National Research Council Canada
Conseil national de recherches Canada

EuroGames16: Evaluating Change Detection in Online Conversation

Cyril Goutte, Yunli Wang, Fangming Liao, Zachary Zanussi, Samuel Larkin, Yuri Grinberg

Digital Technologies, National Research Council Canada

Change Detection in Online Conversation

Online conversation:

- Stream of docs/messages
- e.g. Follow twitter hashtag

Changes impact conversation

- Volume
- Sentiment
- Topic

We want to detect changes

- when they happen [19:31:31],
- as soon as they happen.

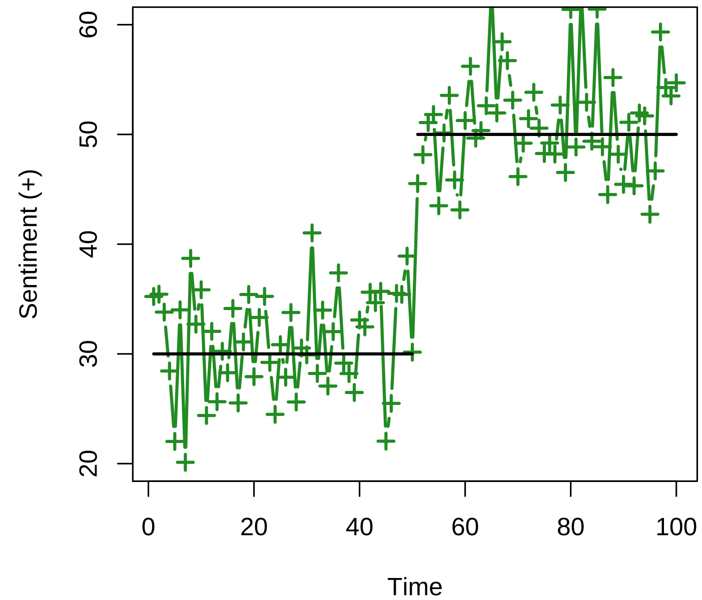
19:30:37	How come no-one is mentioning the blatant foul in the buil
19:30:45	I like #Nainggolan's hair style #WALBEL #Euro2016
19:30:52	That is a top draw save by Courtois. Wales building pressu
19:30:59	#WALBEL @BBCMOTD How many dodgy referee decision
19:31:05	Jordan Lukaku is absolutely dreadful. #WALBEL
19:31:11	Come on #WAL! You can do this. #EURO2016 #WALBEL
19:31:16	How was Kanu not offside? #WALBEL #EURO2016
19:31:21	Tap and dive for Begium #WALBEL
19:31:24	Good game so far, a #wal equaliser is coming #WALBEL
19:31:27	And can we talk about the #referee ? #WALBEL #BELWAL
19:31:31	GOAL!!!!!! Wales 1-1Belgium #WAL #WALBEL
19:31:34	Somehow Belgium players are falling like leaves...get up a
19:31:37	Oh dear Wales who is laughing now #WALBEL
19:31:39	YESSSS WILLIAMS!!!! #WALBEL
19:31:40	RT @AJoToole: Ashley Williams for the equaliser. Come o
19:31:41	GOOOOOOOOOOOOOOOOOOOOOL WILLIAAAAAAAA
19:31:44	Wales 1-1 Belgium (Williams '30) #EURO2016 #WALBEL Fo
19:31:46	Goooooooooooooooooooooal #WALBEL
19:31:48	Wales have equalised!!! 1-1 #WALBEL
19:31:50	ASSSHHHLLLEEEYYY WILLLLIIIAAAMMMSSSS!! #WAL #\

Index

Change Point Detection

Known problem in time series analysis.

