Conceptual Models for ML: Reflections and Guidelines

Arturo Castellanos[1], Alfred Castillo[2], Monica Chiarini Tremblay[1], Roman Lukyanenko[3], Jeffrey Parsons[4], and Veda C. Storey[5]*

[1] William & Mary, Williamsburg VA, USA
[2] Florida International University, Miami, FL, USA
[3] HEC Montreal, QC, Canada
[4] Memorial University, St. John's, NL, CANADA
[5] Georgia State University, Atlanta, GA, USA
Authors are listed in alphabetical order.

Abstract

Many organizations rely on machine learning techniques to extract useful information from large collections of data. Much research in this area has focused on developing and applying machine learning techniques. We propose that using conceptual models can improve machine learning by providing needed domain knowledge to augment training data with domain knowledge. We propose guidelines for doing so.

Keywords: artificial intelligence, ML, conceptual modeling for ML (CMML), data preparation, domain knowledge

1 Introduction

Machine learning (ML) is an essential approach for organizations to extract useful information from large collections of data in order to make informed decisions. Organizations have adopted industry standards such as CRISP-DM and other frameworks to support the development and deployment of ML systems. At the same time, many challenges remain, including interpreting the outcomes of machine learning models, executing machine learning algorithms effectively, and integrating machine learning into organizational processes. As a result of these challenges, organizations continue to struggle to implement successful machine learning solutions [1]. Recent research explores how conceptual modeling can support machine learning by embedding domain knowledge into machine learning algorithms and models [2]. However, research to date has not been successful at identifying needed domain knowledge or how it might be extracted from conceptual models.

2 Conceptual Models and ML

Conceptual models formally describe "aspects of the physical and social world around us for the purposes of understanding and communication" [3]. Modern ML proceeds without support from external knowledge sources and relies heavily on the training data and learning algorithms [4]. Conceptual models can represent specific objectives and goals for an ML project. In addition, the complexity of some of the models (e.g., neural networks, deep learning) has made models opaque and difficult to explain, opening a stream of research in explainable AI [5]. Explainability and transparency are fundamental, particularly in highly regulated fields such as healthcare. For example, it is critical to understanding how a model concluded the likelihood of diagnosis for a patient. Research has explored ways to improve training data by referencing domain ontologies and incorporating rules such as "all birds are animals" into a model. Thus, instead of learning the concept of "animal" from the data, a model can use the domain ontology to infer that an instance labeled as a bird is also an animal [6]. Finally, conceptual models can assist organizations to comprehend the scope of ML in their business operations (i.e., identify the processes affected by ML).

Given the focus of ML on data and algorithms, the quality of data inputs is critical. The AI community has emphasized the need for a data-centric approach that focuses on the consistency of the training data. According to Andrew Ng, "in the dominant model-centric approach to AI, you collect all the data you can collect and develop a model good enough to deal with the noise in the data. The established process calls for holding the data fixed and iteratively improving the model until the desired results are achieved". In the emerging data-centric approach to AI, "you hold the model or code fixed and iteratively improve the quality of the data," Ng added [7].

There are many kinds of conceptual models, including data models, process models, models of business activity and goals, and models of enterprise and systems architecture [8–10]. However, research at the intersection of conceptual modeling and ML remains rare. ML is absent from the highly regarded conceptual modeling agenda set by Wand and Weber and has absent from the discourse in conceptual modeling [10, 11].

Recently researchers proposed using conceptual modeling in ML applications [12–14]. Nalchigar and Yu [15] developed a conceptual modeling framework for analytics, which includes ML. Nalchigar et al. [16] demonstrated the value of using goal modeling to model ML requirements. Maass and Storey explore the pairing of conceptual modeling with ML [13]. Castellanos et al. [17] proposed guidelines for data preparation using conceptual modeling constructs to improve ML outcomes such as performance and process transparency. Work is being done to improve the transparency of ML models [2] and to make explanations more accessible [18].

3 Proposed Guidelines

As ML becomes more widespread, so are concerns that the process of building models remains opaque [19, 20]. Therefore, we propose the following.

- Domain knowledge captured in conceptual models can help expand the scope of conceptual modeling because conceptual models, through their diagrammatic capabilities and labeling, capture some of the semantics of the real world.
- Including conceptual models as a companion to ML algorithms can help increase

the abilities to perform meaningful explainable AI (with respect to ML).

- Conceptual modeling can make the ML process transparent, repeatable, and auditable by providing a systematic method to clean, organize and transform data.
- Research is needed to foster interdisciplinary connections and demonstrate the continued importance and value of conceptual modeling research across multiple applications of the many different types of ML techniques.
- Further work on natural language processing to aid in constructing meaningful conceptual models from existing documentation and process models is needed.

