

ANALYSE



BIG DATA IN CPS / Mai 2017

INHALT

| | | |
|----|--|---|
| 1. | EINFÜHRUNG UND RELEVANZ FÜR CPS | 4 |
| 2. | AKTUELLER STAND IN WISSENSCHAFT UND WIRTSCHAFT | 5 |
| 3. | FORSCHUNGSOPTIONEN | 8 |
| 4. | LITERATURVERZEICHNIS | 7 |

Autoren



Offen im Denken

Shafiq Ur-Rehman
Stefan Gries
Volker Gruhn
Marc Hesenius
Julius Ollesch
Nils Schwenzfeier
Christian Wahl
Florian Weßling

Kontakt



Institut SIKoM+ | Bergische Universität Wuppertal

Rainer-Gruenter-Straße 21
42119 Wuppertal

E-Mail: kontakt@cps-hub-nrw.de
Telefon: +49 202 439 - 1026
Fax: +49 202 439 - 1037

www.cps-hub-nrw.de | www.facebook.com/cpshubnrw | www.twitter.com/cpshub

Gefördert durch



1 EINFÜHRUNG UND RELEVANZ FÜR CPS

Cyber-physische Systeme (engl. Cyber Physical Systems, CPS) zeichnen sich durch eine enge Verknüpfung physischer und informationstechnischer Prozesse aus [1]. Mit der enthaltenen Sensorik und ihrer Vernetzung bieten CPS das Potenzial, völlig neue Einblicke in Abläufe und Verhaltensmuster zu erhalten und daraus eine Optimierung der Prozesse abzuleiten [2]. Diese Systeme versprechen Fortschritte in verschiedenen Bereichen vom Gesundheitssektor bis hin zu intelligenten Stromversorgung. Genauso unterschiedlich sind auch die Daten und Datentypen, die in CPS verarbeitet werden müssen.

Mit dem Hypewort „Big Data“ sind ebenso große Erwartungen verknüpft. Neben der schieren Menge an Daten, die durch aktuelle Technologien produziert und verarbeitet wird, deutet der Begriff auch auf die Natur der Daten. Spielten vor einigen Jahren noch stark strukturierte Daten in der Informationsverarbeitung die Hauptrolle, so haben spätestens seit der Popularisierung von sozialen Netzwerken die unstrukturierten Daten die Oberhand. Dazu gehören Bilder, Videos, zusammenhängende Texte und Beziehungsstrukturen.

Durch die weitere Verbreitung von Sensorik in CPS wird die zu verarbeitende Datenmenge rapide ansteigen – schon heute steigt die Masse der Daten mit einer Rekordrate [3, 4]. Autonome Kamerasysteme in Fahrzeugen und Gebäuden, Messnetzwerke in Fabriken oder mit Sensoren ausgestattete Felder erfordern alle eine durchdachte und leistungsfähige Infrastruktur für Datenspeicherung und -analyse, die mit Big Data Technologien gegeben ist.

Moderne Fabriken überwachen Produktionsprozesse kontinuierlich und zeichnen Sensorwerte parallel zur Produktion auf. Oft werden Werte dabei mehrfach pro Sekunde erfasst und in Data Warehouses hinterlegt. Die hierbei entstehende Datenmenge ist immens und Bedarf einer Analyse, um Erkenntnisse daraus ableiten zu können. Dies ist insbesondere dann der Fall, wenn man vollständig autonom ablaufende Produktionsprozesse betrachtet, deren einzelne Produktionsschritte nahtlos ineinander übergehen. Das Aufdecken von Beziehungen zwischen einzelnen Sensorwerten und somit auch Produktionsschritten ist ein zentraler Aspekt der Big Data-Analyse.

2 AKTUELLER STAND IN WISSENSCHAFT UND WIRTSCHAFT

CPS verändern in vielfältiger Art und Weise unseren Alltag: vernetzte Transportsysteme, Roboter für betreutes Wohnen, intelligente Häuser und intelligente Stromnetze sind nur einige Beispiele. Durch CPS wird die physische Welt eng mit Sensoren, Netzwerkkommunikation und Computing-Komponenten integriert und diese Komponenten haben wiederum komplexe Interaktionen. Die Modellierung und das Management dieser Systeme ist entscheidend. Ziel dieser Komponenten von CPS ist es, eine hohe Datenqualität zu liefern [5, 6].

