

A Preliminary Approach about using nowadays Knowledge Engineering in Artificial Intelligence: a literature overview

Cynthia Vegega¹[0000-0002-5382-7875], María F Pollo-Cattaneo¹[0000-0003-4197-3880] and José A Olivas²[0000-0003-4172-4729]

¹ Grupo de Estudio de Metodologías para Ingeniería en Software (GEMIS)
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires, Argentina

² Soft Management of Internet and Learning (SMILe)
Universidad de Castilla-La Mancha, España
cynthia.vegega@gmail.com, flo.pollo@gmail.com,
joseangel.olivas@uclm.es

Abstract. This paper presents a first literature overview on the nowadays use of Knowledge Engineering in the framework of the development of solutions in the field of Artificial Intelligence. With the assumption that the conceptualization, formalization and modeling of data is fundamental in this type of projects, it is considered that Knowledge Engineering can actively collaborate orienting and guiding part of these activities. In this context, a doctoral research line is proposed about the use of Knowledge Engineering to perform Intelligent Data Analysis. This line of research arises from the fact that existing methodologies in the field of Artificial Intelligence and Data Mining do not incorporate complete domain and context knowledge in data analysis. The incorporation of this knowledge allows contrasting the hypotheses obtained from the data and enriches the analysis reaching better results.

Keywords: Knowledge Engineering, Intelligent Data, Artificial Intelligence

1 Introduction

Knowledge Engineering [1,2] is a discipline that is linked to Artificial Intelligence and is oriented to the construction of Intelligent Systems, which are artifacts that present some intelligent behavior in the human sense. Within Intelligent Systems are Knowledge-Based Systems whose source of knowledge can come from data or domain knowledge. The data may be obtained from the domain itself or could be external, so they may require data from libraries or other databases outside the domain.

Data analysis methods typically use normalized and structured numerical data as inputs, without considering the heterogeneous nature of the data repositories. These methods try to interpret the output, but a complex Intelligent System cannot be analyzed through them. If added to this data analysis, domain knowledge would be enriched and

complement the development of Intelligent Systems. In this sense, Knowledge Engineering can help with the work of data that may be inaccurate, uncertain, have errors, contradictions and biases. If the data is biased, there is a risk of generating an Intelligent System that is not based on reality and produces erroneous results [3], solving a different problem from the one that is to be solved.

Also, the usual practice to find relationships between existing data involves "going blind" in their processing using the application of algorithms or analysis tools, such as statistical tools or Machine Learning [4]. This is done in a unidirectional way with the purpose of being able to interpret the output generated by these algorithms in order to verify if the dataset can be useful or not for the implementation of an Intelligent System. However, this practice is not valid for an Intelligent System due to its complex nature.

In this context, the doctoral research line proposes that, based on existing data and domain knowledge, the process of Intelligent Data Analysis can be systematized using techniques and tools belonging to Knowledge Engineering. This analysis will focus on different elements such as the contrast and establishment of hypotheses between data and domain knowledge, the bidirectional balance between correlation and causation, searching for association rules in both directions, and the construction of formalizable cognitive models, such as taxonomies and ontologies, that enrich the analysis.

2 Motivation and literature overview

Hybrid Intelligence [5,6] allows achieving complex objectives by combining human intelligence and Artificial Intelligence in order to achieve results superior to those that each of them could have achieved separately and to continuously improve by learning from each other. In this sense, the combination of Knowledge Engineering and Machine Learning allows complementing the strengths and weaknesses of both, opening new possibilities for the organization of knowledge and contributing to the creation of hybrid Intelligent Systems. This is why, currently, there are international forums that are beginning to address the importance of the relationship between Knowledge Engineering and Machine Learning, such as AAAI-MAKE [7], KEPS [8], IKE [9] and KEOD [10].

Some works that show the application of Hybrid Intelligence are:

- In [11] a recommendation system for small businesses with Hybrid Intelligence is proposed to prevent the lack of operating funds in the short term that represents a critical problem in this type of organizations. Since the solution using Machine Learning alone was not viable due to the complexity of the problem, the heuristic knowledge of financial experts is integrated. This knowledge is incorporated through a rule-based analysis in order to identify different corrective actions to solve the problem while Machine Learning models are used to predict the probability of success and the cost of each action.

- In [12] a model is presented that combines the results of different agricultural simulations with the knowledge of the experts to make predictions. Artificial Neural Networks are used to make these predictions, while ontologies are used to link the data with the experts' knowledge.
- In [13] a decision support system is designed for the admission process of students to a University in order to admit the right students. Since the current admission process is based on experts who apply their criteria for admitting or not admitting a student, their knowledge was acquired through previous case studies. It combines reasoning based on previous cases and Machine Learning for the analysis of qualifications.

On the other hand, knowledge is usually used anarchically, since there are methodologies such as CRISP-DM [14], SEMMA [15] or P3TQ [16] that define an understanding of the domain, but do not systematize how to use it and do not relate it directly to data analysis. Likewise, the KDD process [17] proposes a data preprocessing and transformation stage to then detect behavioral patterns, but does not incorporate domain knowledge in the data analysis either.

In this line of doctoral research we intend to deconstruct what has been done in the implementation of Intelligent Systems that are currently successful cases, such as the evaluation of hate speech in social networks [18] or INFEDEC 2.1 [19] which is a Knowledge Based System for decision support in the medical field.

In order to make the appropriate decision, the right knowledge needs to be available and accessible in order to be able to make better decisions [17]. When we are faced with Artificial Intelligence to be able to anticipate the future by looking for the right outcome, today, there is a decision of whether we link it to the data that is available or we also incorporate domain knowledge into that data. In this sense, two tasks can be performed from the available data. In the first instance, their manipulation and in the second instance their exploitation.

