Algorithmic Feasibility of Observing Artificial Life Evolution

Janardan Misra

HTS Research, Bangalore, India 560076
janardan.misra@honeywell.com

Abstract

An observation process is a fundamental implicit component of the simulation based studies on artificial-evolutionary systems (AES) by which time-varying entities are identified and their behavior is observed to uncover higher-level "emergent" phenomena. In this paper, we analyze algorithmic feasibility of implementing an observation process and consequent automated discovery of the entities and the evolutionary processes in arbitrary AES models. We characterize the bounds for the worst case computational complexity for the process of discovery of possible presence of entity and population level reproduction with epigenetic development in the child entities involving mutations and heredity in presence of natural selection. In particular, we prove that if entities in an AES simulation are structurally distinguishable, the problem of observability of evolutionary processes is only polynomially harder w.r.t. the entity recognition. The complexity bounds are presented in parameterized form so that for any given AES model, if parameter estimates are known, corresponding bounds can be derived.

Background

Studies on *Artificial Evolutionary Systems* (AES) are recent attempts to complement real-life theories to study the principles underlying the complex phenomena of life without directly working with the real-life organisms. For example, AES studies can complement theoretical biology by uncovering potential evolutionary dynamics (Ostrowski et al., 2007; Lenski et al., 2003).

Observations play a fundamental role in AES research, in particular, for those AES studies, which focus on the problem of the "emergence" of life-like behavior. However, the mechanisms and analysis often employed in AES studies to discover the emergent entities and their life-like behavior remain useful only to the specific models and do not always have the generic perspective. Therefore an important aspect where AES studies demand increasing focus is to study observational processes and mechanisms used in AES studies in their own right resulting into a framework for *automated discovery* of life-forms and their dynamics in simulated environments. With AES studies involving mostly digitized universes and their simulations, it is actually desirable to explore by algorithmic means potentially varied possibilities which these simulations hold yet usually require such detailed observations that it may not always be feasible to carry out for human observers alone. Such an automated discovery of life-forms and the evolving dynamics may bring much promise in AES studies as compared to what could possibly be achieved only with manually controlled observations.

An example of such an automated discovery of life forms is discussed in (Sayama, 1998). In order to observe the living loops in his Cellular Automata (CA) model, another "Observer CA" system is designed and embedded within the simulator software. The observer CA is capable of performing the complex image processing operations on the CA configuration given to it as an input by the simulator CA to automatically identify the living loops of different types. Also recently (Stone et al., 2009) have discussed the integration of artificial life simulations with interactive games-based techniques to study simulation complexity for the behavioral representation of species in fragile or long-vanished landscapes and ecosystems.

However, because of its implicit nature and the multitude of AES models, a precise characterization of the observation process is generally a difficult problem. Importantly it needs to be defined independent of the low-level micro dynamics any specific AES model to permit the study of higher-level observationally "emergent" phenomena. Initial work on systematically studying the observational processes independent of the underlying AES models appeared in Henz and Misra (2007); Misra (2009). In (Henz and Misra, 2007) an observation process is characterized as an abstraction on the model universe for establishing the necessary elements and the level of evolutionary behavior in that model. Based upon this formal characterization, in Misra (2009), it was proved that the task of entity recognition in a simulation, is a NP-hard problem and therefore cannot be completed in polynomial number of steps. In this paper we extend this result further and present computational complexity theoretic analysis for the problem of algorithmic discovery of evolutionary phenomena in AES studies. The presented analysis on observing evolutionary behavior reveals important insights on how computation intensive an automated discovery of life-like phenomena could be.

Related Work To the author's knowledge, there is not much work focusing on the algorithmic feasibility analysis of generic models for AES studies. However, interestingly, for few specific AES models, there exist parallel results. For example, Melkikh (2008) considered the computational analogue of the problem of the origin of species in a genome space under DNA Computing framework (Paun et al. (2006)) and has shown that in absence of a priori information about the possible species of organisms, the underlying computational problem is NP-hard. Similarly, Centler et al. (2008) prove that the problem of computing a reactive chemical organization is NP-hard.

Notations: Set notations: $\setminus$ (set difference), $\mathcal{P}$ (power set), $\rightsquigarrow$ (partial function). Logical operators: $\wedge$ (and), $\neg$ (not), $\Rightarrow$ (implication), $\Leftrightarrow$ (if and only if), $\exists$ (existential quantifier), and $\forall$ (universal quantifier). Programing pseudo code notation: $if \ldots then \ldots$. $\mathcal{N}^+$ is the set of positive integers. For a vector $x = (a_1, a_2, \ldots, a_r), i^{th}$ element (a_i) will be denoted as $x[i]$. Also basic notions from multiset theory (Singh et al., 2007) (e.g., $\uplus$ (multiset join)) and the theory of computational complexity (Papadimitriou, 1994; Cormen et al., 2001) (e.g., 'big-Oh' notation - $\mathcal{O}$[1]) would be used in the formal exposition of the derived results.

