Proceedings of the 51st Hawaii International Conference on System Sciences | 2018

Towards Service-Oriented Cognitive Analytics for Smart Service Systems

Robin Hirt
Karlsruhe Institute of Technology (KIT)
hirt@kit.edu

Niklas Kühl
Karlsruhe Institute of Technology (KIT)
kuehl@kit.edu

Björn Schmitz
Karlsruhe Institute of Technology (KIT)
bjoern.schmitz@kit.edu

Gerhard Satzger
Karlsruhe Institute of Technology (KIT)
gerhard.satzger@kit.edu

Abstract

The development of analytical solutions for smart services systems relies on data. Typically, this data is distributed across various entities of the system. Cognitive learning allows to find patterns and to make predictions across these distributed data sources, yet its potential is not fully explored. Challenges that impede a cross-entity data analysis concern organizational challenges (e.g., confidentiality), algorithmic challenges (e.g., robustness) as well as technical challenges (e.g., data processing). So far, there is no comprehensive approach to build cognitive analytics solutions, if data is distributed across different entities of a smart service system. This work proposes a research agenda for the development of a service-oriented cognitive analytics framework. The analytics framework uses a centralized cognitive aggregation model to combine predictions being made by each entity of the service system. Based on this research agenda, we plan to develop and evaluate the cognitive analytics framework in future research.

1. Introduction

As companies strive to digitize their business, data is produced in vast amounts [1, 2]. However, this data is usually generated and controlled by different entities of a smart service system [3]—for instance, by different suppliers of a supply chain. To derive insights from this distributed data—e.g., to make predictions about the timely delivery of a part—a centralized aggregation of these different data sources is required [4]. Bringing data together opens up manifold opportunities to optimize business processes and entire service systems. However, building centralized analytical solutions requires to process the data of each entity and to understand possible dependencies among it. Moreover, this data may be confidential, may differ in structure and format and may be generated at different points in time.

To address these challenges, we develop a distributed analytics framework based on a cognitive paradigm. The framework allows to combine data from different business entities of a service system to find holistic insights in it. We describe this framework in detail in the remainder of this paper, which is structured as follows: We first identify challenges of developing analytics solutions in smart service systems (section 2). We then review related work and show first propositions on how to cope with the identified challenges (section 3). With this basis, we present our approach in detail (section 4). Comparing the state of the art and our presented approach, we derive a research agenda for cognitive analytics in smart service systems (section 5) and identify research questions that need to be addressed in future research. We finally conclude with an outlook (section 6).

2. Problem Description

According Kaplan et al. [5], data is produced by different entities of a smart service system. Compared to analyzing a data source of one single entity, the aggregation and simultaneous analysis of data sources owned by more than one entity yields considerably more potential to gain insights that have an economic value.

Thus, an aggregation of distributed data sources within a service system is required. Smart service systems are "service systems that are specifically designed for the prudent management of their assets and goals while being capable of self-reconfiguration to ensure that they continue to have the capacity to satisfy all the relevant participants over time" [3, p. 33].

The development of centralized analytical solutions for such systems suffers from three groups of challenges (Figure 1): *Technical* challenges concern the processing of data, the development and continuous improvement of a flexible and modular architecture are of major importance [6]. *Organizational* challenges concern issues of trust and data confidentiality that impede entities from their sharing data [7]. *Algorithmic* challenges concern the data structure as well as the performances and robustness of algorithms.

For each group of challenges, we discuss issues that impede the development of analytical solutions in

URI: http://hdl.handle.net/10125/50090
ISBN: 978-0-9981331-1-9

HICSS

distributed systems, which is especially relevant to utilize the required connectivity in smart service systems.

