold CV | MKL vs {LASSO, SVR, RF}

		positive	negative	wellbeing
DA	LASSO	.31	.11	.30
	RF	.69	.43	.75
	SVR	.58	.35	.62
	MKL	.61	.38	.65
TEXT	LASSO	.53	.23	.55
	RF	.70	.45	.74
	SVR	.60	.32	.62
	MKL	.62	.36	.65
DA+TEXT	LASSO	.49	.18	.54
	RF	.71	.46	.76
	SVR	.60	.34	.62
	MKL	.65	.41	.68

RQ3: Experiment & Results

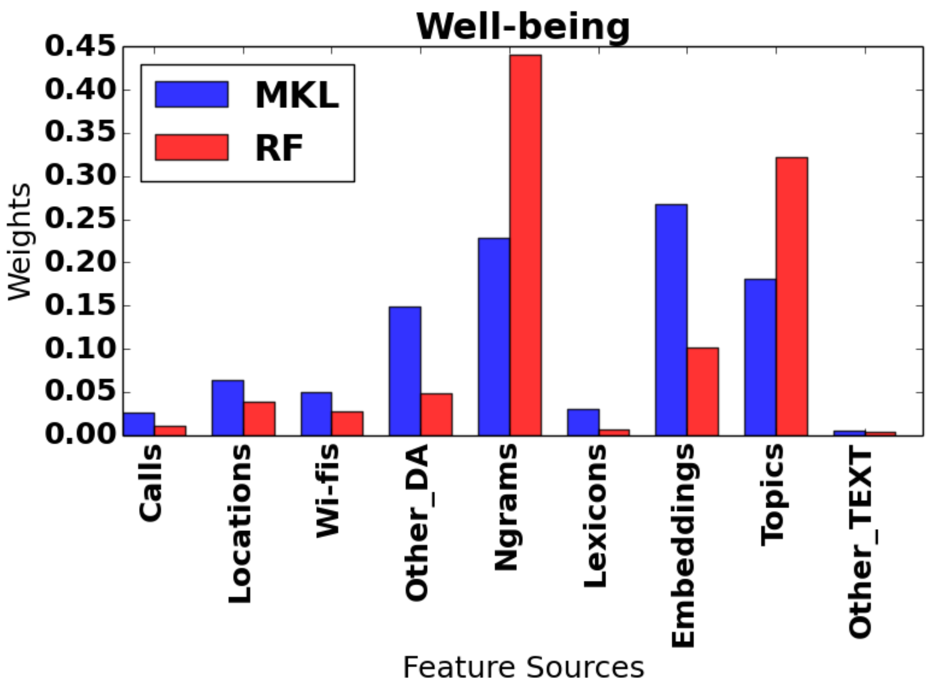
5-fold CV | MKL vs {LASSO, SVR, RF}

		positive	negative	wellbeing
DA	LASSO	.31	.11	.30
	RF	.69	.43	.75
	SVR	.58	.35	.62
	MKL	.61	.38	.65
TEXT	LASSO	.53	.23	.55
	RF	.70	.45	.74
	SVR	.60	.32	.62
	MKL	.62	.36	.65
DA+TEXT	LASSO	.49	.18	.54
	RF	.71	.46	.76
	SVR	.60	.34	.62
	MKL	.65	.41	.68

RQ3: Experiment & Results

5-fold CV | MKL vs {LASSO, SVR, RF}

		positive	negative	wellbeing
DA	LASSO	.31	.11	.30
	RF	.69	.43	.75
	SVR	.58	.35	.62
	MKL	.61	.38	.65
TEXT	LASSO	.53	.23	.55
	RF	.70	.45	.74
	SVR	.60	.32	.62
	MKL	.62	.36	.65
DA+TEXT	LASSO	.49	.18	.54
	RF	.71	.46	.76
	SVR	.60	.34	.62
	MKL	.65	.41	.68



Reference: Tsakalidis, Adam, et al. "Combining heterogeneous user generated data to sense well-being." *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. 2016.

RQ3: Experiment & Results

5-fold CV | MKL vs {LASSO, SVR, RF}

		positive	negative	wellbeing
DA	**LASSO**	.31	.11	.30
	RF	.69	.43	.75
	SVR	.58	.35	.62
	MKL	.61	.38	.65
TEXT	**LASSO**	.53	.23	.55
	RF	.70	.45	.74
	SVR	.60	.32	.62
	MKL	.62	.36	.65
DA+TEXT	**LASSO**	.49	.18	.54
	RF	.71	.46	.76
	SVR	.60	.34	.62
	MKL	.65	.41	.68

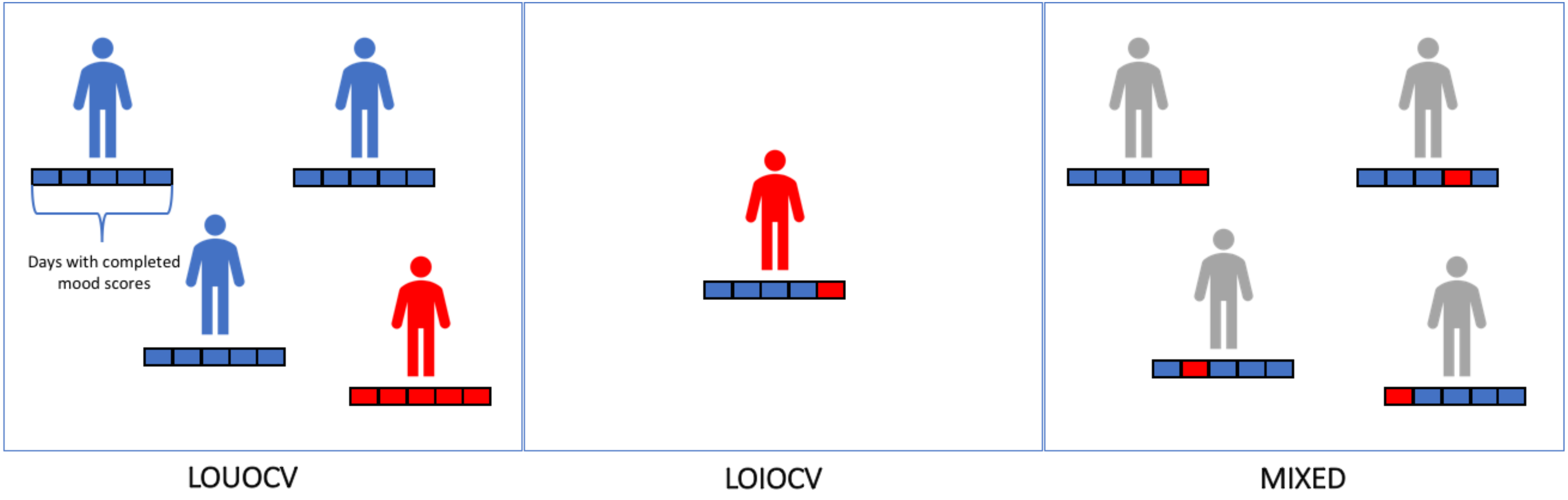
Does this evaluation framework follow a real-world setting?

