"8 Amazing Secrets for Getting More Clicks": Detecting Clickbaits in News Streams Using Article Informality

Prakhar Biyani, Kostas Tsioutsiouliklis, and John Blackmer
Yahoo Labs, Sunnyvale, California, USA
Email: {pxb5080, kostas, johnbl}@yahoo-inc.com

Abstract

Clickbaits are articles with misleading titles, exaggerating the content on the landing page. Their goal is to entice users to click on the title in order to monetize the landing page. The content on the landing page is usually of low quality. Their presence in user homepage stream of news aggregator sites (e.g., Yahoo news, Google news) may adversely impact user experience. Hence, it is important to identify and demote or block them on homepages. In this paper, we present a machine-learning model to detect clickbaits. We use a variety of features and show that the degree of informality of a webpage (as measured by different metrics) is a strong indicator of it being a clickbait. We conduct extensive experiments to evaluate our approach and analyze properties of clickbait and non-clickbait articles. Our model achieves high performance (74.9% F-1 score) in predicting clickbaits.

Introduction

News headlines are often made to look more interesting/appealing than the actual article, in order to attract clicks, and subsequently monetize the landing page. Various strategies such as building suspense, sensation, luring and teasing, and certain stylistic formats are used to make the headlines look more interesting. Though this is a common journalistic practice, it leads to a frustrating user experience in cases where the landing page, i.e., the actual article, is of low quality and significantly under-delivers the content promised in the headline. We call such pages "clickbaits".

Figure 1 shows examples of clickbaits from the Yahoo homepage stream. The examples link to these articles [1] [2] [3]. As can be seen, the headlines highly exaggerate the content of the landing page.

News aggregators such as Yahoo News serve articles from different news sites on their users' homepages. Since low quality articles such as clickbaits decrease user satisfaction and increase abandonment rate, they should be detected in order to be blocked or demoted in the homepage stream. Algorithms for ranking and recommending articles rely significantly on Click Through Rates (CTR) (Bian et al. 2013; Agarwal et al. 2013). The higher the CTR of a page, the higher its position in the stream. However, clickbaits, by nature, have high CTR's and, hence, the approach cannot be used to differentiate them from genuine high-quality pages.

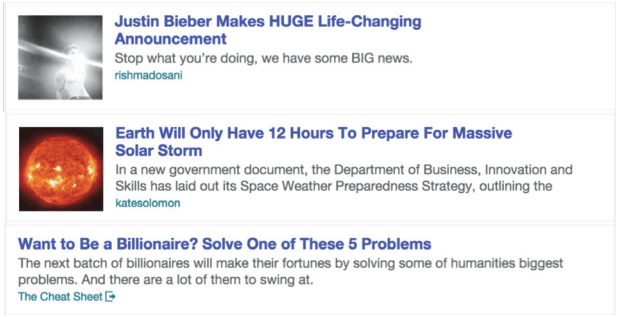

Figure 1: Examples of clickbaits.

A related field of research is identifying bad content on the web, such as spam and fake websites, where features like link-structure (Becchetti et al. 2008) and blacklists of urls, hosts and IPs (Abbasi et al. 2010; Graham 2003) have been found to be useful. However, clickbaits are not spam or fake pages. They can be hosted on reputed news sites. Therefore, these features cannot be used for clickbait detection.

To address the problem of detecting clickbaits, we investigate designing features which are strong indicators of clickbait and combine them in a machine learned model to make automatic inference of whether an article is clickbait or not. We use a variety of features extracted from the body, title, and url of a webpage such as its content, degree of informality and similarity between title and body. Experimental results show that the classifier achieves strong performance.

Contributions of our work as follows: 1. Formalize the concept of clickbait by defining eight types of clickbait. 2. Develop the first automatic machine learning method for identifying clickbaits. 3. Use novel *informality* features for clickbait classification which have not been previously used for similar problems such as identifying spam and fake webpages, and show that these features are the most important

[1] http://news.yahoo.com/justin-bieber-makes-huge-life-changing-184057049.html

[2] http://news.yahoo.com/earth-will-only-have-12-hours-to-prepare-for-081651736.html

[3] http://www.cheatsheet.com/money-career/want-to-be-a-billionaire-solve-one-of-these-5-problems.html

among all the content features in clickbait classification.

