Predicting Controversial News Using Facebook Reactions

Angelo Basile
Faculty of ICT, Univ. of Malta
CLCG, Univ. of Groningen
Groningen, NL
a.basile@student.rug.nl

Tommaso Caselli
CLTL, VU Amsterdam
CLCG, Univ. of Groningen
Amsterdam/Groningen, NL
t.caselli@gmail.com

Malvina Nissim
CLCG, Univ. of Groningen
Groningen, NL
m.nissim@rug.nl

Abstract

English. Different events and their reception in different reader communities may give rise to controversy. We propose a distant supervised entropy-based model that uses Facebook reactions as proxies for predicting news controversy. We prove the validity of this approach by running within- and across-source experiments, where different news sources are conceived to approximately correspond to different reader communities. Contextually, we also present and share an automatically generated corpus for controversy prediction in Italian.

Italiano. *Diversi tipi di eventi e la loro percezione in diverse comunità di utenti/lettori possono dare vita a controversie. In questo lavoro proponiamo un modello basato su entropia e sviluppato secondo il paradigma della "distant supervision" per predire controversie sulle notizie usando le reazioni di Facebook come "proxy". La validità dell'approccio è dimostrata attraverso una serie di esperimenti usando dati provenienti dalla stessa fonte o da fonti diverse. Contestualmente, presentiamo anche un corpus generato automaticamente per la previsione delle controversie in italiano.*

1 Introduction and Background

The explosion of social media (e.g. Facebook, Twitter, Disqus, Reddit, Wikipedia, among others) and the increased interactions with readers-users that traditional newspapers embraced, have transformed the Web in a huge *agora*, where news are shared, opinions are exchanged, and debates arise. On many topics, such as climate change, abortion, vaccination, among others, people strongly disagree. Following the work by Timmermans et al. (2017), we call *controversies* situations where, even after lengthy interactions, opinions of the involved participants tend to remain unchanged and become more and more polarized towards extreme values.

Modeling and understanding controversies may be useful in many situations. Journalists and news agencies may pay additional attention in the framing of a certain news, government officials and policy makers may be more aware of the issues involved in specific laws, social media managers might be more careful, i.e. monitor controversial content, in order to avoid the spreading of hate speech, and the general public may benefit as well thanks to a reduction of the "filter bubble" effect (Pariser, 2011).

Recently, computational approaches on controversy detection have been developed with varying degrees of success (Awadallah et al., 2012; Borra et al., 2015; Dori-Hacohen and Allan, 2015; Lourentzou et al., 2015). Works in the areas of Sentiment Analysis (Zhou et al., 2013; Deng and Wiebe, 2015; Deng et al., 2013; Chambers et al., 2015; Russo et al., 2015), Emotion Detection (Strapparava and Mihalcea, 2007; Strapparava and Mihalcea, 2008; Russo et al., 2011; Pool and Nissim, 2016), and Stance Detection (Mohammad et al., 2016) are, on the other hand, only partially related, as they focus on predicting/classifying the content of a message with respect to specific categories, such as "positive", "negative", "neutral", or "joy", "sadness" (among others), or as "being in favour" or "being against". They may be seen as necessary but not sufficient tools for detecting/predicting controversy (Timmermans et al., 2017).

The main contribution of this work is two-fold: i.) we propose a distant supervised entropy-based

Table 1: Sample rows from the dataset showing how entropy varies in relation to the reactions.

ID	TEXT	LIKE	LOVE	ANGRY	HAHA	WOW	SAD	entropy
1.)	In volo sul Piemonte con biplano anni '30	32	0	0	0	0	0	0.0
2.)	Medico anti vaccini radiato	5700	216	220	36	42	22	0.5
3.)	Piacenza, abbattuto il cinghiale Agostino	125	7	34	33	5	78	1.9

model to predict controversial news; and ii.) we present and share an automatically created corpus to train and test models for controversy detection. At this stage of development, we focused only on Italian, although the methods are completely language independent and can be reproduced for any language for which news are available on Facebook. The remainder of the paper is structured as follows: Section 2 illustrates the methods used to collect the data and develop the entropy-based model. Section 3 reports on the experiments and results both in a within- and across-source setting. Finally, Section 4 draws conclusions and outlines future research. Data and code are made available at `https://anbasile.github.io/predictingcontroversy/`.

