

Text Mining für Online-Partizipationsverfahren: Die Notwendigkeit einer maschinell unterstützten Auswertung

Matthias Liebeck¹ Katharina Esau² Stefan Conrad¹

¹ Institute of Computer Science, Heinrich Heine University Düsseldorf, Germany
{liebeck, conrad}@cs.uni-duesseldorf.de

² Institute of Social Sciences, Heinrich Heine University Düsseldorf, Germany
katharina.esau@uni-duesseldorf.de

This is an author-created version of the article.

The final publication is available at <http://link.springer.com/article/10.1365/s40702-017-0321-6>
DOI: 10.1365/s40702-017-0321-6

Zusammenfassung

Online-Partizipationsverfahren werden in den letzten Jahren vermehrt von Städten und Gemeinden eingesetzt, um ihre Bürger in politische Entscheidungsprozesse einzubeziehen. Der vorliegende Beitrag beginnt mit einer Kategorisierung von Online-Partizipationsverfahren im politischen Kontext in Deutschland und fokussiert auf das Beteiligungsprojekt Tempelhofer Feld in Berlin. Dazu werden die Probleme einer manuellen Auswertung und die Notwendigkeit einer maschinell unterstützten Auswertung von Textbeiträgen aus Partizipationsverfahren beschrieben.

Im Beitrag wird auf die Probleme und Lösungsmöglichkeiten in den drei Analysebereichen Argument Mining, Themenextraktion und Erkennung von Emotionen eingegangen. Für den Bereich Argument Mining wird ein geeignetes dreiteiliges Argumentationsmodell, welches auf das Online-Partizipationsverfahren Tempelhofer Feld der Stadt Berlin angewendet wird, diskutiert. Zudem wird der Einsatz von word embeddings als Features für eine Support Vector Machine zur automatisierten Klassifikation von Argumentationskomponenten evaluiert. Anschließend wird ein Einblick in das Aufgabengebiet der Themenextraktion, dessen Ziel die Erstellung eines groben Überblicks über die diskutierten Themen eines Online-Partizipationsverfahrens ist, gegeben und die Ergebnisse zweier Verfahren werden diskutiert. Danach erfolgt eine Diskussion über die Einsatzmöglichkeiten einer automatisierten Emotionserkennung im Kontext von Online-Partizipationsverfahren.

1 Online-Partizipationsverfahren in Deutschland

Als Online-Partizipationsverfahren werden internetgestützte, meist forenähnliche, Plattformen bezeichnet, in denen Menschen aktiv an einer Entscheidung, z. B. in Form von Diskussionen, Abstimmungen oder mit dem Äußern von Vorschlägen, teilhaben können. Dieser Beitrag fokussiert auf Online-Partizipationsverfahren im politischen Kontext in Deutschland, durch die Bürgerinnen und Bürger aktiv an politischen und administrativen Entscheidungsprozessen teilhaben können. In der Regel sind die Online-Partizipationsprozesse durch die Verwaltung der jeweiligen Stadt oder Gemeinde top-down organisiert. Sie zielen auf das Einbringen von Bürger-Feedback in politische Entscheidungsprozesse ab. Zum Beispiel können neue Vorschläge eingebracht werden, die dann von Politikern in Sitzungen des Gemeinderats besprochen werden können. Überregional setzen auch Parteien Online-Partizipationsverfahren ein, um die Meinungen von Bürgerinnen und Bürger zu berücksichtigen. Zum Beispiel hat die SPD solch ein Verfahren für Formulierungsvorschläge eines Leitantrags für den Bundesparteitag im Dezember 2011 genutzt.

Außerhalb der Politik werden Online-Partizipationsverfahren auch in anderen Kontexten eingesetzt. So kann Online-Partizipation beispielsweise auch im universitären Feld verwendet werden. Escher et al. (2017) beschreiben den Einsatz von Adhocracy¹ zur Neugestaltung der Promotionsordnung der Mathematisch-Naturwissenschaftlichen Fakultät an einer deutschen Universität. In einem mehrstufigen Prozess wurden

¹<https://github.com/liqd/adhocracy>

Änderungsvorschläge für die bisherige Promotionsordnung gemacht, die durch die Teilnehmer kommentiert werden konnten. Zusätzlich konnte über die Textbeiträge individuell abgestimmt (in Form von Zustimmung bzw. Ablehnung) werden. Zur Beteiligung waren alle Professoren, Postdocs und Doktoranden der Fakultät eingeladen, wobei 71 % der Personen die Plattform besucht und 29 % aktiv an dem Beteiligungsverfahren teilgenommen haben. Eingebrachte Vorschläge mit positivem Abstimmungsverhalten wurden in der Neuformulierung berücksichtigt, die vom Fakultätsrat angenommen wurde.

Darüber hinaus kann Online-Partizipation auch in Unternehmen eingesetzt werden, um die Meinungen der Mitarbeiter beispielsweise bei Veränderungen von Firmenrichtlinien oder beim Sammeln neuer Produktinnovationen zu berücksichtigen (Wagenknecht et al., 2016). In Deutschland wird Online-Partizipation inzwischen auch in Schulen, insbesondere im Rahmen des Projekts *Aula*², eingesetzt, um Schülerinnen und Schüler eine Mitgestaltungsmöglichkeit am schulischen Umfeld zu ermöglichen. Für den politischen Kontext in Deutschland beginnt die idealtypische Beteiligung eines Bürgers mit dem Informieren über den Hintergrund des Online-Partizipationsverfahren bzw. dessen inhaltlichen Gegenstand. Oftmals stellen die Verwaltungen hierzu entsprechende Hintergrundinformationen bereit, die dem Benutzer erklären, welche Form der Teilnahme auf der Plattform gewünscht ist und welche rechtlichen und finanziellen Rahmenbedingungen eingehalten werden müssen. Die genaue Beteiligungsform eines Bürgers ist verfahrensabhängig, wobei die Möglichkeiten, Textkommentare zu erstellen und andere Textbeiträge mit einem *upvote* oder *downvote* zu bewerten, zu den gängigen technischen Möglichkeiten der Diskussionsplattformen gehören. Handelt es sich bei der Plattform um ein Verfahren mit Vorschlagscharakter, das heißt, dass die Teilnehmer dazu angehalten sind, eigene Ideen in die Diskussion einzubringen, so würde ein Bürger zunächst vorhandene Beiträge lesen, überprüfen, ob bereits Punkte genannt wurden, die seinen Vorstellungen entsprechen, und diese mit einem *upvote* versehen. Falls der Benutzer stark gegenteilige Formulierungen vorfindet, so sollte er diese mit einem *downvote* versehen und in einem Kommentar sachlich darlegen, warum er anderer Meinung ist. Falls die Meinung des Benutzers nicht vertreten ist oder ihm während des Lesens weitere Ideen einfallen, so kann er einen neuen Textbeitrag eröffnen, in dem er die Diskussion erweitert.

Je nach Zielsetzung und Art des Beteiligungsverfahrens können Städte und Kommunen in einem Verfahren Umsetzungsvorschläge zur Diskussion einstellen. In Ludwigshafen konnten Bürgerinnen und Bürger 2014 anlässlich einer maroden und abrißbedürftigen Hochstraße in einem Online-Verfahren³ über vier vorab geprüfte Varianten zum Neubau (unter anderem Hochstraße vs. Stadtstraße) diskutieren. Die Teilnehmer konnten Kommentare zu sieben verschiedenen Aspekten (Bauzeit, Entwicklungschancen, Kosten, Nachbarn, Raumgefühl, Umwelt und Verkehr) abgeben und dabei abstimmen, welche der Varianten sie in Bezug auf den jeweiligen Aspekt bevorzugen. In dem Meinungsbild der Bürger wurde die Variante „Stadtstraße lang“ in allen Aspekten präferiert (Märker und Rucker, 2016). Das Verfahren war jedoch konsultativ angelegt und die Ergebnisse sind nicht bindend gewesen. Der Stadtrat hat im Anschluss die Bürgerkommentare in seinen Entscheidungsprozess einbezogen und letztlich dieselbe Variante für die Beantragung von Fördergeldern ausgewählt.

In der Stadt Köln sieht der geplante Verfahrensablauf aus Sicht der Verwaltung für den Bürgerhaushalt 2016⁴ vor, dass die Bürger online Vorschläge eingeben und diese diskutieren sowie bewerten. Anschließend werden die 25 bestbewerteten Vorschläge pro Stadtbezirk durch die Verwaltung ausgewertet und dem Rat zur Entscheidung vorgelegt (Behrendt, 2016).

