



School of Engineering

INIT Institut für angewandte
Informationstechnologie

Natural Language Generation aus strukturierten Daten

Es hat sich mehrfach gezeigt, dass neuronale Netzwerke gut im Gebiet der Natural Language Generation funktionieren. Ein Problem, welches sie zurzeit aber noch haben, ist, dass eine grosse Menge an speziell vorbereiteten Daten benötigt werden.

Das Ziel dieser Arbeit ist es festzustellen, ob ein bereits existierendes neuronales Netz, das gut darin ist, Restaurant Reviews zu generieren, auch mit keinen oder nur geringen Anpassungen in der Lage ist, ähnlich gute Resultate im Generieren von Laptop Reviews zu erreichen. Ein weiteres Ziel ist es herauszufinden, wie schwer es ist, ein passendes Datenset für ein neuronales Netz zu erstellen.

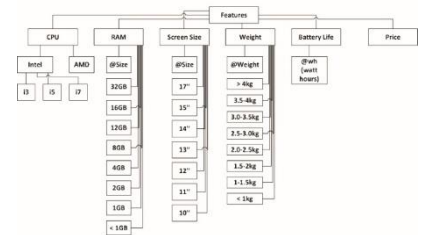
Zuerst wurden Daten von einem Onlinehändler gesammelt, die wichtigsten Informationen extrahiert und das Ganze in einem Datenset zusammengefasst. Anschliessend wurde das existierende neuronale Netz auf dem erstellten Datenset trainiert.

Die Resultate zeigten, dass die Outputqualität eines neuronalen Netzwerks in grossem Zusammenhang mit der Qualität des Datensets steht. Nicht nur die Grösse des Datensets selber, sondern auch Eigenschaften wie die Grösse des Vokabulars (Anzahl einzigartiger Wörter) spielen eine grosse Rolle. Da das erstellte Datenset noch sehr "unsauber" war, führte es dazu, dass auch der Output schlecht war. Ausgehend von diesen Resultaten musste die Problemstellung der Arbeit vereinfacht werden. Die zweite Problemstellung beinhaltet die Frage, wie gut ein neuronales Netz beim Ordnen von ungeordneten Sätzen und beim Füllen von Lücken in Sätzen ist.

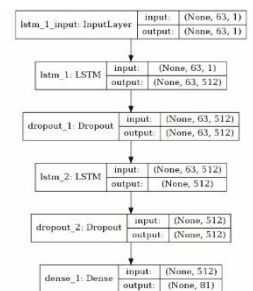


Diplomierende
Tobias Fierz
Jerome Schaub

Dozent
Mark Cieliebak



Der definierte Feature-Vektor



Keras Modell für die Text-Generation