

# Self-Organized Inductive Reasoning with NeMuS

Leonardo Barreto, Edjard Mota

Institute of Computing, Federal University of Amazonas, Manaus - Brazil  
{lapb, edjard}@icomp.ufam.edu.br

## Abstract

Neural Multi-Space (NeMuS) is a weighted multi-space representation for a portion of first-order logic designed for use with machine learning and neural network methods. It was demonstrated that it can be used to perform reasoning based on regions forming patterns of refutation and also in the process of inductive learning in ILP-like style. Initial experiments were carried out to investigate whether a self-organizing approach is suitable to generate similar concept regions according to the attributes that form such concepts. We present the results and make an analysis of the suitability of the method in the process inductive learning with NeMuS.

## 1 Introduction

Neural-Symbolic (NeSy) seeks to develop effective integration between connectionist learning and symbolic reasoning, possibly taking advantage of all statistical methods that can be applied to the features of data perceived or on the logical structure of symbolic information [Lamb *et al.*, 2017]. Moreover, the NeSy Computing community that has sought to integrate the views from AI, cognitive sciences, machine learning, ANN, computational vision and natural language processing and point out the main lines NeSy should go to meet Human-Like Computing (HLC), and in particular the Human-Like AI (HLAI) initiative. Among these points addressed, two of them are of particular interest for our experiment because it falls, somehow, in both aspects.

*Statistical Relational Learning* to integrate explanation and computation of symbolic knowledge in deep networks. Recent results have already shown that this is a possible path [Donadello *et al.*, 2017]. In the same way that the statistics of elements of a propositional (or relational) logic program can be used to induce the semantics of Artificial Neural Networks (ANN) to behave like them, e.g. C-IL<sup>2</sup>P [d’Avila Garcez and Zaverucha, 1999], self-organization can be exploited to produce meaningful maps of concepts in order to ease induction of hypothesis.

*Explainable AI* aims to develop AI models that are intelligible to humans, unlike “black box” models, which are efficient but difficult to extract knowledge, e.g., deep (or sim-

ilar) learning without symbolic interpretation. In this direction, *patterns of concepts* can be used to justify (and explain) “shortcuts” to generate recursive hypothesis from very large sets of relations without the need to compute the entire path to justify it. This is critical when the background knowledge has huge amounts of data. It could be adequately handled as regions of concepts and categories, similar to the human brain map organization. This will allow symbolic deduction to be performed as matching and inductive reasoning to use weights to prune the search space of candidates for hypotheses.

The *Shared Neural Multi-Space* (Shared NeMuS), is a Smarandache’s multispace [Mao, 2007], or the union of  $n$  spaces  $A_1, \dots, A_n$ . As such, each  $A_i$  represents the space of a characteristic distinct from total space and so it is suitable to represent different concepts of logical language. Such structure resembles the self-organization of a brain-like map as it is enhanced with adjustable weights of importance.

NeMuS was proposed in [Mota and Diniz, 2016] as a hierarchy of weighted vectors of logical elements pointing to their occurrences within a set of logical formulae in their normal disjunctive form. This structure of weights was used to generate patterns of refutation so that deduction is treated as a matching problem. The knowledge base, known as background knowledge (BK), is used to compute complementary similarities among literals and form activation regions. The present work brings another proof-of-concept that NeMuS, we claim, is suitable to deal with complex learning tasks that meet at least two of human-like AI endeavour because:

- the *self-organizing tendency of information in our brain, occurring at all levels* is easily mapped into its hierarchy.
- the *sort of segregation of information into distinct parts* needed to learn about objects, relations and rules composing them can be obtained through the adaptation of its (neural) network of weights to form regions of concepts (like refutation, similarities, etc.).

Thus, NeMuS is suitable to self-organization of maps or regions of importance on the logical structure that have a certain *relevance to what one wants to learn* about symbolic knowledge. This can be used, for example, to choose better deduction strategies that help reduce search space. This paper makes the following contributions: it shows how patterns of similarities among (mostly relational) logical formulae can

be determined; it points out how the formation of such patterns can be used, initially, as heuristics to guide the search for consistent hypothesis in inductive learning; it brings a self-organizing perspective that one may be interested to learn about a large set of relational knowledge.

The remainder of this paper is organized as follows: section 2.1 gives some brief background on NeMuS main concepts and its applications on patterns of reasoning and inductive clause learning, section 3 presents the self-organizing models we used in our experiments, section 4 describes the experiment setting up and the preliminary results found, section 5 we briefly contextualize this work in relation to others similar, and finally in section 6 concludes the work presented.

