Intelligentes, spielorientiertes Crowdsourcing für die Verarbeitung von Audiodaten[1]

Simone Hantke[2]

Abstract: Heutige Sprachassistenzsysteme sind ein fester Bestandteil unseres modernen Lebens. Der Erfolg dieser Technologien ist jedoch maßgeblich auf die Menge und Qualität der annotierten Trainingsdaten zurückzuführen. Dafür ist eine große Anzahl an Sprechern und Annotatoren erforderlich, und zudem erhebliche Investitionen notwendig, um diese Datenressourcen zu strukturieren und zu annotieren. Die Verfahren zur Datenerstellung sind kostspielig, zeitaufwendig und mühsam, so dass derzeit qualitativ hochwertig annotierte Daten knapp sind. In dieser Arbeit wird daher eine intelligente, Crowdsourcing-basierte Plattform mit Spielelementen und innovativen maschinellen Lernalgorithmen für die Datensammlung und Annotation vorgestellt. Es wurden diverse Audiodaten gesammelt und die Plattform in einer Vielzahl von Klassifikations- und Spracherkennungsstudien sowie mittels Perzeptionsstudien evaluiert. Durch die eingeführten Verfahren kann künftig die Audiodatenerfassung in erheblichem Maße beschleunigt sowie kostengünstiger und zuverlässiger durchgeführt werden.

1 Einführung und Motivation

Wir leben in technisch aufregenden und schnelllebigen Zeiten. Heutzutage ist es selbstverständlich, dass wir mit sprachgesteuerten Assistenzsystemen in unserem täglichen Leben in Kontakt kommen, da sie in Alltagstechnologien wie ALEXA, CORTANA und SIRI eingebettet sind.

Zu Beginn der Entwicklung automatischer Sprachanalysesysteme beschäftigte man sich mit relativ kleinen Datensätzen, die hauptsächlich zum Aufbau kleiner automatischer Systeme und zu dessen Validierung verwendet wurden [DBB52]. Im Laufe der Zeit hat sich viel geändert: Es wurden robuste und echtzeitfähige Systeme entwickelt und eine breite Palette an Algorithmen und Methoden des maschinellen Lernens unterstützt die Entwickler nun bei der Datenanalyse [Go16]. Der Erfolg von statistisch basierten Methoden zur automatischen Sprachverarbeitung wie z.B. Deep Learning [Sc15a] – welches im maschinellen Lernen schnell allgegenwärtig wurde – hat den Bedarf an größeren Eingangsdaten dramatisch erhöht. Diese Entwicklungen erfordern eine erhöhte Anzahl an Sprechern und Annotatoren und damit eine erhebliche Investition in Datenressourcen.

Obwohl eine große Menge an Daten frei verfügbarer ist, sind diese meist unstrukturiert und es mangelt ihnen an zuverlässigen Annotationen. Die Strukturierung und Annotation der Daten ist jedoch mit einem erheblichen zeitlichen und finanziellen Aufwand verbunden. Eine mögliche Lösung ist in Form einer Technik namens ‘Crowdsourcing’ [Ho06] gekommen. Hierbei werden Annotationsaufgaben an eine nicht spezifische Gruppe von

[1] Englischer Titel der Dissertation: ”Intelligent Gamified Crowdsourcing for Audio Processing”

[2] MISP Group, Technische Universität München, Germany, simone.hantke@tum.de

Personen ins Internet auslagert, wobei die Annotatoren meisten nicht speziell ausgebildet sind. Auf diese Weise kann Crowdsourcing überall auf der Welt und zu jeder Zeit für viele verschiedene Interessensbereiche genutzt werden und bietet sofortigen Zugang zu einer breiten und vielfältigen Palette an Personen mit unterschiedlichen Hintergründen, Kenntnissen und Fähigkeiten.

Defizite der aktuellen Technik Obwohl heutige intelligente Systeme autonomer arbeiten denn je, erfordern sie immer noch menschliche Interaktion. Nicht alle Aspekte der menschlichen Intelligenz können eigenständig verarbeiten werden. Der weiterhin benötigte Annotationsvorgang ist jedoch eine kostspielige, zeitaufwändige und mühsame Angelegenheit, die nicht jeder investieren möchte [Ra10]. Dies hat zu einer Verknappung an annotierten Daten geführt und dadurch das Wachstum und den Erfolg bei der Entwicklung von intelligenten Systemen verlangsamt. Ein effizienterer Prozess der Datensammlung und Datenannotation wäre hier ein großer Gewinn. Es fehlt ein automatisiertes kombiniertes System, um multimodale Daten zu beschaffen, neue Datensätze aufzuzeichnen und, was noch wichtiger ist, die Daten zeit- und kosteneffizienter zu annotieren.

