



Machine Learning zur Vorhersage von Startwerten für Löslichkeitssimulationen

Molekulare Simulationen sind Experimente am Computer, bei denen dynamische oder stochastische Methoden verwendet werden, um die Positionen von Atomen und Molekülen und die daraus resultierenden Energien zu berechnen. Aus diesen lassen sich eine Vielzahl an Systemeigenschaften bestimmen, wie beispielsweise auch die Löslichkeit eines Stoffes in einem anderen.

Bei der Bestimmung der Löslichkeit mit Hilfe von molekularen Simulationen wird die freie Enthalpiedifferenz zwischen den beiden Endzuständen -gelöst und ungelöst- bestimmt. Dafür werden Zwischenzustände eingesetzt, welche sich durch einen Parameter λ beschreiben lassen. Die Anzahl und die Verteilung der λ -Parameter entscheidet über die Genauigkeit der Simulationsergebnisse. Zu jedem λ -Parameter muss jedoch auch eine eigenständige Simulationsreihe durchgeführt werden, was bei einer hohen Anzahl an λ -Zuständen zu einem beträchtlichen Rechenaufwand führen kann.

Um die Genauigkeit der Simulationen zu erhöhen und den Rechenaufwand zu minimieren, ist eine Kenntnis der optimalen Anzahl und Verteilung der λ -Zustände, dem so genannten alchemischen Pfad, notwendig. Ein optimierter alchemischer Pfad kann mit einem am IfT entwickelten Tool ermittelt werden. Die Konvergenz dieses Optimierungsprozesses hängt dabei von der Startverteilung der λ -Zustände ab.

Machine Learning bietet die Möglichkeit Korrelationen zwischen einem Satz an Daten und von System beschreibenden Parametern, Deskriptoren, zu ermitteln. Das Ziel dieser Arbeit soll es sein, mit Hilfe von Machine Learning Kenntnisse über Korrelationen zwischen der optimierten Anzahl und Verteilung der λ -Parameter mit molekularen Deskriptoren zu erlangen, um so durch eine geeignete Startverteilung eine schnelle Konvergenz des Optimierungstools erreichen zu können. Die dafür zu verwendenden Simulationsdaten werden dabei aus bereits durchgeführten Molekularen Simulationen entnommen.

Die Aufgaben umfassen:

- Aufarbeitung der Daten für eine Machine Learning unterstützte Korrelationsanalyse
- Identifikation geeigneter Deskriptoren
- Auswahl einer geeigneten Machine Learning Variante
- Durchführen der Machine Learning gestützten Korrelationsanalyse
- Bewertung der Vorhersagen

Erste Erfahrungen mit Python und Machine Learning sind erwünscht.

Bei Interesse melden Sie sich bitte bei Jan Bode: Ja.Bode@tu-braunschweig.de