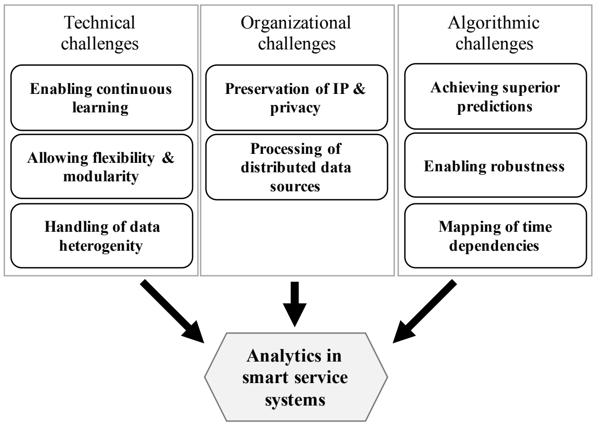

Figure 1. Challenges of analytics in smart service systems

Organizational challenges split in two types of problems: Data is generated by different entities and at different physical or virtual locations [8]. Thus, the *processing of distributed data sources* represents a key challenge for performing a holistic analysis. Considering the example of a smart factory [9], data is gathered across distributed production lines that can comprise thousands of individual sensors. Additionally, sharing data among entities can disclose private information (e.g., process knowledge, production recipes) and intellectual property (IP) [10]. Companies may not be willing to share private data to *preserve IP and privacy*, as they may fear confidentiality breaches or disclosure of IP [11].

Besides organizational challenges, different technical ones arise: Data itself can come in different formats and sizes (*data heterogeneity*). It can be structured (e.g., sensor outputs) or unstructured (e.g. manually written service reports) and stored in different data types (e.g., image, audio). Consequently, each source requires a dedicated processing, which can then reveal different insights [12]. Their combination and aggregation is complex and time-consuming [13]. In conjunction with different data sources, businesses need to process data on different hierarchies and abstraction levels (e.g., via field buses on shop floor or via enterprise resource planning systems on a management level) [14]. Thus, an analytics architecture needs to be *flexible and modular*. Since business processes change over time, analytics solutions have to be flexibly designed to *continuously learn* to adjust to these changes [6].

Distributed analytics also yields different algorithmic challenges: As we aim to combine different data sources to find holistic insights, a superordinate comprehensive analysis should also be *superior in terms of prediction performance* compared to the analysis of single data sources. However, aggregating data that comes from different sources makes a viable prediction difficult. Problems concern the performance of algorithms, their *robustness* and reliability [15]. Additionally, data is produced during different steps of business processes which may be run sequentially or in parallel. Therefore, analytics should be able to cope with input data that has a *time-dependency.*

To solve these problems, we propose a cognitive framework that combines information from independent, distributed subordinate entities by making use of different layers of machine learning. Hereby, we utilize cognitive learning: Every distributed data source is first processed by a machine learning predictor and only the prediction output is transferred to a centralized entity. The centralized entity then processes the prediction outputs of all predictors of distributed entities to make an aggregation based on a cognitive paradigm. Since the framework communicates prediction outputs of entities instead of their raw data, the framework allows to overcome challenge such as data heterogeneity, privacy and velocity amongst others.

3. Related Work

We identify relevant literature about processing distributed data alongside the mentioned challenges in four fields or research: fog computing, service-oriented decision support systems, complex event processing and privacy-preserving data mining. Table 1 gives an overview of these research fields in regards to the described challenges.

In the area of distributed sensor networks, fog computing promises to tackle the problem of the data transfer bottle neck as well as leveraging unused computational capabilities of sensor hubs [16] . Driven by the computational advancements of sensor nodes, fog computing propagates a decentralized, low-latency model for computing on-device and extending business logic onto these nodes. Similar to our approach, distributed entities perform autonomous predictions. However, fog computing does not propagate any directed, ordered way of orchestrating these calculations into an abstract calculation. Its focus lays on processing data on premise and dynamic process logistics throughout distributed entities that are not necessarily interconnected.

In contrast to fog computing, the framework of service-oriented decision support systems promotes an architecture of cloud-based analytics entities [17] (organizational & business perspective). However, it describes systems from a high-level, organizational view and demands a strong communication and coordination of data and processing.

Complex Event Processing (CEP), as a layer built on top of event-driven architecture [18], strives to handle a timely and continuous processing of big data streams (data and performance perspective). Whereas CEP can handle high volume data streams, it does not focus on distribution of data processing or IP and privacy preservation.

