

Masterseminar: Reading Club „Künstliche Intelligenz“,  
WS 2016/17

# **Automatische Humorerkennung in Einzeilern**

Andreas Wolter

Betreuerin: Prof. Dr. Ute Schmid

Professur für Angewandte Informatik, insbes. Kognitive Systeme  
Fakultät Wirtschaftsinformatik und Angewandte Informatik  
Otto-Friedrich-Universität Bamberg

In dieser Arbeit wird die automatische Erkennung von Humor in Einzeilern beschrieben. Dazu werden zunächst die positiven Effekte von Humor erläutert und dann die wichtigsten Humorthorien aus den Gebieten Philosophie, Psychologie sowie Sprachwissenschaft vorgestellt. Es folgt die Beschreibung der Zusammensetzung der beiden verschiedenen Datasets und der Ergebnisse des Experiments, wobei die beiden verwendeten Verfahren der Kreuzvalidierung und der Bayes-Klassifikation vorgestellt werden. Die Arbeit schließt mit einem Ausblick für das Feld Automatischen Humors.

# Inhaltsverzeichnis

<b>1</b>	<b>Positive Wirkung von Humor</b>	<b>4</b>
<b>2</b>	<b>Problemstellung</b>	<b>5</b>
2.1	Die wichtigsten Humorthorien . . . . .	5
2.1.1	Philosophie und Psychologie . . . . .	5
2.1.2	Sprachwissenschaft . . . . .	5
2.2	Humorerkennung in Einzeilern . . . . .	7
2.2.1	Positive und negative Datasets . . . . .	7
2.2.2	Ergebnisse des Experiments . . . . .	9
<b>3</b>	<b>Zukunft im Feld Humorerkennung</b>	<b>11</b>

# 1 Positive Wirkung von Humor

Humor ist essentieller Bestandteil menschlicher Kommunikation, da er den mentalen Zustand von Personen stark beeinflussen kann.

Humor kann starke Emotionen hervorrufen, welche ihrerseits motivierend wirken. Über Emotionen – und damit indirekt über Humor – lassen sich also die Meinung und das Verhalten Anderer direkt beeinflussen [5, S. 609].

Besonders in der Werbung wird die Überzeugungskraft von Humor eingesetzt [5, S. 610].

Indem Witze gegen die Erwartungen und Glaubensgrundsätze der Zuhörer verstoßen, erzeugen sie bei ihnen zunächst Überraschung und bringen sie dann zum lachen [5, S. 610].

Humor bekräftigt auch Kreativität. Situationen, die als lustig empfunden werden, lösen einen Perspektivenwechsel aus und dieser ruft auch neue Denkmuster hervor [5, S. 610].

Kreativität wiederum weist neue Handlungsmöglichkeiten auf und bringt unerwartete Problemlösungen hervor. Kreative Impulse stellen einen der effektivsten Auslöser menschlichen Handelns dar. Computersysteme, die Humor erkennen und erzeugen, können die Gefühle von Nutzern beeinflussen und in ihnen Motivation hervorrufen [5, S. 610].

Maschineller Humorerkennung und -erzeugung kann daher in Werbung, Lern- und Unterhaltungssystemen und allgemein bei Mensch-Computer-Interaktion gewinnbringend eingesetzt werden [5, S. 610].

Für Werbung im Web kann Humor nützlich sein, um zu personalisieren, selektive Aufmerksamkeit zu erhalten und beim Merken von Namen [5, S. 610].

Diese Arbeit beschäftigt sich mit Humor in geschriebener Sprache. Zunächst werden die wichtigsten Humorthorien aus den Feldern Philosophie und Psychologie vorgestellt und dann werden zwei Humorthorien aus dem Bereich Sprachwissenschaft näher betrachtet. Es folgt die Beschreibung der Zusammensetzung der beiden verschiedenen Datensets und der Ergebnisse des Experiments, wobei die beiden verwendeten Verfahren der Kreuzvalidierung und der Bayes-Klassifikation vorgestellt werden. Die Arbeit schließt mit einem Ausblick für das Feld Automatischen Humors.

Die Arbeit folgt hauptsächlich „Computational Humour“ von Strapparava, Stock und Mihalcea (2011) sowie „Making Computers Laugh: Investigations in Automatic Humor Recognition“ von Mihalcea und Strapparava (2005).

## 2 Problemstellung

### 2.1 Die wichtigsten Humorthorien

#### 2.1.1 Philosophie und Psychologie

Aus den Feldern Philosophie und Psychologie stammen die folgenden drei Humorthorien: die *Incongruity Theory*, die *Superiority Theory* und die *Relief Theory* [5, S. 611].