The Formal Structure of the Framework

In this section we will briefly review the axiomatic framework presented in Henz and Misra (2007); Misra (2009). In the ensuing discussion, we will use "AES model" and "model", "Observation process" and "Observer" interchangeably to add convenience in presentation. *Axioms* are used to specify conditions which need to be satisfied in order to draw valid inferences e.g., recognition of entities and their causal relationships.

Observation Process and the Model Universe

Axiom 1 (The Axiom of Observable Life)**.** *Life-like phenomena in a AES model exists only if it can be observed using its simulations.*

In other words, existence of life-like behavior can only be proved with respect to an observation process and associated simulations.

Definition 1 (Observation Process)**.** An observation process is an algorithmic transformation from the underlying AES simulation model to observer abstractions $(Abs_{ind},\ Abs_{dep})$, where Abs_{ind} is the set of process independent abstractions and Abs_{dep} is the set of process dependent abstractions.

Definition 2 (States)**.** Σ: set of observed states of the model across simulations.

Definition 3 (Observed Run)**.** $\mathcal{T} : \Sigma \rightsquigarrow \mathcal{P}(\mathcal{N}^+)$: An observed sequence of states ordered with respect to the temporal progression of the model during its simulation.

$\mathcal{N}^+$ acts as a set of indexes for the states in the sequence. Since a state may appear multiple times in a simulation, subsets of $\mathcal{N}^+$ are used to denote that. Each such sequence represents one *observed run* of the model. We let $\Sigma_{\mathcal{T}}$ denote the set of unique states appearing in a specific run $\mathcal{T}$.

Entity Recognition

Definition 4 (Entity Set)**.** E_s: Multiset of entities observed and uniquely identified by the observer in a state s of the model for a given run $\mathcal{T}$. $E_{\mathcal{T}} = \biguplus_{s \in \Sigma_{\mathcal{T}}} E_s$ is the multiset of entities observed and uniquely identified by the observer across the states in a given run $\mathcal{T}$.

"Tagging" can be used as a mechanism for identifying individual entities whenever there exist multiple entities in the same state which are otherwise indistinguishable.

Axiom 2 (Axiom of Unique Identification of Entities)**.** *An entity must be uniquely identified in a given observed run $\mathcal{T}$.*

Axiom 3 (Axiom of Unique Identification in States)**.** *If two states are identical, i.e., consist of the identical multisets of atomic observable structures, then an observer must identify the same multisets of entities in these states irrespective of their temporal ordering in the observed run $\mathcal{T}$.*

Axiom 4 (Axiom of non-Ignorance)**.** *It must not be true that an observer omits identification of an entity in a state s but in a different state s' identifies it as consisting of the same atomic elements which were also available in s.*

Definition 5 (**Character Space**)**.** An observer should define a set of all possible mutually independent (or orthogonal) and measurable characteristics for possible entities in the model as a multi dimensional character space $\Upsilon = Char_1 \times Char_2 \times \ldots \times Char_d$, where each of $Char_i$ is the set of values for i^{th} characteristic.

Corresponding to each entity $e \in E_{\mathcal{T}}$ there is a point in Υ, say $(v_1, v_2, \ldots v_d)$, where $v_i \in Char_i$.

Observable characteristics need not to be limited to syntactic level or *structural properties* and may also include semantic properties, which are *observable patterns of behaviors* abstracted over a range of states.

Definition 6 (**Distance Measure**)**.** An observer defines a computable clustering distance measure $D : E_{\mathcal{T}} \times E_{\mathcal{T}} \rightarrow$ **Diff**, where **Diff** is the set of values to characterize the observable "differences" between entities in E.

[1]Asymptotic order notation, $\mathcal{O}$, is used to measure the bounds on computational complexity for algorithms and problems. If $f(n) = \mathcal{O}(g(n))$, then f is said to be upper bounded by g for all the values of the input of size n after certain point. Two useful asymptotic properties of $\mathcal{O}$ are: If $f_1(n) = \mathcal{O}(g_1(n))$ and $f_2(n) = \mathcal{O}(g_2(n))$, then $f_1(n) + f_2(n) = \mathcal{O}(\max\{g_1(n), g_2(n)\})$ and $f_1(n) * f_2(n) = \mathcal{O}(g_1(n) * g_2(n))$.

Definition 7 (**Mutation Bound**). Based upon the choice of D, an observer selects $\delta_{mut} \in \mathbf{Diff}$ as a vector such that each element in δ_{mut} specifies an observer-defined threshold on the recognizable mutational changes for corresponding characteristic.

It is important to note that the choice of δ_{mut} critically affects further inferences. For example, a choice of very large values would result in the lack of identification of variability in characteristics among entities. On the other hand, with relatively smaller values for δ_{mut}, it is difficult to recognize persistence of an entity across states under changes.

