

Implementierung eines Opinion Mining Algorithmus für Soziale Medien  
zur Bewertung von Produkteigenschaften

**Bachelorarbeit**

zur Erlangung des akademischen Grades „Bachelor of Science (B. Sc.)“ im Studiengang  
Wirtschaftswissenschaft der Wirtschaftswissenschaftlichen Fakultät  
der Leibniz Universität Hannover

vorgelegt von

Name: Gercke



Vorname: Dennis



Prüfer: Prof. Dr. M. H. Breitner

Hannover, den 11.08.2016

## Inhaltsverzeichnis

<b>Inhaltsverzeichnis .....</b>	<b>I</b>
<b>Abkürzungsverzeichnis .....</b>	<b>II</b>
<b>Abbildungsverzeichnis.....</b>	<b>II</b>
<b>1 Einleitung und Motivation.....</b>	<b>1</b>
<b>2 Theoretische Grundlagen .....</b>	<b>3</b>
2.1 Subjectivity Analysis, Sentiment Analysis und Opinion Mining.....	3
2.2 Algorithmen.....	4
2.3 Soziale Medien.....	5
2.3.1 Social Media Monitoring.....	8
2.3.2 YouTube .....	8
<b>3 Aktueller Forschungsstand.....</b>	<b>9</b>
3.1 YouTube Kommentarsektion .....	10
<b>4 Methodik und Daten .....</b>	<b>13</b>
4.1 Auswahl der Datenquelle .....	13
4.2 Auswahl der Daten .....	15
4.3 Auswahl des Algorithmus .....	16
<b>5 Implementierung der Resultate .....</b>	<b>18</b>
5.1 Beschreibung des in R erstellten Algorithmus .....	18
5.2 Theoretischer Vergleich mit dem MOS-algorithm nach Eirinaki et al. ....	26
5.3 Auswertung für einen Testdatensatz .....	28
<b>6 Diskussion und Limitierungen .....</b>	<b>30</b>
6.1 Erkennung der syntaktischen Satzstrukturen .....	30
6.2 Sprachwissenschaftliche Evaluation der Hilfsdaten.....	33
6.3 Weitere mögliche Verbesserungen.....	36
<b>7 Forschungsausblick .....</b>	<b>38</b>
<b>8 Zusammenfassung.....</b>	<b>40</b>
<b>Literaturverzeichnis .....</b>	<b>43</b>
<b>Anhang .....</b>	<b>46</b>
R Markdown: „Daten vorbereiten und Algorithmus ausführen“ .....	46
R Script: „Split.Comments.R“ .....	47
R Markdown: „Ergebnisse überprüfen“ .....	48
Ehrenwörtliche Erklärung .....	50

## 1 Einleitung und Motivation

Ein zentraler Punkt im Prozess der Entscheidung für eine Alternative und gegen eine Andere, ist die Datenbasis, welche ihr zugrunde gelegt wird. Für die meisten Menschen sind dabei gerade die Informationen besonders wertvoll, welche auf den Meinungen von Anderen basieren. Eine Meinung ist eine (private) Einstellung einer Person, die seine Ansicht, Beurteilung und Wertung zu einer Entität ausdrückt (Khan et al., 2014: 259). Sie ist dabei „(...) zentral für fast alle menschlichen Aktivitäten und eine der Haupteinflussquellen unseres Verhaltens“ (Liu, 2012: 5). Ein Verhalten, welches besonders stark von den Meinungen anderer Personen beeinflusst wird, ist dabei das menschliche Kaufverhalten.

Der allgemeinen Informationsbeschaffung kommt hier zu Gute, dass die Meinungen der Menschen nicht so privat sind wie obiges Zitat impliziert. Dies ist vor allem dem Internet geschuldet, welches es ermöglicht, Informationen in nie dagewesener Quantität und Qualität recht gezielt aufzurufen. Besonders nutzergenerierte Inhalte aus sozialen Medien sind für einen erheblichen Teil dieser Datenflut verantwortlich. Nach Obar & Wildman (2015: 746) sind soziale Medien als Dienste zu verstehen, die auf der Grundlage von Web 2.0 Instrumenten „Internet Apps“ darstellen, welche soziale Netzwerke generieren, deren Nutzen meist primär auf nutzergenerierten Inhalten basiert (Wissen, Fotos, Videos, etc.). Neben diesen - in den Medien selbst als Primärdaten zu verstehenden Inhalten – wird auch eine Masse an ihnen zugeordneten Sekundärdaten produziert (z.B. Kommentare, unter einem Video).

