

Emotionsklassifikation in Texten unter Berücksichtigung des Komponentenprozessmodells

Roman Klinger

Zusammenfassung Ein wichtiger Bestandteil unserer alltäglichen Kommunikation, neben der Mitteilung und Beschreibung von Ereignissen und Fakten, ist der Ausdruck von Emotionen, welcher auch Bestandteil von Hassrede ist: Es wird zum Beispiel Wut zum Ausdruck gebracht, was wiederum bei den Betroffenen Angst, Traurigkeit oder vielleicht auch Überraschung auslösen kann. In der maschinellen Verarbeitung von Sprache haben sich in der letzten Zeit einige konkrete Aufgaben, welche Teil der Emotionsanalyse in Text sind, herauskristallisiert. Diese sind zum einen Klassifikationsaufgaben (welche Emotion drückt ein Text aus?) und zum anderen relationale Strukturlernaufgaben (welche Wörter bezeichnen die Person, die eine Emotion fühlt und welche Wörter lassen auf die Ursache der Emotion schließen?). Wir verschaffen uns in diesem Kapitel einen kurzen Überblick über das Feld und diskutieren im Anschluss etwas genauer, wie sich die Beschreibungen von Emotionen in verschiedenen Domänen unterscheiden und wie Ereignisbeschreibungen mit Hilfe psychologischer Theorien mit Emotionen zusammengebracht werden können. Insbesondere analysieren wir auf Basis des Emotions-Komponenten-Prozessmodells, auf welche Komponenten von Emotionen (subjektives Gefühl, kognitive Evaluation, Körperreaktion, Ausdruck, Motivation) Autor:innen zugreifen, und stellen fest, dass diese Verteilung zwischen sozialen Medien und Literatur unterschiedlich ist. In beiden Domänen spielt aber die kognitive Komponente zur Interpretation von Emotionen eine wichtige Rolle. Dies zeigt auf, dass insbesondere der Ereignisinterpretation Aufmerksamkeit geschenkt werden muss, um implizit kommunizierte Emotionen aufzudecken. Dies motiviert uns, Emotionen mit Hilfe der Appraisaltheorien zu analysieren, welche den Zusammenhang zwischen kognitiven Prozessen und Emotionen erklären. Zu beiden Konzepten – dem Komponentenmodell und den Appraisaltheorien – präsentieren wir Textkorpora und Klassifikationsmodelle.

Roman Klinger
Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, Pfaffenwaldring 5b,
70569 Stuttgart, e-mail: roman.klinger@ims.uni-stuttgart.de

1 Einleitung

Die Emotionsanalyse in Text verknüpft die Forschung des Affective Computing (Picard 1997) mit der maschinellen Verarbeitung von Sprache (Computerlinguistik/Natural Language Processing – CL/NLP). Hierbei wird im Gegensatz zu dem Feld der *Informationsextraktion* (Sarawagi 2008) weniger in den Mittelpunkt gestellt, *was* im Text ausgedrückt wird, sondern eher *wie* etwas wahrgenommen wird, also welche emotionale Konnotation mittransportiert wird. Dies kann sich auf die konkreten Nennungen von Emotionen im Text (zum Beispiel wie eine Figur in einem Buch ein bestimmtes Ereignis empfindet) oder auch auf Schlussfolgerungen über die Emotionen eines Autors eines Texts (zum Beispiel wie sich eine Nutzerin sozialer Medien beim Schreiben eines Tweets gefühlt hat) beziehen. Der erstgenannte Fall bewegt sich an der Schnittstelle zur Informationsextraktion, wobei der zweitgenannte Fall in den Bereich des *Author Profiling* (Rosso et al. 2018) fällt.

In der NLP haben sich einige Kernaufgaben im Bereich der Emotionsanalyse herauskristallisiert. Dies ist zum einen die Erkennung von semantischen Rollen im Kontext von Emotionen, also *wer* fühlt *was* und *warum* (Kim und Klinger 2018; Mohammad et al. 2014), mit einem weiteren Schwerpunkt auf die Erkennung der Emotionsursache (Xia und Ding 2019; Oberländer und Klinger 2020). Zum anderen, vielleicht die Hauptaufgabe der Emotionsanalyse, ist dies die Emotionsklassifikation (oder Regression), bei der Texten (Sätzen, Tweets, Absätze. . .) Emotionsklassen oder Intensitätswerte zugewiesen werden. Diese Aufgaben sind mit der Analyse von Hassrede verwandt – Emotionen werden von einer Ursache ausgelöst und können ein Ziel haben, dies sind auch wichtige Bestandteile bei der Äußerung von Hassrede. Tatsächlich wurde bereits gezeigt, dass maschinelle Lernverfahren, welche Hassrede vorhersagen, von einer gemeinsamen Modellierung profitieren (Plaza-del-Arco et al. 2021; Rajamanickam et al. 2020). Dieser Aspekt stellt eine wichtige Motivation dar, der Emotionsanalyse auch im Kontext der Analyse von Hassrede Aufmerksamkeit zu schenken. Die bestehenden Ansätze, Hassrede und Emotionsanalyse in komputationellen Modellen zu integrieren, bauen nur in Ansätzen auf dem umfangreichen Wissen auf, welches in der Psychologie zu der Struktur, dem Zweck, und dem Ausdruck von Emotionen bekannt ist. Somit ist dieses Kapitel auch als ein Vorschlag zu verstehen, das vorhandene Wissen über Emotionen in der Psychologie und allgemeinen Ansätzen in der Emotionsanalyse auf den spezielleren Fall von Hassrede zu übertragen. Dies kann potentiell zu einer höheren Genauigkeit automatischer Systeme, aber vor allem zu einer größeren Abdeckung führen.

Selbstverständlich berührt das Feld der Emotionsanalyse in Text auch die Emotionspsychologie. Hierbei werden insbesondere Emotionskonzepte „importiert“, also Kategorien, welche als relevant erachtet werden, zur Klassifikation herangezogen. Wir betrachten in diesem Kapitel genau diese Aufgabe der Zuweisung von vorgefertigten Emotionskategorien zu gegebenen Textabschnitten. Übliche Emotionskategorien umfassen solche, wie sie aus den

Theorien von Paul Ekman (1992) folgen, nämlich *Wut*, *Freude*, *Ekel*, *Angst*, *Traurigkeit* und *Überraschung* oder von Robert Plutchik (2001), der neben Intensitätsabstufungen noch *Vertrauen* und *Antizipation* als sogenannte Basisemotionen vorschlägt. Häufig wird die Klassifizierung von Emotionen als eine durchgängige Lernaufgabe verstanden („Ende-zu-Ende“), die möglicherweise durch lexikalische Ressourcen unterstützt wird (siehe die SemEval Shared Task 1 zu „Affect in Tweets“ für einen Überblick über aktuelle Ansätze (Mohammad et al. 2018)). Die Information, was eine Emotion ausmacht, liegt somit alleine in den Daten, welche auf Basis von Beispielen die Emotionskonzepte nutzbar machen. Während dieses Ende-zu-Ende-Lernen und die Feinabstimmung von vortrainierten Modellen für die Klassifikation große Leistungsverbesserungen im Vergleich zu rein merkmalsbasierten Methoden oder wörterbuchbasierten Verfahren gezeigt hat, vernachlässigen solche Ansätze typischerweise allerdings das vorhandene Wissen über Emotionen in der Psychologie. Es gibt nur sehr wenige Ansätze, die darauf abzielen, psychologische Theorien (über grundlegende Emotionskategorien hinaus) mit Emotionsklassifikationsmodellen zu kombinieren. Als bemerkenswerte Ausnahmen sind solche Arbeiten zu nennen, die das Konzept von *Affekt* in den Mittelpunkt stellen, das durch die Kombination von Valenz und Arousal gemessen werden kann (Buechel und Hahn 2017b; Buechel et al. 2021). Dies gilt auch für die oben genannten Arbeiten, welche Emotionsanalyse und Hassrede zusammenbringen – sie modellieren Emotionen auf Basis einer diskreten Menge von Kategorien.

Nun können Emotionen aber unterschiedlich ausgedrückt werden. In diesem Kapitel stellen wir eine Analyse auf Basis der Appraisaltheorien und des Emotions-Komponentenmodells von Scherer vor, welche folgende Definition von Emotionen enthalten (frei übersetzt): Emotionen sind ein synchronisierter Prozess verschiedener Teilsysteme, in Reaktion auf ein relevantes Ereignis (Scherer et al. 2001; Scherer 2005b). Diese Teilsysteme sind eine *Motivationskomponente*, der *Ausdruck*, die *neurophysiologische körperliche Reaktion*, das *subjektive Gefühl* und die *kognitive Bewertung*. Dieser Prozess folgt der Wahrnehmung von relevanten Ereignissen. Wir nehmen daher an, dass alleine Ereignisbeschreibungen für die Kommunikation von Ereignissen ausreichend sind, da die Leser:in dieser Beschreibung empathisch fähig sind, die Emotionen der Teilnehmenden eines Ereignisses zu rekonstruieren. Dies ist auch für Hassrede relevant – ein wiederkehrender Bestandteil dieser ist der Aufruf zu Taten. Um diese bezüglich ihrer (emotionalen) Wahrnehmung einzuordnen, dürften Appraisaltheorien ein sinnvolles Werkzeug sein.