Next, a *Recognition relation* is defined to establish the persistence of entities across states in the presence of mutational changes:

Definition 8 (**Recognition Relation**). An observation process establishes recognition of entities across states of the model with (or without) mutations by defining a partial function $\mathbf{R}_{\delta_{\mathbf{mut}}} \colon E_{\mathcal{T}} \rightsquigarrow E_{\mathcal{T}}$, satisfying following axioms:

Axiom 5. *Entities to be recognized as the same should be observed in successive states.*

Axiom 6. *No two different entities in one state can be recognized as the same in the next state.*

Axiom 7. *If an entity e mutates and in the next state is identified as e', observer might be able to recognize e and e' as the same only if these changes (between e and e') are bounded by δ_{mut}.*

In order to infer meaningful relationship among entities, to be used as a basis for inferring macro level phenomena in the model, an observer needs to first identify "causal" relationships among entities independent of the underlying 'physical laws' or 'micro level dynamics' of the model.

Definition 9 (**Causality**). $C \subseteq \biguplus_{s \in \Sigma_{\mathcal{T}}} E_s \times E_{s+1}$. C establishes the observed causality among the entities appearing in the successive states of a run $\mathcal{T}$.

Since causality is largely an observer and model dependent, it is further refined by defining additional axioms for specific cases, for example, for the case of reproductive causality to infer reproductive relationships among entities (See Axiom 8).

Observing Evolution

In the following discussion we will define components in Abs_{dep} for observing the fundamental evolutionary components: reproduction with mutations and epigenetic developments, heredity, and natural selection.

Reproduction An observation process establishes reproduction by defining causal descendance relationships among the entities across states, whereby parent and the child entities are recognized by the observer as being sufficiently similar and "causally" connected across the states. Formally, we add a new Axiom for the causal relation C defined before:

Axiom 8 (**Reproductive Causality**). *If an entity e in state s is causally connected to entity e' in the next state $s+1$, then there must not be any other entity e'' in state s, which is recognized by the observer as (mutating to) e'.*

In essence, this formulation of causality is an abstract specification which demands observers to identify the entities which have been observed to be causal sources for the appearance of a new entity.

Similar to δ_{mut}, as discussed before, it is important to specify the limits under which an observer can identify whether an entity is a descendant of another entity even though they might not be identical. This limit on observable reproductive mutations is essential while working with models where epigenetic development in the entities can be observed (Mahner and Bunge, 1997). This is because in such models including examples from real life, "child" entity and the "parent" entities may not have identical characteristics the beginning and therefore an observation process needs to wait until whole epigenetic developmental process gets unfolded and only then compare the entities for similarities in their characteristics.

Definition 10 (**Reproductive Mutation Bound**). Based upon the choice of D, the observer selects $\delta_{rep\_mut} \in \mathbf{Diff}$, which will be used to bound reproductive mutational changes for proper recognition.

δ_{rep_mut} assists an observer to establish whether a particular entity could be treated as a "descendant" of another entity or not. It is important to note that the choice of δ_{rep_mut} also critically affects further inferences. For example, small values for δ_{rep_mut} might make it harder to establish reproductive relationships among entities and for such an observer every new entity would seem to be appearing *de novo* in the model. On the other hand choice of very large values would result in the lack of identification of variability in characteristics and thus make it difficult to infer natural selection.

An auxiliary relation Δ is used to determine that the differences due to reproductive mutations are bounded by δ_{rep_mut}.

Definition 11. $\Delta \subseteq E_{\mathcal{T}} \times E_{\mathcal{T}}$ s.t. $\forall e, e' \in E_{\mathcal{T}}$. if (e, e') is in Δ then their differences for each single characteristic $char_i$ must be bounded by $\delta_{rep\_mut}[i]$ and e should not be recognized as mutating to e'.

Based on the thus established notion of "causal" relationships between entities and Δ, we define $\mathbf{AncestorOf}$ relation, which connects entities for which an observer can establish descendance relationship across generations.

Definition 12. $\mathbf{AncestorOf} = (C \cup \mathbf{R}_{\delta_{\mathbf{mut}}})^+ \cap \Delta$

In this definition the transitive closure of $(C \cup \mathbf{R}_{\delta_{\mathbf{mut}}})$ captures the observed causality (C) across multiple states even in cases when "parent" entities might undergo mutational changes ($\mathbf{R}_{\delta_{\mathbf{mut}}}$) before "child" entities complete their "epigenetic" maturation with possible reproductive mutations. Intersection with Δ ensures that causally related parent and child entities are not too different from each other, that is, reproductive mutational changes are under observable limit.

