Identifying Opinion and Fact Subcategories from the Social Web

Ankan Mullick
Bing
Microsoft, India
ankan.mullick@microsoft.com

Surjodoy Ghosh D
Dept. of CSE
IIT Kharagpur, India
ghosh.d.1302@iitkgp.ac.in

Shivam Maheswari
Dept. of CSE
IIT Kharagpur, India
shiv.maheshwari1994@iitkgp.ac.in

Srotaswini Sahoo
Dept. of CSE
NIT Rourkela, India
srotasahoo7@gmail.com

Suman Kalyan Maity
Dept. of CSE
IIT Kharagpur, India
sumankalyan.maity@cse.iitkgp.ernet.in

Soumya C
Dept. of CSE
IIT Kharagpur, India
soumya.cbr@iitkgp.ac.in

Pawan Goyal
Dept. of CSE
IIT Kharagpur, India
pawang@cse.iitkgp.ernet.in

GROUP '18, January 7–10, 2018, Sanibel Island, FL, USA
ACM 978-1-4503-5562-9/18/01.
https://doi.org/10.1145/3148330.3154518

Abstract
In this paper, we investigate the problem of building automatic classifiers to categorize opinions and facts into appropriate subcategories. While working on two English News article datasets and two social media datasets (Twitter hashtag idioms and Youtube comments), we achieve consistent performance with accuracies in the range of 70-85% for opinion and fact sub-categorization. The proposed classifiers can be instrumental in understanding argumentative relations as well as in developing fact-checking systems. It can also be used to detect anomalous behavior such as predominant drunkers or other psychological changes.

Author Keywords
Opinion Classification, Fact Classification, Opinion-Fact Diversity, Opinion-Fact Categorization

ACM Classification Keywords
H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval; H.m [Information System]: Miscellaneous

Introduction
In the online world, people post texts or pictures in social media and comment in online news articles to express their views on some events or different topics of news articles. We can broadly classify a sentence or phrase (tweet, comment etc.) into opinion (statements based on a belief or

Dataset (#)
We have worked on four different datasets: two classical datasets (a,b) and two social media datasets (c,d) - **(a) Multi perspective Question and Answering (MPQA):** We take categorically labeled opinionated sentences from MPQA articles[6]. It contains labeled 1237 opinions and 1232 facts (786 report, 200 knowledge, 179 belief, 51 doubt, 16 perception). **(b) Yahoo news articles:** From [6], we gathered categorically labeled opinionated sentences of yahoo articles. Dataset has 470 labeled opinions and 252 facts (160,18,46,15,13). Details of (a) and (b) in [6]. **(c) Twitter Hashtag Idioms:** We collected 2942 opinionated hashtag idioms from [8] and 1480 were categorically labeled (report, judgment, advise, sentiment) with inter annotator agreement Fleiss κ=0.78. **(d) Youtube Comments:** After collecting opinions from [8], we extended the dataset up to 1540 opinions and labeled them (Inter annotator agreement Fleiss κ=0.71).

view on a fact) or fact (which can be proved true or false). Consider the following sentences: (i) *Senate voted 55-43 to confirm Robert Wilkins to the U.S. Court of Appeals for the District of Columbia* and ii) *McGreevey's lover was being paid 11000 Dollar even though he was wildly unqualified for the position*. While the first sentence is a fact, the second one is an opinion. Interestingly, all the opinion sentences may not be similar. Asher et al. [1] provide four different sub-categories of opinions - report, judgment, advise, sentiment. Similarly, Soni et al. [13] manually classify tweets into 5 fact subcategories - report, knowledge, belief, doubt, perception. However, there are no generic classifiers for categorical classification of opinions and facts as of now. Understanding fine-grained opinion and fact subcategories can be instrumental in many applications including deriving various argumentative relations such as support / attack, as well as understanding if a given sentence is fact-checkworthy.

