

IVIS White Paper
Oktober 2018

IVIS Intelligent Voice Interpreter for Sign-Language

Ein Ansatz zur Inklusionsförderung hörgeschädigter Menschen durch ein autonomes KI-System für eine Gebärdenübersetzungssoftware.

VERSTEHEN SIE 你明白好玩嗎？

GEBÄRDENSPRACHE

KOMPLIKATIONEN

AKTUELLE FORSCHUNGEN

VORAUSSETZUNGEN

BEWEGUNGSERKENNUNG

ÜBERSETZUNGSDATENBANK

EMOTIONSERKENNUNG

LAUTSPRACHE

GEBÄRDENÜBERSETZUNG

SEQUENTIELLE ÜBERSETZUNG

GEBÄRDENAUSGABE

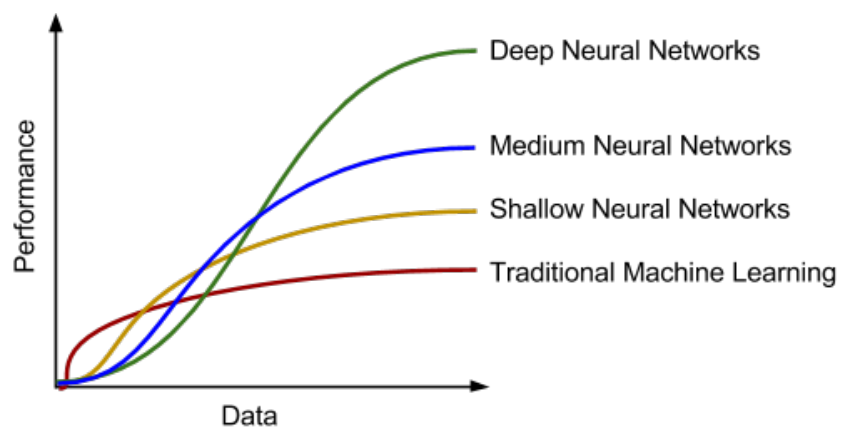
AUTONOME LÖSUNG

FAZIT

VERSTEHEN SIE 你明白好玩嗎? Statistisch gesehen kann nur jeder 1000. deutsche Bürger diesen Satz ohne Hilfsmittel verstehen. Für die anderen 999 Bürger bleibt letztlich nur die Möglichkeit zu versuchen diese chinesischen Schriftzeichen zu übersetzen. Bei einem Satz wie „你明白好玩嗎 = Verstehen Sie Spaß“ ist das mit einem Wörterbuch noch einfach. Bei mehreren Sätzen, ganzen Artikeln oder gesamten Büchern gestaltet sich eine Übersetzung komplizierter. Dafür sind Dolmetscher notwendig, die die komplexen Regeln einer Sprache kennen und eine qualitativ hochwertige Übersetzung anbieten können.

Wer Dienste wie GOOGLE TRANSLATE zur Übersetzung benutzt, wird in den letzten Jahren häufiger bemerkt haben, dass sich die Qualität der Übersetzungen deutlich verbessert hat. Dies ist darauf zurückzuführen, dass sich die Architektur der Übersetzungssoftware grundlegend verändert hat. Statt dem Computer zu versuchen alle Regeln inklusive ihrer Ausnahmen durch Programmcode beizubringen, wird neuerdings auf die künstliche Intelligenz zurückgegriffen. Mit Hilfe sogenannter neuronaler Netze erlernen Computer menschliche Sprachen quasi von selbst. Zwei Personen, die die Sprache ihres Gegenübers nicht beherrschen, können heutzutage dank künstlich intelligenter Software mit einem einfachen Smartphone miteinander kommunizieren.

Hierbei wird künstliche Intelligenz zur Spracherkennung, Übersetzung und eine kaum vom Menschen unterscheidbare Sprachausgabe genutzt. Die Kommunikationsbarrieren zwischen den unterschiedlichen Sprachen scheinen somit gebrochen zu sein.



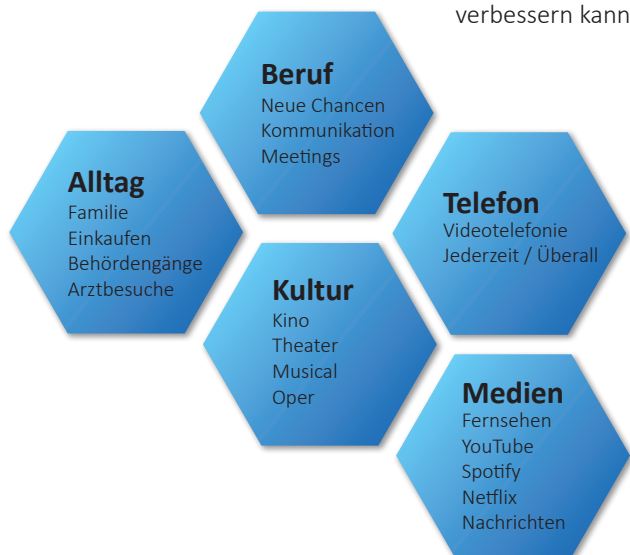
GEBÄRDENSPRACHE

Gehörlose nutzen visuelle Gebärdensprachen, die zur Kommunikation verschiedenster Sachverhalte Handzeichen, Körperhaltung und Mimik statt Laute und Wörter verwenden. Gebärdensprachen haben eigene Grammatiken und ein umfangreiches Vokabular. Viele gelten daher auch als vollwertig anerkannte Sprachen.