Related Work

In Computer Science and related fields, there have been extensive studies on identifying bad quality content on the web, such as spam (Ntoulas et al. 2006; Kolari et al. 2006; Becchetti et al. 2006; Lau et al. 2011) and fake webpages (Abbasi et al. 2010; Abbasi and Chen 2009; Zahedi, Abbasi, and Chen 2015). However, clickbaits are not necessarily spam or fake pages. They can be genuine pages delivering low quality content with misleading titles. In spam detection, various cues extracted from a webpage's content (number of words in a webpage and its title, average length of words and n-grams) (Ntoulas et al. 2006), url (Fetterly, Manasse, and Najork 2004), and link structure (Becchetti et al. 2006) have been used in rule-based, graph-based and machine learning frameworks. Lakkaraju et al. (Lakkaraju, McAuley, and Leskovec 2013) studied the impact of title and community on the success of any content posted in the community. In this work, we use a much larger set of features, in addition to the content features, and show that features such as informality of a web page, its url and similarity between its title and body are strong indicators of it being a clickbait.

In Psychology and in Pragmatics, there have been studies analyzing discourse structure of news headlines for certain properties similar to those of clickbaits, such as sensationalism, luring, dramatization, emotionalism, etc (Molek-Kozakowska 2013; 2014; Blom and Hansen 2015). However, all these works used manual methods for data preparation and analysis, such as interviews and group discussions. In contrast, we develop an automatic machine learning method for identifying clickbaits. Also, we extract features from the body (readable part) of clickbaits, whereas, these studies worked with headlines (titles) only.

Clickbait classification

We define eight types of clickbaits and list them in Table 1.

Problem Statement: Given a webpage (url, title and body), classify it into one of two classes: *clickbait* or *not-clickbait*.

Next, we describe the features used in the classification.

Feature Engineering

We use the following features for classification:

Content Clickbait headlines are made to look appealing by using a certain content type and formatting such as superlative (adjectives and adverbs), quotes, exclamations, use of upper case letters, asking questions, etc. We use presence of these and other indicators as features. We also use unigrams and bigrams form title and body of a page as features. Such features have been found to be very useful for text classification tasks such as subjectivity analysis (Biyani et al. 2014b; 2014a) and sentiment analysis (Biyani et al. 2013; Biyani 2014). Table 2 lists all the content features. For positive and negative words, we use the dictionary created by Hu and Liu (Hu and Liu 2004).

Similarity Clickbait headlines are often misleading and promise substance which is not reflected by the content of the landing page. Therefore, we can posit that the textual similarity between the title and the body of the landing page is lower for clickbaits than it is for non-clickbait pages. So, we use similarity between title and top sentences of the body as features in our model. Specifically, we designed five features corresponding to the similarity between the title and the top one, two, three, four and five sentences of the body (namely Top{1,2,3,4,5}Sim). We used tf-idf encoding to compute the similarity and removed stopwords [4].

Informality and Forward Reference Here we describe Informality and Forward Reference features.

Informality: In general, clickbaits are low quality pages that serve sensational, provoking, gossip-like content. As a result, their language tends to be less formal than that of professionally written news articles. To capture this difference in informality, we compute the following scores from the webpages and use them as features. These scores are indicative of the readability/informality level of a text and have been used previously to measure the informality level and reading difficulty of text (Miltsakaki and Troutt 2008; Mosquera and Moreda 2012; Pérez Téllez et al. 2014; Lahiri, Mitra, and Lu 2011).

1. **Coleman-Liau score (CLScore)** (Coleman and Liau 1975): Computed as $0.0588L - 0.296S - 15.8$ where L is the average number of letters and S is the average number of sentences per 100 words.
2. **RIX and LIX indices** (Anderson 1983): Computed as $RIX = LW/S$ and $LIX = W/S + (100LW)/W$ where W is the number of words, LW is the number of long words (7 or more characters) and S is the number of sentences.
3. **Formality measure (fmeasure)** (Heylighen and Dewaele 1999): The score is used to calculate the degree of formality of a text by measuring the amount of different part-of-speech tags in it. It is computed as $(nounfreq + adjectivefreq + prepositionfreq + articlefreq - pronounfreq - verbfreq - adverbfreq - interjectionfreq + 100)/2$.

