Masterarbeit

Automatische Erstellung von Trainingsdaten für die Bewertung von Freitextaufgaben mittels generative pre-trained language models

Ulrich Birkholz

April 2023 – October 2023

Matriculation Id: 2114780
Course of Study: Praktische Informatik

Reviewer:
Professor Dr.-Ing. Torsten Zesch

FernUniversität in Hagen
Center of Advanced Technology for Assisted Learning and Predictive Analytics
Research Professorship Computational Linguistics
58097 Hagen

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit ohne fremde Hilfe selbstständig verfasst und nur die angegebenen Quellen und Hilfsmittel benutzt habe. Ich versichere weiterhin, dass ich diese Arbeit noch keinem anderen Prüfungsgremium vorgelegt habe.

Hagen, 17. April 2023

.....

Ulrich Birkholz

b

Zusammenfassung

very abstract

d

Inhaltsverzeichnis

Kapitel 1

Einleitung

1.1 Hintergrund und Motivation

1.2 Zielsetzung und Forschungsfragen

Kapitel 2

Theoretische Grundlagen

2.1 Generative pre-trained language models

2.1.1 Funktionsweise

2.1.2 Anwendungsbereiche

2.1.3 Vor- und Nachteile

2.2 Trainingsdaten

2.2.1 Definition und Bedeutung

2.2.2 Herausforderungen bei der Erstellung

2.2.3 Automatisierte Erstellung von Trainingsdaten

2.3 Maschinelles Lernen

2.3.1 Grundlagen von ML-Modellen

2.3.2 Supervised Learning

2.3.3 Bewertung von ML-Modellen

2.4 Datenbasis und Aufgabentypen

2.4.1 Beschreibung der Datenbasis

2.4.2 Aufgabentypen und ihre Charakteristiken

2.4.3 Relevanz für automatische Bewertungssysteme

Kapitel 3

Konzeption der Trainingsdaten

3.1 Erstellung der Prompt

3.1.1 Anforderungen an die Prompt

3.1.2 Formulierung der Prompt

3.1.3 Wahl des Models

3.1.4 Wahl der Hyperparameter

3.1.5 Sicherstellen der Datenqualität

3.2 Automatisierte Erstellung der Trainingsdaten

3.2.1 Datenerhebung und -verarbeitung

3.2.2 Datenaufbereitung und -bereinigung

Kapitel 4

Training und Testen des ML-Modells

4.1 Trainieren des ML-Modells

4.1.1 Auswahl des ML-Algorithmus

4.1.2 Trainieren des Modells

4.2 Testen des ML-Modells

4.2.1 Erstellung und Auswahl der Testdaten

4.2.2 Durchführung der Tests

Kapitel 5

Ergebnisse und Evaluation

5.1 Bewertung des ML-Modells

5.1.1 Performanzvergleich mit manuell erstellten Testdaten

KPIs, Vergleich zwischen dem manuell erstellten Modell und dem Modell, das mit automatisch generierten Testdaten erstellt wurde (Wenn zeit ist könnte man auch vollständig manuell bewertetete Datensätze mit einbringen):

- Genauigkeit, Präzision und Recoil der Auswertung (Am wichtigsten)
  - Manuelle Auswertung aller Bewertungen (nach dem Prinzip der Doppelblindstudie)
  - Einteilung der Qualität von 1 - 10 (mit Begründung)
  - Definition fester Bewertungskriterien um subjektive Bewertungen zu minimieren (TBD).
- F1-Score: 2 * (Precision * Recall) / (Precision + Recall)
- Geschwindigkeit der jeweiligen Systeme (Bewertungen / Sec)
- Ressourcenauslastung

5.1.2 Evaluation der Ergebnisse

5.1.3 Diskussion der Ergebnisse

5.1.4 Interpretation der Ergebnisse

5.1.5 Einschränkungen und Limitationen

Kapitel 6

Zusammenfassung und Ausblick

6.1 Zusammenfassung der Ergebnisse

What was done?

6.2 Fazit

What was learnt?

6.3 Ausblick

What can/has to be/may be done in future research? Impact on other branches of science? society?

Appendix

Abbildungsverzeichnis

Tabellenverzeichnis

Literaturverzeichnis