Zielsetzung und Auflistung der Beiträge In Anbetracht der Defizite modernster Technologien besteht der Hauptbeitrag der Dissertation [Ha19] in der bahnbrechenden intelligenten, spielorientierten Crowdsourcing Plattform iHEARu-PLAY, mit welcher schneller, kosteneffizienter und zuverlässiger Sammlung und Annotation audiovisueller Daten erfolgen können. Als Alternative zu herkömmlichen Crowdsourcing Plattformen geht iHEARu-PLAY über den aktuellen Stand der Technik hinaus und motiviert Teilnehmer, indem es ihnen eine spielerische und unterhaltsame Umgebung bietet, in der die Spieler freiwillig zu wissenschaftlichen Forschungsprojekten beitragen können. Als Kernkomponente enthält die Plattform erweiterte und verbesserte Algorithmen für maschinelles Lernen, um die Genauigkeit der manuellen Annotation mit der Geschwindigkeit und Kosteneffizienz von maschinellem Lernen auf der Basis von Klassifikatoren zu kombinieren. Indem nur dann um manuelle Annotation gebeten wird, wenn das System selbst nicht in der Lage ist, die Daten zu annotieren, kann der manuelle Arbeitsaufwand auf ein Minimum beschränkt werden. Darüber hinaus werden hochqualitativere Annotationen gesammelt, da mehrere Qualitätsbewertungsmerkmale eingeführt werden. Diese berechnen zuverlässig beispielsweise die Intra–Annotator– und Inter–Annotator–Übereinstimmung durch die neu entwickelte Vertrauenswürdigkeitsbewertung eines Annotators.

2 iHEARu-PLAY: Intelligente, spielorientierte Crowdsourcing Plattform

Kern dieser Arbeit ist es die modulare, intelligente, spielorientierte Crowdsourcing Plattform iHEARu-PLAY[3] [Ha15] vorzustellen. Die browserbasierte Plattform läuft auf jedem Standard-PC oder Smartphone und bietet Audio-, Video- und Bildannotation für eine Vielzahl von Annotationsaufgaben sowie audio(-visuelle) Datensammlung und Analysen.

[3] https://www.ihearu-play.eu

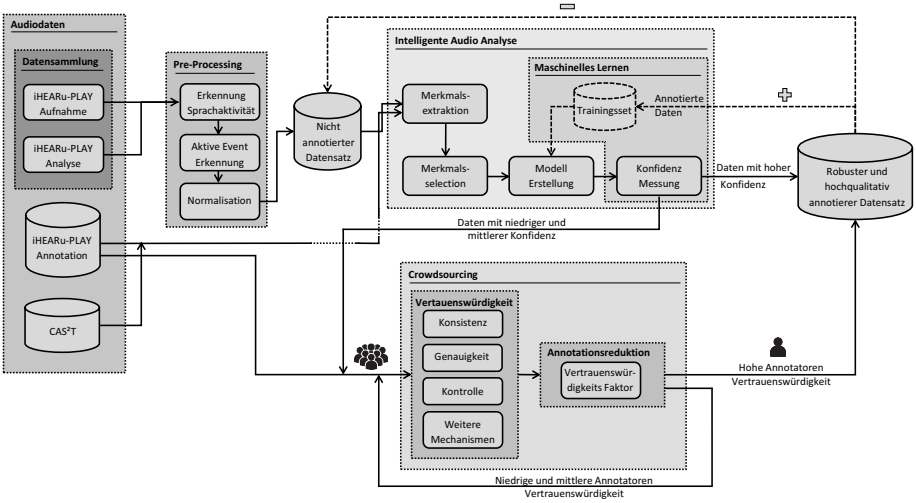

Abb. 1: *In* iHEARu-PLAY *integrierte Interaktion zwischen den Komponenten der intelligenten Audioanalyse, dem aktiven Lernen und Qualitätsmanagement, einschließlich der Vertrauenswürdigkeitsberechnung und der Komponente zur Annotationsreduktion; nach [HZS17, HAS18a].*

Das System hat eine Reihe von Vorteilen gegenüber herkömmlichen Crowdsourcing Plattformen, unter anderem die Anwendung und Integration verschiedener Spielelemente [HAS18b]. In diesem Zusammenhang umfasst die Plattform eine Kombination aus Punkten, Ranglisten, Abzeichen und eine soziale Kommunikationsplattform, um die alltägliche Arbeit des Annotierens zu einer angenehmeren und motivierenden Aufgabe zu machen. Darüber hinaus wurde ein faires, verständliches und offenes Datenschutzkonzept für die Sammlung, Speicherung, Nutzung und Weitergabe von Daten entwickelt, um Missbrauch zu vermeiden und die Anonymität der Annotatoren zu gewährleisten, während gleichzeitig sichergestellt wird, dass unbefugte Dritte keinen Zugang erhalten. Neben diesen Merkmalen ist das Hauptcharakteristikum von iHEARu-PLAY – das es von herkömmlichen Crowdsourcing Plattformen unterscheidet – die Integration mehrerer vertrauenswürdigkeitsbasierter Algorithmen des maschinellen Lernens, die darauf abzielen, die manuelle Annotationsarbeit zu reduzieren. Eine schematische Übersicht über die kombinierten intelligenten Komponenten ist in Abbildung 1 dargestellt.