Using $\mathbf{AncestorOf}$ relation, we now can consider the cases of *entity level reproduction* and *Fecundity*:

Case 1: Entity Level Reproduction We consider the case where instances of individual entities can be observed as reproducing. For a given simulation $\mathcal{T}$ of the model, an observer defines the following $\mathbf{Parent_\Delta}$ relation:

Definition 13. $\mathbf{Parent_\Delta} = \{(p, c) \in \mathbf{AncestorOf} \mid \nexists e \in E_{\mathcal{T}} \,.\, [(p, e) \in \mathbf{AncestorOf} \wedge (e, c) \in \mathbf{AncestorOf}]\}$

The condition in defining $\mathbf{Parent_\Delta}$ is used to ensure that p is the immediate parent of c and thus there is no intermediate ancestor e between p and c. Using $\mathbf{Parent_\Delta}$ relation, in order for the observer to establish reproduction in the model, the following axiom should be satisfied:

Axiom 9 (**Reproduction**). *There should exist at least one instance of reproduction in a simulation $\mathcal{T}$ of the model i.e.,* $\mathbf{Parent_\Delta} \neq \emptyset$.

Since for every $(p, c) \in \mathbf{Parent_\Delta}$, some other $(p', c') \in \mathbf{AncestorOf}$ where p (and/or c) has been observed to change to p' (c') may also be present in the $\mathbf{Parent_\Delta}$, therefore, let $\mathbf{Parent_\Delta^{min}}$ consist of temporally least parent-child pairs (p, c) from $\mathbf{Parent_\Delta}$.

Case 2: Population Level Reproduction - Fecundity Owing to the *carrying capacity* of the environment, which limits the maximum possible size of a population, for natural selection it is the population level collective reproductive behavior (fecundity), which is significant. Therefore in order to ensure that there is no perpetual decline in the size of the population, following axiom should hold:

Axiom 10 (**Fecundity**). *There exist statistically significant number of different generations of reproducing entities in temporal ordering $G_1, G_2, \ldots, G_L$ such that for every generation of reproducing entities, there exists a generation of its descendant entities such that the size of descendant generation is equal or more than the current generation.*

Heredity yet another precondition for evolution, can in general be observed on two different levels: Syntactic level and Semantic level. On *syntactic level*, entity level inheritance is implied by the structural proximity between parents and their progenies ranging over several generations. For syntactic inheritance to persist, design of the model needs to ensure that environment, which controls the reaction semantics of entities, remains approximately constant over a course of time so that structural similarities also result into continued reproductive behavior. On the other hand, the semantic inheritance is implied in terms of *semantic relatedness* between entities, whereby progenies and their parental entities exhibit similarities in their behaviors (e.g., reproduction) under near identical set of environments. This in turn would require an observer to abstract the behavioral (e.g., reproductive) semantics from the observable reactions among entities in the model, which in turn might require non-trivial inferences in absence of the knowledge of the actual design of the model.

Heredity usually requires further mechanisms to reduce possible undoing of current mutations in future generations owing to new mutations. Therefore, in order to establish inheritance in AES models, sufficiently many generations of reproducing entities need to be observed to determine that the number of parent-child pairs where certain characteristics (both syntactic and semantic) were inherited by child entities without further mutations is significantly larger than those cases where mutations altered the characteristics in the child entities. We can express it as the following axiom:

Axiom 11 (**Heredity**). *Let Ω be a statistically large observed subsequence of a run T, then there exists a characteristic $Char_i$ such that the set of entities in Ω, where this characteristics were inherited without (further) mutation is statistically significant.*

Natural Selection Following the idea from (Bell, 2008, page 19) that on evolutionary scale rate of reproduction is the only attribute selected directly and characteristics affecting the rate of reproduction are selected only indirectly, we consider natural selection as a *statistical inference* on *average reproductive success* of a population of reproducing entities over an evolutionary time scale. Towards that we define following necessary and sufficient axioms as generally discussed in the literature (Stearns and Hoekstra, 2000):

Axiom 12 (**Observation on Evolutionary Time Scale**). *An Observer must observe statistically significant population of different reproducing entities, say $\mathbf{\Lambda_{min}}$, for statistically large number of states in a run $\mathcal{T}$.*

Axiom 13 (**Sorting**). *Entities in $\mathbf{\Lambda_{min}}$ should be different with respect to characteristics in Υ and there should exist differential rate of reproduction among these reproducing entities. Rate of reproduction $ror(e)$ for an entity e is the number of child entities it reproduces before undergoing any mutations beyond observable limit.*

Axiom 14 (**Heritable Variation**). *There must exist variation in the inherited mutations in the population of $\mathbf{\Lambda_{min}}$ implying that a significant fraction of the population of all reproducing entities should have at least one unique characteristics.*

Axiom 15 (**Correlation**). *There must be non zero correlation between heritable variation and differential rate of reproduction.*

Yet another important constraint from the evolutionary perspective is that reproduction in a model should not entirely cease because of the (harmful) mutations. Though this constraint is implicitly captured in the axioms 12 and 13, we can still restate it below primarily since this weaker version may enables us to directly argue for the reasons of the absence of evolutionary behavior in a model:

Axiom 16 (Preservation of Reproduction under Mutations). *Some mutations do preserve reproduction. In other words, if there exist reproductive entities in a state s, either some mutants of these entities or their children should continue reproducing further.*

Software Architecture for an Observation Process

An implementation of the observation process discussed so far essentially demands deciding the level of abstraction on which observations need to be carried out with respect to the underlying AES model. Once it is decided by the designer of the model, either of the following two approaches can be considered for the software design:

Source Code Interleaving/Embedding The specified observational processes can be executed by interleaving the programs for the observations and corresponding interferences within the source code of the AES model simulation design itself. Advantage of such interleaving is that the implemented observation process can reuse some of the computational resources (e.g., memory) of the AES model.

