Developing a Decision Support System leveraging Distributed and Heterogeneous Sources: Case-Based Reasoning for Manufacturing Incident Handling

M. van der Pas[1,2,*]

[1]Eindhoven University of Technology, Department of Industrial Engineering & Innovation Sciences, Eindhoven 5600MB, The Netherlands

[2]Semaku B.V., Eindhoven 5617BC, The Netherlands

Abstract

Case-Based Reasoning is a proven method to provide decision support in a manufacturing context. However, data and knowledge relevant for the case representation is often spread over distributed sources, leading to challenges in the case representation and retrieval. Those challenges require different techniques that this PhD project aims to develop. Techniques for data collection and integration during the case representation, as well as similarity measurement during case retrieval. This paper describes the motivating problem, the research methods, and the current state and future plans.

Keywords

Incident Handling, Traceability, Case-Based Reasoning, Semantic Web, Knowledge Graph, Event Graph

1. Introduction

One of the challenges identified for Case-Based Reasoning (CBR) research is the acquisition of cases from heterogeneous and distributed data sources [1]. This challenge certainly also applies to complex manufacturing environments, where CBR can be applied to assist engineers with the handling of quality incidents. In case customers have an issue with a device, they might initiate a (quality) complaint at the company that produced the device. The company should then analyse the complaint and take suitable measures, like containment and corrective action. This complaint handling process is taking an increasing amount of effort, caused by the increasing product and production process complexity. Especially in the semiconductor industry [2], which is the main motivating use case for this project. During the handling process there are already some commonality checks done to find historic complaints related to a new complaint. However, due to the challenges described below only to limited extend.

In manufacturing companies there are many data available about the products and their production process, which can be used to describe a (complaint) case. These data are often spread over many different systems. Not only because of the different types of data that are of interest, but also because of the complexity of the semiconductor production process, consisting

ICCBR DC'22: Doctoral Consortium at ICCBR-2022, September, 2022, Nancy, France

*Corresponding author.

m.c.a.v.d.pas@tue.nl (M. v. d. Pas)

of many different production steps spread over multiple facilities. At the same time it is costly to index all data in a central case base. Therefore, a system that supports engineers with identifying similar cases (historic complaints), will have to deal with distributed and heterogeneous sources. These characteristics introduce specific challenges, and this research will try to solve some of them, focusing on case representation and retrieval phase of the CBR cycle [3].

2. Research Plan, Objectives, and Approach

The goal of this project is to develop a (decision support) system that assists engineers in the handling of manufacturing incidents by providing cases similar to the new case they have at hand. The system should leverage data and knowledge from distributed and heterogeneous sources. As such, answer the following research question: **How to identify related quality incidents in a manufacturing environment with distributed and heterogeneous data and knowledge sources?**

2.1. Sub-projects

The project is divided into four sub-projects. The topic of the first sub-project is the development of a general framework for CBR-based decision support leveraging distributed and heterogeneous sources. The other sub-projects focus on specific components in this framework, for case representation and retrieval. More details about the evaluation of the components and system can be found in subsection 2.2.

I. Framework

How to design a system to find related quality incidents in a manufacturing environment with distributed and heterogeneous data and knowledge sources?
The first sub-project focuses on a framework to support the CBR-cycle, more specifically the case representation based on distributed sources. In distributed decision support systems [4] and CBR systems for knowledge management [5, 6, 7] it is common to use an agent-based approach. Chaudhury et al. [8] proposed a solution for CBR with distributed storage of cases. Similarly, Camarillo [9] proposed a knowledge management framework using CBR in an industrial context. However, both focus on combining cases from distributed case bases, while this research focuses on gathering data from distributed systems for case representation. Therefore, the system should be able to collect and integrate data from heterogeneous sources to describe a new case, refine and enrich the case representation, and provide similar cases back to the user. An overview of the main steps can be found in Figure 1. The system will to a large extend rely on Semantic Web Technologies, which are proven to be suitable for combining data and knowledge driven approaches [10]. The main components of this architecture are investigated as part of the other sub-projects. Once the components are developed, the framework will also be implemented and evaluated with the (source) systems in a case study.

