INTEGRATING CONCEPTS OF ARTIFICIAL INTELLIGENCE IN THE EO4GEO BODY OF KNOWLEDGE

R. Lemmens1,*, S. Lang2, F. Albrecht2, E. Augustijn1, C. Granell3, M. Olijslagers4, C. Pathe5, C. Dubois5, M. Stelmaszczyk-Górksa5

1Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, The Netherlands - (r.l.g.lemmens, p.w.m.augustijn)@utwente.nl
2Department of Geoinformatics, University of Salzburg, Salzburg, Austria - (stefan.lang, florian.albrecht)@plus.ac.at
3Institute of New Imaging Technologies (INIT), Universitat Jaume I (UJI), Spain - carlos.granell@uji.es
4Spatial Applications Division Leuven, KU Leuven, Belgium - marc olijslagers@kuleuven.be
5Department of Earth Observation, Friedrich-Schiller-University, Jena, Germany - (carsten.pathe, clemence.dubois, m.stelmas)@uni-jena.de

Commission IV, WG IV/2

KEY WORDS: Body of Knowledge, Artificial Intelligence, Machine Learning, Geocomputation, Concept model, Ontology.

ABSTRACT:
The EO4GEO Body of Knowledge (BoK) forms a structure of concepts and relationships between them, describing the domain of Earth Observation and Geo-Information (EO/GI). Each concept carries a short description, a list of key literature references and a set of associated skills which are used for job profiling and curriculum building. As the EO/GI domain is evolving continuously, the BoK needs regular updates with new concepts embodying new trends, and deprecating concepts which are not relevant anymore. This paper presents the inclusion of BoK concepts related to Artificial Intelligence. This broad field of knowledge has links to several applications in EO/GI. Its connection to concepts, already existing in the BoK, needs special attention. To perform a clean and structural integration of the cross-cutting domain of AI, a separate cluster of AI concepts was created, which was then merged with the existing BoK. The paper provides examples of this integration with specific concepts and examples of training resources in which AI-related concepts are used. Although the presented structure already provides a good starting point, the positioning of AI within the EO/GI-focused BoK needs to be further enhanced with the help of expert calls as part of the BoK update cycle.

1. INTRODUCTION

In information integration and knowledge sharing, it is a common practice to keep a set of foundational concepts and definitions to serve as a vocabulary. A general term to cover a knowledge domain is a Body of Knowledge (BoK), which is defined by DiBiase et al. (2007) as "a comprehensive inventory of the intellectual content that defines a field". BoKs have been developed for software engineering, project management and for Geographic Information Science and Technology (GIS&T BoK - DiBiase et al., 2007). The Erasmus+ Sector Skills Alliance project EO4GEO (http://eo4geo.eu) started from the GI-N2K project (Vandenbroucke and Vancauwenberghe, 2016) to develop the EO4GEO BoK (Hofer et al., 2020). The EO4GEO initiative extends the European GIS&T BoK with Earth Observation (EO) concepts and concepts about thematic and application domains. With the support of a network of almost 300 experts, both GIS and EO technology are being covered. Thereby, the BoK combines both the perspective of academia and industry. This paper describes the process of extending the set of AI-related concepts, as until now, the BoK did not reflect the increasing relevance of artificial intelligence (AI) within the EO/Geographic Information (GI) domain.

2. EO4GEO BOK - CONCEPT CONTENT AND RELATIONSHIPS

The EO4GEO BoK includes theories, methods, technologies and applications about the EO/GI domain and organises them in knowledge areas (Stelmaszczyk-Górksa et al., 2020). Each included concept has a focused description accompanied by a set of associated skills and a list of key literature references (Dubois et al., 2021). The relationships between concepts in the BoK follow the broader-narrower notion from the Simple Knowledge Organization System SKOS (Miles and Bechhofer, 2009), creating a hierarchical structure in the form of subconcept relations. A concept can be a subconcept of more than one other concept. In addition, the BoK uses other relationship types such as 'Prerequisite' (one concept needs to be known to understand the other) and 'Similarity'. In the EO4GEO project, different tools provide access to the BoK. The Living Textbook tool is the editing tool for the BoK.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLIII-B4-2022-53-2022 | © Author(s) 2022. CC BY 4.0 License.

