Centre for Data Analytics

A Mood-based Genre Classification of Television Content

Humberto Jesús Corona Pampín
Dr. Michael P. O'Mahony

UNIVERSITY COLLEGE DUBLIN, IRELAND

Insight

DCU   NUI Galway Oé Gaillimh   UCC University College Cork, Ireland Coláiste na hOllscoile Corcaigh   UCD DUBLIN   sfi Science Foundation Ireland

Centre for
Data Analytics

Insight

Research Application domains

200+ **RESEARCHERS**
8 **INSTITUTIONS**
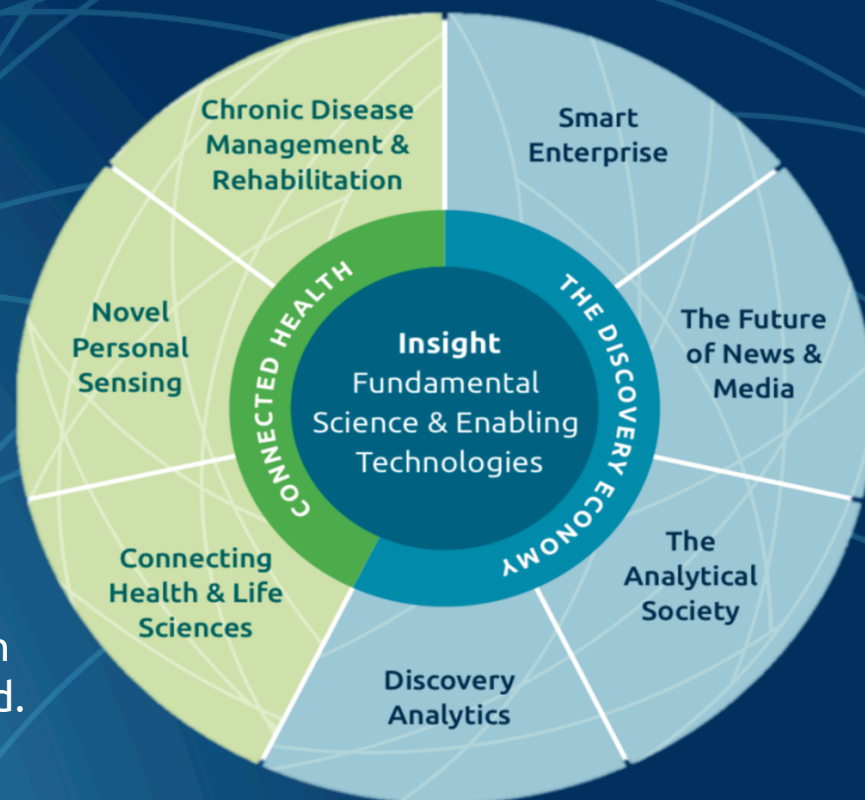30+ **INDUSTRY PARTNERS**
€88M **FUNDING**

Creating a data-driven society

New ways to capture and understand
data from the world around us,

Make better decisions for people,
communities and industries,

Create a more informed society in
a healthier, more productive world.

Chronic Disease
Management &
Rehabilitation

Smart
Enterprise

CONNECTED HEALTH

Novel
Personal
Sensing

THE DISCOVERY ECONOMY

Insight
Fundamental
Science & Enabling
Technologies

The Future
of News &
Media

The
Analytical
Society

Connecting
Health & Life
Sciences

Discovery
Analytics

www.insight-centre.org

DCU

NUI Galway
OÉ Gaillimh

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

UCD
DUBLIN

sfi
Science
Foundation
Ireland

INTRODUCTION

Choice overload. In 2009 there were 2,218 television broadcast stations in the U.S.