Reference: Tsakalidis, Adam, et al. "Combining heterogeneous user generated data to sense well-being." *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. 2016.

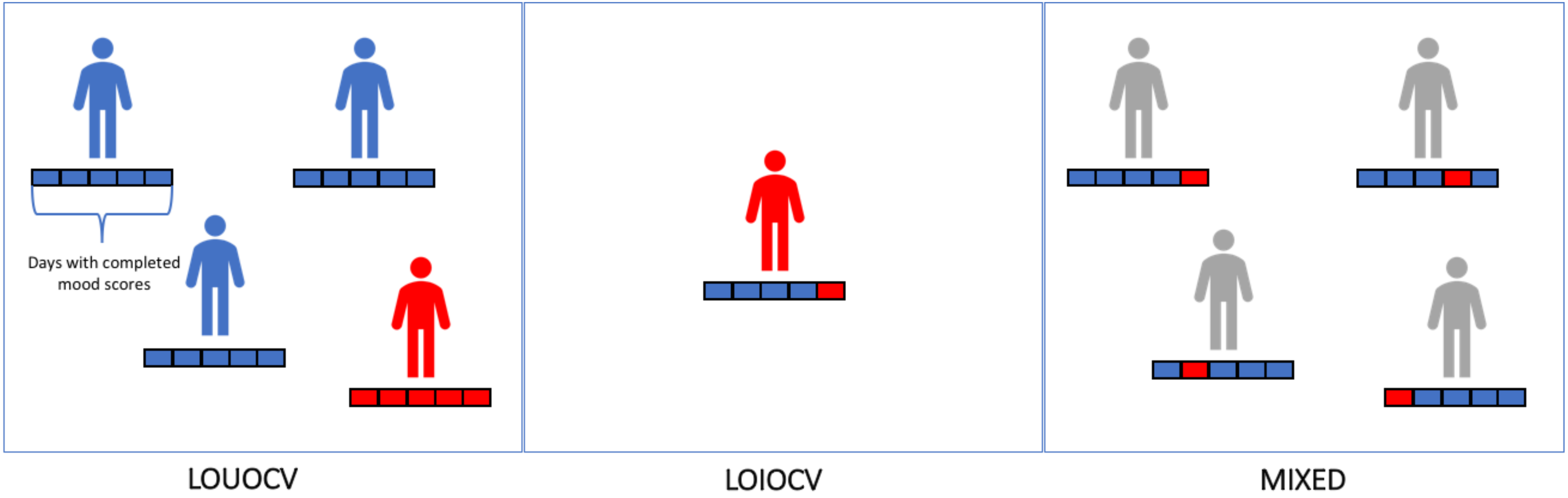
Overview

- Introduction
- Part 1: Nowcasting Political Indices Using Social Media
  - Predicting Election Results with Social Media
  - Nowcasting the Political Stance of Social Media Users
- **Part 2: Nowcasting Mental Health Using Heterogeneous Data**
  - Using Heterogeneous User Generated Content to Sense Well-being
  - **Challenges in Assessing Mental Health using User Generated Content**
- Conclusion & Future Directions