In Abschnitt 3 werden wir zunächst untersuchen, ob sich Emotionsbeschreibungen in sozialen Medien tatsächlich solchen Emotionskomponenten zuweisen lassen, und vergleichen die Verteilung dieser Komponenten mit einer anderen Domäne, nämlich der Literatur. In Abschnitt 4 schlagen wir weiterhin vor, spezifisch die Komponente des kognitiven Appraisal zu nutzen, um Emotionen von Ereignisbeschreibungen aus Text zu extrahieren. Die Haupterkenntnisse dieses Kapitels sind, dass die Modellierung von Emotionskomponenten tatsächlich Aufschluss über die Art der Emotionskommunikation

erlaubt und des Weiteren in einem Großteil der Texte eine kognitive Evaluierung (Appraisal) zur Interpretation der Emotion notwendig ist. Diese Modellierung führt weiterhin zu einer Verbesserung der Emotionsklassifikation in den gegebenen Korpora. Die hier vorgestellten Arbeiten basieren auf bereits erschienenen Publikationen (Casel et al. 2021; Hofmann et al. 2021, 2020). Die zugrunde liegenden Daten können auf der Internetseite <https://www.ims.uni-stuttgart.de/data/emotion> heruntergeladen werden.

2 Hintergrund

2.1 Emotionsmodelle in der Psychologie

Als Bestandteil des menschlichen Lebens sind Emotionen in der Psychologie eingehend untersucht worden. Sie werden allgemein als Reaktionen auf wichtige Ereignisse interpretiert. Die Debatten um ihre Definition haben jedoch (bisher) nie zu einem eindeutigen Konsens geführt, sodass eine vielfältige Literatur zu diesem Thema existiert. Dies hat Auswirkungen auf die computergestützte Emotionsanalyse (in Texten, aber auch darüber hinaus), welche aus einer großen Vielfalt der verfügbaren psychologischen Theorien auswählen kann (Scarantino 2008).

Einige dieser Theorien konzentrieren sich auf die evolutionäre Funktion von Emotionen, und dementsprechend auf ihre Verknüpfung mit Handlungen (Izard 1971; Tooby und Cosmides 2008). Der Kerngedanke ist, dass Emotionen dem Menschen helfen, alltägliche Aufgaben zu bewältigen und sozial relevante Informationen zu kommunizieren, indem sie spezifische physiologische Symptome auslösen. Insbesondere gibt es Verhaltensmuster (z.B. Lächeln), die diskrete Emotionsbegriffe widerspiegeln (z.B. Freude), was darauf hindeutet, dass emotionale Zustände anhand einiger Kategorien der natürlichen Sprache gruppiert werden können. Eine der populärsten Quellen für eine Reihe von grundlegenden Emotionen ist die Theorie von Paul Ekman (1992). Ekman untersuchte die Beziehung zwischen Emotionen und Kultur und betrachtete hierbei insbesondere beobachtbare Kriterien, unter anderem Gesichtsausdrücke. Er behauptete, dass die Gruppe der grundlegenden Emotionen, nämlich Angst, Ekel, Wut, Freude, Traurigkeit und Überraschung, Gesichtsmuskelbewegungen kulturübergreifend unterscheiden lasse (was heutzutage teilweise angezweifelt wird (Gendron et al. 2014)). Als Ergänzung dieses Modells geht Robert Plutchik explizit von der Annahme aus, dass verschiedene Grundemotionen gemeinsam auftreten können, zum Beispiel Vertrauen und Freude, was der Fall sei, wenn Liebe erlebt werde (Plutchik 2001). Solche Emotionsmischungen sowie ein Gegensatz zwischen Angst und Furcht, Freude und Traurigkeit, Überraschung und Antizipation, Vertrauen und Abscheu, wurden des Weiteren in dieses Modell aufgenommen. In der automatischen Verarbeitung

natürlicher Sprache wird meist eine Menge von vier bis acht grundlegenden Emotionen verwendet, wobei Angst, Wut, Freude und Traurigkeit von den meisten Ansätzen geteilt werden (es existieren aber auch Ausnahmen, die eine größere Zahl von Kategorien betrachten, zum Beispiel von Abdul-Mageed und Ungar (2017)).

Ein wichtiger Bestandteil von Emotionen ist ihr prozeduraler, synchronisierter Charakter. Während bereits Ekman die Reaktion auf ein relevantes Ereignis als konstituierendes Merkmal von Basisemotionen nennt, wird dieser Aspekt von Klaus Scherer noch stärker in den Mittelpunkt gerückt (Scherer 2005a). Er definiert Emotionen als eine

„Episode zusammenhängender, synchronisierter Veränderungen in den Zuständen aller oder der meisten der fünf organismischen Subsysteme als Reaktion auf die Bewertung eines externen oder internen Stimulusereignisses, welches als relevant für wichtige Belange des Organismus erkannt wird“ (Übersetzung durch den Autor dieses Kapitels).

Die fünf Komponenten sind die *kognitive Bewertung*, *neurophysiologische körperliche Symptome*, *motorische Ausdrücke*, *motivationale Handlungstendenzen*, und *subjektive Gefühle*. Die *Kognitive Bewertung* (Appraisal) befasst sich mit der Bewertung eines Ereignisses. Wir werden diese fünf Teilsysteme in Abschnitt 3 genauer betrachten. Hier sei zunächst festgehalten, dass der *Affekt*, welcher durch Valenz, Arousal und (in der NLP nur gelegentlich) durch die Dominanz (kurz: VAD) des Erlebten gemessen werden kann, eine Alternative zu diskreten Emotionsmodellen darstellt. Solche dimensionalen Modelle stellen Vektorräume dar, in denen sich verschiedene Emotionen als Punkte wiederfinden (Russell und Mehrabian 1977).

Eine aussagekräftigere, ebenfalls dimensionale, Alternative zum VAD-Modell ist durch den kognitiven Bewertungsprozess (Appraisal) motiviert, der Teil der Emotionen ist. Das Modell von Smith und Ellsworth (1985) führt eine Reihe von Variablen ein, und zwar wie angenehm eine Situation ist, ob man sich verantwortlich dafür fühlt, was geschieht, die Gewissheit, was geschieht, die notwendige Anstrengung mit der Situation umzugehen und die situative Kontrolle der die Emotion wahrnehmenden Person. Smith und Ellsworth zeigen, dass diese Dimensionen besser geeignet sind, um Emotionskategorien zu unterscheiden, als VAD. Dieses Modell betrachten wir in Abschnitt 4 etwas genauer.

2.2 Emotionsklassifikation

Bisherige Arbeiten zur Emotionsanalyse in der NLP konzentrieren sich entweder auf die Erstellung von Ressourcen oder auf die Klassifizierung von Emotionen für eine spezifische Aufgabe und Domäne. Auf der Seite der Ressourcenerstellung ist die frühe und einflussreiche Arbeit von Pennebaker et al. (2001) zu nennen, nämlich die Erstellung eines Wörterbuchs, in

dem die Einträge mit verschiedenen psychologisch relevanten Kategorien verknüpft werden, einschließlich einer Untergruppe von Emotionen. Strapparava und Valitutti machten WordNet Affect verfügbar, um Wörter mit ihrer emotionalen Konnotation zu verknüpfen (Strapparava und Valitutti 2004). Mohammad veröffentlichte das NRC-Wörterbuch mit mehr als 14.000 Wörtern für eine Reihe von diskreten Emotionsklassen und ein Valenz-Arousal-Dominanz-Wörterbuch (Mohammad und Turney 2012; Mohammad 2018). Büchel et al. haben einen methodischen Rahmen entwickelt, um bestehende Affektlexika an spezifische Anwendungsfälle anzupassen (Buechel et al. 2016).