Previous work on TV content classification based on audio and visual features, focusing on genre*

A need for new ways to classify and describe TV content

Supervised genre classification of television programs using a textual representation obtained from transcripts

To build a personalized content-based #recsys

* J. Eggink and D. Bland, "A Large Scale Experiment for Mood-Based Classification of TV Programmes," in *Proceedings of the 2012 IEEE International Conference on Multimedia and Expo*, 2012, pp. 140–145.
* R. S. Jasinschi and J. Louie, "Automatic TV Program Genre Classification Based on Audio Patterns," in *Proceedings of the 27th. IEEE Euromicro Conference.*, 2001, pp. 370–375.

TALK OVERVIEW

1. **Russell's model of affect (valence, arousal)**

2. **Extending Russell's model of affect with ANEW**

3. **Using the Boxfish API to build our dataset**

4. **Preliminary Feature Analysis**
   1. **A year of television content**
   2. **Focus: A year of news content**

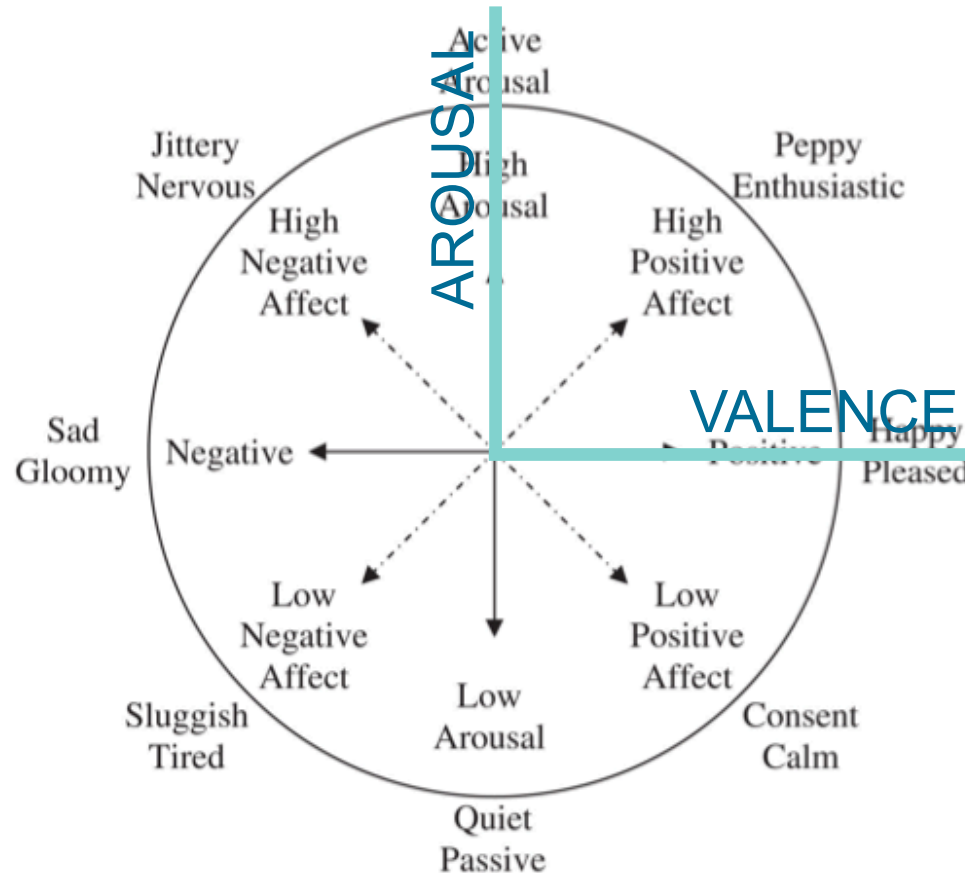5. **Program genre classification approach and Results**