3 Vertrauenswürdigkeitsbasiertes kooperatives Lernen

Trotz der vielversprechenden Möglichkeiten bedeutet die Online–Rekrutierung von Annotatoren immer noch einen großen Aufwand und es müssen mindestens so viele Annotationen gesammelt werden, wie es nicht annotierte Dateninstanzen gibt. In jüngster Zeit wurden mehrere intelligente Ansätze vorgestellt, um die Belastung der Annotatoren durch nicht annotierte Daten zu reduzieren, wobei einer der vielversprechendsten Ansätze das sogenannte AKTIVE LERNEN (AL) ist [Zh15]. Das Konzept von AL basiert auf der Idee, dass der Algorithmus die Klassifikationsgenauigkeit mit so wenig Trainingsdaten wie möglich

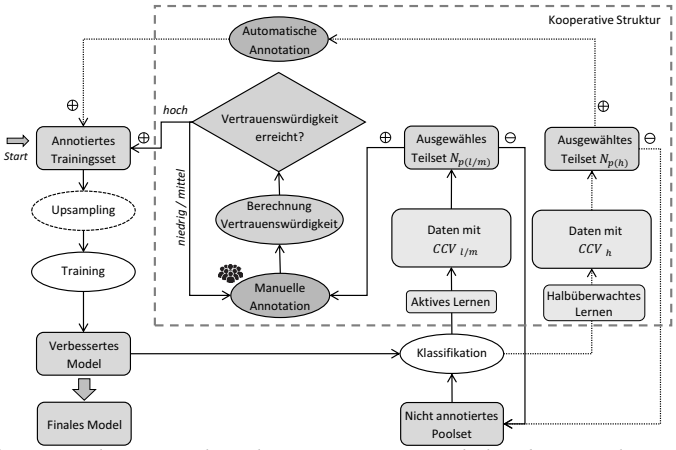

Abb. 2: *Flussdiagramm des TCoL-Algorithmus (Vertrauenswürdigkeitsbasiertes kooperatives Lernen); CCV ist der Vertrauenswert des Klassifikators. Angepasst und erweitert von [Zh15].*

verbessern kann, indem aktiv die Daten ausgewählt werden, bei denen der Algorithmus am sichersten ist [TK11]. Dies bedeutet, dass nur die Daten manuell annotiert werden müssen, für die das trainierte Modell kein Label vorhersagen kann. Dies führt zu einer allgemeinen Reduktion der benötigten Annotationen (CR), wobei eine mindestens gleichwertige Klassifikationsgenauigkeit erzielt werden kann [HZS17].

Daher wird die Anwendbarkeit des entwickelten vertrauenswürdigkeitsbasierten kooperativen Lernalgorithmus (TCoL) untersucht, der AL mit halbüberwachtem Lernen (SSL) auf Basis der Vorhersageunsicherheit bei Support Vector Machines (SVMs) integriert. Dieser ermöglicht eine optimale Aufteilung von manuellen und automatischen Annotationstätigkeiten. Der schematische Ablauf ist in Abbildung 2 dargestellt.

Exemplarische Studie

Der entwickelte TCoL-Algorithmus wurde in einer Reihe von Studien zur Emotionserkennung mit dem *FAU Aibo Emotion Corpus* [St09] evaluiert, wobei sowohl konventionelle als auch auf Vertrauenswürdigkeit basierende Annotationen verwendet wurden. In den Experimenten wurden über 15.000 Audio-Instanzen verwendet mit insgesamt nahezu 7,5 Stunden kontinuierlicher Sprache. 14 Annotatoren (3 weibliche und 11 männliche, im Alter zwischen 20 und 27 Jahren; drei gaben ihr Alter nicht an) annotierten die *FAU Aibo Emotion Corpus* Instanzen in die vorgeschlagenen Emotionsklassen [St09].

Es wurde das IS09-Merkmals-Set [SSB09] angewandt, das auf eine hohe Robustheit für die menschliche Emotionserkennung ausgelegt ist. Die Audiomerkmale wurden mit dem *openSMILE*-Toolkit [Ey13] extrahiert und es wurde die Open-Source-Software für maschinelles Lernen und Data-Mining *WEKA* [Ha09] eingesetzt. Eine SVM mit einem SMO-Algorithmus und einer Komplexitätskonstante C von 0,1 wurde trainiert, um eine Hyperebene zu konstruieren, die die Instanzen verschiedener Klassen trennt. Ein passiver Lernansatz (PL), welcher alle Daten zufällig annotiert, war die Vergleichsbasis.

