



Stelzinger Maximilian

Aufgaben

- Auswahl der Kriterien
- Beschreibung der Vor- und Nachteile von Preprocessing Methoden
- Auswahl und Beschreibung von Anwendungsbeispielen der Preprocessing Methoden in der Industrie

Ergebnisse/Erkenntnisse

- Lokale Verarbeitung wird in Zukunft eine wichtige Rolle einnehmen
- Die Zusammenarbeit von Personen unterschiedlicher Herkunft und die Kooperation zwischen Universität und Industrie ist wichtig für eine erfolgreiche Datenanalyse
- Neue Lern- und Entwicklungsstrategien müssen entwickelt werden
- Preprocessing Methoden arbeiten mit manchen ML Methoden / Klassifikationsalgorithmen besser zusammen

Motivation / Ausgangslage

Da immer mehr Daten auf der Welt entstehen und scheinbar die ganze Welt miteinander verbunden ist, wird in dieser Bachelorarbeit der Datenanalyse beziehungsweise dem Data Preprocessing besondere Aufmerksamkeit geschenkt. Der Datenaufbereitungsschritt benötigt ungefähr 80% des gesamten Zeitaufwands bei einem Data Mining Prozess. Die Schwierigkeit bei diesem Schritt besteht darin, aus den großen Datenmengen wertvolle Informationen oder Muster zu erkennen. Durch eine sorgfältig durchgeführte Datenaufbereitung, ergibt sich allerdings eine zuverlässigere Datenanalyse.

Da diese anwachsenden Datenmengen vielfach am Rande des Netzwerks entstehen, wird die Möglichkeit der lokalen Datenverarbeitung ebenfalls betrachtet. Aufgrund der Bandbreite des Netzwerks und der Sorge über die Privatsphäre von Daten, ist es oftmals nicht notwendig und ineffektiv Daten zu einer entfernten Cloud zu schicken.

Vorgehensweise / Gliederung der Arbeit

Zu Beginn dieser Arbeit werden die wichtigsten Begriffsdefinitionen, die mit der Datenanalyse in Zusammenhang stehen, dargelegt.

Anschließend folgt die Auflistung und Beschreibung der einzelnen Preprocessing Methoden. Data Preprocessing wird in der Wissenschaft, der Industrie und auch von Einzelpersonen verwendet, um eine zuverlässigere Datenanalyse zu erstellen. In dieser Arbeit werden die wichtigsten Data Preprocessing Methoden in industriellen Anwendungen anhand von fünf Beispielen vorgestellt. Diese Beispiele werden anschließend anhand von Kriterien, die vom Autor ausgewählt wurden, bewertet und gegenübergestellt.

Zuletzt wurde ein Fazit, ein Ausblick und eine Empfehlung erstellt.

Ergebnis

Das Data Preprocessing kann häufig nicht getrennt vom gesamten Data Mining / Datenanalyse Prozess bewertet werden. Grundsätzlich gilt, dass die Ergebnisse des Data Minings umso besser sind, je kompakter die Daten nach dem Preprocessing vorliegen. Jedoch muss dieser Schritt sorgfältig und präzise durchgeführt werden.

In industriellen Anwendungen steigen die Anforderungen immer weiter und mit Hilfe von Data Mining / Machine Learning können diese erfüllt werden. Bereiche in denen Data Mining / Machine Learning erfolgreich eingesetzt werden kann, sind beispielsweise Fehlerentdeckung und Qualitätsverbesserung. Des Weiteren wird die lokale Verarbeitung und Speicherung der Daten in Zukunft eine wichtige Rolle spielen.

Kontakt:

Institut für Konstruktionswissenschaften



und Technische Logistik

Prof. Dr. Manfred Grafinger

Tel: +43 1 58801 30723

E-mail: manfred.grafinger@tuwien.ac.at

WWW: <http://www.mivp.tuwien.ac.at>

 **Forschungsbereich**
Maschinenbau Informatik
 **und Virtuelle**
Produktentwicklung

Univ.-Prof. Dr.-Ing. Manfred Grafinger