

Zukunftsthema KI im Gesundheitswesen: Interview mit Tiplu ML-Netz Product Owner Lennart Grosser

Seit einiger Zeit ist Tiplu Mitglied im ARIC e. V.: Der Verein hat es sich zur Aufgabe gemacht, ein interdisziplinäres, anwendungsnahes Knowhow-Center für KI in der Metropolregion Hamburg zu etablieren und damit zum Bindeglied zwischen anwendungsorientierter Forschung und Praxis im Bereich der Künstlichen Intelligenz zu werden.

Da auch wir bei Tiplu fest von dem großen Potenzial künstlicher Intelligenz überzeugt sind, möchten wir gemeinsam mit dem ARIC Aufmerksamkeit und Verständnis für die spannenden und vielfältigen Möglichkeiten schaffen, die KI speziell im medizinischen Bereich bietet. Wir werden zukünftig unsere Ideen und Visionen für das Gesundheitswesen im ARIC einbringen und mit vielen engagierten, klugen Köpfen diskutieren können.


Für einen ersten Einblick in die Themen, die uns bei Tiplu beschäftigen, hat Sabrina Pohlmann vom ARIC ein Interview mit unserem Kollegen Lennart Grosser geführt. Es ging um die Frage, was wir bei Tiplu grundsätzlich machen, um technische Basics und Details unserer Produkte sowie aktuelle und zukünftige Möglichkeiten unseres ML-Netzes und das Potenzial der Nutzbarmachung der darüber zur Verfügung stehenden Daten. Lennart Grosser arbeitet seit vier Jahren bei Tiplu in Berlin. Er ist Product Owner für das Tiplu Machine Learning-Netz und Entwickler im Machine Learning Team.

Euer Kernprodukt heißt MOMO. Was ist das eigentlich?

MOMO ist eine Software, die auf Erlössicherung im Krankenhaus ausgelegt ist. Das heißt: Ein Patient wird im Krankenhaus behandelt und muss nach der Entlassung von der Krankenkasse abgerechnet werden. Es gibt Indikatoren, die ihn für eine Pauschalgruppe klassifizieren, für die das Krankenhaus einen bestimmten Geldbetrag bekommt. MOMO hilft dabei, die Indikatoren für die Pauschalgruppe zu identifizieren und dadurch die Patientenfälle vollständig und korrekt abzurechnen.

Wie ist der herkömmliche Prozess?

Die Kodierfachkräfte, die für die Krankenhausabrechnung zuständig sind, haben oft ein hohes Arbeitspensum und zusätzlich Zeitdruck. Deswegen gibt es immer wieder unvollständige Kodierungen, sodass das Krankenhaus nicht den vollständigen Erlös für die erbrachten Leistungen bekommt. MOMO hilft den Fachkräften dabei, die Indikatoren zu vervollständigen: Es gibt sogenannte OPS- und ICD-Kodes, aus deren Kombination sich die Fallpauschale ergibt. Normalerweise werden diese Codes vom Fachpersonal in der Patientenakte manuell identifiziert. MOMO nimmt die Arbeit ab und macht Vorschläge zu möglichen Codes.

 ICD und OPS-Kodes werden im medizinischen System genutzt, um Diagnosen und Behandlungen eindeutig zu klassifizieren und zu bezeichnen.

Worauf basieren die Vorschläge?

Die Vorschläge für mögliche Kodierungen basieren auf der umfangreichen Dokumentation, die während der Behandlung erstellt wird, also auf schriftlicher Dokumentation, wie beispielsweise diagnostischen Befunden, OP-Berichten, Arztbriefen und der Visitedokumentation, sowie den Laborwerten und Messbefunden oder auch dem Medikationsverlauf.

Hier kommt KI zum Einsatz. MOMO hat zuvor aus einer großen Datenbasis Muster gelernt, also, welche Formulierungen, Textstruktur oder Wortwahl mit welchen Codes im Zusammenhang stehen und kann diese Muster dann in der Dokumentation wiedererkennen.

Wie funktioniert MOMO technisch genau?

MOMO vervollständigt die Kodierung auf verschiedene Arten. Die Machine Learning-Komponente, die sich vor allem auf die Erkennung der Codes fokussiert, ist hier besonders wichtig. Wir haben zwei ML-Modelle. Das eine ist ein Sprachmodell, das andere nutzt strukturierte Daten, um Vorschläge zu generieren. Und es gibt auch noch die regelbasierte Erkennung von Codes.

Was ist für unsere technisch versierten AI-Nerds besonders interessant an dem Modell?

Besonders interessant daran ist, wie wir es entwickeln: Wir nutzen dafür unser Machine Learning-Netz – das ist eine verteilte Daten- und Entwicklungsplattform, die wir aufgebaut haben. Konkret besteht das ML-Netz aus Servern, die in Partnerkrankenhäusern eingerichtet sind und auf denen wir die Daten des Krankenhauses verarbeiten dürfen. Jeder von den Servern ist mit einem zentralen Server bei Tiplu verbunden. Daraus ergibt sich ein Netzwerk mit Datenbeständen aus mittlerweile rund 140 Krankenhäusern.

Wir laden zu keinem Zeitpunkt sensible Informationen herunter, können aber ein verteiltes Machine Learning Training – das sogenannte Föderierte Lernen – durchführen. Dadurch können wir die Daten aller Krankenhäuser verarbeiten, ohne sie zentral sammeln zu müssen. Statt die Daten herunterzuladen, laden wir unser Machine Learning Modell zu den Daten ins Krankenhaus hoch, verarbeiten dort die Daten und laden uns dann das veränderte Modell herunter. Das ist technologisch sehr cool!