Interactive Observations An observation process could also otherwise be programmed as a separate process itself together with the actual AES model simulation process. These two processes could communicate with each other asynchronously by exchanging the messages containing the required information on the state changes by the model simulation process, which then can be used by the observation process independently for drawing the inferences. This keeps the design of both the processes independent of each other, however unlike the earlier option, the observation process requires to have separate resources for itself. Nonetheless, by virtue of the independence between these two processes, simulation cum observation can be carried out in a distributed environment, which can be useful in case of certain AES studies requiring large amount of computational resources to uncover rare and complex phenomena or detailed dynamics not possible to execute on a single machine owing to main memory limitations or CPU speeds.

Computational Complexity

In the next few (sub)sections, we will estimate upper bounds on the worst case time complexity for the problem of establishing axioms dealing with evolutionary components in the framework for arbitrary AES models. For a discussion on the very choice of worst case computational complexity measure, we request reader to refer to the next Section.

Estimates for space complexity, though equally important, will not be addressed. Primary reason for that is that space (memory) requirement is often dependent upon the actual model at hand, the syntactic nature of the entities as determined by an observation process, and is often linear w.r.t. the total number and size of states observed.

An important problem to be considered while providing estimates on the computational complexity is that observed state progression during simulations might not correspond to the actual underlying reaction semantics for a specific entity. In other words, observed states during simulations progress according to the underlying updating rules for the model, which determine which subset of entities would react in any state. However, in the following analysis, we assume that all those entities, which are enabled to react in each state, are indeed allowed to react. In cases where it is not true, an observation process may store state subsequences of finite size where all (or most of) the enabled entities have been observed to react and then merge all the states in each of these subsequences into single meta states, which reflect the effect that most of those entities which can react have actually reacted.

Computational Complexity of Entity Recognition

Following basic result was proved in Misra (2009):

Theorem 1. *The problem of entity recognition using structural (syntactic) constraints is NP-hard.*

Assuming that all the states in a simulation are of comparable size (i.e., having roughly same number of atomic observable elements), let us use $\mathcal{O}(n)$ as the size of any state. Therefore, if the size of a run $\mathcal{T}$ is r, entity recognition using structural constraints in all the states $s_0, s_1, \ldots, s_r$ may require in the worst case $\mathcal{O}(r2^n)$ steps.

In case, where entities do not have overlapping structures, corresponding upper bound is $\mathcal{O}(rn2^n)$ steps.

Computational Complexity of Observing Evolutionary Components

We can now discuss some of the computational complexity theoretic aspects of observing various components of evolution. Also we will use the following notations:

t_c: expected number of time steps required to determine membership of an entity pair in the relation C.

t_Δ: expected number of time steps required to determine membership of an entity pair in the relation Δ.

$t_{\delta_{mut}}$: expected number of time steps required to determine membership of an entity pair in the relation $\mathbf{R}_{\delta_{\mathbf{mut}}}$.

$t_=$: expected number of time steps required to compare two entities for equality checking.

t_D: expected time steps required to compute function D to check the equality (or inequality) of the characteristics of two entities.

We further assume that checking the negation of a condition takes same number of time steps as checking the condition itself. For example, t_Δ would also be the expected number of time steps required to determine that an entity pair is not in the relation Δ.

Computational Complexity of Observing Entity Level Reproduction Establishing the case for the entity level reproduction in the simplest case, where there are no epigenetic developments in the child entities, minimally demands identifying a single instance of a reproducing entity and its progeny in the next state during one simulation. Suppose an observer needs to determine that an entity p in a state s is an instance of a reproducing entity. For this, the observer needs to establish that under the specified definition of the causal relation C, there exists another entity c in the state $s+1$ such that $(p, c) \in C$ and that the reproductive mutations in c with respect to p are bound by δ_{rep_mut}, i.e., $(p, c) \in \Delta$, and that there does not exist any other entity in the state s, which is recognized as mutating to c. This process would at worst take $N_p^{(s)} = t_c + t_\Delta + |E_s| t_{\delta_{mut}}$ steps where $|E_s| t_{\delta_{mut}}$ factor comes owing to the fact that for each of the $|E_s|$ number of entities in the state s, we need to ascertain that it is not mutating to c. Since for a state s, such a reproducing instance may not be found quickly, in the worst case all the entities in the state s might need to be assessed under these steps. Therefore search for an reproducing instance in a state s may take at worst

