Deep Learning for Sentiment Analysis: Case Studies

Francesco Pugliese, PhD

Email **Francesco Pugliese** : neural1977@gmail.com

Case Study: Sentiment Analysis for Economy

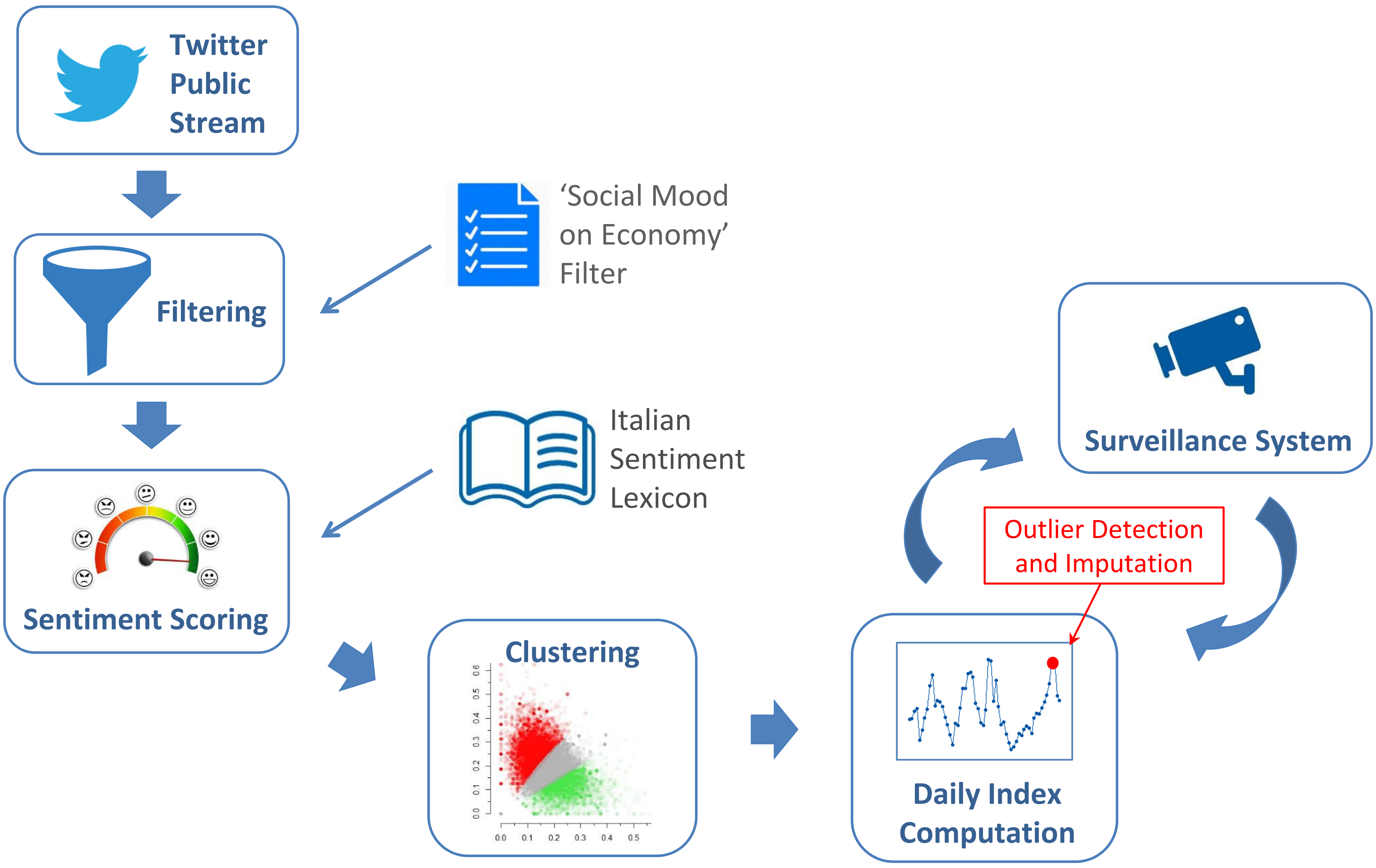
- Nowadays more and more people are using Social Media platforms to find out news, to express their feelings and to share or debate opinions about virtually every possible topic
  - The interest towards Social Media as a means for "measuring" public's mood is still growing

- We are investigating whether social media messages may be successfully exploited to develop *domain-specific* sentiment indices. The aim is to assess the Italian mood about *specific* topics or aspects of life, e.g.
  - the economic situation, the European Union, the migrants' phenomenon, the terrorist threat, and so on

- These new indices would enable high-frequency (e.g. daily) measures of the Italian sentiment about phenomena which are of interest in Official Statistics

- The hope is that such indices could either improve the performance of forecasting models, or enrich existing statistical products (e.g. the BES), or even be disseminated as new statistical outputs in their own right

Processing Pipeline at a Glance



Data Collection and Storage

- **Data Collection Technique**
  - We exploit Twitter's Streaming API to get low latency access to Twitter's firehose and collect samples of public tweets
- **Target Population**
  - Public tweets whose text matches at least one keyword belonging to the filter
- **Sampling Design**
  - The sampling algorithm is a black box, as it is entirely controlled by Twitter's Streaming API. At most a 1% of all the tweets produced on Twitter at a given time can be sampled
- **Data Format**
  - Twitter's Streaming API returns data in JSON format
- **Data Staging and Storage**
  - We temporarily store gathered JSON data as text files inside a staging area residing on a server. Then we periodically load bunches of tweets into an Oracle DB (and remove the corresponding files from the staging area)

Text Processing and Sentiment Analysis

- **Process Granularity**
  - To compute daily index values, we process all the tweets collected in a single day as a single block

- **Input Data**
  - We only analyze the textual content of the tweets. (No information about users is ever accessed: the index only uses *unlinked anonymized* data)

- **Text Cleaning and Normalization**
  - We perform standard NLP pre-processing steps: (i) convert to lowercase, (ii) tokenize running text into words, (iii) apply basic orthographic repairs, (iv) remove URLs, (v) remove non-alphabetic characters (e.g. ‘#’ or ‘@’), (vi) remove stop words, (vi) *if needed*, stem words to get rid of inflected forms

- **Sentiment Analysis Approach**
  - To classify tweets as Positive, Negative or Neutral we chose to adopt an unsupervised, lexicon-based approach
  - We discarded supervised, Machine Learning approaches because we were unable to find large, high quality training sets of human-labeled tweets in Italian

Sentiment Scores: The Lexicon

- Our Sentiment Analysis application involves two sequential steps
  1) Calculate sentiment scores for each tweet
  2) Use these sentiment scores to cluster tweets into three mutually exclusive classes: Positive (P), Negative (N) and Neutral (U)

- To attach sentiment scores to a tweet we leverage an Italian Sentiment Lexicon, namely a vocabulary whose lemmas are associated to pre-computed *positive* and *negative* sentiment scores

- Currently we are using the **Sentix** lexicon [Basile and Nissim 2013]

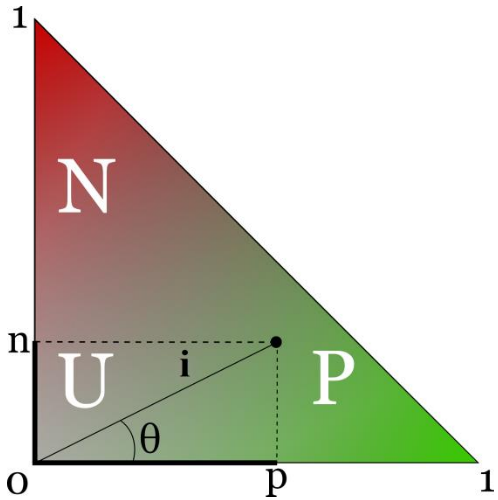
- Since it aligns several existing, independent lexical resources (*WordNet*, *MultiWordNet*, *BabelNet*, *SentiWordNet*) Sentix contains many duplicated lemmas
  - ~75’000 lemmas overall, only ~42’000 unique
    - ➡ To ensure unambiguous and reproducible results we de-duplicated Sentix by averaging atomic sentiment scores of duplicated lemmas

The Sentiment Space

- In Sentix, positive (p) and negative (n) sentiment scores of lemmas are constrained as follows:

$$\begin{cases} p \in [0, 1] \\ n \in [0, 1] \\ p + n \leq 1 \end{cases}$$

- Therefore Sentix maps lemmas to points belonging to the sentiment triangle:



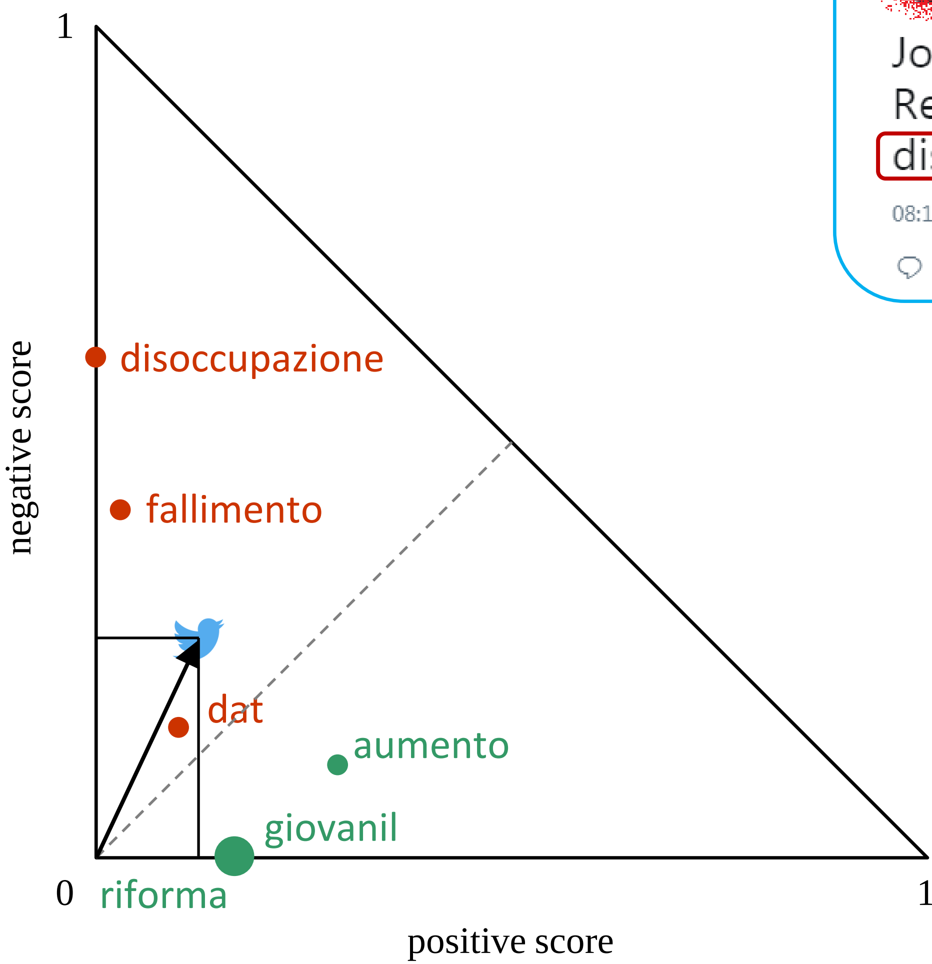
- From (p, n) coordinates we can pass to polar coordinates (i, θ) and derive two *additional* sentiment scores:
  - ✓ **Polarity** $\omega = 1 - 4\theta/\pi$ $\omega \in [-1, 1]$
  - ✓ **Intensity** $i = \sqrt{p^2 + n^2}$ $i \in [0, 1]$

