

Dokumentationsbegleiter, KI für das Gesundheitswesen

Warum wir Deep Learning für unsere Spracherkennungssoftware nutzen

[Nuance Communications](#)

30 September 2019



Jeder Mensch verwendet Sprache auf seine eigene Art – wie er spricht und welche Wörter er nutzt, sind bei jedem Menschen unterschiedlich. Diese Individualität von Sprechern ist von großer Bedeutung, wenn es darum geht, mit Hilfe von Deep Learning und neuronalen Netzwerken bessere Nutzererfahrungen zu erzielen

Deep Learning

Jeder Mensch verwendet Sprache auf seine eigene Art. Wie er spricht und welche Wörter er nutzt ist bei jedem Menschen unterschiedlich. Moderne Spracherkennungssysteme eliminieren diese Variationen durch Training mit Sprachdaten. Zahlreiche Akzente, Altersgruppen und andere denkbare Variationen von Sprachstilen werden ebenso abgedeckt. Das Ergebnis sind sehr stabile Systeme, die für (fast) alle Sprecher gut funktionieren. Wir nennen das eine „sprecherunabhängige“ Spracherkennung. Wofür ist dann Deep Learning?

Sprecherunabhängige Spracherkennung

In einigen Fällen kann man aber auch anders auf die Individualität des Sprechers eingehen. Deep Learning und neuronalen Netzwerken erzielen viel bessere Nutzererfahrungen. Ein Beispiel hierfür ist unsere [Dragon Medical Software](#). Sie wird für gewöhnlich von nur einem einzigen Nutzer verwendet. Dieser Aspekt ermöglicht es uns, durch die Anpassung an jeden Nutzer über die sprecherunabhängige Spracherkennung hinaus zu gehen. [Dragon Medical](#) macht das auf mehreren Ebenen:

- Das Programm macht sich den aktiven Wortschatz des Nutzers zu eigen:
 - Es prüft die in der Vergangenheit erstellten Texte.
 - Untypische, aber von diesem Nutzer verwendete Begriffe werden in das aktive Vokabular aufgenommen.
 - Es lernt die für den Nutzer typischen Sätze, Redewendungen und Textmuster.
- Während jeder Sitzung erfasst das Programm innerhalb von nur wenigen Sekunden, wie die Aussprache von Wörtern erfolgt, und passt sich entsprechend an.
- Damit kann das Programm auch auf den aktuellen Klang der Stimme des Nutzers reagieren. Beispielsweise wenn der Anwender erkältet ist, ein anderes Mikrofon verwendet oder sich das Umfeld verändert hat.
- Während des optionalen initialen Trainings oder nach Abschluss einer jeden Diktatsitzung durchläuft Dragon Medical weitere intensive Lernprozesse. Und passt sich so mit der Zeit an nutzerspezifische Sprachmuster an.

Dragon Medical verwendet durchgängig Deep Learning und neuronale Netzwerke. Sowohl auf der Ebene des Sprachmodells (Erfassung der Häufigkeit von Wörtern und ihrer typischen Kombinationen) als auch auf der des akustischen Modells (die die Phoneme oder kleinsten gesprochenen Einheiten einer Sprache modellieren).

Graphical Processing Units

Diese Modelle sind recht umfangreich und werden bereits während der Entwicklung mit großen Datenmengen trainiert. Neuronale Netze wurden zwar bereits im späten 20. Jahrhundert entwickelt, können aber erst jetzt richtig zum Einsatz kommen, da wir erst heute über die nötige Rechenleistung verfügen. Beim Training unserer Modelle verwenden wir daher auch sogenannte GPUs (Graphical Processing Unit), die ursprünglich für grafische Computeranwendungen, wie Videospiele, entwickelt wurden. Die Berechnung von Bildern und das Training von Deep Neural Networks haben viele Gemeinsamkeiten, da beide Aufgaben Berechnungen vieler Datenpunkte zur selben Zeit erfordern. Für einen schnellen Trainingsprozess werden daher mehrere parallel geschaltete GPUs eingesetzt.

Die erfolgreiche Umsetzung in Dragon Medical ist ein wichtiger Meilensteinnach nach vielen Jahren der innovativen Forschungs- und Entwicklungsarbeit auf dem Gebiet der Spracherkennung und des maschinellen Lernens.

Tags: [Spracherkennung](#)

More Information

Möchten Sie auch Dragon Medical nutzen?

Erfahren Sie welches Potenzial die KI-basierte Spracherkennung für Sie bereit hält!

[Learn more](#)