

Conceptual Modeling Interacts with Machine Learning – A Systematic Literature Review

Moayid Ali Zaidi(✉)

Østfold University College, Halden, Norway

Abstract. Due to the advancement in the digital world, society's expectation towards Machine Learning is very high, especially in Conceptual Modeling. However, the relationship between Machine Learning and Conceptual models are very interesting. Literature in this field has identified the relationship and interaction between Machine Learning and Conceptual Models. However, to the best of our knowledge, there is not a Systematic Literature Review devoted to studying in deep the interaction of these two fields. In this paper, the authors conduct a Systematic Literature Review to get to know how Machine Learning is used in Conceptual Modeling. Results show the deep connection of Machine Learning with Conceptual Models in a solid way, as well as providing challenges and opportunities for future research.

Keywords: Machine Learning · Conceptual Modeling · Systematic Literature Review

1 Introduction

Conceptual Modeling advances the diagrammatic representation of the concepts converges from a broad concept to a very specific point. The goal is to simplify complex information visually. It reduces the gap between planning and requirements to solve the problem. It is the finest way to show how business goals align with the user's expectations visually. Having a long history, Conceptual Modeling has been conducted in the internal organizational settings [1]. The research and practice of the conceptual modeling assumed that the organizations develop these models to support comprehensive and stable internal requirements [2]. Rapid changes occurred in the Information systems (IS) over the last decade, including big data technologies, social media, business analytics, sensors, artificial intelligence, and the internet of things. These variations transformed how the IS designed and used. There is a necessity to innovate new approaches that support these developments in befitting manner [3]. Machine Learning plays a key role in the organizational competitive advantages within the business context [4]. Machine Learning is considered to have an ability to transform organizations and the society in general [5]. Machine Learning is one of the "three trends that

© Springer Nature Switzerland AG 2021
O. Gervasi et al. (Eds.): ICCSA 2021, LNCS 12957, pp. 522–532, 2021.
https://doi.org/10.1007/978-3-030-87013-3_39

organizations must track to gain a competitive advantage" according to Gartner's Hype Cycle for Emerging Technologies 2016 [6]. Machine Learning works without the support of external knowledge; for instance, domain models relies on the learning algorithm and training data. Current research is exploring new ways to scramble the semantics so that rules do not have to be learned from data (training data) [7]. Without external assistant from the users, it mainly focuses on infusing of a learning algorithm. Several issues have been reported working with Machine Learning algorithms and data. The quality of the data is crucial while performing Machine Learning techniques [8]. There is a dire need to improve the ability of Machine Learning techniques to explain the decisions and predictions of the models in effective manner [9]. Historically speaking, the main focus of Artificial Intelligence was symbolic representation while using logical formalism [10]. The first engineering requisite rules such as semantic networks were used to develop AI applications. With the excessive amount of data AI shifted its focus to computationally intensive data-driven approaches (Machine Learning) [11]. Machine Learning iterates over the data to detect the complex relationship among target and input variables. This type of approach is not easily understood by humans. In the contribution of the model's decision some methods give importance to the weight of input features. This type of technique has local interpretable model agnostic explanations with a focus on precise features but it fails to abstract the higher-level concept [12]. Roman [13] proposed the new approach based upon the concept of conceptual modeling, where he used the ML model agnostic approach. This approach can complement existing approaches [14] by combining the knowledge form the application domain and the Machine Learning model. It can deliver the cognitive benefits that facilitate understanding and explanation of the model. Considering the ever-increasing importance of the topic, the authors present the systematic literature review devoted to investigating the intersection of these two fields. To best our knowledge there is no systematic literature review on Machine Learning interacting with conceptual modeling. Our approach provides new insight on this topic. The rest of the paper is organized as follows: Sect. 2 discusses the background and research questions. Section 3 discusses the Research Methods and Research Questions. Section 4 described the review protocol, data sources and search strategy). Section 5 focused on Analysis and Discussion. Finally, Sect. 6 described the conclusion and recommendations for future work.