Es gibt verschiedene Ansätze, politische Online-Partizipationsverfahren zu kategorisieren. Zunächst sind Angebote zum Anliegenmanagement, in denen beispielsweise defekte Beleuchtungen oder Straßennlöcher direkt an die Stadt bzw. Kommune gemeldet werden können, bei der folgenden Betrachtung nicht zu den Online-Partizipationsverfahren zu zählen, da bei ihnen lediglich eine Meldefunktion und keine Partizipation, beispielsweise in Form eines Diskurses mit anderen Bürgern, im Vordergrund steht. Eine Kategorisierung nach Kubicek et al. (2007) führt eine inhaltliche Gruppierung in die Themen *räumliche Planung*, *Finanzen/Haushalt* und *Sozialpolitik* auf, wobei es noch eine weitere Kategorie für *themenbereichsübergreifende Prozesse* gibt. Die als Monitor Online-Partizipation (Gladitz et al., 2017) bezeichnete umfangreiche Befragung in Nordrhein-Westfalen bietet einen systematischen Überblick über Online-Partizipationsverfahren, bei denen Bürger von Städten und Gemeinden die Zielgruppe waren. Bei den 396 befragten Städten und Gemeinden konnte eine Rücklaufquote von 94 % erreicht werden. In der Umfrage wurden die Verfahren in die Kategorien *Bauleitplanung*, *Bürgerhaushalt*, *Stadtentwicklungsplanung*, *Lärmaktionsplanung* und *Sonstiges* unterschieden, wobei zusätzlich erfragt wurde, ob die Kommunen bereits Plattformen zum *Mängelmanagement* eingesetzt haben. Als wichtige Erkenntnisse sind hervorzuheben, dass 32 % der Verfahren Bürgerhaushalte sind, 90 von 130 Städten und Gemeinden bereits mehr als ein Online-Verfahren durchgeführt haben und 79 % der Verfahren darauf abzielen, Ideen von den Bürgern

²<http://aula-blog.website>

³<https://ludwigshafen-diskutiert.de>

⁴<https://buergerhaushalt.stadt-koeln.de/2016>

zu sammeln. Deutschlandweit haben bereits über 460 Kommunen Beteiligungsverfahren zum Haushalt durchgeführt.⁵

Die Regierung Nordrhein-Westfalens verfolgt seit 2014 die sogenannte *Open.NRW-Strategie*, die auf den drei Säulen *Open Data*, *E-Partizipation* und *E-Zusammenarbeit* aufbaut (Beuß, 2017). Die Landesregierung will dadurch die Digitalisierung der Gesellschaft auch in der Verwaltung verankern. Hinter der Säule *Open Data* steht die Absicht, dass die in der Verwaltung erzeugten Daten transparent und möglichst kostenlos für alle Interessierten in maschinell lesbarer Form im *Open.NRW-Portal*⁶ bereitgestellt werden sollen. Unter *E-Partizipation* wird der Einsatz von Online-Partizipationsverfahren zur Einbindung von Bürgern in den politischen Entscheidungsprozess für konkrete Fragestellungen bezeichnet. Als *E-Zusammenarbeit* bezeichnet die *Open.NRW-Strategie* die bessere Vernetzung von Verwaltungsmitarbeitern, die Erstellung fachspezifischer Expertennetzwerke und den Einsatz von elektronischem Anliegenmanagement. Im Rahmen unserer Forschung ist vor allem die Säule der *E-Partizipation* von großer Bedeutung.

Die *Open.NRW-Strategie* wird als die „freiwillige Selbstverpflichtung der Landesregierung, die Ziele des *Open Government* umzusetzen“ (Open.NRW, 2016) bezeichnet und hält damit die Städte und Gemeinden in NRW dazu an, ihre Bürger mittels Online-Partizipationsverfahren in politische Entscheidungsprozesse einzubeziehen. Eine detaillierte Auswertung von Online-Partizipationsverfahren ist unabdingbar, damit die Meinungen aller Bürger berücksichtigt werden können. Dies sieht die Landesregierung ähnlich, da sie im Sachstandsbericht zur Umsetzung der *Open.NRW-Strategie* (Open.NRW, 2016) definiert, dass „nicht die Anzahl der Rückmeldungen [...] das Verfahren erfolgreich [macht], sondern die Expertise der Teilnehmer und die inhaltliche Qualität ihrer Kommentare“.

Seit dem 8. Juli 2016 gilt in NRW das neue *Gesetz zur Förderung der elektronischen Verwaltung in Nordrhein-Westfalen (E-Government-Gesetz Nordrhein-Westfalen - EGovG NRW)*. Für Online-Partizipationsverfahren wird in §18 Abs. 2 EGovG NRW festgehalten: „Die Resultate der elektronischen Öffentlichkeitsbeteiligung sind von der zuständigen Behörde auszuwerten und zu prüfen. Die Ergebnisse durchgeführter Beteiligungen sind bekannt zu geben.“ Somit sind die einzelnen Städte und Gemeinden zusätzlich zur Auswertung der Online-Partizipationsprojekte gezwungen.

Zwei gängige Probleme bei Online-Beteiligungsverfahren sind, dass entweder zu wenig Bürgerinnen und Bürger teilnehmen oder dass die Beteiligungsrate so hoch ist, dass eine extrem große Anzahl an manuell auszuwertenden Textbeiträgen entsteht, deren Auswertung sehr zeitintensiv werden kann.

Beobachtungen der Praxis zeigen, dass die Beteiligungsraten an Bürgerhaushalten bei erstmaliger Durchführung in einer Stadt hoch sind und mit der Zeit abnehmen, wie beispielsweise von 2008 bis 2015 in Köln (Behrendt, 2016). Ein möglicherweise beeinflussender Faktor könnte die Enttäuschung derjenigen Bürgerinnen und Bürger sein, die nicht umgesetzte Vorschläge eingereicht haben. Der *Monitor Online-Partizipation* (Gladitz et al., 2017) verzeichnet in Nordrhein-Westfalen einen Rückgang in Bezug auf die Anzahl der durchgeführten Bürgerhaushalte. Die Bereitstellung dedizierter finanzieller Ressourcen für die Umsetzung der Bürgervorschläge könnte ein weiterer Faktor für die Beteiligungsrate sein. Im Beispiel Köln wurde 2016 für den Bürgerhaushalt ein Budget von 100.000 Euro pro Stadtbezirk zur Verfügung gestellt, das zur Umsetzung der Vorschläge verwendet wird. Die Beteiligungsrate gegenüber dem Vorjahr konnte in Köln um 56 % gesteigert werden (Behrendt, 2016). Als ein Beispiel für ein Verfahren mit großer Beteiligungsrate sei der Bürgerhaushalt 2011 in Bonn⁷ mit mehr als 14 000 Beiträgen zu nennen. Die Auswertungen von Partizipationsverfahren erfolgen prinzipiell durch die Verwaltung oder durch externe Dienstleister, was zum Einsatz von personellen oder finanziellen Ressourcen führt. Für Verfahren mit einer hohen Anzahl an Textbeiträgen gewinnt eine automatisierte Textanalyse, die durch verschiedene Ansätze eine Voranalyse bzw. Aufbereitung der Textbeiträge liefern kann, daher immer mehr an Bedeutung. Derartige Analysen können zu einer Arbeitsreduktion für die Personen führen, die die qualitative Auswertung manuell vornehmen.

Im Folgenden wird auf drei Analysebereiche von Online-Partizipationsverfahren eingegangen: Kapitel 2 widmet sich dem Bereich *Argument Mining*, der automatisierten Extraktion von Argumentationskomponenten. Danach wird in Kapitel 3 der Analysebereich der Themenextraktion vorgestellt und es werden zwei Verfahren zur Themenextraktion auf die Diskussionsbeiträge eines Online-Partizipationsverfahrens angewendet. Anschließend führt Kapitel 4 in die Anwendungsmöglichkeiten der automatisierten Emotionserkennung im Kontext von Online-Partizipationsverfahren ein.