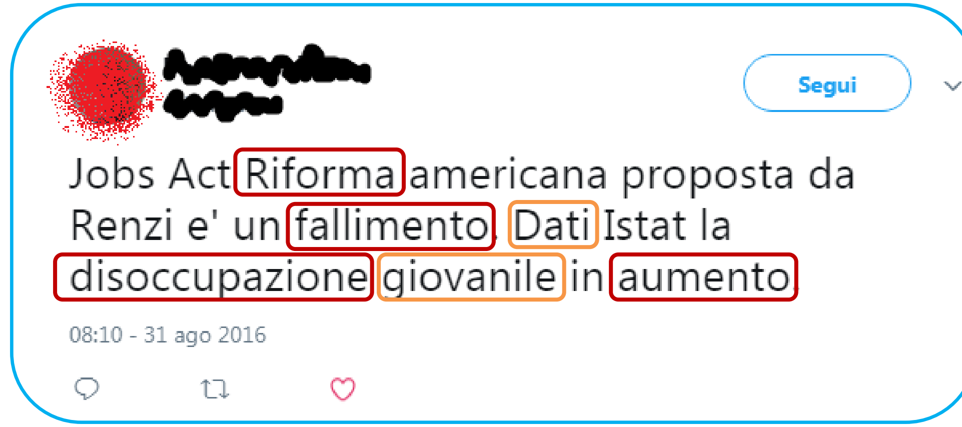
- This way Sentix lemmas are mapped to a 4D sentiment space

lemma	pos	neg	polarity	intensity
caldo	0.25	0.125	0.41	0.28
freddo	0.047	0.297	-0.8	0.3

➡ To enable clustering, tweets too must be mapped to this 4D space

From Word-level to Tweet-level Sentiment Scores



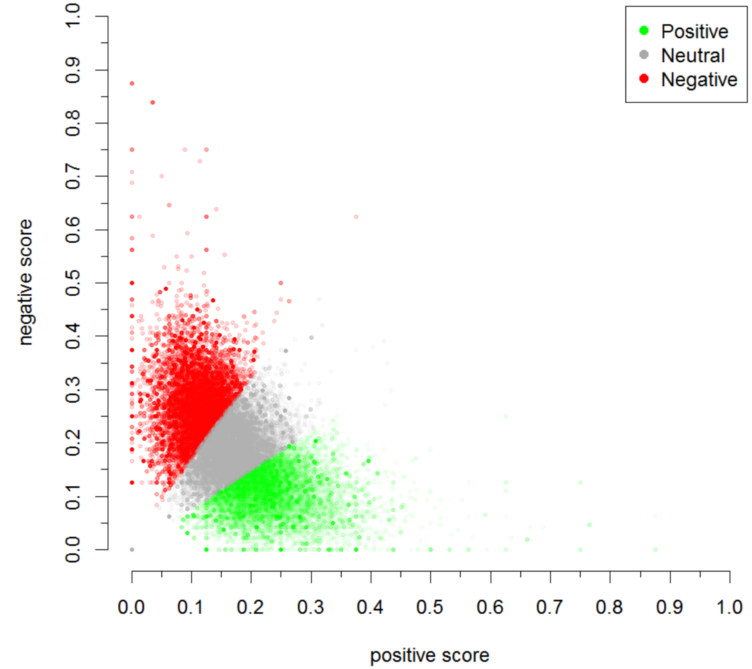


word	pos	neg	polarity	intensity
riforma	0.125	0	1	0.125
fallimento	0.021	0.375	-0.929	0.376
dat	0.063	0.104	-0.312	0.121
disoccupazione	0	0.625	-1	0.625
giovanil	0.125	0	1	0.125
aumento	0.208	0.083	0.516	0.224

pos	neg	polarity	intensity
0.090	0.198	-0.455	0.218

Clustering and Calculation of the Index

- Once sentiment scores (p, n, ω, i) are available for all the tweets of a daily block…
- …we use K-means to cluster them into Positive, Negative and Neutral tweets
  - ✓ to lower the risk of finding a local optimum, we run it 100 times with random starts and pick the best solution
- Lastly we compute the daily index value (S), which depends on the distribution of tweets within the Positive, Neutral and Negative classes



$$\boxed{\mathrm{S} = \bar{\omega}_{\mathrm{i}} = \frac{\sum_t \mathrm{i}_t\, \omega_t}{\sum_t \mathrm{i}_t} = \frac{\sum_{t\in P} \mathrm{i}_t\, \omega_t + \sum_{t\in N} \mathrm{i}_t\, \omega_t}{\sum_t \mathrm{i}_t}}$$

where $\omega_t \stackrel{\text{def}}{=} 0 \;\; \forall t \in \text{Neutral}$

- This index can be seen as the average of polarity (ω) weighted by intensity (i), provided we treat *Neutral* tweets as if their polarity were *zero*. Compared to traditional alternatives:
  - ✓ It is more resilient to tweets' misclassification
  - ✓ It reduces day-to-day volatility

Monitoring and Validation

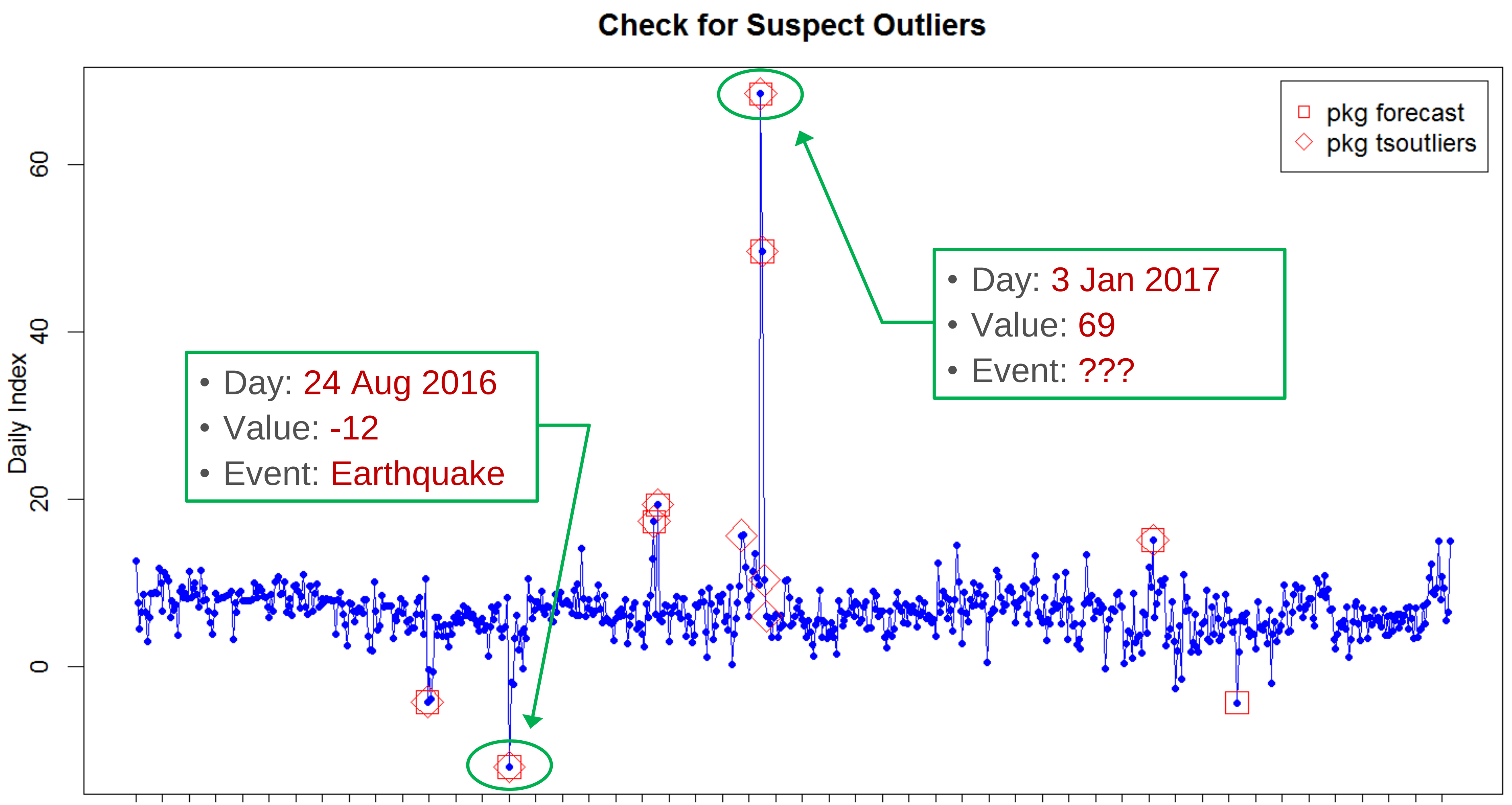
- *No filter is perfect!* Thus we devoted special care to make the index robust against possible contaminations by off-topic tweets that might pass the filter

- We developed a surveillance system, which periodically searches for anomalous values in the daily time series by means of two independent and complementary outlier detection routines
  - Daily values detected as potential outliers cause the system to generate a set of automated diagnostic reports
  - These are then sent to human reviewers in charge of deciding whether the detected values are actually proper data points, or instead truly anomalous

- Truly anomalous data typically arise when an off-topic tweet that happened to pass the filter becomes **"viral"** on Twitter
  - Being re-twitted and quoted thousands of times in a day, viral tweets may have an unduly impact on the daily index and introduce bias

