Hybrid Intelligent System for Leveraging Georeferenced Data and Knowledge*

D. Carrasco[1], Jose A. Olivas[1[0000-0003-4172-4729]] and Pablo L. Higueras[2[0000-0002-3662-7302]]

[1] Department of Information Technologies and Systems, University of Castilla-La Mancha, 13071 Ciudad Real, Spain
[2] Institute of Applied Geology (IGeA), University of Castilla-La Mancha, 13071 Ciudad Real, Spain

Abstract. This work presents a process for developing an intelligent hybrid system designed to effectively leverage georeferenced data and expert knowledge. The effectiveness of this approach is demonstrated in this work through a specific case study, using the proposed system to achieve a powerful tool for mineral prospectivity. The system consists of three main phases: knowledge and valuable data acquisition, modeling, and results representation using prospectivity heat maps. In the initial step, the recovery and representation of expert knowledge for the case of study was conducted. This system design was tested in the Almadén Mercury Mining District, it involved interviewing expert geologists with ages of experience in the area. Afterwards, the gathering of georeferenced data was carried out to enrich the dataset. Following this phase, the modelling was done, first, using unsupervised techniques to unveil the underlying structure and patterns of the information. Later, employing supervised learning and knowledge representation techniques to enhance the results. In the final step, prospectivity maps were created to represent the achieved results to help in decision making.

Keywords: Hybrid intelligent systems, Mineral Exploration, Artificial intelligence.

1 Introduction

In today's world, with the proliferation of smartphones, GPS-enabled devices and various sensors, georeferenced data is growing at an incredible pace. From optimizing traffic in cities to finding best areas for crops in agriculture, the potential behind this data is beyond imagination, industries like transportation, environmental monitoring, agriculture, or mining rely heavily on it to make well informed decisions.
There are a multitude of existing approaches to analyze this data, ranging from traditional commonly used methods to sophisticated advanced artificial intelligence (AI) techniques that employ complex models to extract well-hidden information within the data. Despite the undeniable importance of this data, its full potential remains underexploited, and more efforts are required to fully leverage its capabilities. In addition to advancing in analytical methods, efforts should also be focused on visualization meth-

ods, given that humans comprehend visual representation more effectively than numerical data alone. Heatmaps are powerful techniques to intuitively show the outcomes of georeferenced information studies.

2 Motivation

The development of a system that combines both knowledge and georeferenced data offers numerous benefits, including enhanced generalizability, handling of uncertainty and better interpretability of outcomes. While most works employ either data-driven or knowledge-driven approaches, there are some authors that use data and knowledge integrations at the same time obtaining promising results [5].
In this thesis, an intelligent hybrid system is proposed to leverage georeferenced data using state-of-the-art AI techniques. This method takes advantage of the best of each approach: the scalability of data driven techniques and the domain expertise in knowledge-driven based methods. This multidisciplinary procedure not only enhances the quality of results but also facilitates a deeper understanding of complex geospatial information.

3 Case study: Mineral prospectivity in the Almadén Mining District

The decrease in the number of discovered mineral deposits, caused by the nonstop exploitation of mineral resources for more than three centuries, is a reality. Some authors even argue that valuable minerals like antimony, molybdenum and zinc may be exhausted within a few decades if the extraction continues at the same speed [2].
The scarce resources, combined with the unprecedented increase in demand for minerals in recent years, is leading geologists to seek more efficient methods in the exploration of new mineral deposits. Fortunately, new techniques are emerging in the field of AI to minimize both economic and human efforts in mineral exploration, when finding deposits that would not be found with other methods.
Although mercury has been extracted in numerous locations throughout the Earth's crust, one deposit stands out; this is, without a doubt, the mining deposit of Almadén. It has been estimated that, in this small district southwest of Ciudad Real, 270.000 tons of this metal have been extracted, representing approximately one-third of the total mercury consumed by humanity throughout its history.
There is a lot of known deposit in this area, as well as experts in the area that can guide hybrid intelligent systems and analyze the results.
In literature, the most common approach in this case consists of statistical models, using unsupervised or supervised machine learning (ML) techniques, but in this case, a robust system was created to leverage expert knowledge, as well as georeferenced data of known deposits of the area. This case of study offered a perfect opportunity to test the designed hybrid intelligent system, to discover its capabilities, weaknesses, and limitation. In the following lines, the proposed apprach is explained. The limits of this system

go beyond mineral prospectivity, as it was designed to be capable of leveraging any kind of georeferenced data.