Types of Evaluation



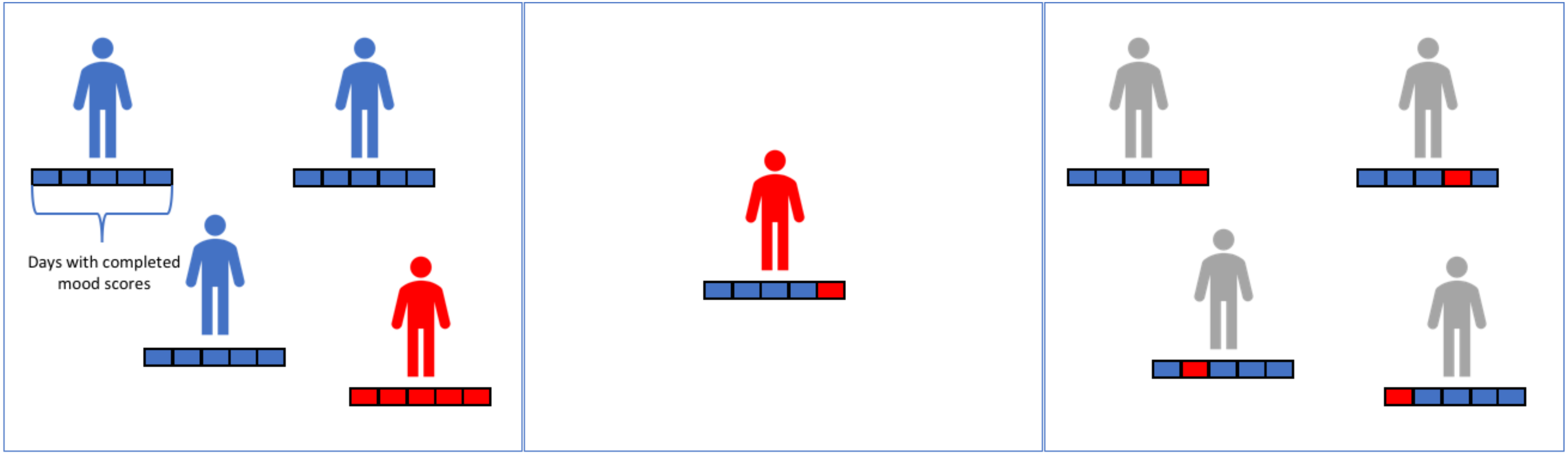
Types of Evaluation



Goal: Generalise to new users

Problem: few users…

Types of Evaluation



LOUOCV LOIOCV MIXED

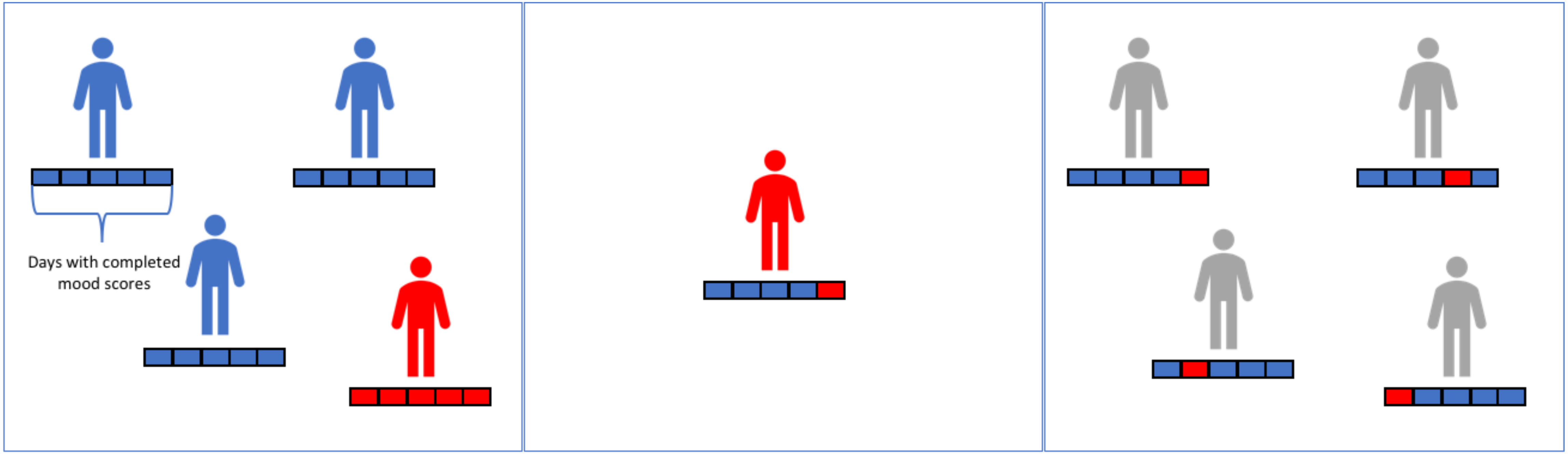
Goal: Generalise to new users

Problem: few users…

Goal: Personalised models

Problem: few instances…

Types of Evaluation



LOUOCV LOIOCV MIXED

Goal: Generalise to new users

Problem: few users…

Goal: Personalised models

Problem: few instances…

Goal: Generalise to certain users only

Problem: identify the user & infer his/her “average” score

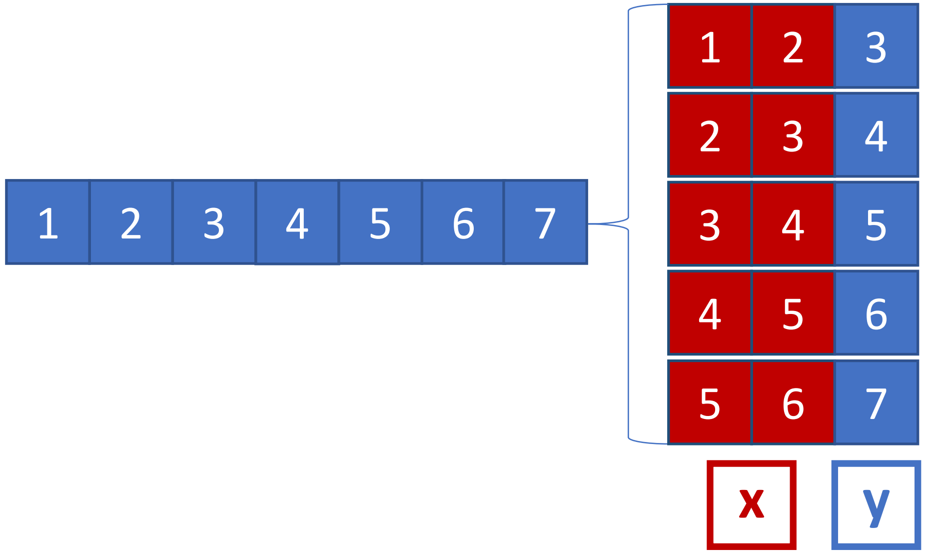
Bias in SOTA

Training on past values of the target

1	2	3	4	5	6	7

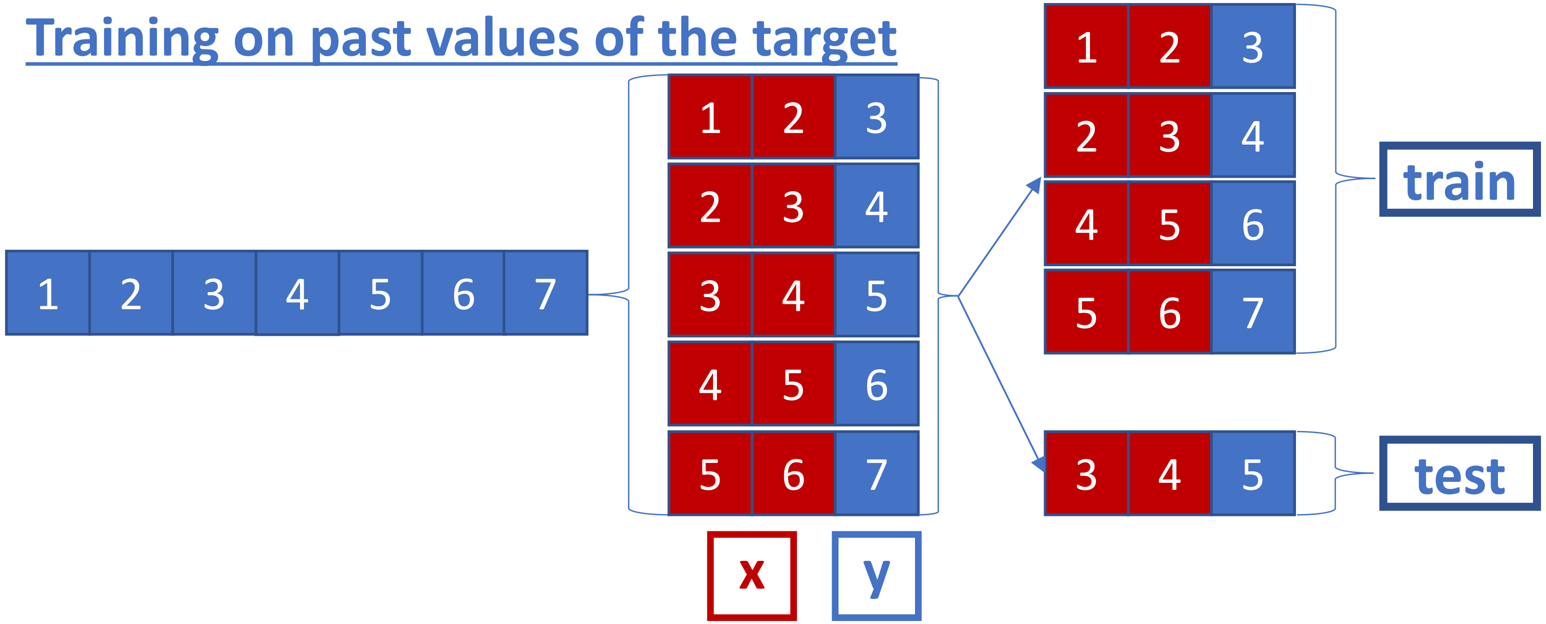
Bias in SOTA

Training on past values of the target



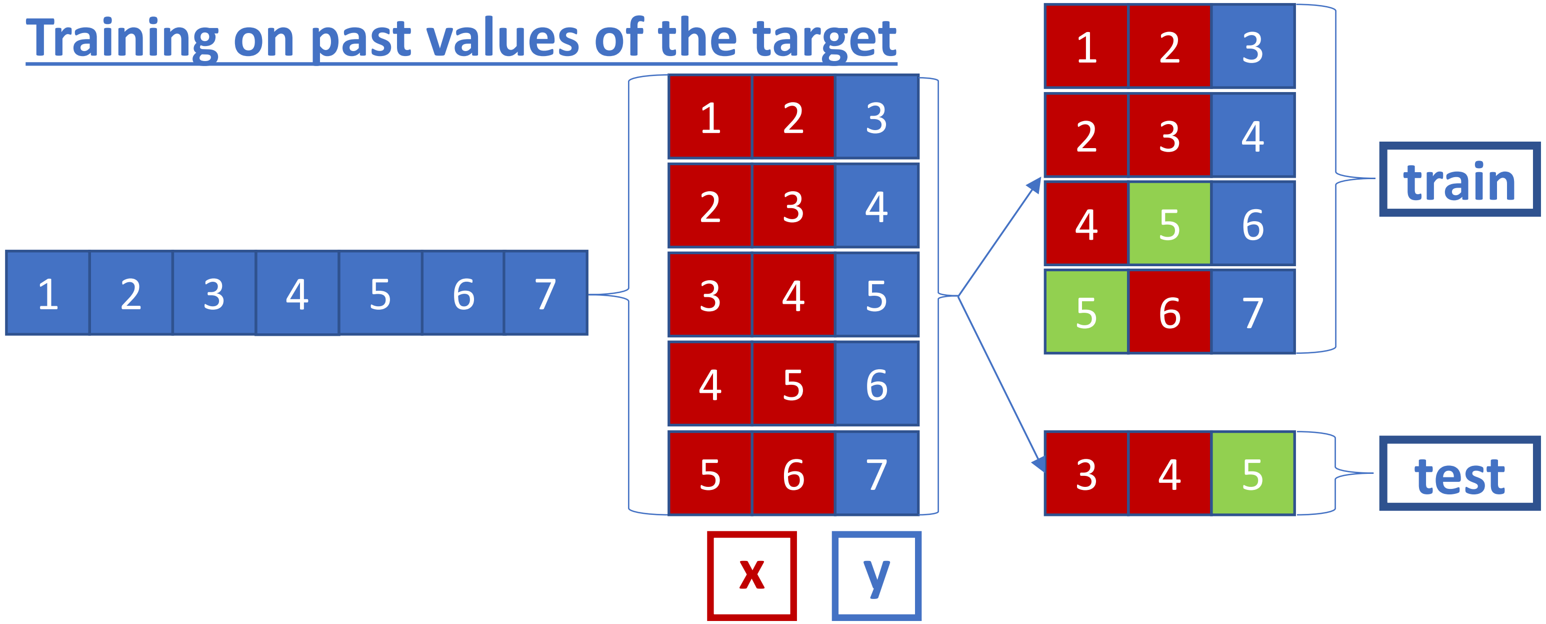
Bias in SOTA

Training on past values of the target



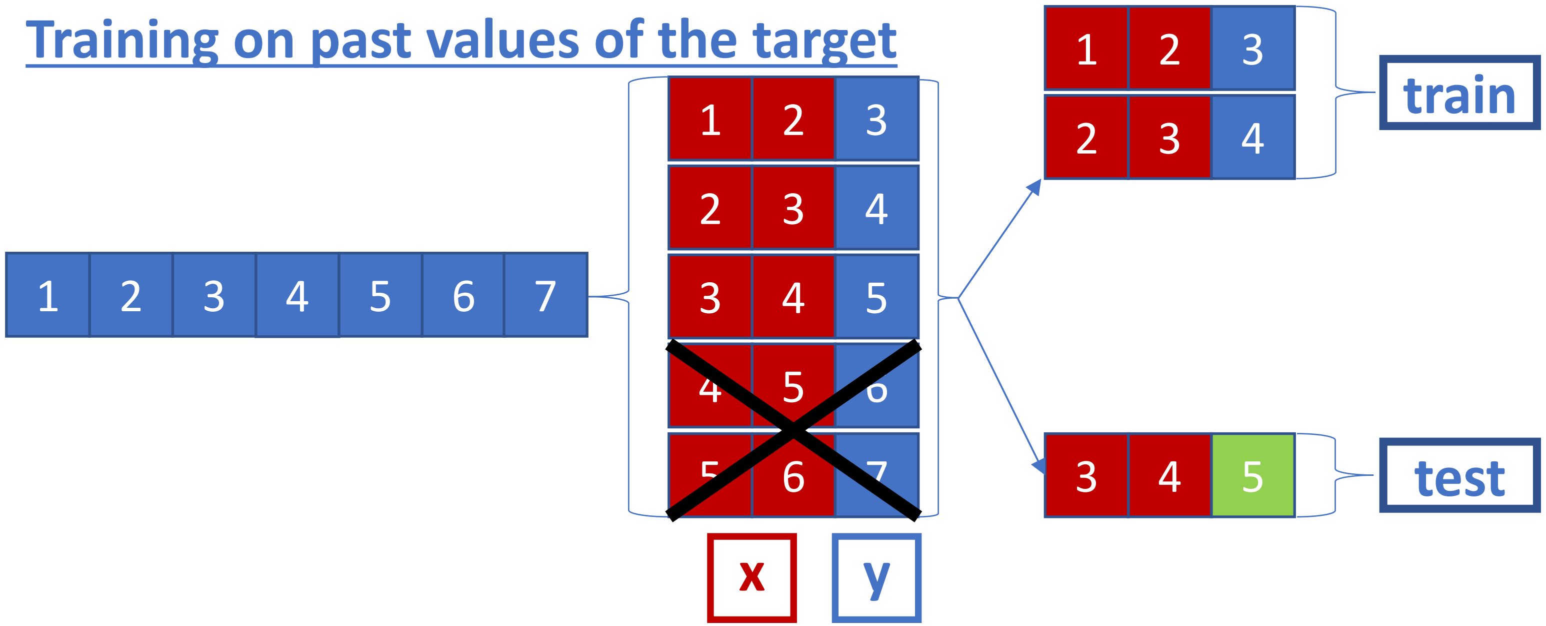
Bias in SOTA

Training on past values of the target



Bias in SOTA

Training on past values of the target



Bias in SOTA

Training on past values of the target

- LOUOCV: new user?

	positive	negative	wellbeing	stress
LOIOCV				
LOUOCV				
AVG				
-mood				

Bias in SOTA

Training on past values of the target

- LOUOCV: new user?

	positive	negative	wellbeing	stress
LOIOCV				
LOUOCV	36.77	31.99	51.08	0.81
AVG				
-mood	113.30	75.27	138.67	1.08

Bias in SOTA

Training on past values of the target

- LOUOCV: new user?

	positive	negative	wellbeing	stress
LOIOCV				
LOUOCV	36.77	31.99	51.08	0.81
AVG	29.89	27.80	41.14	0.70
-mood	113.30	75.27	138.67	1.08

Bias in SOTA

Training on past values of the target

- LOUOCV: new user?
- LOIOCV: test set target: feature in train set!

	positive	negative	wellbeing	stress
LOIOCV	15.96	11.64	20.94	1.07
LOUOCV	36.77	31.99	51.08	0.81
AVG	29.89	27.80	41.14	0.70
-mood	113.30	75.27	138.67	1.08