In the modern scenario, online publishers want to increase user engagements / comments on their news articles / channels by highlighting important sentences. However, there exist no such modeling of the revenue of a newspaper based on sentence types and subtypes over different time periods. Distribution as well as flow of opinion and fact sub-categories can help in modeling the revenue generation. It will also be quite informative to examine how different categories of opinions and facts vary demographically (sex, age, region etc.), for different time frames like days of week (weekdays vs weekends), monthly (start of the month vs end of the month) or hourly (morning vs work hours vs evening vs late night). Demographic patterns of opinion and fact categories can be different for psychogenic people, predominant drunkers and others scenarios than the normal people. For example, we can identify peoples' suicidal tendencies or change in behavior in near future by tracking social media so that we can control situations accordingly. In the field of rumors and fake news problems, it is important to study factuality or opinionatedness and how their different categorical distributions vary for rumors and fake news. At a micro level, one sentence or comment or tweet may contain multiple opinion and fact categories and separating these might be useful to derive actionable insights (Rudra et al. [10] use this for tweet summarization in disaster scenario.). In the above opinion example, "McGreevey's lover was being paid 11000 Dollar" is factual but "even though he was wildly unqualified for the position" is opinionated (judgment). Many research works focused on subjectivity or objectivity of sentence and analysis can be easier if we discover the interaction between various opinion-fact categories in the same / nearby sentences. Another important measurement is to check how sentiment analysis varies with different categories for different demographic features. Example: whether 'report' opinions are mostly neutral and 'judgment' opinions are polarized? Sentiment analysis for different subcategories of opinion categories is also interesting. To design intelligent chat-bot system, categorical classifications are important because it may help to identify mentalities of the person and each questions can be answered accordingly.

This paper takes a first step in this direction as we build two different classifiers - Bagging with Random Forest and Repeated Incremental and Pruning (Rip) for opinion and fact sub-categorizations respectively[1]. Our classifiers achieved high precision, recall, accuracy and ROC across various news and social media datasets.

Related Work

People have been working on opinion mining for the last two decades. Some works [5, 14, 15] have focused on opin-

[1] Due to unavailability of sufficiently labeled categories of opinions and facts, deep neural network produces poor results, and is not reported.

Feature Identification (#)
Our identified features can be broadly classified into three categories -
(i) POS Tag based features: We used *Stanford POS-tagger* for MPQA, Yahoo classical datasets and *CMU POS-tagger* [3] for hashtag idioms and youtube comments[a] - social media data to find no. of nouns, verb, adjectives etc, presence of adverbs etc.
(ii) Dependency parse based features (using Stanford Dependency parser): dobj (direct object), amod (adjective modifier), acomp (adjectival complement) etc.
(iii) Others - no. of characters, presence of wh-words, numbers, strong, weak adjectives, words specific to particular categories.
Different combinations of features are used for social media data (40 features) and classical data (45 features). Fact classification was done only on classical datasets[b].

[a]As CMU POS-tagger works better than Stanford POS-tagger in social media data (e.g. - tweet)[3].
[b]since the manual labeling of fact subcategories on the social media data had a very poor inter-annotator agreement.

ion mining, e.g., subjective vs. objective classification, separating facts from opinions, identifying opinion polarity, etc. Scholz and Conrad [12] extract entropy based word connections to identify word combinations, and analyze opinion tonality of news articles. Soni et al. [13] predicted factuality of tweet text, using keywords from [11].

Some graph based models have been built to identify opinions in a news article. Rajkumar et al. [9] and Mullick et al. [6] built HITS framework, modeling opinions as hub and supporting facts as authority to identify important opinions. [6] also identified top k diverse opinions, showed categorical distributions and classification of opinions into four [1] categories[2]. Mullick et al. [8] built a generic opinion-fact classifier based on classical and social media datasets and also presented how opinionatedness of various sections of news articles differs. None of the prior works have attempted building a generic classifier to identify opinion and fact subcategories that works on both news and social media. We built generic classifiers to classify opinions into four categories [1] - report, judgment, advise, sentiment, and facts into five categories [13] - report, knowledge, belief, doubt, perception.

Experiments
To handle the imbalance of datasets, we first use SMOTE [2] algorithm to make the dataset balanced (corresponding to the maximum count of the instance). After feature extraction[3], using these balanced datasets, various classifiers from Weka [4] - Naïve Bayes (NB), Logistic Regression (LR), Support Vector Machine (SVM), Repeated Incremental and Pruning (JRip, baseline model used in [6]), Logistic Boost, IBK (Instance based learning with parameter k),

[2][6] built a classifier for opinion sub-categorization only on classical news datasets, we use this as a baseline.
[3]Details of the features are in https://goo.gl/U1HrN3.