## 2 Background

### 2.1 Fundamentals of NeMuS structure

For FOL purposes, NeMuS can be defined with five spaces, one for each component of the first-order language as depicted in Figure 1. The upward arrows illustrate the distribution of weights from the bottom up. The hierarchical composition of the FOL is top-down, i.e. the semantics of a first order expression is the composition of the semantics of its subexpressions.

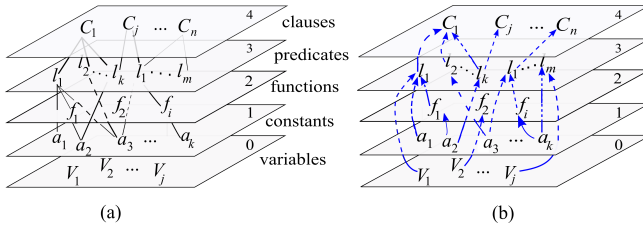


Figure 1: (a) Multi-space hierarchy, and (b) weights among elements

Note that this structure is not limited to only these spaces since clauses may influence possible worlds, possible worlds may define other higher concepts, and so on. For our purposes, we have: 0 (space) for variables, 1 for atomic constants of the Herbrand Universe, 2 for functions (suppressed here), 3 for predicates with literal instances with their own space and 4 to clauses. In what follows vectors are written  $v$ , and  $v[i]$  or  $v_i$  is used to refer to an element of a vector at position  $i$ .

The building block of NeMuS is a 4-place vector, called T-Node, used to describe logical elements. Each element is uniquely identified by an integer code (an index) within its space. In addition, a T-Node identifies the lexicographic occurrence of the element, and (when appropriate) an attribute position.

**Definition 1 (T-Node)** Let  $c \in \mathbb{Z}$ ,  $a, i, h, \in \mathbb{Z}^+$ . A T-Node (target node) is a quadruple  $(h, c, i, a)$  that identifies an object at space  $h$ , with code  $c$  and occurrence  $i$ , at attribute position  $a$  (when it applies, otherwise 1).  $\mathcal{T}_N$  is the set of all T-Nodes.

**NeMuS Binding** is an indexed pair  $(p, w)_k$ ,  $p \in \mathcal{T}_N$ ,  $w \in \mathbb{R}$  and  $k \in \mathbb{Z}^+$ , such that  $n_h(p) = h$ ,  $n_c(p) = c$ ,  $n_a(p) = a$  and  $n_i(p) = i$ . It represents the importance  $w$  of object

$k$  over occurrence  $n_i(p)$  of object  $n_c(p)$  at space  $n_h(p)$  in position  $n_a(p)$ .

**Constant Space (1)** is a vector  $C = [x_1, \dots, x_m]$ , in which every  $x_i$  is a vector of bindings. The function  $\beta$  maps a constant  $i$  to the vector of its bindings  $\vec{x}_i$ , as above.

Functions, predicates and clauses are compounds forming higher spaces. Their logical components are vectors of T-Nodes (one for each argument), and a vector of NeMuS bindings (simply bindings) to represent their instances.

**Compound** in NeMuS is a vector of T-Nodes, i.e.  $x_a^i = [c_1, \dots, c_m]$ , so that each  $c_j \in \mathcal{T}_N$ , and it represents an attribute of a compound logical expression coded as  $i$ .

**Instance Space (I-Space)** of a compound  $i$  is the pair  $(x_a^i, w_i)$  in which  $w_i$  is a vector of bindings. A vector of I-Spaces is a NeMuS *Compound Space (C-Space)*.

A literal (predicate instance), is an element of an I-Space, and so the predicate space is simply a C-Space. Seen as compounds, clauses' attributes are the literals composing such clauses.

**Predicate Space (3)** is a pair  $(C_p^+, C_p^-)$  in which  $C_p^+$  and  $C_p^-$  are vectors of C-spaces.

**Clause Space (4)** is a vector of C-spaces such that every pair in the vector shall be  $(x_a^i, \square)$ .

The following description is not the only way of building a NeMuS structure. For the purpose of this work we assume no function terms, and so only 4 spaces is needed.