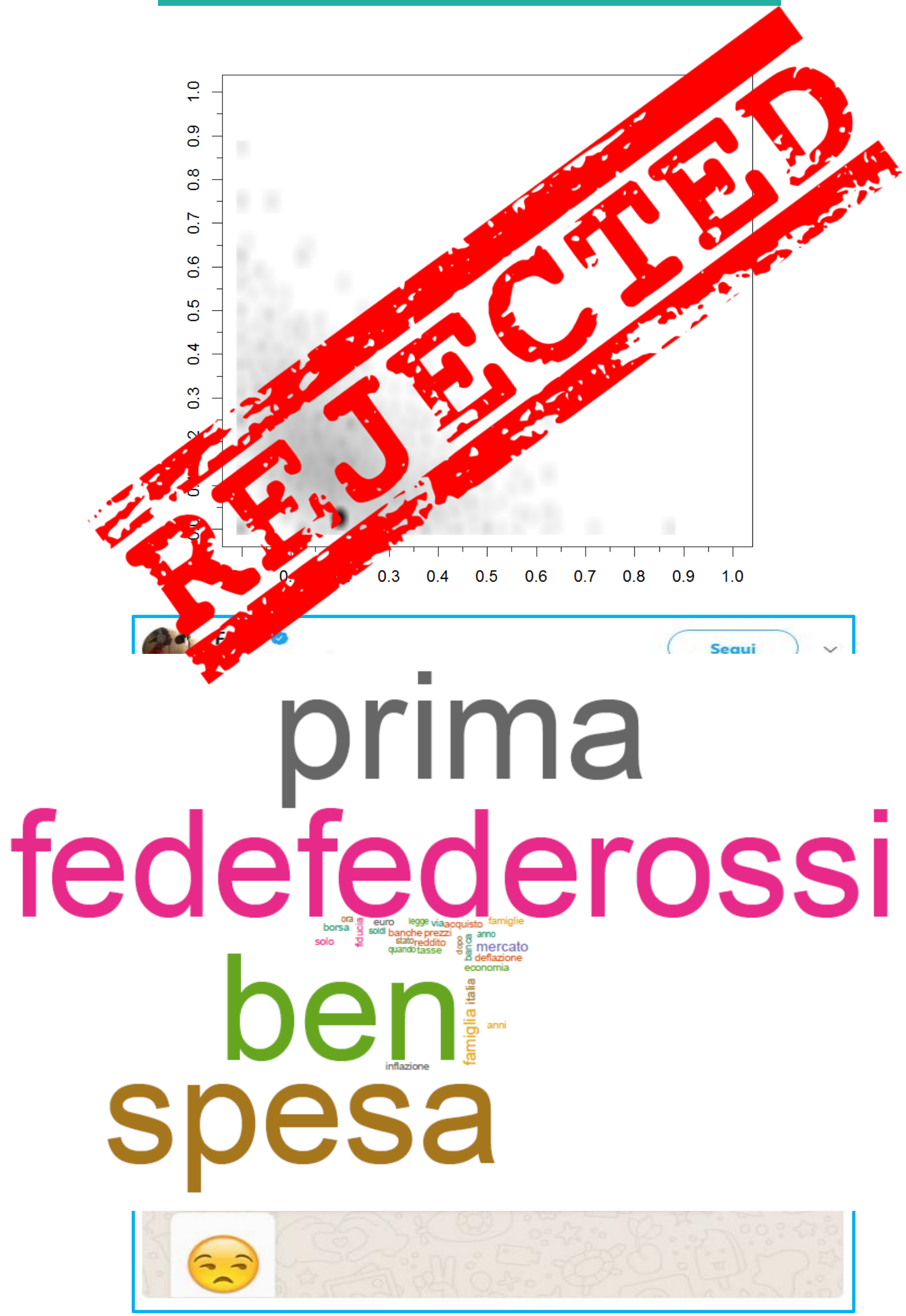
- All the daily index values classified as truly anomalous are eventually imputed via nearest-neighbor interpolation

