



Weniger Aufwand bei der klinischen Dokumentation dank **künstlicher Intelligenz**



Menschliche Intelligenz – vervielfacht

Den Arztberuf ergreift man meist, weil man Menschen helfen möchte – Gesunde sollen gesund bleiben und Kranke geheilt werden. Die klinische Dokumentation ist natürlich ein wichtiger Teil der Patientenversorgung, beansprucht aber Zeit, Aufmerksamkeit und Energie des Arztes, die dieser lieber dem Patienten widmen würde. Unser Ziel ist es, Technologien zu schaffen, die Ärzten und Patienten einen Teil der Dokumentationslast abnehmen, Dokumentationsprozesse optimieren und beschleunigen und damit die Patientenbetreuung und -zufriedenheit verbessern.

In diesem eBook erläutern unsere Experten, wie wir Technologien unter Einsatz von künstlicher Intelligenz und maschinellem Lernen nach dem Vorbild der Arbeitsweise des menschlichen Gehirns schaffen – und was dies für die Zukunft des Gesundheitswesens bedeutet.



Inhalt

Maschinelles Lernen: Wie viele neuronale Netze braucht man für die großen Fische?	4
Was ist „tief“ am „Tiefen Lernen“?	6
Die Zukunft des Gesundheitswesens: wie klinische Umgebungsintelligenz zu besserer Versorgung beiträgt.	11

Maschinelles Lernen: Wie viele neuronale Netze braucht man für die großen Fische?

Von Nils Lenke, Senior Director Corporate Research,
Nuance Communications

Tiefe neuronale Netze (Deep Neural Networks, DNN) dominieren in den letzten Jahren die Diskussionen und Schlagzeilen rund um das Thema maschinelles Lernen. Dabei stehen wir noch ganz am Anfang der Erforschung neuronaler Netze, denn in diesem sich entwickelnden und verändernden Fachgebiet gibt es zahlreiche Ansätze und Herausforderungen.

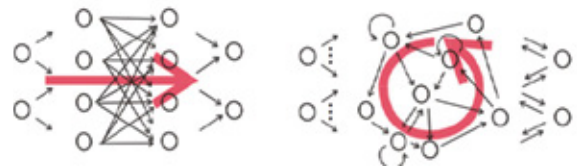
„Standard“-DNN sind unidirektional: Information fließt nur in eine Richtung, nämlich von der Eingabeschicht durch die verdeckten Schichten zur Ausgabeschicht. In der Fachsprache des maschinellen Lernens nennt man das ein DNN vom Typ „Feedforward“. Diese funktionieren dann am besten, wenn alle für das Lernen erforderlichen Informationen gleichzeitig vorliegen. Ein Beispiel ist die Bilderkennung: Das Bild steht sofort zur Verfügung, und das Netz kann das Gesehene in einem Durchgang durch das Netz verarbeiten.

Die Teams hier bei Nuance nutzen DNN für die Verbesserung der Spracherkennung und des Verstehens natürlicher Sprache im Rahmen unserer Arbeiten für eine verbesserte Kommunikation zwischen Mensch und Technologie.

Eine interessante Herausforderung gesprochener Sprache ist, dass sie, im Gegensatz zum Bild, eine zeitliche Dimension hat. Eine Äußerung erstreckt sich über mehrere Sekunden. Was vor einigen Minuten oder sogar Sekunden gesagt wurde, trägt zum Verständnis dessen bei, was jetzt gerade passiert – das nennt man Kontext. Man könnte rein technisch gesehen natürlich bis zum Ende der Äußerung warten, damit das DNN die gesamte Äußerung zur Verfügung hat. Ein Feedforward-Netz hätte dann sämtliche benötigten Informationen, um die Erkennung in einem Durchgang zu erledigen. Dies ist aber bei Dialogsystemen, wie beispielsweise persönlichen Assistenten, unrealistisch. Da Spracherkennung eine rechenintensive Aufgabe ist, müssen die Maschinen bereits am Anfang der Äußerung mit der Arbeit beginnen und versuchen, mit dem Sprecher Schritt zu halten, damit sie wie in einer Konversation zwischen Menschen bald nach dem Ende der Äußerung eine Antwort anbieten können.

Aus diesem Grund analysiert eine Spracherkennungs-Engine jeweils kleinere Spracheinheiten. Und um den Kontext zu erarbeiten, verwenden wir bei Nuance eine spezielle Art von DNN, die so genannten rekurrenten neuronalen Netze (RNN).

