

DIE ZEITSCHRIFT FÜR TRENDS IN TECHNIK UND IT



iT-Lösungen für das
Gesundheitswesen

Gesagt – Getan

Mensch spricht – Maschine schreibt

Große Systemübersicht 2019

580 Produkte von 55 Herstellern

DIAM

Ein Wegweiser für die Radiologie-IT und Imaging-IT

Dosis

Ausführliche Systemübersicht

SONDERDRUCK
Computerbasierte Spracherkennung



GESAGT – GETAN

MENSCH SPRICHT – MASCHINE SCHREIBT

Computerbasierte Spracherkennung hat bemerkenswerte Fortschritte gemacht. Automatische Übersetzungen haben in der Zwischenzeit für die verbreiteten Sprachen ein brauchbares Niveau erreicht und werden bald auch gut genug sein, um nur noch marginal durch Menschen verbessert zu werden. Der entscheidende Durchbruch begann allerdings erst in diesem Jahrzehnt mit den Deep Learning-Methoden auf Basis digitaler neuronaler Netzwerke.



Schriftsteller, Physiker und Oscar-Nominee Sir Arthur C. Clarke (1917–2008) hat nicht nur die Roman-Vorlage für den Film „2001: Odyssee im Welt-raum“ geschrieben, sondern ist auch für die Bemerkung bekannt, dass jede nur weit genug entwickelte Technologie sich nicht von Zauberei unterscheiden lässt. „Sag einen Zauberspruch, und ein in der Nähe befindliches Gerät wird Dir diesen Wunsch erfüllen.“

Amazon Echo hört auf den Namen Alexa, spielt Musik, erzählt Witze und steuert das Smart Home. Apples Siri verarbeitet pro Woche derzeit mehr als zwei Milliarden Befehle und allein in den USA werden bereits mehr als 20 Prozent aller Suchen in Google auf Android-Geräten über verbale Kommunikation abgewickelt. Warum tippen, wenn man auch sprechen kann?

Sprachkommunikation mit automatisierten Systemen hat in den vergangenen Jahren einen großen Sprung nach vorn gemacht, auch wenn trotz des Hypes um die Künstlichen Intelligenzen diese noch lange nicht auf dem Niveau eines menschlichen Pendants sind (Amazon hat ein Preisgeld in Höhe von einer Million US-Dollar ausgeschrieben für den Chatbot, der einen Menschen mindestens 20 Minuten lang in einer Diskussion festhält, bei der dieser den Automatencharakter seines Gegenüber nicht realisiert). Aber obwohl Sprachtechnologien außer in ein paar wenigen hochspezialisierten Anwendungen noch keine Menschen ersetzen können, sind sie inzwischen auf einem ernstzunehmenden Niveau angekommen.

Statistische Methoden verbessern Analyse

Computerbasierte Spracherkennung hat bemerkenswerte Fortschritte gemacht. Automatische Übersetzungen haben in der Zwischenzeit für die verbreiteten Sprachen ein brauchbares Niveau erreicht und werden bald auch gut genug sein, um nur noch marginal durch Menschen verbessert zu werden.

Die Technologie könnte schon weiterentwickelt sein, doch die ersten Versuche verliefen erfolglos. Viele frühe Technologien blieben in der Sackgasse eines regelbasierten Ansatzes stecken, aber Sprachen sind mit vielen Ausnahmen und individuellen Eigenheiten gespickt und zwangen damit jedes regelbasierte System zur Aufgabe.

Entscheidende Verbesserungen zeigten sich erst durch die Anwendung statistischer Methoden und die Analyse großer Datenbestände, um darin Muster aufzufinden und vorhergehende Analyseergebnisse ständig weiter zu verfeinern. Dank steigender Prozessorleistung und vor allem einer explosionsartig zunehmenden Verfügbarkeit an Daten entwickelten sich die Leistungsfähigkeit und Qualität der Spracherkennungssysteme entscheidend weiter.