Nichtsdestotrotz kann bei der Vielfalt beteiligter Systeme und inhärenten Komplexität nicht von 100%ig korrekten Werten ausgegangen werden. Vielmehr besteht ein Großteil des Engineerings daraus, Daten zu plausibilisieren und in höherwertige Informationen zu verdichten. Pre-Processing Verfahren aus dem Umfeld von Big Data wie Integration, Cleansing, Redundanzeliminierung und letztlich multivariate statistische Analysen können Lösungen in solchen Szenarien bereitstellen [7, 8].

Wie beschrieben, werden ermittelte Daten und Sensorwerte oft in ein Data Warehouse eingespielt, dass eine Art zentrale Datenbank für Analysezwecke darstellt. Hierin werden die heterogenen Daten aus verschiedenen Quellen zusammengetragen und verdichtet, um eine gemeinsame Grundlage zur Verarbeitung und Analyse zu schaffen.

Für die Einbindung von CPS werden häufig Machine-2-Machine (M2M) Kommunikationsstandards genutzt. Diese Echtzeitkommunikation wird genutzt um sicherstellen, dass riesige Datenmengen rechtzeitig geliefert werden. Jedoch ist nicht nur die Zeit für die Übermittlung der Daten relevant, letztlich muss auch die Interpretation der Daten zeitnah geschehen – zum Beispiel, wenn es sicherheitsrelevante Daten sind. Hier können Big Data Infrastrukturen wie verteilte Dateisysteme und In-Memory Datenbanken genutzt werden, die auch große Datenbestände schnell genug durchsuchen und analysieren können. Hierbei unterscheidet man im Bereich Datenmanagement drei Ebenen: Programmiermodelle (z.B. MapReduce), Datenbanken (z.B. MongoDB) und Dateisysteme (z.B. Google File System, GFS) [8].

3 FORSCHUNGSOPTIONEN

Gegenwärtig stehen viele Organisationen vor Herausforderungen der Verteilung großer Datenmengen für eine effiziente Entscheidungsfindung und zur Verbesserung der Produktivität [9]. Dabei bestehen Forschungsoptionen im Spannungsfeld zwischen (in der Praxis oft mangelnden) Datenqualität und der Gefahr von ungewollten Enthüllungen und Sicherheitsfragestellungen. Aber auch rein technisch bestehen weitere Herausforderungen um Big Data im CPS-Umfeld einzusetzen.

Big Data für CPS-Anwendungen stellt zunächst neue Kapazitätsanforderungen für die Echtzeit-Handhabung großer Datenmengen [10]. Durch die verschärften zeitlichen Anforderungen von CPS sind klassische Analysewerkzeuge, die beispielsweise in einem Stapelverarbeitungsverfahren arbeiten, ausgeschlossen [11]. Auch Koumpouros [12] weist auf die Herausforderung hin riesige Datenmengen zu verarbeiten, die von Sensoren in Echtzeit im Gesundheitswesen gesammelt werden. Um die Handhabung solcher kritischen Daten zu bewältigen, braucht eine interdisziplinäre Methode, die Performance-Modellierung und Systemdesign mit der Implementierung von einzigartigen Algorithmen kombiniert [7]. Eine eher praktische Herausforderung für Unternehmen ist die Entwicklung bzw. Einführung geeigneter Computing-Plattformen [8].

Mangelnde Datenqualität ist zwar seit jeher ein Problemfeld, mit dem sich Big Data Spezialisten befassen mussten, es wird bekommt durch CPS aber eine neue Wichtigkeit. Denn die Datenqualität hat großen Einfluss auf den Mehrwert von Big Data. Datenqualität wird in der Regel in über Genauigkeit, Vollständigkeit, Redundanz und Konsistenz definiert. „Schlechte Daten“ verschwenden Übertragungs- und Speicherressourcen – können aber darüber hinaus auch zu falschen Entscheidungen führen. Das heißt man benötigt nicht nur die Daten selbst, sondern auch Informationen über die Qualität der vorliegenden Informationen. Anderenfalls lassen sich hieraus keine sicheren Rückschlüsse ziehen.

