

Abschlussarbeit in Bioinformatik

Roman Martin & Hannah Franziska Löchel

AG Bioinformatik (Prof. Dr. Dominik Heider)

loechelh@mathematik.uni-marburg.de

Thema: Webtool-Entwicklung zur Analyse von Chaos Game Representations

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 Ecke ein Punkt markiert, durch weitere Iterationen dieser Vorgehensweise füllt sich das Sierpinski-Dreieck [1].

Neben der Darstellung von zufälligen Zahlenreihen, ist auch die Darstellung anderer Sequenzen wie beispielsweise von Texten möglich. In der Bioinformatik kann diese Vorgehensweise genutzt werden, um biologische Sequenzen darzustellen. Die dadurch entstandene Bilder zeigen charakteristische Muster abhängig von der Sequenz [2,3]. Das R-Paket `kaos` erlaubt die einfache Darstellung mit beliebigen Eingabedaten [4].

Ziel der Arbeit ist die Implementierung weiterer Funktionen für das R-Paket `kaos`. Bestandteil der Erweiterung soll eine interaktive Webseite zur Analyse der entstandenen Bilder sein. Weiterhin soll eine genaue Zuordnung der dargestellten Punkte zu einem Zielwert ermöglicht werden. Am Beispiel von SMILES-formatierten [5] Molekülen (Strings) sollen diese Funktionen getestet werden, um zu prüfen, ob sich so gegebene Zielwerte in den Punkten clustern lassen.

Literatur

[1] H. Juergens, H.-O. Peitgen, und D. Saupe, *The Language of Fractals*, Bd. 263. 1990

[2] H. J. Jeffrey, „Chaos game representation of gene structure“, *Nucleic Acids Res.*, Bd. 18, Nr. 8, S. 2163–2170, 1990.

[3] S. Basu, A. Pan, C. Dutta, und J. Das, „Chaos game representation of proteins“, *J. Mol. Graph. Model.*, Bd. 15, Nr. 5, S. 279–289, 1997.

[4] D. Weininger, „SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules“, *J. Chem. Inf. Comput. Sci.*, Vol. 28, S. 31–36, 1988

[5] H.F. Loechel, D. Eger, T. Sperlea, D. Heider, „Deep Learning on Chaos Game Representation for Proteins“, btz493 575324, 2019.

06.07.2019

Bitte melden Sie sich bei Interesse per Email bei:

Roman Martin - roman.martin@uni-marburg.de

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