Table 1. Comparison of distributed data analytics research domains to proposed approach alongside challenges

	Robustness	Prediction performance	Time dependencies	Data heterogeneity	Flexibility and modularity	Continuous learning	Distributed data sources	IP and privacy preservation
Fog computing	○	◐	○	●	●	○	●	○
Service-oriented decision support	○	○	◐	○	●	○	◐	○
Complex event processing	○	◐	●	◐	◐	○	●	○
Privacy-preserving data mining	○	○	○	○	○	○	◐	●
Service-oriented cognitive analytics	●	●	●	●	●	●	●	●

○=Not addressed, ◐=Partially addressed, ●=Fully addressed

The field of privacy-preserving data mining aims to build accurate models without disclosing precise information about an individual data record (legal perspective). Agrawal & Srikant [19] classify privacy-preserving methods into two categories: value-class membership and value distortion return. In case of the value-class membership, the data is disjoint into more than one set and therefore discretize sensitive information. The main principle is that one entity never has all information about a given state of the data space, but enough to mine knowledge. The value distortion uses a function to mask the actual value by mixing it either with a uniform or a Gaussian random value and therefore make the real value unreadable but keeping he original patterns in data. Thus, for each value x_i a value $x_i + r$ is returned instead, whereas r represents a random variable. In contrast to these two approaches this work focuses on dividing the analytical process itself and, therefore, aggregate more than one value with an abstract prediction output.

Thus, compared to the current state of the art, this approach distinguishes itself from previous research along the described challenges by combining a distributed analytics architecture based on a service-oriented paradigm and a centralized cognitive learning entity.

4. Service-Oriented Cognitive Analytics

To give an overview of the proposed approach, we first illustrate the general architecture that the service-oriented cognitive analytics framework is based on. Then, we describe the concept of a cognitive aggregation method that represents a key contribution of the proposed approach.

4.1. Architecture

First, we describe the architecture of the proposed service-oriented cognitive analytics framework that enables researchers and businesses to perform comprehensive, robust and IP-preserving predictions on variations of heterogeneous, distributed data sources and subordinate predictors. We distinguish between two types of entities and corresponding predictions: subordinate entities and cognitive entities both of which can be considered different entities of a service system. Every subordinate entity is autonomous and generates data which is processed by a subordinate predictor. An example of a subordinate entity is an assembly machine the condition of which is monitored by sensors that measure temperature, rotational speed or similar parameters. In this case, the sensor output represents a data source which is processed by a subordinate predictor. This subordinate predictor could, for instance, predict if the machine is in good condition, or not.

The output of this subordinate predictor is sent to the cognitive entity—alongside with the predictions of other subordinate entities (e.g. other machines or pre-assembly process entities). The cognitive entity then aggregates all subordinate predictor outputs to make a cognitive prediction. The cognitive prediction represents an insight that is latently contained in the distributed data of the service system. In connected smart factories, the aggregation of information from different entities can reveal valuable insights: For instance, the analysis of data about machine conditions (manufacturer), the delivery capability of parts

