

Woche 14: Wie war das nochmal? – Überblick und Ausblick

Skript

Erarbeitet von
Maïke Mayer

Lernziele	1
Inhalt	1
Einstieg	1
Von künstlichen Texten, Bildern und Co.....	2
Diskriminierung durch KI – Teil 2	3
Auf geht's!	3
Quellen	4
Disclaimer	4

Lernziele

- Erinnern relevanter Inhalte der vergangenen Wochen (Dimensionsreduktion, Raum bei Wortembeddings, Loss Function)
- Nachvollziehen der Zusammenhänge zwischen den Inhalten

Inhalt

Einstieg

Wäre es nicht schön, wenn ein KI-gestütztes System für uns einen Aufsatz schreiben würde, damit wir die Zeit für andere Dinge nutzen können? Oder ein Bild für uns malt, weil wir nicht so fit mit Stift oder Pinsel sind? Oder vielleicht sogar für uns komponiert? Verlockend klingt es schon und irgendwie – aber auch noch ein bisschen nach Science Fiction. Es gibt aber durchaus Systeme, die in dieser Richtung schon einiges leisten können, nämlich sogenannte

Generative Modelle. Und mit denen beschäftigen wir uns in der letzten Woche dieses Kurses näher.

Einblendung Illustrationen/Stichwort

Von künstlichen Texten, Bildern und Co.

Wir haben in den letzten Wochen ja schon darüber gesprochen, wie Maschinen Texte oder Bilder verarbeiten. Nun beschäftigen wir uns näher mit Systemen, die Texte, Bilder oder auch Sprache etc. generieren. KI-Systeme oder Neuronale Netzwerke, die für diesen Zweck verwendet werden können, sind Generative Deep Learning Modelle. Das mag insgesamt (noch) futuristisch klingen oder nach starker KI, die an die intellektuellen Fähigkeiten des Menschen heranreicht. Aber lass dich nicht irritieren, auch bei generativen Modellen handelt es sich um schwache Künstliche Intelligenz, die darauf trainiert wurde, eine bestimmte Aufgabe – wie das Schreiben eines Textes – sehr gut zu bewältigen.

Einblendung Illustrationen/Stichwort

Aber bleiben wir nochmal kurz bei Bildern. Für die Generierung von Bildern können beispielsweise Variational Autoencoders verwendet werden. Autoencoder kodieren einen gegebenen Input und reduzieren darüber die Informationen. Das ist ähnlich wie bei der Dimensionsreduktion. Du erinnerst dich vielleicht? Je mehr Features vorliegen, desto mehr Daten werden benötigt, um ein Modell sinnvoll zu trainieren. Dieses Phänomen hast du als „Fluch der Dimensionalität“ schon kennengelernt und du weißt auch, dass die Anzahl der Features beispielsweise über das Zusammenfassen oder auch das Entfernen von redundanten Features verringert werden kann. Ähnlich gehen die Autoencoder auch vor – die wichtigsten Informationen bleiben erhalten. Im nächsten Schritt soll das Netzwerk lernen, aus diesen komprimierten Informationen einen Output zu erstellen, der so nah wie möglich an dem ursprünglichen Input liegt.

Einblendung Illustrationen/Stichwort

Was Variational Autoencoder machen, die von der Architektur her Autoencodern sehr ähnlich sind, erfährst du in dieser Woche. Dabei wirst du auch den Begriff „Latent Space“ kennenlernen, der dem Raum bei Wortembeddings ähnelt. Kurz zur Erinnerung: Um den Kontext von Wörtern bei der Textverarbeitung berücksichtigen zu können, wird für jedes Wort in einem Wörterbuch eine Embedding berechnet, also ein Vektor, der den Standort dieses Wortes im Raum angibt. Wörter, die Ähnliches bedeuten, liegen in dem Vektorraum nah beieinander. Außerdem wird dir übrigens auch der Begriff der Loss Function, oder Verlustfunktion, wieder begegnen. Auch hier nochmal kurz zur Auffrischung: Den Begriff hast du im Rahmen der Optimierung von Modellen im Bereich des Supervised Learnings schon kennengelernt. Der im Training generierte Output wird mit dem gewünschten Output verglichen und die Distanz dazwischen lässt sich über alle Datenobjekte hinweg als Funktion – nämlich als Verlustfunktion – ausdrücken. Ziel des Optimierungsprozesses ist es, die

Verlustfunktion möglichst zu minimieren, da dann der generierte Output und der gewünschte Output sehr ähnlich sind.

