Supplementary Objectives: Analysing The Motivations Behind Semantic Web Machine Learning System Design

Presenter:
Artem Revenko
Semantic Web Company

1 Introduction

The combination of symbolic and sub-symbolic paradigms i.e., neuro-symbolic AI has has gained much attention in the recent years. Especially systems where Semantic Web Technologies are combined with Machine Learning (i.e., SWeML systems) showed an increasing number of publications (3; 2). SWeML systems are used to solve tasks in various applications, including text processing, graph processing, and image-related tasks.

While the primary objective of any SWeML architecture is to fulfil the task at hand, supplementary objectives refer to the underlying motivations guiding specific architectural decisions to achieve desired goals that extend beyond primary system functionalities. This talk investigates the notion of supplementary objectives in Semantic Web Machine Learning systems.

2 Preliminaries - SWeML System Framework

We build our study on top of an existing characterization framework (1), which uses *Patterns and Pattern Types* as main component. A pattern describes the data processing flow of a system, and consists of a series of interconnected (atomic) components: symbolic or non-symbolic data inputs/outputs or data-processing units – machine learning (ML) or knowledge representation and reasoning (KR). Depending on the way these atomic components are connected (e.g., taking one vs taking multiple inputs), SWeML patterns can be classified into pattern types (e.g., two processing units taking in one input and producing the output, combined with a processing unit taking two inputs).

3 Supplementary Objectives

SWeML system designers and engineers have different rationales behind their architectural decisions. These rationals are bound to certain desirable factors which are externally defined, including good performance, quick response time, interpretability, etc. We call these rationales for the usage of specific system components *objectives*.

The primary objective of any system component is usually targeted towards primary system functionalities, i.e., to fulfil the task at hand, however, some components might be associated with *supplementary objectives*.

Supplementary objectives refer to the underlying rationales guiding architectural decisions within a system, which aim at factors exceeding primary system functionalities and entail design choices aimed at enhancing specific performance metrics or user experiences. They are thus defined for actions that modify the system in some way as they stand in comparison to a minimal or previous system.

The First Austrian Symposium on AI, Robotics, and Vision (AIROV24).

When analysing SWeML systems from their pattern perspective, supplementary objectives can be targeted on different levels and by different strategies. At component level, choice of individual components might pursue certain supplementary objective. For example, altering the used model from a deep learning network to a decision tree classifier in an ML component might target the explainability of the outcomes. As another example, system designers might change the exploitation of input data source to better support the generalisation capabilities of a downstream classifier (ML component). On pattern level, the (types of) components or the processing flow itself might be adapted, such as the addition or removal of certain components, or exchanging a KR for an ML component.

In the course of our study we were able to identify a non-extensive list of 7 upper-level supplementary objectives: *Performance* (e.g., recall or on rare classes), *Data* (e.g., reduce size of required training data or better handle noise in data), *Explainability* (including understandability and/or interpretability), *Generalisation* (e.g., to new classes or to new domains), *Execution Time* (e.g. decrease training time or inference time), *Model Tuning* (e.g., for more robustness or to improve feature selection), and *Usability* (from an end-user perspective). These upper-level objectives can be refined in more fine-grained categories, to help better understand and document the motivation behind SWeML system design decisions.

Supplementary objectives mirror the current landscape and research trends of SWeML systems. By analysing how, why, and when different resources and components are used in SWeML systems, we can identify exploitation patterns, as well as desired characteristics of these technologies and provide guidance to developers on how to build more effective resources, especially in the Semantic Web domain. Finally, a deeper understanding of the problems that authors are trying to solve with the use of particular resources and components forms the basis for an empirical evaluation of the effectiveness of their integration into SWeML systems in achieving their goals

4 Literature Analysis

In ongoing work, we conducted a literature study where we focus on the analysis of supplementary objectives for the inclusion of additional inputs as architectural decision in existing work. Our analysis is based on a dataset resulting from a Systematic Mapping Study (1), which analysed 476 SWeML systems. We first pre-select those papers that are relevant for our research question, i.e., those systems with more than one input, which leaves us with 293 systems. For these we check if and which input can be considered additional, i.e., an input consumed by the system that is not necessary to fulfil the task at hand. We identified 178 systems as having at least one optional input, for which we conducted a deeper analysis.

The majority of the analysed systems added inputs were symbolic in nature, indicating a prevalent trend in leveraging symbolic representations to enhance system functionalities. However, it's noteworthy that one fifth of the authors chose to include additional non-symbolic inputs, mostly for systems employing Graph-based deep learning methods or focusing on improving Generalization or Usability. An intriguing finding is the frequent co-occurrence of Explainability and Usability objectives within many systems, highlighting a notable relationship between these two aspects. Furthermore, our analysis revealed a class of hybrid systems, that offer the flexibility of considering either symbolic or non-symbolic inputs as optional. Despite the dominance of Performance objectives, the research landscape showcases a diverse array of goals, with an increasing number of systems pursuing multiple objectives. This trend reflects the evolving demands within the AI field, underscoring the necessity for systems to accommodate a broader spectrum of requirements for effective performance and adaptation.

Acknowledgements

This work was supported in part by the research project OBARIS, which has received funding from the Austrian Research Promotion Agency (FFG) under grant 877389, and by the research project FAIR-AI, which has received funding from the Austrian Research Promotion Agency (FFG) under grant 904624.

References

[1] Breit, A., Waltersdorfer, L., Ekaputra, F.J., Sabou, M., Ekelhart, A., Iana, A., Paulheim, H., Portisch, J., Revenko, A., Teije, A.T., Van Harmelen, F.: Combining machine learning and semantic web: A systematic mapping study. ACM Comput. Surv. **55**(14s) (jul 2023). https://doi.org/10.1145/3586163, https://doi.org/10.1145/3586163

[2] Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., Melo, G.D., Gutierrez, C., Kirrane, S., Gayo, J.E.L., Navigli, R., Neumaier, S., et al.: Knowledge graphs. ACM Computing Surveys (CSUR) **54**(4), 1–37 (2021)

[3] Sarker, M., Zhou, L., Eberhart, A., Hitzler, P.: Neuro-symbolic artificial intelligence current trends. arXiv:2105.05330 (2021)