International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Developing Knowledge-Centric Frameworks for Enhancing Web Search Diversity through Semantic Artificial Intelligence

Garapati Suresh[1], Cherukuri Anusha[2], vankayala Anil Santosh[3]

Submitted: 05/09/2024 **Revised:** 15/10/2024 **Accepted:** 25/10/2024

Abstract: Machine Learning methods, particularly Artificial Neural Networks, have garnered significant interest in both research and practical applications because to their substantial promise in prediction tasks. Nevertheless, these models often fail to provide explainable results, which is an essential criterion in several high-stakes fields such as healthcare or transportation.

Concerning explainability, Semantic Web Technologies provide semantically interpretable tools that facilitate reasoning on knowledge bases. Consequently, the inquiry emerges about how Semantic Web Technologies and associated notions might enhance explanations inside Machine Learning systems. This discussion presents contemporary methodologies for integrating Machine Learning with Semantic Web Technologies, focusing on model explainability, derived from a rigorous literature review. In this process, we also emphasize the areas and applications propelling the study field and examine the methods by which explanations are provided to the user. Based on these observations, we propose avenues for further study on the integration of Semantic Web Technologies with Machine Learning.

Keywords: *Semantic Web Technologies, Machine Learning, Explainability, XAI.*

Introduction

Artificial Intelligence (AI) and, specifically, Machine Learning (ML) approaches have achieved remarkable success in several tasks, including medical diagnosis, credit card fraud detection, and facial recognition. These methods, however, are often opaque and typically fail to provide comprehensible explanations for their predictions [23]. This circumstance is troublesome as it might negatively impact the comprehension, confidence, and administration of ML algorithms [23]. Although not every benign algorithmic action requires comprehensive elucidation, explainability is essential when addressing incomplete problem statements that include safety, ethics, or trade-offs. Moreover, the legal implications of AI accountability enhance the significance of explainable decision-making systems [19].

The term Explainable Artificial Intelligence (XAI) encompasses several methodologies aimed at making machine learning techniques explainable, transparent, interpretable, or understandable in academic discourse. A multitude of study on XAI exists, including literature evaluations of prominent approaches and strategies (see to [2] or [22] for instance). Nonetheless, several approaches depend only on a technical understanding of the opaque machine learning models. Cherkassky and Dhar [14] contend that model explainability is unattainable for such techniques. The authors assert that explainability is significantly reliant on the application of domain knowledge rather than only on data analysis. This concept has been lately modified by many writers who contend that the integration of Semantic Web Technologies might be pivotal in realizing really explainable AI systems [26, 27]. Given that current surveys on XAI have not thoroughly examined this interesting research area, we provide a literature-based assessment of the use of Semantic Web Technologies in conjunction with ML approaches to enhance explainability. We specifically concentrate on three research inquiries:

1. What combinations of Semantic Web Technologies and Machine Learning have been suggested to improve model explainability?
2. Which application areas and activities are particularly significant to this study field?
3. In what manner are model explanations assessed and conveyed to the user?

The subsequent sections of this work are structured as follows. Section 2 offers pertinent background

[1,2,3,4]International School Of Technology And Sciences For Women, A.P, India.

material about the explainability of machine learning systems. Subsequently, Section 3 succinctly outlines the research design prior to providing the main findings of this study. Section 4 delineates the implications for future study based on these observations. Ultimately, Section 5 finishes this study.

Literature Review

The explainability of artificial intelligence is not a new area of investigation. Mueller et al. [39] examined the chronological progression of XAI and demonstrated that the subject has been extensively researched from the 1970s to the early 1990s in relation to Expert and Tutoring Systems. In the subsequent two decades, less study has been conducted in the field. Recently, there has been a revival of interest in the area owing to advancements in Machine Learning and Deep Learning [39].

Notwithstanding the recent proliferation of articles about XAI, there is a lack of consensus on the definition of explainability [34]. This survey distinguishes between interpretable systems, which enable users to examine the mathematical mapping from inputs to outputs, and explainable systems, which elucidate the operational logic of the system, as delineated by Adadi and Berrada [2]. Doran et al. [17] assert that really explainable systems must include reasoning features that use knowledge bases to provide explanations that are comprehensible to humans while remaining impartial. Moreover, it is important to note that interpretability or explainability is contingent not only upon a certain model but also on the knowledge and expertise of its users [24].