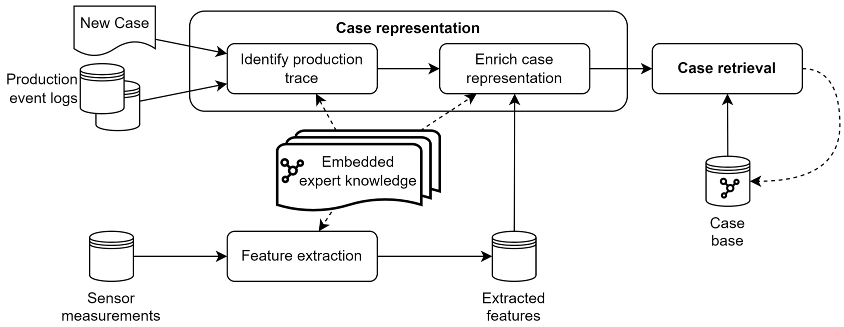

Figure 1: Main steps of the CBR-based decision support for quality incidents.

II. Case Representation: Trace identification

How to identify the entities and events that were involved in the production process of a case?
During the production process often multiple case identifiers are used and production batches are split and merged. For example, when multiple semi-finished goods are assembled into one device. This results in multi-dimensional event data, and introduces fuzziness and uncertainty in the trace. Therefore, it is a challenge to collect relevant data and information to build a case representation. The production trace [11] can serve as the foundation for the case representation. The production trace describes the production process of a device and consists of production events and related entities. It can be represented as an event graph, which is well suited to represent multi-dimensional event data [12], combining the time and relation dimension. In comparison to Esser and Fahland [12], this research aims to enrich the event graphs with knowledge encoded in ontologies. For example, Lee and Park [13] improved the traceability using information about the bill of materials. The developed technique should be able to generate an event graph consisting of events and entities on different levels of aggregation, which describe the production trace for the case at hand.

III. Case Representation: Data integration

How to integrate data from distributed and heterogeneous sources?
The first step of the representation phase introduced by Finnie and Sun [14] is to construct a case description. This research will focus on event and sensor data for describing the case, which are common in the manufacturing domain [15]. As it is costly to integrate and store all sensor data centrally, a method is required to reduce the data volume and dimensionality to be able to integrate it into one case representation. Wang et al. [16] propose to aggregate sensor data to events and integrate those events with graph structured context data. Similarly, the system developed by Gundersen et al. [17] abstracts sensor data to events using pattern matching. Those events are subsequently used in CBR to find similar situations from the past. In a similar way, this research aims to use machine learning techniques to extract features from a series of data points[18]. Those data points are generated by sensors on the production equipment and describe the production process of one device. The extracted features should correlate to quality incidents.

IV. Case Retrieval: Similarity Measurement

How to find similar quality incidents based on heterogeneous incident descriptions leveraging domain knowledge?
The goal of this sub-project is to develop a technique that can be used to retrieve cases similar to a novel case. Camarillo et al. [6] use a predefined set of attributes to describe the case and its context. To deal with the distributed and heterogeneous sources, a more flexible case representation format is required. Therefore, this research aims to use RDF (Resource Description Framework) knowledge graphs and corresponding graph-based similarity measurement techniques. Zhang et al. [19] also used knowledge graphs, but conclude that more work needs to be done on the similarity and knowledge reasoning. Furthermore, domain knowledge is required to conduct proper similarity measurement. This knowledge can be represented by taxonomies or ontologies [6, 20, 21, 5, 22]. There are various standards for describing ontologies/taxonomies using RDF, for example OWL[1] and SKOS[2], which as such can be integrated in the case representation.

2.2. Evaluation of the system

The sub-projects will result in different components of the system, which require different methods and data sets to validate and evaluate their functioning.