Deren Neuronen nehmen Inputs nicht nur von der linken Seite entgegen (siehe linke Abbildung), sondern haben auch Zugang zu ihren eigenen *früheren* Zuständen (oder bei manchen Varianten auch zu denen anderer Neuronen, siehe rechte Abbildung). Diese Feedback-Schleifen bilden eine Art Gedächtnis.



Dies lässt sich am Beispiel von Sprachmodellen (SM) darstellen: Diese sagen auf Basis der letzten soundsoviel gesprochenen Wörter das nächste Wort voraus (idealerweise wäre die Wortzahl nicht festgelegt, sondern variabel). Wenn Sie beispielsweise bereits „Einigkeit und Recht und“ gehört haben, ist „Freiheit“ eine viel wahrscheinlichere Fortsetzung als die meisten anderen Wörter. Unserer Erkenntnis nach arbeiten auf RNN basierende SM deutlich besser als herkömmliche SM.

Nun folgt die Aufgabe des Sprachverstehens (Natural Language Understanding, NLU), also die Zuordnung einer Bedeutung zu den erkannten Wörtern.

Eine Unteraufgabe hierbei ist die Erkennung von „benannten Entitäten“ (named entities). In der Frage



„Welche U-Bahn-Station liegt in der Nähe vom *Hotel Excelsior*?“ wären beispielsweise die beiden kursiv gesetzten Wortgruppen solche benannten Entitäten. Als erster Schritt wird man also ermitteln, ob die Wörter in einer Äußerung zu einer benannten Entität gehören oder nicht. Vor ein bis zwei Jahrzehnten würde man für eine solche Aufgabe beispielsweise HMM (Hidden Markov-Modelle), das damalige „Arbeitspferd“ des maschinellen Lernens, genutzt haben. Sie kamen auch in der Spracherkennung zum Einsatz, bevor man DNN hatte. Mittlerweile hat sich ein anderes mathematisches Modell durchgesetzt, das sich besonders gut für solche zuweisenden oder einordnenden Aufgaben eignet, also in unserem Beispiel die Erkennung der Zugehörigkeit von Elementen zu einer bestimmten Gruppe.

Dieses Modell heißt Conditional Random Fields (CRFs). Im Gegensatz zur zuvor behandelten Aufgabe (Spracherkennung), kann beim Verstehen natürlicher Sprache (NLU) gewartet werden, bis die gesamte Äußerung vorliegt. Die Vorteile, die man durch das gleichzeitige Vorliegen aller Wörter hat, überwiegen den Nachteil der kleinen Zeitverzögerung durch den NLU-Verarbeitungsschritt, denn dieser ist im Vergleich mit der automatischen Spracherkennung (Automatic Speech Recognition, ASR) sehr schnell. CRF-Modelle übertreffen HMM bei solchen Aufgaben deutlich.

Sie haben jedoch eine Schwachstelle. Sie brauchen ein wenig Hilfe bei der Festlegung, nach was sie in den Eingabedaten (der Wortkette) suchen sollen. Sollen Sie nur die Wortbedeutung berücksichtigen oder auch die (grammatische) Wortart? Die Nachbarn zur Linken und zur Rechten, und wie viele davon? Diese so genannte Merkmalsauswahl aber ist etwas, das neuronale Netze gut können: Sie entwickeln sich und lernen, welches die nützlichsten Merkmale sind.

Warum also nicht CRFs (Conditional Random Fields) mit neuronalen Netzen kombinieren? Genau das tut unser NLU-Team bei Nuance.

In diesem Modell – NeuroCRFs – übernehmen die neuronalen Netze die Merkmalsbestimmung und die CRFs den „Rest“.

RNN (rekurrente neuronale Netze) mit ihrer integrierten Erinnerungsfunktion arbeiten besonders gut mit CRFs zusammen. Dies liegt daran, dass sie sich an einen Kontext von variabler Länge „erinnern“ können, während wir bei anderen neuronalen Netzen gezwungen wären, die Länge des Kontextfensters willkürlich festzulegen. Nach einigen weiteren Kniffen und Optimierungen entstand daraus ein Modell, das die bereits recht gute Ausgangsgenauigkeit der CRFs um über 10 Prozent übertrifft. (Zwei meiner Kollegen haben hierüber auf der ASRU im Dezember 2015 [wesentlich] detaillierter berichtet – Marc-Antoine Rondeau und Yi Su: „RECENT IMPROVEMENTS TO NEUROCRFS FOR NAMED ENTITY RECOGNITION“, in Proc. of ASRU 2015, pp. 390-396.)