**Definition 2 (Shared NeMuS)** A Shared NeMuS for a set of coded first-order expressions is a ordered 4-tuple (i.e.  $h$  is = 4),  $\mathcal{N} : \langle \mathcal{N}_v, \mathcal{N}_c, \mathcal{N}_p, \mathcal{N}_C \rangle$ , in which  $\mathcal{N}_v$  is the variable space,  $\mathcal{N}_c$  is the constant space,  $\mathcal{N}_p$  is the predicate space and  $\mathcal{N}_C$  is the clause space.

### 2.2 Inductive Clause Learning (ICL) with NeMuS

The main focus of ILP is symbolic learning of a generic logic formula  $H$ , called a hypothesis, which describes a concept, say  $\alpha$ , not yet defined in BK, having BK data and examples that characterise  $\alpha$  instances or  $\alpha^+$  positive examples, as well as  $\alpha^-$  negative examples. In other words, a formula  $H$  learned is a valid hypothesis if, and only if, the union of it with BK only yields positive deductions from the concept and not the negative ones, or formally

$$BK \cup \{H\} \models \alpha^+, \text{ but } BK \cup \{H\} \not\models \alpha^-$$

In [Mota *et al.*, 2017], Mota, Howe and Garcez showed how to make use of NeMuS to perform the same task of inductive logic programming systems presented on the literature. ICL could be performed in a system called Amao, did not explicitly used the weights, but the "intuitive use" of them was explored to define *linkage patterns* between predicates of the HB and the occurrences of ground terms in positive examples. Meaningless hypotheses are pruned away as a result of *inductive momentum* between predicates connected to positive and negative examples.

The results on inductive learning in Amao show that using its shared structure leads to reliable hypothesis generation in the sense that the minimally correct ones are generated.

More recent ILP solutions proposed the use of meta-programming to define which predicates should appear in hypothesis  $H$ , as [Muggleton *et al.*, 2015], like a bias in traditional ILP. However, this and other ILP approaches have to generate multiple hypothesis candidates, pair with a copy of BK, and test positive and negative deductions. A valid hypothesis search engine will be efficient if there is a partial ordering on the substitutions of hypotheses that subsume others, i.e. are more generic [Nienhuys-Cheng and De Wolf, 1997].

Amao use of NeMuS takes a different approach: rather than generating all possible candidates for a hypothesis, *it considers only those whose predicates of the premises contained terms from the positive examples, while excluding from negative ones*. This is achieved by walking through the Herbrand base, from the terms of the examples, using inverse unification [Idestam-Almquist, 1993].

However, when it comes to generate recursive hypothesis that demand many examples, it behaves like most approaches. This is where the importance of this experiment came about: *by indicating that (relational) literals without a direct chain of attributes can be induced as recursive if the respective regions of their predicates shows that they have connections*. A parallel work to ours, [Mota *et al.*, 2019], is actually making use of the results presented here (for awhile) as and heuristic to provide predicate invention on recursive hypothesis.

In order that Amao meets the desired self-organized behavior when learning (or reasoning), we took an approach that adapts itself to new information which will be very important for future neuro-symbolic applications like non-monotonic reasoning, and so on.

### 3 Self-Organizing Models

In this section, we presented a brief explanation about the Self-Organizing Maps and its variant used in this experiment, the Associative Self-Organizing Maps.

#### 3.1 Self-Organizing Maps

Self Organizing Maps (SOM) is an artificial neural network which transforms a given n-dimensional pattern of data into a 1- or 2-dimensional map or grid. This transformation process is done following a topological ordering, where patterns of data (synaptic or vector weights) with similar statistical features are gathered in regions close to each other in the grid. This learning process can be classified as competitive-based because neurons compete against each other to be placed at the output layer of the neuron network, but only one wins. It is also unsupervised because the neuron network learns only with entry patterns, reorganizing itself after the first trained data and adjusting its weights as new data arrive.

A detailed description of the complete SOM algorithm is presented in [Kohonen, 2001]. In what follows, we provide a summary of the main steps of how the SOM's learning process works.

1. Initialization: at the beginning of the process, all neuron vectors have their synaptic weights randomly generated. Such vectors must have the same dimension of the entry pattern space.

2. Sampling: a single sample  $x$  is chosen from the entry pattern space and fed to the neuron grid.
3. Competition: based on the minimum Euclidean distance criterion, the winning neuron  $i(x)$  is found as follows:

$$i(x) = \operatorname{argmin} \|x - w_j\|, j = 1, 2, \dots, l$$

where  $l$  is the number of neurons in the grid.