Numerous studies within the field of machine learning examine the subjects of explainability and interpretability. Biran and Cotton [10] examine algorithmic and mathematical approaches to interpretable machine learning models, Abdul et al. [1] emphasize explanations from a human-centered viewpoint, and Adadi and Berrada [2] provide a comprehensive assessment that includes evaluation and perceptual elements. Nevertheless, these research often neglect to address how tools like Semantic Web Technologies may enhance the explainability of machine learning systems. Conversely, in the pertinent domain of Data Mining and Knowledge Discovery, the elucidation of data patterns using Semantic Web and Linked Open Data has been comprehensively examined in a study by Ristoski and Paulheim [45]. Although Data Mining, Knowledge Discovery, and Machine Learning exhibit significant overlaps, a comprehensive review of the integration of Semantic Technologies and Machine Learning remains absent. This study focuses on conventional machine learning approaches rather than areas such as Inductive Logic Programming (ILP) [40]. Inductive Logic Programming (ILP) integrates concepts from Machine Learning, using both positive and negative instances, with logical programming to build a collection of interpretable logical rules.

A overview of the use of ontologies inside the ILP framework is available in [35]. Although several scholars categorize ILP as a subset of ML (e.g. [50]), we agree with Kazmi et al. [30] in distinguishing the two domains and concentrate on traditional ML, briefly addressing ILP.

Explainable Machine Learning Models through Semantic Web Technologies

This section succinctly outlines the study design of the survey prior to summarizing the findings of the analysis undertaken. To address the study issues, we conducted a comprehensive literature evaluation by scanning prominent academic databases, including ACM Digital Library, SCOPUS, and peer-reviewed preprints on arXiv. The latter has been included since XAI is a continuously expanding field with several contributions arising from ongoing research. We performed a search using terms associated with three categories: Machine Learning, Semantic Web Technologies, and explainability. 1 The resulting list of articles was assessed for relevance according to their abstracts, while the remaining publications were appraised based on their whole content. A forward and backward search has been performed to enhance the compilation of pertinent research papers.

To elucidate the initial study issue, we classified the relevant models according to their use of machine learning and Semantic Web Technologies. We categorized machine learning methodologies based on their learning paradigms (supervised, unsupervised, reinforcement learning) and evaluated the used Semantic Web Technologies according to their semantic expressiveness. We concentrated on the information that was actually used rather than the foundational representation. If a system utilizes an ontology just for taxonomical knowledge, it is classified as a taxonomy. We adhered to the distinctions made by Sarker et al. [47] in differentiating knowledge graphs from ontologies, wherein the former typically consist of triples predominantly articulated through the Resource Description Framework (RDF), while the latter also incorporates type logics and is commonly represented using the Web Ontology Language (OWL). We examined the second study issue by analyzing the application areas and tasks of the evaluated systems. We address the third study question by defining the format of explanations provided to the user and the criteria for evaluating their quality.

Domains and Applications

The integration of machine learning algorithms with Semantic Web Technologies is influenced by the specific application domains and objectives to be achieved. Table 2 presents a summary of the predominant domains and tasks of the analyzed systems. Concerning the former, it is evident that, while several systems are created outside of a specific domain, health care significantly influences the development of interpretable machine learning systems. In relation to the functions of the evaluated systems, we identified the recommendation job and picture analysis as significantly important.

For conciseness, we restrict the following paragraphs to the healthcare sector and the recommendation assignment. Healthcare systems often integrate classification tasks, such as predicting diagnoses, with taxonomical information derived from medical diagnostic codes or medical ontologies. Jiang et al. [29] employ the hierarchical structure of the International Classification of Diseases (ICD) to incorporate a regularization penalty into their logistic regression, resulting in a sparse model where non-zero features are predominantly concentrated within a restricted number of subtrees rather than being dispersed throughout the entire hierarchy. This method of feature weighting may enhance the explicitness of the algorithmic prediction process (interpretability); yet, it fails to provide explanations and justifications for non-experts (e.g., patients). Chen et al. [12] integrate hierarchical ICD knowledge into a Neural Network architecture to regularize the output layer and extract therapeutically relevant characteristics. Yan et al. [60] use hierarchical links inside an ontology to augment a collection of medical labels by deducing absent parent labels. The designation "right mid lung" is extended to include "right lung," "lung," and "chest." The authors use exclusive connections among labels to identify challenging instances and enhance accuracy. The system may provide input examples analogous to the model output as predictive evidence while analyzing medical photos. KAME [36] is a diagnostic prediction system influenced by [15] that use medical ontologies to acquire embedded representations of medical codes and their hierarchical parent codes. These represent

Semantic Web Technologies for Explainability

The integration of Semantic Web Technologies and Machine Learning has significant promise for enabling explainable models. We identified the alignment of machine learning data with knowledge base entities, referred to as knowledge matching, as a primary difficulty that future research must address. Automated and dependable techniques for knowledge matching are essential. In this regard, Wang et al. [56] advocate for string matching between identified items and ontology classes, whereas Liao et al. [33] offer the automated extraction of ideas and connections from online sources. Additional research in this domain, along with other fields such as semantic annotation, is essential to provide effective and efficient knowledge matching.

