

Transformer

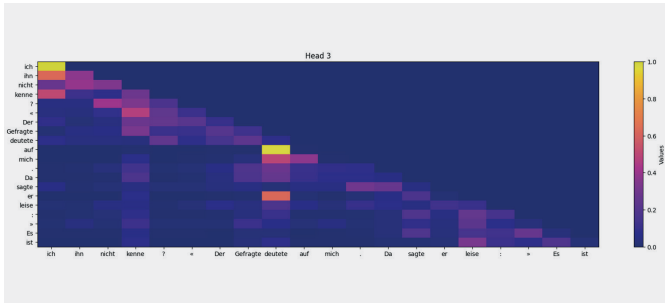


Abb. 1: Self-Attention Matrix des 3. Heads eines der trainierten Modelle

ent-, "fiel ich ihm ein . . . Ich weiß ja auch nicht , wohin sie gehören ! " Er griff zu sich , hielt ihm die beiden Hände entgegen und ließ sie auf ihn los ; er hielt mir die Arme fest an den Mund . Ich zog ihn fort und drückte ihm die Arme aus und riß ihn mit der Rechten , um ihn auf meinen Gürtel , und riß ihn los , um meine Klinge und schlug ihn aus den Hals in die Arme auf den Kopf . Da hinten an die Arme aus

Abb. 2: Generierter Text eines Modells, das mit dem Karl May Datenset trainiert wurde

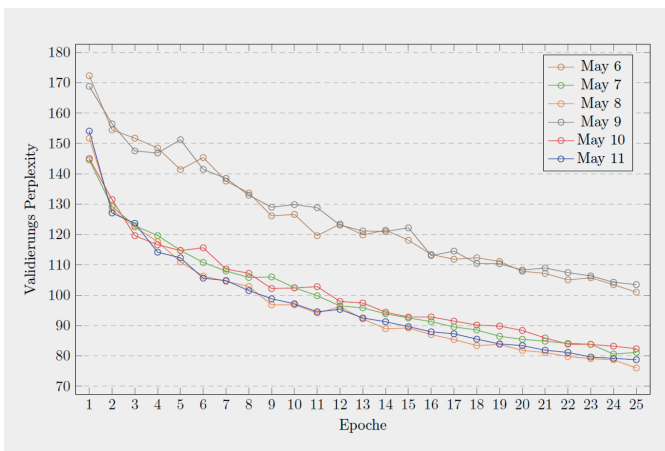


Abb. 3: Verlauf der Perplexities vom Training von verschiedenen Modellen

Problemstellung

Transformer gehören zu den neueren Architekturen in der Machinelearning-welt. Ziel dieser Arbeit ist es, diese Architektur besser zu verstehen und selbst ein Transformer-Modell zu programmieren. Mit diesem Modell soll dann Text im Stil des Datensatzes generiert werden.

Lösungskonzept

Die Modelle sollen mit Pytorch programmiert werden. Die Implementation wird auf Google Collab realisiert. Der Trainingsprozess kann mit Hilfe der GPUs von Google Collab um den Faktor 20 verschleunert werden. Als Datensatz werden das Wikitext2 Datenset und für eine weitere Generation von Modellen 25 Bücher von Karl May verwendet. Zusätzlich wird ein Programm, um mit dem Modell Text zu generieren, geschrieben. Um bestmögliche Resultate zu erzielen, wird auch ein Beam-Search eingesetzt.

Realisierung

Es wurde eine erste Serie von verschiedenen Transformer-Modellen trainiert. Bei diesen Modellen wurden die Anzahl Layer und Heads variiert und untersucht, ob diese Änderungen eine Verbesserung im generierten Text hervorbringen. Anhand der Resultate wurden dann weitere Untersuchungen gemacht.

Ergebnisse

Es wurde herausgefunden, dass die Anzahl der Heads einen Einfluss auf das Resultat haben. Die Heads verschlechtern die Resultate des Modells. Bei einer Embeddingsize von 256 schneiden vier Heads in jedem getesteten Fall besser ab als Modelle, welche mit acht Heads trainiert wurden. In Perplexity ausgedrückt erreicht das beste Modell 59.99. Die Modelle, die mit dem Wikitext2 Datenset trainiert wurden, generierten grösstenteils Zeichenketten anstelle von leserlichem Text. Die Modelle, die mit den Karl May Büchern trainiert wurden, erzeugten leserliche Resultate. Diese Texte folgen einem wagen roten Faden, die einzelnen Sätze machen jedoch nicht immer Sinn.



Diplomand
Vaes Nick

Dozent
Prof. Dr. T. Hunziker

Themengebiet
Nachrichtentechnik/Signal Processing

Projektpartner
Intern