Auf Grund der starken Disparität der visuellen Gebärdensprache gegenüber anderen textuellen Sprachen sind Gehörlose in ihrer Kommunikation sehr eingeschränkt, da sie noch mehr davon abhängig sind, ob ihr Gegenüber dieselbe Gebärdensprache beherrscht oder nicht. Selbst innovative medizinische Implantate können Gehörlosen bei der Kommunikation nur geringfügig weiterhelfen. Für Gehörlose ist es besonders schwer nicht visuelle Sprachen zu erlernen, da sie kein auditives Feedback haben.

Aktuell leben 140.000 Menschen in Deutschland, die auf Gebärdensprachen angewiesen sind. Weltweit sind 70 Millionen Menschen entweder völlig gehörlos oder verfügen nur noch über ein geringes Resthörvermögen. Sie sind für eine Kommunikation mit dem Großteil der Bevölkerung auf Dolmetscher angewiesen.

Die jüngsten Entwicklungen in künstlicher Intelligenz und Computer Vision geben Hoffnung für eine Übersetzungssoftware für Gehörlose, die ihr Leben deutlich verbessern kann und eine Integration in die Gesellschaft erleichtert.



VISUELLE SPRACHE

KOMPLIKATIONEN

Im Gegensatz zu textuellen Sprachen birgt eine Übersetzungssoftware für Gehörlose spezielle Herausforderungen, die zuerst überwunden werden müssen:

Die bereits eingesetzte Übersetzungstechnologie von GOOGLE und Co. befasst sich ausschließlich mit textuellen Sprachen. Gebärdensprachen mögen in ihrer Grammatik zwar einfacher aufgebaut sein als viele andere Sprachen; Es sind jedoch visuelle Sprachen. Daher gibt es kein schriftliches Alphabet, das ein Computer ohne weitere Probleme kodieren und interpretieren könnte. Stattdessen werden die einzelnen Gebärden durch Körperbewegungen, Mimik und Handzeichen in drei Dimensionen dargestellt. Eine Gebärdensübersetzungssoftware muss diese Elemente einer Gebärde zuerst erkennen und unterscheiden können.

EMOTIONEN

Eine der größten Schwierigkeiten liegt darin, dass für die gehörlose Gebärdensprache die emotionale Körpersprache ein essentielles Element darstellt. Jede Sprache verfügt über die Möglichkeit neben der expliziten Aussage eines Textes eine unterstützende implizite Bedeutung durch Emotionen hinzuzufügen. Analog zur emotionalen Aussprache eines geschriebenen Textes, wird die Mimik dazu verwendet Gebärdensprache eine implizite Bedeutung zu geben.

DATENBANK



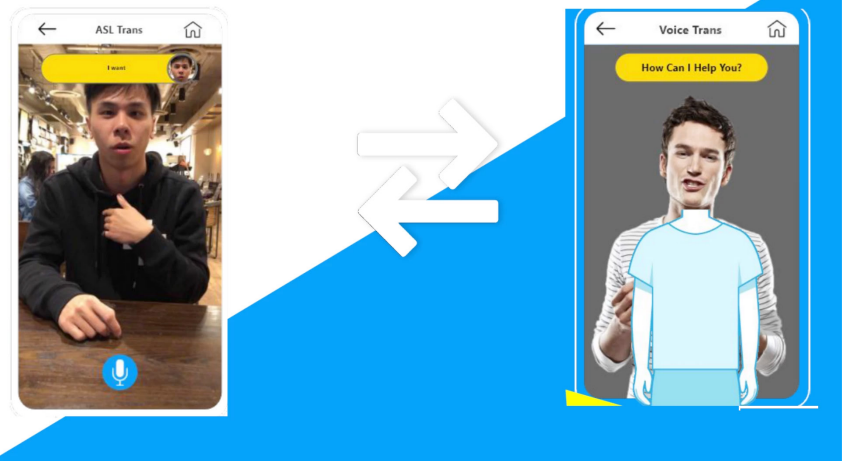
Eine weitere Hürde besteht darin, dass neuronale Netze eine große Menge an interpretierbaren Daten benötigen, um in einer Wechselbeziehung aus Approximation und Fehlerkorrektur die vielseitigen Merkmale einer Übersetzung zu erlernen. Bei textuellen Sprachen gibt es bereits zahlreiche Übersetzungsdatenbanken, die zum Erlernen der sprachspezifischen Merkmale genutzt werden. Je mehr Daten einer künstlichen Intelligenz zum Erlernen zur Verfügung stehen, umso genauer wird ihre Voraussage.

Für die Gebärdensprache gibt es lediglich mangelhafte und unvollkommene Wörterbücher. Sie bestehen meist nur aus einzelnen Gebärden, aber keine zusammengesetzte Sequenz von Gebärden, die zu Sätzen gebildet werden können. Diese unzureichende Datenquelle würde nur eine Wort-für-Wort Übersetzung ermöglichen. Um jedoch auch die notwendigen grammatikalischen Übersetzungsmerkmale zu erlernen ist eine Datenbank mit Übersetzungen gesamter Sätze in Gebärdensprache notwendig.