2 Data and Methodology

We used the Facebook Graph API[1] to download news headlines (including the `description` and `body` fields) from four major Italian newspapers. Of these, two are slightly politically biased (*Corriere della Sera* and *La Repubblica*, both centre/centre-left), two openly biased ones (*Il Manifesto*, left-wing, and *il Giornale*, right-wing), and one news agency (*ANSA*).

Together with each news, we also downloaded all users' reactions.[2] Facebook reactions can be used as a proxy for annotations (Pool and Nissim, 2016), allowing to train a model for predicting the degree of controversy associated to news. On the basis of the definition of controversy previously introduced, our working hypothesis is that if users' reactions fall in two or more emotion classes (not necessarily opposed in terms of "polarity") with high frequencies, the controversy of a news item is higher. Building on this, we assume that entropy can be explanatory in modelling news' controversy: the higher the entropy, the more controversial the news. To better clarify this aspect, consider the data in Table 1. Each sample is the text of a Facebook post, for which we report the reaction breakdown (including `LIKE`), and its overall entropy based on reaction counts. Users expressing different reactions suggest that a text is likely to be controversial as it is shown by the high values of the entropy, as illustrated in examples 2.) and 3.) vs. example 1.).

For each source, namely the newspaper pages mentioned at the beginning of this section, we downloaded a collection of posts which appeared between mid-April and early July 2017. Posts with less than 30 reactions in total were discarded. For each post, we collected: i.) the link to the full article on the source's website (a large majority of the posts include this); ii.) an excerpt of the article (the variable `text`); iii.) additional texts commenting the article, when available (the variable `descriptor`); iv.) the full list of users' reactions. Finally, for a portion of the posts (1024 out of 3595, i.e. 28,48%; column "# body" in Table 2) we downloaded the entire text of the article (the variable `body`).[3]. Table 2 provides an overview of the data collected, including, for each source, the number of Facebook posts, the number of tokens, the number of posts for which the full article was retrieved, the token-post ratio, i.e. the number of tokens per post, and, finally, the average entropy.

Table 2: Dataset shape and average entropy score (avg H) per source.

SOURCE	# POSTS	# TOKEN	# BODY	RATIO TOKEN-POST	AVG H
AgenziaANSA	883	18,635	528	21.10	1.0216
corrieredellasera	594	23,811	124	40.08	0.9135
ilgiornale	1,022	8,665	124	8.47	1.1266
ilmanifesto	752	36,479	124	48.5	0.6195
repubblica	344	7,763	124	22.56	0.9078
total	3,595	95,353	1,024	26.52	0.9386

To further verify the soundness of using entropy as an indicator of controversy, we inspected the top-10 and bottom-10 news in the full dataset

[1] `https://developers.facebook.com/docs/graph-api`

[2] Since February 2016, Facebook users can react to a post not only with a like but by choosing from a set of 5 different emotions: `ANGRY, LIKE, HAHA, WOW, SAD, LOVE`.

[3] The full text of the article is not always available or accessible. Furthermore, there is a monthly limit to the data that can be downloaded. We made sure that the final dataset we used contained, for each source, the same number of posts for which the full body could be downloaded. This constraint did not apply to *ANSA*

Table 3: Sample of entropy-ranked top-5 and bottom-5 posts.