Für die Evaluation und die Entwicklung von Modellen mit Hilfe von maschinellem Lernen steht mittlerweile auch eine große Zahl von annotierten Korpora zur Verfügung. Einige von ihnen enthalten Informationen zu Valenz und Arousal (Buechel und Hahn 2017a; Preoțiu-Pietro et al. 2016), die Mehrheit verwendet jedoch diskrete Emotionsklassen, zum Beispiel zur Kennzeichnung von Märchen (Alm et al. 2005), Blogs (Aman und Szpakowicz 2007), Tweets (Mohammad et al. 2017; Schuff et al. 2017; Mohammad 2012; Mohammad und Bravo-Marquez 2017; Klinger et al. 2018), Facebook-Posts (Preoțiu-Pietro et al. 2016), Nachrichten-Schlagzeilen (Strapparava und Mihalcea 2007), Dialogen (Li et al. 2017), literarischen Texten (Kim et al. 2017) oder Selbstberichten über emotionale Ereignisse (Scherer und Wallbott 1997; Troiano et al. 2019). Wir verweisen die Leser:innen auf unseren Überblickartikel für einen umfassenden Vergleich (Bostan und Klinger 2018).

Die meisten automatischen Methoden, die Textabschnitten Annotationen zuweisen, basieren auf maschinellem Lernen (Alm et al. 2005; Aman und Szpakowicz 2007; Schuff et al. 2017, u.a.). Neuere Systeme basieren häufig auf Transferlernen auf Basis von generischen Repräsentationen (Klinger et al. 2018; Mohammad et al. 2018; Mohammad und Bravo-Marquez 2017). Felbo et al. (2017) schlug z.B. vor, Emoji-Repräsentationen zu nutzen, um Modelle vorzutrainieren. Cevher trainierte erst mit existierenden Korpora und dann mit Daten einer spezifischen Domäne, für die nur wenige Trainingsdaten verfügbar waren (Cevher et al. 2019).

Diese Arbeiten weisen Emotionskategorien oder Affektwerte von Valenz und Arousal zu. In diesem Kapitel beschäftigen wir uns mit Emotionskomponenten und Appraisalannotationen, um nicht nur die wahrgenommene oder ausgedrückte Emotion zu charakterisieren, sondern auch ein Verständnis zu entwickeln, wie diese ausgedrückt wird. Uns ist nur eine damit verwandte Studie bekannt, die Bewertungsdimensionen zur Verbesserung der Emotionsvorhersage berücksichtigt (Campero et al. 2017). In dieser Studie annotierten Probanden 200 Geschichten mit 38 Bewertungsmerkmalen, um zu bewerten, ob eine textbasierte Repräsentation eine fMRI¹-basierte Klassifizierung ergänzt. Abgesehen von dieser Studie haben alle bisherigen maschinellen Ansätze dies ohne Zugang zu Appraisaldimensionen oder Komponenten getan. Nur in einigen wenigen Arbeiten wurden kognitive Komponenten einbezogen,

¹ „functional Magnetic Resonance Imaging“, also ein bildgebendes Verfahren, welches Gehirnaktivität abbildet.

die aus dem OCC-Modell stammen (benannt nach den Initialen der Autoren Ortony, Clore und Collins (Clore und Ortony 2013)). Auf der Grundlage des OCC-Modells entwickelte Shaikh et al. (2009) einen regelbasierten Ansatz zur Textinterpretation. Eine weitere verwandte Arbeit ist die von Alexandra Balahur erstellte Datenbank EmotiNet, welche Handlungsfolgen enthält, die zu bestimmten Emotionen führen. Sie modelliert zwar nicht Appraisal an sich, nutzt diese aber zur Motivation der Verknüpfung von Ereignissen mit Emotionen.

Die einzige uns bekannte Arbeit, die Emotionskomponenten untersucht (allerdings nicht nach dem Emotionskomponenten-Prozessmodell und ohne computergestützte Modellierung), ist eine Korpus-Studie von Fan-Fiction (Kim und Klinger 2019). Sie analysiert, ob Emotionen über Gesichtsbeschreibungen, Körperhaltungsbeschreibungen, das Aussehen, den Blick, die Stimme, Gesten, subjektive Empfindungen oder räumliche Beziehungen der Figuren kommuniziert werden. Dieser Satz von Variablen ist nicht identisch mit den Emotionskomponenten, aber er ist verwandt. Die Autoren stellen fest, dass einige Emotionen bevorzugt mit bestimmten Aspekten beschrieben werden. Motiviert wurde diese Arbeit durch eine linguistische Literaturstudie (van Meel 1995).

Im Gegensatz zu ihrer Arbeit vergleicht unsere Studie zwei verschiedene Domänen (Tweets und Literatur) und folgt dem Prozessmodell der Emotionskomponenten genauer. Außerdem zeigen wir die Verwendung dieses Modells für computergestützte Emotionsklassifikation durch Multi-Task-Lernen und schlagen ein auf die Appraisalkomponente fokussiertes dimensionales Modell vor.

3 Die Emotionskomponenten in Sozialen Medien im Vergleich zu Literarischen Texten²

In diesem Abschnitt untersuchen wir, wie Nutzer:innen von Twitter Emotionen beschreiben und vergleichen dies mit Autor:innen von literarischen Texten. Hierzu ordnen wir Instanzen existierender Emotionskorpora in die verschiedenen Komponenten ein und untersuchen das entstehende Korpus statistisch und nutzen die verschiedenen Annotationen gemeinsam in einem Modell des maschinellen Lernens.

Die betrachteten Komponenten von Emotionen sind die *kognitive Bewertung*, *neurophysiologische körperliche Symptome*, der *Ausdruck*, *motivationale Handlungstendenzen* und das *subjektive Gefühl*. Die *kognitive Bewertung* (Appraisal) befasst sich mit der Bewertung eines Ereignisses. Das Ereignis wird im Hinblick auf seine Relevanz für das Individuum, die Implikationen und die Konsequenzen, zu denen es führen könnte, die möglichen Wege zu Möglichkeiten,

² Dieser Abschnitt basiert auf Casel et al. (2021).

es zu bewältigen und zu kontrollieren, und seine Bedeutung nach persönlichen Werten und sozialen Normen charakterisiert. Die Komponente der *neurophysiologischen Symptome* betrifft automatisch aktivierte Reaktionen und Symptome des Körpers wie Veränderungen des Herzschlags oder des Atemmusters. Die Komponente des *motorischen Ausdrucks* enthält alle Bewegungen, Gesichtsausdrücke, Veränderungen in der Sprache und ähnliche Muster. Handlungen wie Aufmerksamkeitsverschiebungen und Bewegungen in Bezug auf den Ort des Geschehens sind Teil der *motivationale-Handlungstendenzen*-Komponente. Schließlich berücksichtigt die Komponente der *subjektiven Gefühle*, wie stark, wichtig und anhaltend die Empfindungen sind. Scherer (2005b) argumentiert, dass es möglich sei, auf die Emotion, die eine Person erlebt, durch die Analyse der Veränderungen in den fünf Komponenten Rückschlüsse zu ziehen. Er weist auch darauf hin, dass Computermodelle die Emotionskomponenten nicht ignorieren dürfen (Scherer 2009).

Wir nehmen im Folgenden an, dass Emotionen in Texten in einer Art kommuniziert werden, welche den Komponenten folgt. Dies sind Alternativen zu der direkten Nennung des Emotionskonzepts („Ich bin wütend.“). Beispiele sind „Er wollte weglaufen.“, womit die Motivation beschrieben wird, „Sie lächelte.“, womit der Ausdruck beschrieben wird, „Sie zitterte.“, was eine Emotion beschreiben kann, indem die Körperreaktion erklärt wird, „Ich fühle mich schlecht.“, womit das subjektive Gefühl genannt wird, oder „Ich bin unsicher, was geschehen ist.“, womit eine kognitive Bewertung der Situation beschrieben wird und damit bei den Lesenden die Notwendigkeit besteht diese zu interpretieren, um zu verstehen, welche Emotion der/die Autor:in meint.

3.1 Korpusannotation

Wir nutzen zwei verschiedene englischsprachige Korpora unterschiedlicher Domänen für unsere Studie. In diesen Korpora werden wir die Verwendung verschiedener Emotionskomponenten untersuchen. Für die Analyse von Literatur verwenden wir das REMAN-Korpus, welches Texte aus dem Projekt Gutenberg enthält. Dieses Korpus ist bereits mit den Emotionen Wut, Angst, Vertrauen, Ekel, Freude, Traurigkeit, Überraschung und Antizipation sowie „Andere Emotion“ annotiert. Von den 1720 Instanzen annotieren wir 1000 zufällig ausgewählte Instanzen. Zur Analyse von sozialen Medien nutzen wir das Twitter Emotion Corpus (TEC) (Mohammad 2012). Die Emotionskategorien sind Wut, Ekel, Angst, Freude, Traurigkeit und Überraschung. TEC enthält 21.000 Tweets, von denen wir 2041 zufällig gewählte Instanzen annotieren.

