

Automatische Erkennung von Clickbait auf Twitter

Bachelorverteidigung

Sebastian Köpsel

Bauhaus-Universität Weimar

26.05.2016

Was ist Clickbait?


Moderne Medien

Informationsflut

RT @JoelArquillos: How Books Became The Language My Father And I Found Together @davidulin via @buzzfeed


May 10, 2016 2:29:28 PM

A 7-year-old boy who grew out his hair for cancer patients now has cancer




May 10, 2016 1:32:01 PM

Selena Gomez stopped her concert to crumple up a fan's sign




May 10, 2016 9:28:03 AM

RT @BuzzFeedNews: Man Charged With Hacking Celebrities, Stealing Sex Tapes Pleads Guilty

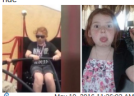


13 short poems that'll tug at the heart strings of any woman who's been body-shamed



May 10, 2016 2:18:02 PM

A girl's scalp was torn off after her hair got caught in a carnival ride



May 10, 2016 11:26:03 AM


What Should You Name Your Puppy?

May 10, 2016 6:30:07 AM


21 totally NSFW photos of massive cocks

May 10, 2016 5:25:04 AM

16 things every book lover has secretly done

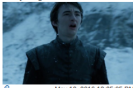


RT @BuzzFeedNews: London Mayor Sadiq Khan Tells Donald Trump He Doesn't Want To Be Exception? To Muslim Ban




May 10, 2016 2:12:59 PM

This Game Of Thrones? fan theory about Bran could change everything




May 10, 2016 12:25:05 PM

This couple is keeping it real by posting hilariously poignant infertility announcements?




May 10, 2016 8:27:02 AM

Welcome to Body Positivity Week! Click here to see all of our Body Positivity Week content:



We need to talk about how fucking amazing Olivia Wilde was in The OC?




May 10, 2016 2:08:01 PM

RT @wordsbydan: I wrote about my body. My bodies. The one I was given, and the ones I made: via @BuzzFeed


May 10, 2016 1:12:07 PM

This plus-size ballerina proves all bodies are beautiful




May 10, 2016 10:29:05 AM

26 Captain America: Civil War? tweets guaranteed to make you laugh



May 10, 2016 7:33:02 AM

Guess how much this hipster fucking food costs



Moderne Medien

Beispiel Überschriften:

- Dieses Mädchen wird dich zum weinen bringen
- Diese Game of Thrones Fantheorie über Bran könnte alles verändern
- Rate wie viel dieses Hipsteressen kostet

Was ist Clickbait?

Gemeinsamkeiten:

- Sensationalistisch
- Irreführend
- Enttäuschend

Was ist Clickbait?

Gemeinsamkeiten:

- Sensationalistisch
- Irreführend
- Enttäuschend

Clickbait Tweets (oder andere Überschriften), die sich eines Tricks bedienen, um den Leser zu einem Klick zu bringen.

Motivation für die Erkennung

Clickbait-Erkennung hat die Ziele:

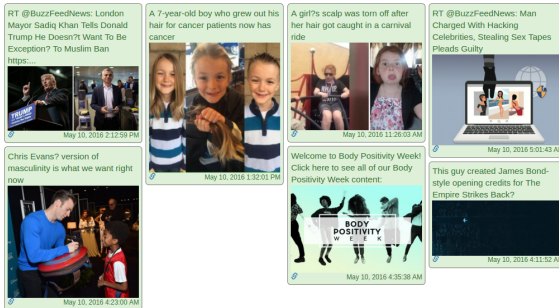
- die Informationsflut zu reduzieren
- als “Spamfilter” zu agieren

