



Master Thesis

Analyse hochdimensionaler MALDI-Daten mittels Convolutional Neural Networks

Analysis of highdimensional MALDI-Data using Convolutional Neural Networks

vorgelegt von/ submitted by: Marvin Neiden

betreut von / supervised by: Prof. Dr. Jan Lellmann (Universität zu Lübeck)

Zusammenfassung

In der bildgebenden Massenspektrometrie wird versucht, die molekulare Verteilung innerhalb einer Gewebeprobe heraus zu finden. Dabei werden in der medizinischen Routine heutzutage klassische Methoden der multivariaten Datenanalyse zur Gewebeklassifikation eingesetzt, um die hochdimensionalen Daten des MALDI-Imaging Verfahren zu verarbeiten. Diese Methoden sind jedoch sehr intensiv bezüglich der Rechenzeit, weshalb nach anderen Methoden gesucht wird.

Methoden des maschinellen Lernens ermöglichen die Analyse und Auswertung von komplexen, großen Datenmengen, und insbesondere Methoden auf der Basis neuronaler Netze haben in den letzten Jahren zahlreiche Anwendungsgebiete revolutioniert.

In dieser Arbeit werden Daten der Massenspektrometrie mittels eines neuronalen Netzes klassifiziert. Zuerst wird anhand eines Datensatzes der Skelettmuskulatur der Ratte gezeigt, dass man die Ergebnisse der konventionellen Methoden reproduzieren kann. Danach wird mit der entwickelten Methode ein Tumor Datensatz klassifiziert. Dabei wird auch getestet, wie anfällig der Klassifikator auf Rauschen ist. Abschließend wird untersucht, inwieweit die verschiedenen Läufe, bei denen die Daten gemessen werden, einen Einfluss auf das Ergebnis haben.

Abstract

The goal of mass spectrometry imaging is to gain information about the molecular distribution of the tissue sample. For this task in medical routine machine learning methods are used to visualize the highdimensional data. The computational time of these methods is high and therefore other algorithms or methods are needed.

Due to the technical advances of the last couple years and the development of potent libraries in the field of deep learning, these methods are getting interesting for the mass spectrometry imaging. In this thesis mass spectrometry data is classified with a neural network. At first it is shown that this approach can reproduce the results of the common machine learning methods with a dataset from the skeletal muscle of a rat. Afterwards the new method is applied to an ovarian cancer dataset. On this dataset the effect of noise on the classifier is studied. Finally the influence of the different measurement runs is looked into.

Ergebnisse / Results

	Klassifikationsgenauigkeit in %	
	Daten normalisiert	Daten nicht normalisiert
Training TAM	99.9156	99.8987
Validierung TAM	96.5110	96.6111
Training TM	99.8717	99.8761
Validierung TM	85.8337 92.0060	86.7153 93.6249

Tabelle 1: Die durchschnittlichen Genauigkeiten der Trauma angrenzenden-(TAM) und Trauma-(TM) Regionen im Vergleich. Bei der Validierung von der TM-Region sind zwei Wahrscheinlichkeiten angegeben. Bei der Ersten sind alle Daten mit einbezogen und die Zweite ist ohne den Ausreißer des Klassifikators von TM-2 auf den TM-3 Daten berechnet.