2 Background

In the development of an Information System (IS), conceptual modeling is the major phase that captures the requirements from the users to support the development of Information Systems. It has led to several kinds of conceptual models including process model, data model, models of business activities, models of enterprise, and system architecture [15]. These models are diagrams that hold graphics and text which are massively used in the development of IS to facilitate communication, improve understanding and guide the activities such as user

interface design, database design, and programming [16]. The conceptual model plays a vital role in the designing of databases, where the Entity-Relationship model has become a standard to represent the structured data. The benefits of conceptual models for information development have been investigated by the researchers. Three basic needs through which the conceptual model arises [17].

2.1 Need to Cope with Complexity

It reduces the complexity of the development of the Information System by concentrating on the relevant aspects, organizing, and structuring the requirements.

2.2 Need for Shared Understanding

In the development of Information Systems, involve many peoples with different backgrounds, expertise, beliefs, and training. That type of diversity creates conflicts and may lead to project failure if it is left unresolved. So they are designed by the general objectives [18].

2.3 Need to Solve Problems

Information Systems are designed to address the organizational and societal needs, although it is very difficult to design a system that has several options. Conceptual modeling supports precise design solutions and analysis. These three needs to be studied in the development of an Information System and Machine Learning. Therefore, these techniques might be useful in resolving challenges, adoption, and development. The potential to apply conceptual modeling to Machine Learning seems to be missing while the existing efforts focused on specific issue.

3 Research Method and Research Questions

The purpose of this study is to structure and characterize the state of the practice on the usage of Conceptual Modeling in Machine Learning. A systematic literature review is used as a research method. The systematic literature review is the means of interpretation, identification, and evaluation of the research relevant to the specific subject or research question. Thus, the main research question driving this study is: What is the state of the practice of the use of Conceptual Modeling in Machine Learning? To address the goal of the paper, three research questions were formulated by the authors. RQ1: How conceptual modeling is used in Machine Learning? RQ2: What are the reported challenges and opportunities in the use of conceptual models in Machine Learning? RQ3: How can Machine Learning facilitate in the conceptual modeling? In the first research question, we discuss which aspects and features in Machine Learning have been reported by the conceptual modeling, while in the second, we examine the potential challenges and opportunities and third describe how ML can facilitate in the Conceptual Modeling.

4 Review Protocol

A review protocol describes the duty that needs to be completed to answer the research questions in a systematic literature review. To find viable literature, the authors used six popular academic databases.

- Google Scholar (https://scholar.google.com/)
- SpringerLink (https://link.springer.com/)
- ScienceDirect (https://www.sciencedirect.com/)
- IEEE Explore (https://ieeexplore.ieee.org/)
- ACM Digital Library (https://dl.acm.org/)
- Wiley Online Library (https://onlinelibrary.wiley.com/)

These databases were chosen because they are among the most relevant source of articles within the extensive field of computing and they are also accessible using institutional accounts.

4.1 Search Strategy

While managing the literature search, it is required to describe the search strategy to find the similar search result. After brainstorming the different terms, the author constructed the search string according to the research questions. To cover the exact topic the string should be simple and comprehensive. The author uses the Boolean operator AND to concatenate the major terms. Hence the resulting query is: ("Conceptual Modeling") AND ("Machine Learning") Zotero was used as a reference management tool to avoid duplication. After the search string, the author chose to filter out the paper based on the inclusion and exclusion criteria. Once the initial results were retrieved, the inclusion and exclusion procedures were applied. The scope of the time in this SLR is set from 2015 to September 2020.

Inclusion criteria

- Papers that are based on the discussion on conceptual modeling and ML.
- Papers, which covers the research questions
- Papers were published in the period 2015 to 2020
- Papers which are written in English

Exclusion Criteria

- Papers contain no relevant data to answer the research question
- The Paper's full text is not available
- Papers that are more than 5 years old
- Duplicated Studies

4.2 Data Extraction

The studies and articles gathered were recorded and stored in a reference manager. This made it simpler to filter different elements based on results. The selection process was recorded as well, and Excel was used to get an additional overview of the selected studies. This made it feasible to attain quantitative data to answer the research questions.