AKTUELLE FORSCHUNGEN

In den letzten Jahren gab es einige Ansätze für eine Gebärden Übersetzungssoftware. Einige Prototypen nutzen dabei Handschuhe zum Tracken der Handgesten. Die meisten Prototypen bieten jedoch keine beidseitige Übersetzung an. Entweder werden Computer Vision Algorithmen verwendet, um Gebärden zu interpretieren und in Sätze für textuelle Sprachen zu übersetzen, oder eine App erkennt die Sprache auditiv und wandelt diese in symbolhafte Gebärdendarstellungen um. Dabei ist eine Darstellung von unterschiedlichen Icons bis hin zu 3D Avataren möglich.

New Yorker Studenten entwickelten ein Prototyp, der eine simultane Übersetzung in beide Richtungen anbietet. Hier werden jedoch nur Icons zum Darstellen der Gebärden verwendet und die gesamte Kopfreion der Gebärde wird ausgeblendet. Für jede Gebärde muss demnach ein individuelles Icon von Hand erstellt werden. Das Auslassen der Kopfreion kann zu Missinterpretationen führen, da die Mimik, einer Gebärde eine vollkommen andere Bedeutung geben kann. Das Hindernis einer unvollständigen Übersetzungsdatenbank ist bei diesen Prototypen nicht überwunden.



Die Forschungen zur Gebärden Übersetzung sind noch nicht ausgereift, bieten jedoch einige gute Ansätze, die gemeinsam zu einer autonomen Übersetzungssoftware gebündelt werden können. Im Folgenden soll untersucht werden, welche Schritte unternommen werden müssen, um die genannten Herausforderungen zu überwinden.

VORAUSSETZUNGEN

Eine autonome Gebärden-Übersetzungssoftware benötigt vor allem eine hohe Reichweite, um zu gewährleisten, dass genügend Daten gesammelt werden, damit ein neuronales Netz die Merkmale der Gebärdensprache identifizieren und erlernen kann. Dafür muss die Software so programmiert werden, dass jeder Mensch sie, ohne zusätzliche Hardware, benutzen kann. Sie muss mit einer simplen RGB Kamera eines Smartphones oder einer Webcam funktionieren.

Motion Capture Anzüge, Tracking-Handschuhe oder eine Multikamera Lösung geben im Einzelfall womöglich ein genaueres Ergebnis, würden aber auf Grund des umfangreicheren Hardwarebedarfs und der daraus resultierenden geringeren Reichweite und somit kleineren Datenquelle zu mangelhafteren Ergebnissen führen.

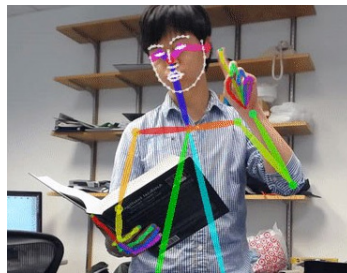


OPEN POSE

BEWEGUNGSKENNUNG

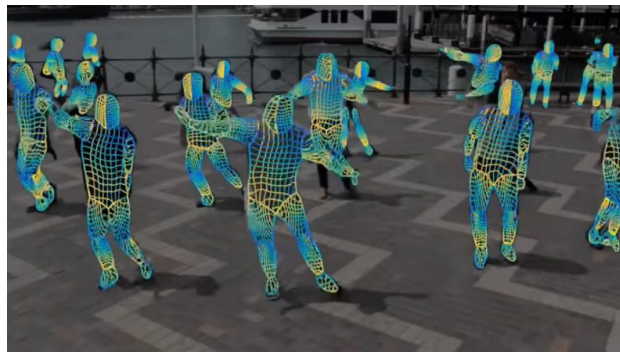
Forscher haben in den letzten Jahren große Fortschritte im Bereich der Computer Vision gemacht. Zwei innovative Algorithmen, die zur Erkennung von Körperbewegungen verwendet werden können sind OPEN POSE und DENSE POSE.

Der OPEN POSE Algorithmus kann mit Hilfe einer einzelnen RGB Kamera sowohl die Körperhaltung (18 – 25 Tracking Punkte), als auch die Handbewegungen (2 mal 21 Tracking Punkte) und Gesichtsmimik (bis zu 70 Tracking Punkten) tracken.



DENSE POSE

Der DENSE POSE Algorithmus nutzt künstliche Intelligenz, um Menschen aus einem RGB pixelbasierten Bild zu extrahieren und jeden dieser Pixel auf eine 3D Oberfläche eines Computer-Generierten menschlichen Körpers abzubilden.



Diese Algorithmen sind zwar auf die zwei dimensionale Fläche des Videosignals limitiert, es gibt jedoch bereits erste erfolgreiche Ansätze, um aus einem einzelnen zwei dimensionalem RGB Video Signal Tiefeninformationen zu extrahieren und auf einen dreidimensionalen Computer Generierten Körper zu riggen.