Data manipulation consists of surveying the available repositories in the domain and from them, being able to transform them into a heterogeneous datalake containing usable data. All available data must be acquired with the objective of converting them into a representation appropriate to the problem being solved. They must be prepared in a certain format and then cleaned so that they can be used and stored in a suitable container.

The objective of *data exploitation* is the complete analysis of the data. Therefore, from the datalake it must be decided whether the data or a part of it follows a trend or not. In the case that part of the data follows a trend, an analysis of the data is performed using statistical models and Machine Learning. However, whether they follow a trend or not, the use of Knowledge Engineering is also proposed in order to contrast and establish hypotheses based on the experience of experts in that area of knowledge. In reference to the contrast and establishment of hypotheses between data and domain

knowledge, Popper [18] defines that in the search for knowledge one cannot dispense with the truth, therefore, it is a matter of finding true theories or those that are close to it, rejecting false theories. The hypotheses that can be established from the data have a degree of confirmation and also have a degree of refutation that can be obtained from the knowledge of the domain, giving more or less credibility to each of them.

Knowledge Engineering allows the acquisition of expert knowledge through direct methods, such as interviews, questionnaires and surveys, or through indirect methods, such as observation of routine tasks, analysis of protocols and grids [19]. This knowledge can be conceptualized taxonomically and ontologically, enriching the data analysis. Soft Computing [20] techniques that are tolerant to imprecision and uncertainty within the knowledge representation can also be used.

2 Conclusions

This article describes a doctoral research line that proposes an Intelligent Data Analysis using techniques and tools belonging to Knowledge Engineering. Knowledge Engineering enriches the data analysis process with domain knowledge. In this way, Machine Learning is used to generate hypotheses that check or contradict with domain knowledge, being able to find a balance between correlation and causation.

As future lines of work and considering that in order to carry out any solution development process in the field of Artificial Intelligence, it would be advisable to be oriented by guidelines, models, methodologies and/or frameworks that guarantee the implementation of a high quality product, it is expected to lay the foundations for the proposal of an integral methodology based on good engineering practices with a view to the fields of Knowledge Engineering and Data Analytics. The proposed framework aims to contribute to strengthen the relationship between specifications and their formal representations in search of a structured and organized implementation of the intervening actors.

References

1. García Martínez, R., Britos, P.: *Ingeniería de Sistemas Expertos*, Editorial Nueva Librería, 2004.
2. Palma, J. T., Paniagua-Aris, E., Martín, F., Marín, R.: *Ingeniería del Conocimiento. De la Extracción al Modelado de Conocimiento*, *Inteligencia Artificial*, vol. 11, pp. 46-72, 2000.
3. Domingos, P: A Few Useful Things to know about Machine Learning. *Communications of the ACM*, 55(10), 78-87, 2012.
4. Olivas Varela, J. A.: *Inteligencia Artificial, Inteligencia Computacional y Análisis Inteligente de Datos*, OBS Business School, 2021.
5. Dellermann, D., Ebel, P., Söllner, M., Leimeister J. M., *Hybrid Intelligence, Business & Information Systems Engineering* 61 (5), 637-64, 2019.
6. Kamar, E., *Directions in Hybrid Intelligence: Complementing AI Systems with Human Intelligence*, *IJCAI*, 4070-4073, 2016.

7. AAAI-MAKE, <https://www.aaai-make.info>, last accessed 08/04/2022.
8. KEPS, <https://icaps21.icaps-conference.org/workshops/KEPS>, last accessed 08/04/2022.
9. IKE, <https://www.american-cse.org/csce2021/conferences-IKE>, last accessed 08/04/2022.
10. KEOD, <https://keod.scitevents.org>, último acceso 08/04/2022.
11. Chapman, P., Clinton, J., Keber, R., Khabaza, T., Reinartz, T., Shearer, C. y Wirth, R, CRISP-DM 1.0 Step by step BI guide. Edited by SPSS, <https://www.the-modeling-agency.com/crisp-dm.pdf>, last accessed 08/04/2022.
12. SAS, Introduction to SEMMA, <https://documentation.sas.com>, last accessed 08/04/2022.
13. Pyle, D., Business Modeling and Data Mining, Morgan Kaufmann Publishers, 2003.
14. Fayyad, U., Piatestky-Shapiro, G. y Smyth, P. The KDD Process for Extracting Useful Knowledge from Volumes of Data. Communications of the acm, 39(11), 27-34, 1996.
15. Montoro Montaroso, A.: Análisis de Sentimientos para la prevención de mensajes de odio en las Redes Sociales, Trabajo Fin de Grado, Universidad de Castilla-La Mancha, Escuela Superior de Informática, 2019.
16. Lago R., Olivas Varela J. A., Megido J. A., Grávalos C., Pérez J.: INFEDEC 2.1: Sistema de Ayuda en la Decisión de Diagnóstico y Tratamiento de Enfermedades Infecciosas. Actas de ESTYLF'96: VI Congreso Español sobre Tecnologías y Lógica Fuzzy, Universidad de Oviedo, pp. 271-275, 1996.
17. Bennet, A., Bennet, D.: The Decision-Making Process in a Complex Situation, In: Handbook on Decision Support Systems 1, pp. 3–20, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.
18. Popper, K. R., La Lógica de la investigación científica, Sánchez Zabala, V. (trad.), Madrid: Tecnos, 1962.
19. García Martínez, R., Rossi, B. & Britos, P.: Metodologías de Educación de Conocimiento para la Construcción de Sistemas Informáticos Expertos, Revista del Instituto Tecnológico de Buenos Aires 21:72-84, 1998.
20. Zadeh, L. A: Soft Computing and Fuzzy Logic. En Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lotfi a Zadeh, p. 796-804, 1996.