CLScore, LIX and RIX are used to gauge the reading difficulty of the text. CLScore and (discretized) RIX index output the approximate US grade level required to comprehend the text whereas (discretized) LIX outputs scores corresponding to five levels of readability: very easy (0-24), easy (25-34), standard (35-44), difficult (45-54) and very difficult (more than 55). In our model, we encode the five levels using integers 0, 1, 2, 3 and 4 respectively. For RIX, the discretization is into thirteen grade levels (0, 1, 2..., 12 and 13 {College Level}) based on whether the score is equal to or greater than 0.2, 0.5, 0.8, 1.3, 1.8, 2.4, 3.0, 3.7, 4.5, 5.3, 6.2, 7.2 respectively with scores below than 0.2 having a grade level of 0. Due to the informal nature of clickbait articles, we posit that, in general, they have a ***lower*** reading grade level (CLScore, RIX and LIX) and formality score than the professionally written news articles.

In addition to the above readability metrics, we also extract some common informality indicators on the web such

[4] http://www.lextek.com/manuals/onix/stopwords1.html

Type	Definition	Example
Exaggeration	Title exaggerating the content on the landing page.	Cringeworthy tattoos that will destroy your faith in humanity.
Teasing	Omission of details from title to build suspense: teasing.	New twist in Panthers star's trial could end his season.
Inflammatory	Either phrasing or use of inappropriate/vulgar words.	Putin Punched at G20 Summit.
Formatting	Overuse of capitalization/punctuation, particularly ALL CAPS or exclamation points.	EXCLUSIVE: Top-Secret Method allowed a mother to break the world record: 12kg in 4 weeks!
Graphic	Subject matter that is salacious or disturbing or unbelievable.	Donatella Versace plastic surgery overload: Waxy face resembles melting candle.
Bait-and-switch	The thing promised/implied from the title is not on the landing page: it requires additional clicks or just missing.	Beers Americans No Longer Drink.
Ambiguous	Title unclear or confusing to spur curiosity.	Hands on: Samsung's iPhone 5 is absolutely beautiful.
Wrong	Just plain incorrect article: factually wrong.	Scientist Confesses: "Global Warming a $22 Billion Scam".

Table 1: Types of clickbait and their examples.

as presence of internet slang [5], swear words [6], and words containing repeated characters (e.g., ooh, aah, etc).

Forward-reference: One of the styles used in framing clickbait headlines is *forward-reference* (Blom and Hansen 2015) to some concept/discourse/entity mentioned in the article (landing page). The forward-reference is used to create a sort of tease or information gap between the headline and the article spurring curiosity among readers and, hence, increasing chances of them clicking on the headlines. Consider the following example headlines.

1. *This Is the Most Horrifying Cheating Story*

2. *He Said He Loved Me...*

In the above examples, (underlined) words "this" and "he" are used to create reference to entities (a story in Example 1 and a person in Example 2) in the respective articles and act as teasers. Blom and Hansen (Blom and Hansen 2015) conducted a study of the usage of forward reference in clickbait news headlines of Danish newspapers and found that forward reference is expressed by demonstrative pronouns, personal pronouns, adverbs and definite articles. Following their findings, we design four features corresponding to the presence of the four grammatical categories in headlines (titles) to explore the effect of forward reference in clickbait classification: Demonstratives (*this, that, those, these*), third person personal pronouns (*he, she, his, her, him*), definite article (*the*) and whether the title starts with an adverb. The corresponding feature names are HasDemonstratives, HasThirdPronoun, HasDefinite and IsStartAdverb.

Url: Urls offer important cues in identifying spam (Fetterly, Manasse, and Najork 2004). To explore if they can be helpful in identifying clickbaits, we extract the following features from urls: frequencies of dash, ampersand, upper case letters, comma, period, equal-to sign, percentage sign, plus, tilde, underscore, and url depth (no. of forward slashes).

[5]http://onlineslangdictionary.com/

[6]http://www.cs.cmu.edu/~biglou/resources/bad-words.txt

Experiments and Results

Data Preparation

Our data comes from different news sites whose pages surfaced on the Yahoo homepage. Sites include the Huffington Post, New York Times, CBS, Associated Press, Forbes, etc. We collected 1349 clickbait and 2724 non-clickbait webpages based on the definitions of clickbait (Table 1). The data came from a period of around one year covering late 2014 and 2015. The articles covered different domains such as politics, sports, entertainment, science and finance. The distribution of clickbaits in different categories is as follows: Ambiguous: 68; Exaggeration: 387; Inflammatory: 276; Bait-and-switch: 33; Teasing: 587; Formatting: 185; Wrong: 33; Graphic: 106. Note that the total number of examples across these categories is more than the total clickbaits as an example can belong to multiple categories.