3.1 Methodology

The proposed approach aims to be a method designed to be used for integrating knowledge and data, constructing a robust hybrid intelligent system that can be useful for georeferenced data leveraging. Helping in decision making in several fields like agriculture, mineral prospectivity, archaeology or fire prevention. The proposed system steps are the following ones:

1. Understanding of the case of study: To deeply understand the case study, an initial interview with experts must be conducted. The purpose of the interview is to identify needs, comprehend the field of study, assess available possibilities, and determine relevant variables.
2. Knowledge acquisition and data gathering: Interviews, surveys or any other knowledge acquisition tool must be used to integrate the maximum amount of valuable information.
3. Knowledge and data integration: To incorporate georeferenced information of different typologies, the area of study must be subdivided into smaller regions of regular shape. A regular grid is often the best option. The granularity of the grid is important for achieving best results, this corresponds to the level of detail to be used, which is something that must be chosen with the aid of experts on the field.
4. Knowledge and data leveraging: When knowledge and data are integrated, several AI techniques can be used to extract its full potential. The proposed ones in this work are the following ones:
   - 4.1. Study of constructions: This analysis explores the characteristics of the elements to achieve a better understanding of the structure of knowledge. The use of grid systems is inspired by the human cognition model developed by the psychologist George Kelly "Personal Construct Theory" [3].
   - 4.2. Observations analysis: Dendrograms generated from the results of hierarchical clustering algorithms are highly valuable as they provide visual representations of the proximity between observations, revealing groups of similar elements.
   - 4.3. Cells clustering using partition-based algorithms to distinguish regions with specific properties, for example, high probability of hosting mineral deposits.
   - 4.4. Supervised Learning and knowledge representation techniques to predict specific targets, like for example areas with high probability of hosting mineral deposits, or susceptible to fire outbreaks.
5. Prospective Mapping using the outcomes of the previous steps. To help in decision making, prospectivity maps are drawn employing the outputs of the models. The use of colour gradients in heat maps enables the identification of areas with high or low levels of potential, with warmer colours typically indicating higher levels of prospectivity and cooler colours representing lower potential. The prospectivity map generated using the outcomes of MinerIA hybrid intelligent system, is presented below (see Fig. 1).

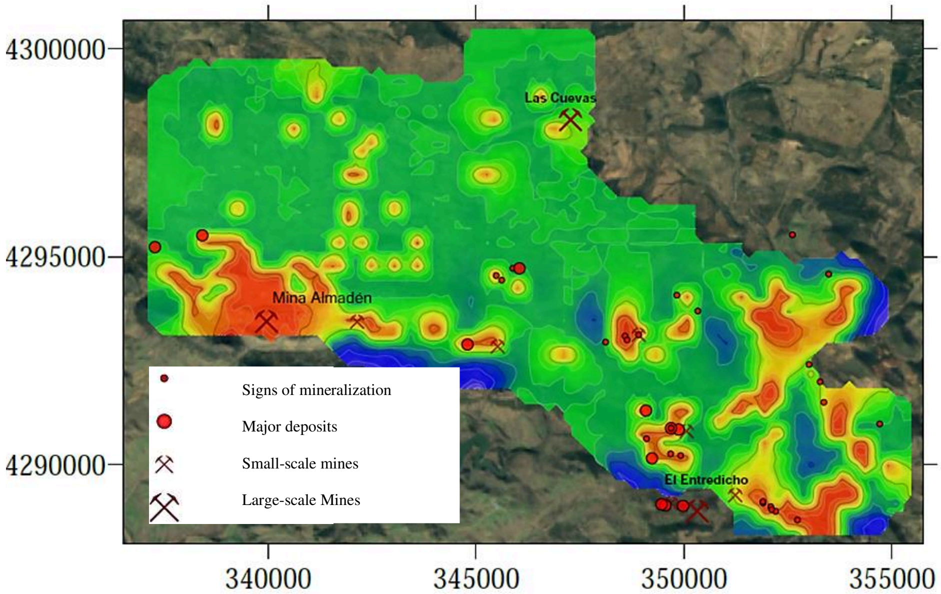

Fig. 1. The final prospectivity heat map, generated using MinerIA proposed system for mineral prospecting in the Almadén Mining districted, is depicted above.