53
It is capable of modelling knowledge domains, providing visual representation of media-rich resources and exporting concept maps, also beyond the scope of the EO4GEO project (Lemmens et al., 2020). Another tool, the BoK Visualisation and Search tool allows exploration of the BoK, following its hierarchical structure and providing access to concept history in previous versions of the BoK (see Figure 2).

3. UPDATING AND EXTENDING THE BOK

Within the EO4GEO project, the updating and extending process is organised in seven working groups (Stelmaszczyk-Górksa et al., 2020). Three groups cover the knowledge areas about geospatial information that were already present in the GI-N2K version of the BoK. Three additional working groups focus on technical and theoretical concepts of EO, while a seventh group focusses on EO application domains. All groups apply the same guidelines to develop concepts and relations within their specific knowledge areas. Additional attention is paid to ensuring cross-workgroup integration of concepts, by merging concepts present in different working groups or by creating cross-workgroup relations. The focus on extending the set of AI concepts in the BoK is exactly such a cross-workgroup initiative. Therefore, a dedicated group “AI4EO4GEO BoK” has been established to create and integrate Artificial intelligence (AI) related concepts. The AI4EO4GEO BoK working group consists of experts from existing working groups as well as additional experts in the field of AI. The latter are invited to contribute as part of an announced call for concepts from the EO4GEO expert database. In addition, all of concepts are published on different media channels: in the EO4GEO newsletter, on the project website and on the social channels of the project and its partners.

The EO4GEO expert database currently has almost 300 experts, and the experts who gave their consent are visible on the list of contributors2 and in the BoK itself along with the concepts they contributed. All concepts in the EO4GEO BoK have permalinks that allow them to be quoted and linked. The concepts are described in accordance with agreed procedures for adding concepts to the BoK. That is, each fully described concept has a knowledge description with references, skills, and a list of contributors. Each concept is interrelated using relationships that constitute an ontological structure. The procedure for adding concepts and their description has been described in earlier publications (Hofer et al., 2020, Dubois et al., 2021, Stelmaszczyk-Górksa et al., 2020).

The following sections will present the concepts that the AI4EO4GEO BoK working group has developed and integrated with those existing in the BoK.

4. CROSS-CUTTING CLUSTER - ARTIFICIAL INTELLIGENCE

Artificial intelligence (AI) as a broad superconcept refers to the capacity of computers to simulate processes characteristic of human intelligence. Thus, AI aims to mimic human reasoning (and action) by digital information assimilation and processing. Cognitive-relevant AI tasks include knowledge representation, automated reasoning and machine learning (ML).

Within ML, deep learning (DL) has emerged as a powerful tool based on nested convolution layers and associated weights and connectors, which has been successfully adopted in automated image analysis. Undeniably, AI has a predominant role across disciplines, including the EO/GI domain and Earth-related fields (Sun et al., 2022).

Due to the boost ML routines have experienced over recent years, both terms AI and ML are often used synonymously. This entails some terminological confusion around the term AI. For example, in the strive for making ML routines, which are often considered ‘black box’, more transparent and explainable, novel hybrid approaches try to merge physical models with learning. Consequently, the terms hybrid AI, explainable AI, or physics-aware AI relate to machine learning approaches rather than to AI as a whole. It is paramount today to encompass other (and partly older) concepts of the (wider) field of AI to enrich ML/DL tasks, for example by incorporating prior scientific knowledge. Merging the (new) data-driven paradigm, which is prior-agnostic and dominated by statistical inferences, with the physical paradigm, which can act as a constraining optimisation, better conditions learning routines and incorporates knowledge-driven inference (Reichstein et al. 2019). Types of AI (narrow AI and general AI to Super AI) may be distinguished by adaptability, performance and proficiency as compared to the human brain. Lately, physics-aware or hybrid AI is promoted as a strategy to better condition ML/DL tasks by employing physical models, principles, or even laws. Machine teaching (Lang et al., 2019) relying on established scientific knowledge is a complementary strategy to machine learning leading to a hybrid approach, where new insights from data are embedded in, and constrained by, the physical reality.