Der Durchbruch begann allerdings erst in diesem Jahrzehnt mit den Deep Learning-Methoden auf Basis digitaler neuronaler Netzwerke. Seitdem gelingt es, zahlreiche Herausforderungen zu meistern: Von Menschen geäußerte Laute sind hochgradig individuell. Sprecher unterscheiden sich in Lautstärke, Tonhöhe, Sauberkeit der Aussprache, Schnelligkeit und Dialekten, um nur einige Eigenschaften zu nennen. Diese können auch menschliche Zuhörer vor Probleme stellen und machen es den Computern damit schwer, Wortgrenzen korrekt zu identifizieren.

Hochentwickelte Sprachmodelle heutiger Erkennungssysteme sind in der Lage, gute Ergebnisse zu liefern. Sie bauen auf Wahrscheinlichkeitsmodellen auf, berechnen damit aufgrund des bereits Erkannten, welches Wort voraussichtlich folgen wird, und sortieren die unwahrscheinlichsten Fälle einfach aus. Die individuellen Eigenschaften eines Sprechers übernehmen heutige Systeme durch kurze Trainingseinheiten. In Kombination mit dem vergleichsweise fest umschriebenen Wortschatz eines medizinischen Fachvokabulars lassen sich Erkennungsraten von mehr als 99 Prozent erzielen – ein gutes Mikrofon und eine ruhige Umgebung vorausgesetzt.

Aber auch daran arbeiten die Entwickler: Microsoft zum Beispiel stellt seinen Partnern das Produkt CRIS zur Verfügung, über das Hintergrund-Störgeräusche effektiv reduziert werden können.

**Konzentration auf das Wesentliche:
Sprachassistenten sind nicht nur
bei Vielschreibern unverzichtbare
Hilfsmittel, die erheblich Zeit und
Aufwand sparen.**

Automatische Übersetzung

Eine besondere Herausforderung bei Übersetzungen ist die Bedeutung eines Wortes, die nicht nur von einer lexikalischen Definition und dem grammatikalischen Kontext abhängt, sondern darüber hinaus auch noch von der Bedeutung der anderen Satzteile.

Ein qualitativ großer Sprung nach vorn wurde aber erst ab dem Jahr 2007 durch die Aktivitäten von Google erreicht: mehr als zwei Billionen Webseiten wurden nach Texten durchsucht, die Übersetzungen sind oder sein könnten. Das Training eines Systems mit zwei parallel-laufenden Texten ermöglicht den Aufbau eines Übersetzungs-Modells für dieses Sprachen-Paar, das konsequenterweise eine Mehrzahl möglicher Übersetzungen in die Zielsprache enthält. Ähnlich wie bei der im vorhergehenden Abschnitt diskutierten Methodik zur Spracherkennung wird dann innerhalb der Zielsprache die Wahrscheinlichkeit dafür ermittelt, dass es sich bei dem möglichen Ergebnis um einen sinnvollen Satz handelt. Die Systeme lernen aus den verfügbaren Datenmengen und bestimmen daraus selbständig das Ergebnis aufgrund des am wahrscheinlich höchsten Rankings.

Google hat seine eigene auf neuronalen Netzen basierende Übersetzungsmaschine für acht Sprachpaare veröffentlicht: Im Vergleich zu den ersten regelbasierten Systemen wie babelfish sind tatsächlich erhebliche Fortschritte sichtbar geworden, aber nach wie vor machen auch diese modernen Systeme Fehler, vor allem bei strukturell unterschiedlichen Sprachpaaren wie Deutsch und Chinesisch. Trotzdem zeigen die Algorithmen auch dabei in den meisten Fällen brauchbare Ergebnisse an.



Das ursächliche Problem bei der Erstellung sinnvoller automatischer Übersetzungen ist das Verständnis für den Textinhalt. Da sich aufeinander beziehende Wörter im Satzbau weit auseinander liegen können, muss dafür ein sogenanntes Attention Model aufgebaut werden, das Wortbezüge auch außerhalb einer direkten Wortfolge registriert und bewertet.

Zusätzlich zu dem oben beschriebenen Modell verbleiben weitere Herausforderungen. Alle heutigen Übersetzungssysteme gehen Satz für Satz vor: Falls die Übersetzung eines bestimmten Satzes von der Bedeutung eines vorhergehenden Satzes abhängt, sind Fehler im Sinn des Wortes vorprogrammiert.