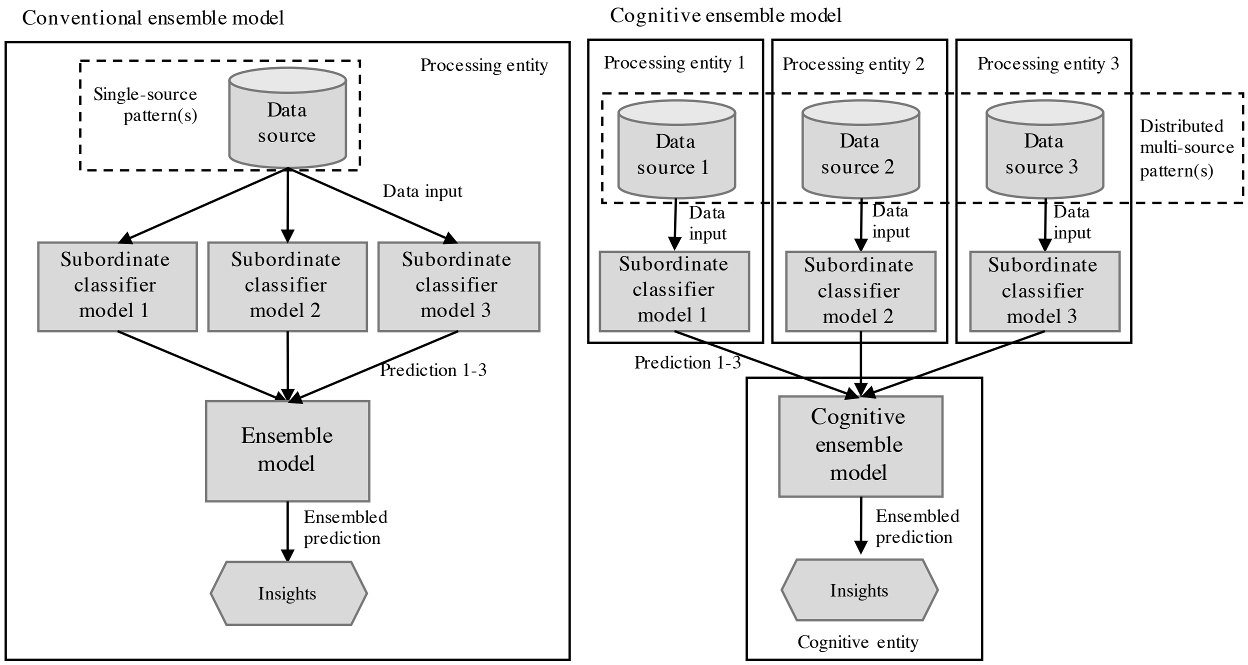

Figure 2. Comparison of conventional and cognitive ensemble model

(supplier), and the demand for the assembled products (customer) can help to optimize production plans and the timely purchasing of parts. Hereby, the cognitive entity learns about significance, relevance and validity of each subordinate prediction and their interdependencies.

Since subordinate entities send prediction outputs instead of raw data to the cognitive entity, their information exchange is IP- and privacy-preserving. Furthermore, the framework ensures that the data is being analyzed at the location where it is produced (sub-predictors) and only the prediction outputs of this analysis are sent. Thus, it is not necessary to transfer huge data streams to make a centralized analysis.

Additionally, every subordinate predictor can be tailored to the analysis of a specific data type (e.g., visual processing for images, natural language processing for texts). The output of a subordinate predictor is always of the same type: a prediction towards a target variable. This facilitates the creation of prediction models, as it focusses on one type of data. In figure 2 we depict a possible architecture of a cognitive ensemble model with corresponding data and information flow. The cognitive prediction itself is based on machine learning and flexibly combines the subordinate predictor outputs without knowing their underlying meaning. No further manual calibration of the cognitive entity is necessary due to the stacked generalization approach.

4.2. Cognitive Aggregation

A key element of a centralized, accumulating analytics model is the aggregation method. This research aims to build a directed data processing framework that aggregates distributed prediction outputs by a cognitive learning layer. Hereby these predictions, that are based on distributed, independent, heterogeneous data sources, are processed by adding more than one layer of machine learning. In the following, the foundations of cognitive learning are laid out.

According to Modha et al., cognitive computing *"aims to develop a coherent, unified, universal mechanism inspired by the mind's capabilities"* [20, p. 62]. One key aspect of cognition depicts the aggregation of several distinguishable input sources. Thus, the question arises, how to combine these sources. A similar problem faces the research field of multimodal fusion, which aims to analyze multimedia content, such as videos, by first separately processing visual and audio content and then aggregating it [21, 22]. Similarly, in this research, we aim to realize cognitive learning by utilizing ensemble learning, a technique that is typically used to increase the prediction performance. Although we focus on ensemble learning, other mechanisms could also be promising to be evaluated [23]. Ensemble learning makes an aggregated prediction by combining predictions of several models (e.g., SVNs, random forest) that have been trained on the same input data. This combination can result in a more accurate aggregated prediction [24]. These aggregating classifiers are also called ensemble classifiers. Ensemble classifiers improve prediction performance by defining an ensemble classification function that minimizes the uncorrelated error among all classifiers. Džeroski & Ženko [25] state that ensemble classification approaches perform better than the selection of a best classifier in a set of single classifiers.

According to Todorovski et al. [26], ensemble classifiers are built in two steps: First, the subordinate classifiers have to be designed and implemented. Once they are ready, an ensemble decision function has to be defined that decides which aggregated prediction is made based on the underlying subordinate classifiers' predictions. They differentiate between three concepts of ensemble functions: voting (Bagging, Boosting), cascading (iterative classification and cascading enrichment of previous prediction) and stacked generalization (applying another layer of ML to the predictions).