Wir annotieren mit zwei Annotator:innen, welche zunächst in zwei Runden trainiert werden. Dabei wird ein moderates Inter-Annotator Agreement von durchschnittlich Cohen's $\kappa=0,68$ erreicht. Beispiele der Annotation finden sich in Tabelle 1.

Komponente	Textbeispiel
Kognitive Bewertung	I can't stop found my old lava lamp!
Physiologisch	She did not know; she trembled. Apparently, i might have alcohol poisoning #stupidgirl
Subjektives Empfinden	Women-women-I love thee! bad day
Motivation	We're going out tonight. Sometimes I wanna take your head and ram it into mirrors.
Ausdruck	An expression of annoyance appeared on the emperor's face. Finals tomorrow... ugh

Tabelle 1: Beispiele des Emotionskomponentenkörpus

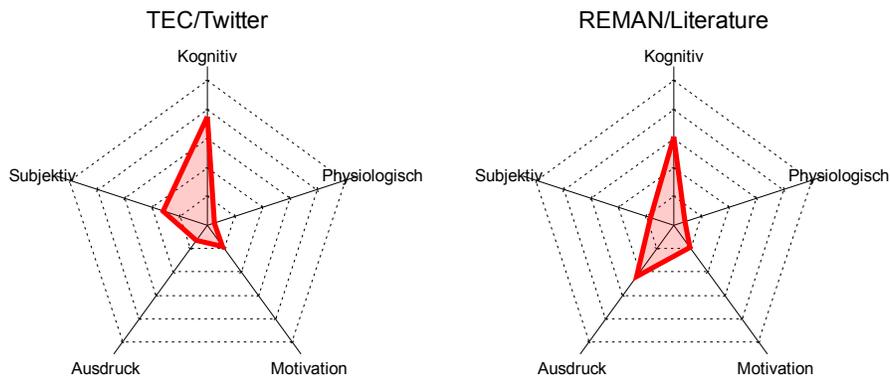


Abbildung 1: Ergebnisse der Emotionskomponentenannotation in Twitter und in Literatur. Desto weiter außen ein Punkt liegt, desto häufiger findet sich eine Annotation dieser Komponente. Für eine emotionsspezifische Analyse verweisen wir die Leser:innen auf Casel et al. (2021).

Die Analyse der Korpora findet sich zusammengefasst in Abb. 1. Wir beobachten, dass die kognitive Komponente in beiden Korpora am häufigsten auftritt. In Twitter ist die zweithäufigste Komponente die subjektive Wahrnehmung, in der Literatur ist es die Ausdrucks Komponente. Bemerkenswert ist die unterschiedliche Ausprägung des subjektiven Ausdrucks – Nutzer:innen sozialer Medien beschreiben diesen deutlich seltener, als er in Literatur auftritt. Dies ist zu erwarten und entspricht dem „Show-don’t-Tell“-Paradigma, nachdem Leser:innen in der Literatur Ereignisse und Emotionen erlebbar gemacht werden sollten, statt sie einfach nur zu beschreiben.

Die Komponenten sind nicht über die Emotionen gleichverteilt. Ekel wird in sozialen Medien häufiger mit Körperreaktionen beschrieben und zur Kommunikation von Angst wird eher das subjektive Gefühl genutzt.

Unter der Annahme, dass die beobachtete Verteilung auch in anderen Domänen als denen, die durch die genutzten Korpora repräsentiert werden,

gilt, ist nun Folgendes festzuhalten: In sozialen Medien, die uns für die Analyse von Hassrede freilich mehr interessieren als die zum Vergleich dargestellte Analyse von Literatur, wird insbesondere die kognitive Komponente von Emotionen genutzt, um affektgeladene Inhalte zu kommunizieren. Somit sollten automatische Systeme die Fähigkeit implementiert bekommen, Ereignisse „empathisch“ zu interpretieren – mit einer expliziten Nennung von Emotionen ist in weiten Teilen der Daten nicht zu rechnen. Eine solche direkte Benennung der Emotion durch die Beschreibung der subjektiven Komponente ist nur die zweithäufigste Art und Weise, Emotionen zu beschreiben.

3.2 Modellierung

Nachdem wir nun ein Korpus zur Verfügung haben, welches mit Emotionskomponenten annotiert ist, werden wir untersuchen, ob diese zusätzliche Sichtweise von Emotionen der automatischen Klassifikation zuträglich ist, welche ja für eine groß angelegte Analyse von sozialen Medien notwendig ist. Wir verwenden verschiedene neuronale Klassifikatoren („deep learning“), welche in Abb. 2 zusammengefasst sind. Entsprechend der aktuell üblichen Konfiguration solcher automatischen Klassifikationssysteme bestehen diese Netze aus mehreren Ebenen. Die unterste Ebene sind vortrainierte BERT-Satzeinbettungen als Eingabemerkmale (Devlin et al. 2019). Diese Einbettungen repräsentieren die Semantik der Eingabetexte in einem Vektorraum. Die darauf aufbauende Netzwerkarchitektur stellt sicher, dass die Wortvektoren nicht isoliert, sondern in der Reihenfolge betrachtet werden, in der sie im Text auftreten. Dies ermöglicht auch die Interpretation von Wortfolgen, wie sie zum Beispiel bei Negationen auftreten. Wir kombinieren hier, inspiriert durch Chen und Wang (2018); Sosa (2017), ein bidirektionales LSTM (Hochreiter und Schmidhuber 1997) in Kombination mit einem CNN.

Diese Architektur verwenden wir in einem Klassifikator zur Vorhersage der Emotionskomponenten (*Cpm-NN-Base*) und der Emotionen (*Emo-NN-Base*). Um zu untersuchen, ob nun das Wissen über Emotionskomponenten der Emotionsvorhersage zuträglich ist, vergleichen wir diese einfache Architektur mit einem Emotionsklassifikator, welcher auf die Komponentenvorhersage Zugriff hat (*Emo-Cpm-NN-Pred*) und mit einem Cross-Stitch-Multi-Task-Learning-Modell (Misra et al. 2016) (*MTL-XS*), bei dem Information aus einer Aufgabe genutzt werden kann, um die andere zu lösen.

Diese konfigurierten maschinellen Lernmodelle stellen nun eine große Menge Parameter bereit, welche optimiert werden müssen, um die vorliegende Aufgabenstellung der Zuweisung von Emotionen bzw. Emotionskomponenten zu Text zu lösen. Hierzu teilen wir die vorliegenden Daten in eine Trainingsmenge (zur Optimierung der Parameter) und eine Testmenge (zur Evaluation der Güte des Klassifikators) auf. Die Evaluationsergebnisse sind in Abb. 3 zusammengefasst. Wir sehen hier, dass das Modell, welches die Komponenten

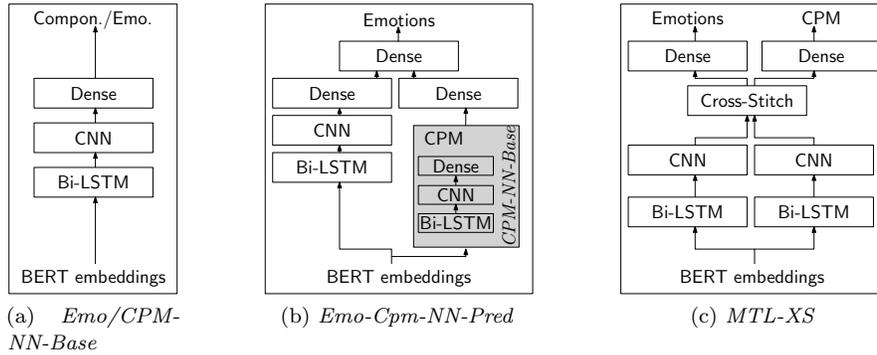


Abbildung 2: Neuronale Modellarchitekturen zur Vorhersage von Emotionskomponenten und Emotionen

nutzt, um Emotionen vorherzusagen (*Emo-Cpm-NN-Pred*), nicht besser ist als das Modell, welches nur Zugriff auf den Text hat (*Emo-NN-Base*). Wenn wir die Vorhersage allerdings durch die Annotation ersetzen (*Emo-Cpm-NN-Gold*) sehen wir eine deutliche Verbesserung – die Vorhersage der Komponenten alleine ist also nicht gut genug, um in einer „Pipeline“ auch Erfolg zu zeigen. Allerdings sehen wir, dass das Multi-Task-Modell (*MTL-XS*) nahezu die selbe Performanz zeigt wie das Modell, welches auf die Komponentennotation zugreift. Wir können also festhalten, dass das Wissen um die Emotionskomponenten auch der Emotionsvorhersage zuträglich ist.