4.3 Selected Primary Studies

The search query provided a total of 3991 results. 25 papers were selected based on their titles, keywords, and abstracts. The papers were filtered by the inclusion and exclusion criteria for this paper. During the filtering process () papers were eliminated because they were not relevant in terms of Research questions and the aim of the paper. Out of 25 results which were selected based on titles, keywords, and abstracts. One is removed because of duplicate. Other discarded because of a lack of connection to the research questions. Although it was approved and used in this paper to answer the questions. The filtering process of the papers is described in Table 1 sorted by their databases.

Table 1. Paper filtering phase

Library	Number of hits	Abstract	Full text
Google Scholar	3050	20	8
ScienceDirect	72	0	0
Springer Link	754	3	2
Wiley Online Library	39	0	0
IEEE Xplore	5	1	0
ACM	71	1	0
Total	3991	25	10

5 Analysis and Discussion

This section deals with answer and findings of research questions. The literature will be compared and discussed, concerning each research question. First, RQ1 will be discussed which feature in Machine Learning have been stated as conceptual modeling. Second, RQ2 investigate the challenges and opportunities. Finally, RQ3 provides a narrow approach to how Machine Learning facilitates conceptual modeling.

RQ1: How conceptual modeling is used in Machine Learning? Conceptual modeling can be used as the standard and effective activity in the different phases of Machine Learning. It can be valuable to improve the understanding of training data, rise the performance and accuracy of Machine Learning algorithms, and recognizing the specific problem. The conceptual modeling benefits in different stages of the Cross-Industry Standard Process for Data Mining for instance understanding of the business, data understanding, modeling, deployment, and evaluation [19,20]. Models are also used to define the experience which is gained by solving Machine Learning tasks [21]. The first phase of CRISP-DM [6] is to understand the requirements and objectives of the project. Conceptual Modeling is used to support this phase. The specific goals cannot be achieved without

understanding the business objectives in Machine Learning projects. Conceptual models describe the goals and objectives of Machine Learning projects. It can assist the managers and other parties to identify the processes which are affected by the Machine Learning. Roman [20] proposed the following steps for conceptual modeling research in Machine Learning.

5.1 Data Understanding

This phase starts with the acquisition of data that is suitable for business problem identification. Modelers review the data, overcome the quality issues, and prepare it for the Machine Learning software. Common notations and modeling grammars such as ER, UML support this phase and well suited for the understanding and modeling of data.

5.2 Data Preparation

This phase involves all the activities which are used to construct the final version of the dataset for the learning algorithms. It includes a different transformation of data by performing the ETL (Extract-Transform-Load) procedure. Previous research in the field of conceptual modeling described the use of process models such as EPCs and BPMN to document the ETL process [20,22]. Conceptual modeling grammar can graphically communicate and better reflect the needs of the Machine Learning process.

5.3 Modeling

In this stage, the dataset emerges from the prior stage is supplied to Machine Learning algorithms for learning, training, and validation. A conceptual model can be used to assist the modeling process and Machine Learning algorithm. For instance, if there is some important information, the modeler may select the inputs manually to ensure that the algorithm examines these inputs. It is accomplished by hard coding some of the data inputs and could also improve the performance of the algorithm.

5.4 Evaluation

This stage includes the interpretation of results of the model generated by the algorithm into the language which is convenient to business users. The conceptual model can highlight the inputs which are necessary for the training model and left the other attributes that exist in the domain. It enhances the understandability of complex Machine Learning models. The notable concern regarding the usage of the Machine Learning algorithm is the lack of transparency i.e. neural networks, random forests, and support vector machines. In the critical and sensitive applications, the “black box” property of these models is a barrier to the wider adoption of Machine Learning [20,23].

5.5 Deployment

The building of the model is not the end of the Machine Learning project. The performance of the model needs to be sufficient if it addresses the business objectives. If the model is working sufficiently then it is introduced into the real world. For instance, to prioritize the case of interest neural networks serve as the decision support tool. Table 2 presents the use of CM for Machine Learning based on the CRISP-DM phases [6].