Based on a literature review we envisage the AI concept cluster in the EO/GI domain to be composed of eight major conceptual elements, see Figure 3. This classification is certainly not the only one possible, but it can be seen as a starting point for discussion and convergence to a wider accepted structure for AI concepts related to the EO/GI domain. As a superconcept, AI is inherently interdisciplinary. It borrows and unites concepts from different contributing disciplines and areas of knowledge of the BoK; therefore, multiple existing elements (concepts and subconcepts) in other areas of the BoK are linked to the AI concept cluster (see Figure 3). For example, when considering Convolutional Neural Networks (CNN) as one of the key strategies within image-related deep learning, this is a powerful tool in image processing and geospatial analysis.

To combine the AI cluster in the actual BoK hierarchy, three types of updates are proposed for the upcoming integrated version of the BoK: adding missing concepts, redirecting relationships to/from existing concepts and restructuring concept descriptions.

---

2 http://www.eo4geo.eu/tools/bok-visualization-and-search/

3 http://www.eo4geo.eu/bok/
5. INTEGRATING THE AI CLUSTER IN THE BOK

The method of getting the AI concepts into the BoK was based on two important starting points, i.e., the low number of existing AI-related concepts in the BoK and the high number of planned AI concepts. To achieve a coherent structure from an AI domain perspective, first a separate mini-BoK was created with only AI concepts. This cluster was then integrated into the existing BoK.

5.1 Integration procedure

The separate cluster is formed by taking established concepts from the already available BoK and adding new concepts. These established and new concepts are then connected into a structure that is consistent in itself. This has given established concepts new relationships with new and established concepts (see Figure 4). To complete the process of updating the BoK, a re-integration has to happen, followed by completing and updating all concept descriptions of the new cluster. The task of re-integration has two major actions. The first action adds the entire cluster to the original BoK. This not only adds all new concepts to the BoK but also causes duplicates for the established concepts. The second major action establishes all new relations and merges the duplicates to ensure consistency of the new cluster’s structure with the established BoK structure. This second action requires going through the concepts along the entire structure of the new cluster. Each concept is evaluated, its relationships checked, and possible duplicates are identified and resolved. Then the concept can be finally marked as integrated. Some cleaning is necessary where the process inserted relations that contradict each other (like loops).

The result of the integration is shown in Figure 5. It was decided to make the top concept of AI (‘Artificial Intelligence (AI) in EO and GI’) a subconcept of the Geocomputation (GC) concept and remove the temporary [AI] code labels from the concept names. Finally, the concepts are taken to the regular procedure of creating descriptions by working group experts for the new concepts and updating of descriptions of established concepts.

5.2 Deep learning and Decision trees concepts - examples of cluster integration

This section presents excerpts of the EO4GEO BoK to show the process of AI concepts integration. The structure and relationships between the concepts are explained, their description is given, and corresponding skills are listed.

The deep learning and decision trees concepts have been added to the GI-N2K version of the BoK along with the Earth observation concepts. The concepts were developed within the work of the Image processing and analysis working group. The Deep learning concept has been directly integrated with the cluster of Artificial Intelligence in EO and GI through the outgoing relationship to AI algorithms; and the Decision trees indirectly through the outgoing relationship to the Random forest concept interrelated to the AI algorithms concept. The concept of AI algorithms was created along with new concepts of AI, and Random forest within the Image processing and analysis working group, indicated with [IP] code labels. The Deep learning concept has four subconcepts: Multilayer perceptron, Long short-term
memory, Recurrent neural networks, Convolutional neural networks (CNNs) (see Figure 6). The Decision trees concept is a prerequisite of the Random forest concept (see Figure 7) and is a subconcept of Image classification and Productions systems concepts (represented in the BoK, but not shown in Figure 7).

As with all fully described concepts in the EO4GEO BoK, these concepts, in addition to relationships, consist of a description of knowledge, references, and skills. These elements are subject to constant updating. As the EO4GEO BoK is an open and freely accessible source of knowledge, it can be improved by, for example, adding annotations or, after registering as an EO4GEO expert, editing selected or adding new concepts. The full description of the Deep learning and Decision trees concepts will be available in the next edition of EO4GEO BoK - version 7.0. This new version will be released during the summer of 2022. The concept elements were written by experts registered in the EO4GEO expert database.