<p>RT @JoelArquillos: How Books Became The Language My Father And I Found Together @davidulin via @buzzfeed</p> <p>May 10, 2016 2:29:28 PM</p>	<p>13 short poems that'll tug at the heart strings of any woman who's been body-shamed</p> <p>the way you speak of yourself the way you degrade yourself into smallness is above</p> <p>— rupi kaur</p> <p>May 10, 2016 2:18:02 PM</p>	<p>RT @BuzzFeedNews: London Mayor Sadiq Khan Tells Donald Trump He Doesn't Want To Be Exception? To Muslim Ban https://...</p> <p>May 10, 2016 2:12:59 PM</p>	<p>We need to talk about how fucking amazing Olivia Wilde was in The OC?</p> <p>"It's a little ritual I have to watch the high tide change whenever something major is about to change in my life"</p> <p>May 10, 2016 2:08:01 PM</p>
<p>A 7-year-old boy who grew out his hair for cancer patients now has cancer</p> <p>May 10, 2016 1:32:01 PM</p>	<p>A girl's scalp was torn off after her hair got caught in a carnival ride</p> <p>May 10, 2016 11:26:03 AM</p>	<p>This Game Of Thrones? fan theory about Bran could change everything</p> <p>May 10, 2016 12:25:05 PM</p>	<p>RT @wordsbydan: I wrote about my body. My bodies. The one I was given, and the ones I made: via @BuzzFeed</p> <p>May 10, 2016 1:12:07 PM</p>
<p>Selena Gomez stopped her concert to crumple up a fan's sign</p> <p>May 10, 2016 9:28:03 AM</p>	<p>What Should You Name Your Puppy?</p> <p>May 10, 2016 6:30:07 AM</p>	<p>This couple is keeping it real by posting hilariously poignant infertility announcements?</p> <p>May 10, 2016 8:27:02 AM</p>	<p>This plus-size ballerina proves all bodies are beautiful</p> <p>May 10, 2016 10:29:05 AM</p>
<p>RT @BuzzFeedNews: Man Charged With Hacking Celebrities, Stealing Sex Tapes Pleads Guilty</p> <p>May 10, 2016 5:25:04 AM</p>	<p>21 totally NSFW photos of massive cocks</p> <p>May 10, 2016 5:25:04 AM</p>	<p>16 things every book lover has secretly done</p> <p>May 10, 2016 5:25:04 AM</p>	<p>26 Captain America: Civil War? tweets guaranteed to make you laugh</p> <p>May 10, 2016 7:33:02 AM</p>

Motivation für die Erkennung

Clickbait-Erkennung hat die Ziele:

- die Informationsflut zu reduzieren
- als “Spamfilter” zu agieren



Methoden

- Warum automatisch/mit maschinellem Lernen?
 - Manuell erstellte Regeln und Ausdrücke konnten sich nicht durchsetzen
- Warum Twitter?
 - Alle großen News-Herausgeber vertreten
 - Große Verbreitung
 - Gute API

Maschinelles Lernen

Überwachtes Lernen

- Eine Menge realer Objekte, das Korpus, wird auf Attributvektoren abgebildet
- Alle Objekte/Attributvektoren besitzen Klasse
- Aus Menge dieser Attributvektoren, dem Testset, wird ein Klassifikator trainiert
- Fähigkeit des Klassifikators das gelernte Konzept zu erkennen mit Trainingsset getestet

Beitrag dieser Arbeit

- 1 Korpus + Annotation
- 2 Modell + Attribute
- 3 Experimente

Beitrag dieser Arbeit

- 1 Korpus + Annotation
- 2 Modell + Attribute
- 3 Experimente
 - Attributgruppen und Einzelattribute
 - Herausgeber
 - Lernkurve
 - Persönliche Präferenzen
 - Attributauswahl
 - Fehleranalyse

Beitrag dieser Arbeit

- 1 Korpus + Annotation
- 2 Modell + Attribute
- 3 Experimente
 - Attributgruppen und Einzelattribute
 - Herausgeber
 - Lernkurve
 - Persönliche Präferenzen
 - Attributauswahl
 - Fehleranalyse
- 4 Webservice und Demo