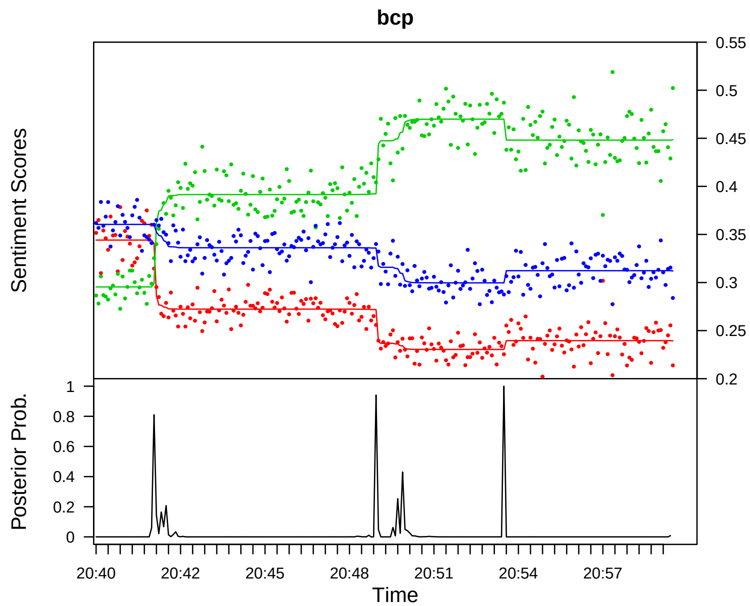
- Assuming time series of sentiment scores:
  - E.g. $s^+(t)$ positive sentiment.
- A Change Point is a location where the underlying stochastic process changes,
  - E.g. go from moderately happy ($s^+ \sim 30\%$) to quite happy ($s^+ \sim 55\%$).



- Various strategies:
  - Batch vs. Online (all data vs. one point at a time)
  - Parametric vs. Non-parametric (e.g. Gaussian vs. no assumption)
  - Univariate vs. Multivariate (e.g. count vs. 3 sent. scores)

Baseline Batch CPD Algorithms

bcp: Bayesian CP detection

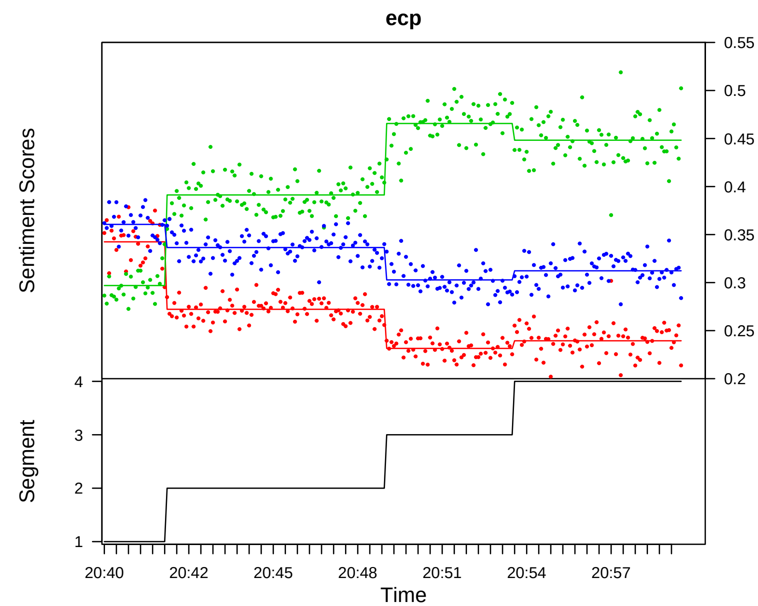


Infering posterior probability of change

- Batch
- Gaussian assumption
- Multivariate

[1] Barry&Hartigan (1993) J.Amer.Stat.Assoc.