Eine der größten Datenquellen für sowohl primäre als auch sekundäre nutzergenerierte Daten dieser Art, stellt die Videoplattform „YouTube“ dar. Der Dienst ist heute eine der meistbesuchten Seiten der Welt (Alexa.com, 2016) und bietet in Bezug auf die Auswertung von sekundären Inhalten - die Kommentare zu den Videos als Textdaten in natürlicher Sprache - einige charakteristische Vorteile, welche auch für diese Arbeit genutzt werden.

Besonders die Auswertung dieser textbasierten Daten, wird zuletzt sowohl von der Forschung als auch von Unternehmen als Herausforderung und Chance gleichermaßen wahrgenommen. Die zunehmende Flut an Informationen macht jedoch computerbasierte Werkzeuge notwendig, um die Auswertung der potenziell wertvollen Informationen in diesen Daten nutzbar zu machen. Als solche sind z.B. Sentiment Analysen (SA) und Opinion Mining's (OM) zu verstehen, die meist durch den Einsatz von Algorithmen automatisiert durchgeführt werden. Obwohl oft unreflektiert synonym verwendet, weisen die Begriffe in ihrer Herkunft und den inhärenten Methoden einige Unterschiede auf.

Während eine SA besonders dann durchgeführt wird, wenn durch Natural Language Processing (NLP) eine positive oder negative Ausprägung im Gesamtkontext ermittelt werden soll (Nasukawa & Yi, 2003: 70), so ist der Gegenstand dieser Arbeit eher ein OM. Dieses wird als wesentlich zielgerichteter verstanden und wendet „information retrieval methods“ an, in dem

es aus einem Input sowohl Attribute, als auch wertende Beschreibungen ermittelt werden um diese zu einer Meinungen über die zu untersuchende Entität zu aggregieren (Dave, Lawrence & Pennock (2003: 519). Um diese Aufgabe bewältigen zu können, werden jedoch ebenfalls Methoden benötigt, um natürliche Sprache auszuwerten und ihre positive oder negative Ausprägung zu bestimmen.

Das Konzept des OM soll in dieser Arbeit genutzt werden, um einen Teilbereich des „Social Media Marketing“, das „Monitoring“ des Erfolgs von Werbekampagnen für Produkte in sozialen Medien zu unterstützen. Dieser Kontext stellt die auf der Verarbeitung von natürlicher Sprache basierenden Methoden vor eine besondere Herausforderung, da die in sozialen Medien im Internet verwendete Sprache ebenfalls von besonderer Natur ist.

In der Arbeit soll nachfolgend aus einer großen Datenmenge von sekundären Inhalten - Textdaten, aus der Kommentarsektion des Social Media (Video-)Dienstes YouTube - eine Basis zur Bewertung von Produkteigenschaften erzeugt werden. Um potenziell wertvollen Daten aufzubereiten und als Basis zu einer solchen Bewertung bereitzustellen, reicht das Vorhandensein eines theoretisch funktionierenden Algorithmus allerdings nicht aus.

Das Ziel der Arbeit ist es daher, einen nicht in einer formalen Notation vorliegenden OM Algorithmus – den „Maximum Opinion Score (MOS)-algorithm“ nach Eirinaki et al. (2012: 1186) - in die Syntax der definierten Programmiersprache eines Statistiktools zu übersetzen, damit er erfolgreich auf Daten angewendet kann, welche Testweise aus sozialen Medien gesammelt wurden. Die Arbeit leistet somit einen praktischen Beitrag zur Evaluation der Anwendbarkeit eines bis dato nur theoretisch funktionierenden Konzepts. Es soll demnach die Fähigkeit des Algorithmus beurteilt werden, die Meinungen in den Daten zuverlässig zu identifizieren und mit hinreichender Genauigkeit zu bestimmen. Diese Beurteilung der Eignung als Datenbasis für Produktbewertungen wird durch einen stichprobenartigen Vergleich mit der Identifizierung und Bestimmung von Meinungen durch einen Menschen gezogen (Human Agreement).