⁵<http://www.buergerhaushalt.org>

⁶<https://open.nrw>

⁷<http://bonn-packts-an-2011.de>

2 Argument Mining

Für die Politiker bzw. Ratsmitglieder ist es wichtig zu wissen, welche Themen die Bürger in dem Portal besprechen und wie sie ihre Meinungen argumentativ darlegen. Falls es sich dabei um ein Online-Partizipationsverfahren mit Vorschlagscharakter handelt, enthält eine optimale Auswertung die Vorschläge bzw. Änderungswünsche der Bürger und die dazugehörigen Begründungen, warum die Vorschläge umgesetzt werden sollen. Das Meinungsbild wird durch Zusammenfassungen der Kommentare der anderen Diskussionsteilnehmer vervollständigt, die insbesondere weitere Begründungen und Stellungnahmen (z. B. zustimmende oder kritisierende) beinhalten. Dabei ist zu beachten, dass die Meinung jedes Bürgers gleich wichtig ist; dies muss bei der Auswertung berücksichtigt werden. Folglich sollte eine finale Auswertung von Verfahren schlussendlich immer von Menschen durchgeführt werden, damit etwaige Klassifikationsfehler bei einem automatisierten Machine Learning-Verfahren korrigiert werden können. Dennoch sollten solche Verfahren eingesetzt werden, um die Arbeitslast durch eine maschinelle Vorarbeit zu reduzieren.

Der vorliegende Beitrag fokussiert auf das Online-Partizipationsprojekt „*Tempelhofer Feld*“⁸ aus Berlin. Das Projekt beschäftigt sich mit dem Gelände des ehemaligen Flughafens Berlin Tempelhof, der den Flugbetrieb im Jahre 2008 eingestellt hat. Die dadurch mitten in Berlin entstandene etwa 300 Hektar große Freifläche trägt den Namen *Tempelhofer Feld*, dient den Besuchern vor allem zur Erholung und Freizeitgestaltung und trägt als Kaltluftentstehungsgebiet zur Luftqualität Berlins bei. Um diesen Charakter zu erhalten, wurde 2014 ein Gesetz, das sogenannte *ThF-Gesetz*, erlassen, das die Weiträumigkeit des Feldes schützt, indem es Neubauten und Erweiterungen an bestehenden Gebäuden untersagt. Das Ziel der top-down organisierten Online-Plattform war die Sammlung von Vorschlägen und Ideen, die das Feld, unter Einhaltung des ThF-Gesetzes, für die Besucher verbessern können. Der Zeitraum von November 2014 bis März 2015 war dafür vorgesehen, Vorschläge in den sieben Bereichen *Bewirtschaftung*, *Erinnerung*, *Freizeit*, *I love THF*, *Mitmachen*, *Natur* und *Was vergessen?* einzureichen. Insgesamt wurden 340 Vorschläge (so genannte *Proposals*) gemacht, die mit knapp 1400 Kommentaren diskutiert wurden, was eine große Menge an auszuwertendem Textmaterial darstellt. Anders als bei den Beteiligungsverfahren zum Bürgerhaushalt, erfolgte die Zusammenfassung der Beiträge durch mehrere Offline-Veranstaltungen in Form von thematischen Werkstätten, in denen das Textmaterial durch engagierte Bürger zusammengefasst wurde.

Aus der bisherigen Forschung im Bereich der Argumentationstheorie hat sich in den letzten Jahren das Forschungsfeld Argument Mining entwickelt, das sich insbesondere mit der automatisierten Extraktion von Argumentationskomponenten aus Texten beschäftigt. In den letzten Jahren sind vermehrt Publikationen zu den größten Computerlinguistik-Konferenzen rund um die *Association for Computational Linguistics (ACL)* erschienen und inzwischen gibt es seit vier Jahren unter dem Dach der ACL einen jährlichen Workshop mit dem Namen *Argument(-ation) Mining*. Argument Mining wird in verschiedenen Anwendungsgebieten, beispielsweise in juristischen Texten (Palau und Moens, 2009), Essays (Stab und Gurevych, 2014), Nachrichten (Eckle-Kohler et al., 2015) oder Social Media (Addawood und Bashir, 2016) angewendet. Die bisherige Forschung umfasst mehrere verschiedene Textsprachen, wobei die englische und die deutsche Sprache von den meisten Arbeitsgruppen erforscht werden. Die für maschinelle Lernverfahren verwendeten Features der bisherigen Arbeiten sind großteils sprachunabhängig und können daher auch für andere Sprachen eingesetzt werden. Das Hauptproblem ist momentan jedoch die geringe Anzahl an annotierten Korpora, da deren Erstellung, insbesondere die Erstellung eines Codebuchs für ein geeignetes Annotationsschema, die Schulung der Codierer und der Codierprozess, sehr ressourcenaufwändig ist. Für einen detaillierteren Überblick zum aktuellen Forschungsstand im Bereich Argument Mining wird der Leser auf Habernal und Gurevych (2016) verwiesen.

Für die automatisierte Analyse von Online-Partizipationsverfahren mit Vorschlagscharakter bedarf es eines passenden Argumentationsschemas, das auf Textbeiträge mit dem Ziel, aus diesen automatisiert Argumentationskomponenten zu extrahieren, angewendet wird. Dabei sollte die Wahl des Argumentationsschemas durch praxisnahe Fragestellungen motiviert sein, um einen Praxiseinsatz zu ermöglichen. Liebeck et al. (2016) beschreiben dafür drei Fragestellungen, die sich in dem Argumentationsschema wiederfinden lassen: (i) Was sind die von den Bürgern gemachten Vorschläge über deren Umsetzung die Politiker entscheiden können? (ii) Was sind Gründe für und Gründe gegen die Umsetzung dieser Vorschläge? (iii) Wie viele Bürger drücken eine Pro- oder Contra-Haltung gegenüber den Vorschlägen aus?

Liebeck et al. (2016) führen dazu ein drei- bzw. vierteiliges Argumentationsschema für solche Beteiligungsverfahren ein, welches aus *major positions* (Vorschlägen), *premises* (Begründungen) und *claims* (Positionierungen), die sich in *pro claims* (Pro-Positionierungen) und *contra claims* (Contra-Positionierungen) aufteilen lassen, besteht. Das Argumentationsmodell und die Relationen zwischen den einzelnen Argumentationskomponenten ist in Abbildung 1 visualisiert.

⁸<https://tempelhofer-feld.berlin.de>

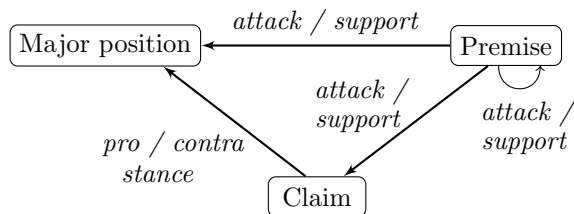


Abbildung 1: Argumentationsmodell für politische Online-Partizipation aus Liebeck et al. (2016)

Um das Argumentationsmodell und die einzelnen Argumentationskomponenten besser verstehen zu können, dienen das folgende Proposal⁹ und ein dazu abgegebener Kommentar als Beispiel. Das Diskussthemema ist die explizite Einführung eines Bereichs für motorisierte Geräte auf dem Tempelhofer Feld. Im Proposal wird ein Vorschlag gemacht und eine unterstützende Begründung geliefert. Im darauffolgenden Kommentar wird Zustimmung geäußert und auch eine Begründung geliefert. Danach wird zudem ein neuer Vorschlag gemacht, welcher direkt mit einem Argument für die Umsetzung untermauert wird.

Proposal:

[*major position*: Ein Bereich in dem motorisierte Geräte (Modellflugzeuge, -autos usw.) benutzt werden können, im Rest des Geländes dann eben nicht.] [*premise*: Weil diese Geräte Lärm machen und evtl. zum Photographieren und Filmen (Beobachten anderer) genutzt werden können.]

Kommentar:

[*claim pro*: kleine Elektrodrohnen just for fun, warum nicht.] [*premise*: Verhindern kann man es eh nicht, ein Verbot nützt also nichts.]

[*major position*: Irgendwelche großen Brummer oder lautere Modellflieger am besten konzentriert dort wo sie es jetzt auch schon tun: Nordstartbahn Westseite, dort ist die Tempelhofer-Damm-Lärmseite,] [*premise*: da stört es kaum und die Freunde der fliegenden Objekte können sich treffen, fachsimpeln und ihre neueste Technik testen und herzeigen.]

Da im Partizipationsprojekt Tempelhofer Feld Zustimmungen und Ablehnungen zu bzw. von Vorschlägen häufig textuell ausgedrückt werden, ist es notwendig, dass diese Zustimmungen und Ablehnungen im Argumentationsschema berücksichtigt werden.

