



Human-centric algebraic machine learning

AML Complete Foundations

Deliverable D2.3

WP2 - Fundamentals of Interactive Algebraic Machine Learning



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

Consortium



ALGEBRAIC AI



Chantalimaud
Foundation



German
Research Center
for Artificial
Intelligence



TECHNISCHE UNIVERSITÄT
KAISERSLAUTERN



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: **D2.3**
 Title: **AML Complete Foundations**
 Month: 36
 Work Package: WP2 - Fundamentals of Interactive Algebraic Machine Learning
 Work Package leader: Champalimaud Foundation
 Deliverable leader: Champalimaud Foundation
 Deliverable type: Report (R)
 Dissemination level: Public (PU)
 Date of submission: 2023-09-28
 Version: v1.2
 Status: Finished

Version history

Version	Date	Responsible	Author/Reviewer	Comments
v1.0	2021.02.16	CF	Fernando Martin-Maroto and Gonzalo G. de Polavieja	arXiv manuscript 1
v1.1	2022.05.25	CF	Fernando Martin-Maroto and Gonzalo G. de Polavieja	arXiv manuscript 2
v1.2	2023.09.14	CF	Fernando Martin-Maroto and Gonzalo G. de Polavieja	final

Executive summary

[Algebraic Machine Learning \(AML\)](#) is a machine learning technique based on algebraic representations of data. Unlike statistical learning, [AML](#) algorithms are robust regarding the statistical properties of the data, can use soft and hard constraints and are parameter-free. 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 have the potential to minimize bias and facilitate trust while better enabling distributed collaborative learning.

This is *Deliverable D2.3 - AML Complete Foundations* of the ALMA project, consisting initially in the submission of a manuscript of AML Foundations to a journal. We have found more suitable the submission of three manuscripts.

In the manuscript Finite Atomized Semilattices, we extend finite semilattices to atomized semilattices by adding additional elements (atoms). We give the full crossing operator to include new information into the model and define redundant atoms. In the manuscript Semantic Embeddings in Semilattices, we define semantic embeddings into semilattices and introduce four different kinds of embeddings, and show how to use atomized semilattices to characterize them. In Infinite Atomized semilattices, we extend atomized semilattices to the infinite case. In this report, we give an explanation of the main results of these submitted manuscripts.

The manuscripts are:

F. Martin-Maroto and G. G. de Polavieja, Finite Atomized Semilattices, submitted. See also [arXiv:2102.08050 \(link\)](#)

F. Martin-Maroto and G. G. de Polavieja, Semantic Embeddings in Semilattices, submitted. See also [arXiv:2205.12618 \(link\)](#)

F. Martin-Maroto, Antonio Ricciardo, David Mendez and G. G. de Polavieja, Infinite Atomized Semilattices, submitted.

Contents

Executive Summary	4
Contents	5
Acronyms	6
1 Mathematical foundations of AML I: Finite atomized semilattices	7
2 Mathematical foundations of AML II: Sematic embeddings in semilattices	7
3 Mathematical foundations of AML III: Infinite atomized semilattices	8

Acronyms

AML Algebraic Machine Learning.

We are submitting three manuscripts for publication. In the following, we give a description of each of these manuscripts.

1 Mathematical foundations of AML I: Finite atomized semilattices

In our manuscript Finite Atomized Semilattices we introduce a new methodology to describe and compute semilattices. Finite atomized semilattices extend finite semilattices with additional elements (atoms). In addition to the atomic representation of semilattices, the methodology also consists in the full crossing operator which computes congruences of a semilattice model expressed in its atomic form. This is central to the methodology because it allows the modification of a semilattice to incorporate a new fact to be “learned”. The atomic representation differs considerably from the more common representations based on graphs. The atomic representation allows the calculation of very large semilattices and is ideal to work with evolving models. We depart from the most basic practice of naming elements in the semilattice as it is no longer convenient when models change. Instead, a set of generator constants and a term from the freest model (the term algebra) is used and the atomic representation provides the information needed to establish the mathematical properties of the element represented with the term. A term identifies an element across evolving models. In the manuscript, various properties of the atomic representation are described. The connection of the atomic representation and a subdirect product representation is established as well as the mapping of the atoms with subdirectly irreducible components of the semilattice algebra. Another concept introduced in Finite Atomized Semilattices is the concept of redundant and non-redundant atom. The non-redundant atoms can generate the redundant atoms so we only need to worry about the first when doing calculations. This is useful in practice as there are exponentially many more redundant atoms than non-redundant.