Anomalous Values: One Example

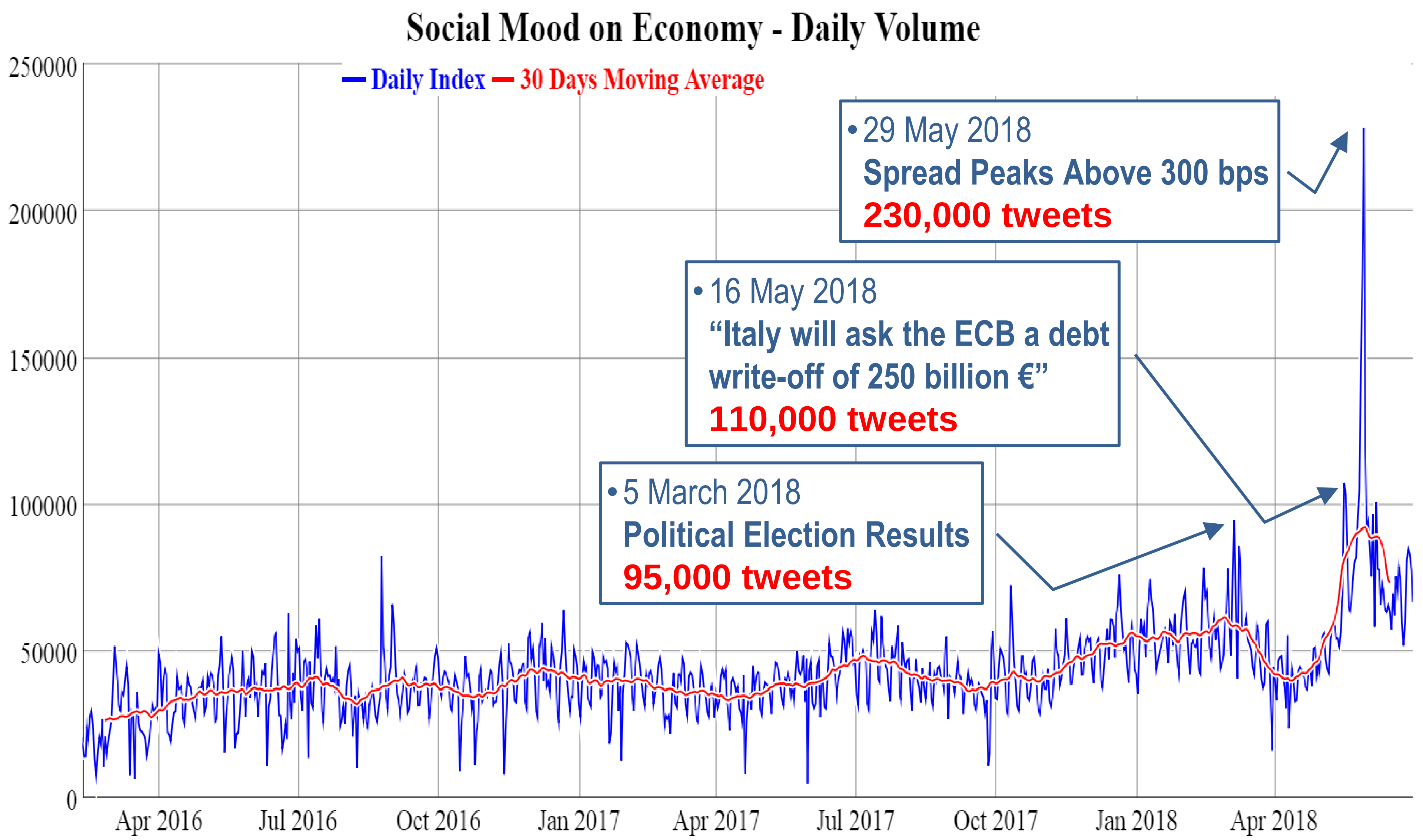


24 Aug 2016

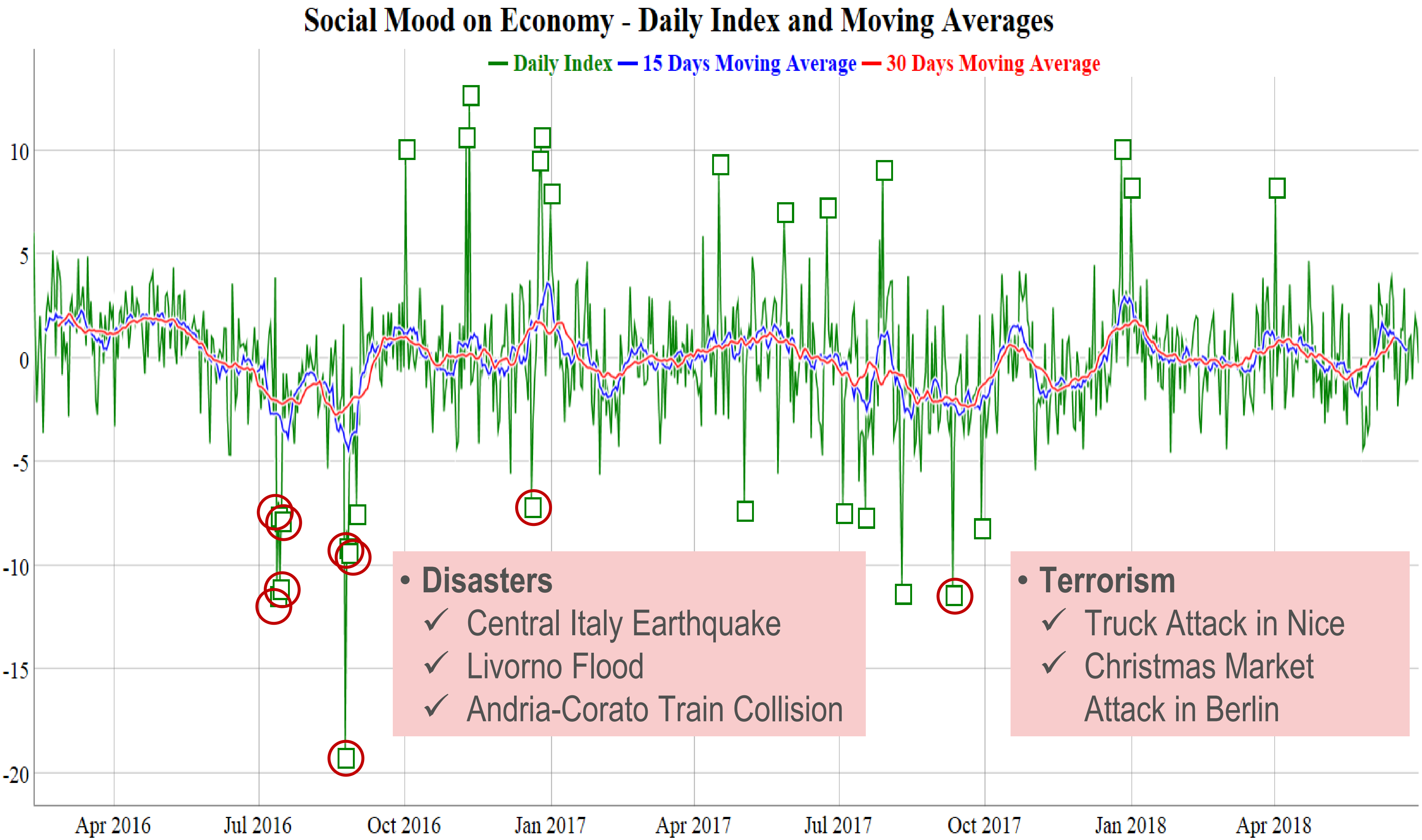
3 Jan 2017



Volume Burst in 2018 Post-Election Crisis



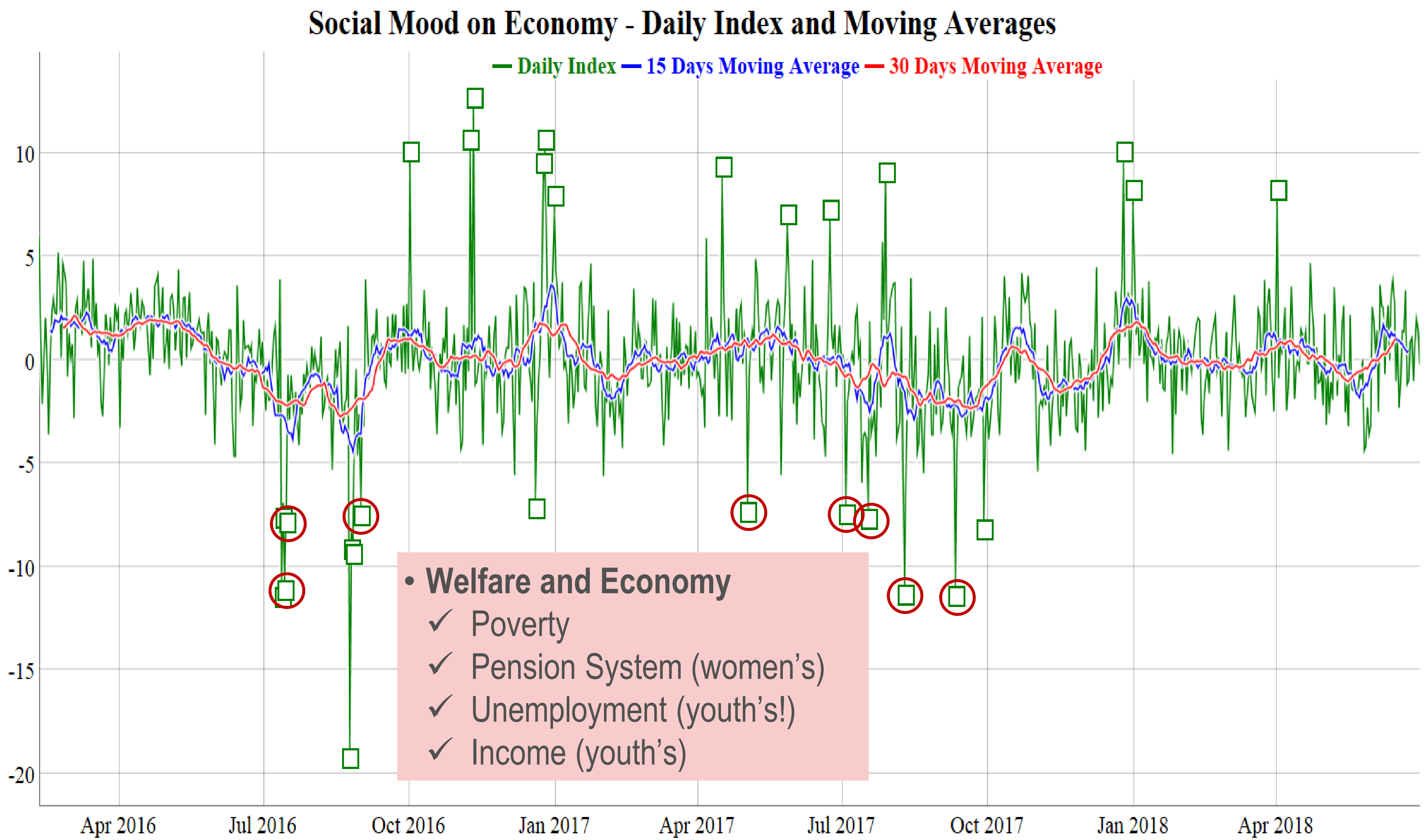
Valleys: Disasters and Terrorism



- **Disasters**
  - ✓ Central Italy Earthquake
  - ✓ Livorno Flood
  - ✓ Andria-Corato Train Collision

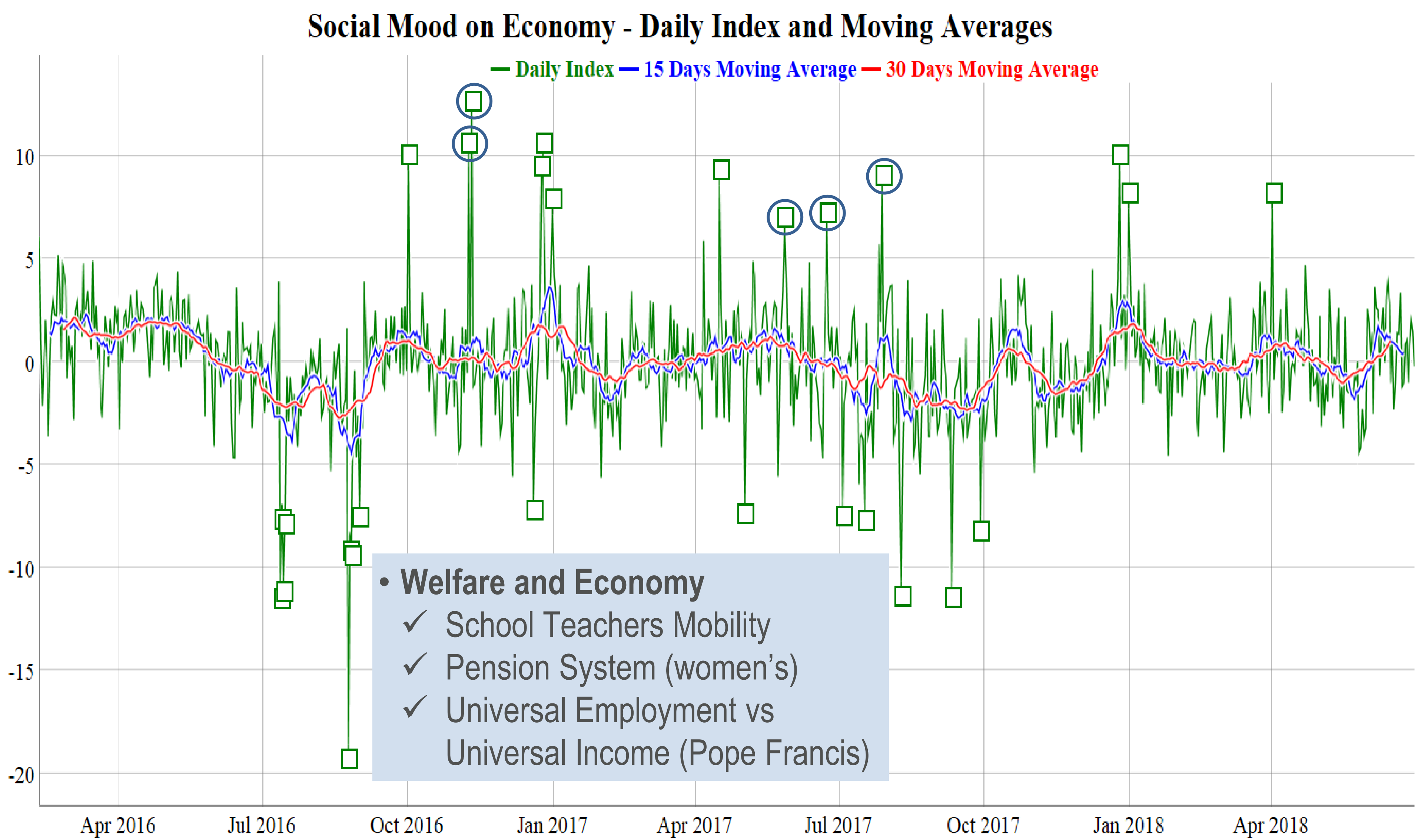
- **Terrorism**
  - ✓ Truck Attack in Nice
  - ✓ Christmas Market Attack in Berlin

Valleys: Welfare and Economy



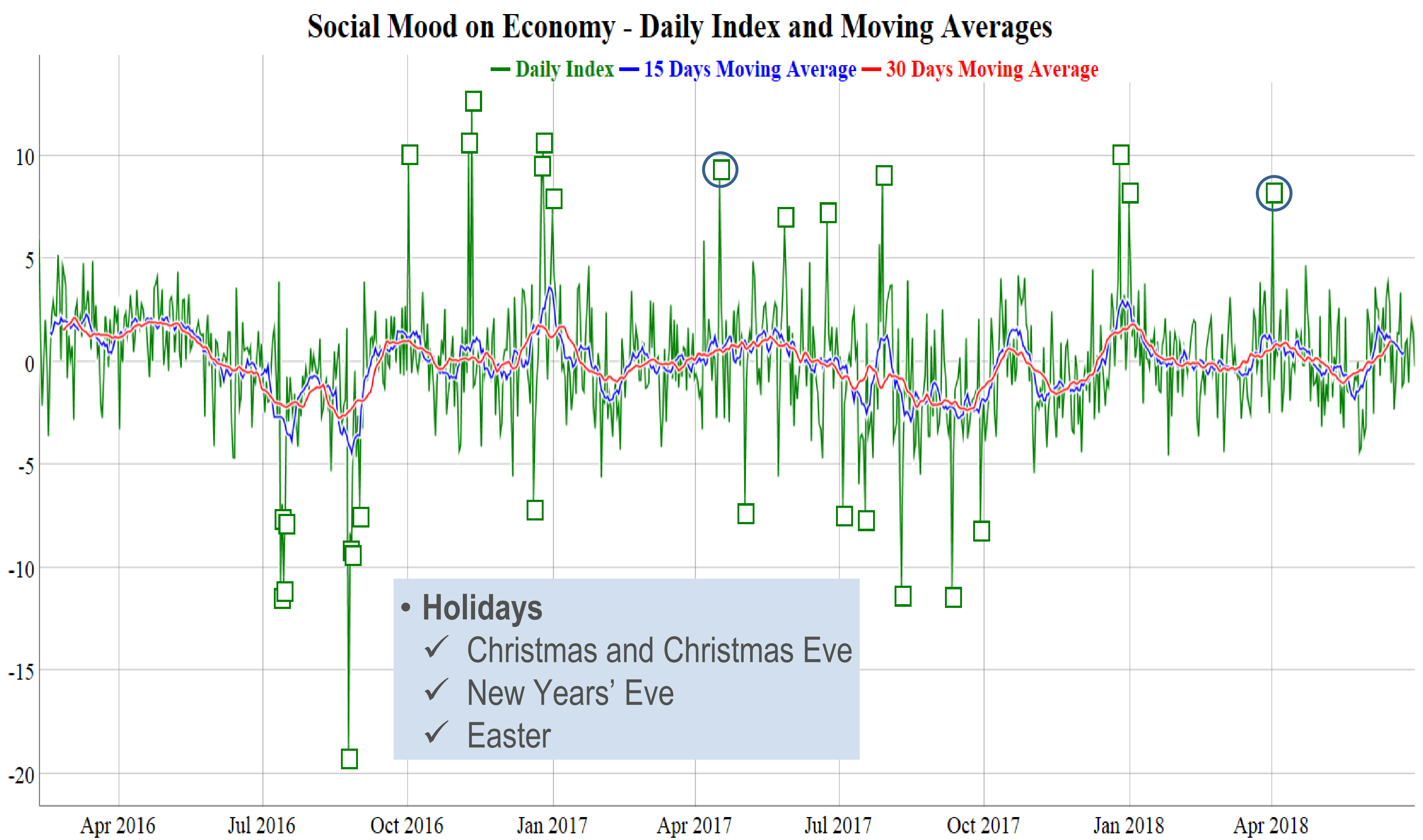
- **Welfare and Economy**
  - ✓ Poverty
  - ✓ Pension System (women's)
  - ✓ Unemployment (youth's!)
  - ✓ Income (youth's)