Die Modelle wurden zunächst mit 200 zufällig ausgewählten Instanzen trainiert, während die restlichen Daten als nicht annotierter Datenpool verwendet wurde. Bei jeder Iteration wählte der Algorithmus 200 Instanzen für die manuelle Annotation aus dem Pool-Set aus. Es wurde ein zufälliges Upsampling angewandt, bei dem Kopien vorhandener Instanzen zu den Klassen hinzugefügt werden, die eine nur geringe Anzahl von Instanzen haben [Pr00]. Je nach Algorithmus wurde nach dem AL-Schritt ein SSL-Schritt durchgeführt, bei dem automatisch 200 Instanzen aus dem Pool-Set annotiert wurden. Der Lernprozess jedes Algorithmus wurde gestoppt, wenn das Leistungsplateau erreicht wurde. Dies wurde mit Hilfe des Unweighted Average Recalls (UAR) bewertet. Um statistische Ungenauigkeiten auszuschließen, wurde jeder Algorithmus außerdem 20 Mal mit zufällig ausgewählten Anfangs-Trainingsinstanzen wiederholt.

Ergebnisauszug und Diskussion

Abbildung 3 zeigt, dass bei allen Ansätzen das sequentielle Hinzufügen von manuell annotierten Instanzen zu einem anfänglichen Trainingsset zu einer kontinuierlichen Verbesserung des Klassifikators führt. Zudem steigt der UAR bei allen Algorithmen zunächst mit der Anzahl der manuellen Annotationen steil an und stagniert anschließend. Die Ergebnisse zeigen, dass alle konventionellen Ansätze PL mit einem UAR von 54,17 % übertreffen. PL erforderte 15.000 Annotationen, um einen maximalen UAR von 57,18 % für die vertrauenswürdigkeitsbasierten Experimente (TCoL) und 54,17 % für die konventionellen zu erreichen. Der TCoL Ansatz reduzierte die Annotationslast erheblich, wobei die relative Kostenreduktion CR bis zu 72 % betrug, während der UAR bei 71,17 % lag. Während die einzelnen (T)AL und SSL Techniken ungefähr den gleichen UAR wie die (T)CoL Algorithmen erreichten, war eine wesentlich höhere Anzahl von Annotationen erforderlich. Daher hat der entwickelte TCoL Algorithmus klare Vorteile gegenüber PL, AL und SSL Ansätzen, da er die Anzahl der erforderlichen manuellen Annotationen effektiv reduziert.

4 Datensammlungen

Aktuelle Technologien bieten die Möglichkeit über das Internet massenhaft neue Daten zu sammeln, wobei in Laptop-PCs, Tablets und Smartphones eingebettete Mikrofone genutzt werden. Aufgrund dessen ist es heutzutage möglich, Sprachdaten unter realen Bedingungen (z.B. mit verschiedenen Mikrofontypen oder Hintergrundgeräuschen) von einer großen Anzahl an Sprechern mit unterschiedlicher geographischer oder kultureller Herkunft, Sprachen oder Altersgruppen zu sammeln. In diesem Zusammenhang wurde in IHEARU-PLAY eine Aufnahmefunktion integriert, die eine multimodale, großflächige Datenerfassung auf ressourcenschonende Weise ermöglicht. Die Anwender können ein breites Spektrum an Aufnahmeaufgaben durchführen und haben die Möglichkeit, das interaktive, webbasierte Sprachklassifizierungs-Framework VOILA zu nutzen [Ha18b]. Dabei können Nutzer den Klassifizierungsprozess verbessern, indem sie qualitativ hochwertig annotierte Sprachdaten zur Verfügung stellen und gleichzeitig ihre Stimme in Bezug auf verschiedene Sprechereigenschaften analysieren lassen. Wann immer Daten jedoch spezielle Anforderungen erfüllen sollen, müssen diese auf konventionelle Weise erhoben werden. In diesem Zusammenhang wurden vier Datensätze gesammelt, von denen zwei nun näher beleuchtet werden.

