

## Übung 11: Modelloptimierung, Punktevergabe und Bonuspunkte

In dieser Übung geben wir weitere Informationen zur Modelloptimierung, Punktevergabe, zu den Bonuspunkten und der Struktur der Abgabe. **Die Abgabe ist zwischen dem 14.02.2024 und dem 28.02.2024 (23:59) im ISIS-Kurs möglich.**

### Modelloptimierung

Bitte optimieren Sie Ihren Spracherkenner abschließend in mindestens einer Hinsicht. Dazu können Sie die Verbesserungsvorschläge aus **Aufgabe10.pdf** zur Hilfe nehmen. Weitere Ideen listen wir ganz unten in diesem Übungsblatt. Sie haben aber auch völlige Freiheit, um an dieser Stelle Ihre eigenen Ideen zur Verbesserung des Systems auszuprobieren.

Bitte erstellen Sie schließlich eine PDF-Datei mit dem Namen **uebung11.pdf**. Protokollieren Sie darin, welche Anpassungen Sie vorgenommen haben, um Ihre besten Spracherkennungsleistungen zu erzielen, sowie die erreichte Wortfehlerrate - sowohl vor als auch nach Ihren jeweiligen Verbesserungen.

Tabelle 1: Modelloptimierung

Modell	# DNN-Parameter	DNN-Accuracy	Test-WER
Baseline			

Erstellen Sie dazu bitte eine Ergebnisübersicht in der Struktur von Tabelle 1. Geben Sie darin an, welches DNN-Modell Sie in jedem der (mindestens zwei Versuche) verwendet haben (*Modell*), die Anzahl der Modellparameter (*# DNN-Parameter*) sowie die Genauigkeit des DNN und die WER auf dem Testdatensatz (*DNN-Accuracy* und *Test-WER*). Der erste Versuch sollte auf jeden Fall eine Auswertung Ihres eigenen, ersten Systems sein, das den Übungsblättern möglichst genau entspricht (*Baseline*). In den anderen Versuchen benennen Sie bitte das System mit einem Namen nach Wahl und beschreiben Sie in Ihrem PDF Ihre jeweiligen Verbesserungen.

## Punktevergabe und Bonuspunkte

Es gibt 11 Aufgabenblätter mit je 10 Punkten, zusätzlich einmalig 10 Bonuspunkte für das Erkennungsergebnis des besten Systems. Die so erreichte Punktzahl wird auf 110 nach oben begrenzt und durch 110 geteilt, um die Prozentnote der Übung zu bestimmen. Abhängig von der Performance des Erkenners werden folgende Bonuspunkte vergeben:

Wortfehlerrate (%)	$\leq 10$	$\leq 9$	$\leq 8$	$\leq 7$	$\leq 6$	$\leq 5$	$\leq 4$	$\leq 3$	$\leq 2$	$\leq 1$
Bonuspunkte	1	2	3	4	5	6	7	8	9	10

## Hinweise zur Abgabe

Bitte geben Sie uns mit der Abgabe auch Ihr, bezüglich der Wortfehlerrate bestes, trainiertes Spracherkennungs-DNN, Ihr Baseline-Erkennnermodell, und Ihr Modell aus der Torch-Intro-Übung jeweils als **pk1**-Datei ab. Wegen der maximalen Dateigröße des ISIS-Kurses geben Sie bitte nur Ihren Code im **ISIS-Kurs** ab und laden Sie die trainierten DNNs bitte in die **TU-Cloud** hoch.

Für die Abgabe im **ISIS-Kurs** nutzen Sie bitte folgende Struktur:

- **Uebung-ASE**
  - info-gruppe.txt
  - uebung1.py
  - ...
  - uebung11.pdf
  - **data**
    - \* TEST-MAN-AH-3O33951A.wav
    - \* TEST-WOMAN-BF-7O17O49A.npy
  - **dataset**
    - \* train.json
    - \* dev.json
    - \* test.json
  - **recognizer**
    - \* tools.py

- \* feature\_extraction.py
- \* hmm.py
- \* model.py
- \* train.py
- \* utils.py
- **torch\_intro**
  - \* **dataset**
    - train.json
    - dev.json
    - test.json
  - \* **local**
    - model.py
    - train.py
    - utils.py

Um trainierte DNN-Modelle in die **TU-Cloud** hochzuladen, nutzen Sie bitte die folgende Struktur:

- **trained**
  - baseline\_model.pkl
  - best\_model.pkl
- **torch\_intro**
  - **trained**
    - \* model.pkl

Alle **fetten** Elemente sind Ordner. Bitte versuchen Sie nicht, den Datensatz mitzuschicken. Stellen Sie sicher, dass alle einzelnen Übungen in dieser Ordnerstruktur funktionieren und auch alle Plots angezeigt werden.

Bereiten Sie auch eine Datei **info-gruppe.txt** vor, welche alle Gruppenmitglieder mit Name, Matrikelnummer und E-Mail-Adresse beinhaltet.

Laden Sie den Code im ZIP-Dateiformat hoch (wenn Ihre Gruppennummer 1 ist, sollte Ihre ZIP-Datei **Group1.zip** heißen) und die trainierten DNN-Modelle (als **Group1DNN.zip** bezeichnet) in die TU-Cloud hoch, bis spätestens Mittwoch, den 28.02.2024, um 23:59 Uhr.

**TU-Cloud Link:** <https://tubcloud.tu-berlin.de/s/9G8CWqKpnznztqt>

**Alle späteren Abgaben können wir nicht mehr bewerten. Auch alle Abgaben, die von der oben definierten Struktur abweichen oder bei denen die Infos zur Gruppe fehlen, werden nicht gewertet.**

## Optimierung des Erkennungssystems

Um Zeit zu sparen, können Sie ein Subset aus dem vorhandenen Validierungsset erstellen. Wenden Sie Ihre Optimierungsstrategien auf dieses Validierungssubset an. Wenn dies die Leistung des Spracherkenners verbessert, werten Sie Ihre Optimierungsstrategie auf dem gesamten Datensatz aus.

Beachten Sie zunächst die Optimierungsideen aus **Aufgabe10.pdf**. Außerdem gibt es beispielsweise folgende Optimierungsmöglichkeiten:

### DNN-Modelltopologie

Entsprechend der Vorlesung oder eigenen Ideen können Sie verschiedene DNN-Topologien in Betracht ziehen. Denkbar wären zum Beispiel:

- ein Convolutional Neural Network (**CNN**)
- ein Recurrent Neural Network mit Long-Short Term Memory (LSTM)-Schichten (**RNN**)
- oder ein Transformer Encoder-Modell auf Basis des Attention-Mechanismus (**ATT**)

### HMM-Modelltopologie

Eine weitere Möglichkeit, bessere akustische Modelle zu bekommen und gleichzeitig eine Balance zwischen den Pfadkosten innerhalb der Modelle und zwischen den Modellen zu erreichen, ist die Anpassung der Anzahl Zustände pro Wortmodell. Mehr Zustände erlauben z.B. eine detailliertere Modellierung des Sprachsignals. Aber Achtung - dazu müssen Sie auch das DNN neu trainieren!

### Anpassung der Merkmale

Es ist auch möglich, optimierte, oder ganz andere Merkmale für die Spracherkennung zu verwenden, und z.B. die Kontextgröße am Eingang des DNNs zu variieren.