

Automated Discussion Analysis in Online Participation Projects

by Matthias Liebeck, M.Sc., Department of Computer Science, Heinrich Heine University Düsseldorf

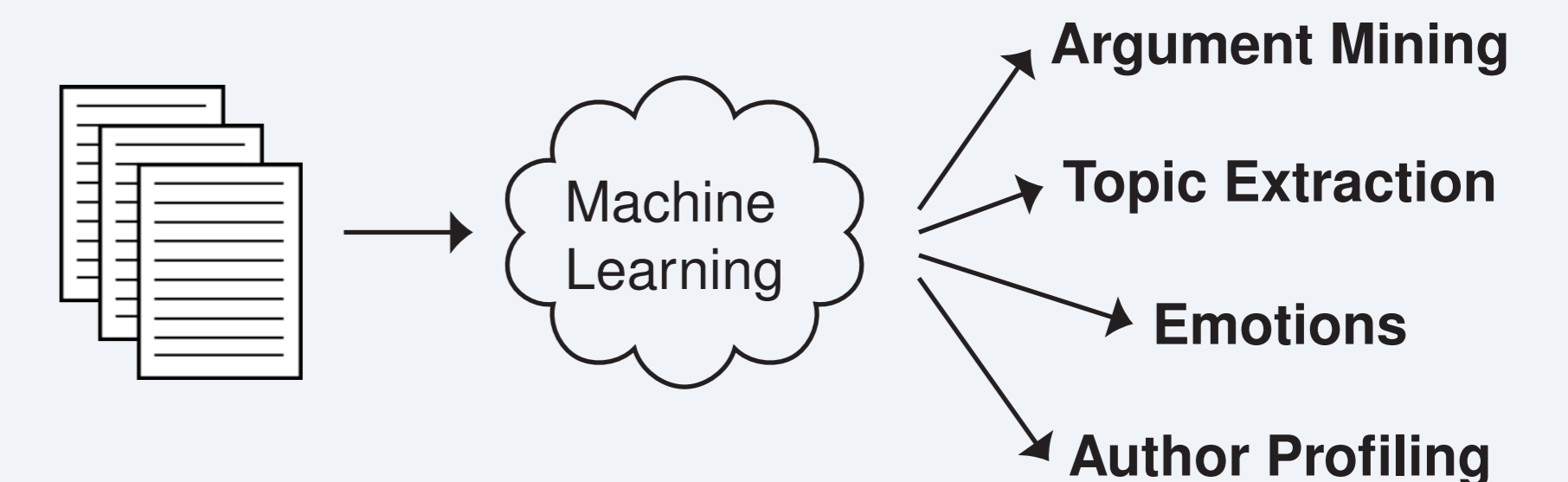
Motivation and Central Research Topics

Motivation:

- Successful online participation projects may have thousands of text contributions.
- The users and the organizers can easily be overwhelmed.
- A manual evaluation is very time consuming.
- ⇒ Use natural language processing to automatically assist in the evaluation.

Our research focused on three research topics:

- **Argumentation:** Which suggestions and ideas are contributed by the citizens? Which reasons do the citizens provide for the realization of their suggestions?
- **Topic Extraction:** Automatically extract an overview of the discussion topics
- **Author Profiling:** Predicting demographic attributes of users based on their text content in order to determine underrepresented groups



Approach and Current Status

Argument Mining:

- Argument model for online participation and an annotated dataset (Liebeck et al. 2016)
- Examples:
 - major position: *We hope that children's playgrounds in the open-air space will be compatible with the ThF law in the future.*
 - premise: *As a father of two children I am missing one or two children's playgrounds on the field.*
 - claim: *I am rather critical of the creation of simple playgrounds even though I have children myself.*
- Two machine learning tasks:
 - Subtask A: Identify argumentative text content
 - Subtask B: Classify argument components
- Two approaches:
 - classical machine learning with feature engineering
 - deep learning
- Results as macro-averaged F_1 :

	Classical ML	Deep Learning
Subtask A	69.71%	68.51%
Subtask B	67.20%	68.59%
- Future work:
 - Sequence tagging
 - Argument linking

Topic Extraction:

- New German lemmatizer (e.g., *Kinderspielplätze* ↦ *Kinderspielplatz*)
- Addressed textual semantic similarity
- Pursued two approaches to extract topics:
 - Latent Dirichlet Allocation applied on *Tempelhofer Feld*:

Topic Words

#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 applied on *Bonn 2011*:

Topic Words

#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

- New topic labeling approach, prototypical implementation
- Future work: open-source framework for topic extraction, interactive topic modeling, and topic visualizations

Selected Publications

- M. Liebeck, K. Esau, and S. Conrad. Text Mining für Online-Partizipationsverfahren: Die Notwendigkeit einer maschinell unterstützten Auswertung. *HMD Praxis der Wirtschaftsinformatik* 54.4. Schwerpunktft „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

Supervisor Team

- Prof. Dr. Stefan Conrad (computer science, HHU)
- Prof. Dr. Martin Mauve (computer science, HHU)
- Prof. Dr. Gerhard Vowe (communication and media studies, HHU)
- Dr. Oliver Märker (ZebraLog)