

Abschlussarbeit in Bioinformatik

Hannah Franziska Löchel

AG Bioinformatik (Prof. Dr. Dominik Heider)
loechelh@mathematik.uni-marburg.de

Thema: Deep Learning zur Klassifizierung von Lebererkrankungen anhand von Mikrobiomen mittels Chaos Game Representation

Als „Chaos Game“ wird ein Algorithmus bezeichnet, der fraktale Strukturen aus Sequenzen erzeugt. Dieser Algorithmus kann z.B. zur Konstruktion des Sierpinski-Dreiecks verwendet werden. Zu Beginn wird ein Zufälliger Punkt im Dreieck gewählt. Von diesem Punkt aus wird an der Hälfte der Strecke zu einer zufälligen Ecke ein Punkt markiert, durch weitere Iterationen dieser Vorgehensweise füllt sich das Sierpinski-Dreieck.

Neben der Darstellung von zufälligen Zahlenreihen, ist auch die Darstellung anderer Sequenzen wie z.B. Texte denkbar. In der Bioinformatik kann diese Vorgehensweise z.B. genutzt werden, um DNA-Sequenzen oder Proteinsequenzen darzustellen. Dadurch entstandene Bilder zeigen charakteristische Muster für den jeweiligen Organismus [1].

Als Mikrobiom werden die Mikroorganismen bezeichnet, die ein Biom besiedeln, z.B. den Darm des Menschen. Dabei spielt die Zusammensetzung, des Mikrobioms eine zentrale Rolle bei verschiedenen Krankheiten, so z.B. auch bei Lebererkrankungen [2], wie HCC [3].

Ziel der Arbeit ist es, basierend auf Mikrobiomdaten Lebererkrankungen zu klassifizieren. Das Encoding der Sequenzdaten soll mittels Chaos Game Algorithmus erfolgen. Dazu soll das R Paket *kaos* [4] verwendet werden.

Literatur

[1] Jeffrey HJ: Chaos game representation of gene structure. *Nucleic Acids Res.* 1990, 18(8):2163-2170.

[2] Hov JR, Karlsen TH: The Microbiome in Primary Sclerosing Cholangitis: Current Evidence and Potential Concepts. *Semin Liver Dis.* 2017, 37(4):314-331.

[3] Best J et al.: The GALAD scoring algorithm based on AFP, AFP-L3, and DCP significantly improves detection of BCLC early stage hepatocellular carcinoma. *Zeitschrift für Gastroenterologie* 2016, 54(12):1296-1305.

[4] Loechel HF, Eger D, Sperlea T, Heider D: Deep Learning on Chaos Game Representation for Proteins. *bioRxiv* 2019, 575324.

Bitte melden Sie sich bei Interesse per Email bei:

24.04.2019

Hannah Franziska Löchel - loechelh@Mathematik.Uni-Marburg.de