Kurz gesagt

Erstens: Modelle für das maschinelle Lernen, insbesondere DNNs, sind zwar für viele Aufgaben geeignet, aber nicht jedes DNN ist in allen Fällen gleich gut. Daher muss Etliches an harter Forschungsarbeit geleistet werden, um herauszufinden, welches Netz man für welchen „Fisch“ – in diesem Fall: welche Aufgabe – am besten einsetzt.

Zweitens: Als Endanwender können Sie nicht wissen, mit welcher Technologie Sie sprechen. Wenn Sie im Laufe der Zeit bereits mehrere Anwendungen mit ASR- und NLU-Technologie erlebt haben, kann es gut sein, dass dies verschiedene Generationen waren. Der einzige Unterschied, den Sie dabei bemerkt haben dürften, war, dass die Systeme immer genauer und leistungsstärker wurden. Und von der Wissenschaft des maschinellen Lernens, wo immer mehr Daten in Wissen umgewandelt werden, können wir noch Einiges erwarten.



Was ist „tief“ am „Tiefen Lernen“?

Von Nils Lenke, Senior Director Corporate Research,
Nuance Communications

Unsere Sprache steckt voller Metaphern – im Alltagsleben genauso wie in der Wissenschaft. Wenn wir beispielweise einen Künstler als „Ikone“ seiner Kunstrichtung bezeichnen, bedeutet dies nicht wörtlich, dass dieser Mensch ein Abbild eines Heiligen ist, sondern wir möchten ausdrücken, wie anbetungswürdig viele seiner Fans ihn finden. Auch um den Begriff des „tiefen Lernens“ (Deep Learning) haben sich einige Metaphern entwickelt. Gibt es wirklich Maschinen, die komplex genug sind, eine menschenähnliche Lernfähigkeit zu entwickeln?

Wie kann etwas so Abstraktes wie das Lernen „tief“ (oder auch „flach“) sein? Diesen Fragen werde ich in diesem Beitrag nachgehen.

In beliebten Filmen wie „Ex Machina“, „Her“ und „The Imitation Game“ wird das Thema des „tiefen Lernens“ mit großer Faszination thematisiert. Ich habe spaßeshalber eine bekannte Suchmaschine mit den Worten „nutzt tiefes Lernen für“ gefüttert. Obwohl ich selbst diese Suche in englischer Sprache ausgeführt habe, wird es auch im Deutschen kaum anders sein: Stichworte wie „tiefes Lernen“ oder „maschinelles Lernen“ bringen hunderte von Suchergebnissen für so verschiedene Anwendungsgebiete wie die Nutzung von Satellitenbildern für geschäftliche Entscheidungen, die Unterscheidung von Erkrankungsstadien anhand von medizinischen Alltagsdaten, die Erschließung von Filmkritiken und die natürliche Sprache. Anscheinend lässt sich maschinelles Lernen auf alle diese verschiedenen Gebiete anwenden. Doch zunächst stellt sich die Frage, woher der Begriff vom „tiefen Lernen“ kommt und was er eigentlich bedeutet. Meiner Meinung nach hat es etwas mit Metaphern zu tun.

Man kann sich in das Thema *vertiefen*, wie Metaphern als *Werkzeug* genutzt werden, um dem Leser eine *glänzende Idee vor Augen zu führen* – sie aber auch auf *Irrwege lenken* und unseren *Blick auf Dinge verstellen* können.

Metaphern dienen nicht dem Zweck, Leser zu beeindrucken, sondern **der Sparsamkeit: man nutzt bereits bekannte Wörter** und recycelt sie im neuen Kontext.

Metaphern sind überall. Ich habe es geschafft, allein im letzten Satz sieben davon unterzubringen (siehe die kursiven Hervorhebungen). Eine Metapher besteht also darin, ein Wort oder einen Begriff aus einem Themenbereich auf einen ganz anderen Sachverhalt anzuwenden. Anstatt jemanden als „Ikone“ zu bezeichnen, könnte ich auch sagen, dass er oder sie eine einflussreiche und verehrte Künstlerpersönlichkeit ist, oder eine neue Bezeichnung mit der gleichen Bedeutung erfinden. Die erste Alternative ist aber umständlich und mit der zweiten vermehre ich die Anzahl der Wörter unserer Sprache. Metaphern dienen anscheinend nicht dem Zweck, Zuhörer oder Leser zu beeindrucken, sondern der Sparsamkeit: Man nutzt bereits bekannte Wörter in einem neuen Kontext und recycelt gleich auch sämtliche damit verbundenen Assoziationen.