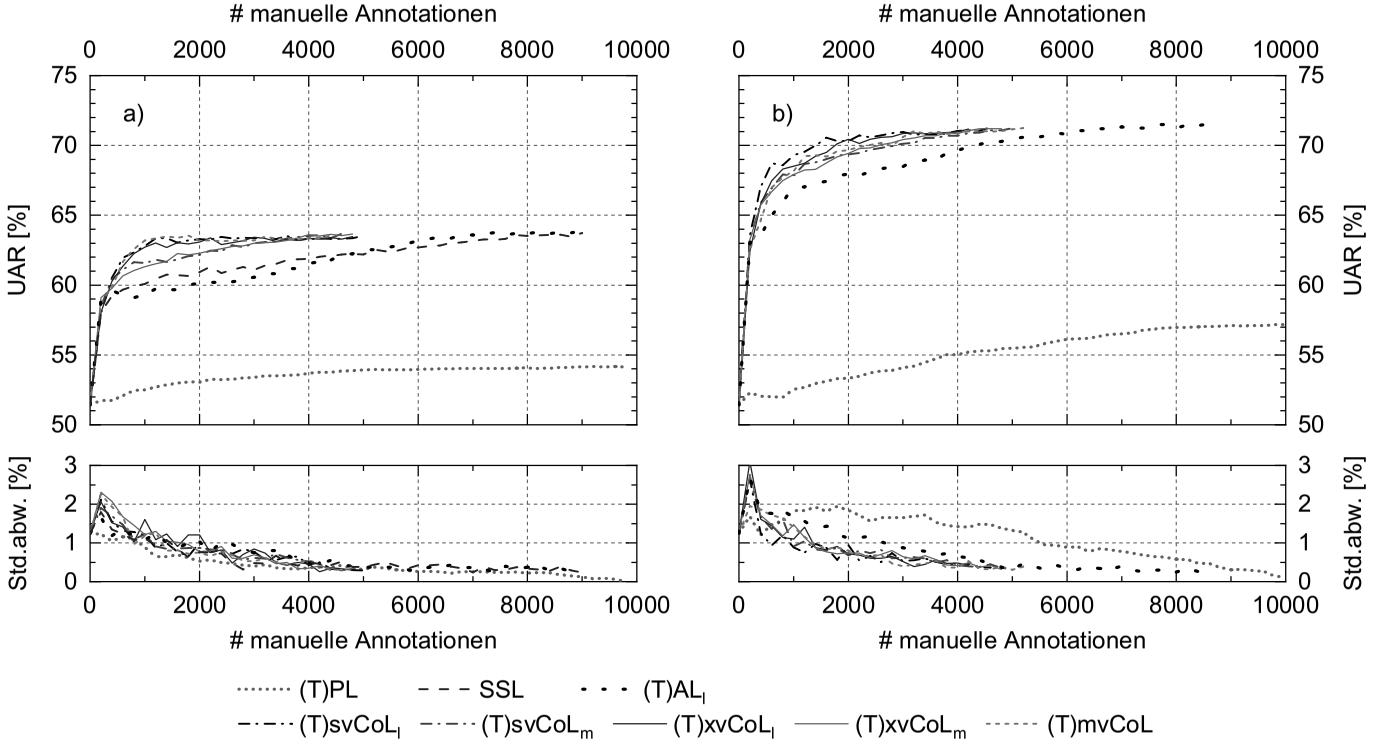

Abb. 3: *Vergleich von Passivem Lernen (PL), halbüberwachtem Lernen (SSL), aktivem Lernen (AL) und den verschiedenen (vertrauenswürdigkeitbasierenden) kooperativen Lernansätzen (T)CoL auf dem FAU Aibo Emotion Corpus mit binären Emotionsklassen [HSS18]. Die Ergebnisse zeigen die über 20 unabhängigen Durchläufe der Algorithmen gemittelten Werte des unweighted average recall (UAR) [%] im Vergleich zur Anzahl der manuell ausgewählten Annotationen. Es werden Ergebnisse mit konventionellen Annotationen (a) und mit auf vertrauenswürdigkeitbasierenden Annotationen (b) gezeigt.*

iHEARu-EAT Der iHEARu-EAT Datensatz ist der erste seiner Art. Es wurden 30 Sprecher beim Verzehr von sechs verschiedenen Lebensmitteln aufgenommen und dabei 1.414 Instanzen und 2,9 Stunden audio-visuelle Daten aufgezeichnet. Es wurde eine 7-stufige Klassifikationsstudie durchgeführt, welche die automatische Klassifizierung der Essbedingungen während des Sprechens demonstrierte. Schließlich ist der Datensatz – mit 14 teilnehmenden Teams – erfolgreich als Eating Condition (EC) Sub-Challenge in die Interspeech 2015 ComParE Challenge [Sc15b] und ist als einziger Datensatz in die ICMI 2018 EAT Challenge [Ha18c] eingeflossen.