Peaks: Welfare and Economy



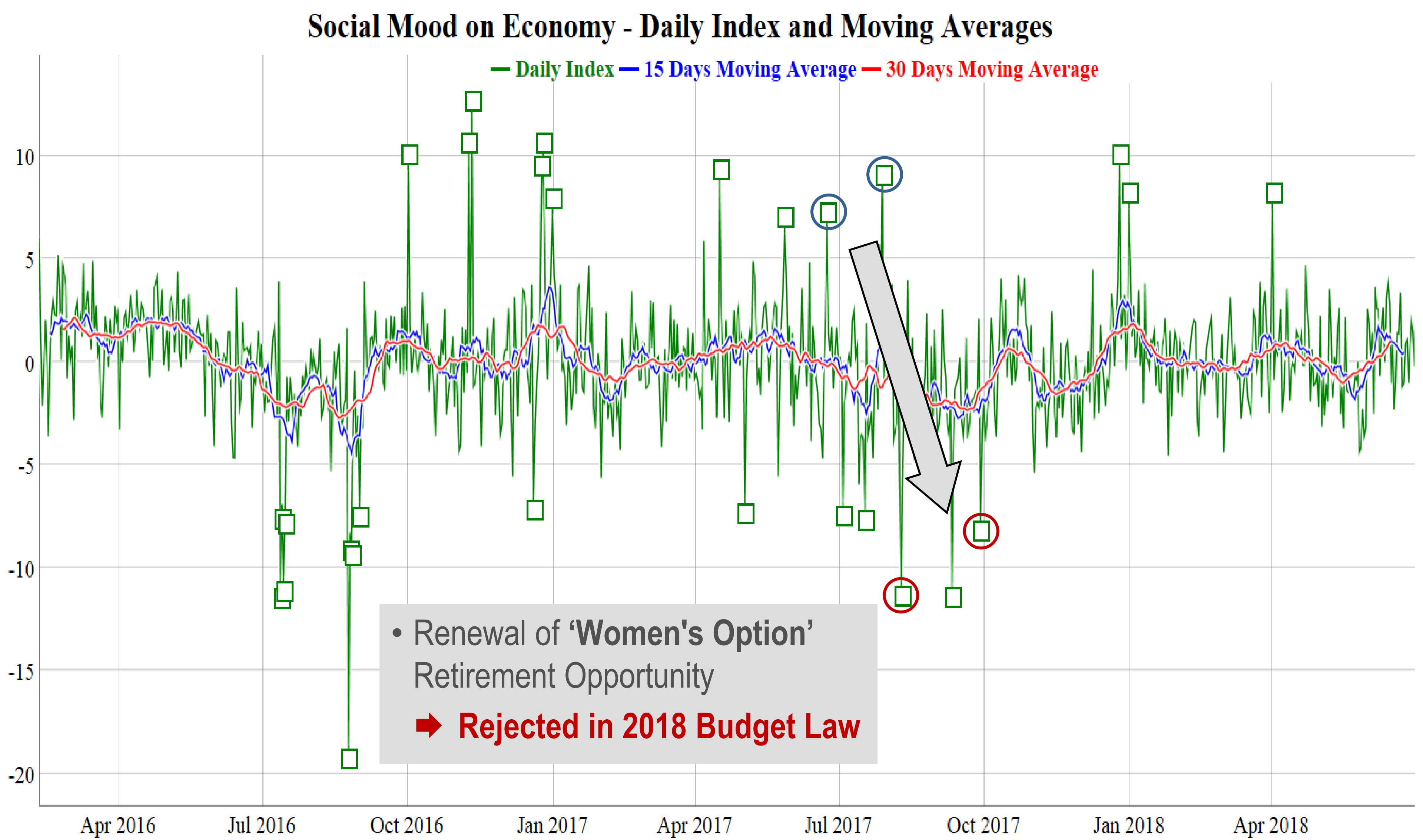
- **Welfare and Economy**
  - ✓ School Teachers Mobility
  - ✓ Pension System (women's)
  - ✓ Universal Employment vs Universal Income (Pope Francis)

Peaks: Holidays



- **Holidays**
  - ✓ Christmas and Christmas Eve
  - ✓ New Years' Eve
  - ✓ Easter

An Interesting Dynamic



Sentiment Analysis (*Ain, et al. 2017*)

- **<u>Sentiments</u>** of users that are expressed on the web has great influence on the readers, product vendors and politicians.
- **<u>Sentiment Analysis</u>** refers to text organization for the classification of mind-set or feelings in different manners such as negative, positive, favorable, unfavorable, thumbs up, thumbs down, etc. Thanks to DL, the SA can be visual as well.

Discovering people opinions, emotions and feelings about a product or service

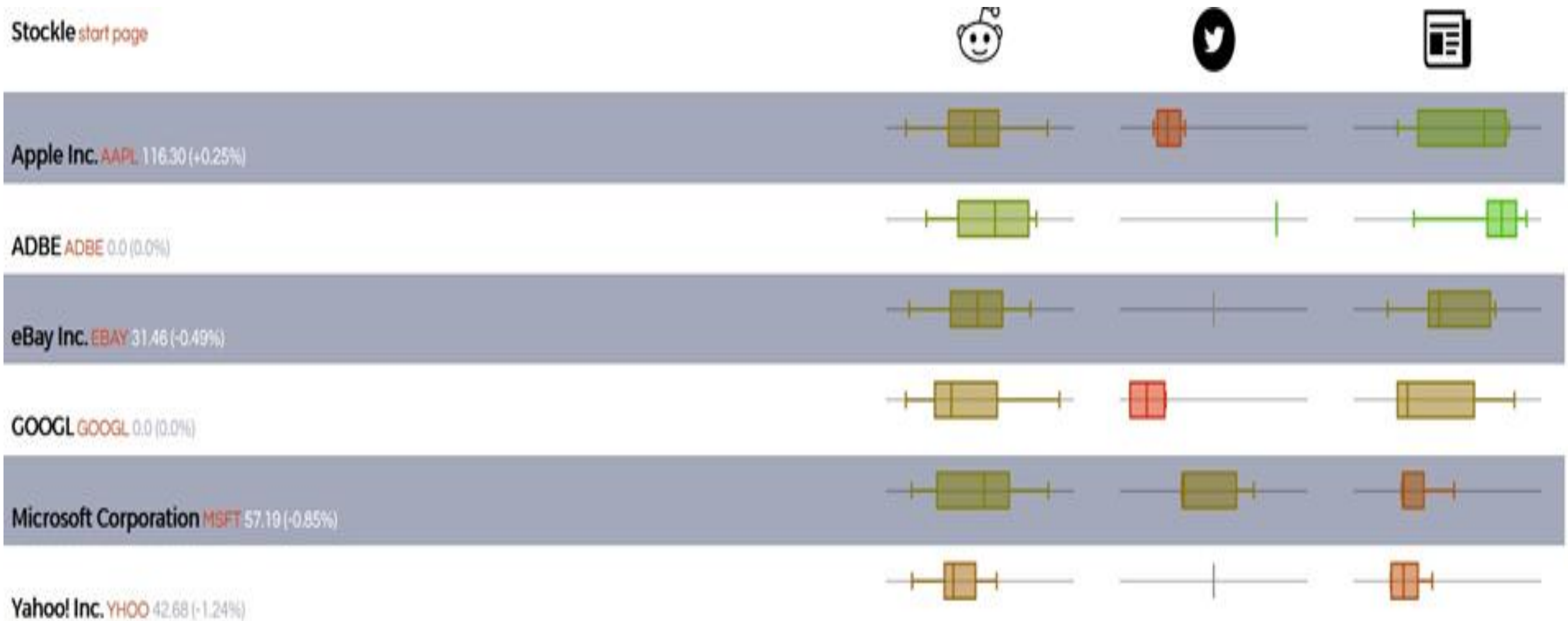
Sentiment Analysis Using Deep Learning Techniques: A Review

Qurat Tul Ain*, Mubashir Ali*, Amna Riaz†, Amna Noureen‡, Muhammad Kamran‡, Babar Hayat* and A. Rehman*
*Department of Computer Science and Information Technology,
The University of Lahore, Gujrat, Pakistan
†Department of Information and Technology,
University of Gujrat, Gujrat, Pakistan
‡Department of Computer Engineering,
EME, NUST, Islamabad, Pakistan

a) Machine learning based techniques: This type of techniques are implemented by extracting the sentences and aspect levels. The features consist of Parts of Speech (POS) tags, n-grams, bi-grams, uni-grams and bag-of-words. Machine learning contains three flavors at sentence and aspect, i.e., Nave Bayes, Support Vector Machine (SVM) and Maximum Entropy.

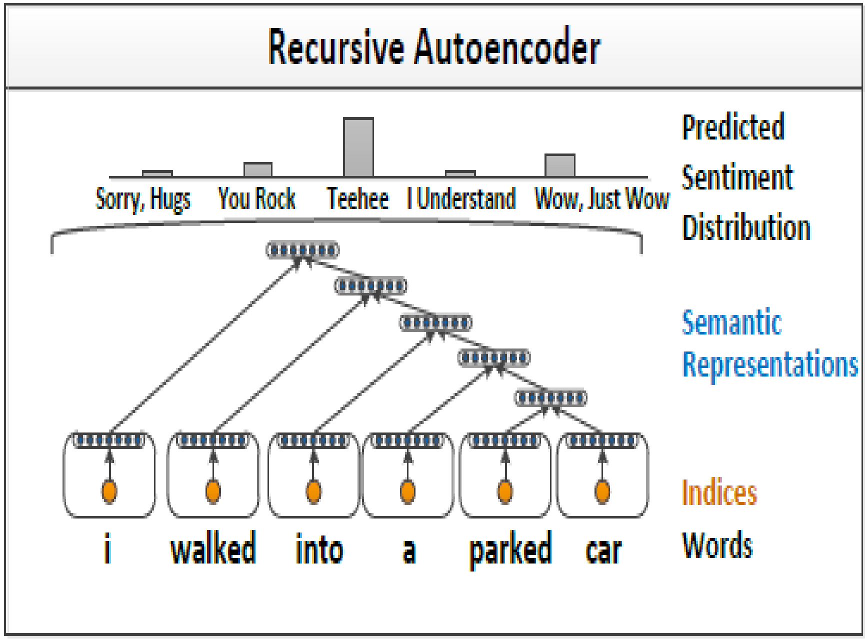
b) Lexicon based or corpus based techniques: These techniques are based on decision trees such as k-Nearest Neighbors (k-NN), Conditional Random Field (CRF), Hidden Markov Model (HMM), Single Dimensional Classification (SDC) and Sequential Minimal Optimization (SMO), related to methodologies of sentiment classification.

