Artificial Life Lecture 17

Information & Life & Evolution
The rough picture:

Part 2: GasNets

Bits (**B**inary dig**ITS** in the computer/information theory sense) are a measure of how much information there is in a program, in a text file, on a floppy disk
--Megabytes, Gigabytes etc.

Genotypes -- real ones of DNA with an alphabet of size 4 CGAT, or artificial ones with a binary alphabet of 0s and 1s -- can also have their information measured in bits.

Vague ideas about genetic information

SO...(*the vague and imprecise argument goes*)...
genotypes can and probably do contain information about something -- I wonder what it is – maybe information about how to build the animal/organism, or maybe information gathered from the environment over many generations, or ....

Can we make any of this vagueness precise ???

Some references

Shannon CE and Weaver W "A Mathematical Theory of Communication" Univ of Illinois Press 1949
The classic intro to Shannon's information theory
-- though be cautious of Weaver.

Susan Oyama "The Ontogeny of Information"
Cambridge University Press 1985
A very thorough critique of naive ideas of DNA as information about the phenotype

More references

RP Worden "A Speed Limit for Evolution"
Journal of Theoretical Biology, 176, pp 137-152, 1995
http://dspace.dial.pipex.com/jcollie/sle/index.htm
Important ideas though be cautious.

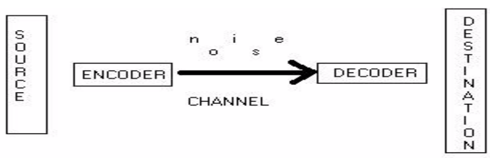
Chris Adami "Introduction to Artificial Life"
Springer-Verlag (TELOS) 1998
Despite general title, rather specialised within Alife field -- **avida** *(same family as* **Tierra***), information theory, fitness landscapes and error thresholds.*

Information

As with most such words, used in lots of different ways,
eg Gregory Bateson
(*'Information is a difference that makes a difference'*)
eg JJ Gibson and ecological perception
(*information in the environment*)
--- but mostly, when people talk of **BITS** of information, they are thinking of **Shannon's** theory.

--- which might be better thought of as a theory of **communication** rather than just of **information.**
--- and is vital to telecomms transmission, CD players etc

Communication



This kind of information *only makes sense* in the context of being sent over a **CHANNEL** from a sender to a receiver

Reducing Uncertainty

Information in bits can best be thought of as a measure of the reduction in uncertainty in the mind of the receiver !

If I am the receiver, and I am expecting from you a single binary digit 0 or 1, and I have no reason to think one is more likely than the other, my uncertainty can be measured as 1 bit *(of Shannon entropy or uncertainty).*

When I receive a 0 (or a 1) my uncertainty is reduced by this 1 bit.

Context-dependence

Likewise if I know the name you are sending me is either John or Jane, equally likely, my uncertainty is still just 1 bit -- and transmitting the string 'John' is worth *just 1 bit of information*..

If the name you are sending me is known in advance to be John or Jane or Anna or Paul (and equally likely) -- then transmitting the string 'john' is worth *2 bits*

--ie it is very much **context-dependent**, on the initially expected (by the receiver) distribution.

Measuring Shannon Uncertainty

Uncertainty gets bigger, the more possibilities there are.

And we would like uncertainty to be additive (when considering 2 uncertainties about 2 unrelated systems)

If the set of possible signals you are expecting to receive has N different possibilities, and the probability of the i^{th} one is p_i, then Shannon entropy (uncertainty)

$$H = -\sum_{i=1}^{N} p_i \log_2(p_i)$$

E.g.

So if a priori probabilities of names were
John $p_1 = 0.5$ Jane $p_2 = 0.25$
Anna $p_3 = 0.25$ Kzwj $p_4 = 0.0$
... all other possibilities 0.0

$H = -\ 0.5 \log_2(0.5) - 0.25 \log_2(0.25) - 0.25 \log_2(0.25)$
$= 0.5 \times 1 \quad + 0.25 \times 2 \quad + 0.25 \times 2$
$= 1.5$ bits

..that's a bit less uncertainty than the 2 bits for 4 equally likely names

Correlation

Same example
John 0.5
Jane 0.25
Anna 0.25

But if we are told the name is male it must be John
and if we are told female it is now 50/50 Jane/Anna

So there is a correlation between sex and names, and in fact here knowing the sex gives us 1 bit of information – reduces our uncertainty by 1 bit.

Correlation entropy

Technically, one definition of information is the *correlation entropy* between two (sets of) random variables.

The *correlation entropy*, or *mutual information*, is zero if there is no correlation between the sets,
-- but becomes positive (measured in bits) where knowledge of one variable (eg sex) reduces uncertainty about the other one (eg name)

<<It's dangerous to think of information as a commodity>>

Information in genotypes

So it is very easy to say that natural genotypes are strings of characters CGAT, just like artificial genotypes are often strings of binary characters 01, and hence 'worth' respectively 2 bits or 1 bit per character.