3.3 Zusammenfassung

Wir haben nun in diesem Abschnitt auf Basis eines neuen Emotionskorpus gelernt, dass das Emotionskomponentenmodell hilfreich ist, um zu verstehen, wie Emotionen beschrieben werden. Wir wissen auch, dass in sozialen Medien wohl subjektive Beschreibungen häufiger sind als in der Literatur, wobei dort häufiger auf den Ausdruck einer Emotion Bezug genommen wird. In beiden Domänen spielt aber die kognitive Komponente – die Interpretation von Ereignissen – die wichtigste Rolle. Dies führt uns zu dem nächsten Abschnitt, in dem wir vorschlagen, Emotionen durch Appraisaldimensionen zu repräsentieren.

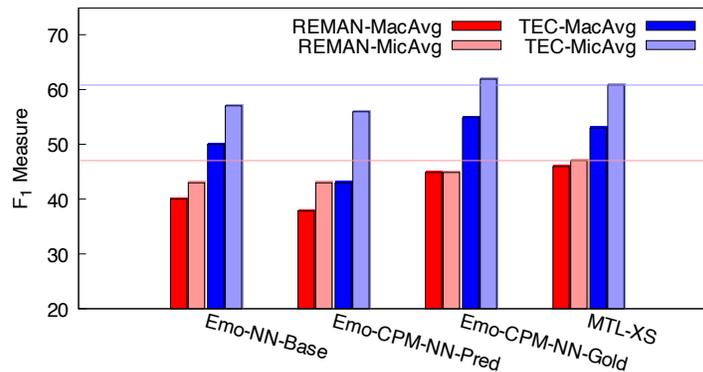


Abbildung 3: Ergebnisse der Emotionsklassifikation unter Berücksichtigung der Emotionskomponenten

4 Klassifikation von Appraisaldimensionen³

Aus diesen Ergebnissen ergibt sich nun die Motivation, insbesondere der Dimension des Appraisals Aufmerksamkeit zu schenken. Uns interessiert jetzt also, wie wir mit komputationellen Mitteln Ereignisse bezüglich ihrer Emotion einordnen können. Ein Beispiel, welches eine solche Einordnung motiviert, ist der Text: „wenn ein Auto ein anderes überholt und ich gezwungen bin von der Straße zu fahren“⁴. Hier wird ein Zwang beschrieben etwas zu tun, was die Person selbst nicht kontrollieren kann und was eher als unangenehm eingeordnet werden dürfte. Hier spielt möglicherweise die generelle Gefahr des Autofahrens eine Rolle, hier dürfte die Person, die den Text verfasst hat, aber vermutlich eher Wut ausdrücken. Ein weiteres Beispiel, aus dem Hatespeech-Korpus von Ross et al. (2016), wäre die Nachricht: „Überall schwangere, muslimische Frauen mit einem Rattenschwanz an Kindern #Islamisierung“. Hierbei wird die/der Autor:in möglicherweise von einer wahrgenommenen fehlenden Kontrolle der Situation geleitet, welche in Kombination mit der individuell als nicht angenehm beurteilten Situation potentiell zu Angst führt. Diese Erkenntnis und Schlussfolgerung kann der (automatischen) Erkennung von Hassrede zuträglich sein. Im folgenden Stellen wir also einen Ansatz vor, wie solche Evaluierungen von Situationen in einem automatischen System genutzt werden können.

³ Dieser Abschnitt basiert auf Hofmann et al. (2020, 2021).

⁴ Ein Beispiel aus dem Korpus von Troiano et al. (2019), übersetzt aus dem Englischen.

Emotion	Appraisal	Text
Wut	Aufmerksamkeit Gewissheit	when my neighbour started to throw rubbish in my garden for no reason.
Ekel	Gewissheit	to watch someone eat insects on television.
Angst	Aufmerksamkeit Aufwand Umstand	when our kitten escaped in the late evening and we thought he was lost.
Schuld	Gewissheit Verantwortung Kontrolle	when I took something without paying.
Freude	Aufmerksamkeit Gewissheit Annehmlichkeit Verantwortung	when I found a rare item I had wanted for a long time.
Traurigkeit	Aufmerksamkeit Gewissheit Aufwand Umstand	when my dog died. He was ill for a while. Still miss him.
Scham	Gewissheit Verantwortung	when I remember an embarrassing social faux pas from my teenage years.

Tabelle 2: Beispiele aus unserem Appraisal-Korpus. Wir nennen die geltenden Appraisal-Dimensionen, alle nicht genannten Variablen sind in der Instanz nicht annotiert worden.

4.1 Korpusannotation

Zur Analyse solcher Beschreibungen benötigen wir ein Korpus, welches vor allem Ereignisse enthält. Wir stützen uns daher auf das enISEAR-Korpus (Troiano et al. 2019), welches von Troiano et al. in Anlehnung an die ursprüngliche ISEAR-Studie (Scherer und Wallbott 1997) entwickelt wurde. Troiano und Kollegen haben Teilnehmer:innen in einer Crowdsourcing-Plattform gebeten, zu einer gegebenen Emotion ein Ereignis zu beschreiben, welches diese Emotion hervorgerufen hat. Dies führte zu 1001 englischen Instanzen, welche gleichverteilt für die Emotionen *Wut*, *Ekel*, *Angst*, *Schuld*, *Freude*, *Traurigkeit*, and *Scham* generiert wurden. Wir annotieren diese für Appraisal-Dimensionen.

Diese Appraisal-Variablen beziehen wir aus der Studie von Smith und Ellsworth, welche gezeigt haben, dass eine relativ niedrige Dimensionalität ausreicht, um Emotionen zu unterscheiden, dabei aber eine größere Mächtigkeit hat als Modelle des Affekts (Smith und Ellsworth 1985). Diese Dimensionen sind, (1) wie angenehm ein Ereignis ist (was wahrscheinlich mit Freude auftritt, aber zum Beispiel nicht mit Ekel), (2) wie viel Aufwand aufzubringen ist (zum Beispiel eher hoch bei Angst), (3) wie sicher die erlebende Person ist, was in der Situation geschieht (niedrig bei Überraschung oder Hoffnung), (4) wie

viel Aufmerksamkeit die Situation bedarf (z.B. niedrig bei Langeweile), (5) wie viel Verantwortung die Person dafür trägt, was geschieht (hoch für das Gefühl von Herausforderung oder Stolz), und (6) wie viel Kontrolle die Person über die Situation hat (z.B. niedrig im Fall von Wut). Im Gegensatz zu der Originalstudie von Smith und Ellsworth sind bei unserer Annotation Ereignisbeschreibungen zu bewerten, welche nicht selbst erlebte Ereignisse darstellen. Daher vereinfachen wir die Annotation zu einer binären Aufgabe, bei der für jede der Dimensionen zu entscheiden ist, ob sie wahrscheinlich für die beschreibende Person galt oder nicht. Des Weiteren teilen wir die Dimension der Kontrolle in „Umstand“ (die Situation konnte von niemandem beeinflusst werden) und die Kontrolle der Person auf. Die von den Annotator:innen zu beantwortenden Fragen waren also (übersetzt aus dem Englischen):

Wahrscheinlich galt für die Person, die das Ereignis beschrieb, zum Zeitpunkt des Ereignisses, dass sie...

- ... dem Ereignis Aufmerksamkeit schenkte. (*Aufmerksamkeit*)
- ... sich sicher war, was geschah. (*Gewissheit*)
- ... physischen oder mentalen Aufwand mit der Situation hatte. (*Aufwand*)
- ... die Situation angenehm fand. (*Annehmlichkeit*)
- ... sich als verantwortlich für das empfand, was geschah. (*Verantwortung*)
- ... die Kontrolle hatte. (*Kontrolle*)
- ... fand, dass das Ereignis von nichts oder niemandem hätte beeinflusst werden können. (*Umstand*)

Jedes Ereignis wurde von drei Annotator:innen markiert, nachdem sie mit Hilfe der Daten aus der Original-ISEAR-Studie (Scherer und Wallbott 1997) angelernt wurden. Hier wurde eine durchschnittliche Übereinstimmung von Cohen's $\kappa=0.67$ erreicht, wobei festzuhalten ist, dass die Annotator:innen keinen Zugriff auf die Emotionsklasse hatten (wenn sie diese kannten, wurde eine Übereinstimmung $> .8$ erreicht, allerdings wäre dies eine Aufgabe, welche nicht einem späteren automatischen System entspräche). Die endgültige Annotation wurde durch einen Mehrheitsbeschluss gebildet, wobei jede:r Annotator:in mit diesem Ergebnis eine Übereinstimmung von $.70-.82\kappa$ zeigte. Beispiele für Daten aus unserem Korpus finden sich in Tabelle 2.