Sensordaten aus CPS sind fast immer fehlerbehaftet und müssen dementsprechend behandelt werden. Sharma et al. [7] stellen hier einige neue Forschungsansätze für CPS vor, etwa die automatische Analyse von Datenflüssen und Vorhersage von Mustern in CPS-Daten um diese nach und nach zu Plausibilisieren. Auch gibt es in CPS-Domänen wie der produzierenden Industrie Bestrebungen Datenformate und Übertragungstechnologien weiter zu standardisieren, was wiederum der Datenqualität zugutekäme [13].

Nicht zuletzt sind Datenschutz und -sicherheit zentrale Felder, welche für Big Data in allen Bereichen relevant sind. Insbesondere steht die Befürchtung im Raum, dass eine große Datensammlung - wenngleich anonymisierter Daten - Rückschlüsse auf Nutzer und personenbezogene Informationen zulässt [13]. An dieser Stelle herrscht weiterhin Forschungsbedarf um sichere und transparente Techniken zu entwickeln, die privaten wie kommerziellen Nutzern von Big Data Datenschutz gewährleisten. Nicht zuletzt sind hier auch internationale Abkommen und die jeweiligen Gesetzgeber gefordert. Aus CPS-Sicht besteht die Besonderheit darin, dass viele Akteure in den Systemen autonom agieren, zukünftige Regelungen also auch „maschinenlesbar“ umgesetzt werden müssen.

4 LITERATURVERZEICHNIS

- [1] E. A. Lee, „CPS foundations,” in Des. Autom. Conf. [DAC], 47th ACM/IEEE, S. 737-742, 2010, 2010.
- [2] M. Book, V. Gruhn und M. Kleffmann, Cyber Physical Systems - Potenzial und Kompetenzen in NRW, 2013.
- [3] R. Villars, C. Olofson und M. Eastwood, „Big data: What it is and why you should care,” IDC, 2011.
- [4] I. Hashem, I. Yaqoob, N. Anuar, S. Mokhtar, A. Gani und S. Khan, „The rise of “big data” on cloud computing: Review and open research issues,” Information Systems, 47, pp.98-115., 2015.
- [5] K. Srinivasa, G. Siddesh und K. Mishra, „Influence of Big Data on Cyber- Physical Systems,” in Cyber-Physical Systems: A Computational Perspective , G. Siddesh, G. Deka, K. Srinivasa und L. Patnaik, Hrsg., Chapman and Hall/CRC., 2015, pp. 515-524.
- [6] P. Derler, E. A. Lee und A. S. Vincentelli, „Modeling Cyber-Physical Systems,” Proceedings of the IEEE, 100 [1], , 2012.
- [7] A. Sharma, F. Ivančić, A. Niculescu-Mizil, H. Chen und G. Jiang, „Modeling and analytics for cyber-physical systems in the age of big data,” ACM SIGMETRICS Performance Evaluation Review, 41[4], , pp. 74-77, 2014.
- [8] H. Hu, Y. Wen, T. Chua und X. Li, „Toward scalable systems for big data analytics: A technology tutorial,” IEEE Access, Nr. 2, pp. 652-687, 2014.
- [9] N. Kshetri, „The emerging role of Big Data in key development issues: Opportunities, challenges, and concerns,” Big Data & Society, 2014.
- [10] L. Zhang, „A framework to specify big data driven complex cyber physical control systems,” in Information and Automation [ICIA], 2014.
- [11] L. Zhang, „A framework to model big data driven complex cyber physical control systems,” in Automation and Computing [ICAC] , 2014.
- [12] Y. Koumpouros, Big Data in Healthcare. Healthcare Administration: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications, 2014.
- [13] L. Wang, M. Törngren und M. Onori, „Current status and advancement of cyber-physical systems in manufacturing,” Journal of Manufacturing Systems, Nr. 37, pp. 517-527, 2015.
- [14] M. Chen, S. Mao und Y. Liu, „Big data: A survey,” Mobile Networks and Applications, Nr. 19, pp. 171-209, 2014.
- [15] R. Viereckl, A. Koster, S. Jursch und D. Ahlemann, „Connected Car Study 2015: Racing ahead with autonomous cars and digital innovation,” 2015.

Gefördert durch

