Artificial Chemistries

The Quest for Complexity

Martin Henz

Seminar of the School of Computing
National University of Singapore

Overview

- Life
- Complexity
- The Chemistry of Life
- Artificial Chemistries
- Conclusions and Personal Remarks

Attempts to Define Life

- Physiological
- Metabolic
- Biochemical
- Genetic
- Thermodynamic

- Phys: functional: breathing, moving, etc
- metabolic: exchange of materials between org and surrounding
- biochem: capability to store hereditary information in nucleic acid molecules.
- genetic: Equate life with evolution: Any system that achieves evolution is called alive. Defining features: multiplication, variation and heredity. J. Maynard Smith, E. Szathmáry: The Origins of Life, 1999
- thermodynamic: define living systems in terms of their ability to maintain low levels of entropy.

Some Facts of Life

- Fossil records show single-celled life existed 3.8 billion years ago [Schidlowski, 1988]
- all terrestrial life seems to share these early ancestral roots
- carbon-based, aquatic, self-replicating structures
- replication through nucleic acid polymers coding for proteins

Earth formed about 4.6 billion years ago.

M. Schidlowski: A 3,800-million-year isotopic record of life from carbon in sedimentary rocks. nature 333:313-318, 1988.

Liquid water may have appeared only 3,900 million years ago, which makes the time for development of life very short, possibly zero.

Major Transitions

1. Replicating molecules (RNA?) $\Rightarrow$ populations of molecules enclosed in membranes (bilayered fatty acids?)
2. Independent replicators $\Rightarrow$ chromosomes
3. RNA world $\Rightarrow$ DNA and protein
4. Prokaryote $\Rightarrow$ eukaryote
5. Asexual clones $\Rightarrow$ sexual populations
6. Protists $\Rightarrow$ animals, plants, fungi
7. Solitary individuals $\Rightarrow$ colonies
8. Primate societies $\Rightarrow$ human societies, the origin of language, second *memetic* evolution

1. Manfred Eigen, Peter Schuster: The Hypercycle: A Principle of Natural Self-Organization, Springer, Berlin, 1979.
   Evidence for the chemical possibility: Miller-Urey experiment, 1953, Stanley L. Miller, “A Production of Amino Acids Under Possible Primitive Earth Conditions”, Science, Vol 117, No. 3046, pp 528–529, 1953.

2. in today’s chromosomes (sequences of DNA), all genes are linked in a strand (in prokaryotes one strand only). That means that when all genes are replicated, all are replicated. This forces cooperation between genes.

3. RNA can function both as genes and as enzymes.
   RNA world is a way around the “chicken-egg-problem” (genes (DNA) require enzymes (proteins), enzymes require genes)
   idea from Francis Crick 1968 speculation that RNA can catalyze its own self-replication (still unproven)
   Francis Crick: “Foreword” xi-xiv, The RNA World, R.F. Gesteland

and J.F. Atkins, eds. Cold Spring Harbor Laboratory Press, 1993. Later Crick wasn't so sure: F. Crick and L.E.Orgel: Directed Panspermia, p 341–346, 19 Icarus, 1973.
Term "RNA world was introduced by Walter Gilbert, Harvard, in 1986.
DNA and protein interactions are the basic ingredients of prokaryotes (bacteria): reproduction through mitosis

4. disappearance of the rigid cell wall, chromosomes, organelles (mitochondria and chloroplasts)

5. this step gives rise to much more rapid evolution. recombination. only when there is sex, we can talk about species. see chapter 7 in Maynard Smith, Szathmary.

6. embryonic differentiation, gene regulation

7. cooperation/selfishness within societies

8. Susan Blackmore “Meme Machine” meme-gene co-evolution

Components of Evolution

- multiplication, variation, heredity
- information processing (coding, error correction, translation)
- separation (membranes, populations)
- integration (chromosomes, organelles, sex, multicellular organisms, colonies)
- co-evolution (meta, gene-meme)

- novel types of inheritance systems, or development from limited heredity to unlimited heredity
- populations: isolation by geography, habitat, seasonal, mechanical, hybrid inviability, hybrid infertility
- gives rise to cooperation and conflict, selfish gene: Richard Dawkins. Group selection vs individual selection

Overview

- Life
- *Complexity*
- The Chemistry of Life
- Artificial Chemistries
- Conclusions and Personal Remarks

Notions of Complexity

- descriptive
- ontological
- algorithmic complexity (Kolmogorov/Chaitin):
  $K(s) = \mathit{min}\{|p| : s = C_T(p)\}$.
- information theoretic complexity (Shannon):
  using mutual entropy, conditional on "universe"

- several different methods are needed to describe a phenomenon in a reasonably complete way: photon, consciousness (first person view, third person view)

- organized as a system of many non-identical components who themselves have systems-like properties (being further decomposable etc), and whose mutual interactions bring forth a kind of collective behaviour which is different from the behavior of the parts. Von Neumann seemed to have this notion of complexity in mind, when he talked about a complexity threshold beyond which self-replication is possible (he could not characterize in detail).

