

## Ensemble-Methoden für Spracherkennung

In den letzten Jahren haben Automatic Speech Recognition (ASR) Systeme grosse Fortschritte beim Transkribieren von gesprochener Sprache zu Text gemacht. Für vorgelesene Sprache können moderne Systeme mit Word Error Rates (WER) von weniger als 2% nahezu perfekte Transkripte erstellen. Dieselben Systeme haben jedoch Mühe, für spontane Sprache, wie Meetings oder Telefongespräche, lesbare Transkripte zu erstellen. In dieser Arbeit wird untersucht, ob die WER und somit Transkript-Qualität verbessert werden kann, indem die Hypothesen (Outputs) von mehreren ASR-Systemen kombiniert werden. Es wurden die drei besten Systeme aus der englischen Version vom CEASR-Korpus verwendet, um unterschiedliche Ensemble-Methoden zu implementieren. Diese Ensemble-Methoden wurden mit der WER und SemDist Metrik bewertet.

Um die Hypothesen der Systeme vergleichbar zu machen, müssen sie in sogenannten Alignment einander angeglichen werden. Dazu wird eine Primärhypothese ausgewählt, an welche die verbleibenden Hypothesen jeweils in einem phonetischen Alignment-Paar angeglichen werden. In dieser Arbeit wurde der n-align Algorithmus entwickelt, welcher eine beliebige Anzahl von solchen Alignment-Paaren zu einem einzelnen Alignment kombiniert. Des Weiteren wurde das Alignment UI Tool implementiert, mit welchem das Alignment visualisiert und dessen Qualität schnell beurteilt werden kann.

Die beste, implementierte Ensemble-Methode verwendet Gradient Boosting, um auf Wortbasis zu lernen, in welcher Situation welche Systeme am ehesten richtig liegen. Dadurch konnte die WER auf dem CEASR-Korpus um 5 % auf spontaner Sprache und um 1–2 % auf vorgelesener Sprache reduziert werden. Schlussendlich demonstriert diese Arbeit, dass Ensemble-Methoden für ASR Potenzial haben. Dennoch ist von einer produktiven Nutzung aus Kosten/Nutzen Gründen abzuraten.

Da während dieser Arbeit vieles zum generellen Umfeld von Ensemble-Learning für ASR beobachtet und präsentiert wurde, dienen die Erkenntnisse aus dieser Arbeit als Grundlage für die Entwicklung eines weiteren, besseren Ensemble-Prozesses.



Diplomierende  
Lars Mosimann  
Ralph Scheu

Dozent  
Mark Cieliebak

### Reference

such is my passage engaged on the steamer

### Alignment map



### Detail



Ausschnitt aus dem Alignment Tool  
für die Beurteilung von Alignments

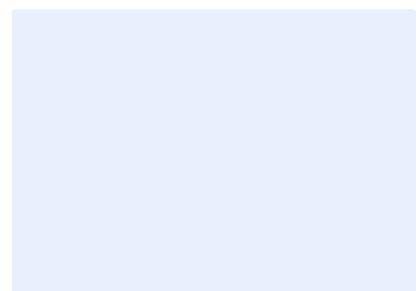


Bild klein 2.