In this map, areas with warmer colours indicate zones identified by MinerIA as having higher prospectivity potential. The map highlights the three largest known deposits in the mining district, with Mina Almadén being the world's largest mercury mineralization. Additionally, smaller deposits and signs of mineralization are represented as dots, with the dot size reflecting the magnitude of the mineralization. It is worth mentioning that the system has successfully categorized the majority of the mineralized zones as having significant potential, while assigning green or blue colours to zones with lower potential that may be discarded in the search for mercury in the area. This confirms its potential as a powerful and promising tool in advanced mineral prospecting. Moreover, the map highlights areas with high potential that lack known deposits, suggesting the possibility of discovering new mineral-rich mercury deposits in these regions.

4 Conclusions

The implementation of this hybrid intelligence system achieved promising results in mineral exploration. By integrating machine learning algorithms with expert knowledge, it has been demonstrated than more accurate results can be achieved.
The generation of prospectivity maps to visualize the results, in the form of heat maps, has been essential for understanding the results of the system. Overlaying georeferenced data on a geographic map with the predictions of the models as a gradient of

colours, offers the expert a powerful tool for decision making. In the case of mineral prospectivity, targeting specific areas with higher probability of hosting mineral deposits can lower the economic costs of a mineral exploration campaigns substantially. Furthermore, the capabilities of the proposed system may extend beyond mineral exploration. It holds promise as a versatile tool in diverse study cases, including archaeological prospection, identifying optimal locations for crops in agriculture, predicting areas with high risk of fire, and many other fields relying heavily in georeferenced data. Its ability to combine the power of machine learning with human expertise offers a handful of new opportunities to fully exploit the potential of georeferenced information.

5 Future work

Some of the future lines of work that will be studied in the near future include:

- Expanding the variety of case studies to demonstrate the generalizability and adaptability of the proposed method to different fields.
- Incorporating techniques such as soft computing [4] to handle uncertainty, like the work of Zhang, where fuzzy logic is used in mineral prospectivity [6].
- Investigating the utilization of tabular Generative Artificial Neural Networks (GANs), like recently discussed in literature [1], to create synthetic data to balance heavily unbalanced datasets.
- Incorporating additional sources of georeferenced data, such as satellite images.

Acknowledgments

This work has been partially supported by FEDER and the State Research Agency (AEI) of the Spanish Ministry of Economy and Competition under grant SAFER: PID2019-104735RB-C42 (AEI/FEDER, UE).

References

1. Bourou, S., El Saer, A., Velivassaki, T. H., Voulkidis, A., & Zahariadis, T.: A Review of Tabular Data Synthesis Using GANs on an IDS Dataset. Information, 12(9), 375. (2021).
2. Henckens, M. L., van Ierland, E. C., Driessen, P. P., & Worrell, E.: Mineral resources: Geological scarcity, market price trends and future generations. Resources Policy, 49, 102-111 (2016).
3. Kelly, G.: Personal construct psychology. Norton, New York (1955).
4. Wilhelm, Y., Reimann, P., Gauchel, W., & Mitschang, B.: Overview on hybrid approaches to fault detection and diagnosis: Combining data-driven, physics-based and knowledge-based models. Procedia Cirp, 99, 278–283 (2021).
5. Zadeh, L.A.: Fuzzy logic. Computer, 21(4), 83-93. (1988).
6. Zhang N.: Mineral prospectivity mapping with weights of evidence and fuzzy logic methods. Journal of Intelligent & Fuzzy Systems, 29(6), 2639-2651 (2015).