Bias in SOTA

Training on past values of the target

- LOUOCV: new user?
- LOIOCV: test set target: feature in train set!

	positive	negative	wellbeing	stress
LOIOCV	15.96	11.64	20.94	1.07
LOUOCV	36.77	31.99	51.08	0.81
AVG	29.89	27.80	41.14	0.70
-mood	113.30	75.27	138.67	1.08

Inferring test set labels (LOIOCV)

Bias in SOTA

Training on past values of the target

- LOUOCV: new user?
- LOIOCV: test set target: feature in train set!

	positive	negative	wellbeing	stress
LOIOCV	15.96	11.64	20.94	1.07
LOUOCV	36.77	31.99	51.08	0.81
AVG	29.89	27.80	41.14	0.70
-mood	113.30	75.27	138.67	1.08

Inferring test set labels (LOIOCV)

x	1	2	3	4	5	6	7
y	1	2	3	4	5	6	7

Bias in SOTA

Training on past values of the target

- LOUOCV: new user?
- LOIOCV: test set target: feature in train set!

	positive	negative	wellbeing	stress
LOIOCV	15.96	11.64	20.94	1.07
LOUOCV	36.77	31.99	51.08	0.81
AVG	29.89	27.80	41.14	0.70
-mood	113.30	75.27	138.67	1.08

Inferring test set labels (LOIOCV)

x	1	2	3	4	5	6	7
y	1	2	3	4	5	6	7

Bias in SOTA

Training on past values of the target

- LOUOCV: new user?
- LOIOCV: test set target: feature in train set!

	positive	negative	wellbeing	stress
LOIOCV	15.96	11.64	20.94	1.07
LOUOCV	36.77	31.99	51.08	0.81
AVG	29.89	27.80	41.14	0.70
-mood	113.30	75.27	138.67	1.08

Inferring test set labels (LOIOCV)

x	1	2	3	4	5	6	7
y	1	2	3	4	5	6	7

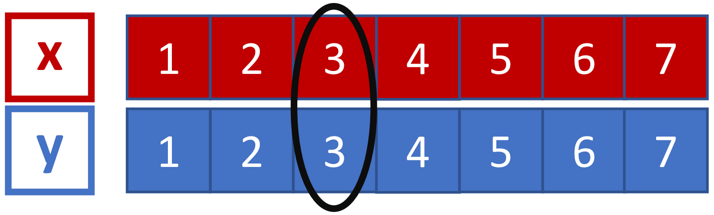
Bias in SOTA

Training on past values of the target

- LOUOCV: new user?
- LOIOCV: test set target: feature in train set!

	positive	negative	wellbeing	stress
LOIOCV	15.96	11.64	20.94	1.07
LOUOCV	36.77	31.99	51.08	0.81
AVG	29.89	27.80	41.14	0.70
-mood	113.30	75.27	138.67	1.08

Inferring test set labels (LOIOCV)



Bias in SOTA

Training on past values of the target

- LOUOCV: new user?
- LOIOCV: test set target: feature in train set!

	positive	negative	wellbeing	stress
LOIOCV	15.96	11.64	20.94	1.07
LOUOCV	36.77	31.99	51.08	0.81
AVG	29.89	27.80	41.14	0.70
-mood	113.30	75.27	138.67	1.08

Inferring test set labels (LOIOCV)

x	1	2	3	4	5	6	7
y	1	2	3	4	5	6	7

Bias in SOTA

Training on past values of the target

- LOUOCV: new user?
- LOIOCV: test set target: feature in train set!

	positive	negative	wellbeing	stress
LOIOCV	15.96	11.64	20.94	1.07
LOUOCV	36.77	31.99	51.08	0.81
AVG	29.89	27.80	41.14	0.70
-mood	113.30	75.27	138.67	1.08

Inferring test set labels (LOIOCV)

x	1	2	3	4	5	6	7
y	1	2	3	4	5	6	7

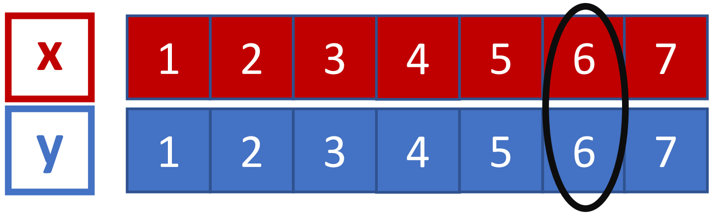
Bias in SOTA

Training on past values of the target

- LOUOCV: new user?
- LOIOCV: test set target: feature in train set!

	positive	negative	wellbeing	stress
LOIOCV	15.96	11.64	20.94	1.07
LOUOCV	36.77	31.99	51.08	0.81
AVG	29.89	27.80	41.14	0.70
-mood	113.30	75.27	138.67	1.08

Inferring test set labels (LOIOCV)



Bias in SOTA

Training on past values of the target

- LOUOCV: new user?
- LOIOCV: test set target: feature in train set!

	positive	negative	wellbeing	stress
LOIOCV	15.96	11.64	20.94	1.07
LOUOCV	36.77	31.99	51.08	0.81
AVG	29.89	27.80	41.14	0.70
-mood	113.30	75.27	138.67	1.08

Inferring test set labels (LOIOCV)

x	1	2	3	4	5	6	7
y	1	2	3	4	5	6	7

Bias in SOTA

Training on past values of the target

- LOUOCV: new user?
- LOIOCV: test set target: feature in train set!

	positive	negative	wellbeing	stress
LOIOCV	15.96	11.64	20.94	1.07
LOUOCV	36.77	31.99	51.08	0.81
AVG	29.89	27.80	41.14	0.70
-mood	113.30	75.27	138.67	1.08

Inferring test set labels (LOIOCV)

x	1	2	3	4	5	6	7
y	1	2	3	4	5	6	7

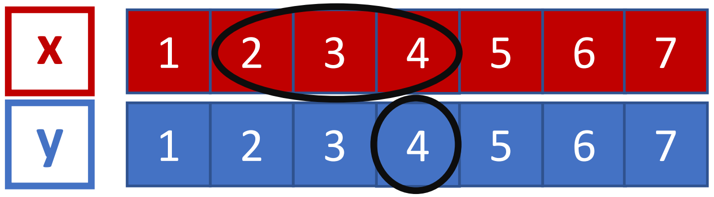
Bias in SOTA

Training on past values of the target

- LOUOCV: new user?
- LOIOCV: test set target: feature in train set!

	positive	negative	wellbeing	stress
LOIOCV	15.96	11.64	20.94	1.07
LOUOCV	36.77	31.99	51.08	0.81
AVG	29.89	27.80	41.14	0.70
-mood	113.30	75.27	138.67	1.08

Inferring test set labels (LOIOCV)



