

Automatisierte Diskussionsanalyse in Online-Partizipationsverfahren

von Matthias Liebeck, M.Sc., Institut für Informatik, Heinrich-Heine-Universität Düsseldorf

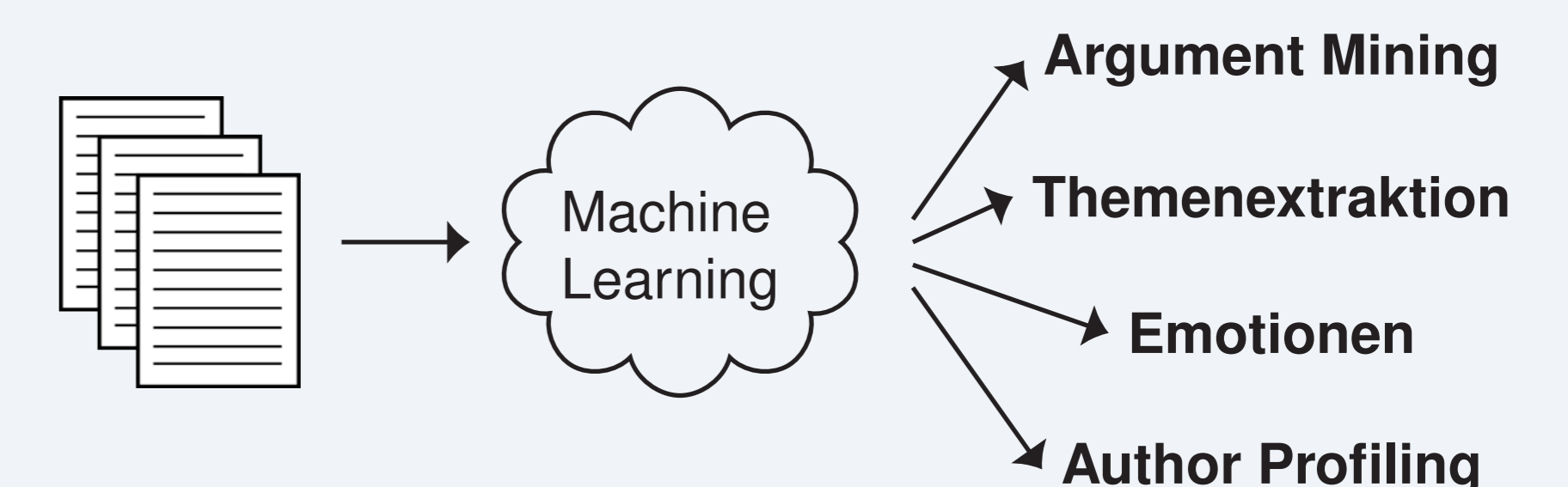
Motivation und zentrale Forschungsthemen

Motivation:

- Erfolgreiche Online-Partizipationsverfahren können tausende Textbeiträge haben.
- Die Benutzer und Organisatoren können schnell überfordert werden.
- Eine manuelle Auswertung ist sehr zeitaufwendig.
- ⇒ Einsatz von *natural language processing* um automatisiert in der Evaluation unterstützen zu können.

Unsere Forschung fokussierte sich auf drei Forschungsgebiete:

- **Argumentation:** Welche Vorschläge und Ideen werden von den Bürgern eingebracht? Welche Gründe führen die Bürger für die Umsetzung ihrer Vorschläge auf?
- **Themenextraktion:** Automatische Erstellung eines Überblicks über die Diskussionsthemen
- **Author Profiling:** Automatisierte Vorhersage demografischer Merkmale der Teilnehmer um unterrepräsentierte Gruppen identifizieren zu können



Ansatz und aktueller Stand

Argument Mining:

- Argumentationsmodell für Online-Partizipation und annotierter Datensatz (Liebeck et al. 2016)
- **Beispiel:**
 - Vorschlag: *Wir hoffen, dass Kinderspielplätze zukünftig mit dem THF-Gesetz im Aussenbereich vereinbar sind.*
 - Begründung: *Als Familienvater von 2 Kindern fehlt mir auf dem Feld der ein oder andere Kinderspielplatz.*
 - Positionierung: *Ich sehe das Anlegen von einfachen Spielplätzen eher kritisch und das obwohl ich selber Kinder habe.*
- Zwei Machine Learning-Aufgaben:
 - Subtask A: Argumentativen Textinhalt erkennen
 - Subtask B: Argumentationskomponenten klassifizieren
- Zwei Ansätze:
 - Klassisches Machine Learning mit Feature Engineering
 - Deep Learning
- Ergebnisse als macro-averaged F_1 :