Experimental Setting

Model Training: We used Gradient Boosted Decision Trees (GBDT) (Friedman 2002) to perform classification experiments. We split the data randomly into training (3000 examples) and testing (1073 examples) sets. We optimized the model on the training set using 5-fold cross validation and evaluated it on the held-out test set. The GBDT parameters were selected by a crude search on the validation set. The best parameter setting corresponded to 500 trees, 20 leaf nodes, 0.1 shrinkage and 0.5 sampling rate.

Feature extraction: We wrote our own scraper and parser to get the headlines and the readable parts (body) from the html source codes of the articles. For part-of-speech tagging, we used the Lingua-EN-Tagger module of CPAN [7]. We used term-frequency to encode word features (unigrams and bigrams) and filtered out tokens with document frequency less than three in the training set. Since the feature dimension was still too high, we used Information Gain (Yang and Pedersen 1996) to rank the features and discarded features with zero information gain. Finally, we got 7677 features. Note

[7]http://search.cpan.org/dist/Lingua-EN-Tagger/Tagger.pm

Feature name	Description	Type
NumWords	Number of words^{t,b}.	N
NumCap	Number of upper case words (excluding acronyms: words with less than 5 characters)^{t}.	N
NumAcronym	Number of acronyms (upper case words with less than five characters)^{t}.	N
{Is,Num}Exclm	Presence/Number of exclamation marks^{t}.	B/N
{Is,Num}Ques	Presence/Number of question marks^{t}.	B/N
IsStartNum	Whether the title starts with a number.	B
HasNumber	Whether the title contains a number(set only if the title doesn't start with a number).	B
IsSuperlative	Presence of superlative adverbs and adjectives (POS tags RBS, JJS)^{t}.	B
{Is,Num}Quote	Presence/Number of quoted words (used "','; excluded 'm, 're, 've, 'd, 's, s')^{t}.	B/N
IsStart5W1H	Whether the title starts with 5W1H words (*what, why, when, who, which, how*).	B
{Is,Num}Neg	Presence/Number of negative sentiment words^{t,b}.	B/N
{Is,Num}Pos	Presence/Number of positive sentiment words^{t,b}.	B/N
{Is,Num}Baity	Presence/Number of the following phrases: click here, exclusive, won't believe, happens next, don't want, you know ^{t,b}.	B/N
HasParenthesis	Presence of parenthesis (round brackets)^{t}.	B
HasMoney	Presence of money mark (dollar, pound signs)^{t}.	B
Unigrams and bigrams	Term frequencies of words and bigrams^{t,b}.	VEC
AvgWordSent	Average words per sentence^{b}.	N

Table 2: Description of content features. Superscripts **t** and **b** imply that the feature is extracted from the title and the body respectively. Feature categories **N** and **B** denote numeric and binary respectively. VEC denotes vector.

that all these optimization were performed on the training data and the held-out test data was completely unseen.

Evaluation: We use precision, recall and F-1 score to evaluate our classifiers. Overall classification performance metrics are computed by taking weighted average of the metrics for individual classes. A class' weight is the ratio of number of instances in it and the total number of instances.

Results

Table 3 presents the results of the 5-fold cross validation on the training set, and applying the trained model on the held out test set. For cross-validation, the results are averaged over the five folds. The performance on the training data is slightly better than that on the test data. This is because of the optimization performed on the entire training data (feature selection, minimum document frequency, etc). We see that the model has an overall strong performance

Class	Precision	Recall	F-1 score
5-fold cross validation on training set			
Clickbait	0.712	0.548	0.619
Non-clickbait	0.804	0.893	0.846
Weighted average	0.774	0.781	0.772
Held out test set			
Clickbait	0.717	0.52	0.603
Non-clickbait	0.775	0.889	0.828
Weighted average	0.755	0.760	0.749

Table 3: Classification performance.

