MACHINE SYMBOL GROUNDING AND OPTIMIZATION

Oliver Kramer
International Computer Science Institute Berkeley
okramer@icsi.berkeley.edu

Keywords: autonomous agents, symbol grounding, zero semantical commitment condition, machine learning, interface design, optimization

Abstract: Autonomous systems gather high-dimensional sensorimotor data with their multimodal sensors. Symbol grounding is about whether these systems can, based on this data, construct symbols that serve as a vehicle for higher symbol-oriented cognitive processes. Machine learning and data mining techniques are geared towards finding structures and input-output relations in this data by providing appropriate interface algorithms that translate raw data into symbols. Can autonomous systems learn how to ground symbols in an unsupervised way, only with a feedback on the level of higher objectives? A target-oriented optimization procedure is suggested as a solution to the symbol grounding problem. It is demonstrated that the machine learning perspective introduced in this paper is consistent with the philosophical perspective of constructivism. Interface optimization offers a generic way to ground symbols in machine learning. The optimization perspective is argued to be consistent with von Glasersfeld's view of the world as a black box. A case study illustrates technical details of the machine symbol grounding approach.

1 INTRODUCTION

The literature on artificial intelligence (AI) defines "perception" in cognitive systems as the transduction of subsymbolic data to symbols (e.g. (Russell and Norvig, 2003)). Auditory, visual or tactile data from various kinds of sense organs is subject to neural pattern recognition processes, which reduce it to neurophysiological signals that our mind interprets as symbols or schemes. The human visual system has often referred to as an example for such a complex transformation. Symbols are thought to be representations of entities in the world, having a syntax of their own. Even more importantly, symbols are supposed to be grounded by their internal semantics. They allow cognitive manipulations such as inference processes and logical operations, which made AI researches come to believe that thinking can be referred to as the manipulation of symbols, and therefore could be considered to be computations (Harnad, 1994). Cognition becomes implementation-independent, systematically interpretable symbol-manipulation.

However, how do we define symbols and their meaning in artificial systems, e.g., for autonomous robots? Which subsymbolic elements belong to the set that defines a symbol, and – with regard to cognitive manipulations – what is the interpretation of a particular symbol? These questions are the focus of the "symbolic grounding problem" (SGP) (Harnad, 1990), and the "chinese room argument" (Searle, 1980), both of which concentrate on the problem of how the meaning and the interpretation of a symbol is grounded in action. Several strategies have been proposed to meet these challenges. For a thorough review cf. (Taddeo and Floridi, 2005).

From my perspective, the definition of a symbol depends on intrinsic structure in the perceived data and on its interpretation, which is entirely of a functional nature. "Functional" here means target-oriented: the intention to achieve goals and the success in solving problems must guide the formation of meaning and thus the definition of symbols. Hence, it seems reasonable to formulate the definition of symbols as a target-oriented optimization problem, be-

cause optimal symbols and their interpretations will yield optimal success. In many artificial systems, symbols are defined by an interface algorithm that maps sensory or sensorimotor data onto symbol tokens, detecting regularities and intrinsic structures in perceived data. Optimizing a symbol with regard to the success of cognitive operations means optimizing the interface design. From this perspective the whole learning process is self-organized, only grounded in some external feedback. This viewpoint is consistent with Craik's understanding of complex behaviors and learning: "We should now have to conceive a machine capable of modification of its own mechanism so as to establish that mechanism which was successful in solving the problem at hand, and the suppression of alternative mechanisms" (Craik, 1966). The optimization model offers mechanisms that allow to ground symbols with regards to external feedback from problem solving processes.

