

Bayesian Network Modeling for Evolutionary Genetic Structures

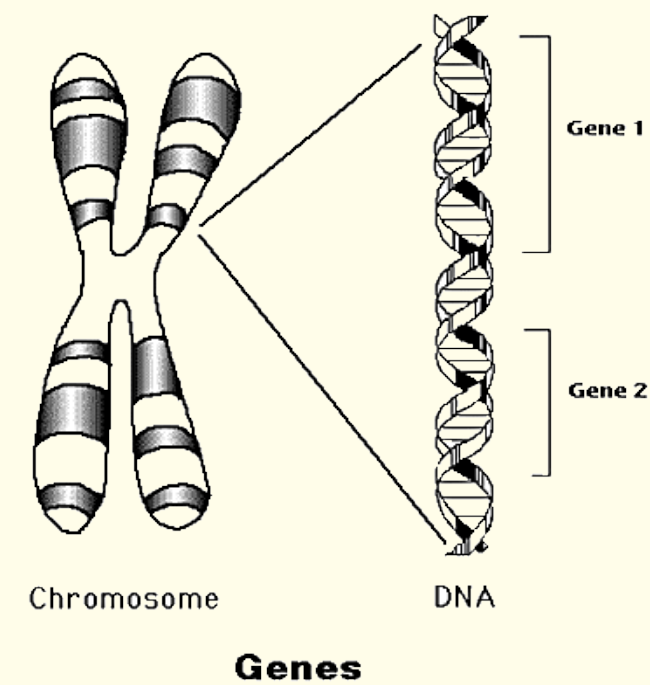
Lisa Jing Yan, Nick Cercone

lisayan@cse.yorku.ca, ncercone@yorku.ca

York University, Canada

Survival of the Fittest!

Evolutionary theory reveals those stronger genetic characteristics, but the process takes a number of generations to assess.



Questions:

1. Which gene makes us more adaptable?
2. Can we see the inter-connections among these genes?
3. Can we apply this information into other areas and have insights?

Contributions

In an artificial life environment, two species compete to survive over n-generations.

- a. First, we provide a useful and unique set of meaningful data, based on ALGAE (Artificial Life & Genetic Algorithm Expression), but also simplify the complexity for later analysis;
- b. Then, we learn a graphical model to describe various hidden connections among "genes" derived from simulated genotype data to reveal what the critical factors/inter-connections are.

Analysis

BANANA provides a graphical connection of these 12 factors with 26 arcs, showing that: *a)* Defense ability (DA) is the major factor in survival (with 8 arcs constraints); *b)* Energy lost in fighting (LA) is the second important factor (with seven arcs constraints); *c)* The relationships and dependencies also indicate that speedy attack ability (SA), and the energy cost of survival (EF) (with six arcs each). We therefore see that combat occupies a central role and different level of importance of each gene in survival, in a hostile environment with competition for survival. The BN reveals this hidden rule of survival embedded in ALGAE. That is, only certain gene combinations will allow a species to survive. **Defense comes first, and attack skills or energy status affects the 'battle period'.**

A successful individual's gene composition does not explain the reason for its success. The data merely reveals the principle; however, BN describes the causal relations among the factors and how these connections influence the way the whole diagram works. It shows the reality of why this species could continue to live and thrive. For ALGAE, it is: who adapts and stays to the last, survives!

ALGAE: Artificial Life Genetic Algorithm Expression



We build a simple ecology system, ALGAE, which includes environmental factors and individual genetic factors, to determine the survival ability in the evolutionary process.

Fig. 32-bit chromosome descriptors

Gene	Description	Bit Site	Gene	Description	Bit Site
SP	SPecies type	0	CA	Action Characteristics	13-15
SL	Life Span	1-4	CR	Capricious Rate	16-18
VF	Vision Field	5-6	SA	Attack Speed	19-21
TM	Transition Movement	7-8	DA	Defend Ability	22-24
CM	Motion Characteristic	9-11	LA	Attack Loss	25-27
LM	Motion Loss	12	EF	Food Efficiency	28-31

Two species compete for resources to survive, and each individual genotype is represented as a 32-bit chromosome which is a combination of the above factors:

G: [SP, SL, VF, TM, CM, LM, CA, CR, SA, DA, LA, EF].