Table 1: Comparison of 10-fold cross validation testsn: Precision (P), Recall (R), Accuracy (A), Receiver Operating Characteristic (ROC) for classification of opinions into subcategories (report, judgment, advise, sentiment) for MPQA and Yahoo articles

Dataset	MPQA				Yahoo			
Classifiers	P	R	A(%)	AUC	P	R	A	AUC
NB	0.36	0.35	34.9	0.64	0.43	0.49	49.5	0.78
LR	0.43	0.43	34.9	0.68	0.58	0.47	44.5	0.69
SVM	0.61	0.61	61.7	0.74	0.56	0.57	56.6	0.71
JRip	0.74	0.71	70.1	0.73	0.63	0.70	70.1	0.65
IBK	0.79	0.79	79.2	0.86	0.68	0.69	69.1	0.80
RF	0.87	0.83	83.7	0.91	0.74	0.73	73.1	0.90
Bg+RF	**0.89**	**0.84**	**84.3**	**0.96**	**0.75**	**0.74**	**73.8**	**0.91**

Random Forest (RF), Bagging with RF (Bg+RF) were used to classify opinions and facts into their respective subcategories. The performance is measured in terms of Precision (P), Recall (R), Accuracy (A) and Receiver Operating Characteristic (ROC). Results for opinion sub-categorization are shown in Tables 1 and 2. It is clearly seen that Bagging with Random Forest (Bg+RF) produces best Precision, Recall, Accuracy and ROC for opinion sub-classification.

Table 2: Comparison of 10-fold cross validation testn: Precision (P), Recall (R), Accuracy (A), Receiver Operating Characteristic (ROC) for classification of opinions into categories (report, judgment, advise, sentiment) for idiom hashtags and Youtube comments

Dataset	Hashtags Idioms				Youtube Comments			
Classifiers	P	R	A(%)	ROC	P	R	A	AUC
NB	0.49	0.48	45.7	0.73	0.39	0.35	34.7	0.63
LR	0.46	0.51	51.3	0.83	0.47	0.47	47.3	0.72
SVM	0.60	0.60	60.5	0.77	0.58	0.59	58.6	0.72
JRip	0.78	0.76	76.4	0.89	0.57	0.51	51.8	0.48
IBK	0.77	0.78	77.5	0.85	0.63	0.62	62.1	0.63
RF	0.82	0.81	81.7	0.93	0.74	0.73	73.8	0.81
Bg+RF	**0.83**	**0.82**	**81.9**	**0.95**	**0.75**	**0.75**	**74.4**	**0.91**

Table 3: Information Gain (IG) and One Attribute Evaluation (OAE) for MPQA Opinions

Feature	IG	OAE(%)
no. nouns	0.32	44
verb	0.28	41.3
det	0.27	40.6
nsubj	0.25	39.5
no. adj	0.23	39.3

Table 4: IG and OAE for Yahoo Opinions

Dataset	Yahoo	Yahoo
Feature	IG	OAE(%)
no. nouns	0.22	44
verb	0.22	41.3
no. adj	0.21	43.4
no. mark	0.24	39.9
no. det	0.18	40.1

Table 5: IG and OAE for Hashtag Idiom opinion

Feature	IG	OAE(%)
no. nouns	0.21	38.3
verb	0.21	38.3
str adj	0.18	35.5
no. char	0.16	40.1
word	0.15	37.1

Table 6: IG and OAE for Youtube comment (YC) opinions

Feature	IG	OAE(%)
nsubj	0.19	40.0
verb	0.21	39.7
dobj	0.20	40.2
pronoun	0.17	40.2
noun	0.16	39.0

For fact classification, we used several classifiers but only top seven classifiers have been shown – apart from NB, LR, Rip we used Multi-class Classifier (MCC), Logistic Iterative Boost (LIB), Bagging with Random Forest (Bg+RF) and Sequential Minimal Optimization (SMO).

Table 7: Comparison of 10-fold cross validation testn: Precision (P), Recall (R), Accuracy (A), Receiver Operating Characteristic (ROC) for classification of facts into categories (report, knowledge, belief, doubt,perception) for MPQA and Yahoo articles

Dataset	MPQA				Yahoo			
Classifiers	P	R	A(%)	ROC	P	R	A	ROC
NB	0.54	0.59	58.7	0.64	0.55	0.50	50.4	0.59
LR	0.60	0.67	66.5	0.72	0.59	0.63	62.6	0.67
SMO	0.49	0.64	64.2	0.64	0.58	0.69	69.7	0.62
MCC	0.61	0.67	66.8	0.72	0.59	0.64	64.7	0.68
LIB	0.64	0.67	67.0	0.72	0.57	0.68	68.4	0.66
Bg+RF	0.63	0.67	66.7	0.71	0.56	0.67	66.5	0.67
Rip	**0.69**	**0.71**	**71.0**	**0.73**	**0.60**	**0.70**	**70.2**	**0.69**

For fact classification, we get the best results for Rip classifier but not for Bg+RF. Thus, we are getting two different classifiers for automatic classification of opinions and facts into categories. Tables 3, 4, 5 and 6 show Information gain and One attribute evaluation for top five features of different datasets in case of opinion classification. For fact classification, information gain and one attribute evaluation of different features for MPQA and Yahoo datasets are shown in Table 8. We see that while POS tag based features are very helpful, features from all three categories constitute the top five. For opinion classification, no. of nouns, adjectives, presence of verbs, nominal subject (nsubj) are important but for fact classification, presence of fact words, nouns, adverbs, clausal complement (ccomp) are important.