4. Synaptic adaptation: after finding the winning neuron (Best-Matching Unit or BMU), all synaptic weights of each neuron vector are adjusted:

$$W_j(t+1) = W_j(t) + \eta(t)\theta_j(t)[x(t) - W_j(t)]$$

where  $t$  represents the current instant,  $\eta(t)$  is the learning rate which gradually decreases with time  $t$ , and  $\theta_j(t)$  is the neighborhood function which determines the grade of learning of a neuron  $j$  according to its relative distance to the winning neuron.

5. Repeat steps 2 to 4 until no significant change happens in the topological map or achieve the number of epochs predefined.

#### 3.2 Associative Self-Organizing Maps

The A-SOM is described in [Johnsson *et al.*, 2009] and can be considered as a SOM which learns to associate its activity with additional ancillary inputs from a number of additional SOMs. It consists of an  $I \times J$  grid of neurons with a fixed number of neurons and a fixed topology. Each neuron  $n_{ij}$  is associated with  $r + 1$  weight vectors, where  $w_{ij}^a \in \mathbb{R}^n$  is used for the main input and  $w_{ij}^1 \in \mathbb{R}^{m_1}, w_{ij}^2 \in \mathbb{R}^{m_2}, \dots, w_{ij}^r \in \mathbb{R}^{m_r}$  are used for the ancillary inputs.

The following equations show the synaptic adaptation in the main  $w_{ijk}^a$  and ancillary  $w_{ijl}^p$  weight vectors where  $\alpha$  is the learning rate (that decreases after each iteration),  $G$  is the neighbourhood function,  $c$  is the BMU index,  $y_{ij}^a$  is the main neuron activity and  $y_{ij}^p$  the ancillary activity. At the end of each train epoch, all the weight vectors are normalized.

$$w_{ijk}^a(t+1) = w_{ijk}^a(t) + \alpha(t)G_{ijc}(t)[x_k^a(t) - w_{ijk}^a(t)]$$

$$w_{ijl}^p(t+1) = w_{ijl}^p(t) + \alpha(t)x_l^p(t)[y_{ij}^a(t) - y_{ij}^p(t)]$$

### 4 Setting Up the Experiment

The experiment has two parts. The first consists of training a SOM map of concepts,  $S_C$ , to learn about a predicate base that describes a family genealogy using only the NeMuS constant space. The second part consists of modeling an Associative SOM, train it with the NeMuS constant and predicate spaces provided by the knowledge base (KB) and compare the two approaches. Some predicates present in the KB are described below.

father(Jake, Bill)	mother(Alice, Ted)
father(Jake, John)	mother(Alice, Megan)
mother(Matilda, John)	father(John, Harry)
mother(Matilda, Bill)	father(John, Susan)
father(Bill, Ted)	mother(Mary, Harry)
father(Bill, Megan)	mother(Mary, Susan)

Table 1: The first 12 unit clauses present in the KB.

#### 4.1 SOM training and induction

The SOM  $S_C$  is generated and trained using as inputs the NeMuS constant space. The entire NeMuS structure is generated when a knowledge base is compiled. The codes for each logical element are inserted in a very efficient hashed corpus so that they are uniquely identified with space, instance, and the attribute (if it is the case) they belong to. For detailed description, we refer to [Mota and Diniz, 2016]. Our script just selects the NeMuS constant space to feed SOM training phase and it yields a map as shown in Figure 2. The circles represent the *father* predicate instances and the  $\times$ 's represent the *mother* predicate instances.

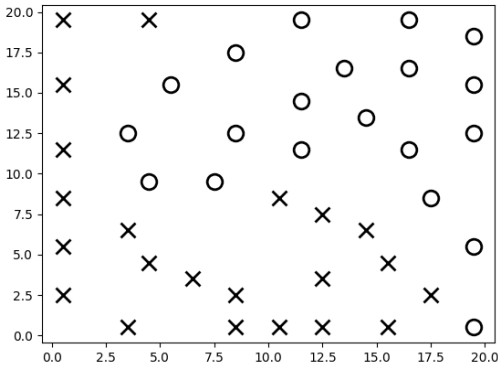


Figure 2: The 20x20 SOM grid trained using only with the NeMuS constant space.

After the training, we used  $S_C$  to induce rules that define predicate targets such as *ancestor*. In this experiment, we used Python language and the Jupyter environment and its MiniSOM library. In the following step-by-step description of how *self-organized induction* works,  $\mathcal{N}$  is a NeMuS space instance,  $E$  positive and negative examples and  $V_e$  is a vector of examples.