While only observing regularities and invariances, a cognitive system (or "agent", for short) is able to act and to predict without internalizing any form of ontological reality. This machine learning motivated point of view is related to constructivism, in which the world remains a black box in the sense of Ernst von Glasersfeld (Glasersfeld, 1987). From his point of view, his experience, i.e., the high-dimensional information perceived by sense organs, is the only contact a cognitive system has with the ontological reality. It organizes its "experience into viable representation of a world, then one can consider that representation a model, and the "outside reality" it claims to represent, a black box." (Glasersfeld, 1979). A viable representation already implies a functional component. The representation must fulfill a certain quality with regard to the fulfillment of the agent's target.

2 SYMBOLS AND THEIR MEANING

Before we can discuss the problem SGP in greater detail, let me review the current state of play. As pointed out above, according to mainstream AI claims that cognitive operations are carried out on a symbolic level (Newell and Simon, 1976). In this view I assume that an autonomous agent performs cognitive operations with a symbolic algorithm, i.e., based on an algorithm that operates on a symbolic level. An agent is typically part of the actual or a virtual world. It is situated in a real environment and this is referred to as "embodied intelligence" (Pfeier and Iida, 2003). An embodied agent should physically interact with its environment and exploit the laws of physics in that environment, in order to be deeply grounded in its world. It is able to perceive its environment with various (e.g., visual or tactile) sensors that deliver high-dimensional data, e.g., a visual system or tactile sensors. These sensory information is the used to build its cognitive structures.

In the following I assume that an artificial agent uses an interface algorithm I that performs a mapping $I : D \rightarrow S$ from a data sample $d \in D$ to a symbol $s \in S$, i.e., the I maps subsymbolic data from a high-dimensional set D of input data onto a set of symbols S. The set of symbols is subject to cognitive manipulations A. The interface is the basis of many approaches in engineering, and artificial intelligence – although not always explicitly stated. The meaning of a symbol $s \in S$ is based on its interpretation on the symbolic level. On the one hand symbols are only tokens, which may be defined independent of as their shape (Harnad, 1994). On the other hand, the effect they have on the symbolic algorithm A can be referred to as the meaning or interpretation of the symbol. Formally, a symbolic algorithm A performs (cognitive) operations on a set of symbols S, which is then the basis of acting and decision making.

In this context Newell and Simon (**?**) stated that "a physical symbol system has the necessary and sufficient means for general intelligent action". Even if we assume this to be true and if we have the means to implement these general intelligent algorithms, the question of how we can get a useful physical symbol system remains unanswered. How are symbols defined in this symbol system, how do they get their meaning? Floridi emphasizes that the SGP is an important question in the philosophy of information (Floridi, 2004). It describes the problem of how words get assigned to meanings and what meaning is. Related questions have been intensively discussed over the last few decades (Harnad, 1987; Harnad, 1990). Harnad argues that symbols are bound to a meaning independent of their shape (Harnad, 1990). This meaning-shape independence is an indication that the ontological reality is not reflected in the shape of a symbol and is consistent with. The ability to organize the perceived input-ouput relations is independent of the shape of a symbol. This can also be observed in a lot of existing machine learning approaches for artificial agents.

While it may not be difficult to ground symbols in one way or other, finding an answer to the question of how an autonomous agent is able to solve this task on its own thereby elaborating its own semantics renders much more difficult. In biological systems, genetic preconditions and the interaction with the environment and other autonomous agents seem to be

the only sources this elaboration is based on. Therefore, the interpretation of symbols must be an intrinsic process to the symbolic system itself without the need for external influence. This process allows the agent to construct a sort of "mental" representation that increases its chances of survival in its environment. Harnad derived three conditions from this assumption: First, no semantic resources are preinstalled in the autonomous agent (no innatism, or nativism respectively), second, semantic resources are not uploaded from outside (no externalism), and third, the autonomous agent possesses its own means to ground symbols (using sensors, actuators, computational capacities, syntactical and procedural resources, etc.) (Harnad, 1990; Harnad, 2007). Taddeo and Floridi called this the "zero semantical commitment condition" (Taddeo and Floridi, 2005). .