Die Basis für die erste Humorthorie, die *Incongruity Theory*, bildet die Vermischung zweier verschiedener Interpretationsmuster in einer einzigen Aussage. Dabei ist eines der beiden Interpretationsmuster dem Zuhörer zunächst verborgen und zeigt sich erst im Verlauf des Witzes. Diese Theorie dient auch als Grundlage für die beiden sprachwissenschaftlichen Theorien *Semantic Script-Based Theory of Humour* und der *General Theory of Verbal Humour* [5, S. 611].

Die zweite Theorie aus den Bereichen Philosophie und Psychologie ist die *Superiority Theory*. Sie besteht darin, dass sich der Hörer durch Humor überlegen fühlt. Überlegen entweder gegenüber Anderen oder auch gegenüber sich selbst und zwar zu dem Selbst, dass der Hörer vor dem Verstehen des Witzes war [5, S. 611].

Die dritte und letzte Theorie aus den Bereichen Philosophie und Psychologie ist die *Relief Theory*. Laut dieser Theorie umgeht Humor bestimmte Zensoren, die den Hörer davon abhalten, als verboten empfundene Gedanken zu wahrzunehmen. Das Umgehen dieser Zensoren seitens des Hörers ruft in diesem Erleichterung hervor und bringt ihn dadurch zum Lachen [5, S. 612].

#### 2.1.2 Sprachwissenschaft

Da sich diese Arbeit mit Humor in geschriebener Sprache beschäftigt, werden neben den Humorthorien aus den Bereichen Philosophie und Psychologie auch Humorthorien auf linguistischer Ebene benötigt. Davon sollen die beiden bedeutendsten Theorien vorgestellt werden: die *Semantic Script-Based Theory of Humour* und die *General Theory of Verbal Humour*.

#### **Semantic Script-Based Theory of Humour**

Die *Semantic Script-Based Theory of Humour* ist eng verwandt mit der oben beschriebenen *Incongruity Theory*. In dieser linguistischen Humorthorie werden Witze als Skriptopposition dargestellt. Die Struktur des Witzes besteht dabei aus zwei Teilen: der Vorbereitung und der Pointe. Die Vorbereitung bietet mindestens zwei verschiedene Deutungsmöglichkeiten, von denen nur eine offensichtlich ist. Die Pointe lässt aber nur die

weniger offensichtliche Deutungsmöglichkeit zu und aus dieser überraschenden Auflösung entsteht der komische Effekt [5, S. 612].

Laut dieser Theorie müssen komische Texte die folgenden zwei Bedingungen erfüllen: Sie müssen erstens zwei verschiedene Deutungsmöglichkeiten aufweisen und zweitens müssen die Deutungsmöglichkeiten zueinander gegensätzlich sein [5, S. 612].

Ein Beispiel für zwei gegensätzliche Deutungsmöglichkeiten ist folgender Witz: “The first thing that strikes a stranger in New York is a big car” [5, S. 612].

Die Phase der Vorbereitung bietet mit *strike* zwei verschiedene Deutungsmöglichkeiten: entweder *strike* im Sinne von *beeindrucken* oder im Sinne von *anfahen*. Die beiden Deutungsmöglichkeiten stehen im Gegensatz zueinander, denn *beeindrucken* ist eher positiv konnotiert und *anfahen* negativ. Damit sind beide Voraussetzungen für die *Semantic Script-Based Theory of Humour* erfüllt. Die Deutungsmöglichkeit *beeindrucken* ist offensichtlicher und wird deswegen vom Hörer zunächst gewählt. Die Pointe verlangt aber die weniger offensichtliche Deutungsmöglichkeit *anfahen* [5, S. 612].

### General Theory of Verbal Humour

Die *Semantic Script-Based Theory of Humour* bezieht sich hauptsächlich auf semantische Aspekte. Die *General Theory of Verbal Humour* nimmt diese als Basis und erweitert sie mit weiteren sprachwissenschaftlichen Kategorien. Die folgenden sechs Haupt-Wissensgebiete werden auf Witze angewandt. Sie sind absteigend von abstrakt (oben) bis konkret (unten) sortiert [5, S. 613]:

1. Skriptopposition: dies ist die abstrakteste Kategorie. Sie basiert auf der *Semantic Script-Based Theory of Humour*, die besagt, dass der Vorbereitungsteil des Witzes zwei gegensätzliche Deutungsmöglichkeiten bietet.
2. Logischer Mechanismus: er bietet mögliche Auflösungsmechanismen für die Unvereinbarkeit verschiedener Skripte.
3. Situation: sie legt den Kontext des Witzes fest, also den Ort der Handlung und die Teilnehmer.
4. Zielgruppe: sie beschreibt die Personengruppe, über die sich im Witz lustig gemacht wird.
5. Erzählstrategie: sie ordnet die Art des Witzes ein; also ob es sich um einen Dialog, eine Scherzfrage oder eine einfache Erzählung handelt.
6. Sprache: sie beschreibt die sprachwissenschaftlichen Aspekte wie Morphologie, Lexikon, Syntax, Semantik, Pragmatik etc. Sie bildet somit die konkreteste Kategorie.

Es folgt Beispiel, an dem die genannten Kategorien erläutert werden: “How many Poles does it take to screw in a light bulb? Five. One to hold the light bulb and four to turn the table he is standing on” [5, S. 613].

Die Skriptopposition besteht darin, dass ein Gegensatz zwischen dem erwarteten Verhalten und der tatsächlichen Antwort besteht. Das erwartete Verhalten wäre, dass nur eine Person benötigt wird, um die Glühbirne auszutauschen. Aber die Pointe bietet eine dazu gegensätzliche Auflösung, indem fünf Personen benötigt werden. Der Logische Mechanismus ist eben diese Auflösung, die das erwartete Verhalten praktisch umkehrt. Die Situation ist das Wechseln einer Glühbirne. Die Zielgruppe sind die Polen. Die Erzählstrategie ist Scherzfrage. Sprachwissenschaftliche Aspekte wären zum Beispiel, dass pragmatisch eine rhetorische Frage gestellt wird oder dass auf syntaktischer Ebene ein Fragesatz und zwei Aussagesätze vorliegen [5, S. 613].

## 2.2 Humorerkennung in Einzeilern

Die vorgestellten Humortheorien bilden die Grundlage für das Experiment in: Strapparava, Stock und Mihalcea: „Computational Humour“ (2011).

Ziel des Experiments ist es, Humor in sogenannten Einzeilern zu erkennen. Einzeiler besitzen eine einfache Syntax; sie bestehen meist nur aus einem Satz. Sie haben eine Länge von 10 bis 15 Wörtern und weisen einen kreativen Sprachgebrauch auf [5, S. 624]. Ein Beispiel für einen Einzeiler ist: “A clean desk is a sign of a cluttered desk drawer.” [5, S. 626])

Es soll geprüft werden, ob automatische Textklassifikation ein geeigneter Ansatz ist, komischen von nicht-komischem Text zu unterscheiden [5, S. 624].

Beim Experiment sollen positive (humorvolle) and negative (nicht-humorvolle) Beispiele durch Textklassifikation eingeordnet werden [5, S. 625]. Im Folgenden werden die beiden Datasets, aus denen die Beispiele jeweils entnommen werden, vorgestellt.

### 2.2.1 Positive und negative Datasets

#### Positiver Dataset

Zur Erstellung des positiven Datasets kommt ein Web-basierter *bootstrapping*-Prozess zum Einsatz. Als erster Schritt wurden zunächst per Hand zehn Einzeilern gesammelt, die dem Prozess als *seed* dienen. Über diesen *seed* findet der Algorithmus dann im zweiten Schritt per Websuche Seiten, auf denen mindestens einer der Einzeiler aus dem *seed* enthalten ist. Als dritter Schritt werden auf diesen Webseiten dann weitere Einzeiler erkannt und zum *seed* hinzugefügt.

Dabei gibt es zwei Einschränkungen, die dafür sorgen, dass möglichst nur Einzeiler in den *seed* aufgenommen werden und keine Datensätze, die gar keine Einzeiler sind. Die erste Einschränkung ist thematisch und sie verlangt, dass in einer URL folgende Schlüsselwörter vorkommen müssen, damit die betreffende Seite weiter auf Einzeiler hin durchsucht wird: “oneliner, one-liner, humor, humour, joke” [3, S. 532 f.]. Die zweite Einschränkung bedient sich der HTML-Struktur. Die im dritten Schritt des *bootstrapping*-Prozesses durch Websuche gefundene Seite wird weiter durchsucht, indem die tag-Struktur des Einzeilers, über den die Seite gefunden worden war, analysiert wird. Wenn sich der Einzeiler in einem Kindelement einer Aufzählungsstruktur befindet, wird der Inhalt

gleichnamiger Kindelemente auf der selben Hierarchieebene als Einzeiler herausgefiltert [3, S. 533].