In Abb. 4 zeigen wir eine aggregierte Statistik der Annotation. Wir sehen, dass die Ergebnisse in weiten Teilen den Erkenntnissen von Smith und Ellsworth entsprechen (Smith und Ellsworth 1985, siehe Tabelle 6). Dies ist ein Indiz dafür, dass die Annotationsaufgabe der Bewertung von durch andere Personen beschriebenen Ereignissen tatsächlich möglich ist. Dies ist eine wichtige Erkenntnis, da ansonsten auch automatische Systeme, welche auf den Daten aufbauten, nur einen fragwürdigen Nutzen hätten.

Die am häufigsten annotierte Klasse ist *Gewissheit*, gefolgt von *Aufmerksamkeit*. *Wut* und *Angst* erfordern *Aufmerksamkeit*, *Schuld* und *Scham* dagegen nicht; *Ekel* und *Wut* zeigen die höchste Assoziation mit *Gewissheit*, im Gegensatz zu *Angst*. *Verantwortung* und *Kontrolle* spielen bei *Schuld* und

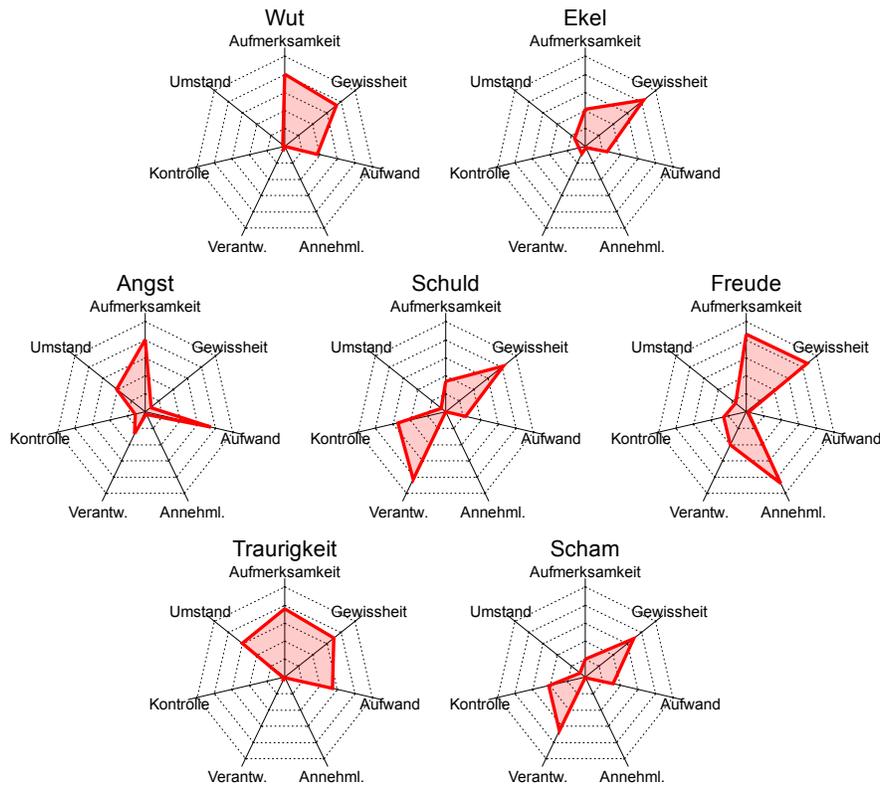


Abbildung 4: Ergebnisse der Appraisalannotation

Scham die größte Rolle, wobei diese Dimensionen bei Schuld deutlich stärker ausgeprägt sind. *Freude* hängt stark mit *Annehmlichkeit* zusammen. *Angst* hat eine klare Verbindung zu *Aufwand*, gemeinsam mit *Traurigkeit* ist sie auch durch die Abwesenheit von *Kontrolle* gekennzeichnet.

4.2 Modellierung

Wir wenden uns nun der Frage zu, ob und wie gut die Appraisaldimensionen automatisch aus Text vorhergesagt werden können.⁵ Dazu verwenden wir

⁵ Aus Platzgründen diskutieren wir hier nicht, inwieweit die Appraisalvorhersagen der Emotionsklassifikation zuträglich sind und verweisen die Leser:in auf unsere vorgehende Publikation (Hofmann et al. 2020). Die hier dargestellten Ergebnisse stammen aus den Experimenten, welche wir in Hofmann et al. (2021) durchgeführt haben.

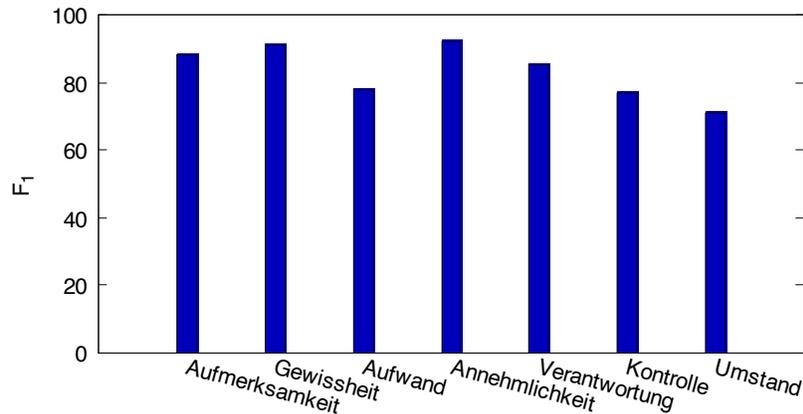


Abbildung 5: Ergebnisse der Vorhersage der Appraisaldimensionen mit Hilfe einer 3×10 -fach Kreuzvalidierung

das vortrainierte Sprachmodell RoBERTa-base (Liu et al. 2019) mit der Abstraktion, wie sie durch ktrain (Maiya 2020) zur Verfügung gestellt wird. Um bei der verhältnismäßig kleinen Datenmenge ein möglichst zuverlässiges Ergebnis zu erlangen, verwenden wir eine 3×10 -fach Kreuzvalidierung und geben die Durchschnittswerte an.

Diese Ergebnisse zeigen wir in einer Zusammenfassung in Abb. 5. Die Performanz, gemessen als F_1 , also dem harmonischen Mittel zwischen Vollständigkeit und Genauigkeit, liegt zwischen 0,71 und 0,92 und ist somit für alle Dimensionen zufriedenstellend. Die höheren Werte werden für die Klassen erreicht, welche in den Daten häufiger auftreten, sodass erwartet werden kann, dass die Erstellung größerer Korpora auch zu weiteren Verbesserungen führen dürfte.

4.3 Zusammenfassung

In diesem Abschnitt des Kapitels haben wir nun eine Operationalisierung der kognitiven Komponente von Emotionen vorgestellt. Diese Komponente verknüpft Kognition mit Emotion und stellt somit einen wichtigen Aspekt in der automatischen Analyse dar. Unsere Ergebnisse basieren auf einer manuellen Korpusannotation und einem komputationellen Modell, welches auf maschinellem Lernen basiert. Sowohl die Korpusanalyse als auch die Modellierung zeigen, dass die gewählten Dimensionen eine schlüssige Alternative zu etablierten Emotionsanalyseverfahren darstellen. Diese Repräsentation mit Hilfe der Appraisaldimensionen ist insbesondere für eine automatische Verarbeitung von Ereignisbeschreibungen geeignet, welche, wie wir in Abschnitt 3 gesehen haben, besonders häufig in sozialen Medien und in der Literatur auftreten.

Aus der Arbeit von Plaza-del-Arco et al. (2021) wissen wir, dass die Erkennung von Hassrede in sozialen Medien durch Emotionserkennung unterstützt wird, automatische Klassifikationsverfahren also insbesondere eine höhere Abdeckung erreichen und weniger Hassrede unerkannt bleibt. Unsere hier vorgestellten Studien motivieren nun, diese Arbeiten fortzusetzen und Appraisaltheorien und die Erkennung von Hassrede zu integrieren. Wir erwarten, dass insbesondere implizit formulierte Hassrede, in den Beschreibungen von Situationen und Ereignissen genutzt werden, von diesem Ansatz profitiert.

5 Fazit

In diesem Kapitel haben wir die Verwendung des Emotionskomponentenprozessmodells und von Appraisaltheorien für die Emotionsanalyse diskutiert und entsprechende Ressourcen und maschinelle Lernverfahren vorgestellt. Wir konnten damit einen Beitrag zu dem Verständnis leisten, wie Emotionen ausgedrückt und (maschinell) erkannt werden. Daraus ergeben sich nun eine Reihe weiterer Forschungsfragen.

