

Protokoll Informatik Diplomprüfung Praxis

Zeitraum: SS 2004

Fächer: Data-Mining Algorithmen, Datenexploration, Artificial Intelligence

Note: 1,3

Data-Mining Algorithmen

F: Welche großen Themengebiete hatten wir in der Vorlesung?

A: Clustering, Classification, Association

F: Worum geht es denn bei Assoziation?

A: Warenkorbanalyse, Transaktionen, Frequent Itemsets, Assoziationsregeln

F: Was ist eine Assoziationsregel?

A: Body, Head jeweils Mengen von Items, Support, Confidence

F: Wie sind Support und Confidence definiert?

A: siehe Vorlesung

F: Wie findet man häufige Teilmengen?

A: Apriori Eigenschaft, Apriori Algorithmus

F: Wie viel Datenbankdurchläufe braucht der Apriori Algorithmus, um Häufige Teilmengen der Länge n zu finden?

A: n

F: Wo liegt die Schwäche des Apriori-Algorithmus?

A: Zu viele Kandidaten werden generiert, ging auch noch darum, wieviele

F: *Malt kleines Beispiel mit 3-elementigen Mengen auf* Das hier sind die häufigen 3-Itemsets, machen sie mal einen Apriori Schritt

A: *mach*

F: Wieviele $(n - 1)$ -Itemsets müssen überprüft werden, um zu sehen ob alle Subsets von einem n -Itemset häufig sind?

A: n

F: Wie kann ich das berechnen der Häufigkeiten verbessern?

A: Supportwerte im Hash Tree

F: Was gibt es noch für Verbesserungen?

A: FP-Tree, FP-Growth Verfahren oberflächlich erläutert

F: Wie kann man Regeln aus den Häufigen Teilmengen generieren

A: siehe Vorlesung

DE

F: Welche Ähnlichkeitsmaße hatten wir denn für Bilder?

A: Pixelweise, Transformationsbasiert

F: Was ist die Idee/"Semantik" der Transformation?

A: Distanz im Feature-Raum entspricht Ähnlichkeit der Bilder

F: Was hatten wir denn an Ähnlichkeitsmodellen für Formen?

A: Symmetrische Flächendifferenz, Morphologische Distanz, ... jeweils kurz erklärt

F: Einige dieser Distanzen eignen sich ja auch für mehrstufige Anfragen

A: Filterschritt erklärt, warum, Lower-Bounding-Property

F: Was ist, wenn der Filter immer Null ist?

A: Dann werden alle Elemente der Datenbank im Verfeinerungsschritt berücksichtigt

F: Wie funktioniert die Anfragebearbeitung?

A: Sequentieller Scan $O(n)$, besser Indexstruktur mit angestrebt $O(\log n)$

F: Gut, dann suchen Sie sich doch mal eine Indexstruktur aus

A: R-Baum, Aufbau, Anfragebearbeitung, Überlapp, Effizienz mit einigen Rückfragen

F: Machen Sie Mal ein Beispiel

A: Beispiel gemacht mit Rechtecken und Baumstruktur

F: Und wie funktioniert hier jetzt eine Bereichsanfrage?

A: Anfragekreis gemalt mit den Rechtecken usw.

F: Und eine Nächste-Nachbarn Anfrage

A: Siehe Skript

Introduction to Artificial Intelligence

F: Ein wichtiges Thema in KI sind Suchverfahren. Was gibt es da?

A: Uninformiert (aufgezählt), Informiert, Startzustand, Nachfolgerfunktion Goal-Test

F: Was ist die Idee bei uninformatierten Suchverfahren?

A: "Best-First-Search", Schätze Distanz zum Ziel mit Heuristik

F: Wie funktioniert A*-Suche?

A: Kombiniere Pfadkosten und Restkostenschätzung, Vollständigkeit, Optimal, Optimal effizient bei optimistischer Heuristik

F: Wie kann man mit A* Breitensuche, Uniform Cost, Tiefensuche simulieren?

A: UCS mit Heuristik=0, Breitensuche mit uniformen Kosten und Heuristik=0, Tiefensuche ist unvollständig, die verwendete Heuristik ist also nicht admissible

F: Wie funktioniert die Suche bei Spielen?

A: Gegner mitmodellieren, MinMax Algorithmus, Min/Max Knoten, optimiere worst-case outcome

F: Wie kann man den Minimax mit AlphaBetaPruning verbessern?

A: siehe Buch

F: Machen sie mal ein Beispiel dazu

A: *Beispiel*