Einblendung Illustrationen/Stichwörter

Aber nochmal zurück zu den Generativen Modellen. Du lernst in dieser Woche auch noch sogenannte Generative Adversarial Networks kennen. Das sind im Prinzip zwei konkurrierende Neuronale Netzwerke, die versuchen, einander zu schlagen bzw. zu übertrumpfen – du darfst also gespannt sein. Außerdem schauen wir uns noch Text-to-Image Modelle an, die Bilder basierend auf einer Texteingabe generieren, und auch auf die Generierung von Text selbst werden wir einen Blick werfen.

Einblendung Illustrationen/Stichwort

Bei den Generativen Modellen bleiben wir auch für die praktischen Anwendungsbeispiele in dieser Woche. Wir betrachten ChatGPT näher und betrachten worauf man hier vielleicht achten sollte. Aber auch über die Vorteile von generativen Modellen und über mögliche Herausforderungen, offene Fragen und Risiken werden wir sprechen. Ein mögliches Risiko – nämlich das der Diskriminierung – sollte dir bereits bekannt vorkommen ...

Einblendung Illustrationen/Stichwort

Diskriminierung durch KI – Teil 2

Mit diskriminierenden Entscheidungen von Künstlicher Intelligenz haben wir uns im Rahmen dieses Kurses schon beschäftigt. Das war ziemlich am Anfang im Bereich der Ethik, aber vielleicht sagt dir das Stichwort „Amazon“ noch etwas. In dieser Woche werden wir uns mit dem Themenbereich der Diskriminierung durch Künstliche Intelligenz von der rechtlichen Seite beschäftigen. Wir werfen beispielsweise einen näheren Blick auf Artikel 3 des Grundgesetzes (als verfassungsrechtliche Basis des Gleichbehandlungsgrundsatzes) und auf das Allgemeine Gleichbehandlungsgesetz. Außerdem beschäftigen wir uns mit verschiedenen Typen der Diskriminierung durch KI, wie beispielsweise der Diskriminierung durch vorbelastete Trainingsdaten (Stichwort „Amazon-Beispiel“).

Einblendung Illustrationen/Stichwörter

Quelle [1]

[Auf geht's!](#)

In dieser Woche beschäftigen wir uns also mit Generativen Modellen, die zur Generierung von Bildern oder Texten verwendet werden können. Im Bereich des Rechts beschäftigen wir

uns noch einmal mit dem Thema der Diskriminierung durch Künstliche Intelligenz. Und abschließend werfen wir nochmal einen genaueren Blick auf generative Modelle wie ChatGPT und ihre Vorteile, aber auch die damit verbundenen Herausforderungen und Risiken.

Einblendung Videotitel

Also nicht einfach unreflektiert Texte eines Generativen Modells übernehmen ... Aber jetzt erstmal viel Spaß in Woche 14!

Quellen

- Quelle [1] Hamilton, I. A. (2018, October 11). Amazon hat eine Künstliche Intelligenz für Bewerbungen gebaut, die Frauen diskriminierte. Business Insider.
<https://www.businessinsider.de/wirtschaft/amazon-hat-eine-ki-gebaut-die-frauen-diskriminierte-2018-10/>

Disclaimer

Transkript zu dem Video „Woche 14: Wie war das nochmal? – Überblick und Ausblick“, Dr. Maïke Mayer.

Dieses Transkript wurde im Rahmen des Projekts ai4all des Heine Center for Artificial Intelligence and Data Science (HeiCAD) an der Heinrich-Heine-Universität Düsseldorf unter der Creative Commons Lizenz [CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/) veröffentlicht. Ausgenommen von der Lizenz sind die verwendeten Logos, alle in den Quellen ausgewiesenen Fremdmaterialien sowie alle als Quellen gekennzeichneten Elemente.