Three Hypotheses about the Geometry of Mind

Ben Goertzel[1] and Matthew Ikle[2]

[1] Novamente LLC, Rockville MD
[2] Adams State College, Alamosa CO

Abstract. We present a novel perspective on the nature of intelligence, motivated by the OpenCog AGI architecture, but intended to have a much broader scope. Memory items are modeled using probability distributions, and memory subsystems are conceived as "mindspaces" – geometric spaces corresponding to different memory categories. Two different metrics on mindspaces are considered: one based on algorithmic information theory, and another based on traditional (Fisher information based) "information geometry". Three hypotheses regarding the geometry of mind are then posited: 1) a *syntax-semantics correlation* principle, stating that in a successful AGI system, these two metrics should be roughly correlated; 2) a *cognitive geometrodynamics* principle, stating that on the whole intelligent minds tend to follow geodesics in mindspace; 3) a *cognitive synergy* principle, stating that shorter paths may be found through the composite mindspace formed by considering multiple memory types together, than by following the geodesics in the mindspaces corresponding to individual memory types.

1 Introduction

One of the many factors making AGI research difficult is the lack of a broadly useful, powerful, practical theoretical and mathematical framework. Many theoretical and mathematical tools have played important roles in the creation and analysis of contemporary proto-AGI systems; but by and large these have proved more useful for dealing with *parts* of AGI systems than for treating AGI systems holistically. And the general mathematical theory of AGI [6], though it has inspired some practical work [7] [12], has not yet been connected with complex AGI architectures in any nontrivial way. This paper gives a rough sketch of a novel theoretical framework intended to fill tis gap. While the framework has been developed largely in the context of a quest to understand and improve the dynamics of the OpenCog [5] AGI architecture (see [8] for some concrete OpenCog algorithmics directly related to the present ideas), it is intended to be much more broadly applicable.

For a more extensive presentation of these ideas, see `http://goertzel.org/papers/MindGeometry_agi_11_v2.pdf`. Two important background notions from that longer version are omitted here: 1) the ideas presented here are meant to be interpreted in terms of a general formal model of intelligent agents called SRAM (Simple Realistic Agents Model), presented in [4] and inspired by the

J. Schmidhuber, K.R. Thórisson, and M. Looks (Eds.): AGI 2011, LNAI 6830, pp. 340–345, 2011.

simpler agents model in [6]; 2) The multiple types of memory critical for general intelligence (declarative, procedural, episodic, attentional, intentional) may be modeled using category theory. The memory store corresponding to each memory type is a category, and then conversion from one memory type to another (e.g. declarative to procedural) is carried out using functors.

2 Metrics on Memory Spaces

We begin by explaining how to define geometric structures for cognitive space, via defining two metrics on the space of *memory store dynamic states*. Specifically, we define the dynamic state or *d-state* of a memory store (e.g. attentional, procedural, etc.) as the series of states of that memory store (as a whole) during a time-interval. Generally speaking, it is necessary to look at d-states rather than instantaneous memory states because sometimes memory systems may store information using dynamical patterns rather than fixed structures.

It's worth noting that, according to the metrics introduced here, the above-described mappings between memory types are topologically continuous, but involve considerable geometric distortion – so that e.g., two procedures that are nearby in the procedure-based mindspace, may be distant in the declarative-based mindspace. This observation will lead us to the notion of cognitive synergy.

Information Geometry on Memory Spaces. Our first approach involves viewing memory store d-states as probability distributions. A d-state spanning time interval (p, q) may be viewed as a mapping whose input is the state of the world and the other memory stores during a given interval of time (r, s), and whose output is the state of the memory itself during interval (t, u). Various relations between these endpoints may be utilized, achieving different definitions of the mapping e.g. $p = r = t, q = s = u$ (in which case the d-state and its input and output are contemporaneous) or else $p = r, q = s = t$ (in which case the output occurs after the simultaneous d-state and input), etc. In many cases this mapping will be stochastic. If one assumes that the input is an *approximation* of the state of the world and the other memory stores, then the mapping will nearly always be stochastic. So in this way, we may model the total contents of a given memory store at a certain point in time as a probability distribution. And the process of learning is then modeled as one of *coupled changes in multiple memory stores*, in such a way as to enable ongoingly improved achievement of system goals.

Having modeled memory store states as probability distributions, the problem of measuring distance between memory store states is reduced to the problem of measuring distance between probability distributions. But this problem has a well-known solution: the Fisher-Rao metric, which has been extended by Dabak [1] to handle nonparametric distributions. This metric is reviewed in the long version of this paper, together with the idea of bringing Fisher information together with imprecise and indefinite probabilities as discussed in [2]. For instance an indefinite probability takes the form $((L, U), k, b)$ and represents an envelope of probability distributions, whose means after k more observations lie in (L, U) with probability b. The Fisher-Rao metric between probability distributions is naturally extended to yield a metric between indefinite probability distributions.