Andererseits ist genau das der Vorteil für die medizinische Sparte, da hier der im Vergleich zum üblichen Sprachgebrauch seltene spezielle Wortschatz klar umrissen ist und umfangreiches Trainingsmaterial zur Verfügung steht. Die entsprechenden Lexika wie RadLex und SNOMED sind ausformuliert und stehen zumindest teilweise in unterschiedlichen Sprachpaarungen zur Verfügung.

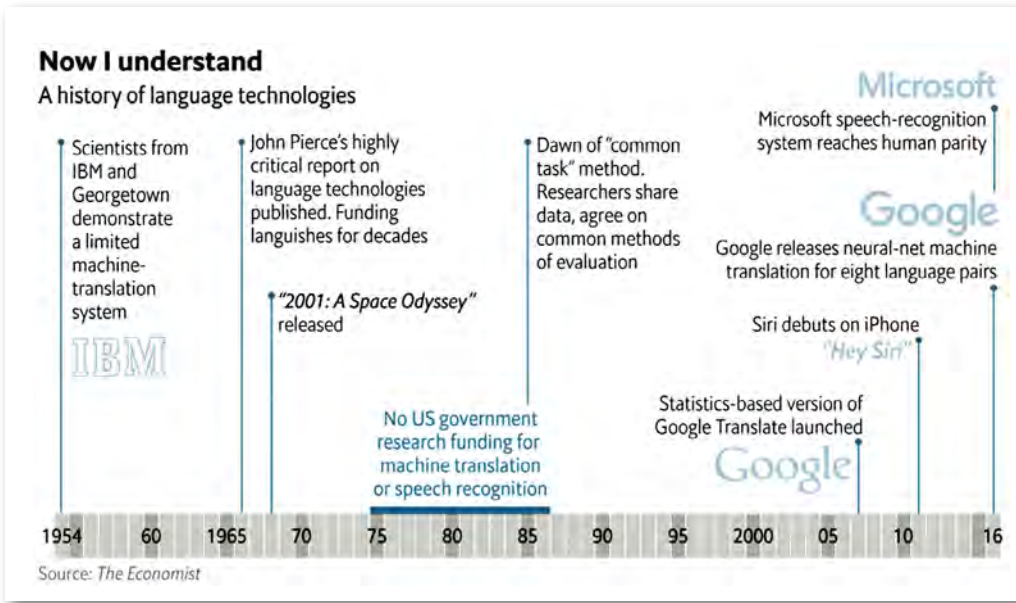
In einem weiteren Schritt kann damit nicht nur die mehrsprachige Kommunikation auf Basis strukturierter Informationen und auf Expertenebene erfolgen, sondern auch eine automatische Über-

setzung für Patienten. Diese werden beispielsweise mit dem Begriff „Gallensteine“ weit mehr anfangen können als mit der in der Radiologie üblichen Interpretation „Sichtbare Gruppe biliärer hyperechogener Kalzifikationen im dorsalen Schallschatten“ eines Ultraschallbilds der Galle.

Semantik – Der Sinn des Gesagten

Erst das Zusammenwachsen unterschiedlicher Technologien ermöglicht den Einsatz heutiger Persönlicher Assistenten. „Wie kalt ist es gerade in Frankfurt?“, eine der knapp zwei Milliarden gestellten Fragen, die Apples Siri in der Woche beantwortet, bedingt das Erkennen des Schlüsselworts „kalt“, das in Zusammenhang mit der physikalischen Eigenschaft Temperatur gebracht werden muss, die Kenntnis einer Stadt namens Frankfurt und deren geographischen Lage, sowie die Verbindung zu der Datenbank des Deutschen Wetterdienstes oder einer vergleichbaren Institution, die Auskunft über die aktuellen Temperaturdaten geben kann. Jede dieser Anfragen wird in den Regelkreis des neuronalen Netzwerks eingespielt und verfeinert die Lernkurve des Systems.

An ihre Grenzen kommen die Systeme bei Folgefragen, die ihren Kontext überschreiten: „In welchem italie- >>



Start mit Hindernissen: Nach ersten Gehversuchen in den Fünfziger Jahren und einer Durstphase in den Achtzigern entwickeln sich Technologien rund um die Sprache zügig weiter. Die Verfügbarkeit verschiedener kombinierbarer Mehrwerte führt heute zu einem exponentiellen Wachstum.

nischen Restaurant können meine Frau und ich heute essen gehen?" resultiert in der Regel in einer brauchbaren Antwort, die Folgefrage „Wie weit ist es von ihrem Büro entfernt?“ schon nicht mehr.