	Klassisches ML	Deep Learning
Subtask A	69.71%	68.51%
Subtask B	67.20%	68.59%

- Zukünftige Arbeiten:
 - Sequence tagging
 - Argument linking

Themenextraktion:

- Neues Verfahren zur Grundformreduktion (z. B. *Kinderspielplätze* → *Kinderspielplatz*)
- Semantische textuelle Ähnlichkeit
- Zwei Ansätze zur Themenextraktion verfolgt:
 - Latent Dirichlet Allocation angewendet auf das *Tempelhofer Feld*:

Thema	Wörter
#1	Gärtner, Nutzung, Allmende, Beet, Raum
#2	Zugang, Kind, Spaß, Spielplatz, Nutzung
#4	Baseball, Softball, Amerikaner, Team, Trinkwasserbrunnen
#7	Harald Juhnke, Straße, Weg, Way, Zeit
#10	Musik, Kultur, Bühne, Wäldchen, Eingang
#11	Biergarten, Bier, Jahr, Tankstelle, Form

- Latent Dirichlet Allocation angewendet auf *Bonn 2011*:

Thema	Wörter
#1	Straße, Auto, Kontrolle, Parkplatz, Uhr
#2	Moderation, Teilnehmer, Liebe, Teilnehmerin, Buch
#3	Orchester, Beethovenhalle, Oper, Gebäude, Halle
#4	Schule, Kind, Stelle, Schüler, Kindergarten
#5	Kind, Musikschule, Familie, Eltern, Geld
#6	Stadthaus, Wasser, Rheinaue, Fahrt, Gewinn
#7	Bürger, Bad, Geld, Freibad, Verwaltung
#8	Verwaltung, Leistung, Dank, Haushalt, Beitrag
#9	Verein, Frau, Museum, Frauenmuseum, Kunst
#10	Ampel, Verkehr, Rahmen, Kreuzung, Rat
#11	Leistung, Geld, Einsparung, Kürzung, Beispiel
#12	Kürzung, Mensch, Angebot, Kind, Leistung
#13	Oper, Theater, Hund, Preis, Bürger
#14	Bus, Linie, Bahn, Innenstadt, Minute
#15	Karneval, Kultur, Mitarbeiter, Veranstaltung, Leute

- Neuer Ansatz für Topic Labeling, prototypische Implementierung
- Zukünftige Arbeit: open-source framework zur Themenextraktion, interaktivem Topic Modeling und Themenuisualisierungen

Ausgewählte Publikationen

- M. Liebeck, K. Esau, and S. Conrad. Text Mining für Online-Partizipationsverfahren: Die Notwendigkeit einer maschinell unterstützten Auswertung. HMD Praxis der Wirtschaftsinformatik 54.4. Schwerpunkt „Online Participation“, pp. 544-562, 2017
- K. Esau, M. Liebeck, and C. Eilders. Mining Arguments in Online Participation: Möglichkeiten und Grenzen manueller und automatisierter Inhaltsanalyse zur Erhebung von Argumentkomponenten. Polkomm 2017
- P. Modaresi, M. Liebeck, and S. Conrad. Exploring the Effects of Cross-Genre Machine Learning for Author Profiling in PAN 2016. CLEF 2016 Evaluation Labs and Workshop – Working Notes Paper, pp. 970-977, 2016
- M. Liebeck, K. Esau, and S. Conrad. What to Do with an Airport? Mining Arguments in the German Online Participation Project Tempelhofer Feld. Proceedings of the 3rd Workshop on Argumentation Mining, pp. 144-153, 2016

Betreuungsteam

- Prof. Dr. Stefan Conrad (Informatik, HHU)
- Prof. Dr. Martin Mauve (Informatik, HHU)
- Prof. Dr. Gerhard Vowe (Kommunikations- und Medienwissenschaft, HHU)
- Dr. Oliver Märker (Zebralog)