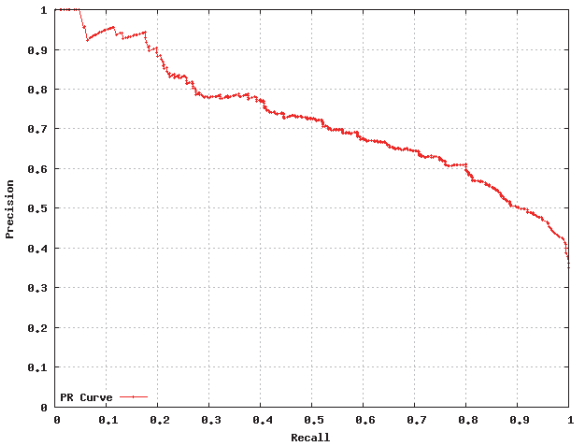

Figure 2: Precision-recall curve for the Clickbait class.

with 0.755 precision and 0.760 recall on the test set. For the Clickbait class, the precision and recall are 0.712 and 0.548 respectively. For the Non-clickbait class, the corresponding numbers are 0.752 and 0.842 respectively. The better performance on the Non-clickbait class can be attributed to the significantly more number of instances in it.

To analyze the classifier performance on the Clickbait class, we plot the precision-recall curve (Figure 2). The area under the curve is 0.723. We see that for precision over 70%, the recall is less than 58%.

Note that the metrics shown in Table 3 are calculated using the classification threshold of 0.5. To see how the metrics vary for different classification thresholds, we plot the threshold curve (Figure 3). We get the highest accuracy (76%) for thresholds between 0.2 and 0.5. Precision increases at a slower rate with threshold. However, recall drops steadily with increased threshold. At a threshold of approx. 0.3, we get equal precision and recall of 66%. Depending on the application requirements (high precision or high recall), the classifier can be operated at the desired threshold.

Next, we investigate the effect of different types of features on classification by using one feature type at a time in the GBDT model. Table 4 presents these results. We see that the informality and forward reference features give the best performance, followed by content features, url features, and similarity features. Furthermore, the combined performance

Class	Pr.	Re.	F-1	Pr.	Re.	F-1	Pr.	Re.	F-1	Pr.	Re.	F-1
	Similarity			URL			Content			Informality & forward ref.		
Clickbait	0.48	0.12	0.24	0.578	0.376	0.456	0.681	0.445	0.539	0.69	0.488	0.572
Non-clickbait	0.667	0.907	0.769	0.718	0.852	0.779	0.749	0.888	0.813	0.762	0.883	0.818
Weighted average	0.602	0.646	0.584	0.669	0.686	0.666	0.725	0.733	0.717	0.737	0.745	0.732

Table 4: Performance of classification models built using one feature type at a time.

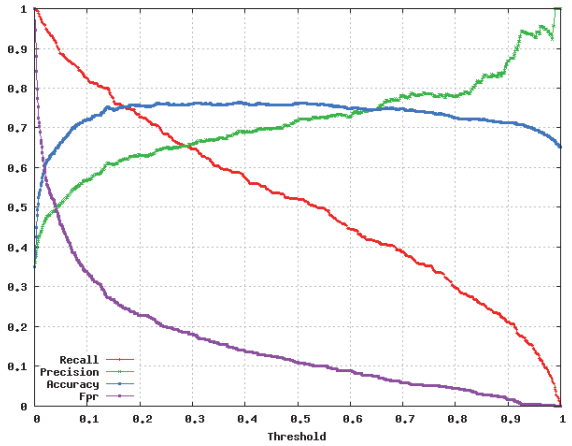

Figure 3: Threshold curve for the Clickbait class. FPR in the legend is false positive rate.

Rank	Feature	Rank	Feature
1	FMEASURE	21	t_"you"
2	NumCap	22	UrlDot
3	IsQuesMark	23	UrlDash
4	RIX	24	HasDemonstrative
5	IsExclm	25	HasNumber
6	NumWordsTitle	26	NumNegBody
7	CLScore	27	t_"one"
8	UrlDepth	28	UrlUnderscore
9	t_"just"	29	IsStartNumber
10	Top3Sim	30	t_"reason"
11	NumSlangBody	31	LIX
12	Top5Sim	32	NumBadBody
13	Top2Sim	33	Is5W1H
14	Top4Sim	34	b_"reason"
15	NumWordsBody	35	b_"just"
16	t_"this"	36	IsStartAdverb
17	b_"photos"	37	Top1Sim
18	b_"image"	38	t_"why"
19	UrlDigit	39	IsNegBody
20	HasDefinite	40	Url_Cap

Table 5: Forty most important features. "t_" and "b_" before a phrase means that the phrase feature is extracted from title and body respectively.

of all the features (Table 3) is better than the performances of all the individual types of features.