- minimal length of program for Turing Machine T; problem: random strings have highest complexity!

- information transmission capacity Adami used Shannon complexity to find the information transmission capacity for genomes. $Rl = c$, where R is mutation rate and l is the length of the genome.

Overview

- Life
- Complexity
- *The Chemistry of Life*
- Artificial Chemistries
- Conclusions and Personal Remarks

The Chemistry of Life

- watery solution of carbon molecules
- separation between information and function: DNA and proteines
- DNA codes the assembly of proteines out of 20 amino acids using four different nucleic acids
- double stranded-ness of DNA allows for error correction
- DNA can be transcribed to mRNA, which is transported to ribosomes
- tRNA (anticodon) carrying amino acids dock successively to mRNA and build up protein

- water absolutely necessary: hydrophile/hydrophobe reactions such as protein folding; osmosis, transportation; RNA world: problems with water.

- compare with RNA world; three nucleic acids code for one amino acid: 64 possible, actually: 20; redundancy; roughly: more common amino acids are coded by multiple combinations of 3 nucleic acids (Huffman code)

- error correction allows "unlimited heredity", whereas RNA based heredity is limited to very simple organisms such as RNA viruses (influenza, Polio)

Complexity of Life

- descriptive: classical biology
- ontological: system-theoretic approach in biology
- Kolmogorov/Chaitin complexity: DNA sequences could be taken as length of descriptions
- information theoretic: treat inheritance between generations as information transmission

- morphology, cytology etc

- rich structure on many levels of abstraction

- problems of redundancy and noise; Simplest known organism: mycoplasma genitalium; 480 genes. Human 3000 million bases (30,000 genes) Drosophila (fruit fly) 165 million bases (15,000 to 25,000 genes) Nematode (roundworm) 100 million bases (11,800 to 13,800 genes) Yeast (fungus) 14 million bases (8355 to 8947 genes) E. coli (bacteria) 4.67 million bases (3237 genes) H. influenzae (bacteria) 1.8 million bases M. genitalium (bacteria) 0.58 million bases (400 genes)

- Adami computes information processing capacity during evolution. Idea: view genome inheritance from on generation to the next as an information channel with noise. Information processing is viewed as the learning capacity of the genome, under very rigid assumptions can be quantified. See Christoph Adami: Artificial Life, 1998.

Overview

- Life
- Complexity
- The Chemistry of Life
- *Artificial Chemistries*
- Conclusions and Personal Remarks

Artificial Chemistries: Motivation

- Theory of evolution has the problem of having only one instance available for study
- use artificial chemistries to enable evolutionary processes
- "life as it could be"
- Goals: scientific study of evolution, optimization, artificial intelligence

Karl Popper suggests that evolution theory is not scientific but metaphysical, lack of repeatability

Artificial Chemistries: Overview

- Wet artificial life
- Rewriting systems
- Cellular automata
- Self-replicative codes

Wet Artificial Life

- Goal: reproducing the RNA world: (sets of) RNA strings that catalyze their own replication.
- Quest for missing link in origin of life. Milestones:
  - Bartel, Szostak: Evolution using "tag" molecules, polymerase chain reaction, reverse transcription,
  - Wright, Joyce: Use tag for targeted PCR, but complex reactions involving double-stranded DNA

M.C. Wright, G.F. Joyce. Continuous in vitro evolution of catalytic function. Science 276, 614. 1997. Emphasis: no interaction of experimenter (except occasional dillution in a fresh reaction solution); other approaches: core-and-shell self-reproduction P.L. Luisi, P. Walde, T. Oberholzer. Enzymatic RNA syntehsis in self-reproducing vesicles: An approach to the construction of a minimal cell. Ber. Bunsenges. Phys. Chem. 98, 1160. 1994.

Rewriting Systems: Principle

- Define set of molecules S.
- Define set of reaction rules R of the form

  $$r = s_1 + \cdots s_n \ \rightarrow \ s'_1 + \cdots s'_m$$

  between molecules $s_i \in S$.
- Define reactor algorithm, for example stochastic molecular collisions

Rewrite Systems: Examples

- Chemical rewriting system on multisets (ARMS), developed by Suzuki, Tanaka, similar to the Chemical Abstract Machine (Berry, Boudol), additional rule order
- Chemical casting model (CCM), developed by Kanada and Hirokawa, goal-driven, used for optimization
- Lambda-calculus (AlChemy), used by Fontana and Buss, β reduction as single reaction rule

- cooling/heating rules, ion rule inspired by CCS, able to model oscillating chemical systems
- chemical casting model: problems con be encoded using links (traveling salesman etc), quality defined by local order degree used for optimization
- self-replication easy, different levels of organization appear with different variants of the experimental setup

Cellular Automata: Background

- introduced by John von Neumann to study self-replicative automata
- used to study both
  - self-replication (nanotechnology, space exploration), and
  - reproduction (evolution)

Cellular Automata: Definition

A cellular automaton is a tuple (L, S, N, f), where

- L is a regular lattice of cells with dimensionality d,
- S is a finite set of states,
- N is a finite set (of size $|N| = n$) of neighborhood indices such that for all $i \in N$ and $c \in L$ we have $c + i \in L$,
- $f : S^n \rightarrow S$ is a transition function.

