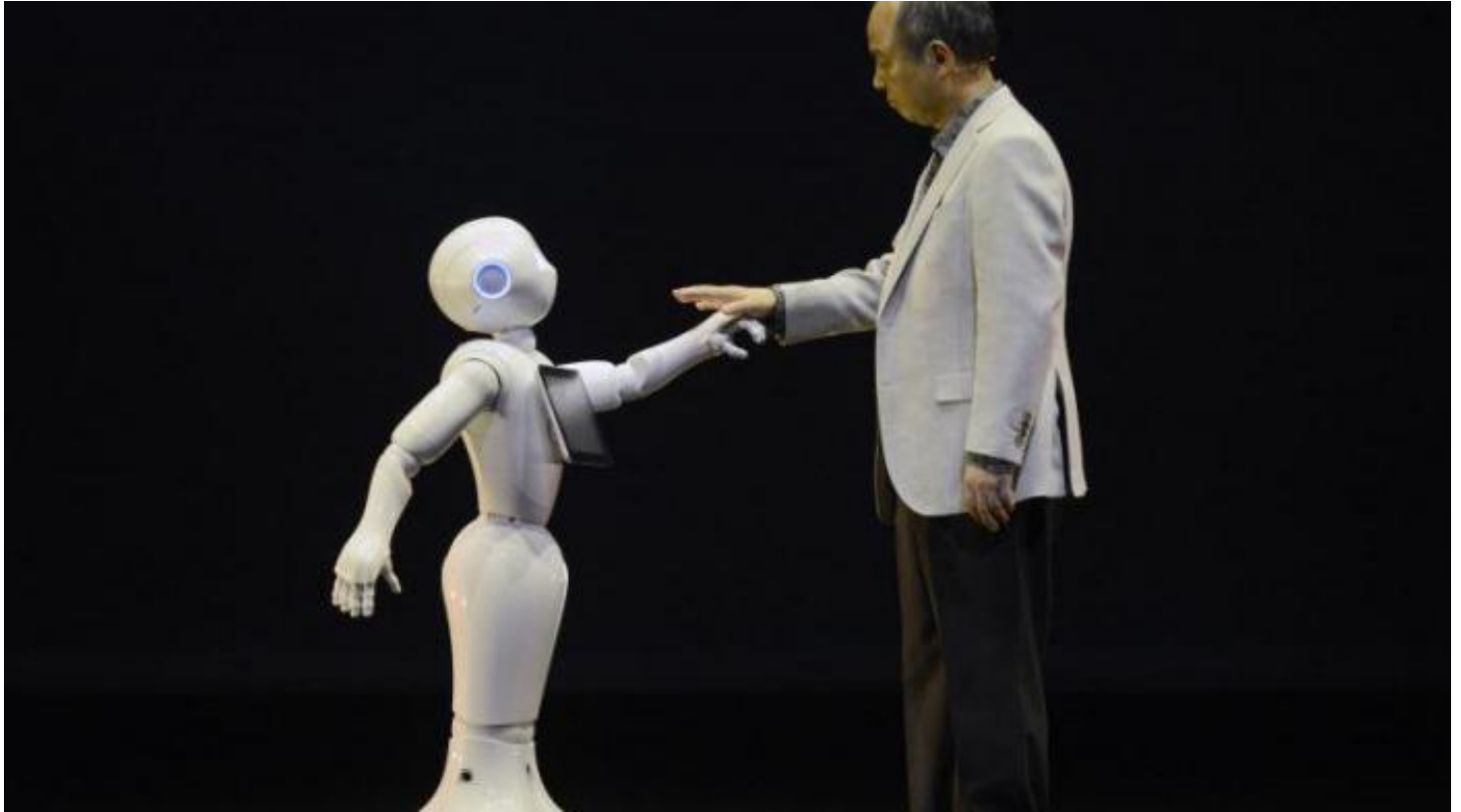


Wenn der Roboter Wünsche von den Augen abliest

20.10.2017 09:16 Uhr

Hans-Arthur Marsiske



(Bild: dpa, Franck Robichon/Archiv)

In Bielefeld tauschen sich derzeit Forscher zur Interaktion zwischen Menschen und künstlichen Agenten aus. Die Spracherkennung hat große Fortschritte gemacht, aber die Worterkennung alleine reicht nicht. Augenaufschläge spielen auch eine Rolle.