Teil I: Das „Lernen“ beim „tiefen Lernen“

Gemäß Definition der meisten Wörterbücher ist „lernen“ – der Duden definiert es als „sich Wissen, Kenntnisse aneignen; seinem Gedächtnis einprägen, Fertigkeiten erwerben“ usw. – etwas, das Menschen tun. Wenn wir also das Wort „lernen“ auf Tiere, Substanzen oder gar Gerätesysteme im Internet der Dinge anwenden, sprechen wir bereits in Metaphern, denn wir sprechen von einem menschlichen Verhalten, zu dem ein Bewusstsein gehört. Vielleicht haben Sie schon von Formgedächtnislegierungen gehört. Dinge aus diesen Metallen haben die interessante Eigenschaft, nach einer Verformung in ihre Ausgangsform zurückzukehren, wenn sie erwärmt werden. Die Metapher für dieses Verhalten ist naheliegend (und hilft beim Einprägen des Begriffs). Die Wikipedia setzt dieses Mittel bei der [Beschreibung von Formgedächtnislegierungen](#) ein, wobei die Metaphern durch Anführungszeichen kenntlich gemacht sind:

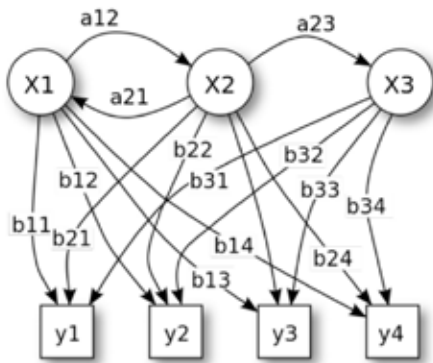
Formgedächtnislegierungen können sich (...) an zwei Formen (...) „erinnern“. Damit das Bauelement beim Abkühlen seine definierte Form wieder einnimmt, muss es (...) „trainiert“ werden.

Vermutlich wird niemand diese Formulierungen wörtlich nehmen und glauben, dass die Metallatome kleine Gehirne besitzen, die sich „erinnern“ oder die „trainiert“ werden können. Wie aber ist es mit dem „maschinellen Lernen“ von Computerprogrammen? Ist das auch rein metaphorisches „Lernen“? Oder sind sie komplex genug für eine menschenähnliche Lernfähigkeit? Und warum stellen wir uns diese Frage überhaupt, anders als bei der Legierung?

Ein Grund ist natürlich, dass Computer komplexer sind und viele Menschen sie nicht gut genug verstehen. Schon ganz zu Beginn sprach man über Computer in Metaphern: In den 1950er Jahren erfanden die Medien den Begriff „Elektronengehirn“. Die Science Fiction bescherte uns, völlig unbelastet von Trivialitäten wie der technischen Machbarkeit, einen bunten Strauß von „denkenden“ Maschinen und Robotern, die sich fest in der Populärkultur etabliert haben. In der westlichen Kultur trafen sie auf alte Legenden vom „künstlichen Leben“; vom Golem aus Lehm über den Homunculus der mittelalterlichen Alchemisten bis zum Monster aus Mary Shelleys „Frankenstein“.

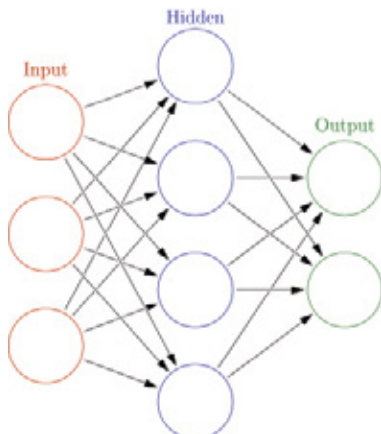
Bei der Entscheidung, ob maschinelles Lernen (ML) nun ein Lernen mit oder ohne Anführungszeichen ist, hilft ein kurzer Überblick über die Thematik.

Beginnen wir mit einem mathematischen Modell, das viele Jahre lang das Rückgrat vieler ML-Systeme war: dem Hidden Markov Model (HMM).



Das obige, noch nicht sehr komplexe Bild besteht aus Zuständen (x) und möglichen Übergängen zwischen ihnen (a), die über bestimmte Wahrscheinlichkeiten und Zuordnungen (b) zu Eingabezuständen (y) verfügen. Das Modell „lernt“ die Wahrscheinlichkeiten durch „Training“ an vielen Beispielen dessen, was das Modell wiedergeben soll, beispielsweise Wörter oder ihre akustischen Bausteine, die so genannten Phoneme. Wir sind uns vermutlich einig, dass „lernen“ hier metaphorisch gemeint ist, denn diese Modelle unterscheiden sich nicht grundlegend von den Atomen einer Formgedächtnislierung.