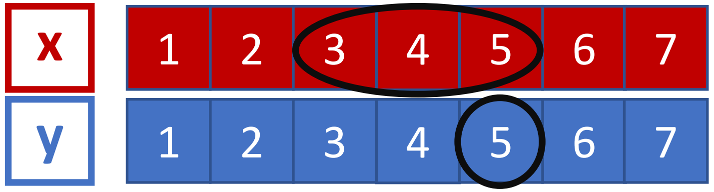
Bias in SOTA

Training on past values of the target

- LOUOCV: new user?
- LOIOCV: test set target: feature in train set!

	positive	negative	wellbeing	stress
LOIOCV	15.96	11.64	20.94	1.07
LOUOCV	36.77	31.99	51.08	0.81
AVG	29.89	27.80	41.14	0.70
-mood	113.30	75.27	138.67	1.08

Inferring test set labels (LOIOCV)



Bias in SOTA

Training on past values of the target

- LOUOCV: new user?
- LOIOCV: test set target: feature in train set!

	positive	negative	wellbeing	stress
LOIOCV	15.96	11.64	20.94	1.07
LOUOCV	36.77	31.99	51.08	0.81
AVG	29.89	27.80	41.14	0.70
-mood	113.30	75.27	138.67	1.08

Inferring test set labels (LOIOCV)

x	1	2	3	4	5	6	7
y	1	2	3	4	5	6	7

Bias in SOTA

Training on past values of the target

- LOUOCV: new user?
- LOIOCV: test set target: feature in train set!

	positive	negative	wellbeing	stress
LOIOCV	15.96	11.64	20.94	1.07
LOUOCV	36.77	31.99	51.08	0.81
AVG	29.89	27.80	41.14	0.70
-mood	113.30	75.27	138.67	1.08

Inferring test set labels (LOIOCV)

x	1	2	3	4	5	6	7
y	1	2	3	4	5	6	7

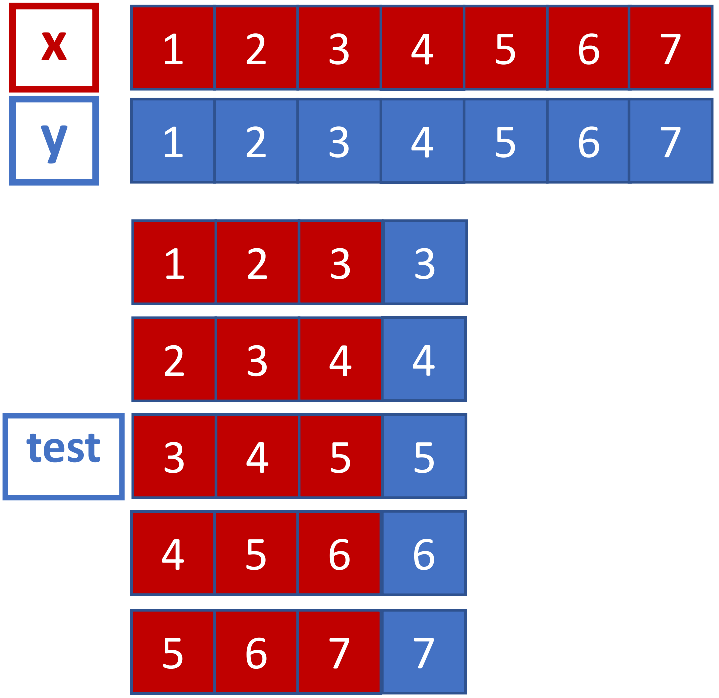
Bias in SOTA

Training on past values of the target

- LOUOCV: new user?
- LOIOCV: test set target: feature in train set!

	positive	negative	wellbeing	stress
LOIOCV	15.96	11.64	20.94	1.07
LOUOCV	36.77	31.99	51.08	0.81
AVG	29.89	27.80	41.14	0.70
-mood	113.30	75.27	138.67	1.08

Inferring test set labels (LOIOCV)



Bias in SOTA

Training on past values of the target

- LOUOCV: new user?
- LOIOCV: test set target: feature in train set!

	positive	negative	wellbeing	stress
LOIOCV	15.96	11.64	20.94	1.07
LOUOCV	36.77	31.99	51.08	0.81
AVG	29.89	27.80	41.14	0.70
-mood	113.30	75.27	138.67	1.08

Inferring test set labels (LOIOCV)

x	1	2	3	4	5	6	7
y	1	2	3	4	5	6	7

	x	x	x	y
	1	2	3	3
	2	3	4	4
test	3	4	5	5
	4	5	6	6
	5	6	7	7

67% of the input is the same!

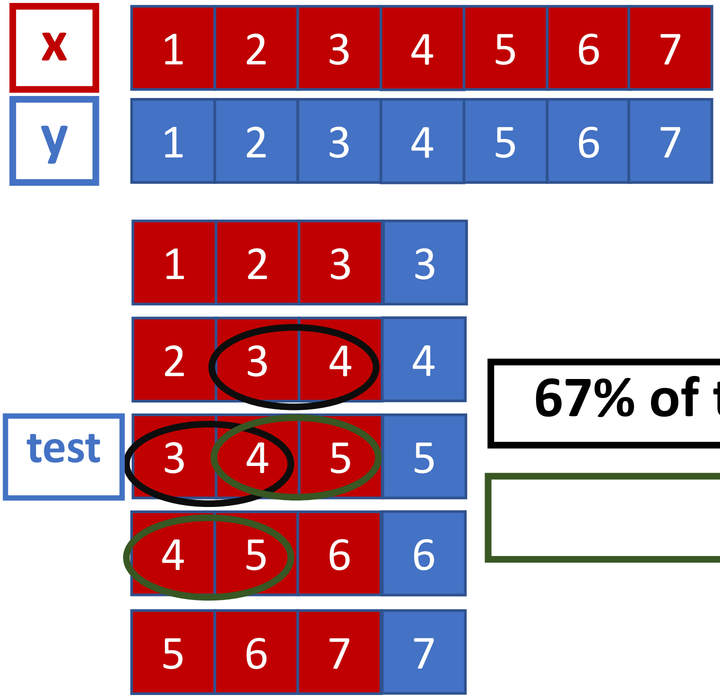
Bias in SOTA

Training on past values of the target

- LOUOCV: new user?
- LOIOCV: test set target: feature in train set!

	positive	negative	wellbeing	stress
LOIOCV	15.96	11.64	20.94	1.07
LOUOCV	36.77	31.99	51.08	0.81
AVG	29.89	27.80	41.14	0.70
-mood	113.30	75.27	138.67	1.08

Inferring test set labels (LOIOCV)



67% of the input is the same!

Again!