Die häufigste Form der Interaktion zwischen Menschen dürfte nach wie vor das Reden sein. Daher beschäftigen sich auch in der **International Conference on Human-Agent Interaction (HAI 2017)[1]**, die gegenwärtig in Bielefeld stattfindet, etliche Beiträge mit der sprachlichen Verständigung zwischen Mensch und künstlichem Agenten.

Rapide Verbesserung der Spracherkennung

Alexander Waibel (Carnegie Mellon University/Karlsruhe Institute of Technology) gab einen Einblick in seine Forschungen zur Spracherkennung und -verarbeitung. Er verwies auf die von ihm im Jahr 1987 vorgestellten Time Delay Neural Networks (TDNN), die eine wichtige Grundlage für diese Techniken darstellten und jetzt als Convolutional Neural Networks (CNN) wiederentdeckt würden. "Ich hätte es allerdings niemals für möglich gehalten, dass sich die

Rechenkapazität von Computern innerhalb meines Berufslebens um den Faktor eine Milliarde steigern würde", sagte er.

Ein Mensch spreche ungefähr 500 Millionen Wörter in seinem Leben, heutige Computer würden mit einem Vielfachen davon trainiert. Das habe entscheidend dazu beigetragen, dass aus dem ersten Übersetzungsprogramm, das 1991 über einen Wortschatz von lediglich 500 Wörtern verfügte, mittlerweile Systeme geworden sind, die gesprochenen Text simultan in eine andere Sprache übersetzen können – und dabei sogar die Interpunktion ziemlich gut erkennen, wie Waibel in einem kurzen Video zeigte. Das cloud-basierte Verfahren läuft seit 2012 in Karlsruhe in vier Hörsälen.

Mehr als nur Wörter

Natürlich stoßen auch diese Systeme immer noch an Grenzen, etwa bei Wörtern wie "cloud-basiert", die aus verschiedenen Sprachen zusammengesetzt sind oder Verben, die nach den Regeln einer anderen Sprache konjugiert werden, etwa wenn jemand erklärt, etwas "gedownloaded" zu haben. Auch mehrdeutige Sätze können automatische Übersetzer vor Probleme stellen. Waibel zitierte den Satz: "If the baby doesn't like the milk, boil it." Einem Menschen ist klar, dass die Milch gekocht werden soll, nicht das Baby. Aber das erschließt sich nur aus dem allgemeinen Wissen, nicht aus der Grammatik.



In einer Gesprächsrunde können Blicke und Zugewandtheit darauf hindeuten, wer als nächstes redet.

Zudem ist es mit der Sprache allein auch nicht getan. Menschen kommunizierten praktisch immer multimodal, sagte Waibel, das heißt, auf mehreren Ebenen gleichzeitig. Neben der gesprochenen Sprache seien auch Gestik, Blickrichtung, die Umgebung und viele andere Faktoren von Bedeutung. Durch Beachtung solcher Signale gelingt es Menschen zum Beispiel bei Gruppengesprächen, einander nicht ins Wort zu fallen.

Zugewandtheit zur Ablaufferkennung

Ryo Ishii (NTT Corporation) berichtete von Experimenten, bei denen ein Computer anhand der Kopfbewegungen der Teilnehmer einer vierköpfigen Gesprächsrunde erkennen sollte, wann ein Redebeitrag zum Ende kam und wer als nächstes das Wort ergreifen würde. Natürlich seien auch Blickrichtungen, Atmung und Mundbewegungen wichtige Indikatoren, so Ishii. Doch allein auf Grundlage der Kopfbewegungen sei es möglich gewesen, den nächsten Rednerwechsel 1,5 Sekunden vorher zu erkennen.