Algorithmic Distance on Memory Spaces. A conceptually quite different way to measure the distance between two d-states, on the other hand, is using algorithmic information theory. Assuming a fixed Universal Turing Machine M, one may define $H(S_1, S_2)$ as the length of the shortest self-delimiting program which, given as input d-state S_1, produces as output d-state S_2. A metric is then obtained via setting $d(S_1, S_2) = (H(S_1, S_2) + H(S_2, S_1)/2$. This tells you the computational cost of transforming S_1 into S_2.

There are variations of this which may also be relevant; for instance [13] defines the generalized complexity criterion $K_\Phi(x) = min_{i \in N}\{\Phi(i, \tau_i)|L(p_i)) = x\}$, where L is a programming language, p_i is the i'th program executable by L under an enumeration in order of nonincreasing program length, τ_i is the execution time of the program p_i, $L(x)$ is the result of L executing p_i to obtain output x, and Φ is a function mapping pairs of integers into positive reals, representing the trade-off between program length and memory. Via modulating Φ, one may cause this complexity criterion to weight only program length (like standard algorithmic information theory), only runtime (like the speed prior), or to balance the two against each other in various ways.

Suppose one uses the generalized complexity criterion, but looking only at programs p_i that are given S_1 as input. Then $K_\Phi(S_2)$, relative to this list of programs, yields an asymmetric distance $H_\Phi(S_1, S_2)$, which may be symmetrized as above to yield $d_\Phi(S_1, S_2)$. This gives a more flexible measure of how hard it is to get to one of (S_1, S_2) from the other one, in terms of both memory and processing time.

One may discuss geodesics in this sort of algorithmic metric space, just as in Fisher-Rao space. A geodesic in algorithmic metric space has the property that, between any two points on the path, the *integral of the algorithmic complexity* incurred while following the path is less than or equal to that which would be incurred by following any other path between those two points. The algorithmic metric is not equivalent to the Fisher-Rao metric, a fact that is consistent with Cencov's Theorem because the algorithmic metric is not Riemannian (i.e. it is not locally approximated by a metric defined via any inner product).

3 Three Hypotheses about the Geometry of Mind

Now we present three hypotheses regarding generally intelligent systems, using the conceptual and mathematical machinery we have built.

Hypothesis 1: Syntax-Semantics Correlation. The informational and algorithmic metrics, as defined above, are not equivalent nor necessarily closely related; however, we hypothesize that on the whole, systems will operate more intelligently if the two metrics are well correlated, implying that geodesics in one space should generally be relatively short paths (even if not geodesics) in another.

This hypothesis is a more general version of the “syntax-semantics correlation” property studied in [10] in the context of automated program learning. There, it is shown empirically that program learning is more effective when programs with similar syntax also have similar behaviors. Here, we are suggesting that an intelligent system will be more effective if memory stores with similar structure and contents lead to similar effects (both externally to the agent, and on other memory systems). Hopefully the basic reason for this is clear. If syntax-semantics correlation holds, then learning based on the internal properties of the memory store, can help figure out things about the external effects of the memory store. On the other hand, if it doesn’t hold, then it becomes quite difficult to figure out how to adjust the internals of the memory to achieve desired effects.

The assumption of syntax-semantics correlation has huge implications for the design of learning algorithms associated with memory stores. All of OpenCog’s learning algorithms are built on this assumption. For, example OpenCog’s MOSES procedure learning component [10] assumes syntax-semantics correlation for individual programs, from which it follows that the property holds also on the level of the whole declarative memory store. And OpenCog’s PLN probabilistic inference component [2] uses an inference control mechanism that seeks to guide a new inference via analogy to prior similar inferences, thus embodying an assumption that structurally similar inferences will lead to similar behaviors (conclusions).

Hypothesis 2: Cognitive Geometrodynamics. In general relativity theory there is the notion of “geometrodynamics,” referring to the feedback by which matter curves space, and then space determines the movement of matter (via the rule that matter moves along geodesics in curved spacetime) [11]. One may wonder whether an analogous feedback exists in cognitive geometry. We hypothesize that the answer is yes, to a limited extent. On the one hand, according to the above formalism, the curvature of mindspace is induced by the knowledge in the mind. On the other hand, one may view cognitive activity as approximately following geodesics in mindspace.

Let’s say an intelligent system has the goal of producing knowledge meeting certain characteristics (and note that the desired achievement of a practical system objective may be framed in this way, as seeking the true knowledge that the objective has been achieved). The goal then corresponds to some set of d-states for some of the mind’s memory stores. A simplified but meaningful view of cognitive dynamics is, then, that the system seeks the shortest path from the current d-state to the region in d-state space comprising goal d-states. For instance, considering the algorithmic metric, this reduces to the statement that at each time point, the system seeks to move itself along a path toward its goal, in a manner that requires the minimum computational cost – i.e. along some algorithmic geodesic. And if there is syntax-semantics correlation, then this movement is also approximately along a Fisher-Rao geodesic.