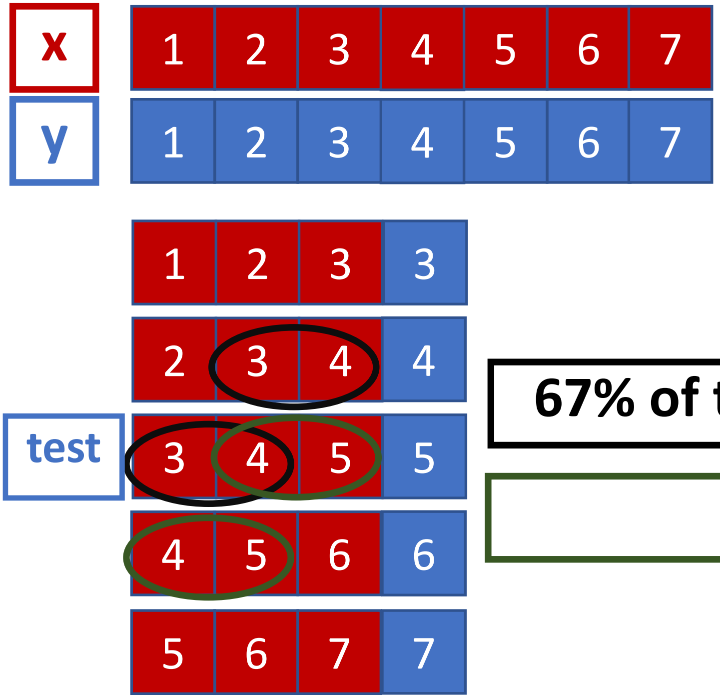
Bias in SOTA

Training on past values of the target

- LOUOCV: new user?
- LOIOCV: test set target: feature in train set!

	positive	negative	wellbeing	stress
LOIOCV	15.96	11.64	20.94	1.07
LOUOCV	36.77	31.99	51.08	0.81
AVG	29.89	27.80	41.14	0.70
-mood	113.30	75.27	138.67	1.08

Inferring test set labels (LOIOCV)



67% of the input is the same!

Again!

Q: What if our target is also temporally correlated?

Bias in SOTA

Training on past values of the target

- LOUOCV: new user?
- LOIOCV: test set target: feature in train set!

	positive	negative	wellbeing	stress
LOIOCV	15.96	11.64	20.94	1.07
LOUOCV	36.77	31.99	51.08	0.81
AVG	29.89	27.80	41.14	0.70
-mood	113.30	75.27	138.67	1.08

Inferring test set labels (LOIOCV)

x	1	2	3	4	5	6	7
y	1	2	3	4	5	6	7

	1	2	3	3
	2	3	4	4
test	3	4	5	5
	4	5	6	6
	5	6	7	7

67% of the input is the same!

Again!

Q: What if our target is also temporally correlated?

A: Artificially easy task – unrelated to mental health!

Bias in SOTA

Training on past values of the target

- LOUOCV: new user?
- LOIOCV: test set target: feature in train set!

	positive	negative	wellbeing	stress
LOIOCV	15.96	11.64	20.94	1.07
LOUOCV	36.77	31.99	51.08	0.81
AVG	29.89	27.80	41.14	0.70
-mood	113.30	75.27	138.67	1.08

Inferring test set labels (LOIOCV)

- Overlapping instances (input)
- What if the target is also correlated?

	positive	negative	wellbeing	stress
MODEL	64.02	60.03	65.06	45.86
Random				

Bias in SOTA

Training on past values of the target

- LOUOCV: new user?
- LOIOCV: test set target: feature in train set!

	positive	negative	wellbeing	stress
LOIOCV	15.96	11.64	20.94	1.07
LOUOCV	36.77	31.99	51.08	0.81
AVG	29.89	27.80	41.14	0.70
-mood	113.30	75.27	138.67	1.08

Inferring test set labels (LOIOCV)

- Overlapping instances (input)
- What if the target is also correlated?

	positive	negative	wellbeing	stress
MODEL	64.02	60.03	65.06	45.86
Random	64.22	60.88	64.87	45.79

Bias in SOTA

Training on past values of the target

- LOUOCV: new user?
- LOIOCV: test set target: feature in train set!

	positive	negative	wellbeing	stress
LOIOCV	15.96	11.64	20.94	1.07
LOUOCV	36.77	31.99	51.08	0.81
AVG	29.89	27.80	41.14	0.70
-mood	113.30	75.27	138.67	1.08

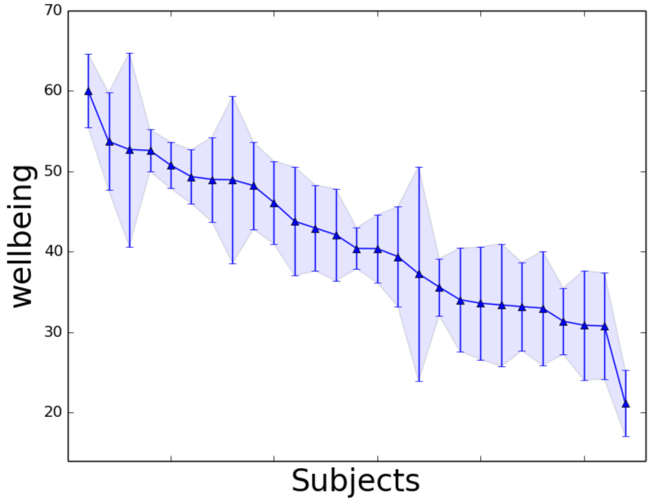
Inferring test set labels (LOIOCV)

- Overlapping instances (input)
- What if the target is also correlated?

	positive	negative	wellbeing	stress
MODEL	64.02	60.03	65.06	45.86
Random	64.22	60.88	64.87	45.79

Predicting users instead of mood scores (MIXED)

- identify the user?



Bias in SOTA

Training on past values of the target

- LOUOCV: new user?
- LOIOCV: test set target: feature in train set!

	positive	negative	wellbeing	stress
LOIOCV	15.96	11.64	20.94	1.07
LOUOCV	36.77	31.99	51.08	0.81
AVG	29.89	27.80	41.14	0.70
-mood	113.30	75.27	138.67	1.08

Inferring test set labels (LOIOCV)

- Overlapping instances (input)
- What if the target is also correlated?

	positive	negative	wellbeing	stress
MODEL	64.02	60.03	65.06	45.86
Random	64.22	60.88	64.87	45.79

Predicting users instead of mood scores (MIXED)

- identify the user?

	positive	negative	wellbeing	stress
MIXED	0.43	0.25	0.48	0.02
LOIOCV				
LOUOCV				

Bias in SOTA

Training on past values of the target

- LOUOCV: new user?
- LOIOCV: test set target: feature in train set!

	positive	negative	wellbeing	stress
LOIOCV	15.96	11.64	20.94	1.07
LOUOCV	36.77	31.99	51.08	0.81
AVG	29.89	27.80	41.14	0.70
-mood	113.30	75.27	138.67	1.08

Inferring test set labels (LOIOCV)

- Overlapping instances (input)
- What if the target is also correlated?

	positive	negative	wellbeing	stress
MODEL	64.02	60.03	65.06	45.86
Random	64.22	60.88	64.87	45.79

Predicting users instead of mood scores (MIXED)

- identify the user?

	positive	negative	wellbeing	stress
MIXED	0.43	0.25	0.48	0.02
LOIOCV	-0.03	-0.04	-0.03	-0.08
LOUOCV	-4.19	-1.09	-4.66	-0.67

Reference: Tsakalidis, Adam, et al. "Can We Assess Mental Health through Social Media and Smart Devices? Addressing Bias in Methodology and Evaluation." (to appear in *ECML PKDD '18/ADS*)

Future Directions

Can we assess mental health through UGC…?