Algorithm 1: ALGAE

1. Initial AWord environment, $n \times m$, randomly set up barriers and food supplies; Initialize two populations;
2. Evolution starts;
Until one population is extinct or goes over the *max* generation number.

The program iterates to mimic generational evolution over lengthy time frames. Species members experience genetic variations throughout the process, and the survivors remain to reveal which specific genes have adapted. With the recording the whole evolutionary process, next, we examine how these remaining genes correlate to ensure successful survival.

BANANA: Bayesian Network ANALysis

- Data Processing

$$G_i: \text{SP SL VF TM CM LM CA CR SA DA LA EF}$$

- Learning algorithm: E-algorithm
E-algorithm is a topology learning algorithm, which is a combination of constraint-based and MDL score-based methodology.

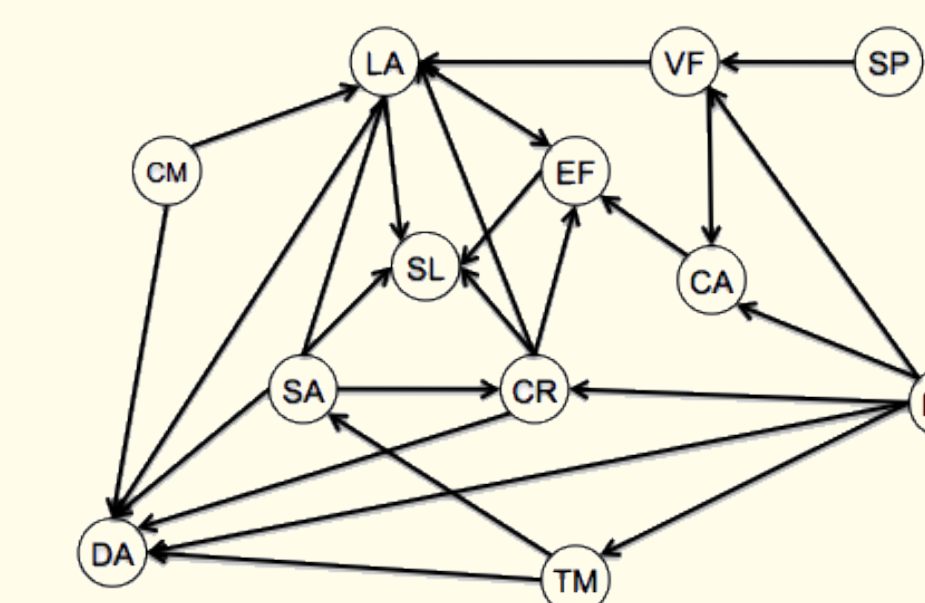
Algorithm 2: E-algorithm

1. Construct fully connected graph G_i ;
2. Conduct order-0 conditional independence (CI) test to eliminate low mutual information arcs, according to given threshold; Obtain G_0 ;
3. Conduct order-1, order-2 CI test to further remove unnecessary arcs according to a t -value generated by χ^2 test in a given degree of confidence σ . Obtain G_1, G_2 ;
4. **repeat** Ascertain each node X_i 's candidate parents Π_i :

- a. sort its potential parent nodes ascending based on their mutual information;
- b. use *MDL* to score each candidate structure, and perform Branch&Bound to decrease the search space of all possible combination of nodes structures;

until min *MDL* score structure is found.;

Results



SP: SPecies type(1); CA: Action Characteristics(3); SL: Life Span(4); CR: Capricious Rate(6); VF: Vision Field(4); SA: Attack Speed(5); TM: Transition Movement(3); DA: Defend Ability(6); CM: Motion Characteristic(2); LA: Attack Loss(7); LM: Motion Loss(5); EF: Food Efficiency(4);

References

- [1] J. Ji, C. Liu, J. Yan, N. Zhong. Bayesian Networks Structure Learning and Its Application to Personalized Recommendation in a B2C Portal. In *WI 2004*.
- [2] D. Heckerman, D. Geiger, D.M. Chickering. Learning Bayesian Networks: The Combination of Knowledge and Statistical Data. *Machine Learning*, 20(3): 197-243, September 1995.

Acknowledgements

We would like to thank the anonymous reviewers for helpful comments and suggestions. This research was supported by NSERC grant CRD 349345-06.