	TOPIC	TEXT
TOP	Incident	Fuggono dall'aereo in fiamme ma si fermano per scattare un selfie a pochi metri dall'aereo
	25th April	#25Aprile #Anpi: ""Festa di tutti gli italiani"". Roma divisa, due celebrazioni
	Gender/LGBTQ	"Genere: Sconosciuto". E il Canada gli dà' ragione
	Immigration	Emergenza #migranti, nave Rio Segura arrivata a Salerno. A bordo 11 donne incinte, 256 minori e 13 neonati #FOTO
	Animals	#Piacenza, abbattuto il cinghiale #Agostino. Da giorni nel parco urbano di Galleana, avrebbe caricato il personale
BOTTOM	25th April	#25aprile, ecco i musei statali aperti'
	Movies	"La La Land" meritava la statuetta del miglior film, andata poi a "Moonlight"?'
	Sport	Il Presidente della Sampdoria Massimo Ferrero è raggiante per la vittoria nel derby di Genova'
	Arts	Quando Eugenio Corti morì, il 4 febbraio 2014, Sébastien Lapaque, sul quotidiano parigino Le Figaro, lo definì "uno degli immensi scrittori del nostro tempo"'
	Arts	New York New York ricostruisce i legami artistici dal '28 a metà anni '60"

sorted by entropy (high values on top, high controversy) and manually assigned them to a topic. Table 3 illustrates the results for the top 5 and bottom 5 posts, in terms of entropy score. In addition to identifying a different distribution of topics according to degrees of controversy, we also observed that in some cases, the entities and the specific event mentions interact to generate controversy. For instance, in the case of the "*25th April*" topic[4], the controversial news involves a political actor (i.e. *ANPI*, the National Association of Italian Partisans), and divisions on the celebration of this day, while the non-controversial news reports on museums being open on that day. The entropy score appears to capture this distinction.

3 Experiments

We use the *ANSA* dataset to develop our model. The rationale behind this is that, being *ANSA* a news agency, the texts should be more objective and the controversy should depend on the event itself rather than by its framing in a specific, potentially biased, community. We treat this task as a regression problem, and use mean squared error (MSE) to measure the performance of our system. As baseline, we use a dummy regressor which always predicts the mean entropy of the train dataset: considering that the values range between 0 and 2.9, with a standard deviation of 0.4, a system that always predicts the mean entropy is already performing reasonably well. Furthermore, this is in line with the average entropy values of each dataset, ranging from 0.6195 (Table 2, *Il Manifesto*) up to 1.1266 (Table 2, *Il Giornale*).

[4]April 25th is a national holiday in Italy to celebrate the end of World War II.

Settings We use two main settings. Firstly, the data for training and testing the model originates from the same Facebook page, and we use cross-validation. Secondly, we train and test across pages, so as to investigate the model's portability across potentially different communities. This second setting can shed light on the issue of *perspective bias*, as controversy around a specific topic or entity could exist in one domain (or, in this case, in one community as proxied by Facebook pages) and not in another one. In both settings, we run our best model, developed as described below.

Features For predicting the entropy of the reactions to a given text, we built a system using a sparse feature representation and an SVM regressor, with the *scikit-learn* LinearSVR implementation (Buitinck et al., 2013). We used a tf-idf vectorizer to represent the text as both word and character n-grams.

As sentiment might contribute to controversy prediction (Dori-Hacohen and Allan, 2015), we also extended the features with coarse-grained prior polarity information derived from Sentix (Basile and Nissim, 2013), a resource for Italian automatically mapped from the English SentiWordNet (Esuli and Sebastiani, 2006). We represent each token with the absolute values of its polarity (which in Sentix ranges from -1 to +1). This allows us to ignore the specific positive/negative values, and get a more abstract representation on the subjectivity relevance of a token: high values indicate that the text is rich of subjectivity relevant tokens; 0 means that the text is merely objective. For each post, we then compute the average polarity and encoded it into a separate vector. Missing words in the lexicon are simply skipped.

Model development For development, as mentioned, we only used *ANSA*. We experimented with different features and different sizes of texts. In particular, we ran experiments using: i.) only the `text` variable; ii.) a combination of the `text` and the `descriptor` variables; and iii.) a combination of the `text`, the `descriptor`, and the `body` variables. Furthermore, these three basic settings have been extended with the polarity values from Sentix. To fine tune the parameters, a grid-search of the model using a 10-fold cross-validation was conducted. Table 4 reports the results of the different models as well as of the baselines.

