# A Generic Opinion-Fact Classifier with Application in Understanding Opinionatedness in Various News Sections

Ankan Mullick[1], Shivam Maheshwari[2], Soumya C[3], Pawan Goyal[4], Niloy Ganguly[5]
Department of Computer Science and Engineering, IIT Kharagpur, India
{[1]ankan,[4]pawang,[5]niloy}@cse.iitkgp.ernet.in,
{[2]shiv.maheshwari1994,[3]soumya.cbr}@iitkgp.ac.in

## ABSTRACT

In this paper, we build a generic opinion-fact classifier to detect opinions and facts from online news articles and social media datasets such as Youtube comments and idiom hashtags. We further use this classification model to compare opinionatedness of various news article sections. The proposed classifier produces better results than the existing methods over four different datasets, and the opinion fraction of various sections of news articles provides very interesting patterns.

## CCS Concepts

•**Information systems** → **Information Extraction;** *Retrieval effectiveness;*

## Keywords

Opinion detection, Text classification

## 1. INTRODUCTION

Advertisement based revenue is the major goal of any online publishing house. Number of advertisements or value of a single advertise mostly depends on viewer count and number of people engaging in that publishing content through commenting or sharing or tweeting. So studying various topological characteristics and statistics of published content (article or video) for a certain time frame and how revenue changes depending on those factors, is very important. Comments or sentences in an article can either be fact (which can be proved true or false) or opinion (statements based on a belief or view on a fact). So opinionatedness or factuality of an article, different sections of news papers (e.g., sports or politics or business type etc.) and overall comments can be easily calculated for different time scales - weekly or monthly, and we can measure how opinion dynamics parameter for different time scales changes the revenue function. We can also examine effects of a sudden event in opinion dynamics.

Earlier works are mostly focused on opinion mining [12, 10] or predicting certainty of tweet [14] using keywords. Rajkumar et al. [11], Mullick et al. [9] worked on opinion extraction but the aim was to detect top 3 or 5 important and diverse opinions. [9] also built an automatic classifier to classify opinions into further subcategories. [2] analyzed how opinion and factual stories get shared across different news sharing media. We have used [9, 11] as baselines for opinion detection. But none of the work focused on identifying opinion dynamics of an article or different sections (e.g., sports, politics etc.) of distinct news papers and youtube comments. In this paper, we develop an automatic classifier to classify opinion-fact and calculate opinionatedness of news articles from different sections of 'The Guardian'.


WWW 2017, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4914-7/17/04.
http://dx.doi.org/10.1145/3041021.3054270

## 2. DATASETS

Experiments for opinion and fact classification were done on four different datasets (only in English Language): i). standard Multi-Perspective Question and Answering (MPQA), ii). Yahoo newspaper articles, and the remaining two are social media datasets - iii). Twitter hashtag idioms and iv). YouTube comments. The first two datasets contain 535 and 120 articles respectively, with the fraction of opinions being 0.486 and 0.527 respectively. Below, we discuss how the final two datasets were created.

1549 idiom hashtags were taken from the idiom dataset used by Maity et al. [7], later 2877 idioms were extracted from raw Twitter hashtags dataset using the algorithm proposed in [7] making a total dataset of 4426 idioms, and were labeled by 3 annotators as opinionated or factual. Among 4426 idioms, 2942 were labeled as opinions. Inter-annotator agreement Fleiss $\kappa$ is 0.77. For the final dataset, 50 Youtube video details were chosen randomly from Sikdar et al. [13] and total 1005 comments were labeled (651 opinions and 354 facts) by 2 annotators. Inter-annotator agreement Fleiss $\kappa$ is 0.72. Further, we crawled 5 categories of 'The Guardian' news article from July 2016 : 1. Business 2. Sports 2. Politics 4. Environment 5. Editorials. Statistics are in Table 2.

## 3. EXPERIMENTS

**Feature Identification:** Feature encoding was done in python with an initial set of 21 features for MPQA dataset and 26 features for other datasets[1]. Selected features can be classified into 3 broad groups - POS tag based, dependency parse based and other type. POS-tagged based features were generated by *Stanford POS-tagger* [8] for MPQA, Yahoo datasets and *CMU POS tagger*[5] for hashtag idioms and Youtube comments. *Stanford Dependency parser*[4] was used to find dependency parse based feature values in the expressions. Other than the above two, some intuition based features were used, like presence of wh-words, strong and weak adjectives, and words specific to any of the defined classes and subclasses.