Wir verwenden im Folgenden den frei verfügbaren Datensatz aus Liebeck et al. (2016), der als *THF Airport ArgMining Corpus* bezeichnet wird. Die in dem Datensatz annotierten Argumentationskomponenten haben unterschiedliche Längen. Sie können entweder satzübergreifend sein, ganze Sätze umfassen oder nur Satzteile. Wir folgen dem vorgeschlagenen Evaluationsaufbau, der sich zunächst auf Annotationen auf Satzebene beschränkt und die Argumentationskomponenten pro claim und contra claim als claim zusammenfasst. Neben diesen Annotationen enthält der Datensatz auch Textteile, die keine Annotationen gemäß dem Argumentationsschema erhalten haben. Diese werden als *non-argumentative* bezeichnet und umfassen beispielsweise Begrüßungen, Schlussformeln oder Verständnisfragen. Da ein Kommentar sowohl argumentative Passagen als auch nicht-argumentative Passagen enthalten kann, ist es sinnvoll, die entsprechenden Textpassagen genau zu codieren.

Der Datensatz besteht aus 72 Proposals und 575 Kommentaren, deren argumentative Textteile jeweils gemäß dem Argumentationsschema durch drei Codierer codiert wurden. Bei einer Probecodierung auf einer Testmenge, bestehend aus 82 Textbeiträgen, wurde eine Intercoder-Reliabilität von 92.4% Krippendorffs α_u für argumentative bzw. nicht-argumentative Textpassagen und 78% Krippendorffs α_u für die Codierung der vier Annotationskomponenten erreicht.

Die Klassifikation von Argumentationskomponenten auf Satzebene lässt sich als mehrstufiger Klassifikationsprozess betrachten, der in Abbildung 2 visualisiert ist. In *Subtask A* wird zunächst entschieden, ob ein Satz argumentativ oder nicht-argumentativ ist. Argumentative Sätze werden danach in *Subtask B* weiter in major positions, claims und premises unterschieden. Dabei werden momentan pro claims und contra claims zusammengefasst, da es nur wenige contra claims im Datensatz gibt und dadurch ein 4-Klassen-Problem deutlich schwieriger zu lösen ist als ein reduziertes 3-Klassen-Problem. Die genaue Trennung in vier Argumentationskomponenten kann schließlich in *Subtask C* erfolgen, in dem die claims wieder in pro

⁹<https://tempelhofer-feld.berlin.de/i/tempelhofer-feld/proposal/253>

claims und contra claims klassifiziert werden. Es ist anzumerken, dass die drei zum Einsatz kommenden Klassifikatoren jeweils individuell trainiert werden müssen und dabei auch unterschiedliche Features und unterschiedliche Klassifikatoren pro Klassifikationsproblem verwendet werden können.

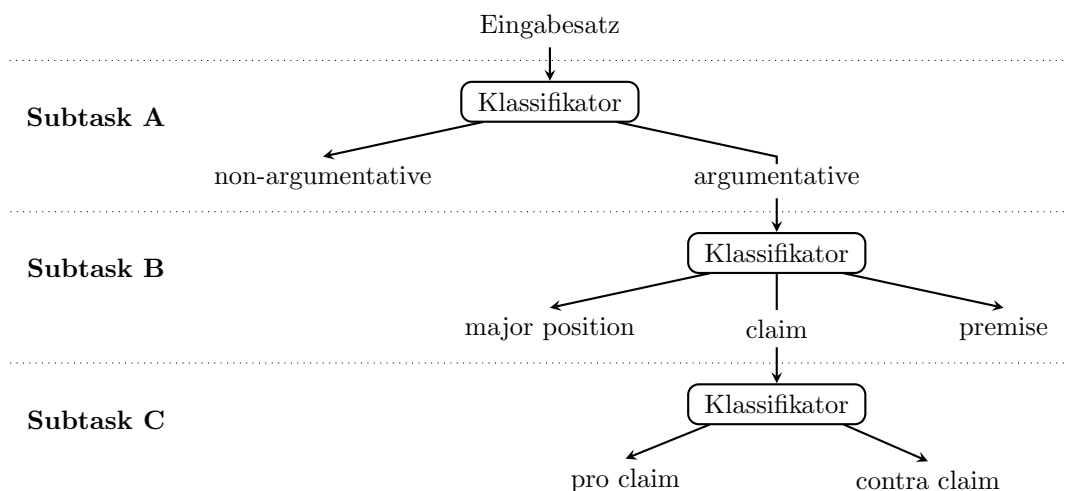


Abbildung 2: Mehrstufiger Klassifikationsprozess für einen Eingabesatz

Im Versuchsaufbau von Liebeck et al. (2016) werden für Subtask A und Subtask B unterschiedliche Mengen an Sätzen betrachtet. In Subtask A wird zwischen Sätzen ohne codierte Argumentationskomponenten und Sätzen mit mindestens einer codierten Argumentationskomponente unterschieden. Für Subtask B werden nur die Sätze betrachtet, in denen exakt eine Argumentationskomponente codiert wurde, um eine Single-Label-Klassifikation zu ermöglichen.

Um Textbeiträge für maschinelle Lernverfahren verwenden zu können, müssen die Textbeiträge zunächst mittels einer *Natural Language Processing Pipeline* aufbereitet werden. Dazu werden die Textbeiträge der Reihe nach durch mehrere Schritte mit weiteren Informationen angereichert: Zunächst zerlegt ein Tokenizer jeden Textbeitrag in einzelne Wörter. Anschließend werden durch einen *Sentence Splitter* automatisch Satzgrenzen erkannt, um den Text in Sätze zu zerlegen. Danach kann auf Satzebene für jedes einzelne Wort die entsprechende Wortart durch einen *POS-Tagger* bestimmt werden. Da im THF Airport ArgMining Corpus die einzelnen Datenpunkte bereits Sätze sind, reicht in unserem Fall und den hier betrachteten Features eine Tokenisierung und die Bestimmung der Wortarten durch *OpenNLP*¹⁰ aus.

Im Rahmen dieses Beitrags werden *Unigramme*, *word embeddings* und deren Kombination als Features für ein maschinelles Lernverfahren, genauer genommen für eine *Support Vector Machine (SVM)* (Cortes und Vapnik, 1995), betrachtet.

Als Unigramm wird eine Häufigkeitsverteilung über die auftretenden Wörter in einem Text bezeichnet. Um eine vergleichbare Repräsentation zwischen mehreren Texten zu ermöglichen, wird eine feste Vektorlänge verwendet, indem sich die Häufigkeitsverteilung auf ein Vokabular, d. h. die Menge aller Wörter in den Trainingsdaten, bezieht. In unserem Fall wird die Verteilung über die Wörter eines Satzes betrachtet.

Word embeddings (Mikolov et al., 2013) beschreiben ein Modell zur semantischen Einbettung von Wörtern in einen Vektorraum mit fester Dimension k . Das Ziel beim Erstellen dieser Einbettung ist, dass semantisch ähnliche Wörter näher beieinander liegen als semantisch unähnliche Wörter. Üblicherweise wird die Distanz zweier eingebetteter Wörter über das Kosinus-Maß berechnet. Da word embeddings in den letzten Jahren in der Computerlinguistik immer mehr an Bedeutung gewonnen haben, werden sie im Rahmen dieses Beitrags für den Einsatz im Argument Mining in Bezug auf den THF Airport ArgMining Corpus evaluiert. Die dafür benötigten word2vec-Modelle (Mikolov et al., 2013) wurden mittels *gensim* (Řehůřek und Sojka, 2010) auf den Textinhalten der deutschen Version von Wikipedia, auf dem Dump vom 20.11.2016, trainiert. Als Dimension der word embeddings wurden Evaluationen mit $k \in \{100, 200, 300\}$ durchgeführt, welche gängige Größen für die embeddings sind, die in vielen Publikationen verwendet werden. Wir verwenden sie in unseren Experimenten, um die Auswirkungen der Dimensionen auf die Klassifikationsergebnisse vergleichen zu können. Für einen konkreten Korpus kann der Wert zwar noch genauer optimiert werden, jedoch wird dies in der Regel nicht gemacht, um overfitting zu vermeiden. Um die unterschiedliche Anzahl an Wörtern pro Satz und damit eine unterschiedliche Anzahl an möglichen word embeddings pro

¹⁰<https://opennlp.apache.org>

Features	Subtask A			Subtask B		
Unigramme	64,71			64,3		
	$k = 100$	$k = 200$	$k = 300$	$k = 100$	$k = 200$	$k = 300$
Embeddings	61,58	64,50	64,23	60,00	60,80	62,09
Unigramme + Embeddings	67,50	67,03	67,38	55,10	64,58	54,65

Tabelle 1: Durchschnittliche F_1 -Werte in Prozent für die verschiedenen Features zur Klassifikation von Argumentationskomponenten

Satz auf eine feste Länge zu bringen, wird üblicherweise, und auch in dieser Arbeit, pro Satz der Zentroid der verwendeten word embeddings als Repräsentant verwendet.