1. Normalize NeMuS constant space  $\mathcal{N}_c$  into  $\hat{\mathcal{N}}_c$
2. Instantiate  $S_C$  and train it with  $\hat{\mathcal{N}}_c$
3. Verify the partition of predicate regions (plot).
4. Induction over  $n$  positive or negative examples of  $E$ .
  - (a) generate NeMuS random vectors for each example  $v_e \in E$
  - (b) for each  $v \in V_e$ , do
    - i. if  $v = p(a, b)$  is a positive example then:

- A. let  $\beta_1 \cup \beta_2$  be a copy of all bindings in  $\mathcal{N}$ , in which  $a$  occurs as 1st argument and  $b$  occurs as 2nd.
  - B. train  $v_i$  with BMU weights of  $\beta_1$  and  $\beta_2$
  - C. select the BMU for the induction vector  $v_i$
- ii. if  $v = \neg q(c, d)$  is a negative example:
- A. let  $\beta_1^-$  be from a copy of  $\mathcal{N}$  that ignores all bindings of  $c$  that it occurs as 1st argument and  $\beta_2^-$  the ones that  $d$  occurs as 2nd.
  - B. train  $v_i$  with BMU weights of  $\beta_1^- \cup \beta_2^-$
  - C. select the BMU for the induction vector  $v_i$
- (c) extract the new rule based on the neurons closest to the positive induction vector
- i. If  $v_i$  is close to neurons representing the same predicate, then it assumes this characteristic as well.
  - ii. If  $v_i$  and  $v_j$ , representing the same rule target and, at end of training, they are located in different regions, then the rule is characterized by the union of these different concepts.

For example, *ancestor(a, b)* and *ancestor(c, d)* are located in the mother and father regions respectively. Therefore, we can say that an ancestor can be a father or a mother. The Figure 3 shows the SOM after the induction *ancestor(X, Y)* knowing *ancestor(Jake, John)* and *ancestor(John, Harry)*. The triangle  $v_0$  represents the induction vector of *ancestor(Jake, John)* and  $v_1$  represents the example *ancestor(John, Harry)*. From the organization of the map, we see that both vectors are near to *father* instances so we can assume that Jake is father of somebody that is ancestor of Harry.

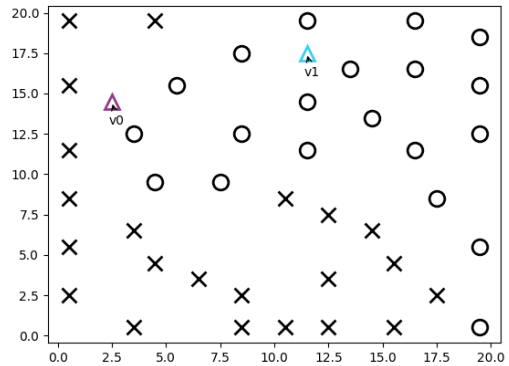


Figure 3: The 20x20 SOM grid after induce the *ancestor* rule only with positive examples.

We have run other vector examples and the results were similar to this one, that for lack of space it is not possible to show. However, the full potential of NeMuS structure of weights is their combination and both maps just considered the bindings of constants from the given vector examples. Next, we present an approach to do that.

## 4.2 A-SOM training

In this part, we modeled a 20x20 A-SOM called  $S_A$  using as main inputs the NeMuS constant space and the NeMuS predicate space as ancillary inputs. Both spaces were generated from the knowledge base before the experiment. We chose A-SOM to combine the different views and notations of the same base.

We trained the  $S_A$  map with two approaches. The first using only the NeMus constant space as well as the first part of the experiment to compare both final results. Then, we trained using the two spaces described in the last paragraph. The comparison of these maps is showed in the Figures 4 and 5. Like the first part, the circles represent the *father* predicate instances and the  $\times$ 's represent the *mother* predicate instances.

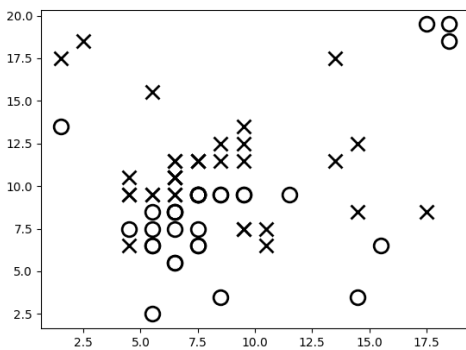


Figure 4: The 20x20 A-SOM grid trained using only with the NeMuS constant space.

