

Untersuchung der Parameter von Kerndichteschätzung für Transinformation

Verfahren zur Erkennung von Abhängigkeiten zwischen Attributen sind ein grundlegendes Werkzeug vieler Techniken zur Datenanalyse. Das Auswählen von unabhängigen Attributen ist eine zentrale Methode des maschinellen Lernens. Umgekehrt ist es für die Erkennung von Ausreißern nützlich, Attribute mit starken Zusammenhängen zu finden. Ein generelles Maß für solche Abhängigkeiten ist die *Transinformation*. Allerdings ist die Transinformation ein ideales Maß und kann auf Datensätzen nicht berechnet, sondern nur geschätzt werden.

Aufgrund der praktischen Relevanz von Transinformation existieren bereits viele Schätzer mit verschiedenen Eigenschaften bezüglich Qualität und Laufzeit. Ein beliebter Schätzer von Transinformation basiert auf *Kerndichteschätzung* von Wahrscheinlichkeitsdichten. Jedoch wurde bisher nicht hinreichend untersucht, wie sich die verschiedenen Möglichkeiten der *Kernform* und *Bandbreite* auf die Qualität der Schätzung auswirken. Zusätzlich interessant sind dabei die Auswirkungen auf die Laufzeit des Verfahrens. Die Forschungsfragen sind entsprechend:

- Welche Schätzqualität für Transinformation lässt sich durch verschiedene Kernformen und Bandbreiten erreichen? Wie verhält sich dies zu den in der Literatur gewählten Parametern?
- Wie verhalten sich die Konfigurationen bei verschiedenen Daten? Gibt es eine „optimale“ Konfiguration?
- Wie wirken sich die Parameter auf die Laufzeit aus? Gibt es Konfigurationen, die besondere Optimierungen erlauben?
- Welche Kompromisse zwischen Qualität und Laufzeit existieren?

Hieraus ergeben sich die folgenden Aufgaben:

- Implementierung des Schätzers für Transinformation basierend auf Kerndichteschätzung mit variabler Kernform und Bandbreite.
- Untersuchung der verschiedenen Kernformen und Verfahren zur Bestimmung von Bandbreiten.
- Evaluation der verschiedenen Konfigurationen anhand echter Datensätze sowie selbst entworfenen synthetischen Daten.

In dieser Arbeit beschäftigen Sie sich mit Schlüsselaspekten der Datenanalyse. Sie erlangen hierbei sowohl praktische Kompetenzen als auch theoretisches Verständnis von grundlegenden Analysetechniken. Für die Bearbeitung sind Programmierkenntnisse unerlässlich. Zusätzliches Vorwissen aus dem Bereich Datenanalyse ist nicht notwendig, aber hilfreich.

Ansprechpartner

Michael Vollmer
Am Fasanengarten 5

michael.vollmer@kit.edu
76131 Karlsruhe

+49 721 608 47336
Geb.: 50.34

Raum: 340