Furthermore, we identified a particular emphasis on certain machine learning approaches and Semantic Web Technologies. Further research is required in the domains of explainable reinforcement learning and clustering. In this setting, we see that efforts across many disciplines and activities remain relatively separated, despite the availability of technologies such as linked data for integrating disparate domains. Certain previous studies recognize the need to broaden the scope of tasks executed by explainable systems or their areas of application. Additional authors 42, 47]. Consequently, the domains of ontology and knowledge graph learning, together with The alignment of the knowledge base is crucial for achieving this objective. Future endeavors must thus identify methods to alleviate the anticipated deficiency of the interconnectivity of data and the heightened complexity of these systems. Ultimately, we emphasize the need for future endeavors to pursue really explicable systems that integrate reasoning with externally sourced information comprehensible to humans. To attain this objective, forthcoming explanation systems must guarantee that the explanations provided accurately reflect the underlying machine learning method. Moreover, such methodologies should elucidate not only the relationship between an output and a pertinent representation but also the process by which this representation has been derived. For instance, it is insufficient to validate the detection of a human face only by asserting that the eyes, mouth, and nose have been identified and that these attributes constitute a human face (e.g., inferred from ontology). An adequately explainable system must elucidate the rationale for the recognition of these properties. This argument pertains to the issue of user engagement, which is further upon below. contemplate the use of more data [60] or extra intricate prior knowledge [41]

Human-Centric Explanations

As explanations constitute a kind of social interaction, their effectiveness and quality largely rely on their intelligibility and comprehensibility as perceived by the user. An explanation is only beneficial if the user comprehends it. This analysis demonstrates that the structure and presentation of explanations vary considerably across existing systems, with many failing to provide explanations in natural language. Consequently, we assert that the domain of Natural Language Processing (NLP), and specifically Natural Language Generation (NLG), provides a valuable foundation. For instance, Vougiouklis et al. [53] produce natural language writings using Semantic Web triples utilizing Neural Networks. Furthermore, Ell et al. [20] convert SPARQL searches into comprehensible English language for non-experts. More broadly, the domain of (Visual) Question Answering might serve as a source of inspiration, since inquiries and responses are often articulated in natural language [56]. Furthermore, we assert that explanations must be flexible and participatory to maximize user benefit. Structured knowledge bases enable users to examine and engage with explanations in several formats. For instance, the user might explore several potential explanations or go into a specific explanation to get more detailed reasons that influenced a forecast. Khan et al. [31] propose a framework that facilitates follow-up inquiries. Likewise, Bellini et al. [9] want to include the capability for users to rectify their system in a continuous feedback loop. Sarker et al. [47] consider this line of action a significant objective for future research. Nonetheless, an agreement about the precise manner of interaction seems to be lacking. To identify best methods for presenting and engaging with explanations, future research must integrate findings from a broader array of disciplines. Current research [1, 2] indicates an expanding array of various and multidisciplinary efforts focused on the issue of human-comprehensible explanations applicable in this setting.

Conclusion

Explainability and interpretability have become crucial prerequisites for several machine learning systems. This study demonstrates, via a comprehensive literature analysis, that the interplay between machine learning and Semantic Web Technologies has promising prospects for model explainability. We examined the predominant methodologies in supervised and unsupervised learning, emphasizing the significance of the health care sector and recommendation tasks as pivotal influences in the study area. The literature review indicated that the incorporation of background information inside the machine learning paradigm does not diminish prediction performance; rather, it often enhances it. Ultimately, we presented instances of certain types of explanations, including natural language and rule-based statements. Simultaneously, we emphasized that significant success in the examined field is contingent upon advancements in several research issues. This encompasses technological inquiries such as automated methods for knowledge matching and advancements in knowledge base learning. Additional problems pertain to the development of adaptive and interactive systems. Ultimately, future studies must develop more stringent assessment procedures. We assert that addressing these inquiries and further investigating the integration of structured information, reasoning, and Machine Learning may facilitate the development of really explainable systems.