Deep learning
As described by Ghanbari, et al. (2022) deep learning is a subfield of AI and machine learning and is the fastest-growing trend in data analysis mainly due to the increasing data accessibility and computational processing powers (Ghanbari et al., 2021; Kussul, et al., 2017; LeCun et al., 2015; Ma et al., 2019; Reichstein et al., 2019; Zhang, et al., 2016; Zhu et al., 2017).

Decision trees
The Decision trees concept (Berta and Albrecht, 2022) describes a data mining technique that is used in different disciplines including Remote Sensing. It is derived from data only and represents the data driven or empirical approach, which requires large and good quality of input data. Therefore it does not require a thorough knowledge of the research area, e.g., spectral response of each land cover class needed for classification (Breiman, et al, 1984; Debeljak and Džeroski, 2011; Jensen, 2005; Quinlan, 1986).

The full knowledge descriptions are available in the EO4GEO BoK.

Skills: Knowledge was transferred into skills formulated as verb + concept relevant statements, using active voice according to the EO4GEO taxonomy, which was developed on the basis of a revised version of the taxonomy of educational objectives prepared by Anderson and Krathwohl (2001) and based on Bloom’s taxonomy (Bloom et al., 1956; Hofer et al., 2020). Skills represent the ability to apply knowledge and use expertise to complete tasks and solve problems in a certain occupational profile (by European Qualifications Framework (Cedefop, 2022a) and European Credit System for Vocational Education and Training (Cedefop, 2022b)). The skills defined for the two example concepts are below. They can be used (individually or in combination) to specify an occupational profile, a job offer or as learning outcomes of a course/curriculum.

Deep learning
- Describe how deep learning works
- Compare different deep learning approaches in EO image classification
- Identify programming languages (like Python, R, and C++) and the main open-source libraries (like OpenCV, PyTorch, TensorFlow, Google Colab, Github, Scikit-learn) that are common for deep learning
- Apply deep learning methods on EO data within online processing platforms like Google Earth Engine Cloud Computing, Amazon Web Service, Microsoft Azure, or Sentinel Hub
- Analyse the EO Image processing tools required for preparing EO data for deep learning
- Apply different DL approaches in EO imagery for classification, detection, or regression

Decision trees
- Understand the role of pruning for reducing overfitting when applying decision trees for various classification purposes
- Understand the types of decision trees and their output
- Understand the advantages and shortcomings of decision trees
- Identify the most popular decision tree algorithms
- Apply decision trees to classify land cover in an EO image

Knowledge: The knowledge description is formulated as an abstract about the concept and consists of preferably 250 to 500 words. It is prepared using recognized, state-of-the-art literature.
6. APPLICATION IN TRAINING MATERIAL AND TRAINING ACTIONS

As part of the EO4GEO project, training materials were developed in the domain of GI and EO, which are openly available on the EO4GEO website\(^4\). The materials are tagged with concepts from the BoK, which make them easily findable (see Figure 8) and facilitate referencing between training materials, training actions and labour market related items such as occupational profiles and job content. The materials are available for training actions. Several of those are reported on the EO4GEO website\(^5\) as well. The sections below provide examples of training materials and training actions offered by EO4GEO.

Figure 8. Interface for searching training materials and training actions based on concepts in the EO4GEO BoK.

6.1 Training material examples

Classification: Decision Trees: In this lecture\(^6\), the main steps required for building a Decision Tree are introduced. Two procedures used to select the best variables for splitting the tree nodes are discussed by making use of a practical example. These procedures include Gini impurity and information gain.

The learning outcomes are as follows:

- Explain how information gain and Gini impurity are calculated.
- Present the main advantages and disadvantages of decision trees classifier.
- Describe the concept of over-fitting and under-fitting.
- Define the main solutions that can be applied to avoid decision trees over-fitting.

The content of the material is tagged with BoK concepts as shown in Figure 9.

