ALMA-952091

ALMA

Human-centric algebraic machine learning

AML-DL Consistency Checker - Initial

Deliverable D3.8

WP3 - AML Description Language

European Commission

This project has received funding from de European Union's Horizon 2020 research and innovation programme under grant agreement No 952091.

Consortium

ALGEBRAIC AI · Champalimaud Foundation · EPROSIMA The Middleware Experts · DFKI German Research Center for Artificial Intelligence · FIWARE

Inria · Technische Universität Kaiserslautern · uc3m · VTT

Project

Title:	ALMA: Human Centric Algebraic Machine Learning
Acronym:	ALMA
Coordinator:	Proyectos y Sistemas de Mantenimiento SL (eProsima)
Grant agreement:	952091
Call:	H2020-EIC-FETPROACT-2019
Program:	Horizon 2020
Start:	01 September 2020
Duration:	48 months
Website:	https://alma-ai.eu
E-mail:	alma@eprosima.com
Consortium:	**Proyectos y Sistemas de Mantenimiento SL (eProsima) (EPROS)**, Spain **Champalimaud Foundation (CF)**, Portugal **ALGEBRAIC AI (ALG)**, Spain **German Research Center for Artificial Intelligence (DFKI)**, Germany **Inria (Inria)**, France **TU Kaiserslautern (TUK)**, Germany **FIWARE Foundation e.V. (FF)**, Germany **Technical Research Centre of Finland (VTT)**, Finland **Carlos III University (UC3M)**, Spain

Deliverable

Number:	**D3.8**
Title:	**AML-DL Consistency Checker - Initial**
Month:	9
Work Package:	WP3 - AML Description Language
Work Package leader:	Algebraic AI (ALG)
Deliverable leader:	Algebraic AI (ALG)
Deliverable type:	Report (R)
Dissemination level:	Public (PU)
Date of submission:	2021-05-31
Version:	v0.3
Status:	Finished

Version history

Version	Date	Responsible	Author/Reviewer	Comments
v0.1	2021-05	ALG	Fernando Martin-Maroto	Draft
v0.2	2021-05	ALG	Nabil Abderrahaman-Elena	Draft
v0.3	2021-05	ALG	Fernando Martin-Maroto	Final review

Executive summary

Algebraic Machine Learning (AML) is a novel machine learning technique based on algebraic representations of data. Unlike statistical learning, AML algorithms are robust regarding the statistical properties of the data and are parameter-free. AML systems can retain what they learn, including sociocultural ethical practices for human–artificial intelligence systems. The aim of the EU-funded ALMA project is to study the mathematical properties of AML and to leverage them to develop a new generation of interactive human-centric machine learning systems. These systems are intended to minimize bias in order to facilitate trust while better enabling distributed collaborative learning.

The AMLDL Consistency Checker is a software tool to study the consistency of algebraic semantic embeddings into semilattices. Given a semantic embedding expressed using the AML Description Language, it can determine if the embedding is consistent or not. If the embedding is not consistent the AMLDL Consistency Checker indicates the potential primary and secondary sources of the inconsistency.

This deliverable contains the AMLDL Consistency Checker as well as examples of consistent and inconsistent embeddings.

Contents

Acronyms

AML	Algebraic Machine Learning.
AMLDL	AML Description Language.
AMLDL-CC	AMLDL Consistency Checker.
ML	Machine Learning.

1 AMLDL Consistency Checker

AML[1] is a Machine Learning (ML) method that relies on calculating semantic embeddings into algebras with idempotent operators such as semilattices [2]. The user should describe the ML problem at hand using a set of algebraic order equations and inequations. These equations are referred as the "embedding" and the solutions of the embedding as the "model". If the set of equations has at least one solution then the embedding is said to be consistent. Inconsistent embeddings have no solutions, i.e. they have no representation as semilattice models.

The AMLDL-CC is a software tool to study the consistency of algebraic semantic embeddings into semilattices. The AMLDL-CC can determine if an embedding expressed in AMLDL is consistent or not. In case that the embedding is inconsistent, the AMLDL-CC pinpoints the potential primary source of the inconsistency and the possible secondary sources. The primary source is a negative duple that cannot be satisfied because it contradicts other duples of the embedding. The AMLDL-CC will indicate the AMLDL instruction from which said negative duple originates. To help understand how the inconsistency arises, the AMLDL-CC also shows the AMLDL instructions that produce duples which, collectively or individually, contradict the primary source duple. If the embedding contains more than one inconsistency the AMLDL-CC will find and display all of them.

Further details about semantic embeddings are provided in *Deliverable D3.1 - AML-DL Specification - Initial* and [3].

1.1 Components of the AMLDL-CC

The AMLDL-CC combines an interpreter and an analyzer:

- An interpreter for AMLDL that transforms AMLDL code into a set of positive and negative duples.
- An analyzer that uses these duples to determine if there is at least one semilattice model compatible with them.

If there is at least one semilattice model that satisfies all the duples then the code is considered consistent.

1.1.1 The AMLDL interpreter

The AMLDL specification is presented in *Deliverable D3.1 - AML-DL Specification - Initial*. The interpreter fully implements that specification, transforming the AMLDL instructions into a set of duples. Those duples represent order relations, with each duple describing the element on its left-hand-side, right-hand-side, and whether such order relation is positive, $<$, or negative, $\not<$.

1.1.2 The consistency analyzer

The methodology and mathematics necessary to determine the consistency of an embedding are fully described in [1, sec. 2.6].

In [1, sec. 2.6], two directed graphs are used: one containing the constants (minimum units to describe the components of the problem) and the terms (idempotent summation of constants); and a second one, the dual, that mirrors the first graph with reverted relations. The dual contains constants and atoms but not terms as every term in the master algebra has a constant as its dual. A model for the dual is first created making use of the positive and negative duples from the interpreter. The edges of the graphs are established using the following steps.

- The first step is to satisfy the reverted negative relations, $\not<$. This may involve adding some *dual-of-atoms* to the dual graph.

- Reverted positive relations, <, are also satisfied by simply adding the necessary edges and using transitive closure of the graph.
- After transitive closure, all reverted relations are satisfied if and only if the embedding is consistent.

To determine the consistency of an embedding, the AMLDL-CC implicitly calculates the dual. The embedding is consistent if it can build a model for the dual until the end.

2 Demonstrator

This deliverable contains the AMLDL-CC v0.1. Two consistent and two inconsistent semantic embeddings are also provided as examples.

2.1 Instructions

Instructions for installation and usage are includes in the `README.md` file together with the program.

2.2 Examples

Two consistent embeddings are provided for illustrative purposes. The same two embedding are also provided with an additional AML-DL code line that turns them inconsistent.

The two examples included correspond to the algebraic semantic embedding of the n-queens completion problem, as described in [1], and the embedding of a cyclic group [3].

3 Future work

3.1 Current state

The current AMLDL-CC is fully functional and in use. It can assess the consistency of hard and complex semantic embeddings such as the ones provided as examples.

3.2 Potential future enhancements

Further enhancements are planned, with special consideration with newcomers to this technology. In particular, it is important to make it an easy and friendly tool for the users. Additionally, new functionalities for debugging of semantic embeddings could be developed and integrated in this tool.

References

[1] Fernando Martin-Maroto and Gonzalo G. de Polavieja. "Algebraic Machine Learning". In: *arXiv:1803.05252* (2018).

[2] Stanley Burris and H. P. Sankappanavar. *A course in universal algebra*. Clarendon Press, 1981.

[3] Fernando Martin-Maroto and Gonzalo G. de Polavieja. "Semantic Embeddings in Semilattices". In: *In preparation* (2021).

