

LIA RepLab systems

Topic Detection & Priority Filtering

Polarity Detection

J.V. Cossu
University of Avignon, France

CLEF - RepLab'2013

24 September 2013

Participants

- **J.-V. Cossu**, B. Bigot, L. Bonnefoy, M. Morchid, X. Bost, G. Senay, R. Dufour, V. Bouvier, J.-M. Torres-Moreno, M. El-Bèze

Idea

- Adaptation of LIA systems from other tasks or domains
- Check out the performances of these systems on RepLab dataset
- Light system adaptation

Participation to each subtasks

- Topic Priority (6 systems)
- Topic Detection (1 system)
- Filtering (7 systems)
- Polarity for Reputation classification (7 systems)
- 3 different merge when several systems

Idea

- Adaptation of a Theme Detection system in Speech Recognition outputs
 - Topic detection (10) in spoken conversations (call centre)
 - Noisy environment, automatic transcriptions with many errors

Common points

- Tweets look like a conversation ?
- There are mistakes, abbreviations...

Aim

- Robustness : can a system built for 10 topics be running with more than 3000 topics ?
- Find the biggest number of correct topics with global models
- Without external data or extra features

Used features (f_n)

- Uni-grams
- Bi-grams
- Distance bi-grams (distance = 1)
- Metadata: Language, Author, ...
- Lower-cased words

Maximum a Posteriori [Hazen, 2011]

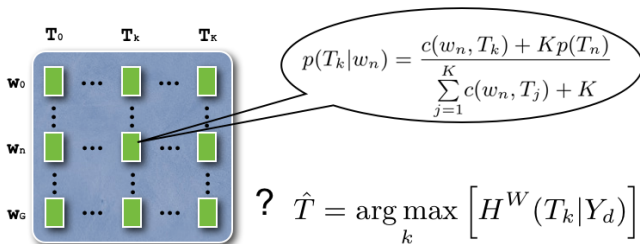
Feature purity computed on the train set:

$$p(T_k|f_n) = \frac{c(f_n, T_k) + Kp(T_n)}{\sum_{j=1}^K c(f_n, T_j) + K}$$

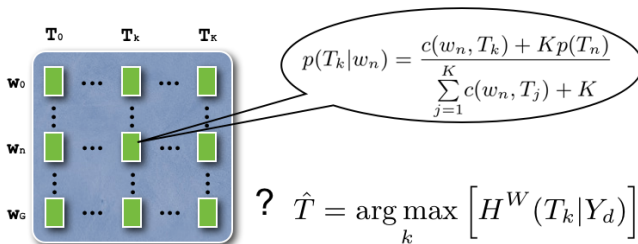
$c(w_n, T_k)$ is the number of times w_n appears in all the train with theme T_n

K =number of topics

$p(T_n)$ probability of topic n



now playing		listening, playing alicia's songs		now playing/listening	
jovi	0.022	alicia	0.027	whitney	0.021
bon	0.022	keys	0.026	houston	0.020
np	0.009	np	0.013	np	0.013
nowplaying	0.007	nowplaying	0.008	adele	0.012
listening	0.006	ft	0.007	nowplaying	0.009
bad	0.004	usher	0.007	you	0.006
always	0.004	by	0.006	love	0.006
bed	0.004	listening	0.005	by	0.007
by	0.004	song	0.005	listening	0.005



now playing		listening, playing alicia's songs		now playing/listening	
jovi	0.022	alicia	0.027	whitney	0.021
bon	0.022	keys	0.026	houston	0.020
np	0.009	np	0.013	np	0.013
nowplaying	0.007	nowplaying	0.008	adele	0.012
listening	0.006	ft	0.007	nowplaying	0.009
bad	0.004	usher	0.007	you	0.006
always	0.004	by	0.006	love	0.006
bed	0.004	listening	0.005	by	0.007
by	0.004	song	0.005	listening	0.005

