

Verbindung manueller und automatisierter Inhaltsanalyse zur Erhebung von Argument- komponenten in Online-Partizipationsverfahren: *aktuelle Ergebnisse und Ausblick*

Katharina Esau & Matthias Liebeck

Gelsenkirchen, 28. Oktober 2016

Heinrich-Heine-Universität Düsseldorf



Fachhochschule
für öffentliche Verwaltung
NRW



Wenn ein Beteiligungsverfahren gut läuft, dann produziert es große Textmengen:

The image displays a screenshot of a public participation platform. The main header reads "Beteiligen Sie sich an der Kommentierung des Entwurfs zur Leitentscheidung Braunkohle". Below this, a navigation bar includes "STARTSEITE", "KOMMENTARE UND STELLUNGNAHMEN", "SIE LEITENTSCHEIDUNG", and "INFORMATIONEN".

A central banner features a woman's portrait and the text: "Willkommen auf der Plattform „Eine nachhaltige Perspektive für das Rheinische Braunkohlenrevier“". Below the banner, a "Statistik" section shows: 119 Stellungnahmen, 6 Studien & Vorträge, 1216 Teilnehmer, 17300 Bewertungen, and 1403 Kommentare.

The right side of the screen shows a detailed discussion thread titled "TEMPELHOFFER FELD" with the subtitle "Entwicklungs- und Pflegeplan". The thread includes a "Das Tempelhofer Feld" section with a description of the area and a "Veranstaltungen" section listing events like "Bewertung Gb, 04, 05, 06" and "Regelmäßige Besprechungen 5, 2. Februar".

Fragen, die von Forschung und Praxis an die Texte gestellt werden, z. B.:

- Welche Vorschläge wurden eingebracht?
- Wie kontrovers wurden einzelne Vorschläge diskutiert?
- Wurden Debatten von einzelnen Nutzern dominiert?
- Wie argumentativ verliefen die Diskussionen? Welche Argumente wurden angeführt?
- Wie emotional wurde zu bestimmten Themen diskutiert?



- *Lösungsansatz*: Verbindung von manueller und automatisierter Inhaltsanalyse. Insbesondere Verfahren des maschinellen Lernens haben sich als vielversprechend erwiesen (Scharkow 2011).
- Dabei wird im Gegensatz zu älteren automatisierten Verfahren (v. a. Wortauszählungen), manuell codiertes Material verwendet, um einer Software („Maschine“) „beizubringen“ Textabschnitte sinnvoll zu vorab festgelegten und codierten Kategorien zuzuordnen.
- *Zielsetzung*: Perspektivisch soll die Software selbstständig Textinhalte klassifizieren und dabei möglichst analog zur Arbeit der Codierer vorgehen.

- *Kommunikationswissenschaft*: erste Versuche Textinhalte unterstützt durch maschinelles Lernen (teil-)automatisiert zu analysieren (z. B. Scharnow 2012, Sommer et al. 2014).
- *Informatik/Computerlinguistik*: hier beschäftigt sich der Bereich Argument Mining mit der automatisierten Analyse von Argumentkomponenten in Texten (z. B. Habernal/Gurevych 2016).
- Darauf lässt sich aufbauen, aber die bisherigen Verfahren sind selten für die Analyse von Online-Diskussionen geeignet und sind nicht für Online-Partizipationsverfahren konzipiert.
- Bisher kein Ansatz, der Argumentationsstrukturen automatisiert erhebt, so dass Forschung und Praxis der Online-Partizipation damit arbeiten können.

- Es gibt kein einheitliches Argumentationsmodell, auf das sich Forscher einigen können (Habernal et al., 2014)
- Die Wahl des Modells/der Kategorien hängt von der Fragestellung und v. a. von der Textart (z.B. Nachrichtenartikel, persuasive essays etc.) ab.
- Problem: in Online-Diskussionen meist keine perfekt strukturierten Argumentationsverläufe!
- Daher wurde bei der Kategorienbildung aufbauend auf Toulmin (1958) ein vereinfachtes Argumentationsmodell gewählt, bestehend aus: Vorschlägen (major position), Pro- und Contra-Positionierungen (claim) und Begründungen (premise).

- Datengrundlage für die quantitative Inhaltsanalyse war ein Beteiligungsverfahren der Stadt Berlin „Entwicklungs- und Pflegeplan Tempelhofer Feld“ (N=1524¹ Textbeiträge).
- Probecodierung von 82 Textbeiträgen (Reliabilität zwischen drei Codierern betrug im Mittel 78 % Krippendorffs α).
- Für die finale manuelle Inhaltsanalyse wurde ein Sample von 647 Textbeiträgen zufällig ausgewählt und von den drei Codierern codiert.
- In dem Sample wurden 337 Vorschläge, 256 Pro-Positionierungen, 94 Contra-Positionierungen und 745 Begründungen codiert.