Der zweite und der dritte Schritt des *bootstrapping*-Prozesses werden dann solange wiederholt, bis der Dataset groß genug ist, sprich circa 16000 Einträge [5, S. 625]. Es folgen Beispiele für Datensätze aus dem positiven Dataset:

“Take my advice; I don’t use it anyway.” [5, S. 626]

“I get enough exercise just pushing my luck.” [5, S. 626]

“Beauty is in the eye of the beer holder.” [5, S. 626]

### Negativer Dataset

Als negativer (nicht-komische) Dataset wurde nach Beispielen gesucht, deren Struktur möglichst ähnlich zu der der Einzeiler ist, bis auf den Umstand, dass sie eben keine komischen Inhalte aufweisen. Deswegen wurde die Länge der negativen Datensätze auf 10 bis 15 Wörter beschränkt. Dies dient dazu, um zu vermeiden, dass nur anhand der Textlänge zwischen komischen und nicht-komischen Datensätzen unterschieden wird [5, S. 626]. Der negative Dataset setzt sich aus Inhalten von insgesamt vier verschiedenen Korpora zusammen [5, S. 626]:

**Reuters** Das erste Korpus besteht aus Überschriften der Nachrichtenagentur *Reuters*. Sie besitzen eine einfache Syntax und sind oft so formuliert, dass sie die Aufmerksamkeit des Lesers wecken, zum Beispiel: “Trocadero expects tripling of revenues” [5, S. 626].

**Sprichwörter** Das zweite Korpus besteht aus Sprichwörtern aus einer Online-Sammlung. Im Allgemeinen sind Sprichwörter kompakt und man kann sie sich leicht merken. Diese beiden Eigenschaften macht sie ähnlich zu Einzeilern. Ein Beispiel für dieses Korpus ist: “Beauty is in the eye of the beholder” [5, S. 626].

**BNC** Im dritten Korpus finden sich Sätze aus dem *British National Corpus* (BNC). Bei diesen Sätzen gibt es – im Gegensatz zu den beiden Korpora *Reuters* und Sprichwörter – keine zusätzliche Kreativität. Die Sätze wurden so ausgewählt, dass sie inhaltlich ähnlich zu den Einzeilern sind. Ein Datensatz aus diesem Korpus ist: “The train arrives three minutes early” [5, S. 626].

**OMCS** Der letzte Korpus setzt sich aus Inhalten des *Open Mind Common Sense*-Korpus (OMCS) zusammen. Dies ist eine Sammlung von englischen Aussagen allgemeiner Gültigkeit, die von Freiwilligen zusammengetragen wurden. Es handelt sich meist um einfache Sätze, die Erklärungen in Wörterbucheinträgen ähneln, nur dass sie in einer weniger spezifischen Sprache geschrieben sind. Diese Sammlung wurde in den negativen Dataset aufgenommen, da Witze oft mit unseren Erwartungen, die sich in allgemeingültigen Aussagen widerspiegeln, brechen. Ein Beispiel aus diesem Korpus ist: “Humans generally want to eat at least once a day” [5, S. 626].



## 2.2.2 Ergebnisse des Experiments

### Kreuzvalidierungsverfahren

Das Experiment wurde mit einem zehn-schichtigen Kreuzvalidierungsverfahren durchgeführt [5, S. 629]. Dies dient dazu, zu verhindern, dass der Lernalgorithmus zufällig anhand eines einzigen Attributs klassifizieren lernt. Beim Kreuzvalidierungsverfahren werden die positiven und negativen Datasets zusammengenommen und aus dieser Gesamtmenge zufällige etwa gleich große Paare bestehend aus Trainings- und Testmengen entnommen [2, S. 1138]. Dazu wird die Gesamtdatenmenge ( $D$ ) in  $k$  Untermengen ( $S$ ), die sogenannten *folds*, aufgeteilt:  $S_1 \dots S_k$ . Nun wird der Lernalgorithmus  $k$ -Mal angewendet, indem die Vereinigungsmenge aller Untermengen außer  $S_i$  ( $D \setminus S_i$ ) als Trainingsmenge herangezogen wird und  $S_i$  als Testmenge [4, S. 249].