Need to re-evaluate mental health models!

- Extremely important for practitioners
- Conclusions reached might be wrong!

Future Directions

Can we assess mental health through UGC…?

Need to re-evaluate mental health models!

- Extremely important for practitioners
- Conclusions reached might be wrong!

Proposal for future directions:

Larger datasets, transfer learning, user data…

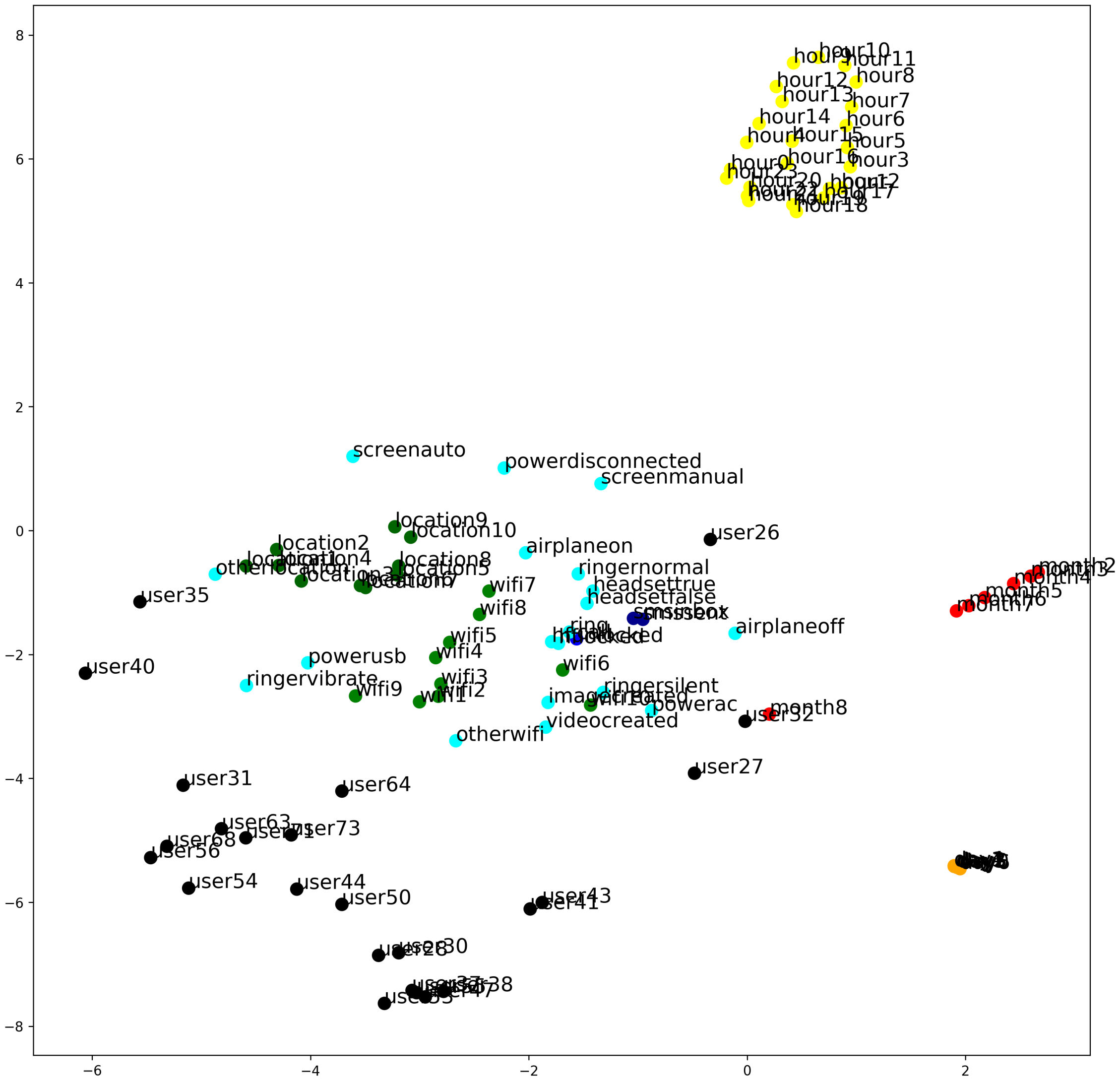
Different users, different behaviour

Future Directions

Can we assess mental health through UGC…?

Need to re-evaluate mental health models!

- Extremely important for practitioners
- Conclusions reached might be wrong!



Proposal for future directions:

Larger datasets, transfer learning, user data…

Different users, different behaviour

Latent feature representations:

$$J = \log p_w(D=1|f,h) + \sum_{\tilde{f}\sim Q_{\text{noise}}} \left[\log p_w(D=0|\tilde{f},h)\right]$$

Overview

- **Introduction**
- **Part 1: Nowcasting User Political Stance Using Social Media**
  - Predicting Election Results with Social Media
  - Nowcasting the Political Stance of Social Media Users
- **Part 2: Nowcasting Mental Health Using Heterogeneous Data**
  - Using Heterogeneous User Generated Content to Sense Well-being
  - Challenges in Assessing Mental Health using User Generated Content
- **Conclusion & Future Directions**

Conclusion & Future Directions

Social Media & Smart Devices

- Rich resources of user-specific information
- Real-time sensors of political/health indices?
- Ethical implications?

High importance:

- work under a real-world setting

Future Work

- Heterogeneous data representation
- Asynchronous modelling
- Multi-domain applications

Published Material and Coverage

[1] Tsakalidis, Adam, et al. "Predicting elections for multiple countries using Twitter and polls." *IEEE Intelligent Systems* 30.2 (2015): 10-17.

- Press Release (University of Warwick): *Can we predict our political future?*
- Financial Times (Chris Tighe): *Tweets analysed for clues to UK general election result*
- Motherboard (Victoria Turk): *Attempting to Predict an Unpredictable Election*

[2] Tsakalidis, Adam, et al. "Nowcasting the Stance of Social Media Users in a Sudden Vote: The Case of the Greek Referendum." *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. ACM, 2018.

- Efimerida ton Syntakton (Kostas Zafiropoulos): Twitter threatens Opinion Polls (Greek)

[3] Tsakalidis, Adam, et al. "Combining heterogeneous user generated data to sense well-being." *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. 2016.

[4] Tsakalidis, Adam, et al. "Can We Assess Mental Health through Social Media and Smart Devices? Addressing Bias in Methodology and Evaluation." *arXiv preprint arXiv:1807.07351*(2018). (to appear in ECML PKDD 2018/ADS)

Acknowledgements

Supported by the EPSRC through the University of Warwick's CDT in Urban Science and Progress (EP/L016400/1) and through The Alan Turing Institute (EP/N510129/1).

EPSRC
Pioneering research and skills

Thank you!

Any questions?

@adtsakal

atsakalidis@turing.ac.uk