Feature importance: We analyze the feature importance by ranking features using Information gain (Yang and Pedersen 1996). Table 5 lists the top 40 features. We see that FMEASURE is the most important feature and three of the seven informality features (FMEASURE, RIX and CLScore) are in top 10 followed by NumSlangBody, LIX and NumBadBody which shows that informality features are indeed one of the most important features for identifying clickbaits. Also, the number of upper case words, presence of question marks and exclamation marks in headlines (titles), and the length of the title (#words) are the most important content features. We see that the most important forward reference feature is HasDefinite followed by HasDemontrative and IsStartAdverb. Top3Sim is the most important similarity feature. Interestingly, url features such as UrlDepth, UrlDigit, UrlDot and UrlDash are ranked high, which implies that urls are good indicators of a page being clickbait or not. We note that certain unigrams from the title and body (said, just, after, you, and image) are ranked high. To analyze their relationship with the clickbait class, we sampled a few articles containing these unigrams in their titles: 1) "Take years off your face in ***just*** 60 seconds." 2) "8 things ***you*** don't want to ask George Clooney's new wife when they come home." As we see, the examples are clickbait and these words are used to *enhance* the impact of the headlines.

Top words and bigrams for individual classes: Feature ranking using Information Gain gives us important (word) features with respect to the class variable {+1,-1} but not for individual classes. To get top words/bigrams for individual classes, we compute word rankings for individual classes using their tf-idf scores. Specifically, for a term t and a class c, we compute the term frequency of t by counting its number of occurrences in the instances belonging to c and multiply the term frequency with the inverse document frequency of t (calculated from the entire corpus) to get the tf-idf score of t for c. Table 6 reports the top ten words and bigrams from the title and body for examples from the two classes. We see that top phrases in the two classes are very different in nature. For the non-clickbait class, top phrases are entities, news subjects and bigrams used to report news such as "according to" and "he said". In contrast, top phrases for the clickbait class are not hard news terms.

Performance on individual clickbait categories: Table 7 reports true positive rates (recall) for the eight clickbait categories, i.e., percentage of instances belonging to a category classified as clickbait. For example, 238 instances *Exaggeration* are classified as clickbait by the classifier so true positive rate will be 238/(#examples in *Exaggeration*) = 238/387 = 61.18% (refer to the Data Preparation Section for distribution of clickbait examples in different categories). Note that since the classifier is binary (clickbait vs non-clickbait), we cannot calculate precision for individual categories as once an instance is classified as clickbait, we do not know

Clickbait				Non-clickbait			
title		body		title		body	
word	bigram	word	bigram	word	bigram	word	bigram
video	the most	image	view this	obama	hong kong	police	he said
photo	faith in	view	this image	ebola	grand jury	ebola	according to
reason	in humanity	people	if you	police	islamic state	state	islamic state
reveals	how to	love	have to	man	in iraq	people	new york
just	about the	photos	the same	report	a new	health	the united
woman	your faith	video	this is	ferguson	kim kardashian	president	the first
watch	the real	show	going to	russia	boko haram	obama	united states
faith	one of	years	you can	state	found in	government	the world
secret	you think	life	the most	american	charged with	military	the country

Table 6: Top ten words and bigrams in title and body for clickbait and non-clickbait classes based on tf-idf scores.

Category	# examples classified as clickbait	True positive rate
Ambiguous	35	51.47%
Exaggeration	238	61.18%
Inflammatory	144	52.17%
Bait-and-switch	16	48.48%
Teasing	335	57.07%
Formatting	115	61.17%
Wrong	15	45.45%
Graphic	43	40.57%

Table 7: True positive rate of different clickbait categories.

in which of the eight categories it has been classified. Also, since the test set contains only 1073 examples and some of the categories have too few examples, we computed TPR by performing 5-fold cross validation on the entire data.

An interesting observation is that TPRs vary hugely across the categories. The *Graphic* category has the lowest TPR which can be attributed to the fact that major clickbait indicators of *Graphic* pages are non-textual such as images and videos whereas our features are extracted from the textual content of the webpages. *Bait-and-Switch* and *Wrong* also have very low TPR which is mainly because of fewer examples in them. Also, we posit that *Wrong* is harder to predict because, even though they contain wrong information, they convey it in a factual manner, similar to non-clickbait articles. We also note that *Formatting* and *Exaggeration* have the highest TPRs. This is consistent with the feature ranking (Table 5) where we found that question marks, exclamation marks, upper case letters and the unigrams "just", "reason", and "you" are among the top ranked features. These features are used for exaggerating and formatting clickbait headlines.