EmotAsS Der EmotAsS Datensatz enthält spontan aufgezeichnete emotionale Sprache von 17 geistig, neurologisch und/oder körperlich eingeschränkten Personen [Ha17]. Der Datensatz enthält ca. 12.700 Instanzen mit knapp 11 Stunden Material. Die automatische Analyse der spontanen Emotionen dieser Personen wurde mit dem Ziel durchgeführt, ein sprachgesteuertes, arbeitsplatzintegriertes Assistenzsystem für Menschen mit Einschränkungen zu entwickeln [Ha18a]. Es wurde gezeigt, dass diese atypische Sprachanalyse eine anspruchsvolle Aufgabe ist, die jedoch mit den in der Arbeit vorgestellten Methoden bewältigt werden konnte. Der Datensatz ist schließlich in die renommierten Interspeech 2018 ComParE Challenge [Sc18] eingeflossen.

5 Perzeptionsstudien

Neben den vorgestellten Arbeiten zur Entwicklung und Evaluierung der intelligenten spieleorientierten Crowdsourcing Plattform und zur Sammlung verschiedener Sprachdatensätze wurde iHEARu-PLAY auch für die Durchführung von Perzeptionsstudien genutzt, welche die Vielseitigkeit der Plattform demonstrieren: (i) die Bewertung der automatischen Erkennung des Kontexts und der wahrgenommenen Emotion von Hundebellen [HCS18], (ii) die Untersuchung der menschlichen Wahrnehmung von Stimmzügen bei synthetischen Stimmen [Ba18], (iii) die Bewertung der Wahrnehmung von Emotionen in der Singstimme [Pa18], und (iv) die Untersuchung der menschlichen Wahrnehmung von Sprachemotion in Störgeräuschen [Pa17].

6 Zusammenfassung und Ausblick

Heutige maschinelle Lernalgorithmen sind von der Verfügbarkeit qualitativ hochwertig annotierter Trainingsdaten abhängig. Obwohl es aufgrund moderner Technologien eine Vielzahl verschiedener, frei verfügbarer Daten gibt, können diese häufig nicht in ihrer Rohform übernommen werden, da sie oft unstrukturiert sind und es an zuverlässigen Annotationen fehlt. In Anbetracht der Defizite modernster Technologien zur effizienten Datenannotation waren die wichtigsten Beiträge der Arbeit die bahnbrechenden Einführungen, die mit der vorgestellten intelligenten, spielorientierten Crowdsourcing Plattform iHEARu-PLAY verbunden sind. Die Plattform ermöglicht eine schnellere, kosteneffizientere und zuverlässigere Datenerfassung und Datenannotation, als bisher möglich. In diesem Zusammenhang leistet die vorliegende Forschungsarbeit vier wesentliche Beiträge:

1. Den Entwurf und die Realisierung einer intelligenten Crowdsourcing Plattform insbesondere zur Steigerung der Qualität und Effizienz der Annotation und Sammlung von audiovisuellen Daten für das maschinelle Lernen.

2. Die Einbeziehung und Erweiterung von über den Stand der Technik hinausgehenden Algorithmen des maschinellen Lernens und daraus resultierende neue Verfahren zur Bewertung der Verlässlichkeit geleisteter Annotationen.
3. Die Erstellung und Sammlung von neuen Datensätzen für das Lernen intelligenter audiovisueller Mustererkennungsverfahren, die von Social-Media-Websites übernommen und konventionell in Studien gesammelt werden.
4. Bewertung und Evaluierung der vorgestellten Verfahren, Umsetzungen und Sammlungen von Daten und Annotationen.

Zusammenfassend lässt sich sagen, dass die im Rahmen dieser Arbeit durchgeführte Forschung vielversprechende Ergebnisse und neuartige Methoden geliefert hat, indem ein automatisiertes, kombiniertes und robustes System zur Verfügung gestellt wurde, um neue Datensätze zu beschaffen, aufzuzeichnen und vor allem auf eine sehr effiziente Weise zu annotieren. Als Ergebnis können zuverlässige Trainingsdaten für die Audioverarbeitung schneller, ressourcenschonender und damit kostengünstiger gesammelt werden.

Zukünftige Arbeiten könnten auf den entwickelten vertrauenswürdigkeitsbasierenden Algorithmen des maschinellen Lernens basieren. Durch Modellierung eines Annotators könnte es den intelligenten Algorithmen ermöglicht werden nicht nur selbst zu entscheiden, wann manuelle Annotationsunterstützung erforderlich ist, sondern auch, handverlesene Annotatoren um ihre Meinung zu bitten, abhängig von den bestehenden Vertrauenswürdigkeitsbewertungen jedes Annotators. Schließlich soll iHEARu-PLAY für externe Forscher öffentlich zugänglich gemacht werden, damit diese ihre eigenen Aufgaben und Datensätze über die neue Schnittstelle hochladen, sammeln, erstellen, bearbeiten und löschen können. In dieser Hinsicht werden sich die Ergebnisse dieser Arbeit auf die nächste Generation intelligenter maschineller Lernansätze auswirken. iHEARu-PLAY wird letztlich den langfristigen Mangel eines geeigneten, effizienten und intelligenten Datenerfassungs- und Annotationswerkzeugs im Bereich des überwachten maschinellen Lernens beheben.