- point lattice: discrete subgroup of Euclidean space, other structures allowed (hexagonal)

- dimensionality d

- i of course d-dimensional vector

Classes of Automata

- usually neighborhood is defined using “radius” r
- rule tables are large. Example: $|S| = 8$, $d = 1$, $r = 2$ (5 neighbors). Then there are $8^5 = 32768$ neighborhood states and 8^{32768} different transition functions.
- problem: find the right chemistry
- spacial isotropy
- quiescent state yields quiescent state

- spacial isotropy: all planar rotations of a neighborhood should map to the same state.
  von Neumann constructed 2-dimensional CA with 29 states and the vN neighborhood, which is computationally universal, but asked for construction universality (unfinished work, completed by Pesavento, 1995, using 32 states, and others). A. Ray Smith showed that for self-replication, only computational universality suffices.

Behavior Classification

- Wolfram investigated 1d automata with $|S| = 2$, $r = 1, 2$
- Wolfram class I: limit point behavior,
- Wolfram class II: limit cycle behavior,
- Wolfram class III: “uniformly chaotic” behavior,
- Wolfram class IV: none of the above

Karel Culick and Shen Yu showed that Wolfram's classification is undecidable; Class IV most interesting; complex patterns; self-organization; conjectured to be capable of universal computation. Conway's game of life is of Class IV. look at `http://www.rendell.uk.co/gol/tm.htm` for a turing machine built in game of life.

Langton's Experiments

- How to find "interesting" rule sets?
- Langton analyses CA for $d = 1$, $|S| = 2$, $r = 2$
- probability of getting a non quiescent state is significant
- self-replicative structures for $d = 2$

class I, II, III with IV between II and III; look at urlhttp://cell-auto.com/links for applets

construction universality: be able to construct any other automaton

not necessary for self-replication

Complexity in Cellular Automata

- Defined mostly negatively: absence of unstructured chaos, periodicity
- emphasis on self-replication,
  quest for simplest self-replicating structure

Self-replicative Code

- computer viruses
- Coreworld
- tierra
- avida
- amoeba

Coreworld

- Developed by Steen Rasmussen out of “Core War”, a game where programmers compete for the memory of a virtual machine.
- Uses circular memory, Core War instruction set, called Redcode
- Introduce noise through a stochastic copy instruction.
- Self-replicating programs can be easily written, but are not stable.
- Different “ecologies” emerge from different parameter settings.
- Coreworld generally fragile under mutation, consequence of instruction set

tierra

- developed by Tom Ray
- 32 instructions, vaguely based on Intel i860
- pattern based addressing, using `nop0` and `nop1`.
- scheduling using slicer queue
- memory management using reaper queue

Behavior of tierra

- seeded with self-replicating code (80 instructions)
- different "species" emerge:
  - parasites
  - immune hosts
  - symbionts
  - "cheaters"
  - "super-parasites"

avida

- developed by Christoph Adami
- based on tierra, but two-dimensional structure
- each cell contains circular program
- refined rules for code sharing and migration
  - replication by copying code into neighbor
  - facing

Other Chemistries for Self-replicative Code

- amoeba
  - target: evolution of self-replication; smallest self-replicators are only 5 instruction long
- sanda
  - similar to avida, but allows targeted evolution by regulating the speed of execution in a cell
  - parallel implementation allows large-scale experiments

Complexity in Self-Replicative Code

- Tom Ray uses mostly descriptive complexity, no attempt for quantitative analysis
- Most developed quantitative analysis done by Christoph Adami on avida
- uses Shannon complexity, based on average string length in population, frequency of genotypes, volatility of instructions

Overview

- Life
- Complexity
- The Chemistry of Life
- Artificial Chemistries
- *Conclusions and Personal Remarks*

Shortcomings of Life

- protein synthesis is slow
- higher life forms are stuck on earth
- communication bandwidth narrow
- duration of generation long (decades for humans)
- brain is slow
  - 5ms between firing of connected cortical neurons
  - $1 m/s$ speed of signal transmission within neuron
  - neurons are large (4 microns)

60ms for adding a single amino acid; average human protein has 450 aas, so 27 seconds for assembling one single protein! No wonder, cells are so complicated (parallelism).

no wonder 10^{11} neurons with 10^4 fanout are needed for a real-time intelligent brain!

feature size in modern VLSI: 130 nano meter: 30 times smaller!

We Should Do Better

- understand and exploit evolution
- understand and exploit meta-evolution
- exploit feature size and *speed*
- Hard Artificial Life: third evolution

look at the major factors in natural life: information processing, separation, integration

many of the "impressive facts" about carbon-based life are artifacts of the limitations apparently inherent in carbon-based aquatic chemistry. Let us study "life as it could be"