3 MACHINE LEARNING INTERFACES

How does the interface algorithm I define the symbolic system? In artificial systems it may be implemented by any machine learning algorithm that is transferring subsymbolic to symbolic representations. From the perspective of cognitive economy and dimensionality reduction respectively, it makes sense that $|S| << |D|$, i.e., the dimensionality of data in D is high while the dimensionality of symbols is low, in many cases one. Assigning unknown objects to known concepts is known as classification, grouping similar objects into a cluster is known as clustering, finding low-dimensional representations for high-dimensional data is known as dimensionality reduction. Symbolic grounding is closely related to the question that arises in machine learning of how to map high-dimensional data to classes or clusters, in particular because they are able to represent intrinsic structures of perceived data, e.g., to detect regularities and invariances. The concept of labels, classes or clusters in machine learning is very similar to the concept of symbols.

In the past, many symbol grounding related work exclusively concentrated on neural networks, see (A. Cangelosi, 2002). However, alternative algorithms used for machine learning and data mining are well suited for these types of tasks, in particular because they find legitimacy in statistics. They are complemented by runtime and convergence analyses in computer science, and they have been proven to work well in plenty of engineering applications. One may assume that these techniques are also an excellent basis for interface algorithms that perform a symbol grounding oriented transformation of subsymbolic to symbolic representations.

The classification task is strongly related to what von Glasersfeld described as "equivalence and continuity", the process of assimilation object-concept relations, i.e., to shape a present experience to fit a known sensorimotor scheme (Glasersfeld, 1979). To recognize such constructs the agent has to abstract from sensory elements that may vary or that may not be available at some time. Glaserfeld's "sensory elements" or "structures in the world" correspond to the high-dimensional features in machine learning. Similar arguments hold for the unsupervised counterpart of classification, i.e., clustering, as I will demonstrate in the following. By means of the machine learning algorithms the agent organizes its "particles of experience".

With regard to its intrinsic structure, classification, clustering and dimension reduction yield a reasonable discretization into meaningful symbols. The mapping will be data driven and statistically guided. From the point of view of statistics a data driven interface may be the most "objective" interface. From the point of view of the agent the question arises if statistical objectivity of his symbols is consistent with his needs. A biological agent is defined by his neurophysiological system, an artificial agent by his sensors and algorithms, and perception is a process that depends on the individual learning history. Consequently, the induced bias is not consistent with statistical objectivity. It is questionable if solely statistical objectivity leads to goad-driven symbol grounding.

4 INTERFACE OPTIMIZATION

Feedback is one necessary means for the internal self-organizing process of knowledge acquisition and organization. Sun calls this intrinsic intentionality when he points out that symbols "are formed in relation to the experience of agents, through their perceptual /motor apparatuses, in their world and linked to their goals and actions" (Sun, 2000). I assume that the agent's target is known and the fulfillment can be measured. To bound the symbols and their subsymbolic correlate to meanings, the interface is optimized with regard to the agent's target. The design of interface I can be formulated as an optimization problem: we want to find the optimal mapping $I^* : D \rightarrow S$ with regard to the success f_A of the symbolic algorithm A. The optimal interface I^* maximizes the success f_A, i.e.,

$$I^* = \arg\max_I \{f_A(I) | I \in \mathcal{M}\},$$

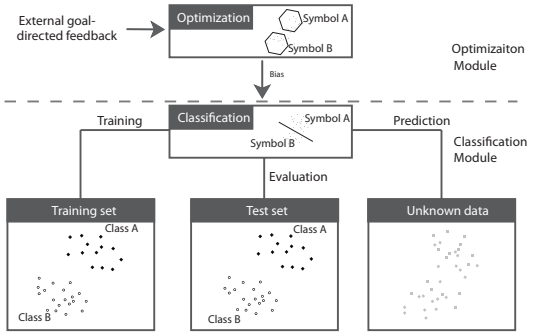

Figure 1: Classic data flow model for classification tasks (lower part), complemented by the optimization module that biases the classification task with regard to an external goal-directed feedback (upper part).