Vor einigen Jahren wurden in der ML-Forschung jedoch die HMM gegen einen anderen Modelltyp ausgetauscht, der zunächst in den 1990ern an Popularität gewann. Nach einer gewissen Zeit verschwand dieses Modell jedoch fast gänzlich, hat sich aber mittlerweile wieder durchgesetzt (wir werden später noch darauf eingehen). Der Name des Modells – neuronales Netz (NN) – war an sich schon problematisch. Wie man im nächsten Bild sehen kann, besteht es aus mehreren Ebenen von Knoten. Diese Knoten sind den Neuronen unseres Gehirns nachempfunden: Der Input kommt von links herein, so wie auch Neuronen (elektrische) Reize über ihre Dendriten erhalten. Dann passiert eine Verarbeitung, und der entsprechende Output geht, ähnlich wie bei den Axonen der Neuronen, nach rechts hinaus, wo er der Input für die nächste Ebene wird.



Die Verarbeitung im Inneren des Neurons ist in der Regel eher trivial, beispielsweise die Ermittlung des Maximums aus den Inputs oder eine Summenbildung. Um dieses Prinzip für eine ML-Aufgabe, beispielsweise die Bilderkennung, einsetzen zu können, weist man jedem Input-Knoten ein Pixel z. B. eines Schwarzweißbildes zu, und jedem Output-Knoten eine Kategorie eines Objekts, das erkannt werden soll („Baum“, „Kuh“ usw.). Wie bei den HMM wird auch dieses Modell mit Bildern und bekannten richtigen Ergebnissen trainiert (das Bild zeigt eine Kuh), indem man die Input- und Output-Knoten auf die entsprechenden Werte setzt. Dann wird eine sogenannte Rückpropagierung angewendet, die von rechts nach links arbeitet (im Arbeitsmodus arbeitet das Modell ansonsten von links nach rechts) und bei der den Bögen zugewiesene Wahrscheinlichkeiten so angepasst werden, dass bei der Berechnung von links nach rechts der entsprechende Input zum korrekten Output führen würde.

Wie man sieht, ist das gesamte Modell nicht viel komplexer als HMM, jedenfalls nicht so sehr, dass man plötzlich davon sprechen könnte, dass neuronale Netze (NN) denken oder lernen. Reale Modelle haben natürlich mehr Knoten (mehrere tausend), aber dennoch bestehen grundsätzliche Unterschiede zu den Neuronen in echten Gehirnen: letztere haben immer noch viel mehr Neuronen, arbeiten analog und kombinieren elektrische mit chemischen und sogar genetischen Effekten; Eigenschaften wie z. B. das Bewusstsein sind weitgehend unerforscht. Meiner Meinung nach sind NN einem Gehirn nicht ähnlicher (oder unähnlicher) als HMM. Da aber das Wort „neuronal“ im Namen vorkommt (alternative Ideen wie „Wahrnehmungsnetzwerk“ oder Ähnliches sind nicht viel besser), trägt das Modell die Last einer dieser Metaphern: Man denkt, ein „Elektronengehirn“, das künstliche „Neuronen“ enthält, müsse auch „lernen“ können.

Teil II: Das „Tiefe“ beim „tiefen Lernen“

Aber da ist ja auch noch der Begriff des „Tiefen“ im „tiefen Lernen“, was ist damit? Die wörtliche Bedeutung bezieht sich auf die räumliche Ausdehnung. Ein See oder das Meer kann tief sein. Und dann gibt es noch viele metaphorische Bedeutungen, wenn zum Beispiel „tiefblau“ für eine sehr intensive oder dunkle Farbvariante steht. Und auch in Kombination mit dem Denken ist die Metapher beliebt. Gedanken sind „tiefgründig“ oder man hat ein „tiefes Verständnis“ oder „vertieft“ sich in etwas.

