

Visual Analytics zur Auswertung von Daten für die Industrie 4.0

Dieter Meiller, Florian Niewiera

Elektrotechnik, Medien und Informatik, Ostbayerische Technische Hochschule Amberg-Weiden

Zusammenfassung

Die Entwicklungen innerhalb des vorgestellten Projektes fokussieren die Ausarbeitung eines innovativen Analyse-Tools, welches für die Visualisierung und Optimierung von Produktionsprozessen und deren Daten eingesetzt werden soll. Für die Auswertung sollen Technologien der Bereiche Visual Analytics, Machine Learning und Clustering zum Einsatz kommen und weiterentwickelt werden. Die interaktive Darstellung wird durch Adaption und Anwendung von Algorithmen aus dem Schwarmverhalten realisiert und mit der automatisierten Auswertung gekoppelt. Durch den Einsatz des Systems soll es ermöglicht werden, die Effizienz, Qualität und Übersicht innerhalb eines Prozesses oder der Produktion zu steigern. Dies ist besonders im Kontext der derzeitigen Entwicklungen im Themenbereich Industrie 4.0 erstrebenswert.

1 Einleitung

Heutzutage werden Anlagen in modernen digitalen Fabriken bereits überwacht und ferngesteuert. Dabei werden große Datenmengen für das Monitoring und die Steuerung generiert und an die Maschinenschnittstellen übermittelt. Der nächste Schritt ist eine Analyse dieser Daten, um unter anderem zielgerichtete Vorhersagen von Produktionsvariablen zu ermöglichen.

Die im Rahmen dieses Projektes entstehende Lösung soll als eine sich abwechselnde Kombination aus maschineller Auswertung und interaktiver Visualisierung konzipiert und realisiert werden.

Innerhalb dieses Kurzberichtes werden die ersten Ergebnisse und umgesetzte Ideen, sowie geplante Arbeiten auf diesem Gebiet vorgestellt. Die Ergebnisse zielen insbesondere auf Anwender aus dem Bereich der kleinen und mittelständischen Unternehmen ab, welchen durch die Arbeiten und Resultate die Vorteile von Industrie 4.0 nutzbar gemacht werden sollen.

Veröffentlicht durch die Gesellschaft für Informatik e.V. 2016 in
S. Franken, U. Schroeder, T. Kuhlen (Hrsg.):
Mensch und Computer 2016 – Kurzbeiträge, 4. - 7. September 2016, Aachen.
Copyright © 2016 bei den Autoren.
<http://dx.doi.org/10.18420/muc2016-mci-0220>

2 Aktuelle Anforderungen im industriellen Umfeld

Der Wunsch und eine bestehende Notwendigkeit nach Optimierungen, Visualisierungen und der Handhabung von großen Datenaufkommen ist bei den Unternehmen ausgeprägter als jemals zuvor. Technologien, die mit der Einführung der Industrie 4.0 nutzbar gemacht werden können, könnten das Potential haben, bestehende Probleme zu lösen und Produktionsprozesse intelligenter und einfacher zu gestalten.

Der Kundensupport von Helpdesk Centern ist beispielsweise eines dieser Probleme. Eine innovative Speichermöglichkeit, sowie die automatische Analyse von Kundenanfragen könnten eine schnellere und verlässlichere Grundlage bilden, um eine, für jede Anfrage eindeutige, Lösung aufzuweisen. Hierbei könnte eine Zeitreduzierung bei der Bearbeitung der Kundenanfragen zu einer gesteigerten Effizienz, einer höheren Kundenzufriedenheit und weniger Aufwand für die Mitarbeiter führen. Ein weiteres denkbare Aufgabenfeld für die intelligente Analyse von Daten ist die Weiterentwicklung und Optimierung von Produktionsabläufen und Maschinenparametern. Dies würde beispielsweise dabei helfen, geringere Prozesszeiten oder materialschonendere Werkzeugeinsätze zu erreichen.