Sub-project II The data collection and integration solution can be validated using simulated or actual (from a semi-conductor use case) production events. However, there exists no data set with 'known-good production traces'. Therefore, the aim is to identify and reconstruct a number of traces for a validation data set, with the help of engineers.

Sub-project III The feature extraction technique can be evaluated using sensor data collected from equipment that is used in the production process. After most production steps a quality check is done. The results of this check can be used to validate if the derived features are indeed correlated to quality incidents.

Sub-project IV The aim is to evaluate the graph-based case comparison technique, using a data set from the semi-conductor use case described in the introduction (handling of customer complaints). The data set consists of historic complaints, traceability data (production events from different Manufacturing Execution Systems (MES)), and data from Product Life cycle Management (PLM) systems. In practice there are already some commonality checks done by the engineers, which can be used as a benchmark.

The integrated system The preferred method of evaluating the integrated system, which integrates the techniques developed in the sub-projects, is to combine the data sets used to evaluate those sub-projects. However, the challenge is that only a very small portion of produced devices result in a complaint, which in turn are only detected months to years after production. Therefore, it will be difficult to construct a data set with relevant sensor and complaint data. A

[1] https://www.w3.org/TR/owl2-primer/
[2] https://www.w3.org/TR/skos-primer/

possible solution is to use simulated data, based on the data collected before. An alternative is to find a different use case, in which quality incidents occur with higher frequency, such that it is possible to construct a data set that contains related sensor and incident data.

3. Progress Summary

At the time of submission, most work is done on defining the framework for data collection and integration (sub-project I) in the context of the MAS4AI project. In future, the developed framework will be evaluated in a case study, using data and information sources from a manufacturing company. Next to the work on sub-project I, two case studies are in progress which look into modelling event graphs based on manufacturing events, and feature extraction from sensor data, respectively contributing to sub-project II and III. Both focus on a specific step in the semi-conductor production process.

Acknowledgements

This project is supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 957204, the project MAS4AI (Multi-Agent Systems for Pervasive Artificial Intelligence for assisting Humans in Modular Production).

References

[1] A. K. Goel, B. Diaz-Agudo, What's Hot in Case-Based Reasoning, in: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, 2017, pp. 5067–5069.

[2] J. Aelker, T. Bauernhansl, H. Ehm, Managing complexity in supply chains: A discussion of current approaches on the example of the semiconductor industry, Procedia CIRP 7 (2013) 79–84. doi:10.1016/J.PROCIR.2013.05.014.

[3] A. Aamodt, E. Plaza, Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches, AI Communications 7 (1994) 39–59. doi:10.3233/AIC-1994-7104.

[4] K. Sycara, D. D. Zeng, Multi-Agent Integration of Information Gathering and Decision Support, in: 12th European Conference on Artificial Intelligence, John Wiley & Sons, Ltd., 1996, pp. 549–553.

[5] P. Reuss, K.-D. Althoff, A. Hundt, W. Henkel, M. Pfeiffer, Multi-Agent Case-Based Diagnosis in the Aircraft Domain, in: Proceedings of the ICCBR 2015, Frankfurt, 2015, pp. 43–52.

[6] A. Camarillo, J. Ríos, K.-D. Althoff, Knowledge-based multi-agent system for manufacturing problem solving process in production plants, Journal of Manufacturing Systems 47 (2018) 115–127. doi:10.1016/j.jmsy.2018.04.002.

[7] W. Luís, J. Carlos, E. Ferreira, F. Gonzalez, C. Gomes, R. M. Lorenzo, W. Luı´s, L. Mikos, C. Espı´ndola, E. Ferreira, A combined multi-agent and case-based reasoning approach to support collaborative nonconformance problem solving in the thermoplastic injection moulding process, https://doi.org/10.1080/09511920903440321 23 (2010) 177–194. doi:10.1080/09511920903440321.