$$\begin{aligned}\mathbf{T}_{rp} &= \sum_{p \in E_s} N_p^{(s)} = |E_s| N_p^{(s)} = |E_s|(t_c + t_\Delta + |E_s| t_{\delta_{mut}}) \\ &\leq 2^n(t_c + t_\Delta + 2^n t_{\delta_{mut}}) = \mathcal{O}(2^n \max\{t_c, t_\Delta, t_{\delta_{mut}} 2^n\})\end{aligned}$$

steps, where $|E_s| \leq 2^n$. Since finding such a state s, where a reproducing entity may be present itself may require search into a potentially large state subsequence of a run, it might take $\mathcal{O}(r) * \mathbf{T}_{rp} = \mathcal{O}(r 2^n \max\{t_c, t_\Delta, t_{\delta_{mut}} 2^n\})$ steps to establish the entity level reproduction, where r is the number of states in the state subsequence used in the search assuming that all the states are of comparable sizes. Therefore we have

Proposition 1. *Given the sets of entities in each state, additional time steps required for observing entity level reproduction, without epigenetic development in the child entities and mutational changes in the parent entities, in an AES is upper bounded by $\mathcal{O}(r 2^n \max\{t_c, t_\Delta, t_{\delta_{mut}} 2^n\})$, where r is the number of states observed before first instance of entity level reproduction is recognized.*

The case where entities do not have overlapping structures, total number of entities in a state are restricted by the number of atomic structures, that is, $|E_s| \leq n$. Therefore we have the following corresponding corollary:

Corollary 1.1. *Given the sets of entities in each state, additional time steps required for observing entity level reproduction in an AES where entities do not have overlapping structures is upper bounded by $\mathcal{O}(rn \max\{t_c, t_\Delta, t_{\delta_{mut}} n\})$.*

Next let us consider the general case of entity level reproduction with epigenetic developments in child entities and mutational changes in the parent entities. Towards that we have the following result:

Theorem 2. *Given the sets of entities in each state, additional time steps required for establishing an entity level reproduction is upper bounded by $\mathcal{O}\left(r 2^n \max\left\{t_{\delta_{mut}}, t_c 2^n, t_\Delta 2^n, t_= r^3 2^{3n}\right\}\right)$.*

The case where entities do not have overlapping structures, we have the following corresponding bound: $\mathcal{O}\left(rn \max\left\{t_{\delta_{mut}}, t_c n, t_\Delta n, t_= r^3 n^3\right\}\right)$

Computational Complexity of Observing Fecundity In order to establish fecundity having recognized an entity level reproduction, the first problem for an observation process is to determine the temporal granularities for the generations of the reproducing entities especially when there may exist different types of reproducing entities with different rates of reproduction. In that case, requirement is to determine how many entity types need to be considered. Towards this, the observation process could initially scan a constant number of states to collect all different kinds of reproducing entities together with their rates of reproductions. Based upon the initial estimates on these rates of reproductions, it may consider their least common multiple as the granularity for a generation and ignore other new types of entities while aiming to establish the fecundity axiom. However in case such initial estimates do not yield sufficient support for the fecundity and more reproducing entity types need to be considered, backtrack step is necessary. This process need to continue till statistically significant number of states have been observed to get support for the fecundity axiom or to assume it to be statistically unsatisfiable in that simulation.

Let us first consider the case of single state reproduction without any epigenetic developments. In this case, we have:

Proposition 2. *Given the set of entities in each state, the worst case computational complexity of observing fecundity without epigenetic development is upper bounded by $\mathcal{O}(L 2^{2n} \max\{t_c, t_\Delta, t_{\delta_{mut}}, L/2^{2n}\})$ where L is the number of generations of the reproducing entities.*

Next, we consider the more general case involving epigenetic developments in the child entities:

Theorem 3. *Given the set of entities in each state, the worst case computational complexity of observing fecundity is upper bounded by $\mathcal{O}(L \max\{t_{\delta_{mut}} 2^n,\ t_c 2^{2n},\ t_\Delta r_\pi 2^{2n},\ t_= r_\pi^4 2^{4n}, L\})$, where r_π is the maximum of the lengths of the reproduction cycles of the different types of observed reproducing entities across these generations.*

In a special case of *replication* (with epigenetic development) involving no reproductive mutations in the child entities and no parental mutations would only demand identification using syntactic equivalence between entities and counting the entities belonging to various reproductive types only in last state of each generation. The worst case complexity for such process is upper bounded by $\sum_{1\leq i\leq L}(|E_{i\lambda}| * k * t_=) \leq L * 2^n * 2^n * t_= = \mathcal{O}(Lt_= 2^{2n})$, where $E_{i\lambda}$ is the multiset of entities in the last state of the i^{th} generation and k is the number of different types of reproducing entities in each generation.

Also the case where entities do not have overlapping structures, we have the following corresponding bound:
$\mathcal{O}\left(Ln \max\{t_{\delta_{mut}}, t_c n, t_\Delta r_\pi n, t_= r_\pi^4 n^3, L\}\right)$.