Seeing that the performance of the classifier varies highly across different categories, it would be interesting to explore whether building separate classifiers for each clickbait category using features specific to that category, improves the performance. We plan to investigate this in the future.

We conducted an additional experiment to support our hypothesis that informality features generally have lower values for clickbait than non-clickbait articles. We computed the mean, median and mode of the feature values for the two classes and compared them. Table 8 reports the results. We see that CLScore, RIX and FMeasure have higher values for the three statistics for non-clickbait class. For LIX, the mean is higher for the non-clickbait class, however, the median and the mode are the same as that of the clickbait class. This is consistent with the lower feature importance of LIX (Table 5) as compared to the other three features. For comparing the means of the four features across the two classes, we also conducted a two-sided t-test and found that the null hypothesis was rejected with more than 95% confidence.

Feature	Mean	Median	Mode
CLScore	$8.76^{c}, 9.73^{nc}$	$8.93^{c}, 9.80^{nc}$	$8.60^{c}, 13.5^{nc}$
RIX	$6.48^{c}, 7.16^{nc}$	$6^{c}, 7^{nc}$	$6^{c}, 8^{nc}$
LIX	$2.26^{c}, 2.50^{nc}$	$2^{c}, 2^{nc}$	$2^{c}, 2^{nc}$
Fmeasure	$91.20^{c}, 104.89^{nc}$	$79^{c}, 92.75^{nc}$	$58.2^{c}, 60^{nc}$

Table 8: Statistics of informality features. Superscripts **c** and **nc** denote clickbait and non-clickbait classes respectively.

Error analysis: Finally, we report some examples that were hard to classify, i.e., borderline cases. The following (non-clickbait) example has a confidence of 50.006% of being a clickbait: *Humiliation: Man proposes to girlfriend with the help of 99 iPhones, still gets rejected*. As we see, this example is framed like a clickbait but the content on the landing page justifies the title and, hence, it is not a clickbait. The following clickbait example is a borderline false negative with a confidence of 46.69% being clickbait: *Pedro: The reason I rejected Manchester United and joined Chelsea*. This example is a case of *bait-n-switch* where the landing page does not have the content promised in the title. As discussed previously, this type of clickbait is difficult to detect.

Conclusion and Future Work

We defined 8 types of clickbait and presented a machine learning model built on a variety of features for detecting clickbaits. Extensive experiments show that the model achieves strong classification performance and that ***informality*** is a crucial indicator of the "baity" nature of webpages. Future directions include: 1) Use non-textual features such as images and videos, and user comments on articles, 2) Identify which types of clickbaits are most effective in attracting clicks and devise targeted methods to block them, 3)

Try deep learning to find additional indicators for clickbaits.

References

Abbasi, A., and Chen, H. 2009. A comparison of fraud cues and classification methods for fake escrow website detection. *Information Technology and Management* 10(2-3):83–101.

Abbasi, A.; Zhang, Z.; Zimbra, D.; Chen, H.; and Nunamaker Jr, J. F. 2010. Detecting fake websites: the contribution of statistical learning theory. *MIS Quarterly* 34(3):435–461.

Agarwal, D.; Chen, B.-C.; Elango, P.; and Ramakrishnan, R. 2013. Content recommendation on web portals. *Communications of the ACM* 56(6):92–101.

Anderson, J. 1983. Lix and rix: Variations on a little-known readability index. *Journal of Reading* 490–496.

Becchetti, L.; Castillo, C.; Donato, D.; Leonardi, S.; and Baeza-Yates, R. A. 2006. Link-based characterization and detection of web spam. In *AIRWeb*, 1–8.

Becchetti, L.; Castillo, C.; Donato, D.; Baeza-Yates, R.; and Leonardi, S. 2008. Link analysis for web spam detection. *TWEB* 2(1):2.

Bian, J.; Dong, A.; He, X.; Reddy, S.; and Chang, Y. 2013. User action interpretation for online content optimization. *Knowledge and Data Engineering, IEEE Transactions on* 25(9):2161–2174.