ÜBERSETZUNGSDATENBANK

Die getrackten Daten können dazu verwendet werden, um eine Datenbank für die Gebärdenübersetzung aufzubauen. Zeichnet man eine große Menge textueller Sätze in Gebärdensprache als Video auf und markiert diese mit ihrer textuellen Übersetzung, so können Algorithmen mit Hilfe künstlicher Intelligenz einen logischen Zusammenhang zwischen den textuellen Sätzen und den getrackten Gebärdendaten herstellen.

Die gewonnene Datenbank kann im Anschluss sowohl für die Übersetzung aus der Gebärdensprache in die textuelle Sprache verwendet werden als auch um eine textuelle Sprache in Gebärdensprache zu übersetzen und darzustellen.



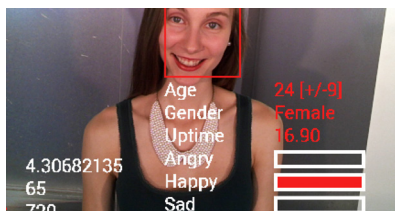
TEXT

Social Media Netzwerke wie Facebook und Twitter verwenden emotionale Texterkennung um Hass Kommentare zu filtern. Algorithmen wie GOOGLE NATURAL LANGUAGE kommen dabei zum Einsatz.

Einige Forscher schlagen fünf Basisemotionen vor (ANGST, WUT, FREUDE, TRAUER UND SCHAM). Weitere Forschungen unterscheiden zusätzlich zwischen INTERESSE-NEUGIER, ÜBERRASCHUNG, EKEL und SCHULD.

AUDIO

VIDEO



INDIVIDUELL

EMOTIONSERKENNUNG

Da Gebärden nicht nur den reinen Inhalt einer Nachricht transportieren, sondern auch eine emotionale Komponente durch die Mimik des Gebärdenden vermitteln, muss eine Gebärdenübersetzungssoftware Emotionen zwingend erkennen können. Hierbei stehen drei interpretierbare Quellen zur Verfügung. Die Emotionserkennung kann entweder visuell, auditiv oder textuell erfolgen.

Unter textuellen Emotionserkennung versteht man, aus einem geschriebenen Satz emotionale Stimmungen zu erkennen. Ähnlich wie der Mensch, kann ein Algorithmus mit Hilfe einer künstlichen Intelligenz, anhand bestimmter negativen oder positiven Wörtern, wie zum Beispiel «HASS» oder «LIEBE», die Sentimentalität eines Satzes erkennen und in einer kontinuierlichen Skala zwischen positive oder negative Satzstimmungen einordnen.

Emotionen sind jedoch vielfältiger und lassen sich nicht nur in rein negative oder positive Kategorien einordnen. Viele Psychologen haben versucht die Vielfalt der menschlichen Emotionen auf eine geringe Anzahl an Basisemotionen zu reduzieren. Diese Eingrenzung der Emotionen ist ein essentieller Aspekt für die Gebärdenübersetzung, da verschiedenste Emotionen genutzt werden, um unterschiedliche Gebärden darzustellen. Für eine Erkennung muss es klare Definitionen geben, welches Merkmal zu welcher Emotion gehört.

Für eine auditive Emotionserkennung kann eine künstliche Intelligenz mit der Eingabe emotions-markierter auditiven Sprachaufnahmen eine Prognose über die Emotion stellen. Dazu werden digitalisierte Tonsignale analysiert und emotionspezifische Muster erkannt. Die Softwareschnittstelle OPEN VOKATURI bietet ihren Nutzern die Möglichkeit Audiodateien auf die Sprachstimmung zu analysieren. Die Ergebnisqualität ist jedoch nicht optimal.

Dieses Erkennungsprinzip kann auch auf die visuelle Emotionserkennung angewandt werden. Statt digitalisierter Tonsignale dienen hierbei Tracking Punkte des Gesichts als Datenquelle. Auch bei der visuellen Emotionserkennung gibt es bereits einige Beispiele. Die Softwarelösung SHORE des FRAUNHOFER-INSTITUT FÜR INTEGRIERTE SCHALTUNGEN ILS kann Emotionen anhand der Mimik erkennen. Sie kann lächelnde Gesichtszüge von traurigen, wütenden oder überraschten unterscheiden. Dazu werden sowohl die Augen als auch die Mundwinkel analysiert. Für eine makellose Erkennung müssen die Gesichtszüge überspitzt dargestellt werden. Die Gebärdensprache bietet hierbei einen Vorteil gegenüber anderen Sprachen. Emotionen werden zum leichteren Verständnis der Gebärden ohnehin überspitzt dargestellt und können somit leichter von Algorithmen erkannt werden.