Figure 9. BoK concepts representing the subjects associated with the training material on Decision Trees. The BoK Concept Chart indicates the major domains to which the associated concepts belong (IP = Image Processing and Analysis; AM = Analytical Methods).

Object-based image analysis: This presentation\(^7\) was prepared for and used in the virtual GEOBIA summer school in 2020.

The learning outcomes are as follows:

- Understand the principles of spatial image analysis.
- Discuss specific advantages of OBIA over pixel-based approach.
- Apply OBIA concepts and analyse complex image content.
- Examine different segmentation and object-based classification methods.

The annotated concepts are shown in Figure 10.

Figure 10. BoK concepts representing the subjects associated with the training material on Decision Trees. The BoK Concept Chart indicates the major domain to which the associated concepts belong (IP = Image Processing and Analysis).

---

\(^4\) http://www.eo4geo.eu/training-material-catalogue/
\(^5\) http://www.eo4geo.eu/training-actions/
\(^6\) http://www.eo4geo.eu/training/classification-decision-trees/
\(^7\) http://www.eo4geo.eu/training/object-based-image-analysis-an-introduction/
6.2 Training action examples

Training actions covering the application of AI in Earth observation have been held in conjunction with the EO4GEO project. Two examples are highlighted below.

A webinar⁸ (see Figure 11), mainly non-technical, for all AI researchers, EO experts, GIS technicians, and technology enthusiasts was organised on 2 March 2021 by Planetek Italia. It covered the basic idea of AI algorithms (random forest, Multi-Layer Perceptron, Convolutional Neural Networks, and Generative Adversarial Networks) and how these algorithms are useful in EO applications.

Figure 11. Webinar on AI for Earth Observation, posted on the EO4GEO website⁹.

The International Summer School entitled “Intelligent Earth Observation”¹⁰, co-organised by the University of Salzburg and UNEP/GRID-Warsaw Centre, took place in June/July in full virtual mode, including a final event at the ISDE / GI Week 2021 hosted by the University of Salzburg.

The Summer School used web-based tools and satellite data processing platforms, following a case-based learning and solution approach. Technical concepts around artificial intelligence in Earth observation were employed to tackle emerging information needs from various thematic fields.

7. DISCUSSION AND FURTHER WORK

The EO4GEO BoK is a living representation of a living knowledge field and needs to be updated regularly. As the BoK serves as a basis for annotating resources, such as training and people’s profile information, it needs to be up to date. Still, an in-depth description of a course may not match for 100 percent with the available descriptions of concepts and their related skills in the BoK. However, the BoK concepts will serve the purpose of topic search and the exploration of topic context. With the integration of AI concepts in the current BoK, as presented in Section 5 of this paper, such exploration is possible through all arbitrary integration of AI concepts in the current BoK, as presented in topic search and the exploration of topic context. With the available descriptions of concepts and their related skills in the BoK, or is it enough to link to other evolving BoKs? To our knowledge, there is no single comprehensive existing ‘AI’ BoK to be reused here, but concepts are described in a variety of sources across the web. With respect to reusing these, there are pros and cons. Examples given in an ML-centric BoK will not provide examples and solutions for application in the geo-domain. On the other hand, including all methods and possible application fields would go far beyond the scope of the EO4GEO BoK. It will be more pragmatic to include generally used concepts as references and only include more elaborate descriptions of those, which are really relevant to EO/GI. Further work on fine-tuning the relationships and creating the concept definitions will be done with experts belonging to the network of the EO4GEO Skills Alliance. A close look at literature and a regular tech trend watch is needed to identify new AI applications and represent them in the BoK.

REFERENCES


http://www.eo4geo.eu/training/actions/intelligent-earth-observation/

http://www.eo4geo.eu/training/the-rise-of-artificial-intelligence-for-earth-observation/

http://www.eduacademy.at/eo4geo/course/view.php?id=8

⁸ http://www.eo4geo.eu/training/the-rise-of-artificial-intelligence-for-earth-observation/

⁹ https://www.eduacademy.at/ eo4geo/course/view.php?id=8

¹⁰ http://www.eo4geo.eu/training-actions/intelligent-earth-observation/