6. **Conclusions + Q&A**

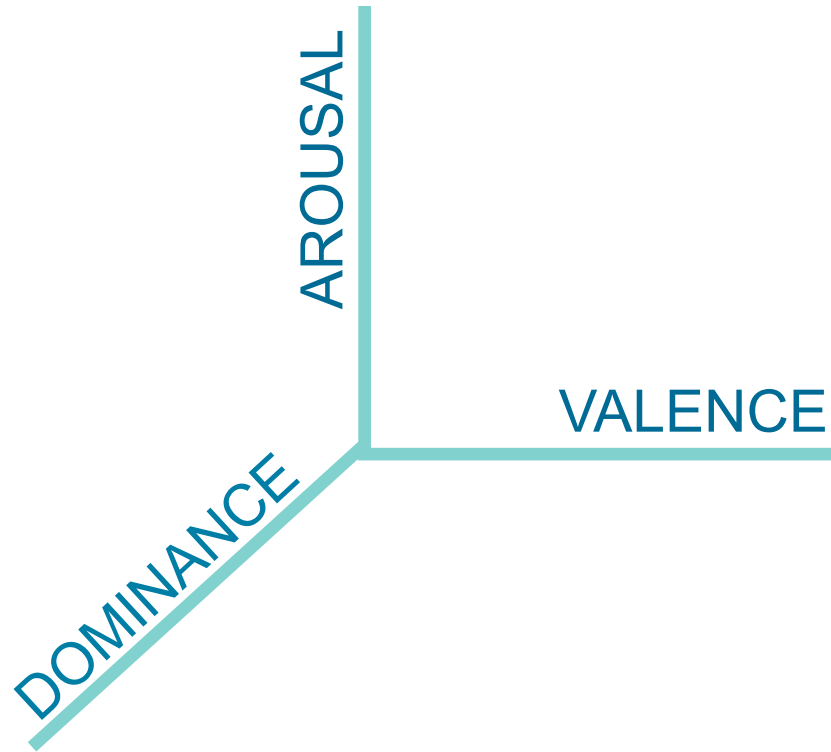# FIRST THINGS FIRST …

RUSSELL'S MODEL OF AFFECT
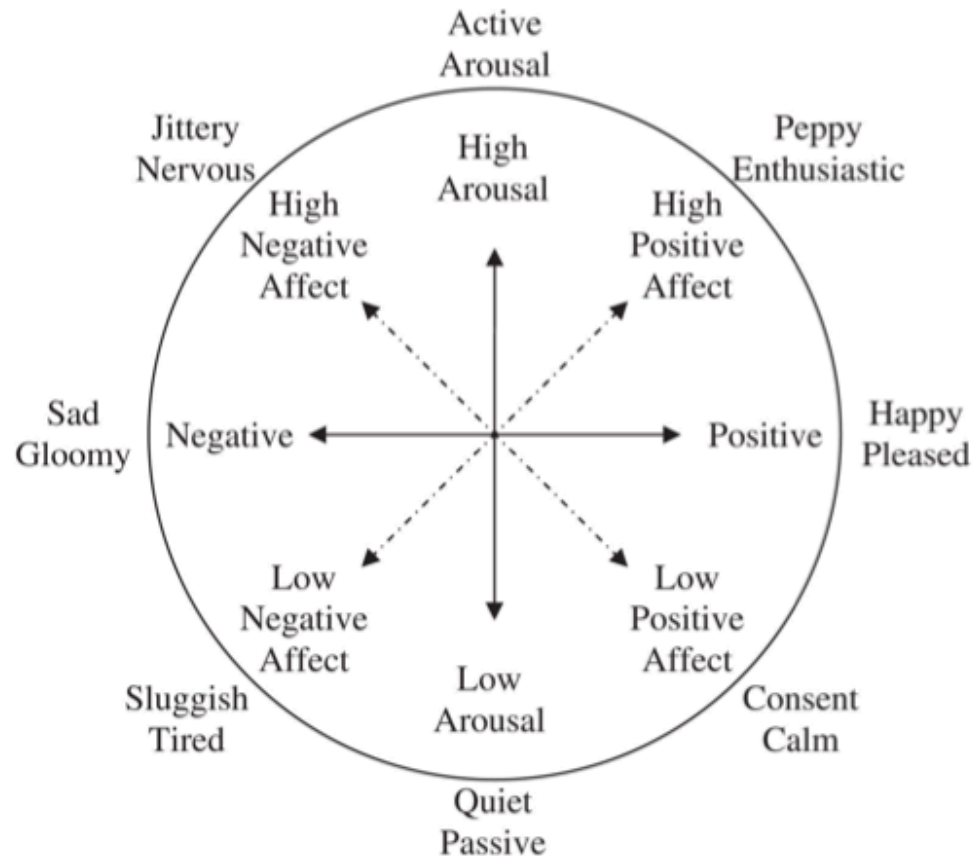
J. A. Russell, "A Circumplex Model of Affect.," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