¹Beschränkung auf 5 von 7 Themenbereichen: nur Themenbereiche, in denen Vorschläge zur Umsetzung eingebracht wurden.

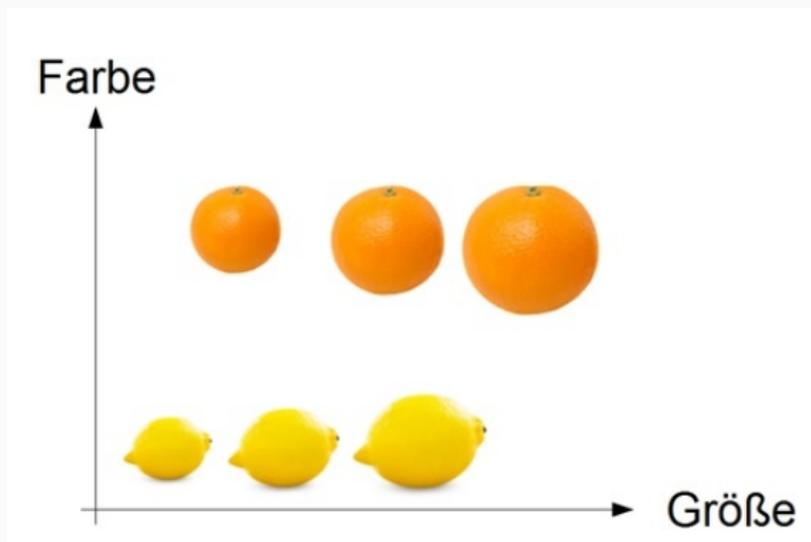
- Die Codierung erfolgte manuell durch das Markieren von Textabschnitten, analog zur Markierung eines Textes mit Textmarkern, in einer dafür geeigneten Weboberfläche:

1	Claim contra Ich sehe das Anlegen von einfachen Spielplätzen eher kritisch und das obwohl ich selbst Kinder habe.
2	Premise Im Umkreis des Feldes sind bereits viele zum Teil sehr schöne Spielplätze vorhanden.
3	Premise Dafür muss meiner Ansicht nach das Feld nicht bebaut werden.
4	Major position Begrüßen würde ich allerdings eine Art Naturspielplatz, der eher temporären Charakter hat und wandelbar ist.
5	Major position Siehe auch: http://de.wikipedia.org/wiki/Naturerfahrungsraum (http://de.wikipedia.org/wiki/Naturerfahrungsraum)

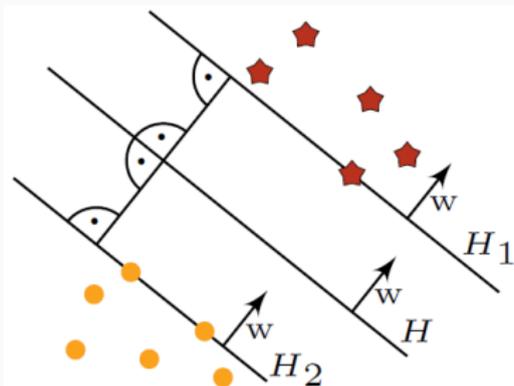
- Im Rahmen unserer Arbeit haben wir bisher zwei Klassifikationsaufgaben auf Satzebene betrachtet:
 - Task A: Enthält ein Satz gemäß unseres Argumentationsmodell argumentativen Inhalt?
 - Task B: Bestimme Argumentationskomponenten in Sätzen mit genau einer codierten Argumentationskomponente
- Perspektivisch wollen wir auch Argumentationskomponenten in Satzteilen und auch satzübergreifend automatisiert erkennen.



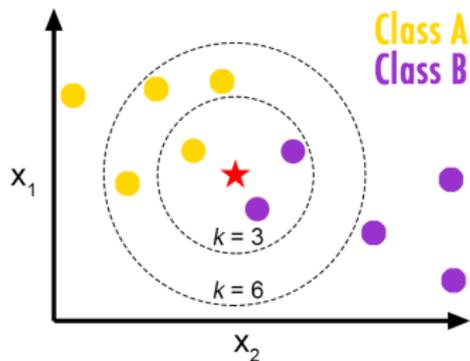
- Um der „Maschine“ die Erkennung von Argumentationskomponenten beizubringen, müssen aus den Texten zunächst Merkmale extrahiert werden.
- Zunächst ein Beispiel für die Erkennung von Zitronen und Orangen:



- Anschließend wird ein Klassifikator verwendet, um, basierend auf den Merkmalen, eine Zuordnung eines Objektes (Bildes) zu einer Klasse (Zitrone / Orange) zu ermöglichen:
- Zwei von uns eingesetzte Verfahren:



Support Vector Machine (SVM)