Der Aufbau der Arbeit ist dabei wie folgt strukturiert: Nachdem die Motivation zu der Arbeit und ihre Relevanz für die Forschung durch ihren praktischen Beitrag dargestellt wurden, wird im folgenden Abschnitt zunächst die theoretische Grundlage für ein Verständnis des Forschungsanliegens als Ganzes vermittelt. Hierfür werden die wichtigsten Begriffe definiert und erläutert, die während der Arbeit von zentraler Bedeutung sind.

Nachfolgend wird der aktuelle Stand der Forschung - speziell zum Forschungsgegenstand der YouTube Kommentare herausgestellt, um die Ergebnisse dieser Arbeit in den Kontext früherer Studien setzen zu können. Daraufhin wird eine detaillierte Beschreibung der zugrundeliegenden Methoden vorgenommen, sowie die Daten(quellen)- und Algorithmus-Auswahl näher erläutert.

Im Hauptteil der Arbeit wird dann die Implementierung des gewählten Algorithmus in die definierte Programmiersprache der Statistiksoftware R beschrieben, sowie die hauptsächlichen Faktoren erläutert, die einen Einfluss auf die Genauigkeit der Erkennung der Meinungen aufweisen. In diesem Schritt wird außerdem die Auswertung einer Stichprobe von durch den Algorithmus beurteilten Daten vorgenommen. Abschließend wird eine Reihe von unterschiedlichen Ansätzen diskutiert, deren Einbeziehung die Genauigkeit eines OM mit dem MOS verbessern könnten. In diesem Zuge werden auch Limitierungen des regelbasierten Ansatzes aufgezeigt, sowie ein Forschungsausblick gegeben, welcher sich sowohl auf die Ergebnisse dieser Arbeit stützt, als auch die Forschungsansätze anderer Autoren untersucht.

## **2 Theoretische Grundlagen**

Um die Ergebnisse dieser Arbeit und ihre Aussagekraft zu verstehen und in den richtigen Kontext setzen zu können, soll zunächst eine Definition der wichtigsten verwendeten Begriffe erfolgen. Die Auswahl konzentriert sich hier auf Begriffe, die entweder keine einheitliche Definition besitzen, oder vermehrt verwendet werden, ohne dass dem Nutzer die genaue Wortbedeutung - mit allen Facetten - hinreichend bewusst ist. Im Folgenden seien demnach die für den Umfang dieser Betrachtung anzuwendenden Arbeitsdefinitionen gültig.

### **2.1 Subjectivity Analysis, Sentiment Analysis und Opinion Mining**

Um die grundlegende Funktionalität des zu implementierenden Algorithmus zu beschreiben gibt es viele unterschiedliche Begriffe. Von einigen Autoren werden die Begriffe (besonders SA und OM) teils synonym verwendet (Liu, 2012: 7), jedoch teils auch nuanciert unterschieden (Pang & Lee, 2008: 8).

Auf der obersten Ebene haben die Begriffe zunächst gemein, dass sie - anders als Sachinformationen - subjektiv sind (Liu, 2012: 16). Dies bedeutet, dass ihre Informationen vor ihrer Verwendung gedeutet werden müssen. Einige Autoren sehen deshalb die „Subjectivity Analysis“ als übergeordnete Ebene von OM und SA an (z.B. Tsytarau & Palpanas, 2012: 481). Diesem Ansatz folgen auch Pang & Lee (2008: 11) die sie als „(...)recognition of opinion-oriented language in order to distinguish it from objective language.“ beschreiben.

Obwohl formalisierte Untersuchungen zu Stimmungen und Meinungen natürlich auch früher stattgefunden haben (Liu, 2012: 7), sind die jeweiligen Erstnennungen der heute geläufigsten Arbeitsdefinitionen Anfang des 21. Jahrhunderts zu finden. Hier beschreiben Das & Chen (2001: 50 f.) den Begriff erstmals im Kontext einer automatischen NLP Analyse, um (Finanz-) Marktstimmungen zu erkennen. Die erste wortwörtliche Nennung erfolgte wenig später durch Nasukawa & Yi (2003: 70), welche die Definition des SA schon im Titel ihrer Arbeit nennen, als „Capturing favorability using natural language processing (...)“.

Weiterhin beschreiben Schultes et al. (2013: 673) - speziell im Kontext von YouTube Kommentaren - eine Limitierung ihrer Arbeit, die von sehr vielen Studien in diesem Bereich geteilt wird und damit zugleich einen Ansatz für zukünftige Verbesserungen aufweist: Wie viele andere Forscher auch, verwendeten auch Schultes et al. (2013: 673) für ihre Analyse Videos aus der „Most Popular“ Kategorie, welche sich im gezeigten User-Verhalten - bezogen auf die Kommentare - massiv von einer Vielzahl der anderen, weniger populären Videos unterscheidet, was die Aussagekraft der Ergebnisse deutlich reduziert.

