

Smart Data Analyse

1. Beispiele für Smart Data Analyse

1.1. (versteckte) Korrelationen erkennen

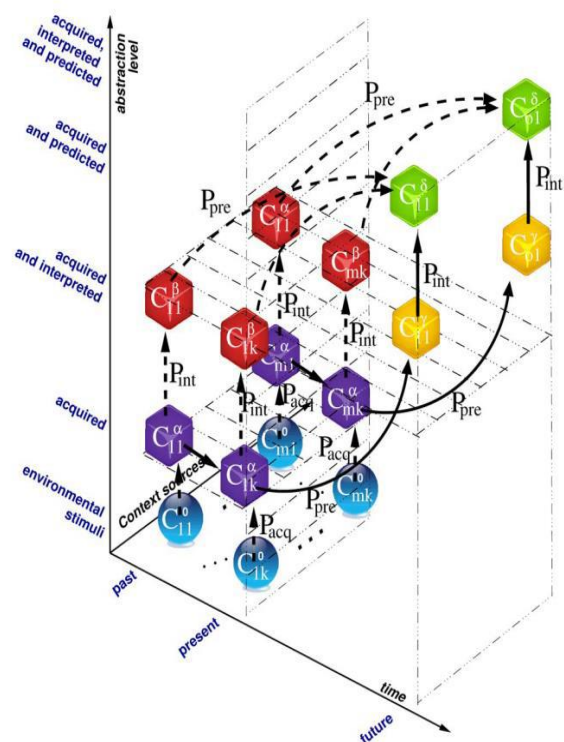
- Beispiel: Für eine bestimmte Auswirkung (z.B. Verschleiß einer Baugruppe, ungewöhnlich hoher Energieverbrauch, schlechte Fertigungsqualität) wird die Ursache gesucht.
 - o In einem großen Datensatz befinden sich Zusammenhänge von Ursache und Wirkung, die bisher unbekannt waren und unentdeckt geblieben sind.
 - o Mit Korrelationsverfahren (insb. Clustering, Itemset- und Sequence-Mining) können die Zusammenhänge entdeckt werden.
- Beispiel: Einsatzfelder einer Maschine sollen klassifiziert werden
 - o Aus einem Datensatz aus verschiedenen Maschinen im Betrieb wird eine Landkarte erzeugt. Die Cluster entsprechen den Klassen und können dann benannt werden.

1.2. Prozessabläufe erkennen

- Wiederkehrende Bedienungen bzw. Nutzungen bzw. Ereignisse und deren möglichen Auswirkungen (Ausfall einer Maschine/Kauf eines Gutes/Entnahme eines Teils vom Lager etc.) sollen erkannt werden. Dies kann z.B. für Optimierung von Prozesse oder subsequente Prädiktion von Ereignissen genutzt werden.
- Per z.B. Process-Mining werden aus einem Datensatz korrelierte Gründe (Sequenz von Aktivitäten etc.) für ein bestimmtes Verhalten ermittelt.

1.3. Prädiktion von Ereignissen

- Aus einem oder mehreren annotierten Datensätzen werden Zusammenhänge zwischen Ursache und Wirkung gelernt.
- Z.B. werden Zusammenhänge in zwei Datensätzen für Wetterbedingungen und Verkehrsdaten erlernt. Damit können per Wettervorhersage Verkehrssteuerungsempfehlungsvorausschauend gegeben werden.



2. Smart Data Analytics: Verfahrensschritte

2.1. Vorverarbeitung

- Rauschentfernung
- Merkmalsextraktion
- Strukturinformationen identifizieren
- Semantisches Wissen hinzufügen
- Erkennung und Entfernung fehlerhafter Daten (z.B. systematischer Fehler korreliert auf andere Größe)
- Normalisierung und Transcoding um einheitlich beschriebene Datensätze herzustellen
- Manuelle, halbautomatische oder automatische Annotation

2.2. Analyse

- Identifikation von Korrelationen
 - Deskriptive interaktive Analyse auf Basis statistischer Methoden
 - Visualisierung von Daten und Statistiken
 - Deskriptive Statistik
 - Korrelationsfunktionen
 - Statistische Tests
- Identifikation von komplexen Korrelationen und Prozessabläufen
 - Hochdimensionale automatisierte Analysen, besonders zeitlich-komplex auf Basis von Maschine Learning
 - Automatisierte Selektion von Variablen und Merkmalen
 - Unsupervised Feature-Learning und Deep-Learning
 - Automatisierte Identifikation von häufigen bzw. seltenen Mengen, Sequenzen und komplexen Prozessabläufen
 - Itemset-, Sequence- und Process-Mining
- Erkennung und Prädiktion
 - Entwicklung und Performance-Bewertung von Erkennungs- und Prädiktionsmodellen auf Basis von Maschine Learning
 - Lerner und Modelle
 - Clustering (K-means, ...)
 - Klassifikation (Decision Trees, SVMs, ...)
 - Regression (lineare und logistische)
 - Neuronale Netze
 - Markov-Modelle

- Petri-Netze
- Ensembles
 - Bagging, Random Forests, Boosting

2.3. Welche Daten können verarbeitet werden?

- Numerische Daten
 - Messpunkte bzw. Messreihen
- Nominale Daten
 - Kategorisierungen
 - Parameter und Einstellungen
- Semantische Daten
 - Triples und Quads, auch in Form von Zeitreihen

2.4. Beispielhafter Verfahrenseinsatz

- Energie:
 - Lastprognose zur Regelung von Stromflüssen in Verteilnetzen
- Industrie:
 - Finden unbekannter Ursachen für Maschinenfehler
 - Erkennung von Anomalien in Maschinenverhalten
 - Prädiktion von Wartung, Produktionsstörungen, Qualitätsprobleme, ...
 - Kostengünstiges Vermessen von neuen Sensoren auch mit hoher Cross-Sensitivität
- Smart City:
 - Erhebung von Stimmungsbildern in sozialen Medien
 - Prädiktion des Mikroklimas bspw. Schwebstoffe