Ein anderes Anwendungsfeld ist die Auswertung von Sensordaten aus Produktionsprozessen: Heutzutage ist es möglich, für sämtliche Produktionsschritte eine Datenbank aufzubauen. Diese bietet dann die Möglichkeit, bereits im Vorfeld optimale Parameter für wechselnde und neue Prozesse auszuwählen, ohne diese in aufwändigen Vorversuchen oder durch erfahrene Profis ermitteln zu müssen. Mit dieser Hilfe wären auch Nicht-Experten in der Lage, Prozesse zu initialisieren, da sie hierbei aktiv vom System unterstützt werden, indem sie neben den notwendigen auch die nahezu optimalen Parameter aus der Datenbank erhalten würden. Ein Anwendungsszenario ist die langfristige Archivierung von großen Datenvorkommen sowie der anschließenden Analyse ausgewählter relevanter Daten in Kombination mit innovativen und informativen Visualisierungsmethoden. Ziel ist die Identifizierung unerkannter Prozesskorrelationen, die Darstellungsverbesserung von Produktionsübersichten oder eine einfachere Erkennung von unbekanntem Fehlerquellen.

3 Aktueller Stand und Umsetzung

Die Hauptaspekte des Analyseprozesses sind das Clustering und die Klassifizierung von Daten mittels unüberwachten maschinellen Lernens. Insbesondere sollen große Datenpools mit hochdimensionalen Daten und einer Vielzahl unterschiedlicher Datentypen verarbeitet werden. Um eine rechenintensive Dimensions-Reduktion zu vermeiden, werden die Daten in einem ersten Schritt mittels Modellen und Algorithmen aus den Bereichen der Schwarmvisualisierung und Partikelschwarm Optimierung (Maltese & Ombuki-Berman 2015) visualisiert, um Selbstorganisation diverser Datenobjekte zu ermöglichen. Dabei werden die Datenobjekte in einem zweidimensionalen Raum dargestellt, wobei dort die Achsen keine Objektmerkmale repräsentieren. Sich ähnelnde Objekte bewegen sich aufeinander zu, wobei die Ähnlichkeit von der Übereinstimmung der jeweiligen Merkmale bestimmt wird (Meiller 2015).

In der ersten Phase des Projekts wurden real aufkommende Produktionsdaten aus CNC-Fräsmaschinen zum Testen und Auswerten gesammelt. Diese Daten umfassen eine Vielzahl von Informationen wie beispielsweise Zeitstempel, Werkzeugpositionsdaten oder Geschwindigkeitsparameter. All diese Informationen werden in einer nahezu echtzeitfähigen NoSQL-Datenbank gespeichert. Im derzeitigen Projektverlauf wird eine Datenbank des Typs MongoDB genutzt (Dede 2013). Bei dieser dokumentenorientierten Datenbank ist es aufgrund der erlaubten Schemafreiheit möglich, unterschiedliche Datentypen und Formate in großen Mengen zu sammeln und zu speichern. Zusätzlich zu dynamischen Maschineninformationen werden auch statische Informationen wie Werkzeugparameter und Werkstückparameter in der Datenbank gespeichert. Für die zweite Phase wurde ein eigenes Framework "DataOcean@ISAC" (Abbildung 1) für die Visualisierung und Datenanalyse implementiert, welches kontinuierlich weiterentwickelt wird. Dieses stellt Prozessdaten und Parameter als individuelle, autonome Partikel in einer web-basierten Umgebung dar. Damit sich diese Datenstrukturen selbst organisieren können, wurden entsprechende Algorithmen implementiert, welche auf Schwarm-Modellen (Reynolds 1987) und Partikel Schwarm Optimierung basieren. Hierdurch finden sich Partikel zu kleinen Gruppen (Clustern) zusammen, die sich auf der Darstellungsfläche durch den „Ocean“ bewegen.