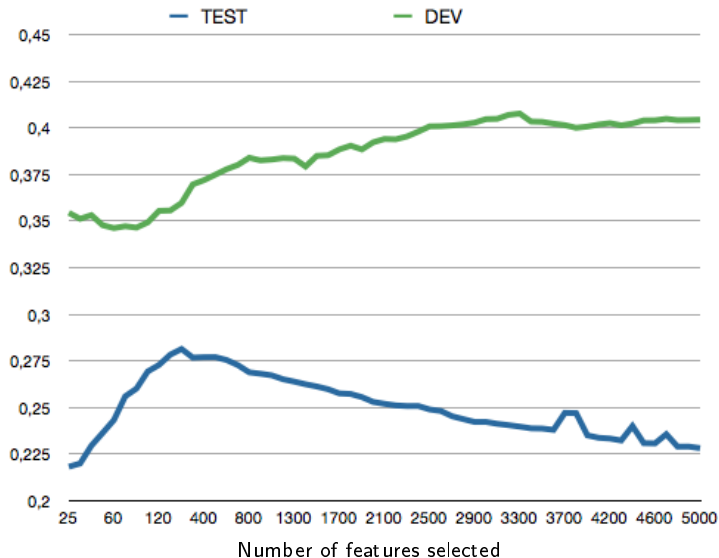
Tweet 281764404670889984 AlvinWongCH RL2013D04E145

#nowplaying adele- skyfall

TOPIC -> Now playing/listening

now playing		listening, playing alicia's songs		now playing/listening	
jovi	0.022	alicia	0.027	whitney	0.021
bon	0.022	keys	0.026	houston	0.020
np	0.009	np	0.013	np	0.013
nowplaying	0.007	nowplaying	0.008	adele	0.012
listening	0.006	ft	0.007	nowplaying	0.009
bad	0.004	usher	0.007	you	0.006
always	0.004	by	0.006	love	0.006
bed	0.004	listening	0.005	by	0.007
by	0.004	song	0.005	listening	0.005

F-Measure



Scores with N Features selected

Run Id	RELIABILITY	SENSITIVITY	F-MEASURE
Run 3	.2187	.3468	.2463
Run 2	.2342	.2730	.2435
Run 1	.3841	.1724	.2280
Run 4	.2538	.2222	.2267
<i>Median</i>	<i>.3659</i>	<i>.2180</i>	<i>.1954</i>
<i>Baseline</i>	<i>.1525</i>	<i>.2173</i>	<i>.1735</i>
Best	.4368	.2499	.2814

Topics correctly found with N Features selected

Run 1	679 topics correctly found
419	"other topics"
40	"mention of a product"
36	"u2 favourite songs"
30	"second hand selling / buying"
25	"4square"
21	"secondhand cars"
19	"nowplaying"

Run 2	648 topics correctly found
335	"other topics"
53	"u2 favourite songs"
46	"jokes"
39	"u2 fans"
37	"4square"
36	"second hand selling / buying"
35	"mention of a product"

Topics correctly found with N Features selected

Run 3	264 topics correctly found
75	"u2 fans"
32	"nowplaying"
31	"u2 favourite songs"
21	"4square"
13	"secondhand cars"
8	"praise for volvo"
7	"mtv"

Run 4	193 topics correctly found
36	"nowplaying"
36	"u2 favourite songs"
25	"4square"
24	"u2 fans"
19	"mention of a product"
17	"secondhand cars"
8	"lyrics"

Merge : Linear combination

- More topics correctly found (742 topics correctly found)
- Worse results regarding F-Measure (.243)

Analyse

- Results situated over the median and the baseline
- System's performance may vary with number of selected features
- No link between system's performance and number of correct topics found
- In terms of correct topics found the system doesn't work
- Topics never seen in the training set
- Confusion between :
 - "Now playing"
 - "Now playing/listening"
 - "Now Playing /Listening - #ARTIST"

Before your questions

- Metrics : severe and not representative of the real performance
- Data annotations :
 - Inter annotator agreement ?
 - Quality of annotations : Pictures on social networks, Pictures posted, Pictures posted in social networks, Pictures posted on social networks

Entity RL2013D04E162 (Maroon 5)

- No UNRELATED tweets in the training set
- 1 UNRELATED tweet in the test set
- All RELATED tweets correctly found
- The UNRELATED one was not found

- Result : Reliability=1 Sensitivity=0 F measure=0
- Really severe in this case
- Right if there was only 1 Related to find in the noise

Any questions ?