Table 2. Conceptual modeling and machine learning [6]

Phase	How CM can contribute to the phase	Suggested grammar
Business understanding	Represent, identify boundary objects	i*, BPMN, UML, BIM
Data understanding	Modeling scripts, extending	ER, UM
Data preparation	Attribute selection, transformation	BPMN, EPCs
Modeling	Representation of domain models	UML, ERD
Evaluation	Transparency, understandability	UML, enterprise model
Deployment	Documentation	i*, BPMN, UML

The conceptual model is used to document the goals and objectives of phase 1. When the deployment occurs, the user can refer to phase 1 for the original goals and objectives. The deployment of Machine Learning in business may occur some changes in the business process. Various models like the process model and enterprise model document these changes and consult with the stakeholders about the effect of the business process [20]. Another interesting approach is one taken by Datameer, a data analytics platform that allows us to define and refine the visual representation of pipeline, while the SKYMIND allows us to translate the Machine Learning models into a set of tasks [24].

RQ2: What are the reported challenges and opportunities in the usage of conceptual models in Machine Learning? The different approach is the usage of deep learning, which is based on the data that represent previously observed process instances. One important drawback of using Deep Learning is that the artificial neural network cannot be easily understood in terms of traditional theory elements, such as causes and constructs [21]. Hossain [25] described that the restricted natural language can serve as a high-level language for conceptual modeling particularly for entity-relationship models, and also mentioned that this approach has the potential to bridge between informal specification and formal representation of conceptual modeling. It involves different aspects such as on front end, natural language understanding; on back end scene generation technology and it also mapping the discrete data to continuous data of the graphic scene [26]. Conceptual Modeling addresses the issue related to handling data integration and curation [27]. The limitation of conceptual modeling in data manipulation via Machine Learning is that it does not focus on the limitation of conceptual modeling methods or grammars and thus, it does not directly

contribute to the progress in conceptual modeling research [28]. There is a dearth of research in designing processes to express deploying Machine Learning for business applications, it is because applying conceptual modeling in Machine Learning is different from traditional Information Systems [29].

RQ3: How can Machine Learning facilitate in the conceptual modeling? Machine Learning introduces the natural path to help modeling systems that are inherently involved with data of high complexity. It could automate or semi-automate the process of knowledge capturing. It can be beneficial where the behavioral aspect of the real system is hard to express in equations analytically or simply as data distribution. With the help of Machine Learning, it is possible to learn and illustrate the vast number of datasets. Machine Learning models were also used to learn the complex patterns of flowing crowds. The necessary point to consider is a mechanism of integration. The development of models is an iterative process from conceptual modeling to implementation (Fig. 1). The preliminary concern is to identify the stages where Machine Learning could be applicable or useful. The main point is that Machine Learning can be utilized at different stages of modeling in different ways as follow [30].

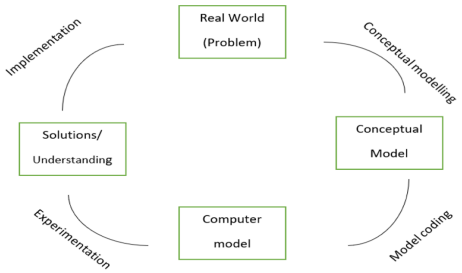

Fig. 1. Key stages of simulation modeling [30]

Sterman [30] consistently stated the importance of developing assistive tools that can understand the complexity and also support the process of learning about systems. At the early stages of problem formulation unsupervised Machine Learning serve as the purpose of knowledge elicitation. For the conceptualization of the system's behavior or structure, unsupervised techniques could be used during the conceptual modeling phase. In the context of the feedback loop, mental models supported the data-driven knowledge learned by Machine Learning. Figure 2 shows the idea, where new data is considered as a form of feedback produced by the new system states. The models continuously re-fitted to the feedback loops to demonstrate the new conditions. In this way, the new states can be learned based on Machine Learning models. While dealing with the large-scale datasets data-driven prediction could be less biased compared to the human modeler.