Sentiment Analysis with Feedback



Recursive Neural Networks (RecursiveNN) (Socher, R., et al., 2011b)

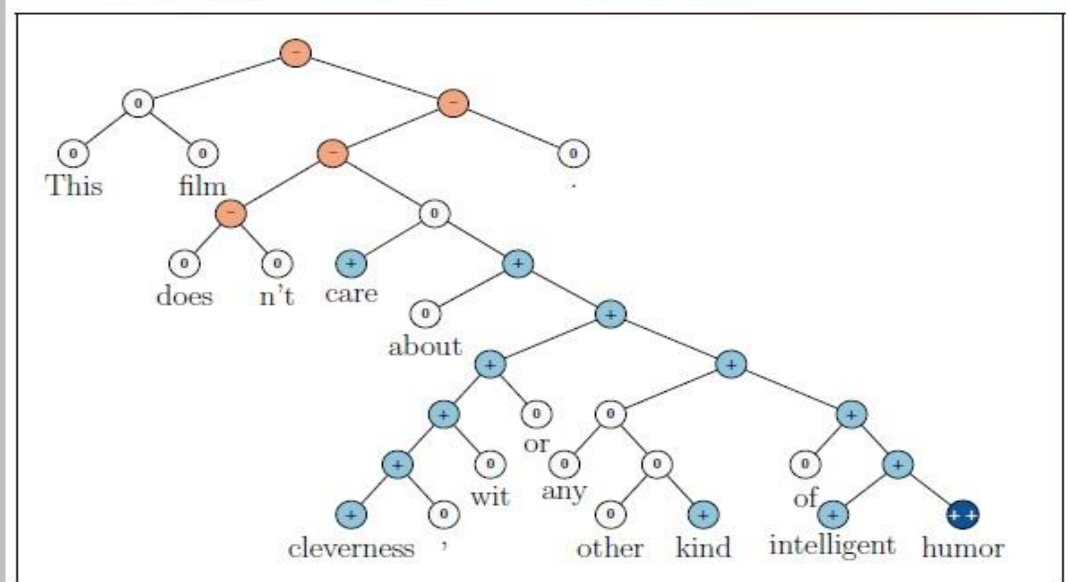
- These models are recursive auto-encoders which learn semantic vector representations of phrases. Word indices (orange) are first mapped into a semantic vector space (blue).
- Then they are recursively merged by the same auto-encoder network into a fixed length sentence representation. The vectors at each node are used as features to predict a distribution over text labels.



Recursive Neural Tensor Networks (RNTN) (Socher, R., et al. 2013)

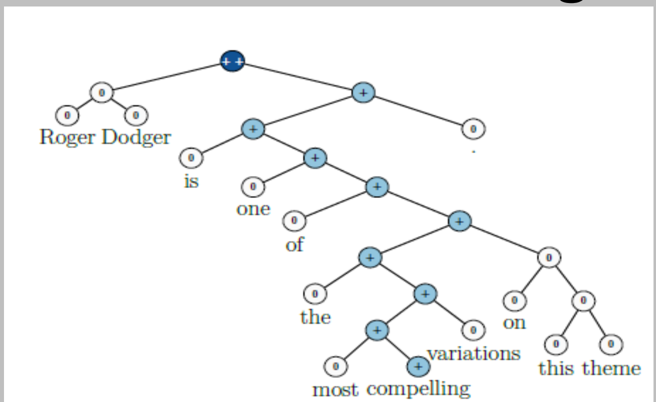
- **The Stanford Sentiment Treebank is the first corpus with fully labeled parse trees that allows for a complete analysis of the compositional effects of sentiment in language.**

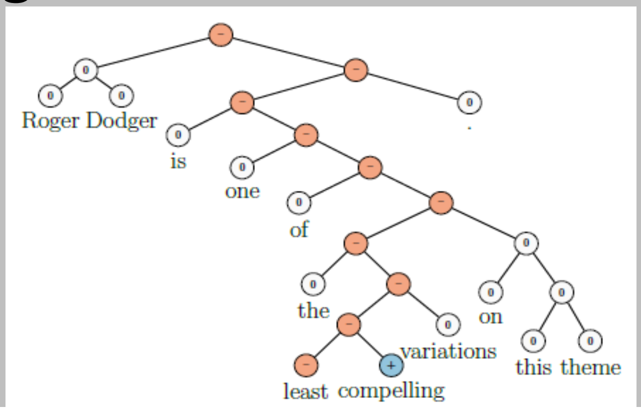
- **<u>RNTNs</u> compute parent vectors in a bottom up fashion using a compo-sitionality function and use node vectors as features for a classifier at that node.**



RNTN – Upside and Downside

- **RNTNS are very efficient in terms of constructing sentence representations.**
- **RNTNs capture the semantics of a sentence via a tree structure. Its performance heavily depends on the performance of the textual tree construction.**
- **Constructing such a textual tree exhibits a time complexity of at least O(n2), where n is the length of the text.**
- **RNTNs are unsuitable for modeling long sentences or documents.**



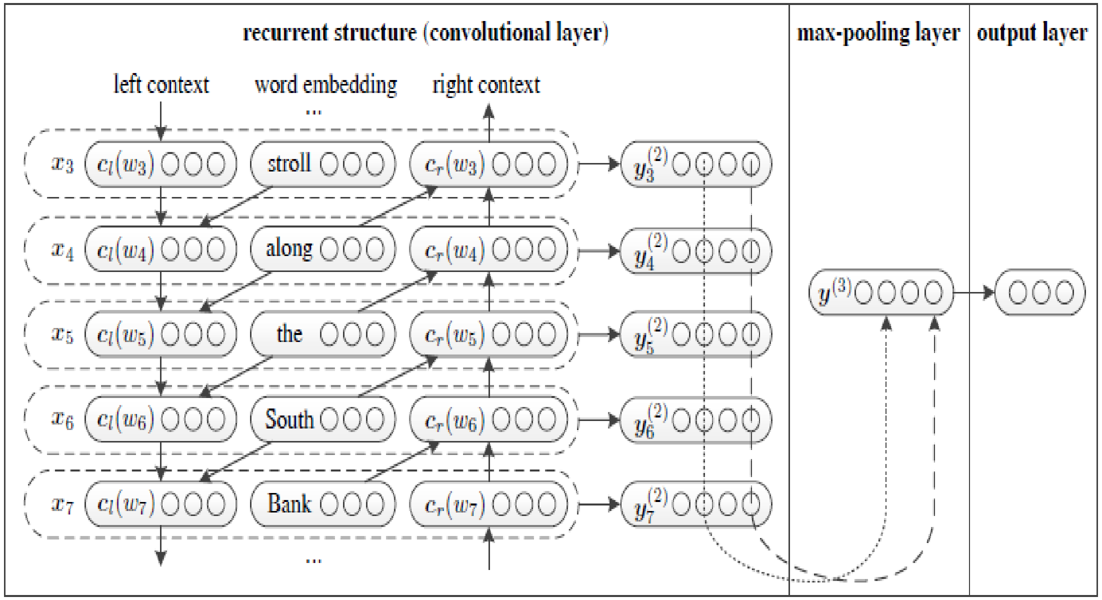


Recurrent Convolutional Neural Networks (RCNN) (*Lai, S., et al. 2015*)

- They adopt a recurrent structure to **capture contextual information** as far as possible when learning word representations, which may introduce considerably **less noise compared** to traditional window-based neural networks.

- The **bi-directional recurrent structure** of RCNNs.

- **RCNNs** exhibit a time complexity of O(n)



RCNN Equations

- RCNNs exhibit a **time complexity of O(n)** , which is linearly correlated with the length of the text length.

$$c_l(w_i) = f(W^{(l)} c_l(w_{i-1}) + W^{(sl)} e(w_{i-1})) \quad (1)$$

$$c_r(w_i) = f(W^{(r)} c_r(w_{i+1}) + W^{(sr)} e(w_{i+1})) \quad (2)$$

$$x_i = [c_l(w_i); e(w_i); c_r(w_i)] \quad (3)$$

$$y_i^{(2)} = \tanh\left(W^{(2)} x_i + b^{(2)}\right) \quad (4)$$

$$y^{(3)} = \max_{i=1}^{n} y_i^{(2)} \quad (5)$$

$$y^{(4)} = W^{(4)} y^{(3)} + b^{(4)} \quad (6)$$

$$p_i = \frac{\exp\left(y_i^{(4)}\right)}{\sum_{k=1}^{n} \exp\left(y_k^{(4)}\right)} \quad (7)$$

- **7 equations** defining all the Neural Network topology
- **Input length** can be variable

RCNN: Feature Extraction

- RCNNs employ a max-pooling layer that automatically judges which words play key roles in text classification to capture the key components in texts.

- The most important words are the information most frequently selected in the max-pooling layer.

- Contrary to the most positive and most negative phrases in RNTN, RCNN does not rely on a syntactic parser, therefore, the presented n-grams are not typically “phrases”.