RUSSELL'S MODEL OF AFFECT

J. A. Russell, "A Circumplex Model of Affect.," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

RUSSELL'S MODEL OF AFFECT

J. A. Russell, "A Circumplex Model of Affect.," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

ANEW DATASET

word	Valence		Arousal		Dominance	
	mean	std.	mean	std.	mean	std.
Abduction	2.76	2.06	5.53	2.43	3.49	2.38
Able	6.74	2.00	4.30	2.17	6.83	2.04
Abortion	3.50	2.3	5.39	2.8	4.59	2.54
Absent	3.69	1.72	4.73	1.76	4.35	1.87
Absurd	4.26	1.82	4.36	2.2	4.73	1.72
Abundance	6.59	2.01	5.51	2.63	5.80	2.16
Abuse	1.80	1.23	6.83	2.7	3.69	2.94
Accept	6.80	2.11	5.53	1.96	5.41	1.92
Acceptance	7.98	1.42	5.40	2.7	6.64	1.91
Access	6.14	1.62	5.07	1.68	6.25	1.53

Table 12: The first ten words of the ANEW dataset.

2,476 words

Annotated with emotional ratings in three dimensions:
valence, arousal and dominance.

Provides a set of normative emotional ratings for the words included.

M. Bradley and P. Lang, "Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings," 1999.

BUILDING A META-FEATURE REPRESENTATION

Lyrics for Michael Jackson's Billie Jean

"She was more like a beauty queen from a movie scene.
⋮
And mother always told me,
be careful who you love.
And be careful of what you do
'cause the lie becomes the truth.

Billie Jean is not my lover,
She's just a girl who claims
that I am the one.
⋮

ANEW words	v_k	f_k
$k=1.$ love	8.72	1
2. mother	8.39	1
3. baby	8.22	3
4. beauty	7.82	1
5. truth	7.80	1
6. people	7.33	2
7. strong	7.11	1
8. young	6.89	2
9. girl	6.87	4
10. movie	6.86	1
11. perfume	6.76	1
12. queen	6.44	1
13. name	5.55	1
14. lie	2.79	1

$$v_{\text{text}} = \frac{\sum_k v_k f_k}{\sum_k f_k}$$

$v_{\text{Billie Jean}} = 7.1$

$v_{\text{Thriller}} = 6.3$

$v_{\text{Michael Jackson}} = 6.4$

P. S. Dodds and C. M. Danforth, "Measuring the Happiness of Large-Scale Written Expression: Songs, Blogs, and Presidents," Journal of Happiness Studies, vol. 11, no. 4, pp. 441–456, Jul. 2009.

BUILDING A META-FEATURE REPRESENTATION

Feature Group	Feature
ANEW features (valence, arousal, dominance)	Minimum value
	Maximum value
	Mean value
	Standard deviation
	Median value
Stylistic features	Num. words
	Num. unique words
	Num. unique ANEW words
	Max. word frequency

THE DATA

BOXFISH DATA - CHANNELS

BOXFISH DATA - API ENDPOINTS

EPG

The /epg/ endpoint provides access to program (TV show) EPG data.