Popular voting algorithms are Bagging and Boosting [27]. During voting, every subordinate classifier makes one prediction which is counted as a vote. The ensemble classifier counts the predictions and decides which prediction is chosen based on the number of votes for each class. Furthermore, voting can be implemented as weighted and unweighted voting. In weighted voting, the weights of classifier reflect their prediction performance. However, the individual weight for each subordinate classifier does not have to be static. Ikeda et al. [28] show that an estimation classifier can also be based on the structure of the underlying information and the resulting performance of each classifier.

Gama & Brazdil [29] propose a cascading ensemble classifier. They iterate over a set of loosely coupled classifiers and add, for each training and testing phase classification, information from the previous machine learning cycles to the dataset. Using this additional information, the classifier performance of the ensemble classifier is significantly improved. However, modeling the information flow and feedback mechanism for the iterative cycle demands a customized analysis of the given problem and therefore is not generalizable.

Ensemble algorithms that are based on stacked generalization add a superordinate layer of machine learning upon the subordinate classifiers. The predictions of the subordinate classifiers are used to train a machine learning ensemble classifier. Simplified, the ensemble classifier learns about how the subordinate classifiers learn ("learning about learning"). Once the subordinate and ensemble classifiers are trained, the ensemble classifier uses the output of each subordinate classifier to make a prediction. Stacked generalization is a common technique to enhance performance and to combine data sources [29, 30].

Compared to voting, stacked generalization does not require a manual weighting of each subordinate classifier and additionally enables the ensemble classifier to contradict every subordinate classifier. Contrary to cascading, stacked generalization, in its basic form, is applicable to any given input of subordinate classifiers. Thus, in this work, we propose a service-oriented cognitive analytics framework that makes use of an extended ensemble learning model and can dynamically combine different data sources that got processed by correspondent classifier algorithms. The combination of the subordinate classifiers follows a stacked generalization approach, where another layer of machine learning is employed. In contrast, we do not learn on variations of a single dataset, but analyze completely heterogeneous data sources that are distributed as depicted in figure 2. We call this variation of ensemble learning cognitive ensemble, as many independent inputs are combined using two or more layers of machine learning to make an aggregative, centralized prediction. The superordinate cognitive layer is part of a service-oriented cognitive architecture that employs interfaces, classifier roles, output and input definitions as well as a prediction object convention. In contrast to Demirkan & Delen [17], this work describes distributed predictions throughout entities that can itself be integrated into superordinate artifacts, e.g., a web service.

5. Research Agenda

With the described motivations and challenges as well as the proposed solution, we see the need of a research agenda to further develop a service-oriented cognitive analytics framework within a larger research project. To set the foundations of this project, we first clarify the research objective by introducing the corresponding research questions. Subsequently, we elaborate on our research methodology which we aim to apply to answer these questions.

5.1. Research Questions

Our primary objective is to develop an approach that copes with the described challenges of performing analytics in distributed systems. Thus, we derive the following three research questions:

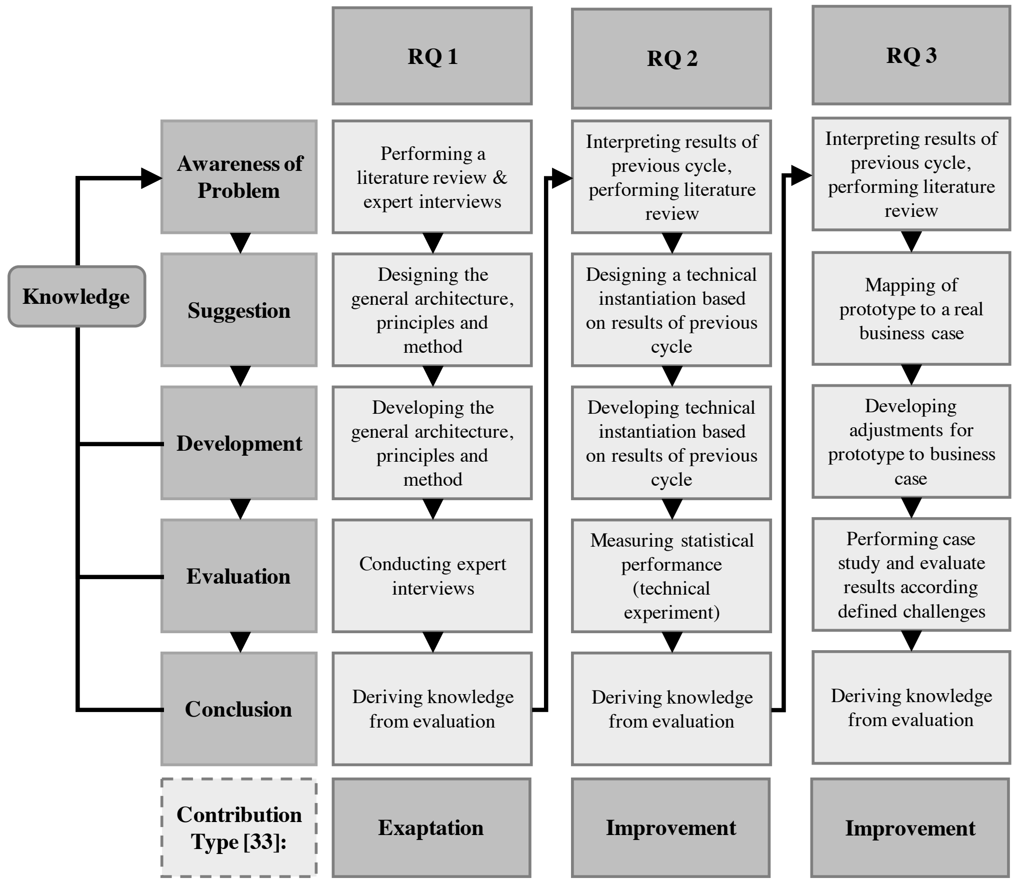

Figure 3. Overview of activities in three consecutive design cycles based on Kuechler & Vaishnavi [31] and the contribution type after Gregor & Hevner [34]

RQ 1. *How can we design a service-oriented cognitive analytics framework to build prediction models that use combinations of heterogeneous, distributed data sources as an input?*

After laying out the architecture and foundations of such a method, research on the feasibility of the proposed architecture needs to be done. For this purpose, we aim to build a prototype based on the results obtained in research question 1.

RQ 2. *Can we achieve robust, IP-preserving, superior predictions based on such a distributed, modular framework that employs a cognitive prediction in comparison to a prediction based on aggregated heterogeneous data sources?*

After evaluating the prototype in terms of general functionality, it needs to be tested in a real-world case study to assess its use to overcome the challenges in a distributed analytics architecture.

RQ 3. *How well does the developed framework cope with the initial challenges of analytics (organizational, technical and algorithmic) in smart service systems?*

5.2. Methodology

To address these research questions, we propose a Design Science Research (DSR) approach [32] with three consecutive design cycles as depicted in figure 3.

The first design cycle aims to set ground for the service-oriented cognitive analytics architecture, principles and method (research question 1). After performing a literature review, we aim to develop a general architecture, corresponding process-flows and principles as nascent design theory artifacts [33]. To evaluate the output of this cycle, we conduct interviews with experts that work in relevant companies which have to cope with distributed analytics problems. The insights of this cycle then either lead to a further refinement of the developed artifacts, or build the basis for the second design cycle that aims to develop a service-oriented cognitive analytics prototype (research question 2).

The developed technical instantiation is evaluated by performing a technical experiment that predicts outcomes on a known data set. Applied evaluation criteria are feasibility, performance of the instantiation and the possibility of IP and privacy-preservation as well as robustness.

The third design cycle combines all previous collected insights and combine them into a holistic artifact that we evaluate by a case study with a real-world business case (research question 3). Hereby we evaluate, whether the developed artifact can overcome the initial challenges, or not.