Computational Complexity of Observing Heredity

Theorem 4. *Given the sets of recognized entities in each state, the worst case computational complexity of observing heredity in an AES is upper bounded by*
$\mathcal{O}\left(r2^n \max\left\{t_{\delta_{mut}}, t_c 2^n, t_\Delta 2^n, t_= r^3 2^{3n+1}, |\Upsilon|^2 t_d 2^n\right\}\right)$

The case where entities do not have overlapping structures, we have the following corresponding bound:
$\mathcal{O}\left(rn \max\left\{t_{\delta_{mut}}, t_c n, t_\Delta n, t_= r^3 n^3, |\Upsilon|^2 t_d n\right\}\right)$

Computational Complexity of Observing Natural Selection Given the sets of recognized entities in each state and the relations $\mathbf{R}^+_{\delta_{\mathbf{mut}}}$, $\mathbf{Parent}^{\mathbf{min}}_{\mathbf{\Delta}}$, $\mathbf{\Lambda_{min}}$, and ror from the earlier steps, additional time steps required for establishing axioms for natural selection are upper bounded as follows:

- The Axiom 12 of Observation on Evolutionary Time Scale: $\mathcal{O}\left(t_= r^3 2^{3n}\right)$.
- The Axiom 13 of Sorting: $\mathcal{O}(r2^n \max\{r2^n, |\Upsilon|\})$.
- The Axiom 14 of Heritable Variation: $\mathcal{O}(r2^{2n} \max\{r^3 2^{2n}, t_d|\Upsilon|\})$.
- The Axiom 15 of Correlation: $\mathcal{O}(r|\Upsilon|2^n)$.

Given the upper bounds for these axioms, the following result is immediate for natural selection:

Theorem 5. *Given the sets of recognized entities in each state and the relations* $\mathbf{R}^+_{\delta_{\mathbf{mut}}}$ *and* $\mathbf{Parent}^{\mathbf{min}}_{\mathbf{\Delta}}$*, additional time steps required for establishing natural selection in an AES are upper bounded by*

$$\mathcal{O}\left(r2^{2n} \max\left\{t_= r^2 2^n, t_d|\Upsilon|, r^3 2^{2n}\right\}\right)$$

Given the estimates for the upper bounds on the time steps required for constructing the entity sets E_Ω, $\mathbf{R}^+_{\delta_{\mathbf{mut}}}$, and $\mathbf{Parent}^{\mathbf{min}}_{\mathbf{\Delta}}$, the bound for the overall computational complexity of observing natural selection can be estimated:

Corollary 5.1. *Overall worst case computational complexity of establishing natural selection in an AES is upper bounded by*
$\mathcal{O}\left(r2^n \max\left\{t_{\delta_{mut}}, t_c 2^n, t_\Delta 2^n, t_= r^3 2^{3n+1}, t_d|\Upsilon|2^n\right\}\right)$

The case where entities do not have overlapping structures, we have the following corresponding bound:
$\mathcal{O}\left(rn \max\left\{t_{\delta_{mut}}, t_c n, t_\Delta n, t_= r^3 n^3, t_d|\Upsilon|n\right\}\right)$

Significance of Results

Before we conclude, it is necessary to discuss why to study these worst case computational complexity bounds? In practice, today, most of the AES studies are carried out with significant manual involvement throughout the simulation process and not all the AES studies are carried out to such an extent that their fullest potential is conclusively explored. However as the field would progress, automated exploration of myriad of possibilities which AES simulation studies could have would also become increasingly important. Such automation necessarily present us with fundamental questions on the hardness and limits of such exploration.

One of well studied questions in the domain of algorithm design and analysis is the computational complexity analysis, which gives an insight on the fundamental resource requirements for the problem at hand with respect to the increasing input size. The precise characterization of the inherent resource requirements resulting from such analysis helps an algorithm designer to devise appropriate strategies to optimally utilize the available resources (e.g., CPU cycles) and also to have an estimate of how much could be achieved with available resources.

Among many possible complexity analysis (e.g., average case analysis, amortized analysis etc.) the one which appears most natural and tractable for AES studies is the worst case analysis considered in this paper. The reason is that other than the worst case analysis, other analyses demand either a unifying AES model or a complete characterization of all the AES models. However, currently known and foreseeable AES models differ so fundamentally from each other in terms of their syntactic structures and semantic rules that it is extremely hard to solve either of the problems of defining a unifying AES model or complete characterization of all possible AES models upon which such analyses could be carried out. Also owing to these irreducible design differences, analysis for one AES model could not be generalized in a meaningful manner for other models and thus an inductive approach of building a theoretical framework starting from specific AES case studies may not yield expected answers. Therefore the only fruitful analysis, which appears feasible is the worst case analysis, which could be performed by rather defining a unifying framework for an observation process independent of the underlying AES models.

