Antragsnummer: 100670478
Name: Thomas Davies Zeitraum: Januar- Juni 2024

Thema Promotionsvorhaben:

Prozessierung von heterogenen Datenstrukturen und Sequenzdaten in komplexen KI-System unter Einbeziehung robuster, interpretierbarer KI-Modelle

Inhalt lt. Anforderung

Um eine besseren Zusammenhang zum vorherigen Bericht zu halten, werden stellenweise Formulierungen des vorherigen Berichts zitiert (rote Farbe).

1. Übersicht der erarbeiteten Inhalte

Das Projektvorhaben der Promotion ist beschreiben als "Prozessierung von heterogenen Datenstrukturen und Sequenzdaten in komplexen KI-System unter Einbeziehung robuster, interpretierbarer KI-Modelle". Hierbei handelt es sich, um den Entwicklungsauftrag grundsätzlich ein KI-Modell zu entwerfen, was besagte Bedingungen der Interpretierbarkeit, Robustheit und Eingabedaten erfüllt.

Heterogene Datenstrukturen haben einen besonderen Stellenwert in der Forschung im Bereich künstlicher Intelligenz, da diese schwierig in stabilen KI Modellen lernbar umzusetzen. Ihre besondere Datenstruktur erlaubt es nicht, Modelle für homogene Datenstrukturen anzuwenden. Ein aktuell prominentes Beispiel solcher Datenstrukturen sind Sequenzdaten, wie zum Beispiel Textdaten in Form von Sprache (Natural Language Processing). Hier haben in jüngster Zeit Modelle wie Transformer (ChatGPT) oder mit Attention verknüpfte Long Short Term Memory (LSTM) rekurrente Netze große Fortschritte gemacht. Diese sind allerdings in ihrer Interpretierbarkeit und Erklärbarkeit stark eingeschränkt. Der Auftrag ist es vergleichbare Modelle mit ähnlichen Fähigkeiten in einer robusten und interpretierbaren Ausführung zu entwickeln.

Es wurden inzwischen verschiedene Aspekte des Projektziels untersucht. Zunächst wird ein völlig neuer Ansatz untersucht prototypbasierte Klassifikatoren mit Hilfe von Differentialgleichungen zu entwickeln. Dabei werden grundlegende theoretische Zusammenhänge untersucht, um eine robuste Entwicklung des neuartigen Algorithmus zu ermöglichen. Durch Netzwerken haben sich Kontakte in den Niederlanden ergeben, welche ähnliche Ansätze untersuchen, wodurch kooperative Synergien entstehen.

Weiterhin wird ein heterogenes Datenmodell untersucht, wo Daten in einem Netzwerk verteilt sind und distributiv über verschiedene Agenten an einem Lernziel gearbeitet wird. Dies ist vor allem interessant im Kontext von Datenschutz und sensiblen Daten wie beispielsweise im medizinischen Bereich. Hier wird untersucht mit Hilfe von besonderen Verschlüsselungstechniken die Lernprozessfortschritte im Schlüsselraum weiter zu verarbeiten, sodass sowohl die Daten als auch das Model vor Angriffen geschützt werden können.

Datum: 30.06.2024

2. Aussage zur Zeitplanung

Die Methodik zur Erreichung der genannten Projektziele lässt sich wie folgt beschreiben. Zunächst gibt es neben der noch weiterhin laufenden Entwicklung solcher Modelle als Blackbox Modelle auch bereits existierende Lösungen (ebenfalls Blackbox Modelle). Diese Lösungen werden gründlich analysiert und es werden Ansätze entwickelt, um Mechanismen der Blackboxmodelle in ein interpretierbares Lernschema zu überführen. Diese Ansätze werden dann in Experimenten untersucht und mathematisch analysiert.

Die Umsetzung erfolgt nach Plan. Inzwischen wurden die sogenannten State-Space Modelle untersucht (Mamba), die sich im innersten Kern auf Differentialgleichungssysteme und Dynamische System zurückführen lassen. Über diese Methoden können Zusammenhänge von Zeitreihendaten unabhängig von einer fixen Länge beschrieben werden. Mit Hilfe dieses Wissens lassen sich potenziell interpretierbare Lernschemen entwickeln.