Die Evaluation der Features erfolgt unter Verwendung des Machine Learning Frameworks *scikit-learn* (Pedregosa et al., 2011). Um Werte für die Parameter der Features und für die Klassifikatoren zu bestimmen, erfolgt eine Gittersuche per zehnfacher Kreuzvalidierung auf dem Trainingsdatensatz. Anschließend wird pro Subtask und Feature jeweils die beste Parameterkonstellation zum Trainieren des Klassifikators verwendet, der im Anschluss auf die Testdaten angewendet wird. Als Evaluationsmaß wird das F_1 -Maß genutzt, welches das harmonische Mittel aus *precision* und *recall* darstellt. Da *precision* und *recall* pro Klasse berechnet werden, bietet sich an, einen Mittelwert über alle F_1 -Werte, den sogenannten *macro-averaged F_1 score*, zu betrachten. Die Klassifikationsergebnisse für Subtask A und B sind in Tabelle 1 aufgeführt.

Die Tabelle zeigt, dass word embeddings alleine immer schlechtere Ergebnisse liefern als Unigramme. Für Subtask A liefert die Kombination aus Unigrammen und word embeddings bessere Ergebnisse als Unigramme alleine. Für Subtask B liefert die Hinzunahme von word embeddings in zwei von drei Fällen deutlich schlechtere Ergebnisse. Ein Blick auf die beiden entsprechenden Konfusionsmatrizen zeigt, dass viele major positions fälschlicherweise als premise erkannt werden. Für $k = 200$ werden zwar leicht bessere Ergebnisse als für Unigramme erreicht, jedoch weisen die Ergebnisse hinsichtlich $k = 100$ und $k = 300$ nicht genug Stabilität auf.

3 Themenextraktion

Eine Themenextraktion kann in Online-Partizipationsverfahren mehrere Aufgaben lösen. Die dabei wichtigste Aufgabe ist eine automatisierte Erstellung eines groben Überblicks über die Diskussionsthemen des Verfahrens. Dies ist nach Abschluss des Verfahrens und auch schon zur Laufzeit sinnvoll, da der Verwaltung während eines laufenden Verfahrens in einem Dashboard stets eine Themenübersicht angezeigt werden kann. Solch einer Übersicht kann beispielsweise zu entnehmen sein, dass sich 30 % der Beiträge mit dem Thema Natur und 10 % mit dem Thema Verkehr beschäftigen und zwar ohne, dass diese Kategorien im Vorfeld definiert werden müssen. Die Erzeugung solch einer Übersicht ist per Hand bereits ab mehreren hundert Beiträgen schwer und bei mehreren tausend Beiträgen nicht mehr praktikabel umsetzbar, da die von den Bürgern diskutierten Themen im Laufe des Verfahrens gewechselt werden können und nicht a priori bekannt sind. Eine Themenextraktion ermöglicht der Verwaltung und den Politikern die Wahrnehmung *latenter* Diskussionsthemen. Wird bei der Analyse eine zeitliche Komponente der Beiträge berücksichtigt, so können während eines Verfahrens Verschiebungen der Themenfelder wahrgenommen werden. Für wiederkehrende Online-Partizipationsverfahren ist es durch den Einsatz solcher Techniken ebenfalls möglich, die Entwicklung der Diskussionsthemen über Jahre hinweg zu vergleichen.

Falls auf einer Plattform Diskussionsbereiche wie, beispielsweise, *Sport*, *Natur* und *Verkehr* vorgegeben sind, so existiert ein weiteres Einsatzszenario für die Themenextraktion in Form einer Moderationshilfe: Mit Techniken des Textclustering kann für jeden Beitrag überprüft werden, ob er möglicherweise vom Autor im falschen Diskussionsbereich erstellt wurde. Wird dies automatisiert erkannt, so kann ein Moderator darüber informiert werden. Solch ein Ansatz setzt jedoch eine gewisse Mindestmenge an Beiträgen in den jeweiligen Diskussionsbereichen voraus, um aus diesen „normalen“ Beiträgen gängige Diskussionsthemen zu extrahieren bzw. zu lernen.

Im Folgenden wird die Erstellung einer Themenübersicht aus Online-Partizipationsverfahren detaillierter behandelt. Als Grundlage dienen dafür die von den Benutzern verwendeten Wörter in den Textkommentaren. Zunächst sei noch angemerkt, dass die Verfahren technisch so angelegt sind, dass sie darauf abzielen, doppelte Inhalte zu vermeiden, indem die Benutzer ihre Zustimmung mit einem upvote und ihre Ablehnung mit einem downvote ausdrücken sollen. Wird beispielsweise vorgeschlagen, einen Spielplatz zu bauen, und 70 Bürger finden diesen Vorschlag gut, so reicht es aus, wenn ein Bürger diesen Vorschlag textuell einreicht

und die anderen Bürger ihre Zustimmung per upvote signalisieren, ohne dass sie dabei textuell das Wort Spielplatz erneut verwenden müssen. Wir treffen die Annahme, dass Beiträge mit einer hohen Beteiligung an Stimmabgaben einen höheren Diskussionsanteil haben als Beiträge ohne Votings und ohne Kommentare, die möglicherweise in der Masse an Textbeiträgen untergegangen sind. Dementsprechend wird im Folgenden die Anzahl an abgegebenen Stimmen als Gewichtung für die Termfrequenzen der auftretenden Wörter verwendet.

Bei der Themenextraktion ist zu überlegen, welche Wörter bzw. Wortarten die meisten semantischen Informationen tragen. Es bietet sich an, Wörter ohne semantische Informationen, sogenannte Stoppwörter wie Artikel, Konjunktionen und Präpositionen, bei der Extraktion herauszufiltern. Dazu kommen datensatzabhängige Stoppwörter, die keine weiteren genaueren Informationen beinhalten. In der nachfolgenden Betrachtung des Tempelhofer Feldes werden für die Themenextraktion nur Nomen und Eigennamen verwendet und allgemeine verfahrensspezifische Wörter, unter anderem *Tempelhof*, *Feld*, *Idee*, *Vorschlag* und *Beispiel*, ausgeschlossen, um die Ergebnisse subjektiv zu verbessern. Eine Möglichkeit, solche Wörter automatisiert zu ermitteln, ist die Suche nach Wörtern, die in fast allen Dokumenten auftauchen.

Bei der Ermittlung von Wortfrequenzen ist eine Lemmatisierung bzw. Grundformreduktion sinnvoll, um verschiedene Flexionsformen eines Wortes als Auftreten seiner Grundform zählen zu können. So werden beispielsweise die Wörter *Bereiche*, *Bereichs* und *Bereich* auf die Grundform *Bereich* zurückgeführt. Zur Lemmatisierung der deutschsprachigen Texte wird IWNLP (Liebeck und Conrad, 2015) eingesetzt, welches Flexionsformen aus der deutschen Version von Wiktionary extrahiert und als Abbildung auf ihre jeweiligen Grundformen anbietet.

Eine der einfachsten Formen der Themenextraktion ist die Betrachtung der am häufigsten auftretenden Wörter. In Abbildung 3 werden die 20 häufigsten Begriffe aus den fünf Bereichen *Bewirtschaftung*, *Erinnerung*, *Freizeit*, *Mitmachen* und *Natur* als Wortwolke visualisiert.¹¹ Ein Nachteil dieser Darstellungsform ist, dass Wortwolken keine semantische Trennung der Begriffe ermöglichen. Darüber hinaus können in einer Wortwolke möglicherweise nur die Wörter eines einzigen Diskussionsthemas auftreten, falls dieses die Diskussion dominiert und andere Diskussionsthemen weniger häufiger besprochen werden.