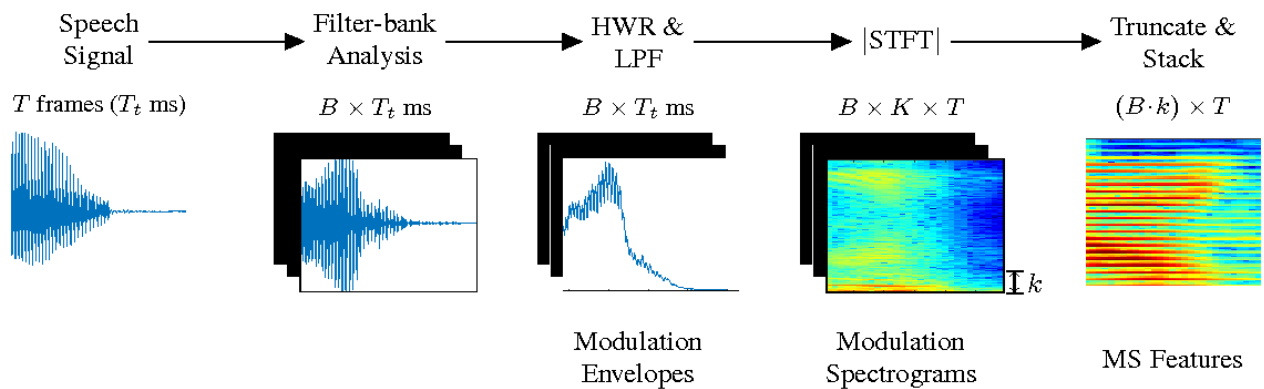
Die größte Herausforderung bei Emotionserkennungsalgorithmen liegt jedoch sicherlich darin, dass jeder Mensch Emotionen unterschiedlich ausdrückt. Hierbei können Geschlecht, Alter oder kultureller Hintergrund einen großen Faktor für eine Missinterpretation darstellen. Sarkasmus und Ironie sind zusätzlich Bereiche, die noch nicht ausreichend erkannt und einwandfrei interpretiert werden können. Da jedoch die Emotionale Komponente bei Gebärden eine essenzielle Rolle spielt, kann vorerst nur eine kombinierende Untersuchung aller drei Emotionserkennungssysteme eine Übergangsregelung darstellen, bis eine vollkommenere Lösung der Emotionserkennung entwickelt wurde.

LAUTSPRACHE

Um Lautsprache in textuelle Schrift umzuwandeln, können gängige Spracherkennungsalgorithmen verwendet werden. Diese besitzen meist bereits eine große Datenmenge zum Trainieren einer künstlichen Intelligenz. GOOGLE, IBM WATSON, AMAZON oder viele weitere große Industriefirmen bieten Spracherkennungsdienste an.

DIGITALISIERUNG

Die Spracherkennung wandelt nach der Digitalisierung eines analogen Audiosignals (digitale Tonaufnahme), die aufgenommene Sprache in Text um. Dabei wird jede Abtastung des Tonsignals durch Anwendung der Fourier Transformation in ihre einzelnen Sinus und Cosinus Wellen zerlegt und in einem Spektrogramm gewichtet dargestellt. Dadurch wird bestimmt wie stark der Anteil an tiefen oder hohen Frequenzen eines Tonsignals ist.



PHONEM

Phoneme sind die kleinste bedeutungsunterscheidende Einheit des gesprochenen Wortes. Grapheme sind die kleinsten bedeutungsunterscheidenden, aber nicht selbst bedeutungstragenden grafischen Einheiten des Schriftsystems einer bestimmten Sprache

Durch künstliche Intelligenz kann man diese gewichteten Frequenzsignale unterscheiden und einem Phonem zuordnen. Nach dieser Analyse erhält man eine sequenzielle Abfolge von Phonemen, die dann zu Graphemen und Wörtern umgewandelt werden. Die einzelnen Wörter werden sequentiell angeordnet. Im Anschluß wird analysiert, ob der Satz einen grammatikalischen Sinn ergibt. Dazu werden die einzelnen Wörter eines Satzes auf ihre syntaktischen Eigenschaften untersucht. Erkennt die künstliche Intelligenz nun einen Fehler, kehrt der Algorithmus zurück zu der Wortfindung aus den Phonemen und betrachtet die zweitwahrscheinlichste Prognose des Wortes aus den Phonemen.

HERAUSFORDERUNG

Eine große Herausforderung liegt darin, dass Menschen üblicherweise nicht perfekt sprechen. Sie benutzen Umgangssprache, Abkürzungen und Dialekte, die den Prozess der Spracherkennung erschweren. Trotz allem haben sich in den letzten Jahren durch eine große Ansammlung an trainierbaren Daten fortschrittliche KI-Algorithmen entwickelt, die wiederum zu besseren Ergebnissen geführt haben.