Biyani, P.; Caragea, C.; Mitra, P.; Zhou, C.; Yen, J.; Greer, G. E.; and Portier, K. 2013. Co-training over domain-independent and domain-dependent features for sentiment analysis of an online cancer support community. In *ASONAM*, 413–417.

Biyani, P.; Bhatia, S.; Caragea, C.; and Mitra, P. 2014a. Using non-lexical features for identifying factual and opinionative threads in online forums. *Knowledge-Based Systems* 69:170–178.

Biyani, P.; Caragea, C.; Mitra, P.; and Yen, J. 2014b. Identifying emotional and informational support in online health communities. In *The 25th International Conference on Computational Linguistics*, 170–178.

Biyani, P. 2014. *Analyzing subjectivity and sentiment of online forums*. Ph.D. Dissertation, The Pennsylvania State University.

Blom, J. N., and Hansen, K. R. 2015. Click bait: Forward-reference as lure in online news headlines. *Journal of Pragmatics* 76:87–100.

Coleman, M., and Liau, T. L. 1975. A computer readability formula designed for machine scoring. *Journal of Applied Psychology* 60(2):283.

Fetterly, D.; Manasse, M.; and Najork, M. 2004. Spam, damn spam, and statistics: Using statistical analysis to locate spam web pages. In *Proceedings of the 7th International Workshop on the Web and Databases: colocated with ACM SIGMOD/PODS 2004*, 1–6. ACM.

Friedman, J. H. 2002. Stochastic gradient boosting. *Computational Statistics & Data Analysis* 38(4):367–378.

Graham, P. 2003. Better bayesian filtering. In *Proceedings of the 2003 spam conference*, volume 11, 15–17.

Heylighen, F., and Dewaele, J.-M. 1999. Formality of language: definition, measurement and behavioral determinants. *Interner Bericht, Center Leo Apostel, Vrije Universiteit Brüssel*.

Hu, M., and Liu, B. 2004. Mining and summarizing customer reviews. In *SIGKDD*, 168–177. ACM.

Kolari, P.; Java, A.; Finin, T.; Oates, T.; and Joshi, A. 2006. Detecting spam blogs: A machine learning approach. In *Proceedings of the 21st National Conference on Artificial Intelligence - Volume 2*, 1351–1356. AAAI Press.

Lahiri, S.; Mitra, P.; and Lu, X. 2011. Informality judgment at sentence level and experiments with formality score. In *CICLing*. Springer. 446–457.

Lakkaraju, H.; McAuley, J. J.; and Leskovec, J. 2013. What's in a name? understanding the interplay between titles, content, and communities in social media. In *ICWSM*, 311–320.

Lau, R. Y.; Liao, S.; Kwok, R. C. W.; Xu, K.; Xia, Y.; and Li, Y. 2011. Text mining and probabilistic language modeling for online review spam detecting. *ACM TMIS* 2(4):1–30.

Miltsakaki, E., and Troutt, A. 2008. Real-time web text classification and analysis of reading difficulty. In *Third Workshop on Innovative Use of NLP for Building Educational Applications*, 89–97. ACL.

Molek-Kozakowska, K. 2013. Towards a pragma-linguistic framework for the study of sensationalism in news headlines. *Discourse & Communication* 7(2):173–197.

Molek-Kozakowska, K. 2014. Coercive metaphors in news headlines: Acognitive-pragmatic approach. *Brno Studies in English* 40(1):149–173.

Mosquera, A., and Moreda, P. 2012. A qualitative analysis of informality levels in web 2.0 texts: The facebook case study. In *LREC workshop: NLP can u tag# user generated content*, 23–29.

Ntoulas, A.; Najork, M.; Manasse, M.; and Fetterly, D. 2006. Detecting spam web pages through content analysis. In *WWW*, 83–92. ACM.

Pérez Téllez, F.; Cardiff, J.; Rosso, P.; and Pinto Avendaño, D. E. 2014. Weblog and short text feature extraction and impact on categorisation. *Journal of Intelligent and Fuzzy Systems* 27(5):2529–2544.

Yang, Y., and Pedersen, J. O. 1996. *Feature selection in statistical learning of text categorization*. Center for Machine Translation, Carnegie Mellon University.

Zahedi, F. M.; Abbasi, A.; and Chen, Y. 2015. Fake-website detection tools: Identifying elements that promote individuals use and enhance their performance. *JASIST* 16(6):448–484.