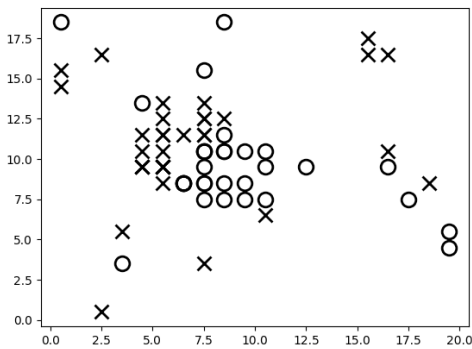


Figure 5: The 20x20 A-SOM grid trained using the NeMuS constant as main inputs and predicate spaces as ancillary inputs.

The arrangement of regions in the associative maps was shown a little different from that presented in the SOM. This was due to the shorter amount of A-SOM training times (1,000 times) compared to SOM (10,000 times).

Because it is a more complex model, which deals with main and auxiliary inputs, it has a much larger number of vectors for synaptic adaptation (the A-SOM has  $n^2(r+1)$  associated weight vectors while the SOM has  $n^2$  vectors, where  $n$  is the map dimension), it would take much longer to train the two models with the same amount of iterations and we did not have enough equipment for a deeper analysis that we expect to explore in a near future.

## 4.3 Dealing with Positive and negative examples

What differentiates the training of the induction vectors from positive and negative examples is the selection of the instances that will be used for the training. If the example is positive, we select only the instances where the constants present in the example appear and respecting the order. Otherwise, we select all other instances. In the end, the positive and even negative induction vectors will be located close to some region of similarity and this is a limitation of our model: interpret the location of the induction vectors of negative examples since it can not be similar to that described section 4.1.

## 4.4 Results

We have presented and experimented a model for reasoning over a knowledge base using Self-Organizing Map (SOM) and one of its variants called the Associative Self-Organizing Map. The A-SOM develops a representation of its input space but also learns to associate its activity with the activities of an arbitrary number of ancillary inputs. In our experiments we connected an A-SOM to a list of main and ancillary inputs provided by the knowledge base using the NeMuS space notation.

As a result, we have presented a embryonic way to recognize patterns among predicates and induce rules over them using the Self-Organising Maps. However, the existence of these regions is yet to be formally proven, although it can be clearly seen in the map plot.

## 5 Related Work

Inductive reasoning has been the building block for the successful development of ILP approaches [Muggleton *et al.*, 2012], as well as in the establishment of NeSy computing as an effective methodology for the integration of machine learning and reasoning [d'Avila Garcez *et al.*, 2019]. Both have the benefit of having logic language as the framework to generate human-interpretable explanations, not present in other ML approaches or artificial neural network (ANN) models of learning.

However, the recent advances in AI as a consequence of the groundbreaking achievements of Deep learning, [Lecun *et al.*, 2015], with its applications, e.g. [Silver *et al.*, 2017] brought attention from ANN side to unveil the "black-box" computations, although they have surpassed human intelligent abilities in some application domains. An avalanche of

works to turn deep ANN with ILP and NeSy-like features have emerged to meet the XAI challenges, e.g. [Evans and Grefenstette, 2018].

Our work brings a new contribution to XAI, although we have not had the time and resources to a massive data set experimentation or mathematical proofs of the existence of such regions before any computation starts. The inductive reasoning endowed with self-organized learning feature points out to a direction in which XAI system will not only be able to explain their computation, but also to give intuitive justification for reasoning with shortcuts like the one presented here and to evolve its learning and reasoning mechanisms as expected for a human-like AI system.

## 6 Conclusion and Future Works

This paper has shown some preliminary results of exploring NeMuS weighted structure to generate patterns of similarities from a small set of relational logical formulae. Those patterns can be used as a strategy to find recursive rules in a more efficient way. Although not yet integrated within Amao platform, the results presented here do support Amao learning strategy when dealing with potential recursive hypothesis. The patterns of similarities among relational logical formulae indicates how they can be used for inductive learning purposes.

For the lack of time, it was not possible to experiment the use of similar regions of concepts to guide predicate invention. Furthermore, the self-organizing approach brought interesting aspects that may be exploited to experiment like dynamic knowledge bases and non-monotonic learning and reasoning based on maps of concepts.