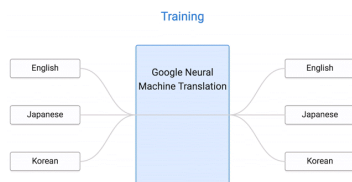
INPUT & OUTPUT

GEBÄRDENÜBERSETZUNG

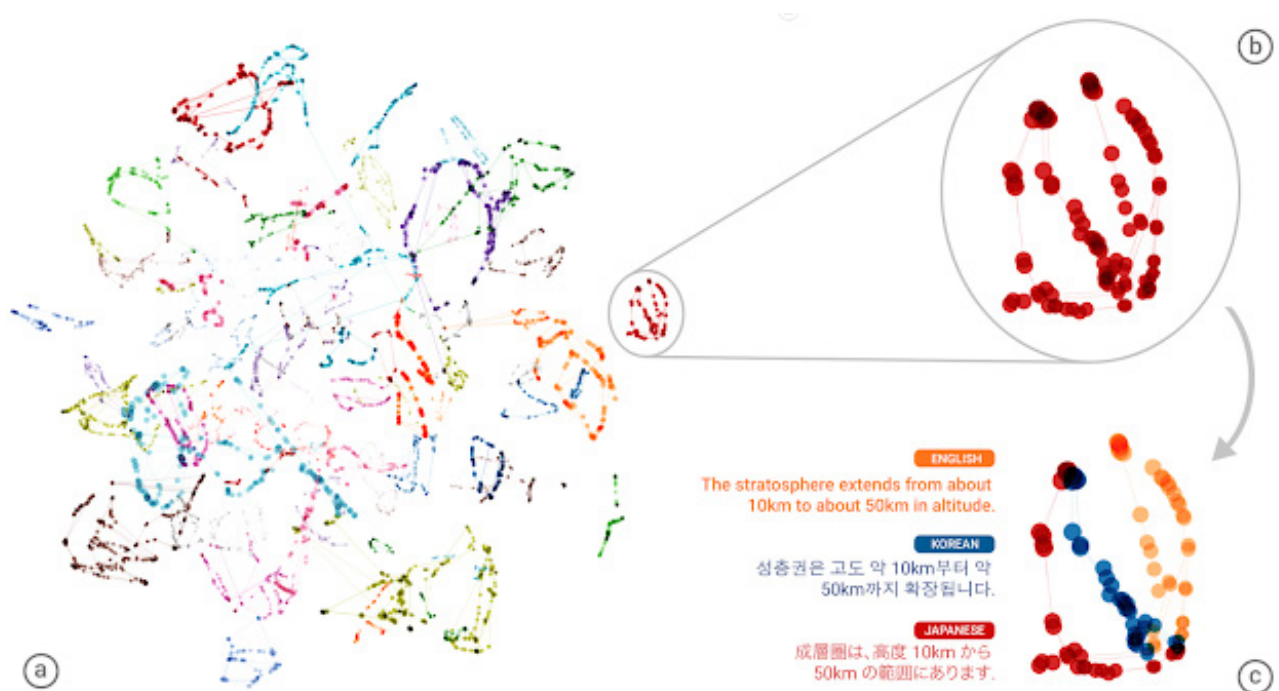
Nun wurden alle benötigten Bereiche zur erfolgreichen Übersetzung zwischen einer visuellen Gebärdensprache und einer textuellen Sprache abgedeckt.

Auf der Seite der textuellen Sprache verfügt der Übersetzungsalgorithmus neben dem zu übersetzenden textuellen Satz, weitere Informationen wie der grammatikalischen Syntax und der erkannten emotionalen Stimmung, entweder durch den Text selbst, durch eine Analyse der Lautsprache oder der Körperhaltung und der Gesichtszüge. Auf der Seite der visuellen Gebärdensprache verfügt der Übersetzungsalgorithmus über die Informationen der getrackten Punkte im dreidimensionalen Raum, die jeweils einer Gebärde oder einem Gebärdensatz entsprechen. Des Weiteren lassen sich emotionale Stimmungen der Gebärden in der Mimik erkennen.

UNIVERSALSPRACHE



Moderne KI-Übersetzungsalgorithmen übersetzen nicht von einer Sprache in eine andere, sondern kodieren die Eingabesprache in Vektoren einer vermuteten Universalssprache. Diese Vektoren werden im drei-dimensionalen Raum dargestellt. Wörter mit ähnlicher Bedeutung verschiedener Sprachen werden im Raum an ähnlichen Koordinaten gefunden. Somit wird eine Verknüpfung (Word Embeddings) dieser Wörter hergestellt. Diese Vektoren dienen wiederum als Quelle für die Dekodierung in eine Ausgabesprache.

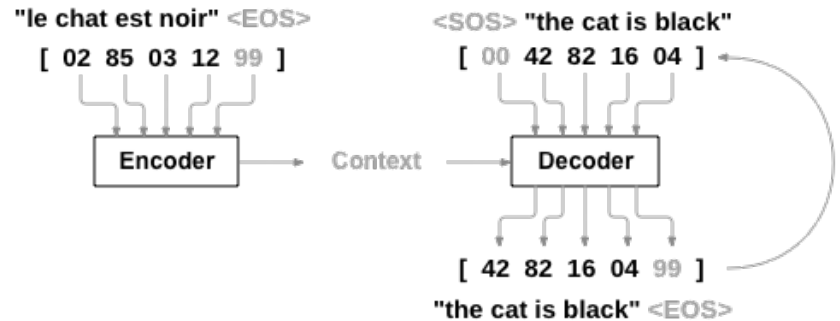


VERSTEHEN

Im Gegensatz zu bisherigen Ansätzen zur maschinellen Übersetzung, gelingt es diesem Algorithmus die Bedeutung eines Satzes einer Eingabesprache zu verstehen und in jede gewünschte Ausgabesprache zu übersetzen, die der Algorithmus beherrscht, selbst wenn zwischen der Eingabe und Ausgabesprache kein direkt verknüpfter Datensatz existiert.

SEQUENTIELLE ÜBERSETZUNG

Übersetzungsalgorithmen untersuchen nicht nur jedes einzelne Wort individuell, sondern auch die umgebenden Wörter und die Position des Wortes im Satz. Mit Hilfe von Recurrent Neural Networks (RNN), im speziellen LSTM Netzwerken (Long-Short-Term-Memory), kann eine Analyse von sequentiellen Daten erfolgen.