6. Outlook

Smart services produce vast amounts of data that are distributed across different locations and even across legally independent organizations. However, to exploit the economic potential of analyzing this distributed data, a centralized aggregation is required. To address this challenge, we propose a service-oriented cognitive analytics framework. The contributions of this work, which sets the foundation for further research, is three-fold:

First, we describe the challenges of performing data analytics in smart service systems by outlining three major groups (technical, organizational and algorithmic). Second, by comparing these challenges to the current state of the art, we propose a novel, service-oriented cognitive analytics framework that is based on a variation of ensemble machine learning. As a basic principle, the approach is based on entities in a smart service system that send information in form of prediction outputs instead of raw data. These prediction outputs are then processed by a cognitive aggregator that finds interdependencies and patterns in this data.

Third, having introduced the architecture and the aggregation function of the framework, we propose a research design to develop and evaluate the service-oriented cognitive analytics framework in future research. We derive three research questions and propose a design science approach to design, develop and evaluate the proposed paradigms based on three consecutive design cycles.

Besides these contributions, this work has limitations. We assume entities in a smart service system are willing to share prediction outputs rather than raw data. Due to the nature of a research agenda, this work in its current state is conceptual and a further quantitative definition and evaluation is needed. Thus, as a next step, we perform the first design cycle that aims to create a general service-oriented cognitive analytics nascent design theory that consists of an architecture, process model and principles. By conducting expert interviews with practitioners in the field of distributed analytics, we plan to evaluate and further refine the outlined problem perspectives.

The service-oriented cognitive analytics framework allows to conduct analytics in smart service systems across distributed entities and therefore enables to find holistic insights. A promising field of research lies ahead.

References

[1] H. Chen and V. C. Storey, “Business Intelligence and analytics : From Big Data To Big Impact,” *Mis Q.*, vol. 36, no. 4, pp. 1165–1188, 2012.

[2] T. H. Davenport, “Competing on analytics,” *Harv. Bus. Rev.*, vol. 84, no. 1, p. 98, 2006.

[3] S. Barile and F. Polese, “Smart Service Systems and Viable Service Systems: Applying Systems Theory to Service Science,” *Serv. Sci.*, vol. 2, no. 1–2, pp. 21–40, 2010.

[4] C. Chong, S. P. Kumar, and S. Member, “Sensor Networks: Evolution, Opportunities and Challenges,” vol. 91, no. 8, 2003.

[5] G. Allmendinger and R. Lombreglia, “Four strategies for the age of smart services,” *Harv. Bus. Rev.*, vol. 83, no. 10, pp. 131–145, 2005.

[6] J. Lee, H. A. Kao, and S. Yang, “Service innovation and smart analytics for Industry 4.0 and big data environment,” *Procedia CIRP*, vol. 16, pp. 3–8, 2014.

[7] D. Miorandi, S. Sicari, F. De Pellegrini, and I. Chlamtac, “Internet of things: Vision, applications and research challenges,” *Ad Hoc Networks*, vol. 10, no. 7, pp. 1497–1516, 2012.

[8] M. Cannataro, A. Congiusta, A. Pugliese, D. Talia, and P. Trunfio, “Distributed data mining on grids: Services, tools, and applications,” *IEEE Trans. Syst. Man, Cybern. Part B Cybern.*, vol. 34, no. 6, pp. 2451–2465, 2004.

[9] D. Lucke, C. Constantinescu, and E. Westkämper, “Smart Factory - A Step towards the Next Generation of Manufacturing,” in *Manufacturing Systems and Technologies for the New Frontier*, M. Mitsuishi, K. Ueda, and F. Kimura, Eds. London: Springer London, 2008, pp. 115–118.

[10] Y. Lindell and B. Pinkas, “Privacy Preserving Data Mining,” *Priv. Preserv. Data Min.*, pp. 1–25, 2000.

[11] M. Jensen, “Challenges of Privacy Protection in Big Data Analytics,” *IEEE Int. Congr. Big Data*, pp. 235–238, 2013.

[12] R. Fay, U. Kaufmann, A. Knoblauch, H. Markert, and G. Palm, "Combining visual attention, object recognition and associative information processing in a NeuroBotic system," *Lect. Notes Computer Science*, vol. 3575 LNAI, pp. 118–143, 2005.

[13] H. Baars and H.-G. Kemper, "Management Support with Structured and Unstructured Data - An Integrated Business Intelligence Framework," *Inf. Syst. Manag.*, vol. 25, no. 2, pp. 132–148, 2008.

