

NLP Assignment 4

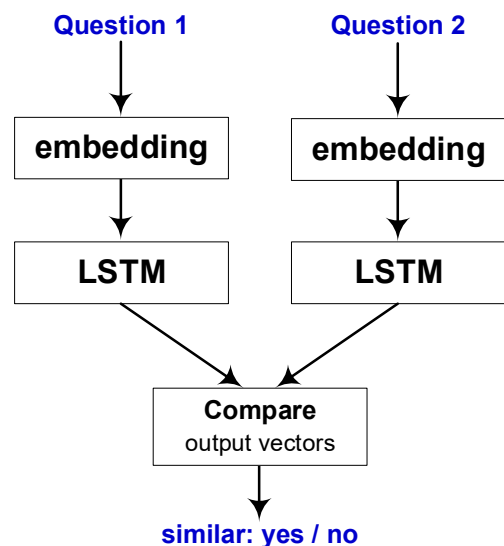
Dieses Assignment wird bewertet !

Die Zusammenarbeit unter Studierenden ist erwünscht, so lange sich diese auf die Diskussion von Konzepten oder Problemen mit TensorFlow/Keras konzentrieren. Das Kopieren von Code ist nicht erlaubt, in diesem Falle werden alle Lösungen der beteiligten Parteien mit 0 Punkten bewertet; dazu werden alle Lösungen (manuell und automatisiert mit Plagiat-Checkern) auf Gruppenarbeit und Kopien untersucht.

In diesem Assignment geht es um die Überprüfung, ob zwei Fragen inhaltlich gleich sind. Insbesondere in der gegenwärtigen Zeit des Online-Lernens ist es nützlich zu wissen, ob Fragen von zwei verschiedenen Studierenden gleich beantwortet werden können.

Entwickeln Sie dazu mit *Tensorflow/Keras* ein System, welches für zwei vorgegebene Fragen feststellt, ob diese inhaltlich gleich sind oder nicht. Trainieren Sie Ihr Modell mit den Daten `Questions_train.csv` und optimieren Sie die Hyperparameter des Modells mit Hilfe von `Questions_eval.csv`. Diese Datensets enthalten pro Zeile zwei Fragen und die Angabe, ob die beiden gleich sind (*is_duplicate*) sowie IDs für das Fragenpaar und die einzelnen Fragen.

Verwenden Sie für Ihr Modell ein Siamese-Netzwerk; dies ist ein Netzwerk, welches zwei Input-Kanäle entgegennimmt, aber gemeinsame Parameter verwendet und optimieren Sie diese mit einer geeigneten *Loss-Funktion*.



Hinweis: Überprüfen Sie vor der Abgabe ob Ihr Code fehlerfrei ausgeführt werden kann!

Abgabe : In Form eines Jupyter Notebooks
per Email an dominik.frefel@fhnw.ch bis 8. Juni 2020