This particular way of counting bits only buys you anything if you have no a priori expectation of what the next symbol may be -- typically only true for random or junk DNA !

But maybe you can be careful and set the context up carefully so that it makes Shannon sense ??

Does this make sense?

…but maybe you can be careful and set the context up carefully so that it makes Shannon sense ??

ONE COMMON SUCH ATTEMPT:

"The genotype (many poss variations) is the sender"
"The developmental process is the channel"
"The phenotype (many poss variations) is the receiver"
"culture, environment etc, makes a difference but is from one perspective just noise"

Nature vs Nurture

There is some limited sense in which there is a bit of truth in this -- if I know that some DNA is either human or chimpanzee DNA, then roughly speaking I know that (in the appropriate womb) it will develop into either a human or chimpanzee genotype -- that is worth just 1 bit of information to know which is which !

But see Oyama ref. for considerable discussion, particulary on nature vs nurture.

Does DNA give information ?

But many would like to say that human genotype of 3x10^9 characters of alphabet size 4 contains up to 6x10^9 bits (or maybe 6x10^8 bits if 90% is junk ??) of information describing the phenotype.

I have never yet seen any rigorous way of setting up the Shannon information theory requirements -- sender, channel, receiver, **decrease in uncertainty** -- for this to make sense, and one can doubt whether it is possible.

But Robert Worden makes an attempt (see refs)

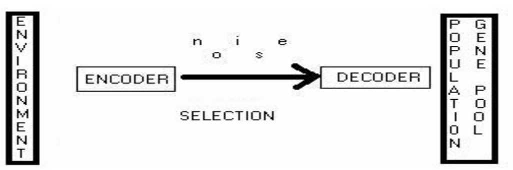
A Speed Limit for Evolution ?

Worden talks of **GIP** (*Genetic Information in the Phenotype*) as a property of a population, measured in bits.

"A population with some trait narrowly clustered around a central value *[inherently improbable in the absence of selection*] has more GIP than another population in which the same trait is more widely spread *[inherently more random*]"

"Halving the width of spread of some trait increases the GIP by 1 bit" --- How does the population increase its GIP store of bits -- why, through **selection!**

From Environment to the Genepool



This is how Worden sets up the channel of information, so as to talk about Shannon information (or uncertainty).

The Speed Limit

IFyou have selection as channel for 'transmitting information', then it will depend on the **selection pressure** --

"An intuitive rationale...if an animal has $2^3 = 8$ offspring and only one survives to maturity, then this event conveys [maximum] 3 bits of information from the environment into the surviving genotype and phenotype" -- Worden

If you have a large popn, and you divide on the basis of fitness into 8 octiles ('eighths') -- then selecting the top eighth can [in the right context] convey 3 bits of information.

Caution

All this is to be taken with a *lot of salt*. BUT Worden's argument attempts to justify such a limit
-- $\log_2$(selection pressure) bits of information down this 'channel' per generation, maximum, this speed limit holding with infinite or finite popns, mutation, recombination etc etc.

Typically $\log_2(S)$ is at most a few bits per generation

This has implications for Artificial Evolution

Worden's deductions …

The first part of Worden's paper is mathematical, justifying the speed limit in different contexts.

He then draws some *[rather speculative...]* conclusions on such matters as 'Design Information in the Brain'

"Common ancestor of humans and chimps was around 5,000,000 yrs ago, say 350,000 generations, at say $\log_2(1.5)$ bits per generation gives something of order of 100,000 bits of GIP…"

…deductions (ctd)

...with some other assumptions (difference in probability to survive and reproduce, dependent on intelligence/brain development, not more than +- 10%) ... useful genetic info in the brain, compared to our common ancestor with chimps, increased by max 40,000 bits or 5Kbytes -- not very much -- cf computer programs" – says Worden.

Implication -- not all that much design difference!

HEALTH WARNING: this is all rather dodgy, see earlier warnings about information as a commodity !

Avida

A different approach. Chris Adami, looking at Avida, artificial organisms of variable genotype length, and variable complexity, evolved in a virtual soup in a computer.

In "Introduction to Artificial Life" he looks similarly at *"the stochastic transfer of information from an environment into the genome"*

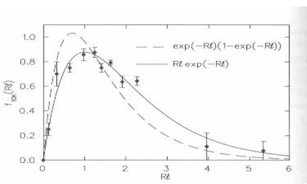
He investigates the maximum rate at which the population [genepool] *absorbs* information from the environment.

Error threshold

The maximum rate, experimentally, is somewhere near the error threshold, mutation around 1 per genotype.

A tradeoff between informational storage and informational transmission -- too much mutation, stored 'info in the genotype' gets lost --

too little mutation, too little variation in the population for selection to use for transmission of info environment->genepool.



'Around 1 bit per generation'

Once again, comparably to Worden, from inspection of fitness graphs it looks like an upper limit well below 1 bit per generation, in terms of increase in effective genotype length.

BUT bear in mind earlier warning about being rigorous with Shannon information -- it is still not clear how rigorously Adami's model stands up to this kind of description.