References

[1] Alirezaie, M., L□angkvist, M., Sioutis, M., Lout_, A.: Semantic Referee: A neuralsymbolic framework for enhancing geospatial semantic segmentation. Semantic Web Journal (2019)

[2] Batet, M., Valls, A., Gibert, K.: Performance of ontology-based semantic similarities in clustering. In: Proceedings of the 10th International Conference on Artificial

Intelligence and Soft Computing. pp. 281{288. Springer, Berlin, Heidelberg (2010)

[3] Batet, M., Valls, A., Gibert, K., S_anchez, D.: Semantic clustering using multiple ontologies. In: Arti_cial Intelligence Research and Development - Proceedings of the 13th International Conference of the Catalan Association for Arti_cial Intelligence. pp. 207{216. IOS Press, Amsterdam, The Netherlands (2010)

[4] Abdul, A., Vermeulen, J., Wang, D., Lim, B.Y., Kankanhalli, M.: Trends and trajectories for explainable, accountable and intelligible systems: An HCI research agenda. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. pp. 582:1{582:18. CHI '18, ACM, New York, NY, USA (2018)

[5] Adadi, A., Berrada, M.: Peeking inside the black-box: A survey on explainable arti_cial intelligence (XAI). IEEE Access 6, 52138{52160 (2018) Aditya, S., Yang, Y., Baral, C.: Explicit reasoning over end-to-end neural architectures for visual question answering. In: Proceedings of the Thirty-Second AAAI Conference on Arti_cial Intelligence. New Orleans, Louisiana, USA (2018)

[6] Ai, Q., Azizi, V., Chen, X., Zhang, Y.: Learning heterogeneous knowledge base embeddings for explainable recommendation. Algorithms 11(9), 137 (2018)

[7] Alirezaie, M., L□angkvist, M., Sioutis, M., Lout_, A.: A symbolic approach for explaining errors in image classi_cation tasks. In: IJCAI Workshop on Learning and Reasoning. Stockholm, Sweden (2018)

[8] Bellini, V., Schiavone, A., Di Noia, T., Ragone, A., Di Sciascio, E.: Knowledgeaware autoencoders for explainable recommender systems. In: Proceedings of the 3rd Workshop on Deep Learning for Recommender Systems. pp. 24{31. DLRS 2018, ACM, New York, NY, USA (2018)

[9] Biran, O., Cotton, C.: Explanation and justi_cation in machine learning: A survey. In: Proceedings of the IJCAI-17 Workshop on Explainable AI (XAI). pp. 8{13. Melbourne, Australia (2017)

[10] Brynjolfsson, E., Mitchell, T.: What can machine learning do? Workforce implications. Science 358(6370), 1530{1534 (2017)

[11] Che, Z., Kale, D., Li, W., Bahadori, M.T., Liu, Y.: Deep computational phenotyping. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 507{516. KDD '15, ACM, New York, NY, USA (2015)

[12] Chen, J., Lecue, F., Pan, J.Z., Horrocks, I., Chen, H.: Knowledge-based transfer learning explanation. In: Sixteenth International Conference on Principles of Knowledge Representation and Reasoning. pp. 349{358. Tempe, AZ, USA (2018)

[13] Cherkassky, V., Dhar, S.: Interpretation of black-box predictive models. In: Measures of Complexity, pp. 267{286. Springer, New York (2015)

[14] Choi, E., Bahadori, M.T., Song, L., Stewart, W.F., Sun, J.: GRAM: Graph-based attention model for healthcare representation learning. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.pp. 787{795. KDD '17, ACM, New York, NY, USA (2017)

[15] Clos, J., Wiratunga, N., Massie, S.: Towards explainable text classi_cation by jointly learning lexicon and modi_er terms. In: IJCAI-17 Workshop on Explainable AI (XAI). pp. 19{23. Melbourne, Australia (2017)

[16] Doran, D., Schulz, S., Besold, T.R.: What does explainable AI really mean? A new conceptualization of perspectives. In: Proceedings of the First International Workshop on Comprehensibility and Explanation in AI and ML 2017 co-located with 16th International Conference of the Italian Association for Arti_cial Intelligence (AI*IA 2017). Bari, Italy (2017)

[17] Doshi-Velez, F., Kim, B.: Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608 (2017)