	RCNN
P	well *worth* the; a *wonderful* movie; even *stinging* at; and *invigorating* film; and *ingenious* entertainment; and *enjoy* .; 's *sweetest* movie
N	A *dreadful* live-action; Extremely *boring* .; is *n't* a; 's *painful* .; Extremely *dumb* .; an *awfully* derivative; 's *weaker* than; incredibly *dull* .; very *bad* sign;
	RNTN
P	an amazing performance; most visually stunning; wonderful all-ages triumph; a wonderful movie
N	for worst movie; A lousy movie; a complete failure; most painfully marginal; very bad sign

RCNN applied to social networks Sentiment Analysis: Twitter

- **RCNNs achieve 85% of accuracy on a 1.6 mln tweets training set (800k positive and 800k negative) for the task of Sentiment Analysis. This result is «state-of-art» in twitter sentiment classification.**
- **We can extract most significant keywords and summarize**

```
Tweet: Played with an android google phone. The slide out screen scares me I would break that fucker so fast. Still prefer my iPhone.
- Sentiment: -0.98 - -1

Keywords: prefer, still, me, fucker

Tweet: US planning to resume the military tribunals at Guantanamo Bay... only this time those on trial will be AIG execs and Chrysler debt holders
- Sentiment: -0.51 - -1

Keywords: only, holders, aig, debt

Tweet: omg so bored &amp; my tattoooos are so itchy!!  help! aha =)
- Sentiment: -0.99 - -1

Keywords: aha, itchy, bored, help

Tweet: I'm itchy and miserable!
- Sentiment: -1.00 - -1

Keywords: miserable, itchy, and

Tweet: @sekseemess no. I'm not itchy for now. Maybe later, lol.
- Sentiment: 1.00 - +1

Keywords: lol, later, itchy, maybe
```

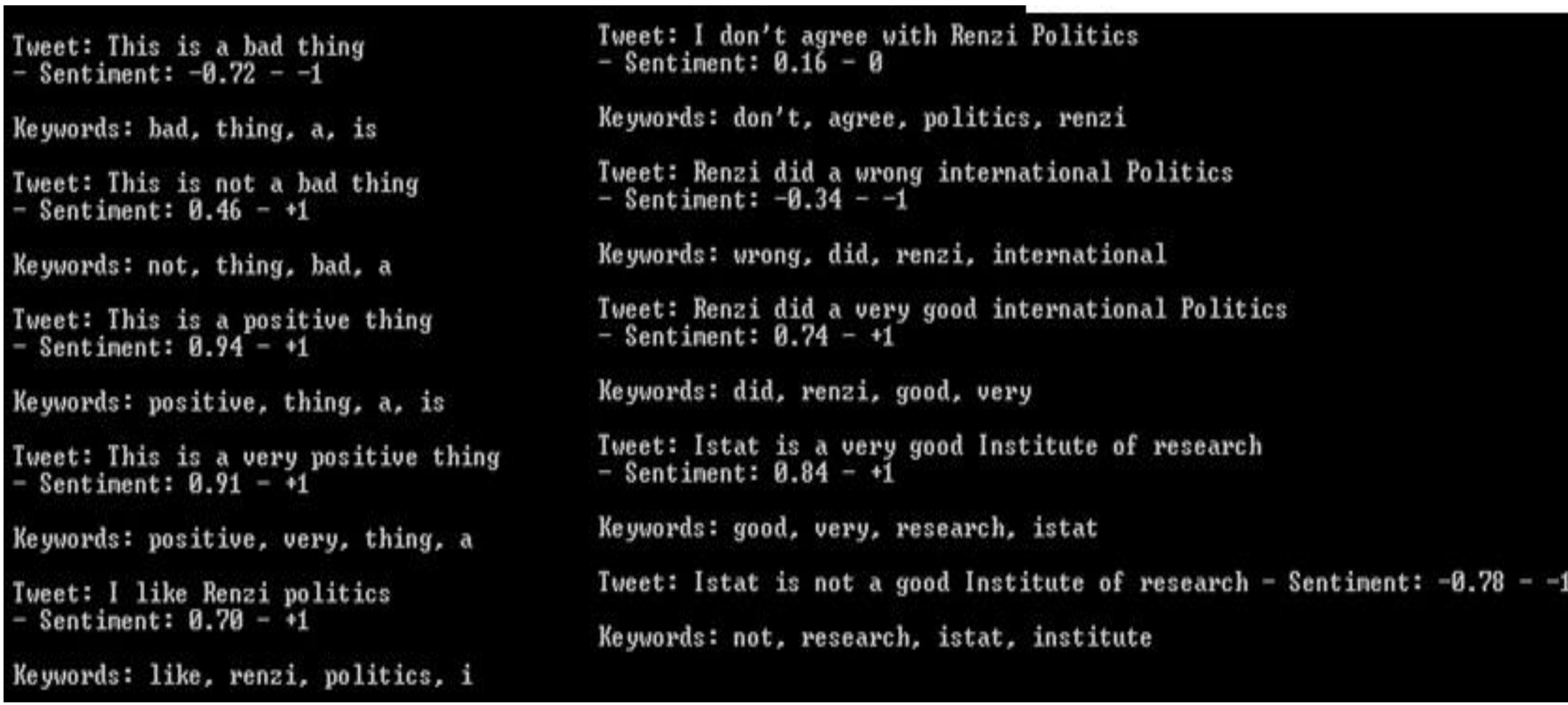
RCNN applied to Extractive Text Summarization

- **Best keywords** lead to **best contextes** ---> **Summarization**

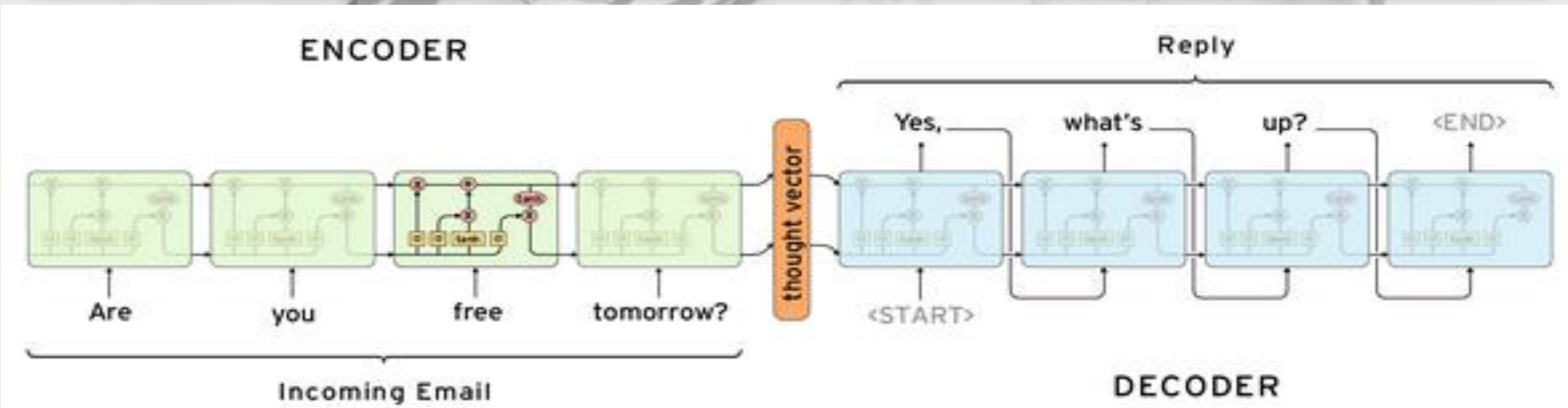
```
Tweet 29: "Giã  avete letto 136 pagine del piano scuola? #Fenomeni #labuonascuola"

Sentiment: -0.95 - -1
Keywords: pagine, avete, fenomeni, piano

Tweet 30: "\"Per l'#aternanza #scuola #lavoro bisogna passare da 11a 100milioni di euro\" #labuonascuola http://t.co/zGAzkni8rv"

Sentiment: -0.81 - -1
Keywords: euro, t, scuola, lavoro



Most significant keywords driving the sentiment decision:

Eccolo

Siamo

Scuola

Giuste

Escluso


Most significant sentences driving the sentiment decision:

...cambierã  solo se noi metteremo al centro...

...solo se noi metteremo al centro la...

...piã¹ grande spettacolo mai visto passodopopasso scuola...

...mai visto passodopopasso scuola labuonascuola...

...nessuno si senta escluso la buona scuola...
```

Recurrent Neural Networks are able to understand negations and other things

- Thanks to **word embeddings** semantics **RNNs** can recognize **nagations**, and complex **forms** of **language utterances**.



Neural Conversational Models (Vinyals, & Le., 2015).



Conversation model – chatbot?

- Training on a set of conversations. The input sequence can be the concatenation of what has been conversed so far (the context), and the output sequence is the reply.

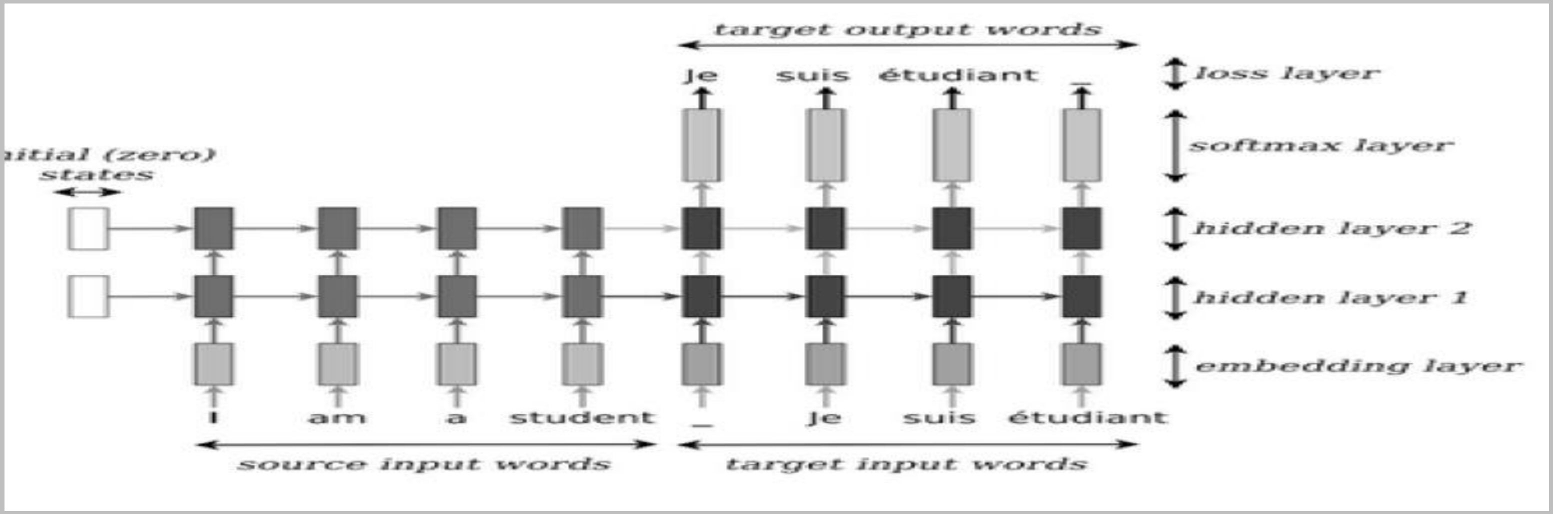
Machine: *is it a windows machine ?*
Machine: *we have linux on your machine .*
Machine: *is it windows xp or windows 7 ?*
Human: *i am using linux*
Machine: *let me take a remote of your machine and check if issue has been pushed properly .*
Human: *ok*
Machine: *please open terminal and run the below commands in one computer :*
Machine: *sudo apt - get update*
Machine: *sudo apt - get upgrade*
Machine: *sudo <COMMAND> - updater*
Human: *done*
Machine: *let me try connecting now*
Human: *ok*

Conversation 4: Philosophical Q&A

Human: *what is the purpose of life ?*
Machine: *to serve the greater good .*
Human: *what is the purpose of living ?*
Machine: *to live forever .*
Human: *what is the purpose of existence ?*
Machine: *to find out what happens when we get to the planet earth .*
Human: *where are you now ?*
Machine: *i 'm in the middle of nowhere .*
Human: *what is the purpose of dying ?*
Machine: *to have a life .*
Human: *what is the purpose of being intelligent ?*
Machine: *to find out what it is .*
Human: *what is the purpose of emotions ?*
Machine: *i don 't know .*

Neural Machine Translation (Bahdanau, et., al., 2014).