Further question, which may arise to the reader is how could these results be used in practice? To discuss this, let us informally interpret the presented theorems:

Entity Recognition Theorem 1 could be interpreted as stating that if one has a large and complex simulation for an AES model, it will be computationally expensive to automatically determine the kind of entities, which would emerge over time without externally supplied meta information.

Evolutionary Components On the other hand the remaining theorems state that if entities are structurally distinguishable (i.e., the case of non overlapping structures), once they are identified in a simulation (automatically or otherwise), determining whether evolutionary processes are effective on these entities can be checked in computationally less-expensive manner.

Further, the parameterized form of the results could be used to determine resource bounds for specific AES models having estimates for the required parameters. For example, if in a given AES model entity recognition is feasible in polynomial number of time steps and observed entities do not have overlapping structures, in that case an automatic discovery of natural selection and other evolutionary components could also be carried out using only polynomial number of time steps. On a different note, the specified axioms and proof steps provide practical guidance on implementing the actual observation process, which, once designed could as well be used as reusable component for different AES models with minor changes.

Conclusion

The work on formal characterization of the observational processes can be seen as an attempt to fulfill the need for explicitly separating the design of the AES models from the abstractions used to describe their dynamic progression and the discovery of life-like behavior. We consider evolutionary behavior, as one such characteristic property of life-like phenomena and discuss basic components for observing evolutionary behavior in AES models.

Computational complexity theoretic analysis of the entity recognition as well as establishing evolutionary behavior reveals that an automated discovery of life-like phenomena could be computationally intensive in practice and techniques from the fields of pattern recognition and machine learning in general can be of significant use for such purposes.

The presented work can be further extended by considering other macro level emergent properties including metabolic processes (Bagley et al., 1992), structural and reactive complexity (Adami et al., 2000), self organization (Kauffman, 1993), autonomy and autopoisis (Zeleny, 1981). Associated computational complexity theoretic analysis can be further refined and strengthened by considering classes of models for which most of the parameters have precise bounds compared to the generic analysis presented in this paper.

References

Adami, C., Ofria, C., and Collier, T. C. (2000). Evolution of biological complexity. *Porc. of National Academy of Science*, 97:4463–4468.

Bagley, R. J., Farmer, J. D., and Fontana, W. (1992). Evolution of a metabolism. In *Artificial Life II*, pages 141–158, Redwood City, CA. Addison-Wesley.

Bell, G. (2008). *Selection: The Mechanism of Evolution*. Oxford University Press.

Centler, F., Kaleta, C., di Fenizio, P., and Dittrich, P. (2008). Computing chemical organizations in biological networks. *Bioinformatics*, 24(14):1611.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2001). *Introduction to Algorithms, Second Edition*. MIT Press, Cambridge, MA, USA.

Henz, M. and Misra, J. (2007). Towards a framework for observing artificial life forms. In *Proc. of the 2007 IEEE Symposium on Artificial Life (IEEE-ALife'07)*, pages 23–30. IEEE Computational Intelligence Society.

Kauffman, S. A. (1993). *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press.

Lenski, R. E., Ofria, C., Pennock, R. T., and Adami, C. (2003). The evolutionary origin of complex features. *Nature*, 423:139–144.

Mahner, M. and Bunge, M. (1997). *Foundations of Biophilosophy*. Springer.

Melkikh, A. V. (2008). Dna computing, computation complexity and problem of biological evolution rate. *Acta Biotheoretica*, 56(4):285 – 295.

Misra, J. (2009). Algorithmic feasibility of entity recognition in artificial life. In *Proc. of the 10th European Conference on Artificial Life (ECAL'09))*, volume 5778 of *LNAI*. Springer (In Press).

Ostrowski, E. A., Ofria, C., and Lenski, R. E. (2007). Ecological specialization and adaptive decay in digital organisms. *American Naturalist*, 169:E1–E20.

Papadimitriou, C. (1994). *Computational Complexity*. Addison-Wesley, Reading, MA.

Paun, G., Rozenberg, G., and Salomaa, A. (2006). *DNA Computing: New Computing Paradigms*. Springer-Verlag.

Sayama, H. (1998). *Constructing Evolutionary Systems on a Simple Deterministic Cellular Automata Space*. PhD thesis, Department of Information Science, Graduate School of Science, University of Tokyo.

Singh, D., Ibrahim, A. M., Yohanna, T., and Singh, J. N. (2007). An overview of the applications of multisets. *Novi Sad Journal of Mathematics*, 37(1):73–92.

Stearns, S. C. and Hoekstra, R. F. (2000). *Evolution–An Introduction*. Oxford University Press.

Stone, R., White, D., Guest, R., and Francis, B. (2009). The Virtual Scylla: an exploration of serious games, artificial life and simulation complexity. *Virtual Reality*, 13(1):13–25.

Zeleny, M., editor (1981). *Autopoiesis: A Theory of Living Organization*. North Holland, New York.