Im Roman „Per Anhalter durch die Galaxis“ nannte Douglas Adams einen Computer „Deep Thought“. (Der kann nur eine einzige Frage beantworten und braucht dafür 7,5 Millionen Jahre. Die Antwort ist bekanntlich „42“.) Später lieh sich ein Student und zukünftiger IBM-Mitarbeiter diesen Namen für einen Schachcomputer, der 1996 Berühmtheit erlangte, weil er Weltmeister Gary Kasparov schlug. Und dann änderte ein findiger Marketing-Experte bei IBM den Namen in

„Deep Blue“ – hier haben wir die gesamte metaphorische Entwicklung vom tiefen Meer über „tiefblau“ zum „tiefen Denken“ in nur zwei Wörtern (das Logo von IBM ist bekanntlich blau). Eigentlich ist auch ein schachspielender Computer nicht besonders „tief“ – der Algorithmus ist in seiner schlichten Brutalität eher flach: Deep Blue war erfolgreich, weil er eine große Rechenleistung für die Berechnung möglicher Züge hatte und seine Chips speziell auf die Auswertung von Möglichkeiten ausgelegt waren. Aber die „tiefe“ Metapher hatte sich festgesetzt. (Nur am Rande: Konkurrent „Deep Fritz“ ist auch heute noch als Schachcomputer auf dem Markt.) Eine weitere „tiefe“ Berühmtheit ist DeepQA: Hier handelt es sich um Textrecherche auf der Basis von Wikipedia-Artikeln und Ontologien, die so computerisiert wurde, dass sie im Spiel „Jeopardy“ ihre menschlichen Herausforderer schlagen konnte. Bekannt wurde das System als „Watson“.

Begonnen hatte diese Entwicklung vor Jahren mit der Entscheidung der Wissenschaftler, die mittlere „flache“, verdeckte Schicht eines neuronalen Netzes weiter auszubauen, die Knoten komplexer zu gestalten und das Ergebnis „tiefes neuronales Netzwerk“ (DNN) oder auch „deep belief network“ zu nennen. Liegt es nicht nahe, dass technisch nicht versierte Menschen bei diesen Begriffen zu den bekannten metaphorischen Assoziationen neigen?

Ok. Nun haben wir den Begriff ausreichend entzaubert und wissen, dass wir es mit einer mathematischen Modellierung zu tun haben, die sich nicht vollständig von HMM unterscheidet. Und dass diese Systeme nur im metaphorischen Sinne lernen. Aber deshalb müssen wir das Konzept nicht geringschätzen. Ganz im Gegenteil.

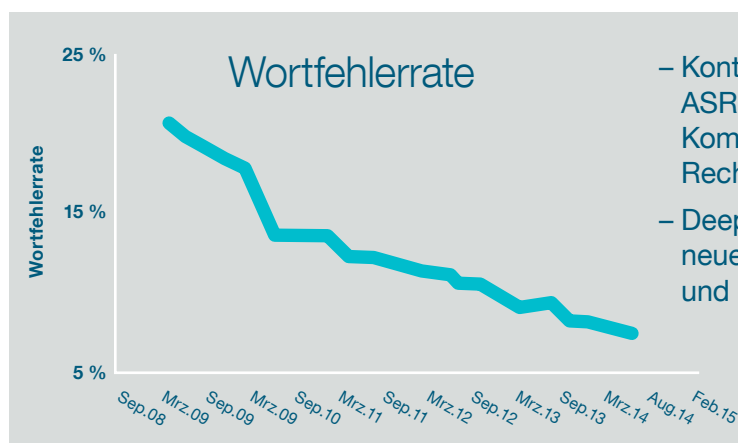
Wir bei Nuance konnten mit DNNs die Genauigkeit unserer automatischen Spracherkennung (ASR) verbessern und die Fehlerrate verringern. Ich spreche von der Technologie, die das Herzstück unserer Cloud-basierten Angebote und von Dragon® NaturallySpeaking darstellt, einem System, das mittlerweile in die 13. Generation geht. In den letzten 20 Jahren konnten wir mit jeder neuen Version die Fehlerraten weiter senken. Mit dem HMM-Modell war das zuletzt immer schwieriger geworden, denn nach jahrzehntelanger Optimierung waren seine Möglichkeiten ausgereizt.

DNNs waren in den letzten Jahren bei vielen unserer Produkte die **größte Innovationskraft**.

DNNs haben also nicht nur sofort geringere Fehlerraten ermöglicht, sondern weisen auch ein weites Feld noch unerprobter Möglichkeiten auf – verschiedene Topologien, Anzahl der Schichten und Knoten, Strukturierung und Training der Knoten usw. – sodass genügend vielversprechendes Potenzial für die nächsten Jahre vorhanden ist. Auch in der Sprachsynthese verbessern DNNs die Verknüpfung zwischen den linguistischen Merkmalen des zu synthetisierenden Texts und den akustischen Parametern der auszugebenden Sprache, beispielsweise der Prosodie. In der Stimmbiometrie tragen sie zur Verbesserung der Sprechererkennung bei. Es ist nicht übertrieben zu sagen, dass DNNs in den letzten Jahren bei vielen unserer Produkte die größte Innovationskraft waren.