Unsatisfactory conclusion

Information theory and evolution is still a wide open area for **rigorous** research -- there is all too much of the non-rigorous kind.

It looks like Worden is onto something with his Speed limit, but not rigorously defined yet.

This may well be easier to do with artificial evolution, as indeed Adami makes a start on this.

Engineering spin

My hope: for binary genotypes of length 1000, with a search space of size 2^{1000}, suppose merely 2^{600} genotypes corresponded to successful designs, then finding a successful design equates to 400 bits.

Is something like Worden's speed limit relevant, eg 1 bit per generation (I believe so)? This would be very much an upper speed limit, like the speed of light. Here, minimum 400 gens.

How can we arrange our artificial evolution so as to get as close as reasonable to this speed limit? An important, and very open, research question.

Practical Relevance

You have to expect to run your GA on a serious problem for a seriously long time – inexperienced people frequently underestimate!

Eg the GA+CTRNN exercise, the posted example code runs 100 different trials for *every* individual, and runs for 10,000*20 individuals.

A total of 20,000,000 trials, each one running for 50/0.1 = 500 update-steps. Total 10 billion update-steps.

This is *normal* !!!!!! This is what computers are for !!!!! Optimise your code.

A brief taste of GasNets

Context: relatively recently discovered that the workings of the brain depend *not only* on "electrical signals down wires between neurons" *but also* on "gases produced in varying amounts, diffusing locally through the 3-D volume of the brain" – and modulating the behaviour of those neurons that the gas impinges on.

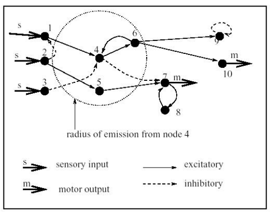
Nitric Oxide being just one example.

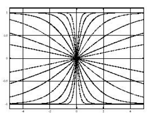
This adds a new angle onto the mechanisms of the brain – and inspired a variant ANN -- GasNets

The quick picture

E.g.: Husbands, P., Smith, T.M.C., Jakobi, N. and O'Shea, M. Better Living Through Chemistry: Evolving GasNets for Robot Control. Connection Science, 10(3-4):185-210.

Neurons are modelled as having a position in 2-D, and potentially emitting/reacting to diffusing gas



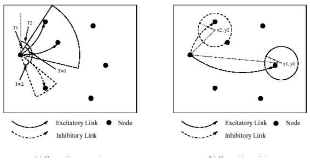


"Reaction" = modulating the sigmoid transfer function at one neuron

Evolving GasNets

Genotypes specified the NN characteristics through a sort-of developmental process: Gen specified for each neuron

1. Where in a 2-D plane it was located
2. Which 'directions' it can link to nearby neurons (2 +ve, 2 –ve links)
3. Under what circumstances (and how far) it may emit a 'gas'
4. …and various other more typical characteristics



Temporal aspect of the Gases

A node will start emitting gas when its activation exceeds a (genetically specified) threshold, and continue doing so until it falls below that threshold.

This 'generation of gas' builds up the local concentration over time – and when production stops, the gas decays away over time. The gas modulates the behaviour of nearby nodes

Roughly speaking: in these models, this appears to be the *only* aspect where there are genuinely time-varying aspects to the ANN

Criticism: Is there 'time' anywhere else?

As far as I can see: No!!

There are recurrent connections, but there is no sense of 'how long' it takes for an 'electrical' signal to pass from one node to another; except purely as an accidental artefact of whatever time-update-step is used.

These are not CTRNNs, there is no proper treatment of them as dynamical systems changing over real-time – except solely for one aspect of the 'gas diffusion' (as per previous slide).

The comparisons with non-GasNets

Several papers published comparing performance, evolving for robot control systems on a task where time-issues are crucial:-- GasNets versus non-GasNets

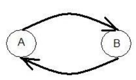
But these studies are, in my opinion, crucially flawed. The non-GasNets are just the same GasNets with 'the gas turned off'. And the gas was the only part where real-time was handled in a reasonably principled fashion.

Regrettably, they made the same mistake that SO MANY people make – misunderstanding time in ANNs

How to think of time in Dynamical Systems

If you are modelling control systems as a DS, then *think of the physics* first.

All nodes in a CTRNN are varying and affecting each other *continuously in real time*



If A is influencing B, and also B is influencing A, then do NOT think of this as a ~~cycle~~ where first A changes, then B, then A, then B.

They are both changing continuously and simultaneously

The computational cycle

The only such cycle is in your computational approximation – and this needs to be understood properly.

$$\frac{\delta A}{\delta T} = f(A,B) \qquad \frac{\delta B}{\delta T} = g(A,B)$$

In the computation, you set δT to something as close to zero as reasonable (rule of thumb: 1/10 of the shortest timescale of significance in what you are modelling). Then there are *cycles* in your simulation – but *not* in what is being modelled.

Changing your δT to a different reasonable value (eg 1/10 smaller again) should *never* materially change your results

www.informatics.susx.ac.uk/users/inmanh/easy/alife10/TimeSteps.html