Das Problem hierbei ist, dass die Proportion der Label in der Testmenge, also ob ein Datensatz Teil des positiven oder negativen Datasets ist, sich nicht unbedingt in der Gesamtmenge widerspiegelt. Deswegen wird eine Stratifizierung durchgeführt, um das Verhältnis der Label in der Testmenge zu dem des Trainingssets anzugleichen [2, S. 1139]. Im konkreten Experiment ist das stratifizierte Kreuzvalidierungsverfahren zehn-schichtig. Heißt, dass es insgesamt zehn Untermengen (*folds*) gibt und dass bei diesen das Verhältnis der Label angeglichen wurde.

### Bayes-Klassifikator

Um die Datensätze anhand inhaltlicher Eigenschaften in einen der beiden Datasets einzuordnen, wurde ein Bayes-Klassifikator verwendet. Dabei wird die Wahrscheinlichkeit, welcher Kategorie ein bestimmter Datensatz angehört, über die vorher festgelegten und verknüpften Wahrscheinlichkeiten von Wörtern und Dokumenten berechnet [3, S. 534].

Beim Bayes-Klassifikator wird vereinfachend angenommen, dass die Attributewerte, in diesem Fall also die Einzelwörter, voneinander unabhängig sind.

Der Bayes-Klassifikator basiert auf dem Satz von Bayes:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (2.1)$$

Das bedeutet, dass  $B$  einen Datensatz darstellt, von dem nicht bekannt ist, ob er zur Klasse  $A$  gehört ( $P(A)$ ). Also, ob es sich um einen Einzeiler (Klasse  $A$ ) handelt, oder nicht. Ziel ist es nun, über die Beobachtung von  $B$  die Wahrscheinlichkeit  $P(A|B)$  zu bestimmen, ob  $B$  zur Klasse der Einzeiler ( $A$ ) gehört.

Der Satz von Bayes bildet die Grundlage für probabilistisches Lernen, indem Vorwissen und Beobachtungen in die Gleichung miteinfließen [1, S. 211].

In Tabelle 2.1 sind die jeweiligen Genauigkeiten der Humorerkennung für den Bayes-Klassifikator verzeichnet [5, S. 630]. Dabei ist *Reuters* dasjenige Korpus, das sich am meisten von Einzeilern unterscheidet und das *British National Corpus* dasjenige, das den Einzeilern am meisten ähnelt.

	Reuters	BNC	Sprichwörter	OMCS
Bayes-Klassifikator	96.67	73.22	84.81	82.39

Tabelle 2.1: Genauigkeit der Humorerkennung für den Bayes-Klassifikator (in %)

### 3 Zukunft im Feld Humorererkennung

In dieser Arbeit wurde die automatische Erkennung von Humor betrachtet. Dabei wurden Datensätze durch einen Bayes-Klassifikator in positive und negative Datasets eingeteilt. Diese simple Einteilung kann als Basis für eine allgemeinere Erkennung von Emotionen dienen.

Die automatische Erkennung und Produktion von Humor kann bei Schnittstellen zur Mensch-Computer -Interaktion sinnvoll eingesetzt werden, um Computer menschlicher oder einfach sympathischer erscheinen zu lassen. Die maschinelle Produktion von Humor hilft auch beim Lernen von Fremdsprachen, da über Humor die Merkfähigkeit gesteigert werden kann. Maschinelle Produktion von Humor kann auch zur Erzeugung von Neuheit eingesetzt werden, um automatisch und personalisiert neue Werbung generieren zu können.

# Literatur

- [1] Wynne Hsu. „Bayesian Classification“. In: *Encyclopedia of Database Systems*. Hrsg. von Ling Liu und M. Tamer Özsu. Boston, MA: Springer US, 2009, S. 210–214.
- [2] Ron Kohavi. „A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection“. In: *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*. Bd. 2. (Montréal, Québec). San Francisco, CA: Morgan Kaufmann, 20.–25. Aug. 1995, S. 1137–1145.
- [3] Rada Mihalcea und Carlo Strapparava. „Making Computers Laugh: Investigations in Automatic Humor Recognition“. In: *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*. (Vancouver, British Columbia). Hrsg. von Raymond Mooney u. a. Stroudsburg, PA: Association for Computational Linguistics, 6.–8. Okt. 2005, S. 531–538.
- [4] „Stratified Cross Validation“. In: *Encyclopedia of Machine Learning*. Hrsg. von Claude Sammut und Geoffrey I. Webb. Boston, MA: Springer US, 2010, S. 928.
- [5] Carlo Strapparava, Oliviero Stock und Rada Mihalcea. „Computational Humour“. In: *Emotion-Oriented Systems: The Humaine Handbook*. Hrsg. von Roddy Cowie, Catherine Pelachaud und Paolo Petta. Berlin, Heidelberg: Springer, 2011, S. 609–634.