Eine große Herausforderung in der Übersetzung von Gebärden liegt in ihrer kontinuierlichen Darstellung. Die Trennung textueller Sätze werden durch Satzzeichen dargestellt. Einzelne Wörter werden durch das Leerzeichen getrennt. Zwischen den einzelnen Gebärden gibt es jedoch keine Pause sondern einen reibungslosen Übergang, der je nach Gebärdenkombination variiert.

Es wird ein weiteres Recurrent Neural Network benötigt, um die einzelnen Gebärden einer sequentielle Abfolge der getrackten Daten zu erkennen. Die Schwierigkeit hierbei besteht darin, dass es im Gegensatz zu auditiven oder textuellen Sätzen keine Leerzeichen oder Sprachpausen zwischen den Wörtern gibt. Dazu muss die künstliche Intelligenz auf ein Wörterbuch an getrackten einzelnen Gebärden zurückgreifen und untersuchen, wann eine Gebärde abgeschlossen ist und wann die nächste Gebärde beginnt.

GEBÄRDENAUSGABE Als letzten Schritt kann nun die Ausgabe in der gewünschten Sprache erfolgen. Mit Hilfe synthetisch erzeugter WAVENET Stimmen ist es möglich täuschend echte Sprachausgaben aus Texten zu erzeugen.

Zur Darstellung von Gebärden werden die ermittelten Tracking Punkte verwendet und auf einen drei dimensional Körper geriggt. Durch Interpolation können die Übergänge zwischen den Gebärden dargestellt werden. Selbst Emotionen können auf das Gesicht eines dreidimensionalen Avatars übertragen werden und somit eine gebärdentreue Simulation realisiert werden.



AUTONOME LÖSUNG

Der größte Vorteil einer künstlichen Intelligenz liegt darin, dass sie eigenständig lernen kann, indem sie sich Iteration für Iteration verbessert und einem Ideal annähert. Der bisherige Ansatz einer Gebärdenübersetzungssoftware verknüpft die nötigen Elemente einer Gebärdenübersetzung, ist aber von einer menschengemachten großen Gebärdendatenbank abhängig.

Für eine autonome Variante müsste vorerst eine Grundlage einer solchen Datenbank geschaffen werden, die ähnlich wie ältere GOOGLE TRANSLATE Versionen zunächst keine hervorragenden Ergebnisse liefert. Durch eine Korrekturmöglichkeit des Nutzers kann die Anwendung von ihren Fehlern lernen und sich eigenständig verbessern.

Zur Erstellung einer solchen selbstlernenden Software benötigt es anfangs einen vorläufigen Datensatz. Dieser könnte aus unterschiedlichen Quellen zusammengestellt werden.



Sobald eine Grundlage geschaffen ist, kann die Anwendung durch Fehlerbehebungen des Nutzers lernen. Ist letztlich eine bereits ausgereifte Datenbank entstanden, kann der Algorithmus neben der Fehlerkorrektur durch den Nutzer auch selbst aus seinen eigenen Fehlern lernen.

Das künstliche Intelligenz Komplex sucht sich immer wieder neue, unbekannte Texte zum Übersetzen aus dem Internet. Durch einen Übersetzungsversuch in eine Gebärde und der darauffolgenden Rückübersetzung in die Originalsprache, kann der Algorithmus aus seinen eigenen Fehlern lernen. Es wird der Originaltext mit dem doppelt übersetzten Text verglichen und dabei eine Fehlerkorrektur durchgeführt. Dadurch ist die künstliche Intelligenz auf keine weiteren Eingaben des Menschen angewiesen. Sie lernt selbstständig.

FAZIT Technologische Fortschritte in den Bereichen der künstlichen Intelligenz haben es in den letzten Jahren ermöglicht, dass die Kommunikation zwischen den unterschiedlichsten Sprachen auf der Erde zusammengewachsen ist. Unglücklicherweise werden die visuellen Gebärdensprachen von dieser Technologie bisher ausgeschlossen. Weltweit über 70 Millionen Gehörlose, die für eine Kommunikation auf Dolmetscher angewiesen sind, würden am effektivsten von einer Übersetzungssoftware profitieren. Aktuelle Prototypen können jedoch auf Grund ihrer zu geringen Reichweite und unzureichender Verknüpfung aller nötigen Elemente, die Herausforderungen einer Gebärdensübersetzung nicht ausreichend überwinden. Es wurde ein Ansatz zur gesamtheitlichen autonomen Gebärdensübersetzung vorgestellt, die eine Inklusion hörgeschädigter Menschen in der Gesellschaft vorantreiben würde. Dabei werden drei große Herausforderungen bewältigt. Die Körperbewegungserkennung durch Computer Vision Algorithmen, eine kombinatorische Emotionserkennung textueller, auditiver und visueller Quellen und die Erschaffung einer selbstlernenden Datenbank zur Gebärdensübersetzung.

