Zürcher Hochschule
für Angewandte Wissenschaften

School of
Engineering

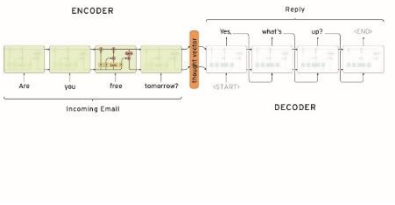
InIT Institut für angewandte
Informationstechnologie

Erstellung von Dialog-Systemen mittels Sequence-To-Sequence Learning

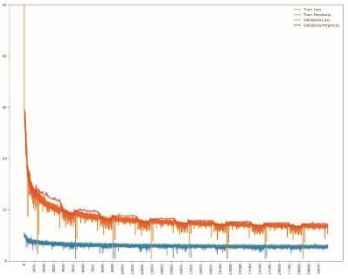
Diplomierende
Dirk Von Gruenigen
Martin Weilenmann

Dozierende
Mark Cieliebak
Stephan Neuhaus

In der folgenden Arbeit untersuchen wir den Aufbau eines End-To-End-Dialogsystems mit Hilfe von Recurrent Neural Networks und der Sequence-To-Sequence-Lernmethode.
Das Ziel dieser Arbeit ist es, herauszufinden ob ein kleineres Modell, welches auf einer GPU trainiert wurde, noch vergleichbar gute Ergebnisse wie ein grösseres hervorbringen kann.
Die ersten Kapitel beschreiben die Entwicklung des Softwaresystems, die Datensätze sowie Methodik zur Durchführung der Experimente.
Die Analyse der daraus resultierenden Modelle erfolgt unter mehreren Aspekten. Als Erstes analysieren wir den Lernprozess. Dabei hat sich gezeigt, dass die Struktur und der sprachliche Ursprung der Daten einen spürbaren Einfluss auf den Lernprozess der Modelle haben.
Anschliessend evaluieren wir die Leistung unserer Modelle, unter anderem mit einer neu vorgeschlagenen Metrik basierend auf der Sent2Vec-Bibliothek, um die semantische Ähnlichkeit numerisch zu messen.
Die Analysen zeigen, dass unsere Modelle einige Schwierigkeiten haben. Ein grosses Problem ist, dass solche Modelle dazu tendieren, generische Antworten zu erzeugen, was zu einer Verschlechterung der Ergebnisse führt. Dies stimmt aber nicht mit unserem subjektiven Eindruck überein, dass die Modelle mit der Zeit besser werden. Um dieses Verhalten erklären zu können, untersuchen wir, wie sich die von den Modellen verwendete Sprache im Laufe der Zeit entwickelt. Diese Analyse zeigt auf, dass die Sprachvielfalt mit fortlaufendem Training grösser wird und der Anteil generischen Antworten sinkt.
Anschliessend folgt ein Vergleich mit anderen Systemen, nämlich mit Cleverbot und den Ergebnissen aus dem Paper "Neural Conversational Model" von Vinyals and Le. Dabei zeigt sich, dass unsere Modelle vergleichbar gute Ergebnisse wie die anderen Systeme erbringen, sofern die Dialoge nicht zu komplex werden. Die Gründe für die schlechteren Ergebnisse bei komplexen Dialogen vermuten wir in der Grösse der Modelle sowie der kurzen Trainingszeit.
Obwohl die Ergebnisse nicht ganz unseren Erwartungen entsprachen, sind wir insgesamt zufrieden. Unsere Modelle sind in der Lage, teilweise sinnvolle und interessante Antworten zu erzeugen, wie zum Beispiel "i m a bot" als Antwort auf die Frage "what are you?".



Architektur eines Sequence-To-Sequence-Modelles. Der Encoder (links) erzeugt dabei aus den Worten des Eingabesatzes eine interne Repräsentation. Diese wird an den Decoder (rechts) weitergereicht und dieser erzeugt damit dann den Ausgabesatz.



Entwicklung der Loss (blau) und Perplexität (orange) während des Trainings. Beide messen, wie gut das Modell die tatsächliche Wahrscheinlichkeitsverteilung während des Lernvorgangs annähert. Tiefere Werte sind besser, das Modell verbessert sich also.