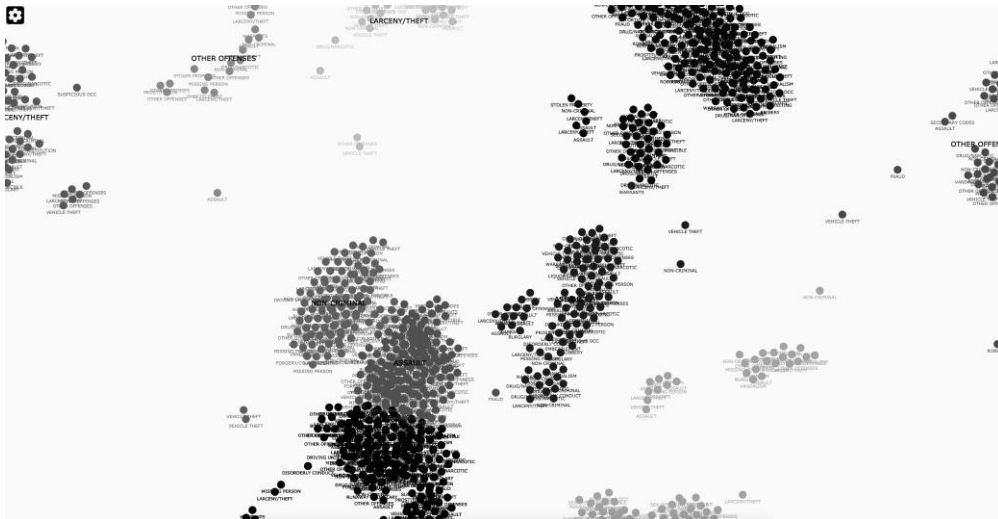


Abbildung 1: Clusterbildung während einer laufenden Instanz bei der Schwarmvisualisierung „DataOcean@ISAC“

Diese Cluster sind durch zusammengehörige Merkmale gekennzeichnet, welche sie gegenüber anderen Clustern abgrenzen. Als weiterer Schritt wurde die Visualisierung um einen neuartigen interaktiven k-Means (Shafeeq & Hareesha 2012) Algorithmus zur Bildung von Clusterzentren erweitert, der es dem Anwender ermöglicht, durch einfaches Klicken oder Tippen Schwerpunkte zu erschaffen, die das jeweilige Zentrum der Schwärme ausfindig machen und sich dort positionieren (Meiller & Niewiera 2016). Die Clusterzentren separieren die Schwarmteile voneinander und zeigen ihre größte Gemeinsamkeit an.

4 Zusammenfassung und Ausblick

Aktuelle Förderausschreibungen und Anfragen aus der Industrie zeigen, dass Bedarf zur Handhabung und Auswertung von großen Datenmengen besteht, um die Effizienz in verschiedenen Bereichen der Produktion zu steigern. Die ersten Ergebnisse dieses Projektes veranschaulichen, dass die Darstellung von Daten in Form von Partikelschwärmen in Kombination mit interaktiven Clusterverfahren eine vielversprechende Möglichkeit ist, aufschlussreiche Daten übersichtlich zu visualisieren und zu analysieren. Dieser Bericht erwähnt nur eine Möglichkeit der Realisierung des Clustervorgangs und der Schwarmbildung. Im Fortlauf des Projektes werden weitere Verfahren getestet und mit der aktuellen Implementierung verglichen: Da der K-Means Algorithmus nur (hyper-) sphärische Cluster ermöglicht, sollen zukünftig dichte-basierte Algorithmen eingesetzt werden (Briant 2007). Auch sollen andere Schwarm-Algorithmen getestet werden, wie beispielsweise Ameisen-Algorithmen (Colorni 1992).

Referenzen

- Dede, E. et al. (2013) *Performance evaluation of a mongodb and hadoop platform for scientific data analysis*, in Proceedings of the 4th ACM workshop on Scientific cloud computing, ACM, S. 13-20.
- Colorni, A. et al. (1992). *An Investigation of some Properties of an "Ant Algorithm"*, in PPSN, S. 509-520.
- Birant, D., & Kut, A. (2007). *ST-DBSCAN: An algorithm for clustering spatial-temporal data*, Data & Knowledge Engineering, 60(1), S. 208-221.
- Maltese, J., and Ombuki-Berman, B. (et al.) (2015). *Co-operative Vector-Evaluated Particle Swarm Optimization for Multi-Objective Optimization*, in Conf.proc. Computational Intelligence 2015, IEEE Symposium Series on IEEE.
- Meiller, D., & Niewiera F. (2016). *Data Visualization and Evaluation for Industry 4.0 using an interactive k-Means Algorithm*, in Conf. Proc. WSCG 2016 – 24th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision.
- Meiller, D. (2015). *Diving into the Data Ocean*, International Federation for Information Processing 2015, Interact 2015, Part IV, LNCS 9299, S.465-468.
- Na, Wenbo. (2015). *Research on well production prediction based on improved extreme learning machine*, in International Journal of Modelling, Identification and Control, S.238-247, 2015.
- Reynolds, C. W. (1987). *Flocks, Herds and Schools: A Distributed Behavioural Model*, in In ACM SIGGRAPH computer graphics 21 (4), S. 25-34.
- Shafeeq, A., & Hareesha, K.S.(2012). *Dynamic Clustering of Data with Modified k-Means Algorithm*, International Conference on Information and Computer Networks (CICIN 2012), IPCSIT vol.27, pp.221-225,2012