Ein in der Zukunft potenzieller Störeinfluss für alle sozialen Medien wird darüber hinaus von Jindal & Liu (2008: 219) formuliert, welche das „Opinion Spamming“ für eine Problematik halten, welche immer relevanter werden wird. Während sie in ihrer Studie vor allem „product review sites“ dahingehend untersuchten, bleibt sowohl die Verbesserung der Erkennung als auch das Durchsuchen anderer Medien eine Aufgabe, der sich die Forscher zukünftig widmen wollen.

Das Bestätigen ihrer angewendeten Methoden und Ergebnisse auf andere Medien als das bereits untersuchte ist dabei ein auch von anderen Autoren oft genannter Ansatz. Unter anderem Krishna et al. (2013: 128) planen somit die Ergebnisse ihrer Bestimmung einer öffentliche Meinung - welche zuerst durch YouTube Kommentare erprobt wurde - auch mit Daten aus anderen sozialen Medien zu vergleichen.

Für soziale Medien im Allgemeinen, könnte zukünftig auch die Einbeziehung der heute nahezu inflationär gebrauchten Emoticons verwendet werden. Durch ihre festgelegte Bedeutung sollten sie - ähnlich wie die zusätzliche Auswertung des Titels eines Reviews, dessen sich Eirinaki et al. (2012: 1185) als zweite Quelle zur Beurteilung des Gesamtkontext bediente - dazu in der Lage sein, die Erkennung von Meinungen massiv zu verbessern. Die Nutzung von Akronymen wie LOL oder OMG - welche ihre Hochzeit Anfang des 21 Jahrhunderts erlebten - werden langsam aber sicher durch den Gebrauch dieser Emoticons abgelöst, wodurch ein Einbeziehen dieser Daten in den Hintergrund rückt. Anzumerken ist allerdings, dass sowohl die Emoticons als auch die Synonyme zumindest im sozialen Medium YouTube weniger häufig verwendet werden als z.B. auf Facebook. Dies ist vermutlich damit zu begründen, dass YouTube selbst keine animierte Auswahl dieses Kommunikationsmittels zur Verfügung stellt, was eine zukünftige Anwendbarkeit also durchaus möglich erscheinen lässt.

## **8 Zusammenfassung**

Das Ziel der Arbeit war es, einen nicht in einer formalen Notation vorliegenden OM Algorithmus – den „MOS-algorithm“ nach Eirinaki et al. (2012: 1186) - in die Syntax der definierten Programmiersprache eines Statistiktools zu übersetzen, um ihn in der Praxis Testweise auf Daten aus sozialen Medien anwenden zu können.

Im Endergebnis ist zunächst festzustellen, dass die Implementierung des theoretischen MOS-Algorithmus nach Eirinaki et al. (2012: 1186) in weiten Teilen erfolgreich gewesen ist. Es ist somit gelungen, die Beschreibung der regelbasierten Prozessschritte zur Durchführung eines OM, in die Statistiksoftware R zu überführen und dabei die Hauptkomponenten und grundsätzlichen Funktionalitäten korrekt abzubilden. Dabei ist aus verschiedenen Gründen auf die vollständige Implementierung aller Teilkomponenten des Ansatzes verzichtet worden. Hierzu zählen unter anderem besonders die Teile, welche sich nur in der theoretischen Betrachtung als relevant herausgestellt haben und in ihrem Nutzen für die praktische Anwendung des Algorithmus als Ganzes, von anderen Faktoren deutlich dominiert wurden.

Die Ergebnisse der Implementierung wurden im Folgenden genutzt, um Testweise ein OM in Social Media Daten durchzuführen. Hierfür wurde eine der größten Quellen für sowohl primäre, als auch sekundäre nutzergenerierte Daten verwendet - aus der Videoplattform YouTube. So wurde anhand von textbasierten Kommentaren unter den Top 50 englischsprachigen Videos zu drei verwandten Smartphones, ein OM praktisch durchgeführt. Dieses sollte die Fähigkeit des Algorithmus beurteilen, Meinungen korrekt zu erkennen, um so in Zukunft die Datenbasis für eigenschaftsbasierte Produktbewertungen im Rahmen eines „SMM“ bereitzustellen zu können.

