

Protokolle zum Colloquium (29.05.2017)

Student: Korbinian Schmidhuber, Betreuerin: Annemarie Friedrich

„Disambiguierung eines japanischen Aspekt-Markers mithilfe von Parallel-Korpora“

Korbinian plant, Englisch-Japanische Parallel-Korpora einzusetzen, um den jeweiligen englischen Übersetzungen die korrekte Interpretation spezieller ambiger Japanischer Suffixe (Aspekt-Marker) zu entnehmen, die dann zum trainieren eines Classifiers zur Desambiguierung dieser Marker verwendet werden können. Als solche Parallelkorpora bieten sich zum Beispiel der Wikipedia-Korpus, der Basic-Sentences-Korpus oder „Wachturm“ Ausgaben in Englisch und Japanisch an.

Bei der Aspekt-Desambiguierung beschränkt er sich für seine Bachelorarbeit auf den Suffix „te-iru“, welcher je nach dem vorangestellten Verb entweder eine fortschreitende oder eine abgeschlossene Situation beschreiben kann.

Das größte Problem stellt sich für ihn aktuell bei der s.g. Alignierung (Zuordnung) der Verben zwischen Ausgangssatz und Übersetzung mit den Programmen GIZA++ und fast_align, die nur bei etwa 30% der Sätze eine Zuordnung finden.

Student: Dayyan Smith

„Exploring the effects of regularization on a Neural Network in the context of Fake News Detection“

Im Rahmen der s.g. „Fake News Challenge“ optimiert Dayyan ein Neuronales Netzwerk für Headline-Body Stance detection. Die Fake News Challenge soll Teilnehmer dazu motivieren, herauszufinden, wie KI-Technologien, insb. Neuronale Netzwerke, zur Auffindung gefälschter Nachrichten beitragen können.

Stance detection beschäftigt sich mit der Erkennung der semantischen Haltung einer Aussage (Nachrichtenartikel) gegenüber einer anderen Aussage (Überschrift). Das Neuronale Netzwerk besteht aus zwei Bidirektionalen Rekurrenten Autoencodern, jeweils für die Überschrift und den Text. Die von den Autoencodern generierten semantischen Repräsentationsvektoren von Überschrift und Text werden über zwei Hidden Layer in einen Softmax-Layer geleitet, der dann zwischen mehreren Labels entscheiden kann: Agreement, Disagreement, Discussion, Related, Unrelated.

Dayyan hat dieses Netzwerk mit verschiedenen Regularisierungsmethoden trainiert, und konnte bereits erste Ergebnisse präsentieren: Die auf SVM basierte künstliche Score-Baseline der Challenge von 75% wurde konsistent durch Dayyans Netzwerk geschlagen, welches tlw. 83 % erreichte.

Allerdings hatte die Einführung von Regularisierung bisher nicht den gewünschten Effekt: Statt durch bessere Abstraktionen höhere Genauigkeiten zu erzielen, kollabierte durch L2-Regularisierung der Score des Modells auf unter 50%. Als nächstes möchte Dayyan sich der Analyse dieses Effektes, sowie der Evaluation der Auswirkung von Dropout of die Leistung des Modells widmen.