Ich sagte oben zwar, DNNs seien nicht komplex (in dem Sinne, dass sie schwerlich ein Bewusstsein oder echte Intelligenz beherbergen), damit meinte ich aber nicht, dass sie leicht zu erfinden oder zu nutzen wären. Ganz im Gegenteil. Wie oben erwähnt, gab es neuronale Netze bereits in den 1990ern, aber damals standen ihrem Erfolg zwei Probleme im Weg.

Erstens: Wenn man sie auf großen Datensätzen trainieren wollte und die Anzahl Knoten und Schichten eine nennenswerte war, dauerte das Training sehr lang – zu lang für die damals verfügbare Hardware. Zweitens: Das Training ergab oft ein Modell, das zwar besser war als „benachbarte“ andere Modelle, aber wenn man den globalen Suchraum ansah, hätte es noch ganz andere und viel bessere geben müssen. Ob das Training in einem solchen „lokalen Optimum“ stecken blieb oder doch das globale Optimum fand, hing von Zufallsfaktoren im Frühstadium des Trainingsprozesses ab.



- Kontinuierliche Senkung der ASR-Wortfehlerrate ~18 %/Jahr: Kombination aus Algorithmen, Daten und Rechentechnik
- Deep Learning (Tiefes Lernen) sorgt für neue Leistungssteigerungen bei der ASR und Bedeutungsextraktion.



Der Durchbruch von DNNs war erst möglich, als beide Probleme von Pionieren wie Geoffrey Hinton und Yoshua Bengio gelöst wurden. Verbesserungen bei der Hardware waren sicher hilfreich gegen das erste Problem, aber weiter kam man durch Ideen zur parallelen Abarbeitung und zur Verwendung grafischer Recheneinheiten (GPUs, also spezieller Chips, die ursprünglich für die Computergrafik entwickelt worden waren). Schon davor war das Problem lokaler Optima durch das Prinzip des Pretrainings gelöst worden, also einen Verarbeitungsschritt, in dem das Modell in einen Zustand versetzt wurde, in dem es wahrscheinlicher (und schneller) ein globales Optimum erreichte.

Was bringt die Zukunft?

Nicht nur sind diese Probleme jetzt gelöst und NN funktionieren im Allgemeinen, sondern es eröffnen sich auch Bereiche für weitere Forschungen und Verbesserungen. Die in der Spielebranche eingesetzten GPUs werden immer leistungsstärker, und davon profitieren auch DNNs. Die Beschleunigung des Trainings ist nicht nur für praktische Anwendungen hilfreich, sondern trägt indirekt auch zum Fortschritt bei der Algorithmusentwicklung bei: Solange DNN-Trainings an großen Datenmengen mehrere Wochen oder Monate dauerten (wie noch vor einigen Jahren), waren Experimente teuer und der Fortschritt langsam. Da man solche Trainings mittlerweile in Tagen oder gar Stunden durchführen kann, lassen sich auch neue Ideen viel einfacher testen.

Dennoch wissen ich und meine Forscherkollegen, dass noch viel Arbeit auf uns wartet. So ist es beispielsweise schwierig, GPUs für alle Trainings Schritte eines DNN zu nutzen, weil das Netzwerk stark verflochten ist. Da der Output eines „Neurons“ potenziell von vielen anderen Neuronen und den Input-Daten abhängt und das Training keine rein lokale Angelegenheit (und dadurch parallelisierbar) ist, müssen große Datenmengen zwischen Rechenknoten übertragen werden, was den Zeitgewinn durch die GPUs wieder zunichtemachen kann. Wie werden wir das lösen? Außerdem wurde zu Beginn des Einsatzes von DNNs in der Spracherkennung das sprecherunabhängige Modell mit einer großen Menge Daten trainiert, die nahezu alle Dialekte und persönlichen Sprechstile wiedergeben mussten. Die meisten praktischen Systeme setzen jedoch eine zweite, sprecherabhängige Trainingsmethode ein, die das Basismodell an den jeweiligen Sprecher anpasst. Je nachdem, ob Sie nur wenige Sekunden oder viele Stunden Sprachproben zur Verfügung haben, wurden verschiedene Methoden genutzt. Da all dies für HMM-Basismodelle entwickelt wurde, muss es jetzt für DNNs angepasst werden.

Und vieles mehr.

Es wartet also noch viel Arbeit im Bereich DNNs auf uns, aber eine sehr aufregende. Auch wenn wir uns dabei nicht von den „tiefen“ Metaphern begeistern lassen.