Table 4: Results for the cross-validated *ANSA* dataset.

DATA	BASELINE	MODEL	+ SENTIX
text	0.24	0.154	0.155
text+descriptor	0.24	0.146	0.148
text+descriptor+body	0.24	0.146	0.148

The best model shows an improvement of 0.094 MSE with respect to the baseline when extending the variable `text` with `descriptor` and `body`. The use of the variable `text` alone still beats the baseline, but obtains a lower score than the models which include both the `descriptor` and the `body` variables. The extensions with the polarity scores from Sentix decrease the model performances (though still outperforming the baselines). We believe that this behaviour is mainly due to noise in the resource itself and calls for better and more context-oriented sentiment lexicons in Italian. Table 5 summarises the features of the best model, which is based on a combination of the three text variables only: `text`, `descriptor`, and `body` (whenever available), represented as word and character n-grams, ignoring the polarity vectors. This model was used on the reminder of the datasets.

Results on the test set Table 6 illustrates cross-validated results for the newspaper datasets. For comparison and completeness, we report also the results of the cross-validation on the full test set, with and without the extension of the data with *ANSA*.

With the exception of *Il Giornale*, our model always beats the baseline, confirming the validity of the designed approach. Extending the newspaper dataset with the data from *ANSA*, we can observe a reinforcement of the predicting power of the model, with a range between 0.04 to 0.1 points with respect to the corresponding baselines. The positive effect on *Il Giornale* dataset can be due to an extension of the number of tokens, since *Il Giornale* is the dataset with the lowest token-post ration (8,47 tokens per post), which clearly affects our model.

Table 5: Best model's settings and features.

PARAMETER	VALUE
SVR C	10
character ngrams	(2,3)
character binary features	True
character normalization	l2
character sublinear tf	False
word ngrams	(1,3)
word binary features	False
word normalization	l2
word sublinear tf	True

Table 6: Cross-validated results on all datasets.

	BASELINE	STD	MODEL	STD
ilgiornale	**0.21**	0.03	0.22	0.04
ilgiornale+ansa	0.23	0.04	**0.19**	0.03
ilmanifesto	0.15	0.04	**0.11**	0.04
ilmanifesto+ansa	0.24	0.04	**0.14**	0.03
repubblica	0.22	0.07	**0.18**	0.07
repubblica+ansa	0.24	0.04	**0.15**	0.04
corrieredellasera	0.24	0.06	**0.16**	0.06
corrieredellasera+ansa	0.24	0.03	**0.14**	0.04
full_dataset	0.24	0.02	**0.17**	0.03
full_dataset-ansa	0.24	0.03	**0.17**	0.04

Cross-source results in Table 7 are less clear-cut. In these experiments, it clearly emerges that our model works in the large majority of cases, although with no big gains over the baselines. All datasets fail to beat the baseline when predicting controversy on *Il Giornale* and, on the contrary, training on *Il Giornale* only fails to beat the baseline when testing on *La Repubblica*. This suggests that either there must be a difference in the wording used by *Il Giornale* with respect to the other datasets, or that the controversy is affected by perspective bias associated to different communities.

On the other hand, slightly politically oriented newspapers (*La Repubblica* and *Il Corriere della Sera*) and the *ANSA* news agency tend to have a homogeneous behavior, being able to correctly predict controversy in highly politically oriented

news (see results for *Il Manifesto* in Table 7). As a matter of fact, the more the post/token ratio is similar between different sources, the better the model works in predicting controversy. For instance, *Il Corriere della Sera* and *Il Manifesto* have a very similar post/token ratio (40,08 and 48,5, respectively) and not surprisingly both cross-source experiments beat the baseline.

Table 7: Cross-source results on all datasets.