**Classification:** After feature extraction and dataset balancing (using SMOTE[3]), Weka implementations[2] of various classifiers [6]: Naive Bayes (NB), Logistic Regression (LR), Support Vector Ma-

[1]Datasets and features are in goo.gl/5dCld1
[2]http://www.waikato.ac.nz/ml/weka/

Table 1: Comparison of 10-fold cross validation Precision (P), Recall (R), Accuracy (A), Area Under Curve (AUC) for classification of sentences into opinion and fact for idiom hashtags, Youtube Comments, MPQA and Yahoo articles

| Dataset | Hashtags Idioms | | | | Youtube Comments | | | | MPQA | | | | Yahoo | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Classifiers | P | R | A(%) | AUC | P | R | A(%) | AUC | P | R | A(%) | AUC | P | R | A | AUC |
| NB | 0.73 | 0.70 | 69.9 | 0.8 | 0.61 | 0.59 | 58.7 | 0.60 | 0.57 | 0.53 | 53.2 | 0.59 | 0.68 | 0.64 | 67.2 | 0.73 |
| LR | 0.76 | 0.76 | 76.1 | 0.83 | 0.61 | 0.61 | 60.6 | 0.66 | 0.59 | 0.58 | 58.4 | 0.57 | 0.66 | 0.63 | 65.7 | 0.69 |
| SVM | 0.75 | 0.74 | 75.6 | 0.69 | 0.63 | 0.63 | 62.7 | 0.63 | 0.60 | 0.60 | 59.9 | 0.57 | 0.70 | 0.69 | 69.8 | 0.67 |
| HITS | 0.71 | 0.68 | 68.5 | 0.67 | 0.61 | 0.62 | 59.2 | 0.62 | 0.60 | 0.61 | 60.6 | 0.59 | 0.71 | 0.66 | 70.1 | 0.73 |
| OP-D | 0.72 | 0.70 | 70.2 | 0.8 | 0.63 | 0.62 | 62.1 | 0.63 | 0.62 | 0.64 | 63.7 | 0.63 | 0.73 | 0.71 | 71.5 | 0.75 |
| RF | 0.8 | 0.81 | 82.1 | 0.89 | 0.74 | 0.73 | 73.8 | 0.81 | 0.58 | 0.59 | 59.1 | 0.61 | 0.69 | 0.70 | 70.9 | 0.74 |
| Bg+RF | **0.83** | **0.83** | **82.6** | **0.9** | **0.75** | **0.74** | **74.2** | **0.82** | **0.74** | **0.75** | **74.7** | **0.83** | **0.76** | **0.74** | **73.1** | **0.77** |

Table 2: Opinionatedness and Factuality of different sections of 'The Guardian' news article dataset.

| Dataset section | B | S | P | En | Ed |
|---|---|---|---|---|---|
| no. of article | 89 | 93 | 85 | 111 | 91 |
| no. of sentence | 2871 | 4173 | 3668 | 4184 | 4175 |
| Opinion (%) | 36.7 | 72.6 | 66.2 | 40.5 | 80.2 |
| Fact (%) | 63.3 | 27.4 | 33.8 | 59.5 | 19.8 |

chine (SVM), HITS [11], Opinion Diversity (Op-D) [9], Random Forest (RF), Bagging with Random Forest (Bg+RF) were used to calculate 10-fold cross validation results for opinion-fact classification in terms of precision (P), recall (R), accuracy (A), area under curve (AUC). Results are shown in Table 1.

It is clearly seen that *Bagging with Random Forest* provides the best results for all the datasets for opinion and fact classification. So we use Bagging with Random Forest classifier to calculate opinionatedness of various sections of 'The Guardian' - business (B), politics (P), sports (S), environment (En) and editorials (Ed) after training the model with MPQA news article corpus. The details on the size of these five sections along with the findings using the classifier are shown in Table 2. We see that while editorials have the largest fraction of opinions, which is also very intuitive, articles from sports and politics section are also quite opinionated. Business and environment sections are mostly factual.