[14] J. Wielki, "Implementation of the Big Data concept in organizations – Possibilities, impediments and challenges," *Proc. Fed. Conf. Comput. Sci. Inf. Syst.*, no. September 2013, pp. 985–989, 2013.

[15] M. Saar-Tsechansky and F. Provost, "Handling Missing Values when Applying Classification Models," *J. Mach. Learn. Res.*, vol. 8, pp. 1625–1657, 2007.

[16] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog Computing and Its Role in the Internet of Things," *Proc. first Ed. MCC Work. Mob. cloud Comput.*, pp. 13–16, 2012.

[17] H. Demirkan and D. Delen, "Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud," *Decis. Support Syst.*, vol. 55, no. 1, pp. 412–421, 2013.

[18] D. B. Robins, "Complex Event Processing," *2010 Second Int. Work. Educ. Technol. Comput. Sci.*, p. 10, 2010.

[19] R. Agrawal and R. Srikant, "Privacy-preserving data mining," *Proc. 2000 ACM SIGMOD Int. Conf. Manag. data - SIGMOD '00*, vol. 29, no. 2, pp. 439–450, 2000.

[20] D. S. Modha, R. Ananthanarayanan, S. K. Esser, A. Ndirango, A. J. Sherbondy, and R. Singh, "Cognitive computing," *Commun. ACM*, vol. 54, no. 8, p. 62, 2011.

[21] J. Kludas, E. Bruno, and S. Marchand-Maillet, "Information Fusion in Multimedia Information Retrieval," *Adapt. Multimedial Retr. Retr. User Semant.*, vol. 4918, pp. 147–159, 2008.

[22] P. K. Atrey, M. A. Hossain, A. El Saddik, and M. S. Kankanhalli, *Multimodal fusion for multimedia analysis: A survey*, vol. 16, no. 6. 2010.

[23] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodal Deep Learning," *Proc. 28th Int. Conf. Mach. Learn.*, pp. 689–696, 2011.

[24] T. G. Dietterich, "Machine-Learning Research," *AI Magazine*, vol. 18, no. 4, p. 97, 1997.

[25] S. Džeroski and B. Ženko, "Is combining classifiers with stacking better than selecting the best one?," *Mach. Learn.*, vol. 54, no. 3, pp. 255–273, 2004.

[26] L. Todorovski and S. Džeroski, "Combining classifiers with meta decision trees," *Mach. Learn.*, vol. 50, no. 3, pp. 223–249, 2003.

[27] J. R. Quinlan, "Bagging, boosting, and C4.5," *Proc. Thirteen. Natl. Conf. Artif. Intell.*, vol. 5, no. Quinlan 1993, pp. 725–730, 2006.

[28] K. Ikeda, G. Hattori, C. Ono, H. Asoh, and T. Higashino, "Twitter user profiling based on text and community mining for market analysis," *Knowledge-Based Syst.*, vol. 51, pp. 35–47, 2013.

[29] J. Gama and P. Brazdil, "Cascade Generalization," *Mach. Learn.*, vol. 41, no. 3, pp. 315–343, 2000.

[30] C. Merz, "Using Correspondence Analysis to Combine Classifiers," *Mach. Learn.*, vol. 36, pp. 33–58, 1999.

[31] D. Rao, D. Yarowsky, A. Shreevats, and M. Gupta, "Classifying latent user attributes in twitter," *Proc. 2nd Int. Work. Search Min. user-generated contents - SMUC '10*, p. 37, 2010.

[32] B. Kuechler, V. Vaishnavi, and C. I. Systems, "Theory Development in Design Science Research: Anatomy of a Research Project," *Conf. Des. Sci. Res. Inf. Syst. Technol.*, pp. 1–15, 2007.

[33] K. Peffers, M. Rothenberger, T. Tuunanen, and R. Vaezi, "Design Science Research Evaluation," *Des. Sci. Res. Inf. Syst. Adv. Theory Pract.*, pp. 398–410, 2012.

[34] S. Gregor and A. R. Hevner, "Positioning and Presenting Design Science Types of Knowledge in Design Science Research," *MIS Q.*, vol. 37, no. 2, pp. 337–355, 2013.