TRAIN	TEST	BASELINE	MODEL
ilgiornale	ilmanifesto	0.40	**0.36**
ilgiornale	AgenziaANSA	0.25	**0.24**
ilgiornale	repubblica	**0.26**	0.29
ilgiornale	corrieredellasera	0.28	**0.26**
ilmanifesto	ilgiornale	**0.46**	0.46
ilmanifesto	AgenziaANSA	**0.40**	0.40
ilmanifesto	repubblica	0.30	**0.28**
ilmanifesto	corrieredellasera	0.32	**0.29**
AgenziaANSA	ilgiornale	**0.22**	0.23
AgenziaANSA	ilmanifesto	**0.31**	0.38
AgenziaANSA	repubblica	0.23	**0.21**
AgenziaANSA	corrieredellasera	0.25	**0.23**
repubblica	ilgiornale	**0.25**	0.28
repubblica	ilmanifesto	0.23	**0.23**
repubblica	AgenziaANSA	0.25	**0.23**
repubblica	corrieredellasera	0.23	**0.20**
corrieredellasera	ilgiornale	**0.25**	0.25
corrieredellasera	ilmanifesto	0.23	**0.20**
corrieredellasera	AgenziaANSA	0.25	**0.21**
corrieredellasera	repubblica	0.21	**0.18**

4 Conclusions and Future Work

This paper presents a simple regression model to predict the entropy of a post's reactions based on the Facebook reaction feature. We take this measure as a proxy to predict the *controversy* of news, where the higher the entropy (indicated by highly mixed reactions), the bigger the controversy. We run experiments both within and across communities, exemplified by the Facebook pages of specific newspapers. As a by-product, we have also automatically generated a first reference corpus for controversy prediction in Italian.

The results are promising, given that our model beats the baseline in almost all cases in cross-validation of same source data (see Table 6), and in the large majority of cases when applied cross-sources (see Table 7). At this stage of development, we observed that coarse-grained sentiment values are not useful, although this may depend on the quality of the lexicon employed. Test and training on openly biased datasets (e.g. *Il Giornale*$_{TRAIN}$ - *Il Manifesto*$_{TEST}$, and vice-versa) results in the lowest entropy, suggesting perspective bias in the different community.

The approach we have developed is based on discrete linguistically motivated features. This has an impact in the learned model as it is not able to generalise enough when dealing with low-frequency features and unseen data in the test set. To alleviate this issue, we are planning to model the post representations by using word embeddings.

We are planning to expand the model to account for perspective bias in different communities. News from different sources may be aggregated per event type, for example via the EventRegistry API[5], allowing to explore entropy (and polarisation of reactions) on exactly the same event instance. A first step in this direction would be to detect and match Named Entities to approximately identify similar events. At the reaction-level, the obvious next step is to explore and experiment with *clusters* of reactions (for instance, positive (`LIKE`, `LOVE`, `AHAH`), negative (`ANGRY`, `SAD`), or ambiguous (`WOW`)), instead of treating them all as single and distinct indicators.

Another follow-up is to extend this work to other social media data, such as Twitter. Twitter does not allow for nuances in reactions in the same way that Facebook does, as only one kind of "like" is provided. However, the substantial use of hashtags and emojis might offer alternative proxies to capture a variety of reactions. There is plenty of work on the usefulness of leveraging hashtags as reaction proxies both at a coarse and finer level (Mohammad and Kiritchenko, 2015), but this information, to the best of our knowledge, has not been used to predict likelihood of controversy.

Acknowledgments

One of the authors wants to thank the Spinoza-NWO Project "Understanding Language by Machines" subtrack 3 for making this work possible.

References

Rawia Awadallah, Maya Ramanath, and Gerhard Weikum. 2012. Opinions network for politically controversial topics. In *Proceedings of the first edition workshop on Politics, elections and data*, pages 15–22. ACM.

[5] `http://eventregistry.org`

Valerio Basile and Malvina Nissim. 2013. Sentiment analysis on italian tweets. In *WASSA@ NAACL-HLT*, pages 100–107.

Erik Borra, Esther Weltevrede, Paolo Ciuccarelli, Andreas Kaltenbrunner, David Laniado, Giovanni Magni, Michele Mauri, Richard Rogers, and Tommaso Venturini. 2015. Societal controversies in wikipedia articles. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pages 193–196. ACM.

