

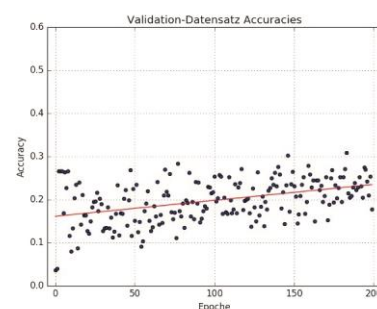
Machine Learning bei Textanalyse: Weiss mein Computer wie alt ich bin?

Informationen zum Autor oder Autorin aus reinem Text zu ermitteln ist heutzutage sehr gefragt in Applikationen der Forensik, Security und Marketing. Mithilfe von Informationen zu Geschlecht und Alter kann beispielsweise auf Webseiten dynamisch Inhalt passend zu den Informationen des Benutzers angezeigt werden. Dies ist nur eine vieler Anwendungen, welche eine Erkennung solcher Eigenschaften möglich macht. Diese Arbeit beschäftigt sich damit, dieses Problem möglichst treffsicher, mithilfe von Machine-Learning-Ansätzen, zu lösen. Es wurde ein Convolutional Neural Network genutzt, um Alter und Geschlecht des Autors eines Textes zu erkennen. Diese Texte wurden in Englisch, Spanisch und Niederländisch verfasst. Dazu wurde prototypisch ein System zusammengestellt, welches ein solches Netz aufbaut und damit ein möglichst gutes Modell antrainiert. Anschliessend wurde mithilfe des Modells die Erkennung der Eigenschaften vorgenommen. Die Erkennung wurde auf die Daten ausgeführt, bei welchen die Informationen schon bekannt sind. So konnte mit den korrekten Daten verglichen werden und die Effektivität des Modells geprüft werden. Am Ende wurde versucht, das System auf die bestmögliche Erkennung zu bringen. Leider konnten in diesem Rahmen noch keine Ergebnisse in Höhe des State of the Art erzielt werden. Recherchen haben ergeben, dass vergangene Forschungen dieser Problemstellung mithilfe von stilistischen und inhaltsbasierten Features gearbeitet haben. Diese Features könnten zum Beispiel in das System eingebracht werden. Weiter braucht das System generell noch mehr Experimentierarbeit, um die erwarteten Ergebnisse zu erfüllen. Dennoch hat das System grosses Potenzial, mindestens zu State of the Art Ergebnissen zu kommen, wenn nicht sogar darüber hinaus.



Diplomandin
Fiona Waser

Dozierende
Mark Cieliebak
Stephan Neuhaus



Accuracies pro Epoche berechnet während der Trainingsphase eines Convolutional Neural Networks, welches Alters- und Geschlechtererkennung des Autors oder der Autorin eines englischen Textes ausführen kann.