with set of interfaces or interface parameterizations $\mathcal{M}$. For this optimization formulation we have to define a quality measure f_A with regard to the symbolic algorithm A. The set S of interfaces may consists of the same algorithm with various parameterizations. In engineering practice the system constructor does not spend time on the explicit design of the interface between subsymbolic and symbolic representations. It is frequently an implicit result of the modeling process, and the system constructor relies on the statistical capabilities of the learning methods. For the adaptation of an optimal interface I^* a clear optimization objective has to be specified. The main objective is to map high-dimensional sensory data to a meaningful, a viable, set of symbols of arbitrary shape. How can this mapping be measured in terms of a feedback f_A from the symbolic algorithm? The feedback depends on the goal of the autonomous agent. If it can be explicitly expressed by a measure f_A, an optimization algorithm is able to evolve the interface. Figure 1 illustrates the optimization approach exemplarily for a classification task. The optimization module biases the classification task with regard to an external goal-directed feedback.

The optimization formulation yields valuable insights into the SGP. Learning and cognitive information processing becomes a two-level mapping, firstly from the space of subsymbolic data D to the space of symbols S, secondly, from there to the meaning of the symbols. Their semantics are implicitly bound to the cognitive process A. During interface design, the first part of the mapping is subject to optimization while the second part guides this optimization process. The whole process yields a grounding of symbols – arbitrary in shape, but based on objectives on the functional level of semantics. Decontextualization, which means to abstract from particular patterns and the ability of a symbol to function in different contexts, is less an interface design problem, but more a problem on the symbolic level. Feedback varies from situation to situation and from context to context. A sophisticated system will be able to arrange the feedback hierarchically, controlling feedback of one level on a higher level. Here, we simplify the model and have only feedback in mind feedback that is necessary to ground symbols.

Not only positive, but also negative feedback lead to self-regulation processes. It guides the optimization process of the underlying machine learning techniques. In general, the following scenarios for feedback acquisition are possible. In the offline feedback response approach the symbolic algorithm runs for a defined time, e.g., until a termination condition is met, and propagates feedback f_A that reflects its success back to the optimization algorithm. If interface design is the only optimization objective the system will adapt the interface to achieve a maximal response. This process might be quite slow if the symbolic algorithm is supposed to run for a long time to yield f_A. The offline feedback does not contradict the zero-semantical commitment condition as the feedback is the only source that guides the agent's perception of the experiential world.

5 A CASE STUDY

To concretize the described optimization perspective, I present an artificial toy scenario that is ought to illustrate the working mechanisms of the proposed approach. I start with an overview:

- **Perception**: Random Gaussian data clouds are produced at different times representing subsymbolic observations $d \in D$
- **Interface**: A clustering approach I clusters the data observations depending on two parameters, leading to an assignment of observations to symbols.
- **Hebbian translation and nterface**: Similar to the Hebbian / STDP learning rule[1] temporal information is used to translate concepts into propositional formulas. Basic inference processes (A) are used to evaluate the interface.
- **Optimization**: Free parameters are optimized, in particular w.r.t. the interface, i.e., kernel density clustering parameters.

[1] STDP means spike-timing-dependent plasticity and is known to be one of the most important learning rules in the human brain.

5.1 Perception

Let us assume that a cognitive agent observes data clouds consisting of N-dimensional points $\mathbf{x} \in \mathcal{R}$. These clouds represent subsymbolic sensory input he perceives in an observation space. The temporal context of appearance and disappearance will induces a causal meaning. Observations that belong to one concept are subsumed to concept $d_i = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ at time t. Such a data cloud is produced with the Gaussian distribution $\mathcal{N}(\nu, \sigma)$[2]. Temporal information like appearance and disappearance of data clouds is determined by a fixed scheme, see Section 5.3.