EPG data provided is for reference and should not be used as the basis of a consumer application. API consumers wishing to build EPG driven applications must obtain a license from an EPG data provider, e.g. FYI Television, Rovi or TMS

METRICS

The /metrics/ endpoint provides real time counts for mentions on TV.

Results can be returned in various levels of granularity, agregated by minute, hour, day, month or year. Metrics are recorded for exact keyword matches against detected entities. Non-entity keywords (e.g. numbers, months and adjectives) are not available.

The availability of certain granularities is restricted to limit the kilobyte size of responses. For example, per minute results for a one year period are not available. If data is required in such detail, multiple smaller (e.g. daily) requests can be made and concatenated by the client.

(PRELIMINAR) FEATURE ANALYSIS

PRELIMINAR FEATURE ANALYSIS, A YEAR OF TELEVISION BROADCAST

1. **Select 8 channels**
   FOX, CNN, E!, SyFy, FOX News, CN, MSNBC, Discovery

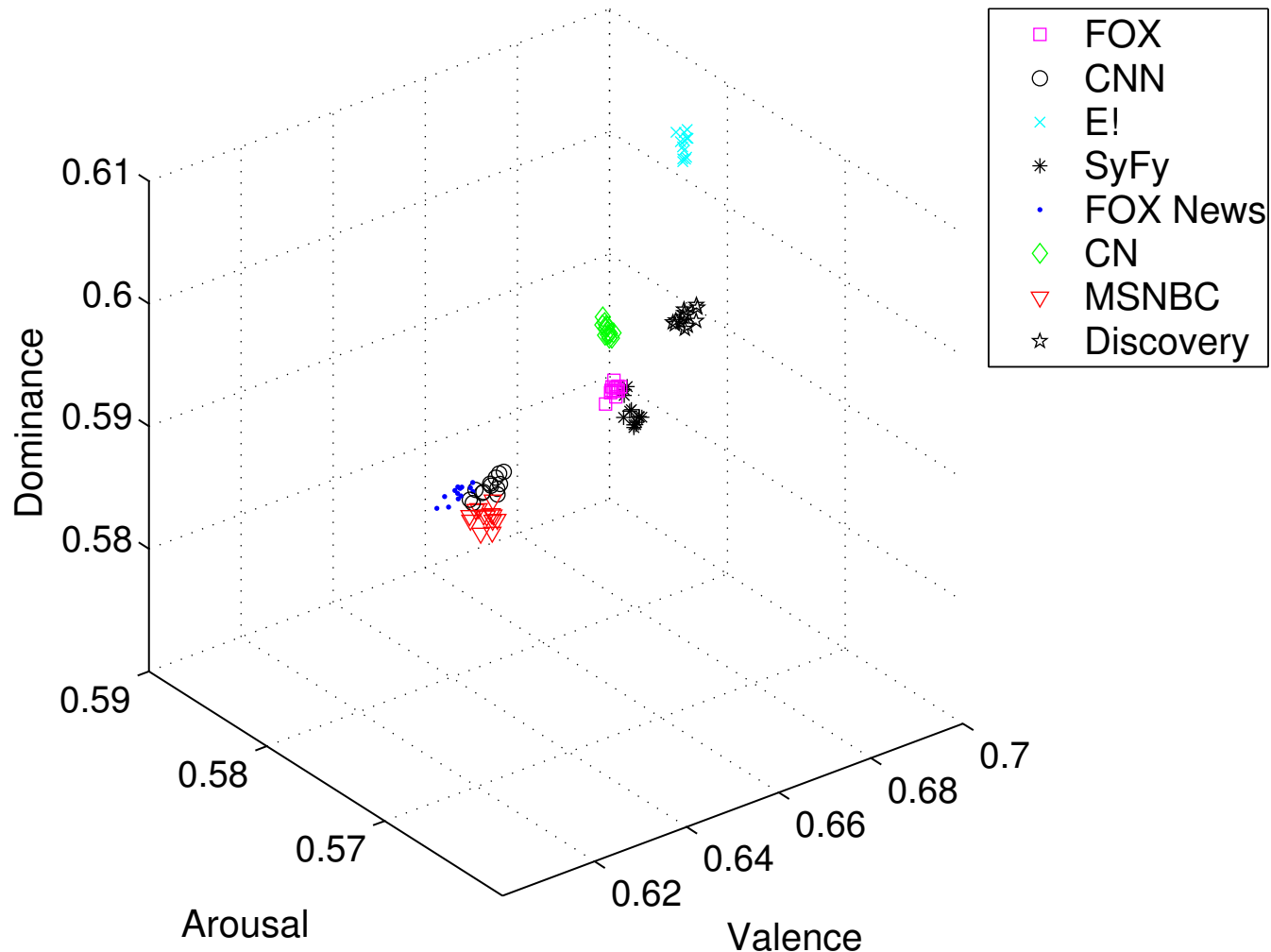2. **Select a year of content (January – December 2013)**