Abbildung 3: Wortwolke aus den 20 häufigsten Begriffen

Eine Möglichkeit der Extraktion von semantisch getrennten Themenbereichen ist der Einsatz von Verfahren aus dem Bereich Topic Modeling. Davon betrachten wir im Folgenden das Verfahren *Latent Dirichlet Allocation* (LDA) (Blei et al., 2003), welches ein generatives probabilistisches Modell ist und zur Bestimmung latenter Themen in Textkollektionen verwendet werden kann. Dabei werden die Annahmen getroffen, dass es k verschiedene Themen gibt und jedes Textdokument zu mehreren Themen gehört. Ein Thema ist dabei als eine Wahrscheinlichkeitsverteilung über die Wörter eines festen Vokabulars, d. h. über die verschiedenen Wörter der Textdokumente, definiert. Je nach Topic Modeling-Verfahren ist die Anzahl der verschiedenen Themen als Parameter anzugeben; im Fall von LDA wird der Parameter k genutzt. Die Wahl des Parameters lässt sich empirisch ermitteln, indem unterschiedliche Ergebnisse für mehrere Parameterwerte berechnet werden und im Anschluss das beste subjektive Ergebnis ausgewählt wird. Es gibt jedoch auch Forschungen (Griffiths und Steyvers, 2004) zu der Frage, wie die Anzahl der zu extrahierenden Themen automatisiert evaluiert werden kann.

In Tabelle 2 werden auszugsweise die Ergebnisse einer Ausführung von LDA auf dem Tempelhofer Feld angezeigt. Als Anzahl der Themen wurde $k = 15$ Themen empirisch ermittelt. Die durch LDA ermittelten Themengebiete spiegeln einige Diskussionsbereiche und Vorschläge wider. So wird in Thema #1 für die Allmende-Gärten, einem Gemeinschaftsgartenprojekt auf dem Tempelhofer Feld, die Thematik der Garten-nutzung angesprochen. Thema #2 behandelt die Verwendung des Feldes für Kinder, Thema #4 die Nutzung

¹¹Die anderen beiden Bereiche wurden in der Betrachtung ausgeschlossen, da sie überwiegend Metadiskussionen enthalten.

für Baseball und Softball und Thema #7 umfasst den Vorschlag, einen Weg zu Ehren des Schauspielers Harald Juhnke auf dem Feld anzulegen. Es ist jedoch auch anzumerken, dass bei einigen extrahierten Themen der semantische Zusammenhang eher gering erscheint und diese Themen daher nicht in der Tabelle aufgeführt sind.

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

Tabelle 2: Auszug aus den von der Latent Dirichlet Allocation identifizierten Themen und den dazugehörigen Wörtern

Die hier aufgeführten Themen werden zwar noch nicht durch ein einzelnes Wort beschrieben, ermöglichen aber dennoch einen Überblick über verschiedene Diskussionsthemen, welcher beispielsweise als Grundlage für eine induktive Kategorienbildung in einer qualitativen Inhaltsanalyse dienen kann.

Bisher wurde nur die Ermittlung der verschiedenen Diskussionsthemen angesprochen, bei der jedoch lediglich die verschiedenen Themen angegeben werden, aber nicht deren Häufigkeit in der Diskussion. Um diese Häufigkeit für eine Gesamtübersicht zu bestimmen, kann für jeden Textbeitrag nachgeschlagen werden, über welche Themen diskutiert wird, und daraus eine Häufigkeitsverteilung ermittelt werden.

Beim Ermitteln von Termfrequenzen sind weitere Vorverarbeitungsschritte denkbar. Gerade in Online-Partizipationsverfahren kann es vorkommen, dass die Sprachqualität einiger Textbeiträge nicht fehlerfrei ist. So können Tippfehler auftreten, die durch den Abgleich mit einem Wörterbuch mittels einer String-Distanz (z. B. der Levenshtein-Distanz, die oft aufgrund ihrer simplen Implementierung verwendet wird) reduziert werden können.

Zudem ist der Einsatz von Wortnetzen sinnvoll, in denen semantische Beziehungen zwischen Begriffen aufgeführt sind. Dazu gehören insbesondere Synonyme (z. B. *Orange* und *Apfelsine*) und hierarchische Ober- und Unterbegriffsbeziehungen. Falls in dem Wortnetz beispielsweise das Wort *Labrador* verwendet wird, so könnte jedes Auftreten dieses Wortes ebenfalls für den Begriff *Hund* gezählt werden. Als deutschsprachiges Wortnetz bietet sich GermaNet (Hamp und Feldweg, 1997) an. Da es in GermaNet für jeden Eintrag des Wortnetzes einen Weg über Oberbegriffsbeziehungen zur Wurzel des Wortnetzes gibt, ergibt sich das Problem, dass beim Zusammenfassen zweier Wörter zu einem Oberbegriff ein Informationsverlust entstehen kann, wenn zu allgemeine Oberbegriffe gewählt werden. Wird GermaNet beispielsweise zur Zusammenfassung von Begriffen in einem Tierforum verwendet, so führt das Ersetzen jeder Tierart durch das Wort *Tier* vermutlich zu einer verminderten Aussagekraft.

4 Erkennung von Emotionen zur Unterstützung von Moderation und Meinungsaggregation

Wie auch in Face-to-Face-Situationen, enthalten onlinebasierte Beteiligungs- und Entscheidungsprozesse neben sachlichen Informationen auch emotionale Inhalte. Bei der automatisierten Textanalyse im Bereich der Online-Partizipation können Emotionen und emotionale Kommunikation aus mindestens zwei unterschiedlichen Blickwinkeln von Interesse sein: (1) Aussagen zu erkennen, die im Verdacht stehen, Emotionen bei anderen auszulösen, und (2) Aussagen zu erkennen, die Emotionen repräsentieren oder darstellen. Das Erkennen des ersten Typs von emotionalen Aussagen kann z. B. bei der Moderation unterstützen, wohingegen der zweite Typ von emotionalen Aussagen, Hinweise über das Meinungsbild zu einem Zeitpunkt geben kann. Bei beiden Problemstellungen sind geeignete Emotionsmodelle und die zuverlässige Analyse von Relationen (Worauf bezieht sich ein emotionaler Inhalt?) zentrale Herausforderungen, für die wir im Folgenden einige Lösungsansätze vorschlagen.

Diskussionen im Internet rühmen sich nicht in erster Linie mit einem zivilen, höflichen und respektvollen Miteinander. Der Einsatz von Moderation kann ein Weg sein, um Zivilität und Respekt in onlinebasierten Partizipationsverfahren zu gewährleisten (Coleman und Götze, 2001; Wright und Street, 2007). Hierbei entscheiden meist Initiatoren und Dienstleister darüber, welche Aussagen in den Diskussionen nicht erwünscht oder problematisch sind. Das können etwa Beleidigungen oder aggressive Äußerungen sein, von denen angenommen wird, dass sie sich negativ auf die Partizipationsbereitschaft auswirken oder dass sie weitere

nicht zivile Äußerungen fördern können. Eine manuelle Moderation kann allerdings sehr zeit- und ressourcenaufwändig sein und ist nicht immer möglich. Es kann daher sinnvoll sein, problematische Äußerungen automatisiert zu erkennen. Aktuell beschäftigt sich die Forschung vermehrt mit „hate speech“, dt. Hassrede im Internet, worunter u. a. sexistische oder rassistische Äußerungen fallen. Um solche mit Hass gegenüber Minderheiten geladenen Aussagen automatisiert identifizieren zu können, schlagen z. B. Waseem und Hovy (2016) eine Liste von problembehafteten Aussagen vor, die sie mit einer logistischen Regression mit character n-grams auf Twitter-Beiträge anwenden. Dieses Verfahren ließe sich bei entsprechender Anpassung auch auf andere Texte übertragen. Weitere Hinweise können Äußerungen mit negativen (Valenz) und zugleich starken Emotionen (Erregung) sein (Barrett, 1998; Scholl, 2013), aber auch Wörterlisten, die Personen mit einer stark abwertenden Bedeutung in Zusammenhang bringen und damit herabwürdigen. Idealerweise sollten alle Äußerungen, die das Potential haben, den Diskurs zu gefährden, identifiziert werden. Danach entscheidet das Moderatorenteam über das Vorgehen im Einzelfall. Zusätzlich kann ein „Melden“-Button aushelfen, um Nutzerinnen und Nutzern die Möglichkeit zu geben, problematische Kommentare selbst zu melden. Bei alledem ist nicht zu vergessen, dass Moderation sich auch negativ, z. B. auf die Bereitschaft zur Beteiligung, auswirken kann (Rhee und Kim, 2009). Am besten geeignet scheint daher eine nicht restriktive und transparente Form der Moderation. Die Herausforderung für Moderatoren liegt nicht nur darin, Raum für alle Interessen und Meinungen zu schaffen, sondern auch darin, vielfältige Ausdrucksformen zuzulassen, um inklusive Beteiligung zu fördern (Trénel, 2009, 255f.).