[8] S. Chaudhury, T. Singh, P. S. Goswami, Distributed fuzzy case based reasoning, Applied Soft Computing 4 (2004) 323–343. doi:10.1016/J.ASOC.2003.10.003.

[9] A. Camarillo González, Knowledge Management framework based on an Agents Network to support Continuous Improvement in Manufacturing integrating Case-Based Reasoning and Product Lifecycle Management, Ph.D. thesis, Universidad Politécnica de Madrid, 2018. doi:10.20868/UPM.thesis.53989.

[10] B. Steenwinckel, D. De Paepe, S. Vanden Hautte, P. Heyvaert, M. Bentefrit, P. Moens, A. Dimou, B. Van Den Bossche, F. De Turck, S. Van Hoecke, F. Ongenae, FLAGS: A methodology for adaptive anomaly detection and root cause analysis on sensor data streams by fusing expert knowledge with machine learning, Future Generation Computer Systems 116 (2021) 30–48. doi:10.1016/j.future.2020.10.015.

[11] R. Schuitemaker, X. Xu, Product traceability in manufacturing: A technical review, Procedia CIRP 93 (2020) 700–705. doi:10.1016/J.PROCIR.2020.04.078.

[12] S. Esser, D. Fahland, Multi-Dimensional Event Data in Graph Databases, Journal on Data Semantics 10 (2021) 109–141. doi:10.1007/S13740-021-00122-1/TABLES/6.

[13] D. Lee, J. Park, RFID-based traceability in the supply chain, Industrial Management and Data Systems 108 (2008) 713–725. doi:10.1108/02635570810883978/FULL/PDF.

[14] G. Finnie, Z. Sun, R5 model for case-based reasoning, Knowledge-Based Systems 16 (2003) 59–65. doi:10.1016/S0950-7051(02)00053-9.

[15] H.-N. Dai, H. Wang, G. Xu, J. Wan, M. Imran, Enterprise Information Systems Big data analytics for manufacturing internet of things: opportunities, challenges and enabling technologies Big data analytics for manufacturing internet of things: opportunities, challenges and enabling technologies, Enterprise Information Systems 14 (2020) 1279–1303. doi:10.1080/17517575.2019.1633689.

[16] H. Wang, P. Nguyen, J. Li, S. Kopru, G. Zhang, S. Katariya, S. Ben-Romdhane, GRANO: Interactive graph-based root cause analysis for cloud-native distributed data platform, Proceedings of the VLDB Endowment 12 (2019) 1942–1945.

[17] O. Erik Gundersen, A. Aamodt, P. Skalle, A Real-Time Decision Support System for High Cost Oil-Well Drilling Operations PhD project View project Reproducible AI View project, Ai Magazine 34 (2012) 21. doi:10.1609/aimag.v34i1.2434.

[18] F. J. Baldán, J. M. Benítez, Multivariate times series classification through an interpretable representation, Information Sciences 569 (2021) 596–614. doi:10.1016/J.INS.2021.05.024.

[19] Y. Zhang, X. Liu, J. Jia, X. Luo, Knowledge representation framework combining case-based reasoning with knowledge graphs for product design, Computer-Aided Design and Applications 17 (2020) 763–782. doi:10.14733/CADAPS.2020.763-782.

[20] G. Cândea, S. Kifor, C. Constantinescu, Usage of Case-based Reasoning in FMEA-driven Software, Procedia CIRP 25 (2014) 93–99. doi:10.1016/J.PROCIR.2014.10.016.

[21] B. Chebel-Morello, M. K. Haouchine, N. Zerhouni, Reutilization of diagnostic cases by adaptation of knowledge models, Engineering Applications of Artificial Intelligence 26 (2013) 2559–2573. doi:10.1016/J.ENGAPPAI.2013.05.001.

[22] S. H. El-Sappagh, M. Elmogy, Case Based Reasoning: Case Representation Methodologies, IJACSA) International Journal of Advanced Computer Science and Applications 6 (2015).