Derzeitig hat sich die Arbeit auf die mathematische Analyse zur sorgfältigen Konstruktion des Algorithmus beschränkt. Eine sorgfältige Analyse der Themen ermöglicht nicht nur, dass der Algorithmus funktioniert, sondern das potenzielle Garantien für die Effektivität des Algorithmus ausgesprochen werden können. Vielmehr ist ein tiefgreifendes Verständnis der bearbeiteten Themen Grundlage für das intuitive konstruieren des Algorithmus.

Für dieses Thema sind die weiteren Schritte nun die mathematische Konstruktion des Algorithmus und dann das Durchführen von Experimenten, um zu erkennen ob der Algorithmus Potenzial hat. Sofern sich die Vermutung durchsetzt, ist zu erwarten dass eine Reihe an Veröffentlichungen zu dem Thema möglich sind, da potenziell ein großes Feld an Anwendungen entstehen kann.

Weiterhin wird noch das Thema der Verschlüsselung und des verteilten Lernen analysiert. Dies erfolgt ebenfalls aus dem Themenfeld der Blackbox Modelle. Auch hier ist das Aufsetzen homomorpher Verschlüsselungsprotokolle die Lösung für sicheres, verteiltes Lernen. Diese Lösung lässt sich auf die interpretierbaren Modelle übertragen. Dafür wurden bislang die mathematischen Grundlagen erarbeitet. Es ist absehbar, dass interpretierbare Modelle einen Vorteil gegenüber den klassischen Blackboxmodellen haben. Diese Analyse wird nun in Zusammenarbeit mit Experten des Bereichs Kryptologie durchgeführt. Ein Abschluss dieses Themas ist zeitnah absehbar und hat ein hohes Potenzial zeitnah eine Veröffentlichung zu produzieren.

3. Weiterführende Bildung

Im aktuelle Halbjahr wurden Lehrveranstaltungen im Bereich Wahrscheinlichkeitsrechnung und Statistik durchgeführt. Darüber hinaus wurden keine weiteren Veranstaltungen zum Aufbau von individuellen Potentialen besucht. Vielmehr wurde an der Lehre der Mathematik Fachgruppe an der Hochschule gezielt mitgewirkt.

Es wurden inzwischen als Gast in Groningen (Niederlande) die Wisci und in Mittweida (Deutschland) die WSOM besucht.

Datum: 30.06.2024

4. Öffentlichkeitsarbeit

In der Untersuchung Prototypen in Hyperboxen darzustellen, wurden Zuarbeiten in Form von Untersuchungen über Hausdorffdistanzmaße durchgeführt und deren Eignung für Weiterentwicklungen des Hyperbox Prototypen Prinzips. Diese Zuarbeiten haben dann für eine Koautorenschaft in der Proceedings of the 15th International Workshop, WSOM+ 2024 geführt mit dem Paper „Hyperbox-GLVQ Based on Min-Max-Neurons“ (https://doi.org/10.1007/978-3-031-67159-3).

Darüber ist geplant in den kommenden Monaten Kryptologiethemen mit KI Themen zu verknüpfen und darin eine Veröffentlichung zu leisten und weiter an einer Lösung für prototypbasiertem Lernen für Zeitreihendaten zu arbeiten.

5. Netzwerke

Bislang gibt es keine Aussicht für eine berufliche Tätigkeit nach der Promotion. Dies lässt sich aber gut erklären, dass ich noch im ersten Jahr meiner Promotion bin. Über die Promotion haben sich Synergien zwischen der *Computational Intelligence* Arbeitsgruppe und der *Kryptologie* Arbeitsgruppe gebildet. Dadurch werden gegenwärtig Themen, die beide Bereiche schneiden, untersucht.

Vor allem die Konferenz in Groningen hat Kontakte ergeben, die womöglich in der Arbeit an den Zeitreihendaten hilfreich sein könnte.

Antragsnummer: 100670478
Name: Thomas Davies Zeitraum: Januar- Juni 2024

_____	_____	_____
Name/Unterschrift Promovend	Name/Unterschrift Prof. HSMW	Name/Unterschrift Externe

Datum: 30.06.2024