3. **Query term occurrences on a weekly period**

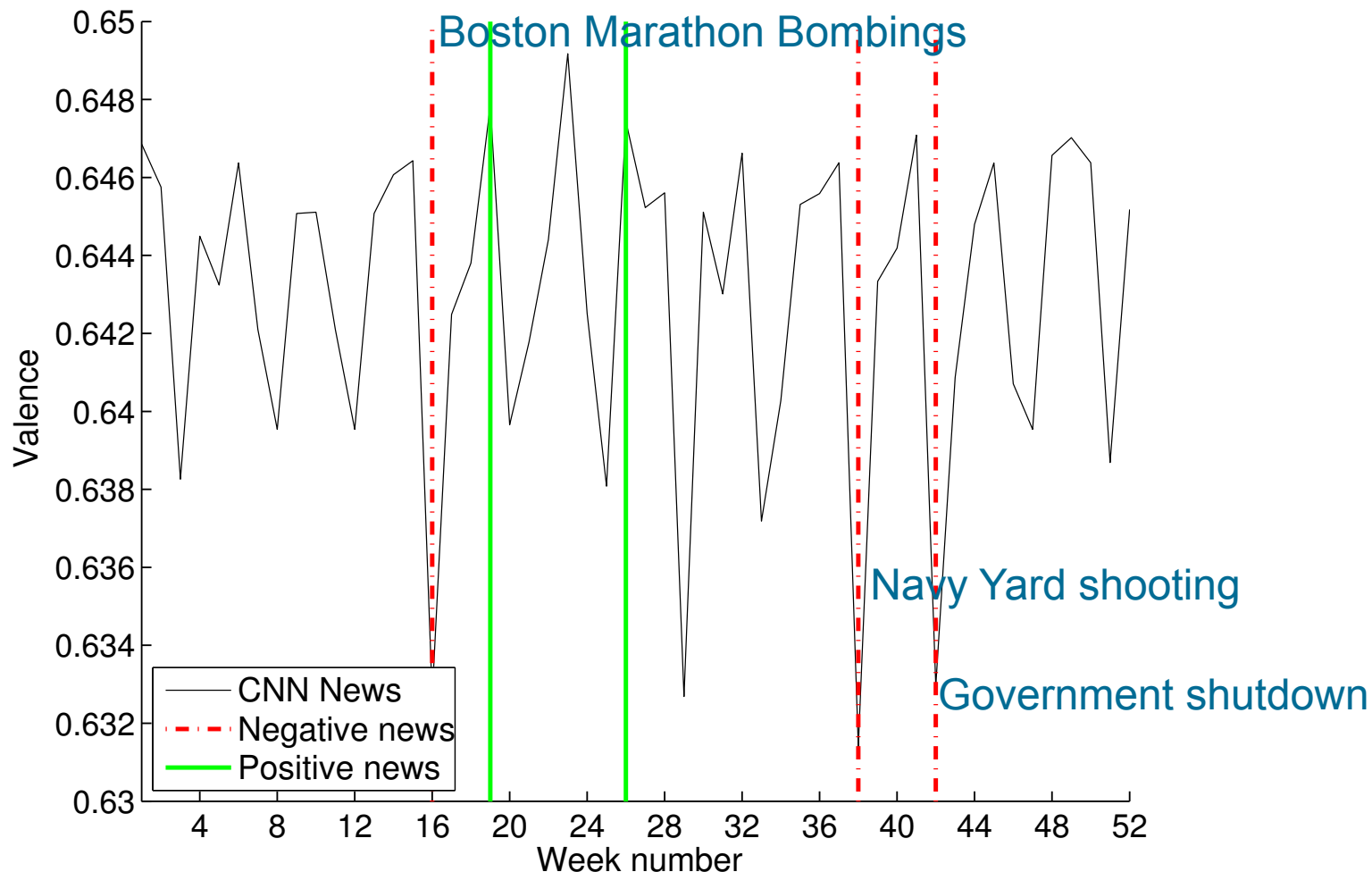4. **Calculate valence, arousal, dominance**

