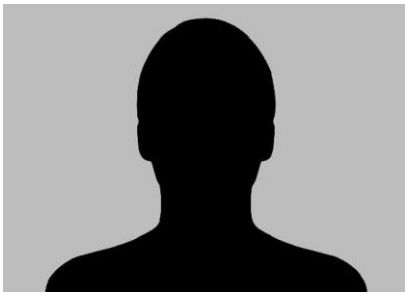
Zürcher Hochschule
für Angewandte Wissenschaften

School of Engineering

InIT Institut für angewandte Informationstechnologie

Quality Assessment of Automatic Speech Recognition Systems



Diplomierende
Fabian Germann
Malgorzata Ulasik

Dozent
Mark Cieliebak

Leistungsstarke Rechenplattformen, Fortschritte in Deep Learning und zunehmende Datenmengen führten zu grossen Fortschritten in automatischer Spracherkennung (ASR - Automatic Speech Recognition) und verbesserter Speech-to-Text-Qualität. Eine zunehmende Nachfrage für ASR-Lösungen ist die Folge. Technologiekonzerne, die ASR-Systeme entwickeln, sind stets bemüht, die bestmögliche Transkribierungsqualität zu erreichen. Unsere Studie bietet einen detaillierten Vergleich aktueller State-of-the-Art-Lösungen, darunter Google Speech-To-Text, IBM Watson und Microsoft Azure. Die Analyse basiert auf Transkriptionen mehrerer bekannter Sprachkorpora wie Timit oder Switchboard. Neben dem Systemvergleich führen wir eine Analyse der Korrelationen zwischen gesprochenen Spracheigenschaften und der Transkriptionsgenauigkeit durch und untersuchen, wie gut die Standardmetrik WER die tatsächliche Transkriptionsqualität widerspiegelt. Die Resultate zeigen, dass die proprietären Cloud-Lösungen die Open-Source-Systeme in praktisch allen Bereichen übertreffen, angeführt von Google, Microsoft und Amazon. Wir stellten ebenfalls fest, dass die schwierigsten Aufgaben der Spracherkennung mit der Übersetzung spontaner Konversationssprachen von nicht Muttersprachlern zusammenhängen. Unsere detaillierte WER-Analyse bestätigt die Annahme, dass diese Metrik nicht immer den Anteil der in der Transkription enthaltenen Informationen widerspiegelt.

Die Heatmap bietet einen Überblick über die grundlegenden Evaluationsergebnisse der am besten abschneidenden Konfigurationen jedes Systems auf allen evaluierten Korpora. Sie zeigt deutlich, dass Korpora mit spontaner Konversationssprache (rechts) bedeutend schwieriger zu transkribieren sind als Korpora mit vorgelesener Sprache (links).