To check as to how much we can trust these results, we take random articles from each of these sections (details of this new sample dataset is shown in Table 3) and annotate the sentences by 2 annotators (Inter annotator agreement using Fleiss $\kappa$ is 0.76) with opinion or fact label. The precision, recall, overall accuracy and area under curve for our classifier, when applied over these sections, is provided in Table 3, which is consistent with the earlier results.

Table 3: Precision (P), Recall (R), Accuracy (A), Area Under Curve (AUC) results on 5 sections of annotated 'The Guardian' dataset.

| Section of the Article | No. of article | No. of sentence | P | R | A(%) | AUC |
|---|---|---|---|---|---|---|
| Business | 16 | 316 | 0.83 | 0.71 | 76.1 | 0.95 |
| Sports | 13 | 179 | 0.82 | 0.80 | 80.1 | 0.89 |
| Politics | 10 | 208 | 0.86 | 0.75 | 88.3 | 0.96 |
| Environment | 18 | 314 | 0.82 | 0.89 | 74.9 | 0.92 |
| Editorials | 15 | 505 | 0.90 | 0.90 | 89.9 | 0.94 |

# 4. CONCLUSION

In this paper we built an opinion-fact classifier that outperforms the existing baselines over a variety of datasets and used the classifier to measure opinionatedness or factuality of various online news sections. Our immediate future step will be to calculate, for different timescale how advertisement based revenue and peoples' engagement (in terms of number of comments in news articles) correlate with this score for different sections of article and for different categories of opinions [1] - report, judgment, advise and sentiment.
**Acknowledgments:** Authors would like to acknowledge Surjodoy Ghosh Dastider (IIT Kharagpur) and Srotaswini Sahoo (NIT Rourkela) for cleaning, labeling datasets and helping in coding.

# 5. REFERENCES


[1] N. Asher, F. Benamara, and Y. Y. Mathieu. Appraisal of opinion expressions in discourse. *Lingvisticæ Investigationes*, 32(2):279–292, 2009.

[2] A. Chakraborty, S. Ghosh, N. Ganguly, and K. P. Gummadi. Dissemination biases of social media channels: On the topical coverage of socially shared news. In *ICWSM*, 2016.

[3] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 2002.

[4] M.-C. De Marneffe and C. D. Manning. Stanford typed dependencies manual. Technical report, 2008.

[5] K. Gimpel, N. Schneider, B. O'Connor, D. Das, D. Mills, J. Eisenstein, M. Heilman, D. Yogatama, J. Flanigan, and N. A. Smith. Part-of-speech tagging for twitter: Annotation, features, and experiments. pages 42–47. ACL, 2011.

[6] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.

[7] S. K. Maity, A. Gupta, P. Goyal, and A. Mukherjee. A stratified learning approach for predicting the popularity of twitter idioms. In *ICWSM*, 2015.

[8] C. D. Manning, M. Surdeanu, J. Bauer, J. R. Finkel, S. Bethard, and D. McClosky. The stanford corenlp natural language processing toolkit. In *ACL*, pages 55–60, 2014.

[9] A. Mullick, P. Goyal, and N. Ganguly. A graphical framework to detect and categorize diverse opinions from online news. *PEOPLES 2016*, page 40, 2016.

[10] A. Qadir. Detecting opinion sentences specific to product features in customer reviews using typed dependency relations. In *Workshop on Events in Emerging Text Types*, pages 38–43. ACL, 2009.

[11] P. Rajkumar, S. Desai, N. Ganguly, and P. Goyal. A novel two-stage framework for extracting opinionated sentences from news articles. *TextGraphs-9*, page 25, 2014.

[12] T. Scholz and S. Conrad. Opinion mining in newspaper articles by entropy-based word connections. *EMNLP*, 2013.

[13] S. Sikdar, A. Chaudhary, S. Kumar, N. Ganguly, A. Chakraborty, G. Kumar, A. Patil, and A. Mukherjee. Identifying and characterizing sleeping beauties on youtube. In *ACM CSCW*, pages 405–408, 2016.

[14] S. Soni, T. Mitra, E. Gilbert, and J. Eisenstein. Modeling factuality judgments in social media text. *ACL*, 2014.