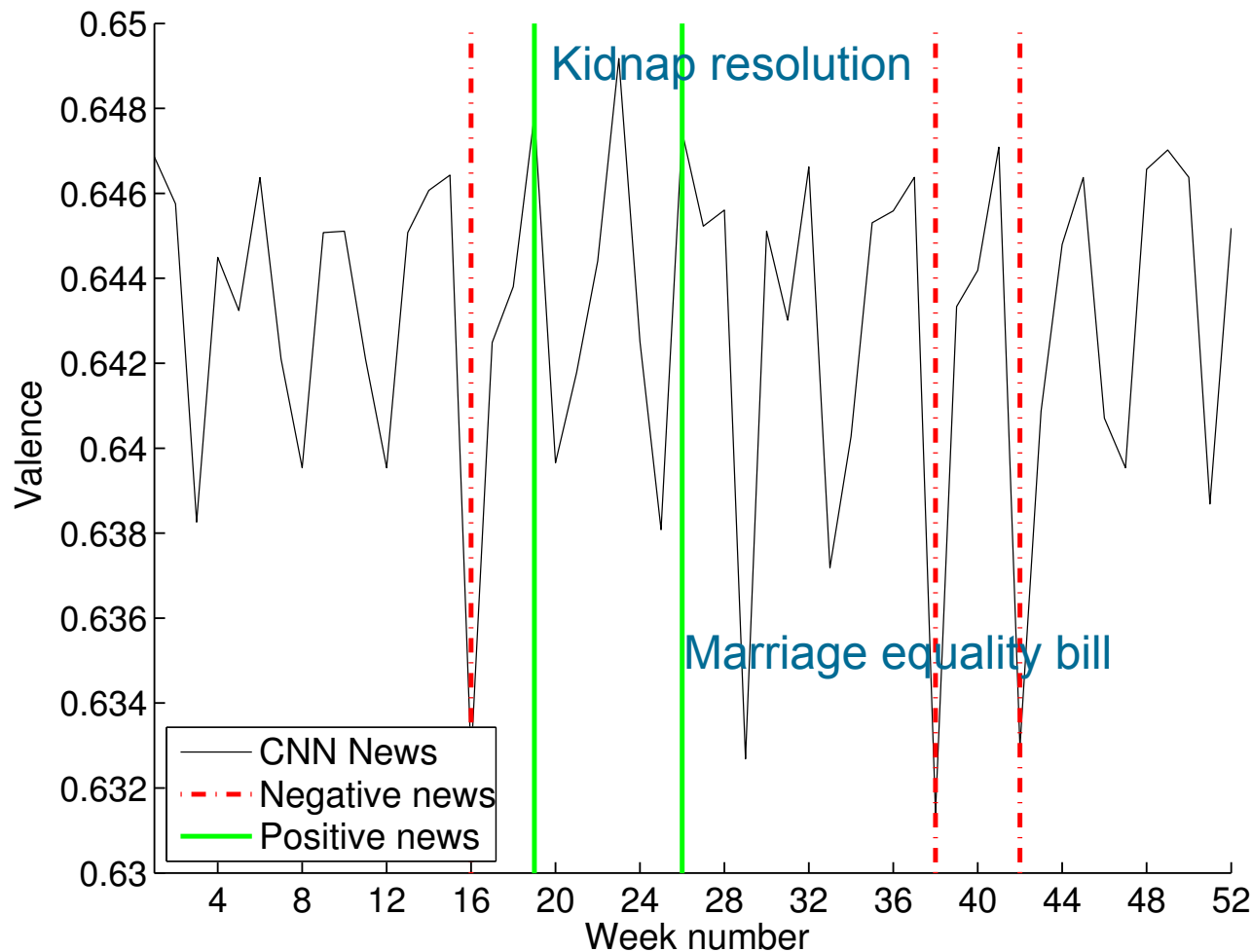5. **Average results over 4 week period (13 periods)**