Das Korpus

Aufbau

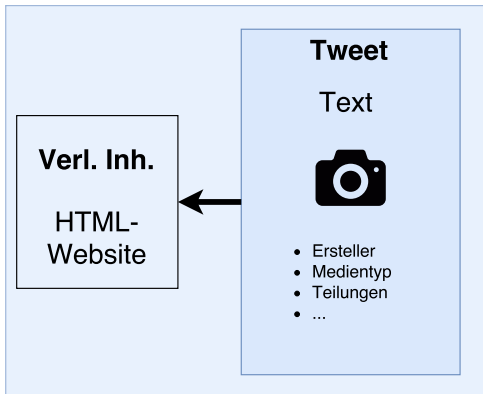
Tweet

Text

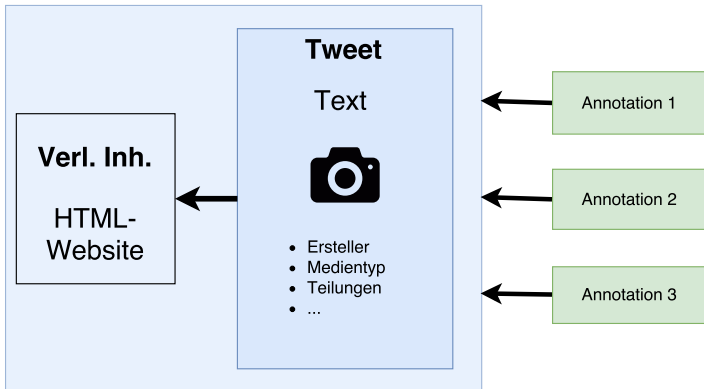


- Ersteller
- Medientyp
- Teilungen
- ...

Aufbau



Aufbau



Annotation

- Bewertung der Clickbait-Stärke
- 4-stufige Bewertungsskala
- Ermöglicht differenzierte Bewertung
- Binarisiert
- Finale Klasse über Mehrheitsentscheid
- Stufen der Bewertung:

Kein	}	Kein-Clickbait
Schwach		
Mittel	}	Clickbait
Stark		

Übersicht

- 20 News-Herausgeber
- 08.06.2015 - 14.06.2015 alle Tweets mit Link gesammelt
- Bilder und die verlinkte Webseite gespeichert
- Pro Herausgeber 150 zufällige Tweets entnommen
- ESPN veröffentlichte nur 142 Tweets (mit Link) in der Woche
- Gesamtgröße von 2992 Tweets
- Binarisiert + Mehrheitsentscheidung:
 - 767 Tweets Clickbait
 - 2225 Tweets kein Clickbait

Das Modell

Attributgruppen

- Tweet
 - General-Inquirer-Wortlisten
 - Downworthy
- Verlinkter Inhalt
- Meta

Bag-of-Words (BOW)

Families who have faced the ...

Zeichen			Wörter		
1	2	3	1	2	3
F	Fa	Fam	Families	Families who	Families who have
a	am	ami	who	who have	who have faced
m	ma	mil	have	have faced	have faced the
...					

Part-of-Speech Tags

POS-Tagging bezeichnet das Zuordnen von Wörtern zu ihrer grammatikalischen Funktion.

Prepare	to	have	your	inbox	carpet	bombed	1
VB	TO	VB	PRP\$	NN	NN	VBD	

¹VB= Verb, Basisform; TO=zu; PRP\$= Personalpronomen; NN= Nomen; VBD= Verb, Vergangenheit

Weitere Attribute

- Flesch-Lesbarkeit
- Einfache Wörter
- Mittlere Wortlänge
- Zeichenlänge des verlinkten Inhalts

Attributauswahl

- Reduziert verwendete Attributmenge
- Vermeidet Overfitting
- Verbessert die Leistung
- Verkürzt Trainingszeit
- Vorauswahl: Nur Attribute die in mehr als 1% des Datensets nicht-triviale Werte besitzen

χ^2 -Attributauswahl

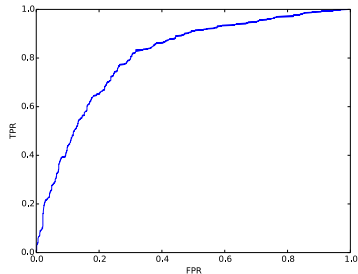
- Pearsons Chi-Quadrat-Test
- Berechnet statistische Abhängigkeit der Attribute
- Testet ob erwartetes Auftreten von Attributwerten = tatsächliches Vorkommen
- Attribute mit größter statistischer Abhängigkeit ausgewählt