Besonders ist außerdem, dass wir die Daten von verschiedenen Krankenhäusern in der gleichen Datenstruktur vereinen. Denn: Grundsätzlich liegen die Daten in Krankenhäusern sehr unterschiedlich vor. Einen OP-Bericht gibt es überall, aber den muss man erstmal finden und als OP-Bericht identifizieren. Die von Tiplu entwickelte elektronische Patientenakte erlaubt es uns, deutlich effizienter mit den Daten zu arbeiten, weil sie für

alle gängigen Krankenhausinformationssysteme immer im gleichen Format vorliegen.



Föderiertes Lernen ist eine Machine Learning-Methode, bei der dezentral vorgegangen wird. Statt, wie meistens üblich, die Daten zentral zu sammeln, wird das Machine Learning-Modell auf verschiedene Geräte geladen, auf denen sich die Daten befinden. Dieses Vorgehen bietet Vorteile für den Datenschutz.

Jetzt wo das Netzwerk schon einmal da ist: Was kann man noch mit den Daten machen?

Neben der Machine Learning-Entwicklung bietet das Netz zum Beispiel die Möglichkeit, Datenanalysen durchzuführen. Man kann Datenstatistiken zusammenführen, zum Beispiel: Wie häufig kommt eine bestimmte Erkrankung über alle Krankenhäuser hinweg vor? Hier haben wir aktuell mehrere Kooperationen laufen.

Einerseits für eine Studie, in der wir zeigen wollen, dass die Daten unseres Netzwerks repräsentativ sind. Dafür vergleichen den Datenbestand des ML-Netzes mit einem Datenbestand des Bundesamts für Statistik und zeigen, dass die Daten der gleichen Verteilung folgen. Dann gibt es noch die PAIRS-Kooperation, die u.a. motiviert durch die Coronapandemie entstanden ist. Hier geht es darum, eine Epidemiefrüherkennung zu entwickeln. Das ist ein Forschungsprojekt mit vielen verschiedenen Partnern. Unser Machine Learning-Netz soll dabei helfen, frühzeitig bestimmte Muster zu erkennen.

Wenn wir mal ins Träumen kommen: Was könnte man theoretisch mit den Daten noch alles machen?

Eine mögliche Nutzung wäre ein Live-Betrachtung der Daten der an das ML-Netz angebotenen Krankenhäuser. Dadurch könnten zum Beispiel zunehmende Erkrankungen wie COVID-19 möglicherweise frühzeitig erkannt werden. Grundsätzlich wäre die Aktualität jeglicher Datenauswertung oder Machine Learning Entwicklungen dadurch viel höher.

So könnte man zum Beispiel Trends identifizieren. Wenn das technisch möglich ist, warum wird es nicht gemacht?

Das müsste gut konzipiert, geplant und mit allen Beteiligten abgestimmt werden. Der notwendige Aufwand für eine solche Lösung ist hoch und es braucht einen gewichtigen Use-Case wie bspw. des Public Health Surveillance als Antrieb.

Seid ihr häufiger mit Widerständen konfrontiert, der sich gegen Data Science oder Machine Learning im medizinischen Bereich richtet?

Die Akzeptanz bei den Krankenhäusern ist grundsätzlich vorhanden. Eine gewisse Skepsis geht zwar schnell in Richtung „Datenschutz – wie machen wir das?“. Aber man muss mit den Leuten arbeiten, über die Themen

sprechen und die Konzepte zu Sicherheit und Datenschutz klarmachen, dann funktioniert es auch. Zum Datenschutz haben wir uns einiges einfallen lassen. Da sind die Leute oft begeistert und wollen mitmachen.

Was habt ihr euch denn neben dem föderierten Modelltraining noch für den Datenschutz einfallen lassen?

Wir pseudonymisieren die auf den ML-Servern im Krankenhaus gespeicherten Falldaten – das bedeutet wir schwärzen oder verändern sensible Inhalte aus der elektronischen Patientenakte, sodass ein Patient nicht mehr identifiziert werden kann. Dafür haben wir einen eigenen Pseudonymisierungsalgorithmus entwickelt, an dem stetig gearbeitet wird, um die Qualität der Pseudonymisierung immer weiter zu erhöhen. Darüber hinaus haben wir verschiedene Maßnahmen getroffen, die den Zugang und die Nutzung des ML-Netzes sichern. Neben allgemeinen Netzwerksicherheitsmechanismen geht es u. a. um die Verhinderung von unbefugtem Herunterladen von Daten. Um anonyme Statistiken oder auch Machine Learning Modelle heruntergeladen zu können, muss ein Freigabeprozess durchlaufen werden, bei dem die gewünschten Dateien geprüft und manuell freigegeben werden müssen.

Wie geht es nun weiter?

Unser Machine-Learning-Netz enthält einen riesigen Datenbestand und wir haben nicht vor, eigennützig darauf sitzen zu bleiben. Wir stellen uns vor, das Potenzial der Daten in gemeinnütziger Art und Weise verfügbar zu machen, beispielsweise für Forschungszwecke – natürlich datenschutzkonform.

Darüber hinaus ist Tiplu immer an Kooperationen interessiert. Wer Interesse hat, darf sich bei uns melden.