Each of these can contribute to data-driven decision making for important, complex real-world problems. However, they will require researchers and developers with knowledge of both conceptual modeling and ML, which is currently rare.

4 Conclusion

Although the need for conceptual modeling to advance ML is now well-recognized, continued efforts are needed to bridge these two areas of research. This paper has reviewed some of the work to date and proposed guidelines for advancing this work.

References

1. Ransbotham, S., Kiron, D., Prentice, P.K.: Beyond the hype: the hard work behind analytics success. MIT Sloan Management Review. 57, 3–15 (2016).
2. Lukyanenko, R., Castellanos, A., Storey, V.C., Castillo, A., Tremblay, M.C., Parsons, J.: Superimposition: Augmenting ML Outputs with Conceptual Models for Explainable AI. In: 1st International Workshop on Conceptual Modeling Meets Artificial Intelligence and Data-Driven Decision Making. pp. 1–12. Springer, Vienna, Austria (2020).
3. Mylopoulos, J.: Conceptual modelling and Telos. Conceptual Modelling, Databases, and CASE: an Integrated View of Information System Development, New York: John Wiley & Sons. 49–68 (1992).
4. Mitchell, T.M.: ML. 1997. Burr Ridge, IL: McGraw Hill. 45, 870–877 (1997).
5. Gunning, D.: Explainable artificial intelligence (XAI). Defense Advanced Research Projects agency. Defense Advanced Research Projects Agency (DARPA), nd Web, 2. (2016).
6. Jones, N.: The learning machines. Nature. 505, 146 (2014).
7. Press, G.: Andrew Ng Launches A Campaign For Data-Centric AI, https://www.forbes.com/sites/gilpress/2021/06/16/andrew-ng-launches-a-campaign-for-data-centric-ai/?sh=92bb8b574f57, (2021).
8. Davies, I., Green, P., Rosemann, M., Indulska, M., Gallo, S.: How do practitioners use conceptual modeling in practice? Data & Knowledge Engineering. 58, 358–380 (2006).
9. Burton-Jones, A., Weber, R.: Building conceptual modeling on the foundation of ontology. In: Computing handbook: information systems and information technology. p. 15.1-15.24. , Boca Raton, FL, United States (2014).

10. Recker, J., Lukyanenko, R., Sabegh, M.A., Samuel, B.M., Castellanos, A.: From Representation to Mediation: A New Agenda for Conceptual Modeling Research in A Digital World. MIS Quarterly. 45, 269–300 (2021).
11. Jabbari, M., Lukyanenko, R., Recker, J., Samuel, B., Castellanos, A.: Conceptual Modeling Research: Revisiting and Updating Wand and Weber's 2002 Research Agenda. AIS SIGSAND. 1–12 (2018).
12. Lukyanenko, R., Castellanos, A., Parsons, J., Tremblay, M.C., Storey, V.C.: Using conceptual modeling to support ML. Presented at the International Conference on Advanced Information Systems Engineering (2019).
13. Maass, W., Storey, V.C.: Pairing conceptual modeling with ML. Data & Knowledge Engineering. 101–123 (2021).
14. Reimer, U., Bork, D., Fettke, P., Tropmann-Frick, M.: Preface of the First Workshop Models in AI. Presented at the Modellierung (Companion) (2020).
15. Nalchigar, S., Yu, E.: Business-driven data analytics: A conceptual modeling framework. Data & Knowledge Engineering. 117, 359–372 (2018).
16. Nalchigar, S., Yu, E., Keshavjee, K.: Modeling ML requirements from three perspectives: a case report from the healthcare domain. Requirements Engineering. 26, 237–254 (2021).
17. Castellanos, A., Castillo, A., Tremblay, M.C., Lukyanenko, R., Parsons, J., Storey, V.C.: Improving ML performance using conceptual modeling. In: AAAI Symposium on Combining ML and Knowledge Engineering in Practice. pp. 1–4. , Stanford, CA (2021).
18. Lukyanenko, R., Castellanos, A., Samuel, B.M., Tremblay, M., Maass, W.: Research Agenda for Basic Explainable AI. In: Americas Conference on Information Systems. pp. 1–8. , Montreal, QC (2021).
19. Hutson, M.: Artificial intelligence faces reproducibility crisis. Science. (2018).
20. Jones, M.: How Do We Address The Reproducibility Crisis In Artificial Intelligence?, https://www.forbes.com/sites/forbestechcouncil/2018/10/26/how-do-we-address-the-reproducibility-crisis-in-artificial-intelligence/#54236c5b7688, (2018).