



Seminarankündigung

Knowledge Discovery

Sommersemester 2009

Prof. Dr.-Ing. Madjid Fathi
Dipl. Wirt.-Inf. Christian Sassenberg
Dipl. Wirt.-Inf. Henning Peuser

Vorbesprechung: Mittwoch 22.4.2009, 11 Uhr, Raum H-A 8107

Inhalt: In vielen Anwendungsgebieten werden heutzutage Datenmengen erfasst, deren Umfänge eine manuelle Analyse zunehmend schwierig, wenn nicht sogar unmöglich machen. Das Forschungsgebiet *Knowledge Discovery in Databases* (KDD) beschäftigt sich mit der Lösung dieses Problems. Es handelt sich hierbei um eine interdisziplinäre Disziplin und verbindet Aspekte aus der Statistik, aus dem maschinellen Lernen sowie aus Datenbanksystemen und behandelt Methoden zur (semi-)automatischen Extraktion von gültigem, neuem und potentiell nützlichem Wissen aus großen Datenbanken. Unternehmen nutzen Methoden des KDD oftmals als Teil ihrer Wissensmanagement Aktivitäten mit dem Ziel Rohdaten in verwertbares Wissen zu transferieren.

Bei Interesse an einem der unten genannten Themen (oder an einem eigenem Thema), kann bereits vor der Besprechung am 22.4.2009 ein Thema einem Studenten zugewiesen werden. Kontaktieren Sie hierzu bitten einen der genannten Ansprechpartner.

Christian Sassenberg
E-Mail: sassenberg@kmcm.uni-siegen.de
Telefon: 0271 740 2642
Raum: H-A 8115

Henning Peuser
E-Mail: henning.peuser@uni-siegen.de
Telefon: 0271 740 2626
Raum: H-A 8116

Themenvorschläge:

Data Mining als Teil des KDD Prozesses

In diesem Grundlagenvortrag soll ein Überblick über den gesamten KDD Prozess gegeben werden mit Hauptaugenmerk auf dem Data Mining Teilschritt. Aufgaben wie Clustering, Klassifikation und Assoziationsanalyse sollen näher beschrieben werden.

Die Bedeutung der Datenvorverarbeitung für den KDD Prozess

Datenvorverarbeitung umfasst die Tätigkeiten, welche in den drei Phasen des KDD Prozesses vor der eigentlichen Datenanalyse anfallen. Ihr kommt eine besondere Bedeutung zu, da die Qualität der KDD Ergebnisse oftmals maßgeblich durch die Vorverarbeitung beeinflusst wird. In diesem Seminarbeitrag soll ein Überblick über die Aufgaben gegeben werden, welche im Rahmen der Datenvorverarbeitung anfallen, incl. einiger Standardverfahren, welche zur Lösung der Aufgaben eingesetzt werden.

Data Mining in der Medizin

Unter den zahlreichen Anwendungsgebieten für KDD hat sich die Medizin als ein in Theorie und Praxis oftmals genutztes Anwendungsgebiet herauskristallisiert. In diesem Seminarbeitrag soll auf die besonderen Anforderungen eingegangen werden, welche an Data Mining für die Medizin bestehen. Konkrete Anwendungsbeispiele sollen dargestellt und erläutert werden.

Variablenselektion zur Verbesserung der Performance von Lernverfahren

Die hohe Anzahl von Spalten in Datentabellen macht in vielen Fällen eine Auswahl der wirklichen relevanten Parameter nötig. In der Literatur hat sich hierfür der Begriff Feature Selection durchgesetzt. Dieser Seminarbeitrag beschreibt die Grundlagen von Feature Selection und stellt die Methoden vor, welche sich über die Jahre gebildet haben.

Privacy Preserving Data Mining

KDD wird häufig auf Datenbanken angewandt, welche persönliche Daten von z.B. Mitarbeitern beinhalten. Im Zuge des Datenschutzes kann der Zugang zu den Daten verwehrt und somit Data Mining Projekte verhindert werden. In diesem Seminarbeitrag sollen typische Probleme und Lösungen für Szenarien präsentiert werden, wo Data Mining auf Hürden des Datenschutzes trifft.

Text Mining

Text Mining bezeichnet die automatisierte Entdeckung relevanter Informationen aus Textdaten. Text-Mining-Systeme liefern im Optimalfall Informationen bzw. Zusammenhänge, von denen Benutzer bisher nicht wussten, dass sie existieren. Ziel dieses Seminarbeitrages ist ein genereller Überblick über Verfahren und Anwendungen aus dem Bereich des Text Minings.

Web Mining

Unter Web-Mining versteht man die Übertragung von Techniken des Data-Mining zur (teil)automatischen Extraktion von Informationen aus dem Internet. In diesem Beitrag soll ein Überblick über typische Fragestellungen gegeben werden, mit denen sich Web Mining beschäftigt, und über Verfahren, mit denen diese Fragestellungen beantwortet werden. An ein oder zwei Fallbeispielen soll der Einsatz von Web Mining verdeutlicht werden.

Visual Data Mining

Visual Data Mining integriert den Benutzer in den KDD Prozess um dessen Fähigkeiten mit der Rechenleistung moderner Computer zu kombinieren. Es soll ein Überblick über die zahlreichen Techniken zur Informationsvisualisierung gegeben werden, welche in den letzten Jahren entwickelt worden sind.

Es ist ausdrücklich erwünscht, dass Studenten auch eigene Themenvorschläge mitbringen, die dann bearbeitet werden können.

Die zu den aufgeführten Themenvorschlägen genannten Inhalte sind nur eine vage Vorgabe. Das heißt ein Student kann zu einem Thema auch Aspekte behandeln, welche in der Beschreibung nicht erwähnt sind. Dasselbe gilt für die letztendlichen Titel der Seminarbeiträge. Da die hier gemachten Vorschläge nur eine grobe Richtung vorgeben, können die letztendlichen, genauen Titel davon abweichen.