5.2 Interface

The machine learning interface has the task to assign the observations to symbols. We employ a clustering approach that assigns each cluster to a symbol. We employ a simple kernel density clustering scheme described in the following. Cluster centers are placed in regions with high kernel densities, but with a least distance to neighbored high kernel densities. For this sake, iteratively the points with the highest relative kernel density (Parzen, 1962)

$$d(\mathbf{x}_j) = \sum_{i=1, i \neq j}^{N} K_{\mathbf{H}}(\mathbf{x}_i - \mathbf{x}_j) > \varepsilon \qquad (1)$$

are identified, with a minimum distance $\rho \in \mathcal{R}^+$ to previously computed codebook vectors $\mathcal{C}$, $d(\mathbf{x}_j, \mathbf{c}_k) > \rho$ for all $\mathbf{c}_k \in \mathcal{C}$, and a minimum kernel density $\varepsilon \in \mathcal{R}^+$. These points are added to set $\mathcal{C}$ of cluster-defining codebook vectors. A symbol s_i corresponds to a codebook vector $c_i \in \mathcal{C}$ with all closest observations resulting in a Voronoi-tesselation of the observation space D. The clustering result significantly depends on the two free parameters ρ and ε that will be subject to the optimization process.

5.3 Hebbian translation and inference

From the temporal information, i.e., appearance, disappearance, and order, logical relations are induced, and translated into propositional logic formulas. This is an important and new step, and probably the most interesting contribution of this toy scenario. We employ two important rules.

1. If two symbols occur at once, e.g., s_1 at t_1 and s_2 at t_2 with $|t_1 - t_2| \leq \theta_1$, this event induces the formula $s_1 \wedge s_2$. Two concepts that occur at once are subsumed and believed to belong together from a logical perspective. This rule can be generalized to more than two symbols.
2. If symbol s_2 at t_2 occurs within time window $[\theta_1, \theta_2]$ after symbol s_1 at t_2, i.e., if $\theta_1 < t_2 - t_1 < \theta_2$, this induces the implication rule $s_1 \rightarrow s_2$. This means, s_2 follows from the truth of s_1 (another interpretation is that s_2 may be caused by s_1).

Translated into propositional logic formulas, inference processes are possible, which represent higher cognitive processes. For this sake a simple inference engine is employed. The logical formulas induced by the Hebbian process are compared to the original formulas that are basis of the data generation process. They are evaluated testing a set of evaluation formulas of the form $(A, \diamond)$ with formula A over the set of symbols, with $\diamond \in \{true, false\}$ corresponding to the data generating set.

5.4 Optimization

The optimization process has the task to find parameter settings for ρ and ε that allow an optimal inference process w.r.t. a set of evaluation formulas. For this sake we employ a $(\mu + \lambda)$-evolution strategy (Beyer and Schwefel, 2002). To evaluate the quality of the interface, we aggregate two indicators: (1) the number N_f of concepts (clusters) that have been found in relation to the real number of symbols N_s, and (2) the number K of correct logical values when testing the evaluation set formulas for feasibility. Both indicators are subsumed to a minimization problem formulation expressed in the fitness function

$$f_A := |N_f - N_s| - K. \qquad (2)$$

The problem of assigning the symbols to correct atoms is solved by trying all possible assignments, the highest K of matching logical values is used for evaluation. From another perspective, f is a measure for the performance of the *higher cognitive process* A.