Wir haben die Appraisalanalyse in Abschnitt 4 mit Hilfe eines dedizierten Korpus von Ereignisbeschreibungen entwickelt. Dies ist für diese isolierte Betrachtung eine sinnvolle Wahl, um genau dieses Phänomen untersuchen zu können. Allerdings ist hiermit noch nicht gezeigt worden, dass die Ereignisbeschreibungen in allgemeineren Korpora, wie solchen, die in Abschnitt 3 Verwendung fanden, auch sinnvoll nutzbar sind. Dieser Fragestellung sind wir zum Teil, aber nur mit automatischen Annotationsverfahren, in Hofmann et al. (2021) nachgegangen. Weiterhin offen ist, ob sich mit Hilfe der Appraisalannotation auch eine Verbesserung der Emotionskategorisierung herbeiführen lässt. Dies konnten wir bisher nicht zeigen, erwarten aber, dass die Entwicklung hinreichend großer Textkorpora diesen methodischen Zusammenhang zukünftig sichtbar machen wird (Hofmann et al. 2020). Festzuhalten ist aber allemal, dass die Appraisaldimensionen, und auch die Emotionskomponenten, eine neuartige Perspektive auf die Erlebnisse von Nutzer:innen sozialer Medien und Figuren in literarischen Texten bieten.

Die Emotionsanalyse fand bereits in der Vergangenheit in anderen Aufgaben Anwendung, auch in der Analyse von Hassrede und Beleidungen. Es ist bereits bekannt, dass auf die Ziele von Hassrede mit Hilfe unterschiedlicher Stilmittel referenziert wird, z.B. durch metaphorische Referenzen auf Tiere, Körperteile oder mentale Zustände (Lemmens et al. 2021; Silva et al. 2016). Wir schlagen hier vor, dem Ausdruck der Emotion weitere Aufmerksamkeit zu schenken, da auch hier implizite Formulierungen verwendet werden könnten, zum Beispiel Beschreibungen von Ereignissen. Dies ist auch aus Sicht der Definition von Hassrede schlüssig: Hassrede enthält häufig einen Aufruf zu Taten. Diese beschriebenen Aktionen stellen Ereignisse dar, welche bezüglich ihrer emotionalen Wirkung berücksichtigt werden müssen. Es sollten also Text-

korpora erstellt werden, welche zur Untersuchung dienen, wie die Emotion in der Hassrede kommuniziert wird (z.B. in Anlehnung an unsere Komponentenanalyse in Abschnitt 3) und welche Bewertungen dieser durch die Autor:innen und Adressaten der Beleidigungen und Hassreden durchgeführt werden (z.B. in Anlehnung an unsere Appraisalanalysen in Abschnitt 4).

Danksagung

Wir bedanken uns bei unseren Koautor:innen vorhergehender Arbeiten, welche zum Teil in dieses Kapitel eingeflossen sind, insbesondere bei Enrica Troiano, Laura Oberländer, Felix Casel, Amelie Heindl, Jan Hofmann und Kai Sassenberg. Diese Arbeit wurde durch die Deutsche Forschungsgemeinschaft unterstützt (Projekte SEAT, KL 2869/1-1 und CEAT, KL 2869/1-2).

Literatur

- Abdul-Mageed, M. und L. Ungar. 2017. EmoNet: Fine-grained emotion detection with gated recurrent neural networks. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vancouver, Canada:718–728. Association for Computational Linguistics. <http://dx.doi.org/10.18653/v1/P17-1067>.
- Alm, C.O., D. Roth, und R. Sproat. 2005. Emotions from text: Machine learning for text-based emotion prediction. *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, Vancouver, British Columbia, Canada:579–586. Association for Computational Linguistics. <http://dx.doi.org/10.3115/1220575.1220648>.
- Aman, S. und S. Szpakowicz. 2007. Identifying expressions of emotion in text. *International Conference on Text, Speech and Dialogue*:196–205. Springer. http://dx.doi.org/10.1007/978-3-540-74628-7_27.
- Bostan, L.A.M. und R. Klinger. 2018. An analysis of annotated corpora for emotion classification in text. *Proceedings of the 27th International Conference on Computational Linguistics*, Santa Fe, New Mexico, USA:2104–2119. Association for Computational Linguistics.
- Buechel, S., J. Hellrich, und U. Hahn. 2016. Feelings from the Past—Adapting affective lexicons for historical emotion analysis. *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)*, Osaka, Japan:54–61. The COLING 2016 Organizing Committee.
- Buechel, S. und U. Hahn. 2017b. EmoBank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis. *Proceedings of the 15th Conference of the European Chapter of*

- the Association for Computational Linguistics: Volume 2, Short Papers*, Valencia, Spain:578–585. Association for Computational Linguistics.
- Buechel, S. und U. Hahn. 2017a. EmoBank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, Valencia, Spain:578–585. Association for Computational Linguistics.
- Buechel, S., L. Modersohn, und U. Hahn. 2021. Towards label-agnostic emotion embeddings. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Online and Punta Cana, Dominican Republic:9231–9249. Association for Computational Linguistics. <http://dx.doi.org/10.18653/v1/2021.emnlp-main.728>.
- Campero, A., B. Felbo, J.B. Tenenbaum, und R. Saxe. 2017. A first step in combining cognitive event features and natural language representations to predict emotions. *Conference on Cognitive Computational Neuroscience*.
- Casel, F., A. Heindl, und R. Klinger. 2021. Emotion recognition under consideration of the emotion component process model. *Proceedings of the 17th Conference on Natural Language Processing (KONVENS 2021)*, Düsseldorf, Germany:49–61. KONVENS 2021 Organizers.
- Cevher, D., S. Zepf, und R. Klinger. 2019. Towards multimodal emotion recognition in german speech events in cars using transfer learning. *Proceedings of the 15th Conference on Natural Language Processing (KONVENS 2019): Long Papers*, Erlangen, Germany:79–90. German Society for Computational Linguistics & Language Technology.
- Chen, N. und P. Wang. 2018. Advanced combined LSTM-CNN model for Twitter sentiment analysis. *2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS)*:684–687. <http://dx.doi.org/10.1109/CCIS.2018.8691381>.
- Clore, G.L. und A. Ortony. 2013. Psychological Construction in the OCC Model of Emotion. *Emotion Review* 5(4):335–343.
- Devlin, J., M.W. Chang, K. Lee, und K. Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Minneapolis, Minnesota:4171–4186. Association for Computational Linguistics. <http://dx.doi.org/10.18653/v1/N19-1423>.
- Ekman, P. 1992. An argument for basic emotions. *Cognition & emotion* 6(3-4):169–200. <http://dx.doi.org/10.1080/02699939208411068>.
- Felbo, B., A. Mislove, A. Søgaard, I. Rahwan, und S. Lehmann. 2017. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, Copenhagen, Denmark:1615–1625. Association for Computational Linguistics. <http://dx.doi.org/10.18653/v1/D17-1169>.

- Gendron, M., D. Roberson, J.M. van der Vyver, und L. Feldman Barrett. 2014. Perceptions of emotion from facial expressions are not culturally universal: evidence from a remote culture. *Emotion* 14(2):251–262. <http://dx.doi.org/10.1037/a0036052>.
- Hochreiter, S. und J. Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9(8):1735–1780. <http://dx.doi.org/10.1162/neco.1997.9.8.1735>.
- Hofmann, J., E. Troiano, K. Sassenberg, und R. Klinger. 2020. Appraisal theories for emotion classification in text. *Proceedings of the 28th International Conference on Computational Linguistics*, Barcelona, Spain (Online):125–138. International Committee on Computational Linguistics. <http://dx.doi.org/10.18653/v1/2020.coling-main.11>.
- Hofmann, J., E. Troiano, und R. Klinger. 2021. Emotion-aware, emotion-agnostic, or automatic: Corpus creation strategies to obtain cognitive event appraisal annotations. *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, Online:160–170. Association for Computational Linguistics.
- Izard, C.E. 1971. *The face of emotion*. New York: Appleton-Century-Crofts.
- Kim, E., S. Padó, und R. Klinger. 2017. Investigating the relationship between literary genres and emotional plot development. *Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, Vancouver, Canada:17–26. Association for Computational Linguistics.
- Kim, E. und R. Klinger. 2018. Who feels what and why? annotation of a literature corpus with semantic roles of emotions. *Proceedings of the 27th International Conference on Computational Linguistics*:1345–1359. Association for Computational Linguistics.
- Kim, E. und R. Klinger. 2019. An analysis of emotion communication channels in fan-fiction: Towards emotional storytelling. *Proceedings of the Second Workshop on Storytelling*, Florence, Italy:56–64. Association for Computational Linguistics. <http://dx.doi.org/10.18653/v1/W19-3406>.
- Klinger, R., O. De Clercq, S. Mohammad, und A. Balahur. 2018. IEST: WASSA-2018 implicit emotions shared task. *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, Brussels, Belgium:31–42. Association for Computational Linguistics. <http://dx.doi.org/10.18653/v1/W18-6206>.
- Lemmens, J., I. Markov, und W. Daelemans. 2021. Improving Hate Speech Type and Target Detection with Hateful Metaphor Features. *Proceedings of the Fourth Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda*, Online:7–16. Association for Computational Linguistics.
- Li, Y., H. Su, X. Shen, W. Li, Z. Cao, und S. Niu. 2017. DailyDialog: A manually labelled multi-turn dialogue dataset. *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume*

