



School of Engineering

InIT Institut für angewandte
Informationstechnologie

Sprach-Klassifikation mit wav2vec 2.0

Wav2Vec 2.0-XLSR-53 ist ein leistungsfähiges Modell, das vortrainiert wurde, um mehrsprachige Sprachrepräsentationen end-to-end zu erlernen. Dialekt-Identifikation (DID) und Akzent-Identifikation (AID) können verwendet werden, um automatische Spracherkennungssysteme (ASR) in Sprachen mit mehreren ausgeprägten Dialekten oder Akzenten zu verbessern.

In dieser Arbeit wird ein Klassifikator auf Basis von wav2vec verwendet, um Sprache zu klassifizieren. Es wird evaluiert, wie das Modell performt, wenn es auf Datensätzen mit geringen Ressourcen trainiert wird. Es werden verschiedene Experimente in den Bereichen AID in Englisch und Spanisch durchgeführt. Zusätzlich wurden Evaluationen auf kurzen Samples durchgeführt. Um die Fähigkeiten von wav2vec weiter zu untersuchen, wird ein Alters- und Geschlechtsklassifikator in deutscher Sprache trainiert. Die verwendeten Korpora wurden aus Mozillas CommonVoice (Common Voice) extrahiert. Trainiert wurde auf 1.5 bis 8 Stunden pro Klasse.

Bei der Klassifikation von sechs Akzenten konnte für Englisch ein durchschnittlicher F1-Score von 0.396 erreicht werden. Bei sieben spanischen Akzenten wurde ein F1-Score von 0.266 erreicht. Im Bereich der Geschlechtsidentifikation wurde ein F1-Score von 0,90 erreicht, während für die Altersidentifikation ein Wert von 0.360 mit einem Macro Averaged Mean Absolute Error (MAEM) von 0.982 erzielt wurde.

Die Ergebnisse zeigen, dass es möglich ist, einen Klassifikator auf wav2vec zu trainieren, allerdings entsprechen die erreichten Scores nicht den gewünschten Werten. Es zeigt sich, dass für das Training eines besseren Klassifikators längere und mehr Audiodaten benötigt werden. Für die Sammlung von Trainingsdaten ist es wichtig, dass die Aufnahmen frei gesprochene Sprache enthalten, die nicht von einem Text abgelesen wurden.



Diplomierende
Pascal Fivian
Dominique Reiser

Dozent
Mark Cieliebak

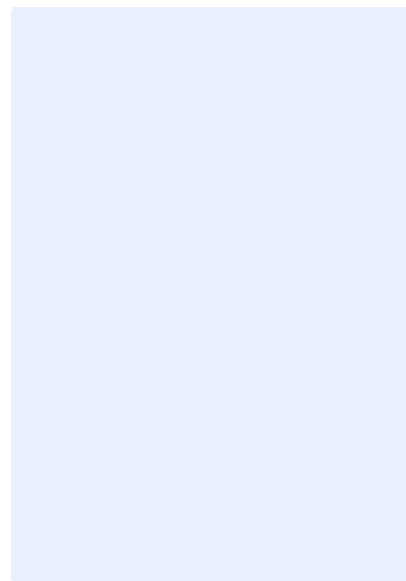


Bild klein 1.