Quellen:

- deep-learning-reduces-ASR-error-rate © artificial-neural-network CC BY Wikipedia
- Hidden Markov Model, CC BY Wikipedia

Die Zukunft des Gesundheitswesens: wie klinische Umgebungsintelligenz zu **besserer Versorgung** beiträgt

Von Joe Petro, Senior Vice President,
Forschung und Entwicklung, Nuance Communications

Täglich vertrauen wir Menschen auf Technologien, um mit anderen Menschen oder mit Informationen und Inhalten in Verbindung zu bleiben. So segensreich moderne Technologie in der Medizin auch ist, kann sie doch zu einem Fluch werden, wenn sie den Arzt zu unnatürlichen Arbeitsabläufen zwingt. Sie sollte dazu da sein, Ärzte und ihre Arbeit zu unterstützen – nicht umgekehrt. Wir stellen uns für die Zukunft eine klinische Umgebungsintelligenz vor, die die menschliche Intelligenz vervielfacht und vorhandene Ressourcen nutzt und Ärzten zeitraubende Routinearbeiten, wie klinische Dokumentation abnimmt, damit diese mehr Zeit mit ihren Patienten verbringen können.

Am Anfang eines Jahres liegt es nahe, sich Gedanken über die Zukunft des Gesundheitswesens zu machen. Ob wir Neuerungen für 2017 erwarten oder größere Prognosen für die nächsten Jahre anstellen, es ist offensichtlich, dass große technologische Fortschritte uns echte kurz- und langfristige Verbesserungen in der Gesundheitsversorgung bringen werden.

Zu unseren Zielen gehört es, dass sowohl Ärzte als auch Patienten davon profitieren. Elektronische Krankenakten und ähnliche Technologien werden immer hilfreicher, sind aber auch dafür verantwortlich, dass Ärzte mehr Zeit mit klinischer Dokumentation verbringen. Dadurch wird der Arbeitstag noch komplizierter und ermüdender, bis hin zum Burnout. Wir brauchen Lösungen, die dem Arzt den Rücken für seine eigentliche Aufgabe freihalten, wovon auch die Patienten profitieren werden.



Wir bei Nuance wollen dazu in Zukunft beitragen, indem wir die menschliche Intelligenz vervielfachen und unsere Ressourcen besser nutzen. Wir wollen eine uns umgebende klinische Intelligenz, die den Menschen (Arzt) proaktiv unterstützt und sich dabei unaufdringlich im Hintergrund hält. Auf diese Weise wird der Dokumentationsprozess vereinfacht und der persönliche Kontakt zwischen Ärzten und Patienten kann konzentriert und ungestört ablaufen. Wir denken, dass diese um Technologie und Inhalte verstärkte menschliche Intelligenz intelligente und fundierte Entscheidungen ermöglicht, die die Gesundheitsversorgung verändern und von denen Patienten und das gesamte Gesundheitswesen profitieren.

Indem wir Ärzte entlasten und ihnen helfen, ihre Möglichkeiten noch besser auszuschöpfen, erreichen wir unser Ziel, dass Ärzte zu ihrer wichtigen Arbeit zurückkehren können: der Versorgung ihrer Patienten.

„Wir stellen uns für die Zukunft eine klinische Umgebungsintelligenz vor, die die menschliche Intelligenz verstärkt und vorhandene Ressourcen nutzt und Ärzten zeitraubende Routinearbeiten, wie klinische Dokumentation abnimmt, damit diese mehr Zeit mit ihren Patienten verbringen können.“

Joe Petro, Senior Vice President
Forschung und Entwicklung, Nuance

Besuchen Sie nuance.de/gesundheit und lesen Sie mehr über Ressourcen, Trends und Erkenntnisse rund um die Zusammenhänge zwischen zufriedenen Ärzten, produktiven Einrichtungen und positiven Ergebnissen für die Patienten.

Über Nuance Communications, Inc.

Nuance Communications, Inc. ist mit der größten Palette von Sprach- und Produktivitätslösungen für Unternehmen und Privatkunden ein weltweit führender Anbieter. Seine Technologien, Anwendungen und Services sorgen für einen deutlich höheren Benutzerkomfort und bieten neue arbeitssparende Methoden zur Interaktion mit Geräten und Systemen. Jeden Tag vertrauen Millionen Anwender und Tausende Unternehmen auf die bewährten Anwendungen von Nuance. Weitere Information erhalten Sie unter www.nuance.de/gesundheit oder unter +49 89 4587 3529. Lesen Sie unseren Mediziner-Blog oder folgen Sie uns auf twitter.com/voice4health