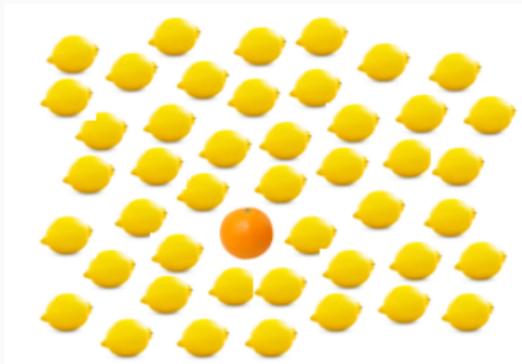


k-Nearest Neighbor (k-NN)

Anstatt Größe und Farbe werden nun andere Merkmale verwendet:

- Unigramme $\overline{A} \quad \overline{B} \quad \overline{C} \quad \overline{D} \quad \overline{E}$
- Bigramme $\overline{A \quad B} \quad \overline{B \quad C} \quad \overline{C \quad D} \quad \overline{D \quad E}$
- Verteilung Wortarten (z.B. Nomen 50%, Verben 10%, Adjektive 20%, ...)
- Verteilung Abhängigkeiten
- Satzlänge
- Punctuation
- Position des Satzes im Text
- Anzahl Links

- Um die Ansätze evaluieren zu können, wurde der Datensatz in eine Trainingsmenge (80%) und eine Testmenge (20%) aufgeteilt.
- Die Klassifikatoren werden auf der Trainingsmenge trainiert und auf der Testmenge mit einem Evaluationsmaß evaluiert.
- Dazu wird nicht die übliche Genauigkeiten verwendet, da sie bei einer Ungleichverteilung der Klassen nicht geeignet ist:



- Stattdessen setzen wir das F-Maß ein.
- Am Beispiel der Erkennung von Zitronen lässt es sich wie folgt veranschaulichen:
 - Precision: Wie viele der als Zitronen erkannten Früchte sind wirklich Zitronen?
 - Recall: Wie viele der Zitronen wurden als Zitronen erkannt?
 - F-Maß = $2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$
- Die besten Ergebnisse unseres ersten Prototypen werden mit einer SVM erreicht:
 - Erkennung argumentativer Sätze: F-Maß = 69,77%
 - Erkennung von Argumentationskomponenten: F-Maß = 68,5%

- Diskussion der bisherigen Ergebnisse
 - Entwicklung eines praxisorientierten Argumentationsmodells für Online-Partizipationsverfahren
 - Erste prototypische Implementierung die bereits gute Ergebnisse erlangt
 - Interdisziplinäre Publikation: M. Liebeck, K. Esau und S. Conrad. *What to Do with an Airport? Mining Arguments in the German Online Participation Project Tempelhofer Feld*. Proceedings of the 3rd Workshop on Argument Mining, 2016
 - Zweiter Konferenzbeitrag zur Begutachtung eingereicht und Publikation in einem kommunikationswissenschaftlichen Journal geplant
- Limitierung
 - Die automatisierte Erkennung erfolgt bisher nur auf Satzebene. In der Praxis sind aber auch einzelne Argumente auf mehrere aufeinanderfolgende Sätze verteilt.

- Weitere Merkmale für das maschinelle Lernen implementieren und evaluieren
- Nach Möglichkeit weitere Datensätze manuell codieren und die Ergebnisse der Verfahren besser bewerten zu können
→ zeitaufwändig
- Desweiteren soll die automatisierte Erkennung von Emotionen neue Erkenntnisse bringen:
 - Praxis: Monitoring von Stimmungen und Stimmungsverläufen (z. B. über Stimmungsbarometer)
 - Forschung: Rolle von Emotionen in der Diskussions- und Meinungsdynamik

- Habernal, I. & Gurevych, I. (2016). Argumentation Mining in User-Generated Web Discourse. Computational Linguistics. <https://arxiv.org/pdf/1601.02403v4.pdf>.
- Scharrow, M. (2011). Zur Verknüpfung manueller und automatischer Inhaltsanalyse durch maschinelles Lernen. Medien & Kommunikationswissenschaft 59: 545-562.
- Scharrow, M. (2012). Automatische Inhaltsanalyse und maschinelles Lernen. Dissertation, epubli GmbH.
- Sommer, K., Wettstein, M, Wirth, W. & Matthes, J. (2014). Automatisierung in der Inhaltsanalyse, Köln: Herbert von Halem Verlag.
- Toulmin, S. (2003 (1958)). The Uses of Argument, Updated edition, Cambridge University Press.

- Screenshot von leitentscheidung-braunkohle.nrw
- Screenshot von <https://tempelhofer-feld.berlin.de>
- Orange: Wikipedia CC3-SA:
<https://commons.wikimedia.org/wiki/File:Orange-Whole-%26-Split.jpg>
- Zitrone: <http://www.veggipedia.nl/consumenten/producten/216>
- KNN: <http://adataanalyst.com/machine-learning/knn/>