Conclusion

In this paper, we investigated the problem of opinion and fact categorical classification across several datasets. To the best of our knowledge, this is the first study which tries to classify facts in classical datasets and opinions in social media and classical datasets into various subcategories. Our proposed classification framework achieves good accuracy, precision, recall and ROC for various datasets. We can now use the proposed classifier to study as to how various kinds of opinions and facts are found across different datasets - (e.g. - opinionated social list hashtags [7]), and how these evolve over time. Our immediate future step is to examine various demographic distributions for different categories of opinion and fact and to be able to use this for various applications involving deep diving into argumentative relations (e.g., finding supporting / attacking / contrastive claims and opinions) or checking if a given sentence is fact checkworthy. Another future direction is to identify smaller units than a sentence by using different discourse markers (e.g., comma) and study how different opinion or fact subcategories in a sentence combine to give overall subjectivity or objectivity. Our aim is to build a generic system to identify categories over different datasets.

Table 8: Information Gain (IG) and One Attribute Evaluation (OAE) for MPQA and Yahoo fact classification

MPQA			Yahoo		
Feature	IG	OAE	Feature	IG	OAE
noun	0.24	42.1	fact words	0.25	45.5
fact words	0.23	39.2	length	0.21	40.1
ccomp	0.18	38.7	prep	0.17	37.2
adv	0.16	33.1	advcl	0.15	30.2
length	0.14	32.1	pro	0.11	30.0

Acknowledgments: We would like to thank Prof. Niloy Ganguly (CSE Dept., IIT Kharagpur) for valuable feedbacks and Microsoft IDC for travel support.

REFERENCES

1. Nicholas Asher, Farah Benamara, and Yvette Yannick Mathieu. 2009. Appraisal of opinion expressions in discourse. *Lingvisticæ Investigationes* 32, 2 (2009).
2. Nitesh V. Chawla and others. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research* (2002).
3. K Gimpel and others. 2011. Part-of-speech tagging for twitter: Annotation, features, and experiments. ACL.
4. Mark Hall, Eibe Frank, G Holmes, Bernhard Pfahringer, Peter R, and Ian H Witten. 2009. The WEKA data mining software: an update. *ACM SIGKDD* (2009).
5. Soo-Min Kim and Eduard Hovy. 2006. Extracting opinions, opinion holders, and topics expressed in online news media text. ACL, 1–8.
6. Ankan Mullick, Pawan Goyal, and Niloy Ganguly. 2016. A graphical framework to detect and categorize diverse opinions from online news. *PEOPLES 2016* (2016), 40.
7. Ankan Mullick, Pawan Goyal, Niloy Ganguly, and Manish Gupta. 2017a. Extracting Social Lists from Twitter. In *ASONAM*.
8. A Mullick, Shivam M, P Goyal, N Ganguly, and others. 2017b. A Generic Opinion-Fact Classifier with Application in Understanding Opinionatedness in Various News Section. In *WWW*. 827–828.
9. Pujari Rajkumar, Swara Desai, Niloy Ganguly, and Pawan Goyal. 2014. A Novel Two-stage Framework for Extracting Opinionated Sentences from News Articles. *TextGraphs-9* (2014), 25.
10. Koustav Rudra, S Ghosh, Niloy Ganguly, Pawan Goyal, and Saptarshi Ghosh. 2015. Extracting situational information from microblogs during disaster events: a classification-summarization approach. In *CIKM*. ACM.
11. Roser Saurí. 2008. *A factuality profiler for eventualities in text*. ProQuest.
12. Thomas Scholz and Stefan Conrad. 2013. Opinion Mining in Newspaper Articles by Entropy-Based Word Connections. *EMNLP* (2013).
13. Sandeep Soni, Tanushree Mitra, Eric Gilbert, and Jacob Eisenstein. 2014. Modeling factuality judgments in social media text. *ACL* (2014).
14. Janyce Wiebe and Ellen Riloff. 2005. Creating subjective and objective sentence classifiers from unannotated texts. In *CICLingLing*. Springer, 486–497.
15. Hong Yu and Vasileios H. 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *EMNLP*.