ecp: Nonparametric CP analysis



Recursive test of differing distributions

- Batch
- Non parametric (no assumption)
- Multivariate

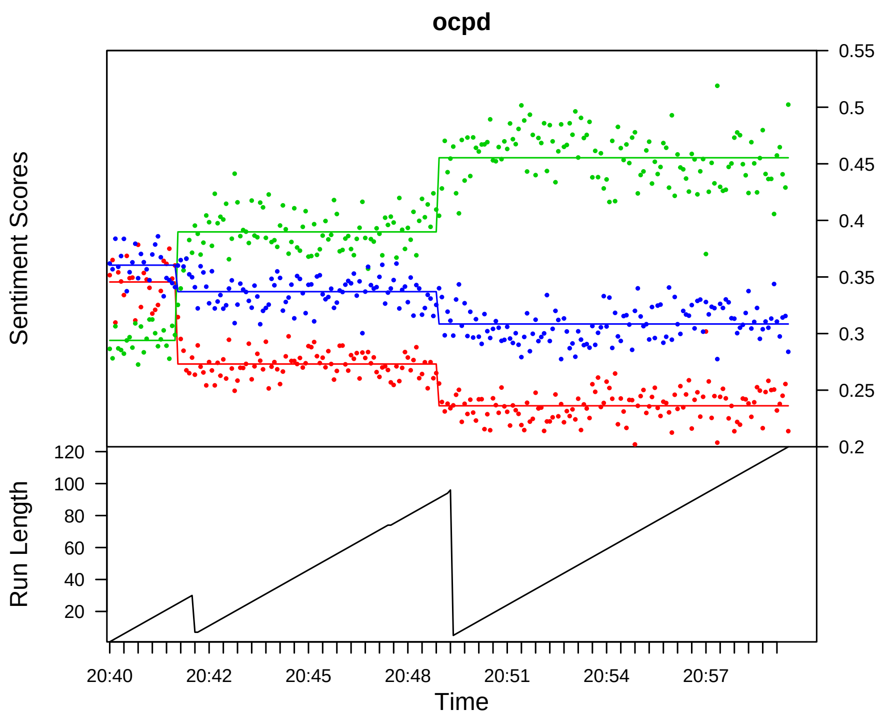
[2] James&Matteson (2014) J. Stat. Soft.

Online Change Point Detection

Model run length, using underlying predictive model (UPM).

E.g.:**ocpd**: Gaussian UPM **ocpd+**: Linear trend + Gaussian noise UPM .

Bayesian inference ⇒ run length distribution.



- Online (one point at a time)
- Gaussian (**ocpd**) or linear (**ocpd+**)
- Multivariate
- `R` package `onlineCPD`

[3] Adams&Mackay (2007) arXiv:0710.3742

Run length drops → change point!

Index

From Document Streams to Time Series

⋮

$\mathbf{c}_i =$`What the fuck is this ref man #WALBEL`, $[t_i =$19:30:53$]$

⋮

$\mathbf{c}_j =$`Best game of the tournament so far.  #EURO2016`, $[t_j =$19:33:00$]$

⋮

- Compute sentiment (polarity) score:[1]
  - $s_i^+ = .130$, $s_i^0 = .190$, $s_i^- = .681$
  - $s_j^+ = .681$, $s_j^0 = .197$, $s_j^- = .123$
- Average message sentiment over fixed bins (e.g. 5 seconds)
  - $\mathrm{s}[\text{19:30:53}] = (.318, .363, .318)$, $n = 25$
  - $\mathrm{s}[\text{19:32:58}] = (.411, .348, .241)$, $n = 139$
- Three time series of positive/neutral/negative sentiment
- Same framework may yield time series of counts, topics, etc.

[1] Kiritchenko, Zhu, Mohammad (2014) Sentiment analysis of short informal texts. *JAIR* 50:723-762.