Using Machine Learning at the process of conceptualization can contribute to lowering the bias of human-based knowledge elicitation. Further, the Machine

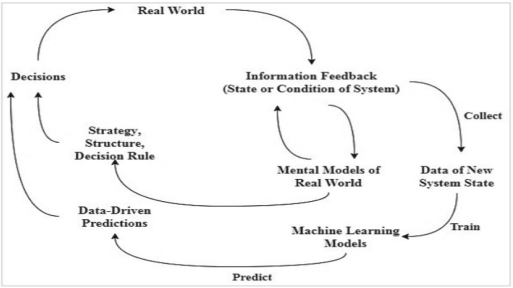

Fig. 2. Data-driven feedback loops; system modeling aided by ML [30]

Learning guided simulation can also help to attain a higher level of model realism. There is a dire need for constructing a high-fidelity representation of real-world problems that brings up new practical aspects for taking advantage of Machine Learning potentials.

6 Conclusion

The future holds many exciting opportunities for assimilating the practice of Machine Learning and Conceptual Modeling. The main motivation for this work was to examine how Conceptual Modeling used in Machine Learning and how Machine Learning interacts with Conceptual modeling. Result show the importance of these two fields paving the way to the new research efforts. Conclusion to our work also identifies the opportunities and challenges in the topic. In addition, many fascinating challenges for Machine Learning and Conceptual Modeling exists which have to be more deeply explored in future.

References

1. Liew, C.F., DeLatte, D., Takeishi, N., Yairi, T.: Recent developments in aerial robotics: a survey and prototypes overview. arXiv:1711.10085 cs, November 2017
2. Castellanos, A., Tremblay, M.C., Lukyanenko, R., Samuel, B.: Basic classes in conceptual modeling: theory and practical guidelines. J. Assoc. Inf. Syst. **21**, 44 (2020)
3. Eckerson, W.W.: Data warehousing special report: data quality and the bottom line. Appl. Dev. Trends **1**, 1–9 (2002)
4. Sabegh, M.A.J., Lukyanenko, R., Recker, J., Samuel, B., Castellanos, A.: Conceptual modeling research in information systems: what we now know and what we still do not know, p. 11 (2017)
5. Khatri, V., Samuel, B.M.: Analytics for managerial work. Commun. ACM **62**(4), 100 (2019). https://doi.org/10.1145/3274277
6. Jones, N.: Using massive amounts of data to recognize photos and speech, deep-learning computers are taking a big step towards true artificial intelligence. Nature **505**, 1–3 (2014)

