



3D-Chaos-Game Repräsentation von Proteinen

Prof. Dr. Dominik Heider
dominik.heider@uni-marburg.de

Zusammenfassung

Als „Chaos Game“ wird ein Algorithmus bezeichnet, der fraktale Strukturen 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.

In der Bioinformatik kann diese Vorgehensweise als Bildgebendes Verfahren genutzt werden, um so DNA-Sequenzen oder Proteinsequenzen darzustellen. Dadurch entstandene Bilder zeigen charakteristische Muster für den jeweiligen Organismus bzw. das jeweilige Protein [1,2].

Ziel der Arbeit ist die Erstellung eines R-Paketes, das Sequenzen in einer 3D Kugel unter Verwendung des Chaos-Games Algorithmus darstellt. Dabei sollte das Paket so generisch sein, dass jedes Alphabet durch das R-Paket repräsentiert werden kann.

Als Use-Case sollen Proteinsequenzen von HIV mit diesem R-Paket untersucht werden. Mit den so gewonnenen Bildern soll mit Hilfe von Deep-Learning evaluiert werden, ob die Chaos-Game Repräsentation zur Vorhersage von Resistenzen [3,4] geeignet ist.

Zur Optimierung der Klassifizierung der HIV-Daten soll die Anordnung der Aminosäuren auf der Oberfläche der Kugel durch einen genetischen Algorithmus [5] angepasst werden.

Literatur

- [1] Jeffrey HJ: Chaos game representation of gene structure. *Nucleic Acids Res.* 1990, 18(8):2163-2170.
- [2] Basu S, Pan A, Dutta C, Das J: Chaos game representation of proteins. *J. Mol. Graph. Model.* 1997, 15(5):279-289.
- [3] Riemenschneider M, Senge R, Neumann U, Hüllermeier E, Heider D: Exploiting HIV-1 protease and reverse transcriptase cross-resistance information for improved drug resistance prediction by means of multi-label classification. *BioData Mining* 2016, 9:10.
- [4] Heider D, Dybowski JN, Wilms C, Hoffmann D: A simple structure-based model for the prediction of HIV-1 co-receptor tropism. *BioData Mining* 2014, 7:14.
- [5] Pyka M, Heider D, Hauke S, Kircher T, Jansen A: Dynamic causal modeling with genetic algorithms. *Journal of neuroscience methods* 2011, 194(2):402-406.

Bitte melden Sie sich bei Interesse per Email bei:

Prof. Dr. Dominik Heider dominik.heider@uni-marburg.de

Stand: 30.11.2017