Index

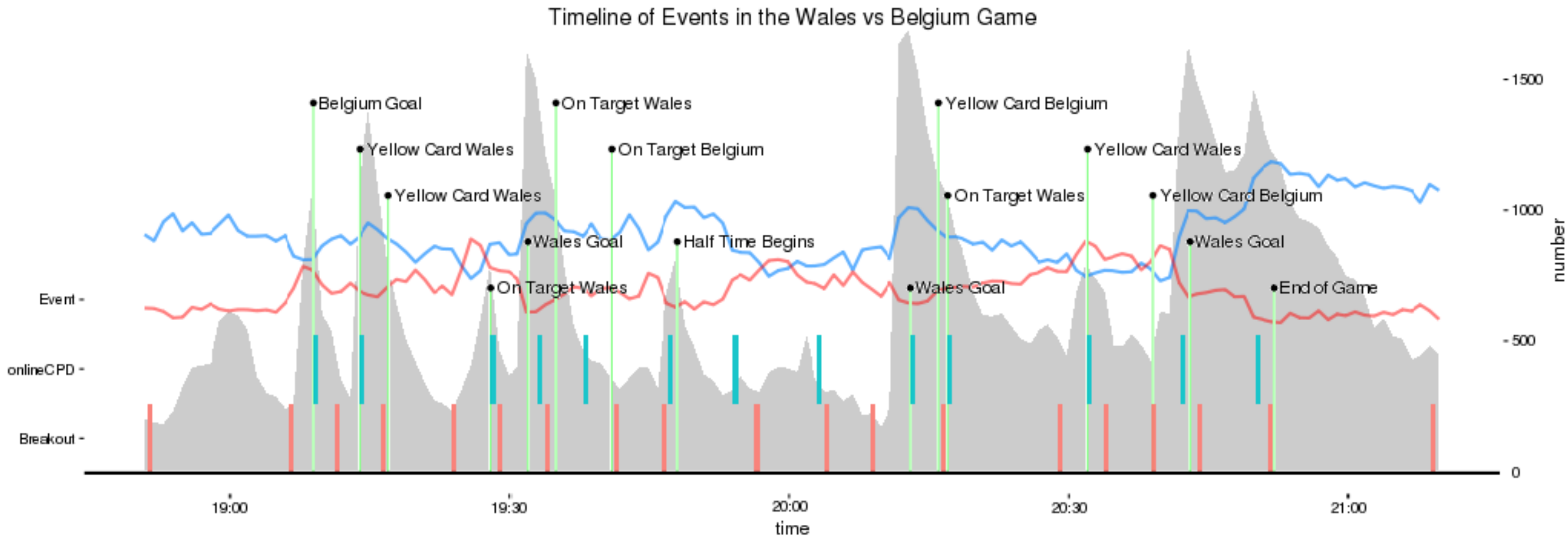
Data Collection: EuroGames16

- 16 games from 2016 UEFA Euro Championship (football/soccer);
- Collected from twitter API during each game +/- a few hours;
- Querying game-specific hashtags (e.g. #FRAPOR #PORFRA);
- Filtering for English + game-specific languages;
- Gold standard events from game reports: goals, on-target, substitutions. . .
- Processed data + gold refs available (find github link in paper).

Game	Hashtag	# Eng.	#Total
-	CROESP	52,953	115k
-	ENGISL	191,384	210k
-	FRAALB	61,748	434k
-	FRAIRL	172,872	665k
-	FRAISL	158,457	721k
-	GERFRA	273,074	496k
-	GERITA	426,381	709k
-	GERPOL	82,132	232k
-	POLPOR	128,079	664k
-	PORAUT	72,644	171k
-	FRAPOR	229,000	1000k
-	PORWAL	287,417	461k
-	RUSWAL	110,165	142k
-	SUIFRA	36,507	468k
-	WALBEL	288,312	379k
-	WALNIR	95,679	115k
Total tweets	-	2.69M	7.04M

Index

Wales-Belgium Timeline (July 1st Q.final)



volume: number of tweets per 60 seconds,
pos/neg: sentiment score (NRC model),
reference: game events, collected from sports site,
breakout: 'breakout' detection (R package)
onlineCPD: online change-point detection (our work)

Experiments: Overall Scores

- Run 4 CPD algorithms on 16 games/datasets from collection.
  - on raw counts (univariate), sentiment scores (multiv.), and both.
- Compute performance as F-score w.r.t. reference game events.
  - Detected changes within +/- 2 bins of events are correct

Input	**bcp**	**ecp**	**ocpd**	**ocpd+**
Count	.3434	.4645	.4250	**.4725**
Sentiment	.2354	.4315	.3047	**.4428**
+ # references	.3918	**.4860**	.3047	.4380
Count+Sentiment	.4251	.4645	.4010	**.5062**

- **ocpd+** does better overall; **ecp** performs almost as well;
- **bcp**/**ecp** work much better when they know the # references (cheating!).