Außerdem lassen sich Aussagen automatisiert identifizieren, die bestimmte Arten und Dimensionen von Emotionen repräsentieren oder darstellen. Solche Informationen aus den Kommentaren von Nutzerinnen und Nutzern können, z. B. in Form eines „Barometers“ oder anderer Visualisierungen des Meinungs- oder Stimmungsbilds, aufschlussreich sein und perspektivisch sogar Abstimmungsverfahren ersetzen oder zumindest ergänzen. Dabei gilt es, spezifische diskrete Emotionen in Verbindung mit bestimmten Subjekten oder Objekten zu erkennen (emotionaler Inhalt bezieht sich auf das Thema, die Position, die zur Diskussion steht). Es können sowohl diskrete als auch dimensionale Emotionsmodelle zur Anwendung kommen, wenn es darum geht, Emotionen zu kategorisieren. Studien im Bereich der Informatik arbeiten bisher vor allem mit diskreten Emotionen. Solche Verfahren können einen Überblick über spezifische Emotionen geben, die in Online-Diskussionen auftauchen. Oftmals wird dabei auf die sechs Basisemotionen Furcht, Freude, Ärger, Traurigkeit, Ekel und Überraschung (Ekman, 1992) zurückgegriffen, die im Kontext von Gesichtsausdrücken erforscht wurden. Häufig wird dieses Modell jedoch unreflektiert auf Texte übertragen, ohne zu begründen, ob und warum eine Übertragung dieses Modells auf Texten sinnvoll ist.

Geeigneter erscheinen uns hier Arbeiten zur automatisierten Bewertung der Stärke von subjektiven Aussagen (z.B. Wilson et al. (2004)), insbesondere von private states (Quirk et al., 1985), zu denen auch Meinungen, Empfindungen und Emotionen gezählt werden. Dabei werden in Wilson et al. (2004) drei maschinelle Lernverfahren, unter anderem auch eine Support Vector Machine, mit mehreren Features, darunter auch Unigrammen, eingesetzt.

In Rahmen des Tempelhofer Feldes haben wir uns die 647 Textbeiträge aus dem THF Airport ArgMining Corpus in Bezug auf Emotionen zunächst in vereinfachter Form angesehen, indem wir grob in positive und negative Emotionen unterschieden haben, die von den Nutzerinnen und Nutzern geäußert werden. Die Kategorien wurden in einem Codebuch ausführlich beschrieben. Obwohl diese Form der Kategorisierung simplifiziert ist und z. B. nicht weiter konkretisiert, welche spezifische Emotion geäußert wird, können bereits durch die Betrachtung von negativen Emotionen möglicherweise störende Auswirkungen auf den Diskurs beobachtet werden. Die folgenden Textabschnitte¹² zeigen negative Emotionen:

[*Negative Emotion*: Die Argumentation finde ich wirklich einfach lächerlich...]

[*Negative Emotion*: Was du willst interessiert keiner solche widerlichen spassverderber wie du bleiben am besten zu hause]

Insgesamt konnten in den Textbeiträgen 127 emotionale Aussagen identifiziert werden, die sich auf 111 Beiträge verteilen, was 17,16 % des codierten Materials entspricht. Daraus lässt sich z. B. folgern, dass die Diskussionen überwiegend sachlich verlaufen sind.

5 Fazit und Ausblick

Dieser Beitrag hat in drei Bereiche der automatisierten Textauswertung eingeführt und deren Möglichkeiten für einen Praxiseinsatz aufgezeigt. Im Bereich Argument Mining wurde der Einsatz von word embeddings

¹²<https://tempelhofer-feld.berlin.de/i/tempelhofer-feld/proposal/253>

auf dem THF Airport ArgMining Corpus evaluiert: Für die Klassifikation von Sätzen in die Kategorien argumentativ und nicht-argumentativ konnte eine Verbesserung gegenüber Unigrammen festgestellt werden. Für die detaillierte Klassifikation von Argumentationskomponenten haben word embeddings nicht den gleichen Effekt erbracht. Auch wenn die Verfahren im Bereich Argument Mining nicht fehlerfrei arbeiten, so ist festzuhalten, dass eine automatisierte Vorverarbeitung zu einer erheblichen Arbeitsreduktion führen kann. In der jetzigen Reife des Systems lässt sich bereits über einen Praxiseinsatz der Argument Mining-Techniken mit einem unserer Kooperationspartner in Form eines Pilotprojektes nachdenken, um die Robustheit und den praktischen Nutzen der Ergebnisse besser bewerten zu können. Wünschenswert wäre ein weiterer Korpus im Bereich der Online-Partizipation, der ebenfalls mit demselben Argumentationsschema annotiert ist. Dadurch könnten die auf dem THF Airport ArgMining Corpus trainierten Machine Learning-Modelle auf einem zweiten Korpus (und auch umgekehrt) evaluiert werden, um die Adaptivität der Verfahren auf Texte eines anderen Verfahrens in derselben Anwendungsdomäne zu überprüfen. Die Annotation solcher Korpora ist allerdings sehr ressourcenaufwändig und datenschutzrechtliche Aspekte verhindern häufig die Weitergabe von Textbeiträgen aus bereits abgeschlossenen Online-Partizipationsverfahren. Um die Klassifikationsqualität möglicherweise zu verbessern und den Einsatz in der Praxis zu optimieren, sehen wir folgende Aspekte, als relevant an, an die zukünftige weiterführende Arbeiten anknüpfen könnten: (i) Wir wollen in Zukunft weitere Features für die Klassifikation von Argumentationskomponenten implementieren und auswerten. (ii) Zusätzlich wollen wir geeignete Features für die Klassifikation von claims nach pro claims und contra claims in Subtask C verwenden. (iii) Eine bisherige Einschränkung des Systems ist die satzbasierte Klassifikation von Argumentationskomponenten. Beispiele aus dem Tempelhofer Feld-Projekt zeigen, dass mehrere aufeinanderfolgende Sätze erst in einer Einheit eine Begründung darstellen oder dass innerhalb eines Satzes mehrere Argumentationskomponenten auftreten können, z. B. eine Positionierung gefolgt von einer Begründung. Um Argumentationskomponenten variabler Länge automatisiert erkennen zu können, werden wir in Zukunft den THF Airport ArgMining Corpus dafür verwenden, um Techniken des Sequence Taggings zu evaluieren. (iv) Anstelle eines klassischen Machine Learning-Ansatzes werden wir zusätzlich die in der Klassifikation zurzeit beliebten Ansätze des Deep Learnings evaluieren. Dabei kommen anstelle von klassischem Feature Engineering neuronale Netze zum Einsatz, bei denen keine Auswahl an semantischen Features getroffen werden muss. Stattdessen muss zwischen verschiedenen Netzarchitekturen und Hyperparametern gewählt werden. Zum aktuellen Zeitpunkt ist aber noch unklar, ob die Größe des Trainingskorpus für Deep Learning ausreicht.

Eine angebrachte Frage für die Extraktion von Vorschlägen ist, wie mit kaum beachteten Textbeiträgen, die möglicherweise interessante oder brauchbare Vorschläge beinhalten, umgegangen werden soll. Im Rahmen einer automatisierten Extraktion ist es zunächst schwierig, ein objektiv messbares Interessanzmaß zu formulieren, ohne dabei auf die Reaktionen der anderen Teilnehmer (in Form von Votings) zurückzugreifen. Beispiele aus der Praxis (z. B. der Bürgerhaushalt 2016 in Köln) zeigen, dass bei der Auswahl der von den Politikern zu berücksichtigenden Vorschläge oft auf die Bewertungen der Plattform zurückgegriffen wird. Bei solch einem Vorgehen kann eine technische Lösung nur bedingt helfen, um untergegangene Vorschläge hervorzuheben. Als Fazit ist daher festzuhalten, dass die automatisiert extrahierten Argumentationskomponenten alle noch durch einen Menschen kontrolliert werden sollten, damit keine Bürgermeinung durch einen Klassifikationsfehler verloren geht. Beim Anzeigen der extrahierten Argumentationskomponenten planen wir, die auf den Plattformen abgegebenen Bewertungen in die Sortierreihenfolge einzubeziehen, um häufig auftretende Vorschläge schneller identifizieren zu können.