Literaturverzeichnis

[Ba18] Baird, Alice; Jørgensen, Stina; Parada-Cabaleiro, Emilia; Hantke, Simone; Cummins, Nicholas; Schuller, Björn: Listener Perception of Vocal Traits in Synthesized Voices: Age, Gender, and Human-Likeness. Journal of the Audio Engineering Society, 66:1–8, 2018.

[DBB52] Davis, KH; Biddulph, R; Balashek, Stephen: Automatic Recognition of Spoken Digits. The Journal of the Acoustical Society of America, 24:637–642, 1952.

[Ey13] Eyben, Florian; Weninger, Felix; Groß, Florian; Schuller, Björn: Recent Developments in OpenSMILE, the Munich Open-Source Multimedia Feature Extractor. In: Proc. of Int. Conference on Multimedia (ACMMM). Barcelona, Spain, S. 835–838, 2013.

[Go16] Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron; Bengio, Yoshua: Deep Learning. MIT press, Cambridge, UK, 2016.

[Ha09] Hall, Mark A.; Frank, Eibe; Holmes, Geoffrey; Pfahringer, Bernhard; Reutemann, Peter; Witten, Ian H.: The WEKA Data Mining Software: An Update. SIGKDD Explorations, 11:10–18, 2009.

[Ha15] Hantke, Simone; Eyben, Florian; Appel, Tobias; Schuller, Björn: iHEARu-PLAY: Introducing a Game for Crowdsourced Data Collection for Affective Computing. In: Proc. of Workshop on Automatic Sentiment Analysis in the Wild, Satellite of Int. Conference on Affective Computing and Intelligent Interaction. Xi'an, P. R. China, S. 891–897, 2015.

[Ha17] Hantke, Simone; Sagha, Hesam; Cummins, Nicholas; Schuller, Björn: Emotional Speech of Mentally and Physically Disabled Individuals: Introducing the EmotAsS Database and First Findings. In: Proc. of INTERSPEECH. Stockholm, Sweden, S. 3137–3141, 2017.

[Ha18a] Hantke, Simone; Cohrs, Christian; Schmitt, Maximilian; Tannert, Benjamin; Lütkebohmert, Florian; Detmers, Mathias; Schelhowe, Heidi; Schuller, Björn: Introducing an Emotion-driven Assistance System for Cognitively Impaired Individuals. In: Proc. of Int. Conference on Computers Helping People with Special Needs (ICCHP). Linz, Austria, S. 486–494, 2018.

[Ha18b] Hantke, Simone; Olenyi, Tobias; Hausner, Christoph; Schuller, Björn: VoiLA: An Online Intelligent Speech Analysis and Collection Platform. In: Proc. of Asian Conference on Affective Computing and Intelligent Interaction. Beijing, China, S. 1–5, 2018.

[Ha18c] Hantke, Simone; Schmitt, Maximilian; Tzirakis, Panagiotis; Schuller, Björn: EAT — The ICMI 2018 Eating Analysis and Tracking Challenge. In: Proc. of Int. Conference on Multimodal Interaction (ICMI). Boulder, USA, S. 1–5, 2018.

[Ha19] Hantke, Simone: Intelligent Gamified Crowdsourcing for Audio Processing. Dissertation, Dissertation, Technische Universität München, 2019.

[HAS18a] Hantke, Simone; Abstreiter, Alexander; Schuller, Björn: Trustability-based Dynamic Active Learning for Crowdsourced Labelling of Emotional Audio Data. IEEE Access, S. 13, 2018. submitted, under review.

[HAS18b] Hantke, Simone; Appel, Tobias; Schuller, Björn: The Inclusion of Gamification Solutions to Enhance User Enjoyment on Crowdsourcing Platforms. In: Proc. of Asian Conference on Affective Computing and Intelligent Interaction. Beijing, China, S. 1–6, 2018.

[HCS18] Hantke, Simone; Cummins, Nicholas; Schuller, Björn: What is my Dog Trying to Tell me? The Automatic Recognition of the Context and Perceived Emotion of Dog Barks. In: Proc. of Int. Conference on Acoustics, Speech and Signal Processing (ICASSP). Calgary, Canada, S. 1–5, 2018.

[Ho06] Howe, Jeff: The Rise of Crowdsourcing. Wired Magazine, 14:1–4, 2006.

[HSS18] Hantke, Simone; Stemp, Christoph; Schuller, Björn: Annotator Trustability-based Cooperative Learning Solutions for Intelligent Audio Analysis. In: Proc. of INTERSPEECH. Hyderabad, India, S. 3504–3508, 2018.