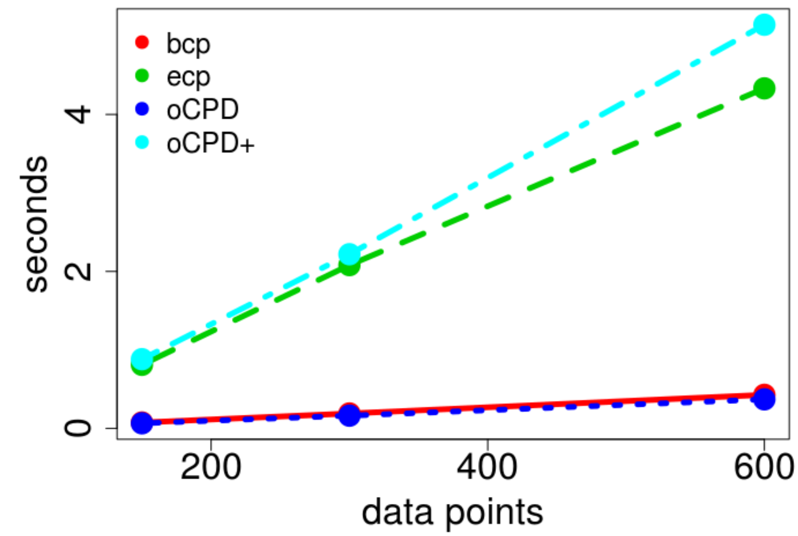
Experiments: Pairwise Comparisons

- Compute wins/ties/losses between pairs of algorithms
  - First number higher when row method wins;
  - Last number higher when column method wins.
- **ocpd+** and **ecp** clearly outperform other two methods;
- Providing correct number of ref. events helps **ecp** a lot (cheating!).

Counts	**ecp**	**ocpd**	**ocpd+**
bcp	2/0/14	5/0/11	2/0/14
ecp	-	11/1/4	**7/1/8**
ocpd	-	-	4/1/11
Sentiment	**ecp**	**ocpd**	**ocpd+**
bcp	1/0/15	5/1/10	0/0/16
ecp	-	12/0/4	**7/1/8**
ocpd	-	-	3/0/13
+ # refs	**ecp**	**ocpd**	**ocpd+**
bcp	0/0/16	4/0/12	0/0/16
ecp	-	14/0/2	**10/0/6**
ocpd	-	-	3/0/13

More Experiments

- Individual game × method results → see paper.
- Can use CPD to detect change in keyword/hashtag usage:
  - More labour intensive (need to specify k/w, #tag);
  - Does not work as well as tracking sentiment → see paper.
- Bin size impacts performance (see paper):
  - Large bin ⇒ short, smooth time series;
  - Small bin ⇒ long, noisy time series;
  - **bcp** better w. large, others with short bins.
- Computational time varies with CPD algo:



Index

Conclusion

- We develop a new framework for detecting change in social media;
  - Change in traffic, Sentiment, topic, etc.
- Works *online*, not retrospective;
  - Can detect changes as they occur (no waiting / re-running).
- Able to detect changes related to $\sim$ half reference events;

Extensions:

- Detect topic change using (online) topic models (tba);
- Re-architecture and extend R package for **ocpd** (ongoing);
- Extend to more data distribution (long-tailed / bursty);
- Optimization to improve computational cost; . . .

NRC is hiring researchers in applied Machine Learning: https://bit.ly/2wlWiuc

The end

Thank you.

Questions?