Verwendete Klassifikatoren

- Naive Bayes
- Logistische Regression
- J48 (C4.5)
- Random Forest

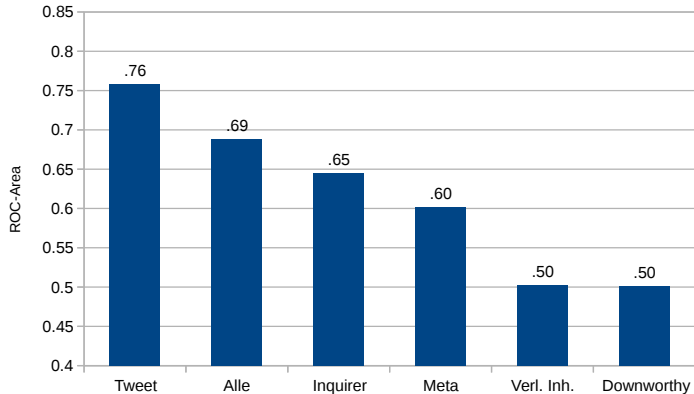
Wie wird Erfolg gemessen?

- Receiver-Operating-Characteristic Area under the Curve
- False Positive Rate zu True Positive Rate item
- Fläche unter der Kurve = ROC-Area
- Schlechtester Wert: 0.5, Bester Wert: 1.0
- Gut für unausgeglichene Klassenverteilungen
- Bezieht alle Schwellwerte ein

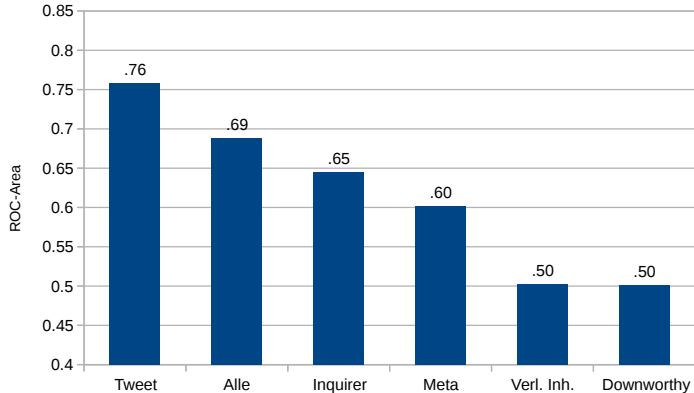


Experimente + Ergebnisse

Vergleich der Attributgruppen



Vergleich der Attributgruppen



Themenunabhängigkeit der Clickbait-Erkennung

Wortattribute der χ^2 Top 100 Tweet-Attribute

- you
- as
- most
- How
- a
- Why
- on
- your
- in

Themenunabhängigkeit der Clickbait-Erkennung

Wortattribute der X^2 Top 100 aller Attribute

Verlinkter Inhalt:

- in
- is
- police
- said
- reported
- a
- State

Tweet:

- you

Einfluss des Erstellers

Ersteller: ROC-Area von 0.673

- Gute Vorhersagekraft (für ein einzelnes Attribut)
- Große statistische Abhängigkeit mit der Clickbait-Klasse

Einfluss des Erstellers

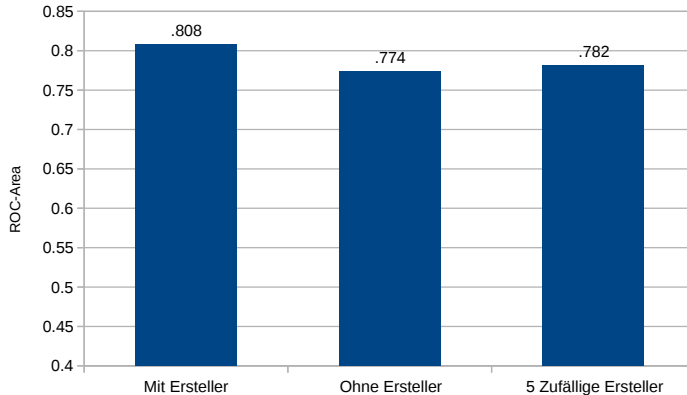
Ersteller: ROC-Area von 0.673

- Gute Vorhersagekraft (für ein einzelnes Attribut)
- Große statistische Abhängigkeit mit der Clickbait-Klasse