These results have been sent to the arXiv as a technical report:

F. Martin-Maroto and G. G. de Polavieja, Finite atomized semilattices, arXiv:2205.12618 ([link](#))

And has been submitted as:

F. Martin-Maroto and G. G. de Polavieja, Finite Atomized Semilattices (submitted)

2 Mathematical foundations of AML II: Sematic embeddings in semilattices

The manuscript Semantic Embeddings in Semilattices explains how to encode a problem in a semilattice and how to find its solutions from the problem description alone. The manuscript uses the atomic representation for calculations and also for the theoretical work. This makes clear the power of the atomic representation not only as a calculation tool but also as a theoretical device. The manuscript starts by explaining the idea of a semantic embedding in a semilattice. Then, four types of embeddings are described: concise, complete,

tight and explicit. An embedding can belong to various of these types at the same time. Their mathematical properties differ considerably so theorems in the document often start with “assume the embedding is complete” or “assume the embedding is not concise”, and so on. The manuscript gives semantic embeddings for four problems. For each problem, we start with the simplest embedding and then build upon it to create better semantic embeddings with more convenient properties. We start with the relatively simple problem of generating images with vertical bars in them, and then we give embeddings for the N-Queen Completion problem, Sudoku and the Hamiltonian Path problem. The main takeaways from the paper are: - It is possible to construct useful semantic embeddings for complex problems in a minimalist algebra such as a semilattice. - For tight embeddings, we can find models for every solution of a problem in the set of irreducible models of the embedding atomized with non-redundant atoms (theorem 24). In other words, for a tight embedding of a problem, the solutions of the problem are given by subsets of the non-redundant atoms of the freest model of the embedding theory. In this way, we have moved from searching for a needle in a haystack to searching for any needle among many needles in a box of matches. - Theorem 22 describes a practical mechanism to make every embedding tight. The embeddings given in the document can be used in practice to find solutions to the various problems considered. For example, it is possible to resolve sudokus and to invent new sudokus using the embedding provided. Particularly interesting is the possibility of finding Hamiltonian paths in practice.

These results have been sent to the arXiv as a technical report:

F. Martin-Maroto and G. G. de Polavieja, Semantic Embeddings in Semilattices, arXiv:2205.12618 ([link](#))

And has been submitted as:

F. Martin-Maroto and G. G. de Polavieja, Semantic Embeddings in Semilattices (submitted)

3 Mathematical foundations of AML III: Infinite atomized semilattices

This manuscript extends the results of the manuscript Finite Atomized Semilattices to the infinite case. We introduce the axiomatization of possibly infinite atomized semilattices and show that it still makes sense in the infinite setting. We also discuss how to keep track of generating elements of a semilattice, which we call constants, across algebras of terms and how our axioms allow us to identify atoms across every model generated by the same constants. The concepts of the finite case are extended to the infinite case: full crossing operator and atom redundancy. In the infinite case, we require two different notions of redundancy: redundancy and weak redundancy. In the case of complete semilattices (which includes finite semilattices), they are equivalent concepts, but not in the general case. In the finitely generated case, we show that semilattices admit atomizations consisting exclusively of non-redundant atoms.

This work may be of interest for applied and pure mathematicians. Most importantly, it provides the necessary basis to extend AML to other algebras beyond semilattices, which has been the main motivation

for this effort. Examples of such algebras (which are intrinsically infinite) include continuous semilattices, semilattices extended with unary operators and structures with multiple idempotent operators. The practical potential of algebraic machine learning on such algebras is considerable. For example, we already use this work for continuous semilattices which allowed us to make AML practical for problems using real-valued data.

These results have been submitted as:

F. Martin-Maroto, Antonio Ricciardo, David Mendez and G. G. de Polavieja, Infinite Atomized Semilattices (submitted)