Sowohl grammatikalisch mögliche als auch unsinnige Varianten steigen exponentiell an, je länger Sätze werden. Automatisierte Systeme orientieren sich unter anderem daran, dass manche Wortkombinationen öfter vorkommen als andere: Substantiv – Verb – Substantiv ist in dieser Kombination weit häufiger als Substantiv – Verb – Verb. Auch hier geht es wieder um Wahrscheinlichkeiten, die das System bewertet und danach entscheidet. Aber die Bedeutung eines Satzes zu erfassen ist weit schwieriger als die Syntax. Die Ausdrücke „Der Junge hat den Ball geworfen“ und „Der Ball wurde von dem Jungen geworfen“ haben trotz unterschiedlicher Struktur dieselbe Bedeutung. „Natural Language Processing“ ist der Überbegriff für eine junge Disziplin,

die sich als Aufgabe die Erfassung von Sprache und deren Sinn gestellt hat. Dazu durchläuft das Ausgangsmaterial (Gesprochenes oder Text) folgende Prozesse:

- Spracherkennung
- Tokenization: Segmentierung des Erfassten in einzelne Wörter und Sätze
- Lemmatisierung: Erkennung von Grundformen der Wörter und Erfassung grammatischer Informationen
- Wortartbestimmung: Erkennen der Funktionen einzelner Wörter im Satz (Subjekt, Verb, Objekt, Artikel, etc.)
- Syntaktische Analyse: Extraktion der Bedeutung von Sätzen und Satzteilen

• Semantische Analyse: Erkennen von Satzzusammenhängen und Satzbeziehungen

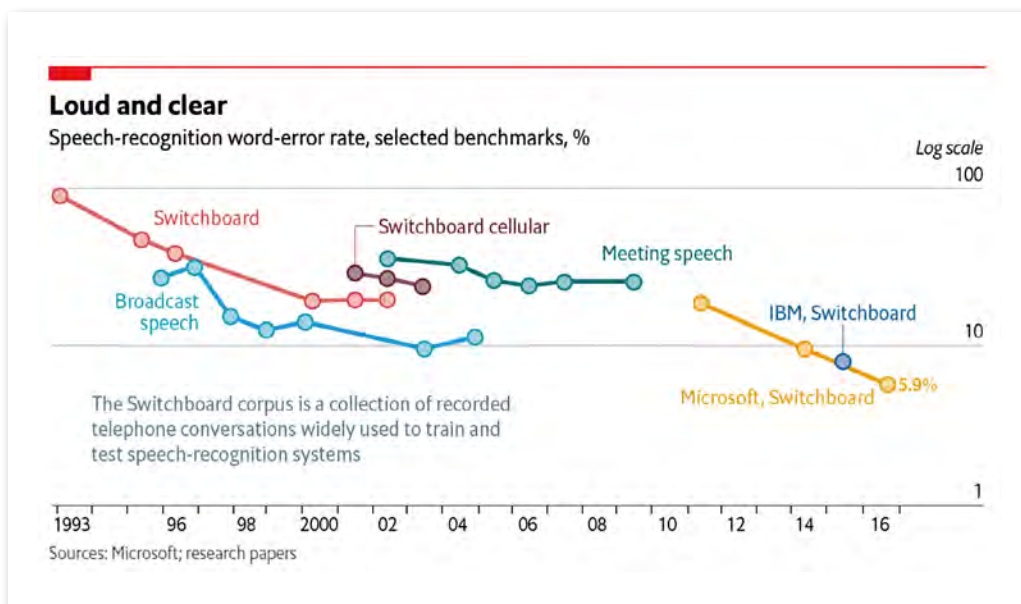
Allerdings gibt es in der menschlichen Sprache oft Mehrdeutigkeiten, so dass selbst ein vollständiger Durchlauf der Prozesskette nicht immer ein eindeutiges Ergebnis liefern wird. Und an der Verwendung von Stilmitteln wie der rhetorischen Frage, Ironie oder Paradoxons scheitern oft auch menschliche Gesprächspartner.

