Gruppe sparc
Nicolas Meisberger
Robin Franzke

Projektplan – Ein optimierter Klassifizierungsmechanismus zum machine learning

Aufgabenbeschreibung:
Es gilt de Machine Learning-Algorithmus HULLER in C zu implementieren. Dieser wird zur Klassifizierung von Datensätzen genutzt. Zunächst wird eine Trainingsmenge (Menge von positiven und negativen Datensätzen) gegeben, und daraus „gelernt“ wie Datensätze zu Klassifizieren sind. Neue Datensätze werden nun automatisch vom Algorithmus positiv oder negativ Klassifiziert.

Bedingungen zur Lösung:
-Das Programm muss Datensätze einlesen können
-Anhand dieser Trainingsmenge muss ein Modell erstellt werden (das was gelernt wurde)
-Mithilfe des Modells und neuen Datensätzen müssen diese Klassifiziert werden
-Eine Erkennungsrate von min X% (Hängt sehr stark von der Trainingsmenge ab)

Grundlegend besteht das System aus 2 Komponenten:
-Dem Lernprogramm, (Trainingsmenge → Modell)
-Dem Klassifizierer (Modell+Datenbasis → Klassifizierter Datensatz)

Tests werden wie im machine learning üblich mit extra Testmengen umgesetzt:
-Zunächst wird mit einer Trainingsmenge ein Modell erstellt.
-Mit diesem Modell wird nun eine Testmenge klassifiziert. Von der Testmenge sind die Klassifizierungen allerdings bekannt. So kann der relative Anteil der richtig Klassifizierten Items in Prozent angegeben werden. Dieser Anteil sollte gegen 100% gehen.

Die Effizienz wird durch einen Vergleich mit libsvm sicher gestellt.
Speicherfehler werden schon während der Entwicklung regelmäßig gesucht und beseitigt.
Mehrere MB Speicherlecks sollten nicht auftauchen.

Teilschritte:

-Lernprogramm
    -Datensätze einlesen
    -Modell erstellen
-Klassifizierer
    -Modell einlesen
    -Datensätze einlesen
    -Klassifizierung
    -Klassifizierte Datensätze ausgeben