QUELLEN

- Cao, Zhe u. a. (2017/04/14): Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, 2016, <<http://arxiv.org/pdf/1611.08050>>
- Christopher Olah (2015): Understanding LSTM Networks, <<https://colah.github.io/posts/2015-08-Understanding-LSTMs/>> [Zugriff 2018-10-05]
- Coldewey, Devin: SignAll is slowly but surely building a sign language translation platform, <<https://techcrunch.com/2018/02/14/signall-is-slowly-but-surely-building-a-sign-language-translation-platform/?guccounter=1>> [Zugriff 2018-09-23]
- Daunt, Joe (11.09.2016): This virtual character can translate speech into sign language, <<http://uk.businessinsider.com/hand-talk-app-translate-speech-sign-language-hugo-brazil-libras-2016-9?R=T>> [Zugriff 2018-09-23]
- Deutscher Gehörlosen-Bund e.V. (Hrsg.): Deutsche Gebärdensprache (DGS)- FAQ, <[http://www.gehoerlosen-bund.de/faq/deutsche%20geb%C3%A4rdensprache%20\(dgs\)](http://www.gehoerlosen-bund.de/faq/deutsche%20geb%C3%A4rdensprache%20(dgs))> [Zugriff 2018-09-22]
- Gehörlosigkeit- FAQ, <<http://www.gehoerlosen-bund.de/faq/geh%C3%B6rlosigkeit>> [Zugriff 2018-09-22]
- Epo (2005): Weltweit sind 70 Millionen Menschen gehörlos- entwicklungspolitik online, <https://www.epo.de/index.php?option=com_content&view=article&id=968:weltweit-sind-70-millionen-menschen-gehoerlos&catid=52&Itemid=100> [Zugriff 2018-09-22]
- Fraunhofer-Institut für Integrierte Schaltungen ILS (Hrsg.) (05.10.2018): Shore: Gesichtserkennung: Schnell, robust und in Echtzeit, <<https://www.iis.fraunhofer.de/de/ff/sse/ils/tech/shore-facedetection.html>> [Zugriff 2018-10-06]
- Güler, Riza Alp/Neverova, Natalia/Kokkinos, Iasonas (2018/02/01): DensePose: Dense Human Pose Estimation In The Wild, 2018, <<http://arxiv.org/pdf/1802.00434>>
- Lernhelfer (Hrsg.) (2010): Phonem in Deutsch | Schülerlexikon | Lernhelfer, <<https://www.lernhelfer.de/schuelerlexikon/deutsch/artikel/phonem>> [Zugriff 2018-10-08]
- Mascha Schnellbacher (2016): KI DeepMind: Bessere menschenähnliche Sprache dank neuronaler Netze, <<https://entwickler.de/online/tools/ki-deepmind-neurale-netzwerke-276374.html>> [Zugriff 2018-09-22]
- Matthias Bastian (2018): Neue Sprachsynthese: Google-KI klingt jetzt wie ein Mensch, <<https://vrod.de/neue-sprachsynthese-google-ki-klingt-jetzt-wie-ein-mensch/>> [Zugriff 2018-10-05]
- Mehta, Dushyant u. a. (2017): VNet, in: ACM Trans. Graph. 36 (2017), S. 1–14, <https://doi.org/10.1145/3072959.3073596>
- Merkert, Johannes (2016): Google: Translate-KI übersetzt dank selbst erlernter Sprache, <<https://www.heise.de/newsticker/meldung/Google-Translate-KI-uebersetzt-dank-selbst-erlernter-Sprache-3502351.html>> [Zugriff 2018-09-22]
- Mikolov, Tomas u. a. (16.01.2013): Efficient Estimation of Word Representations in Vector Space, <<https://arxiv.org/pdf/1301.3781.pdf>> [Zugriff 2018-10-05]
- Rogers, Kaleigh (2018): Augmented Reality App Can Translate Sign Language Into Spoken English, and Vice Versa, <https://motherboard.vice.com/en_us/article/zmgnd9/app-to-translate-sign-language> [Zugriff 2018-09-23]
- Schreibman-Smith, Helen (2008): Why subtitles aren't always enough, <<https://www.theguardian.com/film/filmblog/2008/nov/28/deaf-subtitles-sign-language-film>> [Zugriff 2018-09-23]
- SIL International Publications (Hrsg.): Ethnologue- Languages of the World: Germany, <<https://www.ethnologue.com/country/de>> [Zugriff 2018-09-22]
- Simon, Tomas u. a. (2017/04/25): Hand Keypoint Detection in Single Images using Multiview Bootstrapping, 2017, <<http://arxiv.org/pdf/1704.07809>>
- Zellfelder, Steffen (06.06.2018): Die besten Übersetzer-Apps für Android, <<https://www.pcwelt.de/ratgeber/11-Apps-die-fuer-Sie-uebersetzen-Verstehen-und-verstanden-werden-9658892.html>> [Zugriff 2018-09-22]

AUTOR Christopher Caldwell
Alte Dorfstr. 9
70599 Stuttgart
cc026@hdm-stuttgart.de
Matrikelnummer: 34957

Modul 25304 Aktuelle Themen
Künstliche Intelligenz und ihre Auswirkungen auf Ihre Zukunft

Prof. Dr. Andreas Koch

Masterstudiengang
Audiovisuelle Medien
Sommersemester 2018

Hochschule der Medien

www.hdm-stuttgart.de