- 1: *Long Papers*), Taipei, Taiwan:986–995. Asian Federation of Natural Language Processing.
- Liu, Y., M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, und V. Stoyanov. 2019. RoBERTa: A robustly optimized bert pretraining approach. ArXiv.
- Maiya, A.S. 2020. ktrain: A low-code library for augmented machine learning. arXiv:2004.10703.
- Misra, I., A. Shrivastava, A. Gupta, und M. Hebert. 2016. Cross-stitch networks for multi-task learning. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*:3994–4003.
- Mohammad, S. 2012. #emotional tweets. **SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, Montréal, Canada:246–255. Association for Computational Linguistics.
- Mohammad, S., X. Zhu, und J. Martin. 2014. Semantic role labeling of emotions in tweets. *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, Baltimore, Maryland:32–41. Association for Computational Linguistics. <http://dx.doi.org/10.3115/v1/W14-2607>.
- Mohammad, S. und F. Bravo-Marquez. 2017. WASSA-2017 shared task on emotion intensity. *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, Copenhagen, Denmark:34–49. Association for Computational Linguistics. <http://dx.doi.org/10.18653/v1/W17-5205>.
- Mohammad, S., F. Bravo-Marquez, M. Salameh, und S. Kiritchenko. 2018. SemEval-2018 task 1: Affect in tweets. *Proceedings of The 12th International Workshop on Semantic Evaluation*, New Orleans, Louisiana:1–17. Association for Computational Linguistics. <http://dx.doi.org/10.18653/v1/S18-1001>.
- Mohammad, S.M. und P.D. Turney. 2012. Crowdsourcing a word-emotion association lexicon. *Computational Intelligence* 29(3):436–465.
- Mohammad, S.M., P. Sobhani, und S. Kiritchenko. 2017. Stance and sentiment in tweets. *ACM Trans. Internet Technol.* 17(3):1–23.
- Mohammad, S.M. 2018. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. *Proceedings of The Annual Conference of the Association for Computational Linguistics (ACL)*, Melbourne, Australia.
- Oberländer, L.A.M. und R. Klinger. 2020. Token sequence labeling vs. clause classification for English emotion stimulus detection. *Proceedings of the Ninth Joint Conference on Lexical and Computational Semantics*, Barcelona, Spain (Online):58–70. Association for Computational Linguistics.
- Pennebaker, J.W., M.E. Francis, und R.J. Booth. 2001. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates* 71:2001.
- Picard, R. 1997. *Affective Computing*. MIT Press.

- Plaza-del-Arco, F.M., S. Halat, S. Padó, und R. Klinger. 2021. Multi-task learning with sentiment, emotion, and target detection to recognize hate speech and offensive language. *arXiv:2109.10255*.
- Plutchik, R. 2001. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist* 89(4):344–350.
- Preoțiuc-Pietro, D., H.A. Schwartz, G. Park, J. Eichstaedt, M. Kern, L. Ungar, und E. Shulman. 2016. Modelling valence and arousal in Facebook posts. *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, San Diego, California:9–15. Association for Computational Linguistics. <http://dx.doi.org/10.18653/v1/W16-0404>.
- Rajamanickam, S., P. Mishra, H. Yannakoudakis, und E. Shutova. 2020. Joint modelling of emotion and abusive language detection. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online:4270–4279. Association for Computational Linguistics. <http://dx.doi.org/10.18653/v1/2020.acl-main.394>.
- Ross, B., M. Rist, G. Carbonell, B. Cabrera, N. Kurowsky, und M. Wojatzki. 2016, sep. Measuring the Reliability of Hate Speech Annotations: The Case of the European Refugee Crisis. *Proceedings of NLP4CMC III: 3rd Workshop on Natural Language Processing for Computer-Mediated Communication*. M. Beißwenger, M. Wojatzki, und T. Zesch (Hg.), Volume 17 of *Bochumer Linguistische Arbeitsberichte*, Bochum:6–9.
- Rosso, P., F. Rangel, I.H. Farías, L. Cagnina, W. Zaghouni, und A. Charfi. 2018. A survey on author profiling, deception, and irony detection for the arabic language. *Language and Linguistics Compass* 12(4):e12275. <http://dx.doi.org/https://doi.org/10.1111/lnc3.12275>.
- Russell, J. und A. Mehrabian. 1977. Evidence for a three-factor theory of emotions. *Journal of Research in Personality* 11(3):273–294.
- Sarawagi, S. 2008. *Information Extraction*. Now Publishers.
- Scarantino, A. 2008. The philosophy of emotions and its impact on affective science. *Handbook of emotions*. J. H.-J. M. Lewis und L. F. Barrett (Hg.):3–48. New York: Guilford Press.
- Scherer, K.R. und H.G. Wallbott. 1997. The ISEAR questionnaire and codebook. Geneva Emotion Research Group.
- Scherer, K.R., A. Schorr, und T. Johnstone. 2001. *Appraisal processes in emotion: Theory, methods, research*. Oxford University Press.
- Scherer, K.R. 2005b. What are emotions? and how can they be measured? *Social Science Information* 44(4):695–729. <http://dx.doi.org/10.1177/0539018405058216>.
- Scherer, K.R. 2005a. What are emotions? And how can they be measured? *Social Science Information* 44(4):695–729. [arXiv:https://doi.org/10.1177/0539018405058216](https://doi.org/10.1177/0539018405058216), <http://dx.doi.org/10.1177/0539018405058216>.

- Scherer, K.R. 2009. Emotions are emergent processes: they require a dynamic computational architecture. *Philosophical Transactions of the Royal Society B* 364(1535):3459–3474. <http://dx.doi.org/10.1098/rstb.2009.0141>.
- Schuff, H., J. Barnes, J. Mohme, S. Padó, und R. Klinger. 2017. Annotation, modelling and analysis of fine-grained emotions on a stance and sentiment detection corpus. *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, Copenhagen, Denmark:13–23. Association for Computational Linguistics. <http://dx.doi.org/10.18653/v1/W17-5203>.
- Shaikh, M.A.M., H. Prendinger, und M. Ishizuka. 2009. A linguistic interpretation of the OCC emotion model for affect sensing from text. *Affective Information Processing*. J. Tao und T. Tan (Hg.) 4:45–73. London: Springer. http://dx.doi.org/10.1007/978-1-84800-306-4_4.
- Silva, L., M. Mondal, D. Correa, F. Benevenuto, und I. Weber. 2016. Analyzing the targets of hate in online social media. *Proceedings of the International AAAI Conference on Web and Social Media*:687–690.
- Smith, C.A. und P.C. Ellsworth. 1985. Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology* 48(4):813–838.
- Sosa, P.M. 2017. Twitter sentiment analysis using combined LSTM-CNN models. ArXiv.
- Strapparava, C. und A. Valitutti. 2004. WordNet affect: an affective extension of WordNet. *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal. European Language Resources Association (ELRA).
- Strapparava, C. und R. Mihalcea. 2007. SemEval-2007 task 14: Affective text. *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, Prague, Czech Republic:70–74. Association for Computational Linguistics.
- Tooby, J. und L. Cosmides. 2008. The evolutionary psychology of the emotions and their relationship to internal regulatory variables. *Handbook of emotions*. J. H.-J. M. Lewis und L. F. Barrett (Hg.):114–137. New York: Guilford Press.
- Troiano, E., S. Padó, und R. Klinger. 2019. Crowdsourcing and validating event-focused emotion corpora for German and English. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy:4005–4011. Association for Computational Linguistics. <http://dx.doi.org/10.18653/v1/P19-1391>.
- van Meel, J.M. 1995. Representing emotions in literature and paintings: A comparative analysis. *Poetics* 23(1):159 – 176. Emotions and Cultural Products. [http://dx.doi.org/10.1016/0304-422X\(94\)00013-V](http://dx.doi.org/10.1016/0304-422X(94)00013-V).
- Xia, R. und Z. Ding. 2019. Emotion-cause pair extraction: A new task to emotion analysis in texts. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy:1003–1012. Association for Computational Linguistics. <http://dx.doi.org/10.18653/v1/P19-1096>.