Im Bereich der Themenextraktion sind sowohl Wortwolken als auch semantische Themengruppierungen mit Latent Dirichlet Allocation praxistauglich. Insbesondere bietet es sich an, die LDA-Ergebnisse in Form eines Dashboards, in dem die ersten fünf oder zehn Wörter jedes LDA-Themas gezeigt werden, zu visualisieren. Dabei entsteht jedoch die Frage, wie interpretierbar die Ergebnisse sind, da pro Topic mehrere Begriffe aufgeführt sind. Um die Ergebnisse in noch komprimierterer Form darzustellen, könnten die wahrscheinlichsten Wörter pro LDA-Topic mittels Techniken des Topic Labeling noch verkürzt dargestellt werden, indem die aufgeführten Begriffe durch eine textuelle, automatisch generierte Zusammenfassung ersetzt werden. Daher werden wir in Zukunft Techniken erproben, um die aus Online-Partizipationsverfahren extrahierten Wörter semantisch zusammenzufassen.

Für den Bereich der Emotionserkennung ist zunächst die Frage zu beantworten, ob ein dimensionales oder eher ein diskretes Emotionsmodell für die Analyse von Texten aus Online-Partizipationsverfahren gewählt werden soll. Zu den realistischen Einsatzszenarien für die automatisierte Emotionserkennung zählen unter anderem die Generierung eines Überblicks über das Meinungsbild und die automatisierte Erstellung von Moderationshinweisen, die auf negative Emotionen hinweisen. Diese sind insbesondere für Partizipationsverfahren ohne Prämoderation sinnvoll. Bei einer feineren Unterteilung in Form eines diskreten Modells besteht im Praxiseinsatz die Notwendigkeit die ermittelte Emotionsverteilung für den jeweiligen Anwendungsfall durch einen Menschen zu interpretieren, da die Rohdaten (z. B. dass 30 % der

Teilnehmer die Emotion Überraschung ausdrücken) ohne Kontextwissen nur bedingt Aussagekraft haben. Darüber hinaus ist zu bedenken, dass eine Emotionsanalyse je nach Thema des Partizipationsverfahrens politisch eventuell nicht erwünscht ist. Wir planen, in Zukunft einen Datensatz mit einem diskreten Emotionsmodell zu codieren und dabei die Rolle von Emotionen in der Diskussions- und Meinungsdynamik in Online-Partizipationsverfahren weiter zu untersuchen.

Danksagungen

Die Autoren sind Mitglieder des durch das Land Nordrhein-Westfalen geförderten Fortschrittskollegs „Online-Partizipation“. Rechenkapazitäten wurden vom „Zentrum für Informations- und Medientechnologie“ (ZIM) in der Heinrich-Heine-Universität Düsseldorf bereitgestellt.

Literaturverzeichnis

- Aseel Addawood und Masooda Bashir. 2016. “What Is Your Evidence?” A Study of Controversial Topics on Social Media. In *Proceedings of the Third Workshop on Argument Mining, hosted by the 54th Annual Meeting of the Association for Computational Linguistics, ArgMining@ACL 2016*.
- Lisa Feldman Barrett. 1998. Discrete Emotions or Dimensions? The Role of Valence Focus and Arousal Focus. *Cognition Emotion*, 12(4):579–599.
- Jürgen Behrendt. 2016. Lebendige Demokratie durch transparente und digitale Beteiligungskultur. Vortrag beim Praxis-symposium „Online-Partizipation“ am 13.12.2016, Gelsenkirchen.
- Hartmut Beuß. 2017. Leitfaden zur Mitbestimmung bei der Umsetzung der Open.NRW Strategie.
- David M. Blei, Andrew Y. Ng, und Michael I. Jordan. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022.
- Stephen Coleman und John Gøtze. 2001. *Bowling Together: Online Public Engagement in Policy Deliberation*. Hansard Society.
- Corinna Cortes und Vladimir Vapnik. 1995. Support-Vector Networks. *Machine Learning*, 20(3):273–297.
- Judith Eckle-Kohler, Roland Kluge, und Iryna Gurevych. 2015. On the Role of Discourse Markers for Discriminating Claims and Premises in Argumentative Discourse. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015*, pages 2236–2242.
- Paul Ekman. 1992. An Argument for Basic Emotions. *Cognition and Emotion*, 6(3-4):169–200.
- Tobias Escher, Dennis Friess, Katharina Esau, Jost Sieweke, Ulf Tranow, Simon Dischner, Philipp Hagemeyer, und Martin Mauve. 2017. Online Deliberation in Academia: Evaluating the Quality and Legitimacy of Co-Operatively Developed University Regulations. *Policy & Internet*, 9(1):133–164.
- Peter Gladitz, Sabrina Schöttle, Malte Steinbach, Nadja Wilker, und Theresa Witt. 2017. DIID Monitor Online Partizipation - Zum Stand von Online-Bürgerbeteiligung in den Kommunen Nordrhein-Westfalens. *Kommunalpraxis Wahlen*, (1):30–34.
- Thomas L. Griffiths und Mark Steyvers. 2004. Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(Suppl. 1):5228–5235, April.
- Ivan Habernal und Iryna Gurevych. 2016. Argumentation Mining in User-Generated Web Discourse. *Computational Linguistics*, page (in press).
- Birgit Hamp und Helmut Feldweg. 1997. GermaNet - a Lexical-Semantic Net for German. In *Proceedings of ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, pages 9–15.
- Herbert Kubicek, Barbara Lippa, Hilmar Westholm, und Niels Kohlrausch. 2007. Medienmix in der lokalen Demokratie. Die Integration von Online-Elementen in Verfahren der Bürgerbeteiligung.

- Matthias Liebeck, Katharina Esau, und Stefan Conrad. 2016. What to Do with an Airport? Mining Arguments in the German Online Participation Project Tempelhofer Feld. In *Proceedings of the Third Workshop on Argument Mining, hosted by the 54th Annual Meeting of the Association for Computational Linguistics, ArgMining@ACL 2016*, pages 144–153.
- Matthias Liebeck und Stefan Conrad. 2015. IWNLP: Inverse Wiktionary for Natural Language Processing. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 414–418. Association for Computational Linguistics.
- Tomas Mikolov, Kai Chen, Greg Corrado, und Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. *CoRR*, abs/1301.3781.
- Oliver Märker und Christina Rucker. 2016. Ludwigshafen diskutiert: „Abriss der Hochstrasse Nord ... und was kommt danach?“ Auswertungsbericht zur Öffentlichkeitsbeteiligung www.ludwigshafen-diskutiert.de.
- Open.NRW. 2016. Sachstandsbericht zur Umsetzung der Open.NRW Strategie.
- Raquel Palau und Marie-Francine Moens. 2009. Argumentation Mining: The Detection, Classification and Structure of Arguments in Text. In *Proceedings of the 12th International Conference on Artificial Intelligence and Law, ICAIL '09*, pages 98–107. ACM.
- Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, und Édouard Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, und Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. Longman.
- Radim Řehůřek und Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50. ELRA.
- June Woong Rhee und Eun-Mee Kim. 2009. Deliberation on the Net: Lessons from a Field Experiment. In *Online Deliberation: Design, Research, and Practice*, pages 223–232. CSLI Publications.
- Wolfgang Scholl. 2013. The socio-emotional basis of human interaction and communication: How we construct our social world. *Social Science Information*, 51(1):3–33.
- Christian Stab und Iryna Gurevych. 2014. Identifying Argumentative Discourse Structures in Persuasive Essays. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, pages 46–56.
- Matthias Trénel. 2009. Facilitation and Inclusive Deliberation. In *Online Deliberation: Design, Research, and Practice*, pages 253–257. CSLI Publications.
- Thomas Wagenknecht, Timm Teubner, und Christof Weinhardt. 2016. The Impact of Anonymity on Communication Persuasiveness in Online Participation. In *Proceedings of the International Conference on Information Systems - Digital Innovation at the Crossroads, ICIS 2016*, pages 1–11.
- Zeeraq Waseem und Dirk Hovy. 2016. Hateful symbols or hateful people? predictive features for hate speech detection on twitter. In *Proceedings of the NAACL Student Research Workshop*, pages 88–93. Association for Computational Linguistics.
- Theresa Wilson, Janyce Wiebe, und Rebecca Hwa. 2004. Just how mad are you? finding strong and weak opinion clauses. In *Proceedings of the 19th National Conference on Artificial Intelligence, AAAI'04*, pages 761–767. AAAI Press.
- Scott Wright und John Street. 2007. Democracy, deliberation and design: the case of online discussion forums. *New Media & Society*, 9(5):849–869.