Auch hier gilt wieder: Je begrenzter und strukturierter der verwendete Wortschatz ist, desto einfacher gestaltet >>

Brücke zwischen den Welten: Zielgruppenorientierte automatische Übersetzung unterstützt den inhaltlich eindeutigen Austausch im internationalen Experten-umfeld (Arzt-zu-Arzt-Kommunikation), sowie die verständliche Interpretation im Gespräch zwischen Arzt und Patient.



In den letzten 25 Jahren ist es gelungen, die Fehlerrate von Spracherkennungs-Systemen um mehr als eine Zehnerpotenz zu reduzieren.



sich auch die semantische Analyse für die Computersysteme. Ein korrekter Übergang von frei gesprochenem Text hin zum automatischen Ausfüllen von Struktur-Datenfeldern in einem medizinischen Formular, wie es beispielsweise bei strukturierten Befunden oder in der Onkologie zur Beschreibung von Tumoren verwendet wird, ist aufgrund des klar definierten (und begrenzten) Fachwortschatzes wesentlich einfacher zu realisieren als eine sinnvolle Mensch-Maschine-Kommunikation im täglichen Alltag außerhalb der Berufswelt.

Die genannten Methoden und Technologien werden nicht nur im weiten Feld der automatisierten Sprachverarbeitung angewendet, sondern auch in vielen weiteren Bereichen. Methoden zur Mustererkennung und Deep Learning sind prädestiniert für Anwendungen in bildfokussierten Bereichen wie in der Radiologie zur Analyse von Millionen von Bildern, die in den PACS der Kliniken gespeichert sind. Decision Support-Systeme geben auf Grundlage verfügbarer Informationen Hinweise auf vergleichbare Pathologien und Empfehlungen zum Vorgehen. Die heute verfügbaren Technologien werden in den kommenden Jahren die Arbeitswelt nicht nur in der Medizin entscheidend umprägen: Computersysteme werden den Status von persönlichen Assisten-

ten erreichen und – mit dem Beginn in Nischen wie medizinischen Spezialabteilungen – entweder entscheidende Hinweise aufgrund der Analyse eines für Menschen unüberschaubar gewordenen Datenbestands geben oder Tätigkeiten wie einfache Diagnostik selbständig übernehmen. Der Facharzt übernimmt dann die Rolle der Qualitätssicherung

und kann sich der Diagnostik komplexerer Fragestellungen widmen, für die dann mehr Zeit zur Verfügung steht. ■

Quellen

The Economist
Korean J Radiol. 2017 Jul-Aug;18(4):570-584. English.
Published online May 19, 2017.
<https://doi.org/10.3348/kjr.2017.18.4.570>
Copyright © 2017 The Korean Society of Radiology
Deep Learning in Medical Imaging: General Overview

Über den Autor

Bernd Sadlo ist seit über acht Jahren Key Account Manager und mitverantwortlich für die Produktentwicklung bei DFC-SYSTEMS. DFC-SYSTEMS gehört seit über 20 Jahren zu einem der führenden Anbieter von Sprachlösungen im Gesundheitswesen. Neben etablierten Lösungen für Digitales Diktatmanagement und Spracherkennung entwickelt DFC-SYSTEMS auch Lösungen für die Sprachsteuerung und die semantische Spracherkennung. Gemeinsam mit Partnern aus Forschung und Lehre sowie der Industrie arbeitet das Unternehmen aktuell an Lösungen für die berührungsfreie sprachbasierte Bedienung von Applikationen und medizinischen Devices sowie an einer automatisierten strukturierten Befunderstellung auf Basis „semantischer Spracherkennung“. DFC-SYSTEMS setzt bei der Entwicklung seiner indicda Sprachlösungen auf bekannte Technologien, unter anderem auf die vorstehend genannte Spracherkennungstechnologie der Firma Microsoft.



iT-Lösungen für das Gesundheitswesen



KONTAKT:

DFC-SYSTEMS GmbH Einsteinring 28 · 85609 Aschheim bei München
tel 0 89 / 46 14 87-0 · fax 0 89 / 46 14 87-11 · info@dfcsystems.de · www.dfcsystems.de