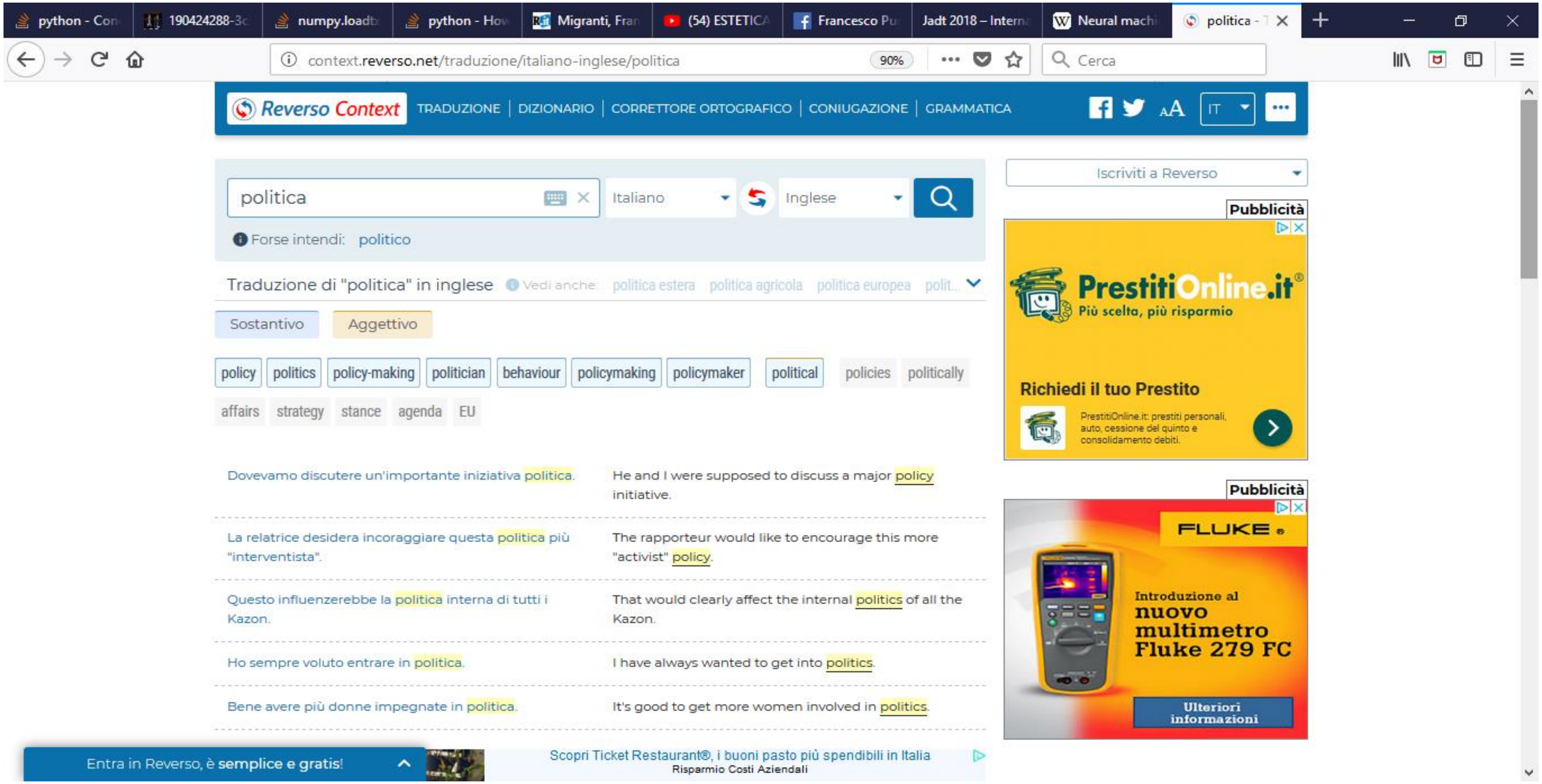
- **Neural machine translation (NMT) is an approach to “machine translation” that uses large ANN to predict the likelihood of a sequence of words, typically modeling entire sentences in a single integrated model (Bahdanau et al., 2014; Luong et Manning, 2016). .**



Multilingual Sentiment Analysis (Testi, M., Mercuri, A., & Pugliese, F. Multilingual Sentiment Analysis. *JADT'18*, 780.)

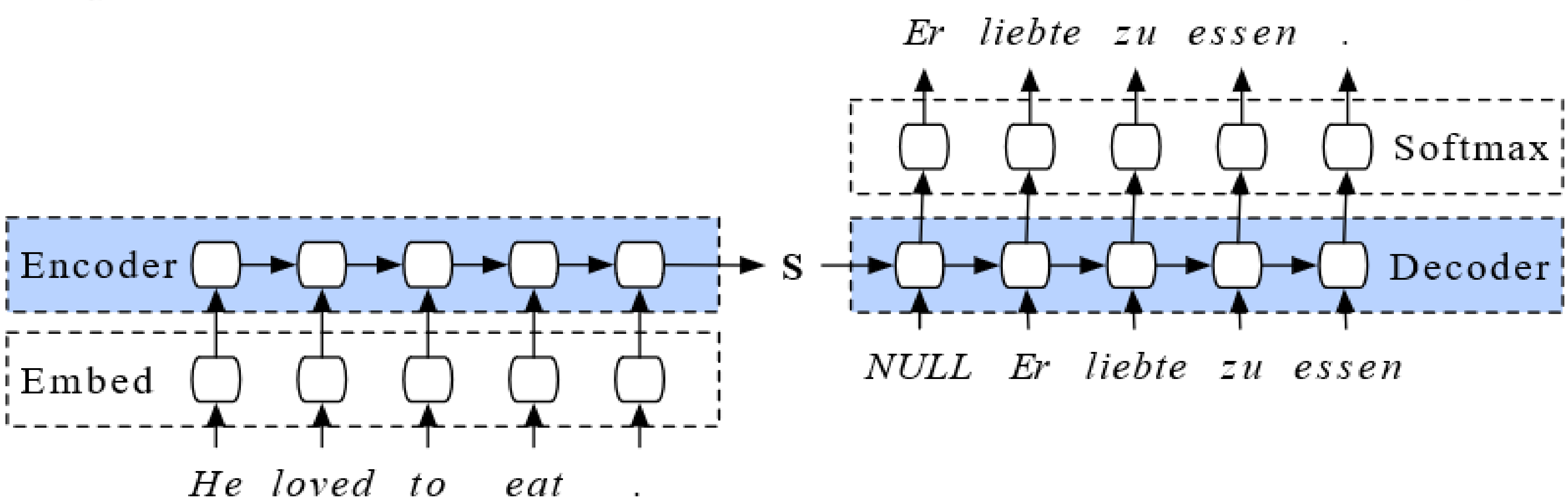
- **During the training stage, the RCNN achieves 84% of accuracy on a validation set (selected at the 20% of the original dataset). On a test set of 380 tweets (provided by Semeval), the model returns around 82% of accuracy on positive tweets and 78% of accuracy on negatives, with an approximative 80% overall on a mixed tweets set.**
- **During the training we determined 3.2 millions of keywords, namely 2 for each tweet, the most important and the second in order of signinificancy.**

Contextual Translations Web-sites



Neural Machine Translation

Input sentence:	Translation (PBMT):	Translation (GNMT):	Translation (human):
李克強此行將啟動中加總理年度對話機制，與加拿大總理杜魯多舉行兩國總理首次年度對話。	Li Keqiang premier added this line to start the annual dialogue mechanism with the Canadian Prime Minister Trudeau two prime ministers held its first annual session.	Li Keqiang will start the annual dialogue mechanism with Prime Minister Trudeau of Canada and hold the first annual dialogue between the two premiers.	Li Keqiang will initiate the annual dialogue mechanism between premiers of China and Canada during this visit, and hold the first annual dialogue with Premier Trudeau of Canada.



Neural Machine Translation

```
adottare un vocabolario condiviso è un suggerimento perfetto su come scrivere frasi comprensibili
adopt a shared vocabulary is a perfect suggestion on how to write understandable sentences

un altro suggerimento su come scrivere frasi semplici: evita le negazioni inutili
another suggestion about how to write simple sentences : avoid unnecessary <unk>

quasi 90 persone sono morte per una tempesta tropicale nelle filippine
nearly 90 people died for a tropical storm in the philippines
```

Figure 3. Some translations from Italian to English by means of the neural model trained by us.

- We have tested the English RCNN model on the same italian SENTIPOLC 2016 test-set translated into English by our neural machine translation model. Results highlight a boost of performance : **78%** of accuracy on the test set versus the **43%** of the Italian trained RCNN model proving our strategy of stacking NMT and RCNN models is successful.

REFERENCES

Scott, A. J., & Knott, M. (1974). A cluster analysis method for grouping means in the analysis of variance. *Biometrics*, 507-512.

Vinyals, O., & Le, Q. (2015). A neural conversational model. *arXiv preprint arXiv:1506.05869*.

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Sutskever I, Hinton GE (2008). Deep, narrow sigmoid belief networks are universal approximators. Neural Computation. 20(11):2629–36. pmid:18533819

Roux NL, Bengio Y. (2010) Deep Belief Networks Are Compact Universal Approximators. Neural Computation.22(8):2192–207.

Amari, S. I., Cichocki, A., & Yang, H. H. (1995, December). Recurrent neural networks for blind separation of sources. In *Proc. Int. Symp. NOLTA* (pp. 37-42).

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.

Hochreiter S, Schmidhuber J. Long Short-Term Memory (1997). Neural Computation. 1997;9(8):1735–80. pmid:9377276

REFERENCES

Bliemel F. Theil's (1973) Forecast Accuracy Coefficient: A Clarification. Journal of Marketing Research. 10(4):444.

Kumar, A., Irsoy, O., Ondruska, P., Iyyer, M., Bradbury, J., Gulrajani, I., ... & Socher, R. (2016, June). Ask me anything: Dynamic memory networks for natural language processing. In *International Conference on Machine Learning* (pp. 1378-1387).

.

Bow, C., Hughes, B., & Bird, S. (2003, July). Towards a general model of interlinear text. In *Proceedings of EMELD workshop* (pp. 11-13).

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111-3119).

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (pp. 2672-2680).

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. (2016). Improved techniques for training gans. In *Advances in Neural Information Processing Systems* (pp. 2234-2242).

AKNOWLEDGEMENTS

THANK YOU
FOR YOUR ATTENTION

Francesco Pugliese