Lars Buitinck, Gilles Louppe, Mathieu Blondel, Fabian Pedregosa, Andreas Mueller, Olivier Grisel, Vlad Niculae, Peter Prettenhofer, Alexandre Gramfort, Jaques Grobler, Robert Layton, Jake VanderPlas, Arnaud Joly, Brian Holt, and Gaël Varoquaux. 2013. API design for machine learning software: experiences from the scikit-learn project. In *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pages 108–122.

Nathanael Chambers, Victor Bowen, Ethan Genco, Xisen Tian, Eric Young, Ganesh Harihara, and Eugene Yang. 2015. Identifying political sentiment between nation states with social media. In *EMNLP*, pages 65–75.

Lingjia Deng and Janyce Wiebe. 2015. Joint prediction for entity/event-level sentiment analysis using probabilistic soft logic models. In *EMNLP*, pages 179–189.

Lingjia Deng, Yoonjung Choi, and Janyce Wiebe. 2013. Benefactive/malefactive event and writer attitude annotation. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 120–125, Sofia, Bulgaria, August. Association for Computational Linguistics.

Shiri Dori-Hacohen and James Allan. 2015. Automated controversy detection on the web. In *European Conference on Information Retrieval*, pages 423–434. Springer.

Andrea Esuli and Fabrizio Sebastiani. 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC'06*, pages 417–422.

Hyunseo Hwang, Youngju Kim, and Catherine U Huh. 2014. Seeing is believing: Effects of uncivil online debate on political polarization and expectations of deliberation. *Journal of Broadcasting & Electronic Media*, 58(4):621–633.

Ismini Lourentzou, Graham Dyer, Abhishek Sharma, and ChengXiang Zhai. 2015. Hotspots of news articles: Joint mining of news text & social media to discover controversial points in news. In *Big Data (Big Data), 2015 IEEE International Conference on*, pages 2948–2950. IEEE.

Saif M Mohammad and Svetlana Kiritchenko. 2015. Using hashtags to capture fine emotion categories from tweets. *Computational Intelligence*, 31(2):301–326.

Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. Semeval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 31–41, San Diego, California, June. Association for Computational Linguistics.

Eli Pariser. 2011. *The filter bubble: What the Internet is hiding from you*. Penguin UK.

Chris Pool and Malvina Nissim. 2016. Distant supervision for emotion detection using facebook reactions. In *Proceedings of the Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media (PEOPLES)*, pages 30–39, Osaka, Japan, December. The COLING 2016 Organizing Committee.

Irene Russo, Tommaso Caselli, Francesco Rubino, Ester Boldrini, and Patricio Martínez-Barco. 2011. Emocause: an easy-adaptable approach to emotion cause contexts. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, pages 153–160. Association for Computational Linguistics.

Irene Russo, Tommaso Caselli, and Carlo Strapparava. 2015. Semeval-2015 task 9: Clipeval implicit polarity of events. In *SemEval@ NAACL-HLT*, pages 443–450.

Carlo Strapparava and Rada Mihalcea. 2007. Semeval-2007 task 14: Affective text. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 70–74. Association for Computational Linguistics.

Carlo Strapparava and Rada Mihalcea. 2008. Learning to identify emotions in text. In *Proceedings of the 2008 ACM symposium on Applied computing*, pages 1556–1560. ACM.

Benjamin Timmermans, Lora Aroyo, Evangelos Kanoulas Tobias Kuhn, Kaspar Beelen, and Gerben van Eerten Bob van de Velde. 2017. Controcurator: Understanding controversy using collective intelligence. In *Collective Intelligence 2017*.

Xujuan Zhou, Xiaohui Tao, Jianming Yong, and Zhenyu Yang. 2013. Sentiment analysis on tweets for social events. In *Computer Supported Cooperative Work in Design (CSCWD), 2013 IEEE 17th International Conference on*, pages 557–562. IEEE.