Schweizer Dialekt Erkenner

Studiengang: BSc in Informatik
Betreuer: Prof. Dr. Erik Graf

In den letzten Jahren hat die natürliche Sprachverarbeitung dank grosser, mehrsprachiger Modelle von Google, Meta AI und OpenAI erhebliche Fortschritte gemacht. Besonders für ressourcenarme Sprachen zeigen diese Modelle grosses Potenzial. Diese Arbeit untersucht die Anwendbarkeit der natürlichen Sprachverarbeitung für Schweizerdeutsch, mit Schwerpunkt auf Dialekterkennung und der Entwicklung eines Speech-To-Text-Systems.

Zusammenfassung

Die Bachelorarbeit «Schweizer Dialekt Erkenner» untersuchte das Potenzial von Natural Language Processing (NLP) zur Erkennung von Schweizerdeutschen Dialekten anhand von Audioaufnahmen. Dabei diente das von Meta AI entwickelte Sprachmodell MMS (Massively Multilingual Speech) als Grundlage, um die Effektivität eines Transfer-Learnings für die Dialekterkennung zu erforschen. Dieses Modell basiert auf Wav2Vec2, das ursprünglich für die automatische Spracherkennung (Speech-to-Text) entwickelt wurde. Im Kern verwendet es einen Transformer, der auf dem Attention-Algorithmus basiert und in den letzten Jahren ein enormes Potenzial für verschiedenste Anwendungen im Bereich des maschinellen Lernens gezeigt hat.

Ziele

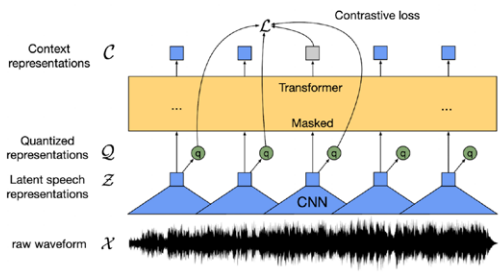
Das Hauptziel der Arbeit bestand darin, relevante Datensätze zu finden und auf geeignete Form zu bringen, um Feinabstimmungsversuche für eine Dialektidentifikation durchzuführen. Als Nebenziel ging es darum herauszufinden ob mit den gegeben Datensätzen auch ein Speech-To-Text System für Schweizerdeutsch entwickelt werden könnte.

Herausforderungen

Die Experimente zeigten, dass die Dialektklassifizierung anspruchsvoll ist und stark von den Daten abhängt. In fast allen Datensätzen waren die Kantone ungleich vertreten. Kleinere Kantone, besonders in der Innenschweiz, hatten so wenige Aufnahmen, dass ihre Dialekte kaum berücksichtigt werden konnten. Auch Verzerrungen bzw. Bias, wie die Geschlechterverteilung, erschwerten die Entwicklung eines robusten Klassifikators.

Ergebnisse

Verschiedene Datensätze und Konstellationen wurden für Feinabstimmungen genutzt, wobei bis zu 70% Genauigkeit erreicht wurde. Die Ergebnisse müssen jedoch mit Vorsicht betrachtet werden, da die meisten Datensätze aufgrund der genannten Verzerrungen ein «zu gutes» Testergebnis zeigen. Trotz der Herausforderungen durch Verzerrungen in den Datensätzen werden in meiner Arbeit ausführlich Massnahmen dagegen diskutiert. Insbesondere im Bereich Speech-To-Text stellte sich die Schwierigkeit heraus, dass die meisten Transkriptionen auf Hochdeutsch vorlagen, während im schweizerdeutschen Datensatz keine standardisierten Richtlinien existierten, was zu erheblichen Variationen in der Verschriftlichung führte. Dennoch konnten trotz dieser Hindernisse interessante Ergebnisse erzielt werden. Über eine Web-Applikation können verschiedene Modelle getestet werden, indem eine kurze Audioaufnahme hochgeladen wird. Anschliessend kann entweder eine Transkription generiert oder eine Dialektklassifizierung vorgenommen werden, wobei die Ergebnisse auf einer Schweizer Karte visualisiert werden.



Architektur von Wav2Vec2 mit Attention Mechanismus

Wahrscheinlichkeitsverteilung der Modell-Inferenz

Florian Leuenberger
Data Engineering
079 524 65 21
leuenberger.florian@gmx.ch