

Feature-Based Applicability Domain Estimation for Structural Kernels

Im Zuge der zunehmenden Bedeutung kernelbasierter maschineller Lernverfahren in der Chemoinformatik stellt sich die Frage, für welche Daten ein gelerntes Modell spezifiziert ist und verlässliche Vorhersagen liefert. Hierfür ist es sinnvoll, die sogenannte *Applicability Domain* des Modells beschreiben zu können. In der Literatur finden sich hierfür vor allem Methoden, die anhand einer deskriptorbasierten Darstellung des Trainingsdatensatzes den Teilraum des *ChemSpace* beschreiben in dem das Modell spezifiziert ist. Diese Ansätze sind allerdings nicht direkt auf kernelbasierte Verfahren übertragbar, da bei diesen der Datenraum in der Regel nicht explizit definiert ist.

Im Rahmen dieser Arbeit sollen deskriptorbasierte *Applicability Domain* Ansätze implementiert und auf ihre Eignung zur Abschätzung der Verlässlichkeit kernelbasierter Vorhersagen untersucht werden. Gegebenenfalls können auch neue Methoden speziell für kernelbasierte *Applicability Domains* entwickelt werden.

Die Implementierung der Algorithmen soll in Java erfolgen und die Chemoinformatik Bibliothek JOELib2 verwenden.

Anforderungen: Erfahrung im Programmieren mit Java. Kenntnisse aus den Bereichen des maschinellen Lernens und der Chemoinformatik sind hilfreich, können allerdings auch im Rahmen dieser Arbeit erworben werden.

Kontakt

Nikolas Fechner
Sand 1, Raum 317
Tel. (07071) 29-77174
nikolas.fechner@uni-tuebingen.de