Ebenfalls mit Vierergruppen hat das Forschungsteam um Seiya Kimura an der Ritsumeikan

University gearbeitet. Hier geht es darum, einen Roboter oder virtuellen Agenten zu entwickeln, der Stellenbewerbern helfen kann, sich auf Gruppengespräche vorzubereiten, die von Firmen mehr und mehr bei der Auswahl neuer Mitarbeiter eingesetzt werden. Dafür muss der Agent seine Aufmerksamkeit auf die übrigen Gesprächsteilnehmer richten, je nachdem, ob er selbst gerade spricht, einem anderen zuhört oder ob gerade niemand etwas sagt.

Als Grundlage für das Training des Agenten wurden Daten aus 15-minütigen Gruppendiskussionen erhoben, bei denen neben den Kopfbewegungen auch Blickrichtungen oder Sprachrhythmen erfasst wurden. Ob der Agent schon gut genug trainiert wurde, um nun wiederum Menschen zu trainieren, wurde nicht ganz klar. Es scheint, dass noch etwas mehr Forschung nötig ist, bevor Roboter als Bewerbungsscoachs antreten können.

Von den Augen ablesen

Die Königsklasse sind ohnehin Roboter, die dem Menschen seine Wünsche von den Augen ablesen. Von einem sehr interessanten Ansatz in dieser Richtung berichtete Siti Aisyah binti Anas, die mit ihren Kollegen an der Eindhoven University of Technology eine Kaffeemaschine mit Augen ausgestattet hat. Es sei nicht schwierig, Geräte zu konstruieren, die auf menschliche Blicke reagieren, sagte sie. Das Problem seien die Menschen, die es nicht gewohnt seien, Maschinen auf diese Weise zu steuern.

Ausgehend von Beobachtungen, wie Menschen untereinander über Blicke kommunizieren, haben die Forscher ihren Kaffeeroboter CoffeePet mit drei verschiedenen Verhalten getestet, von einem starren Blick über einen, der den Bewegungen des Menschen folgt, bis zu einem, der dessen Blickrichtungen folgt. Versuchspersonen wurden vor die Kaffeemaschine gesetzt, wo sie eine Aufgabe mit Legosteinen lösen sollten. Zu CoffeePet wurde ihnen nichts gesagt, sie sollten die Funktion dieses Roboters von selbst erkennen. Das gelang tatsächlich auch allen elf Teilnehmern, wenn CoffeePet seine Augen am menschenähnlichsten bewegte: Sie hielten den Blickkontakt lange genug, sodass der Roboter ihnen einen Kaffee zubereitete, und verstanden auch das Prinzip der Steuerung über Blicke.

In den anderen Verhaltensmodi von CoffeePet erkannten das nur drei beziehungsweise (beim starren Blick) nur einer. Einige Versuchsteilnehmer empfanden



Schau mir in die Augen, CoffeePet!
(Bild: Siti Aisyah binti Anas, Matthias Rauterberg, Jun Hu, Eindhoven University of Technology: "Exploring Gaze-Activated Object With the CoffeePet")

es als unangenehm, von der Maschine beobachtet zu werden, sagte binti Anas. Ob der Kaffee gut genug geschmeckt hat, um als Entschädigung akzeptiert zu werden, verriet sie nicht. Die in der Diskussion ihres Vortrags geäußerte Anregung, die Stärke des Kaffees von der Dauer des Blickkontakts abhängig zu machen, nahm sie aber mit Interesse zur Kenntnis. (*Hans-Arthur Marsiske*) / (**kbe[2]**)

URL dieses Artikels:

<http://www.heise.de/-3865923>

Links in diesem Artikel:

[1] <http://hai-conference.net/hai2017/>

[2] <mailto:kbe@heise.de>

Copyright © 2017 Heise Medien