

A Machine Learning Interface to a Generic Quadratic Program Solver for QSAR and Virtual Screening

Bei der Verwendung von maschinellen Lernverfahren im computergestützten Wirkstoffentwurf tritt insbesondere bei großen Datensätzen häufig das Problem auf, das die Datenbasis sehr unbalanciert ist. Die optimale Vorgehensweise in diesen Fällen ist die Berücksichtigung der Ungleichverteilung durch eine Anpassung der Loss-Funktion. Dies ist bei den vorhandenen Implementierungen von maschinellen Lernverfahren allerdings nur schwierig zu realisieren.

Aus diesem Grund soll eine Schnittstellenbibliothek programmiert werden, die es ermöglicht eigene Formulierungen von maschinellen Lernaufgaben durch die Optimierungstoolbox CPLEX lösen zu lassen. Im Besonderen soll es möglich sein eigene Loss-Funktionen in die Quadratischen Programme der Support Vektor Maschinen zu integrieren. Hierzu müssen die QP Formulierung unter Verwendung einer Java basierten Bibliothek, die am Lehrstuhl entwickelt wird implementiert werden um die entsprechenden Schnittstellen der CPLEX Toolbox ansprechen können.

Die fertige Schnittstellenbibliothek soll getestet werden indem problemspezifische Loss-Funktionen mit den generischen SVM Formulierungen von LibSVM und der auf große Datensätze spezialisierten LibLinear auf Problemen aus dem computergestützten Wirkstoffentwurf verglichen werden.

Voraussetzungen:

Sehr gute Java und Machine Learning Kenntnisse, Erfahrung in der Verwendung und Entwicklung komplexer Softwarebibliotheken

Kontakt

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