[HZS17] Hantke, Simone; Zhang, Zixing; Schuller, Björn: Towards Intelligent Crowdsourcing for Audio Data Annotation: Integrating Active Learning in the Real World. In: Proc. of INTERSPEECH. Stockholm, Sweden, S. 3951–3955, 2017.

[Pa17] Parada-Cabaleiro, Emilia; Baird, Alice; Batliner, Anton; Cummins, Nicholas; Hantke, Simone; Schuller, Björn: The Perception of Emotions in Noisified Non-Sense Speech. In: Proc. of INTERSPEECH. Stockholm, Sweden, S. 3246–3250, 2017.

[Pa18] Parada-Cabaleiro, Emilia; Schmitt, Maximilian; Batliner, Anton; Hantke, Simone; Costantini, Giovanni; Scherer, Klaus; Schuller, Björn: Identifying Emotions in Opera Singing: Implications of Adverse Acoustic Conditions. In: Proc. of Int. Conference on Society for Music Information Retrieval (ISMIR). Paris, France, 2018.

[Pr00] Provost, Foster: Machine Learning from Imbalanced Data Sets. In: Working Notes of the AAAI Workshop on Learning from Imbalanced Data Sets. Austin, USA, S. 1–3, 2000.

[Ra10] Raykar, Vikas; Yu, Shipeng; Zhao, Linda H; Valadez, Gerardo Hermosillo; Florin, Charles; Bogoni, Luca; Moy, Linda: Learning from Crowds. Journal of Machine Learning Research, 11:1297–1322, 2010.

[Sc15a] Schmidhuber, Jürgen: Deep Learning in Neural Networks: An Overview. Neural Networks, 61:85–117, 2015.

[Sc15b] Schuller, Björn; Steidl, Stefan; Batliner, Anton; Hantke, Simone; Hönig, Florian; Orozco-Arroyave, Juan Rafael; Nöth, Elmar; Zhang, Yue; Weninger, Felix: The INTERSPEECH 2015 Computational Paralinguistics Challenge: Degree of Nativeness, Parkinson's and Eating Condition. In: Proc. of INTERSPEECH. Dresden, Germany, S. 478–482, 2015.

[Sc18] Schuller, Björn; Steidl, Stefan; Batliner, Anton; Marschik, Peter B.; Baumeister, Harald; Dong, Fengquan; Hantke, Simone; Pokorny, Florian; Rathner, Eva-Maria; Bartl-Pokorny, Katrin D.; Einspieler, Christa; Zhang, Dajie; Baird, Alice; Amiriparian, Shahin; Qian, Kun; Ren, Zhao; Schmitt, Maximilian; Tzirakis, Panagiotis; Zafeiriou, Stefanos: The INTERSPEECH 2018 Computational Paralinguistics Challenge: Atypical & Self-Assessed Affect, Crying & Heart Beats. In: Proc. of INTERSPEECH. Hyderabad, India, S. 122–126, 2018.

[SSB09] Schuller, Björn; Steidl, Stefan; Batliner, Anton: The Interspeech 2009 Emotion Challenge. In: Proc. of INTERSPEECH. Brighton, UK, S. 312–315, 2009.

[St09] Steidl, Stefan: Automatic Classification of Emotion Related User States in Spontaneous Children's Speech. Dissertation, Dissertation, University of Erlangen-Nuremberg, 2009.

[TK11] Tong, Simon; Koller, Daphne: Support Vector Machine Active Learning with Applications to Text Classification. Journal of Machine Learning Research, 2:45–66, 2011.

[Zh15] Zhang, Zixing: Semi-Autonomous Data Enrichment and Optimisation for Intelligent Speech Analysis. Dissertation, Dissertation, Technische Universität München, 2015.

Simone Hantke erhielt 2011 ihr Diplom in Medientechnik von der Technischen Hochschule Deggendorf, 2014 ihren Master of Science und 2019 ihren Doktortitel jeweils von der Technischen Universität München, einer der deutschen Exzellenz-Universitäten. Während ihrer Promotionszeit war sie zudem als wissenschaftliche Mitarbeiterin am Lehrstuhl für Complex and Intelligent Systems an der Universität Passau und am ZD.B Lehrstuhl für Embedded Intelligence for Health Care and Wellbeing an der Universität Augsburg tätig. Sie arbeitete auf dem Gebiet der Datenverarbeitung und der Spracherkennung, wobei sich ihr Forschungsschwerpunkt auf der Datenerfassung und neuen maschinellen Lernansätzen für robuste automatische Spracherkennung und Sprechercharakterisierung konzentrierte. Simone Hantke hat im Rahmen ihrer Promotion die vorgestellten Verfahren und Ergebnisse erfolgreich in 29 begutachteten wissenschaftlichen Beiträgen veröffentlichen können.