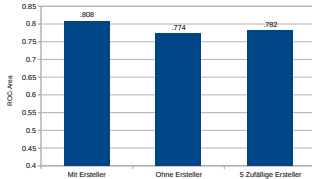
Aber:

- Erzeugt Voreingenommenheit gegenüber dem Tweet
- Bewertet weder Überschrift noch den verlinkten Inhalt

Einfluss des Erstellers



Einfluss des Erstellers



- Ersteller verbessert ROC-Area
- Leistungsgewinn klein
- Unbekannter Stil kein Leistungsunterschied

ROC-Area Bestwerte

- 1 0.826: mit 92 Attributen auf Basis vorheriger Ergebnisse
- 2 0.825: 512 χ^2 -ausgewählte Attribute
- 3 0.814: Alle Tweet-Attribute

Fehleranalyse

- Bewertung von Nicht-Clickbait häufiger korrekt + höhere Konfidenz
- Einige typische Clickbait-Wendungen konnten aufgrund deren vereinzelter Vorkommen noch nicht erkannt werden
- Subtilere Formen der Vorwärtsreferenzierung konnten teils nicht erkannt werden
- Tweets mit Medien häufiger falsch klassifiziert

Herausforderungen

Dies ist möglicherweise das beste Tattoo Cover-Up aller Zeiten



Quelle: <http://www.blogrebellen.de/wp-content/uploads/2015/06/pQoJbeP.jpg>

Webservice und Demo

Server

- Java-Servlet, eingesetzt auf einem Tomcat-Server
- Erwartet GET-Abfragen
- Klassifiziert Tweets (def. über ID) und einfache Texte
- Antwort enthält Konfidenzverteilung und festgestellte Klasse

Demo-Webseite

Demo

Bleacher Report ▾



Apr 18, 2016 7:18:40 AM

At the end of the first half, the Clippers lead the Blazers, 50-42 #NBAPlayoffs



Apr 18, 2016 5:45:46 AM

Dame Dolla vs. CP3: Battle of the point guards, as the Blazers and Clippers face-off in the First Round.



Apr 18, 2016 3:03:51 AM

RT @BR CFB: Class of 2017 WR

The Clippers take Game 1 against the Blazers with a 115-95 victory #NBAPlayoffs



Apr 18, 2016 7:17:54 AM

That Kawhi Leonard defense can bring out all the frustration



Apr 18, 2016 3:24:42 AM

Luke Walton reportedly indicated to Phil Jackson that he is not interested in the Knicks job



Hack-A-Jordan got me like..



Apr 18, 2016 7:07:15 AM

BLAKE is BACK



Apr 18, 2016 4:54:52 AM

After 3 quarters: Spurs - 81 Grizzlies - 51 #NBAPlayoffs



Apr 18, 2016 3:57:25 AM

Looks like Mason Plumlee still hasn't noticed..BLAKE is BACK



Apr 18, 2016 6:22:33 AM

Well, then.



Apr 18, 2016 5:41:03 AM

Uhm..



Apr 18, 2016 5:03:03 AM

Spurs win Game 1 in dominant fashion, defeat the Grizzlies, 106-74 #NBAPlayoffs



Zusammenfassung

Geschafft:

- Korpus mit 2992 Einträgen, von 3 Annotatoren annotiert
- Clickbait-Erkennung: ROC-Area von ≈ 0.83
- Ziel Clickbait-Erkennung über Stil der Überschrift erreicht

Zusammenfassung

Geschafft:

- Korpus mit 2992 Einträgen, von 3 Annotatoren annotiert
- Clickbait-Erkennung: ROC-Area von ≈ 0.83
- Ziel Clickbait-Erkennung über Stil der Überschrift erreicht

Nächste Ziele

- Verbesserung der Repräsentation von Medien im Modell
- Größeres Korpus

Zusammenfassung

Geschafft:

- Korpus mit 2992 Einträgen, von 3 Annotatoren annotiert
- Clickbait-Erkennung: ROC-Area von ≈ 0.83
- Ziel Clickbait-Erkennung über Stil der Überschrift erreicht

Nächste Ziele

- Verbesserung der Repräsentation von Medien im Modell
- Größeres Korpus

Fragen?

Danke für Ihre Aufmerksamkeit!