7. Lukyanenko, R., Parsons, J., Storey, V.C.: Modeling matters: can conceptual modeling support machine learning?, p. 11 (2018)
8. Tsymbal, A., Zillner, S., Huber, M.: Ontology – supported machine learning and decision support in biomedicine. In: Cohen-Boulakia, S., Tannen, V. (eds.) DILS 2007. LNCS, vol. 4544, pp. 156–171. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-73255-6_14
9. Alonso, O.: Challenges with label quality for supervised learning. J. Data Inf. Qual. **6**(1), 1–3 (2015). https://doi.org/10.1145/2724721
10. Castelvecchi, D.: Machine learning is becoming ubiquitous in basic research as well as in industry. But for scientists to trust it, they first need to understand what the machines are doing, p. 4
11. Crevier, D.: AI: The Tumultuous History of the Search for Artificial Intelligence. Basic Books, New York (1993)
12. Cerf, V.G.: AI is not an excuse! Commun. ACM **62**(10), 7 (2019). https://doi.org/10.1145/3359332
13. Ribeiro, M.T., Singh, S., Guestrin, C.: 'Why should i trust you?': explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, pp. 1135–1144, August 2016. https://doi.org/10.1145/2939672.2939778
14. Lukyanenko, R., Castellanos, A., Storey, V.C., Castillo, A., Tremblay, M.C., Parsons, J.: Superimposition: augmenting machine learning outputs with conceptual models for explainable AI. In: Grossmann, G., Ram, S. (eds.) ER 2020. LNCS, vol. 12584, pp. 26–34. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-65847-2_3
15. Henelius, A., Puolamäki, K., Boström, H., Asker, L., Papapetrou, P.: A peek into the black box: exploring classifiers by randomization. Data Min. Knowl. Discov. **28**, 1503–1529 (2014). https://doi.org/10.1007/s10618-014-0368-8
16. Azevedo, C.L.B., Iacob, M.-E., Almeida, J.P.A., van Sinderen, M., Pires, L.F., Guizzardi, G.: Modeling resources and capabilities in enterprise architecture: a well-founded ontology-based proposal for ArchiMate. Inf. Syst. **54**, 235–262 (2015). https://doi.org/10.1016/j.is.2015.04.008
17. Aguirre-Urreta, M.I., Marakas, G.M.: Comparing conceptual modeling techniques: a critical review of the EER vs. OO empirical literature. SIGMIS Database **39**(2), 9–32 (2008). https://doi.org/10.1145/1364636.1364640
18. Lukyanenko, R.: Rethinking the role of conceptual modeling in the introductory IS curriculum, p. 10 (2019)
19. Mark, G., Lyytinen, K., Bergman, M.: Boundary objects in design: an ecological view of design artifacts. J. Assoc. Inf. Syst. **8**(11), 34 (2007). https://doi.org/10.17705/1jais.00144
20. Jackson, J.: Data mining; a conceptual overview. CAIS **8**, 19 (2002). https://doi.org/10.17705/1CAIS.00819
21. Lukyanenko, R., Castellanos, A., Parsons, J., Chiarini, T.M., Storey, V.C.: Using conceptual modeling to support machine learning. In: Cappiello, C., Ruiz, M. (eds.) CAiSE 2019. LNBIP, vol. 350, pp. 170–181. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21297-1_15
22. Fettke, P.: Conceptual modelling and artificial intelligence (2020)
23. Song, I.-Y., Liddle, S.W., Ling, T.-W., Scheuermann, P. (eds.): ER 2003. LNCS, vol. 2813. Springer, Heidelberg (2003). https://doi.org/10.1007/b13244
24. Hall, P., Gill, N.: An Introduction to Machine Learning Interpretability, 2nd edn., p. 62. O'Reilly, Boston (2019)

25. Damiani, E., Frati, F.: Towards conceptual models for machine learning computations. In: Trujillo, J.C., et al. (eds.) ER 2018. LNCS, vol. 11157, pp. 3–9. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00847-5_1
26. Hossain, B.A., Schwitter, R.: Specifying conceptual models using restricted natural language (2018)
27. Wang, J.: Research on spatial conceptual modeling of natural language processing based on deep learning algorithms. J. Phys. Conf. Ser. **1345**, 042090 (2019). https://doi.org/10.1088/1742-6596/1345/4/042090
28. Storey, V.C., Song, I.-Y.: Big data technologies and management: what conceptual modeling can do. Data Knowl. Eng. **108**, 50–67 (2017). https://doi.org/10.1016/j.datak.2017.01.001
29. Lukyanenko, R., Parsons, J., Wiersma, Y., Wachinger, G., Huber, B., Meldt, R.: Representing crowd knowledge: guidelines for conceptual modeling of user-generated content. JAIS **18**(4), 297–339 (2017). https://doi.org/10.17705/1jais.00456
30. Sothilingam, R., Yu, E., Senderovich, A.: Towards higher maturity for machine learning: a conceptual modelling approach. iJournal Grad. Stud. J. Faculty Inf. **5**(1), Art. no. 1 (2019). https://doi.org/10.33137/ijournal.v5i1.33476
31. Elbattah, M.: How can machine learning support the practice of modeling and simulation? – a review and directions for future research. In: 2019 IEEE/ACM 23rd International Symposium on Distributed Simulation and Real Time Applications (DS-RT), Cosenza, Italy, October 2019, pp. 1–7. https://doi.org/10.1109/DS-RT47707.2019.8958703