Future work will focus on such aspects as well as making more efficient use of weighted structures of concepts within Amao and interact more directly with its learning components. This will help to investigate its use on learning and reasoning of complex formulae, as well as dealing with noise, uncertainty and possible worlds. We then aim to incorporate deep learning-like mechanisms for the training of massive structured datasets.

## References

- [d'Avila Garcez and Zaverucha, 1999] Artur d'Avila Garcez and Gerson Zaverucha. The connectionist inductive learning and logic programming system. *Applied Intelligence*, 11(1):59–77, July 1999.
- [d'Avila Garcez et al., 2019] Artur d'Avila Garcez, Marco Gori, Luis C. Lamb, Luciano Serafini, Michael Spranger, and Son N. Tran. Neural-Symbolic Computing: An Effective Methodology for Principled Integration of Machine Learning and Reasoning. *arXiv e-prints*, page arXiv:1905.06088, May 2019.
- [Donadello et al., 2017] Ivan Donadello, Luciano Serafini, and Artur d'Avila Garcez. Logic tensor networks for semantic image interpretation. In *International Conference on Artificial Intelligence*, 2017.
- [Evans and Grefenstette, 2018] Richard Evans and Edward Grefenstette. Learning explanatory rules from noisy data. *Journal of Artificial Intelligence Research*, 61:1–64, January 2018.
- [Idestam-Almquist, 1993] Peter Idestam-Almquist. Generalization under implication by recursive anti-unification. In *International Conference on Machine Learning*, pages 151–158. Morgan-Kaufmann, 1993.
- [Johnsson et al., 2009] Magnus Johnsson, Christian Balke-nius, and Germund Hesslow. Associative self-organizing map. In *IJCCI*, pages 363–370, 2009.
- [Kohonen, 2001] Teuvo Kohonen. *Self-Organizing Maps*. Springer, 3rd edition, 2001.
- [Lamb et al., 2017] Luis C. Lamb, Tarek R. Besold, and Artur d'Avila Garcez. Human-like neural symbolic computing. Dagstuhl Reports 5, Creative Commons, May 2017.
- [Lecun et al., 2015] Y. Lecun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):827–832, 05 2015.
- [Mao, 2007] Linfan Mao. An introduction to Smarandache multi-spaces and mathematical combinatorics. *Scientia Magna*, 3(1):54–80, 2007.
- [Mota and Diniz, 2016] E. de Souza Mota and Yan Brandão Diniz. Shared Multi-Space Representation for Neural-Symbolic Reasoning. In Tarek R. Besold, Luis Lamb, Luciano Serafini, and Whitney Tabor, editors, *NeSy 2016*, volume 1768. CEUR Workshop Proceedings, July 2016.
- [Mota et al., 2017] E. de Souza Mota, Jacob Howe, and Artur d'Avila Garcez. Inductive learning in shared neural multi-spaces. In Tarek R. Besold, Artur d'Avila Garcez, and Isaac Noble, editors, *NeSy 2017*, volume 2003. CEUR Workshop Proceedings, July 2017.
- [Mota et al., 2019] E. de Souza Mota, Ana Schramm, Jacob Howe, and Artur d'Avila Garcez. Efficient predicate invention using shared nemus. In Artur d'Avila Garcez, Freddy Lecue, and Derek Doran, editors, *NeSy 2019*. CEUR Workshop Proceedings, August 2019.
- [Muggleton et al., 2012] Stephen Muggleton, Luc De Raedt, David Poole, Ivan Bratko, Peter Flach, Katsumi Inoue, and Ashwin Srinivasan. Ilp turns 20. *Machine Learning*, 86(1):3–23, Jan 2012.
- [Muggleton et al., 2015] S. H. Muggleton, D. Lin, and A. Tamaddoni-Nezhad. Meta-interpretive learning of higher-order dyadic datalog: predicate invention revisited. *Machine Learning*, 100(1):49–73, 2015.
- [Nienhuys-Cheng and De Wolf, 1997] Shan-Hwei Nienhuys-Cheng and Ronald De Wolf. *Foundations of Inductive Logic Programming*, volume 1228 of *Lecture Notes in Artificial Intelligence*. Springer, 1997.
- [Silver et al., 2017] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel, and Demis Hassabis. Mastering the game of go without human knowledge. *Nature*, 550:354–, October 2017.