And as the system progresses from its current state toward its goal-state, it is creating new memories – which then curve mindspace, possibly changing it substantially from the shape it had before the system started moving toward its goal. This is a feedback conceptually analogous to, though in detail very different from, general-relativistic geometrodynamics.

There is some subtlety here related to fuzziness. A system's goals may be achievable to various degrees, so that the goal region may be better modeled as a fuzzy set of lists of regions. Also, the system's current state may be better viewed as a fuzzy set than as a crisp set. In this case, one may say that the cognition seeks a geodesic from a high-degree portion of the current-state region to a high-degree portion of the goal region.

Hypothesis 3: Cognitive Synergy. Cognitive synergy is a conceptual explanation of what makes it possible for certain sorts of integrative, multi-component cognitive systems to achieve powerful general intelligence [3]. The notion pertains to systems that possess knowledge creation (i.e. pattern recognition / formation / learning) mechanisms corresponding to each multiple memory types. For such a system to display cognitive synergy, each of these cognitive processes must have the capability to recognize when it lacks the information to perform effectively on its own; and in this case, to dynamically and interactively draw information from knowledge creation mechanisms dealing with other types of knowledge. Further, this cross-mechanism interaction must have the result of enabling the knowledge creation mechanisms to perform much more effectively in combination than they would if operated non-interactively.

How does cognitive synergy manifest itself in the geometric perspective we've sketched here? Perhaps the most straightforward way to explore it is to construct a composite metric, merging together the individual metrics associated with specific memory spaces.

In general, given N metrics $d_k(x,z), k = 1 \ldots N$ defined on the same finite space M, we can define the "min-combination" metric $d_{d_1,\ldots,d_N}(x,z) = min_{y_0=x, y_{n+1}=z, y_i \in M, r(i) \in \{1,\ldots,N\}, i \in \{1,\ldots,n\}, n \in \mathbb{Z}} \sum_{i=0}^{n} d_{r(i)}(y_i, y_{i+1})$, which is conceptually similar to (and mathematically generalizes) min-cost metrics like the Levenshtein distance used to compare strings [9]. To see that it obeys the metric axioms is straightforward; the triangle inequality follows similarly to the case of the Levenshtein metric. In the case where M is infinite, one replaces min with inf (the infimum) and things proceed similarly. The min-combination distance from x to z tells you the length of the shortest path from x to z, using the understanding that for each portion of the path, one can choose any one of the metrics being combined. Here we are concerned with cases such as $d_{syn} = d_{d_{Proc}, d_{Dec}, d_{Ep}, d_{Att}}$.

We can now articulate a geometric version of the principle of cognitive synergy. Basically: cognitive synergy occurs when the synergetic metric yields significantly shorter distances between relevant states and goals than any of the memory-type-specific metrics. Formally, one may say that an intelligent agent A (modeled by SRAM) displays **cognitive synergy** to the extent $syn(A) \equiv \int (d_{synergetic}(x,z) - min\,(d_{Proc}(x,z), d_{Dec}(x,z), d_{Ep}(x,z), d_{Att}(x,z)))\, d\mu(x) d\mu(z)$ where μ measures the relevance of a state to the system's goal-achieving activity.

References

1. Dabak, A.: A Geometry for Detection Theory. PhD Thesis, Rice U. (1999)
2. Goertzel, B., Iklé, M., Goertzel, I., Heljakka, A.: Probabilistic Logic Networks. Springer, Heidelberg (2008)
3. Goertzel, B.: Cognitive synergy: A universal principle of feasible general intelligence? (2009)
4. Goertzel, B.: Toward a formal definition of real-world general intelligence (2010)
5. Goertzel, B., et al.: Opencogbot: An integrative architecture for embodied agi. In: Proc. of ICAI 2010, Beijing (2010)
6. Hutter, M.: Universal AI. Springer, Heidelberg (2005)
7. Hutter, M.: Feature dynamic bayesian networks. In: Proc. of the Second Conf. on AGI. Atlantis Press (2009)
8. Ikle, M., Goertzel, B.: Nonlinear-dynamical attention allocation via information geometry. In: Schmidhuber, J., Thorisson, K., Looks, M. (eds.) AGI 2011. LNCS(LNAI), vol. 6830, pp. 62–71. Springer, Heidelberg (2011)
9. Levenshtein, V.: Binary codes capable of correcting deletions, insertions, and reversals. Soviet Physics Doklady 10, 707–710 (1966)
10. Looks, M.: Competent Program Evolution. PhD Thesis, Computer Science Department, Washington University (2006)
11. Misner, C., Thorne, K., Wheeler, J.: Gravitation. Freeman, New York (1973)
12. Schaul, T., Schmidhuber, J.: Towards practical universal search. In: Proc. of the 3rd Conf. on AGI. Atlantis Press (2010)
13. Yi, S., Glasmachers, T., Schaul, T., Schmidhuber, J.: Frontier search. In: Proc. of the 3rd Conf. on AGI. Atlantis Press (2010)