5.5 Results

As a first simple test case 15 logical formulas with 10 atoms (symbols) have been generated. As evaluation set 20 formulas, 10 feasible and 10 infeasible, are used. Each symbol is represented by a data cloud with different parameterizations. A $(15 + 100)$-ES optimizes the parameter of the kernel density clustering heuristic. The optimization is stopped when no improvement could have been achieved for $t = 100$ generations. The experiments have shown that the system was able to evolve reasonable parameters for ρ and ε. The clustering process is able to identify and distinguish between concepts. In most runs the correct

[2] $\mathcal{N}(\nu, \sigma)$ represents Gaussian distributed numbers with expectation value ν, and standard deviation σ.

number symbols have been retrieved from the data cloud, the other runs only differ in at most 3 symbols. In 83 of 100 runs at least 15 formulas of the evaluation set match the observations. A careful experimental analysis going beyond this case study, and an extensive depiction of experimental results and technical details will be subject to future work. Nevertheless, I hope to demonstrate how the machine symbol grounding perspective can be instantiated.

6 CONCLUSIONS

There are many similarities between constructivism and the machine learning perspective. An autonomous agent builds its own model of the perceived environment, i.e., an individual representation depending on its physical makeup and its algorithmic learning capabilities. This point of view blends machine learning with constructivism resulting in some sort of "machine constructivism", i.e., an epistemology of artificial autonomous systems that consist of algorithmic and physical (real or virtual) machine entities. It postulates that machine perception never yields a direct representation of the real world, but a construct of sensory input and its machine. Consequently, "machine objectivity" in the sense of a consistence between a perceived constructed model and reality is impossible. Every machine perception is subjective, in particular with regards to its sensors and learning algorithms. Currently, machine subjectivity is mainly determined by a statistical perspective and biased by a set of test problems in machine learning literature. Input-output relations, i.e., regularities and invariances, are statistically analyzed and mapped to internal machine representations. They are mainly guided by learning algorithms and statistical formulas. However, from a machine learning perspective, the perception of humans and other organisms is determined by their physical and neural composition. A target-oriented optimization process binds symbol grounding to perception and a meaningful construction of representations of the environment. As future technical work, we will conduct a careful experimental evaluation of the proposed artificial case study, and extend the approach to real-world scenarios.

REFERENCES

A. Cangelosi, A. Greco, S. H. (2002). Symbol grounding and the symbolic theft hypothesis. simulating the evolution of language. pages 91–210. Springer.

Beyer, H.-G. and Schwefel, H.-P. (2002). Evolution strategies - A comprehensive introduction. *Natural Computing*, 1:3–52.

Craik, K. (1966). *The Nature of Psychology*. The University Press, Cambridge.

Floridi, L. (2004). Open problems in the philosophy of information. *Metaphilosophy*, 35:554–582.

Glasersfeld, E. (1979). Cybernetics, experience, and the concept of self. *A cybernetic approach to the assessment of children: Toward a more humane use of human beings*, pages 67–113.

Glasersfeld, E. (1987). *Wissen, Sprache und Wirklichkeit*. Viewweg, Wiesbaden.

Harnad, S. (1987). Categorical perception: the groundwork of cognition. *Applied systems and cybernetics*, pages 287–300.

Harnad, S. (1990). The symbol grounding problem. *Physica D*, pages 335–346.

Harnad, S. (1994). Computation is just interpretable symbol manipulation: Cognition isn't. *Minds and Machines*, 4:379–390.

Harnad, S. (2007). Symbol grounding problem. *Scholarpedia*, 2(7):2373.

Newell, A. and Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. In *Communications of the ACM*, pages 113–126. ACM.

Parzen, E. (1962). On the estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33:10651076.

Pfeier, R. and Iida, F. (2003). Embodied artificial intelligence: Trends and challenges. *Embodied Artificial Intelligence*, pages 1–26.

Russell, S. J. and Norvig, P. (2003). *Artificial Intelligence: A Modern Approach*. Pearson Education, 2. edition.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3:417–457.

Sun, R. (2000). Symbol grounding: A new look at an old idea. *Philosophical Psychology*, 13:149–172.

Taddeo, M. and Floridi, L. (2005). Solving the symbol grounding problem: a critical review of fifteen years of research. *Journal of Experimental and Theoretical Artificial Intelligence*, 17(4):419–445.