Zu diesem Zweck wurde ein stichprobenartiger Vergleich der vom Algorithmus gefundenen Meinungen mit der Identifizierung und Bestimmung von Meinungen durch einen Menschen gezogen (Human Agreement). Die Auswertung ergab dabei, dass lediglich 20% der Meinungen korrekt beurteilt werden konnten. Es ist somit zunächst nicht davon auszugehen, dass die durch diese Implementierung des Algorithmus erzeugte Datenbasis die Grundlage für eine verlässliche, eigenschaftsbasierte Produktbewertung im Rahmen eines „SMM“ bilden kann.

Die Ergebnisse der Arbeit sind dennoch von besonderem Interesse für die Verbesserung der Anwendbarkeit von praktischen Umsetzungen regelbasierter Algorithmen. Es konnten wesentliche Faktoren ausgemacht werden, deren Umsetzung es in der Zukunft ermöglichen können, ein OM in sozialen Medien zur eigenschaftsbasierten Produktbewertung mit erheblich besserer Genauigkeit durchzuführen.

Als der kritischste Faktor für die Genauigkeit der Bestimmung von Meinungen, wurden die in den Wortlisten enthaltenen Meinungsworte (Adjektive) ausgemacht. Die verwendeten Listen wurden von den jeweiligen Erstellern teils auf Vollständigkeit, teils auf Anwendung von Daten aus sozialen Medien ausgelegt, beides blieb im Hinblick auf eine möglichst hohe Genauigkeit in der korrekten Beurteilung von Meinungen weitestgehend erfolglos. Zukünftig könnte durch die Erstellung eines „corpus linguistics“ für ein spezifisches soziales Medium und dessen Abgleich mit einem „corpus linguistics“ für „product review sites“, eine prägnante und hoch spezialisierte Wortliste generiert werden, welche somit in der Lage wäre eine Verbesserung der Meinungserkennung um bis zu 43% zu leisten.

Als ein weiteres großes Fehlerpotential in der zuverlässigen Beurteilung einer Meinung innerhalb einer Aussage eines Kommentares, wurde eine fehlende Sensitivität gegenüber den Strukturen und Inhalten eines Kommentares identifiziert. Hier wurde eine substanzielle Herausforderung für regelbasierte Ansätze erkannt, die syntaktischen Satzstrukturen zu erkennen und eine Unterteilung der Kommentare in Aussagen vorzunehmen.

Die in dieser Arbeit vorgenommene Unterteilung stützte sich hauptsächlich auf satzbeendende Trennzeichen, sowie eine sehr begrenzte Auswahl von Konjunktionen. Die so vorgenommene Einteilung von Kommentaren in Aussagen, konnte durch den Testdatensatz als nicht hinreichend ermittelt werden, da auch nach dieser rudimentären Unterteilung ca. 22% aller falsch beurteilten Meinungen aufgrund von zu langen Aussagen als Input getroffen wurden.

Zukünftig könnte durch die Unterstützung von „Maschine Learning Algorithmen“ - in der Vorbereitung der Daten für die spätere Verarbeitung durch einen regelbasierten Algorithmus - demnach eine deutliche Verbesserung in der Meinungserkennung erzielt werden. Die Algorithmen könnten für den speziellen Anwendungszweck der Unterteilung von Kommentaren in Aussagen, z.B. unter Anwendung von Erkenntnissen des „Structural Priming“, sowie der Zuhilfenahme der „Abstraction“ Methode der für die Wortlisten erstellten „corpus linguistics“ trainiert werden. Dies verspricht eine sehr viel zuverlässigere Erkennung und folgerichtige Unterteilung von Kommentaren in Aussagen, als es durch regelbasierte Ansätze möglich wäre.

Die Arbeit leistet somit vor allem dadurch einen praktischen Beitrag in der Evaluation der Umsetzung eines theoretischen Konzepts, als dass durch die Implementierung in eine Statistiksoftware gezielt die Fehlerpotentiale in der Anwendung aufzeigt, sowie die Grenzen regelbasierter Ansätze herausgestellt werden konnten. Die gefundenen Schwachstellen können somit in zukünftigen Arbeiten weiter erforscht werden, um die Genauigkeit regelbasierter Ansätze des OM zur Durchführung eigenschaftsbasierter Produktbewertungen nachhaltig zu verbessern.