PRELIMINAR FEATURE ANALYSIS, A YEAR OF TELEVISION BROADCAST

PRELIMINAR FEATURE ANALYSIS, A YEAR OF NEWS

PRELIMINAR FEATURE ANALYSIS, A YEAR OF NEWS

PRELIMINAR FEATURE ANALYSIS (CONCLUSION)

The valence and dominance dimensions, in particular, can be used to distinguish between the content broadcast by different channels.

However, as mentioned previously, classifying the genre of individual television programs (for use in a recommender system) based on these dimensions may present challenges (given the variance observed within genres), an analysis of which is considered in the next section

A SUPERVISED APPROACH
TO TELEVISION PROGRAM CLASSIFICATION
USING A META-FEATURE REPRESENTATION

PROGRAM GENRE CLASSIFICATION

- **2 weeks of content (February 2014)**

- **Select genres with >= 20 programs**

GENRE	NUMBER OF ITEMS
Animated	
documentary	
horror	
newscast	
reality	
total	384

PROGRAM GENRE CLASSIFICATION (FEATURES)

Feature Group	Feature
ANEW features (valence, arousal, dominance)	Minimum value
	Maximum value
	Mean value
	Standard deviation
	Median value
Stylistic features	Num. words
	Num. unique words
	Num. unique ANEW words
	Max. word frequency

PROGRAM GENRE CLASSIFICATION (RESULTS)

- Naïve Bayes classifier (WEKA machine learning framework)
- Meta-features Approach
- Vector Space Model approach (VSM)

Genre	VSM			Meta-features		
	TP	FP	AUC	TP	FP	AUC
Animated	0.725	0.004	0.942	0.792	0.161	0.885
Documentary	0.554	0.025	0.870	0.447	0.144	0.780
Horror	0.208	0.013	0.649	0.583	0.107	0.865
Newscast	0.976	0.053	0.972	0.707	0.056	0.905
Reality	0.882	0.260	0.845	0.333	0.064	0.729
Weighted Average	0.729	0.084	0.885	0.583	0.115	0.824

CONCLUSIONS

- Television content, at a high level, can be discriminated by the proposed three-dimensional space of affect

- Classifying the genres of individual television programs using this approach in general did not outperform a traditional VSM

- There is evidence to suggest that meta-features based on valence, arousal and dominance values have the potential to contribute to enhanced classification performance

- Future work:
    - Consider using the model of affect in a personalised content-based recommendation approach,
    - Conducting live user studies to understand how individuals respond to mood-based recommendations.

THANKS!

IT'S ALL ON GITHUB
REUSE, REPEAT, REPRODUCE
https://github.com/hcorona/recsystv-2014

CONTACT ME!

humberto.corona@insight-centre.org

http://ie.linkedin.com/pub/humberto-corona/28/943/863

twitter: @totopampin

github: https://github.com/hcorona