

Bayesian Networks Learning for Strategies in Artificial Life

Lisa Jing Yan
lisayan@cse.yorku.ca

Nick Cercone
nick@cse.yorku.ca

Computer Science and Engineering Department
York University
Toronto, ON., Canada, M3J 1P3

ABSTRACT

Evolutionary algorithms have been used to effectively generate solutions to artificial life problems. However, this process may take a number of generations to complete. Research to accelerate evolutionary search has been reported, yet, insights into this evolving process have not been analyzed nor why certain characteristics are more dominant than others. This paper provides a systematic and causal explanation for these findings and why certain genes are superior. We use Bayesian Networks (BNs) to learn a graphical model to represent the learning process in the Artificial Life environment. Bayesian Network ANALysis (BANANA) is then developed, which gives visual representation of the inter-connections among these characteristics and provides information for further insight into genetic fitness.

Categories and Subject Descriptors

I.2.m [Artificial Intelligence]: Miscellaneous; G.3 [Probability and statistics]: Experimental design; G.4 [Mathematical software]: Algorithm design and analysis

General Terms

Experimentation, Algorithms, Design

Keywords

Evolutionary Algorithms, Bayesian Networks, Learning

1. INTRODUCTION

Traditionally, evolutionary algorithms have been used to effectively generate solutions to certain problems; evolutionary theory reveals strong genetic characteristics. However, this process takes a number of generations to complete; being trapped in local optima is a typical shortcoming in evolutionary algorithms. To exclusively research the entire solution space can avoid local optima, but this is time consuming. Research to accelerate evolutionary search has been reported with promising results[1]. However, this evolving process has not been analyzed nor why certain characteristics are more dominant than others.

This paper provides a systematic and causal explanation why these are solutions found and why certain genes are superior. We use Bayesian Networks (BNs) to learn a graphical

model to represent the learning process arising in the Artificial Life environment. This gives a visual representation of the inter-connections among these characteristics, which provides information learned for further insight into genetic fitness.

Bayesian networks [2] are graphical models which can reveal the hidden relations among various features and intuitively provide the causal connections in numerical datasets. We use Bayesian networks to analyze the obtained information and learn the causal/hidden relations among features. First, we simulate an artificial life environment, where two species compete to survive over n -generations. This simulation provides a useful and unique set of meaningful data, based on ALGAE (Artificial Life Genetic Algorithm Expression), and also simplifies the complexity for later analysis. Moreover, a graphical model can be learned to describe various hidden connections among “genes” derived from simulated genotype data to reveal what the critical factors/inter-connections are.

In ALGAE, the process of two tribes of agents exploring the environment and competing for resources available is taken as a random choice of action according to each individual’s characteristics. This randomness does not contain any intelligence; rather the superior genetic characteristics are carried over and survive in the descendants. Our research explore using the factor analysis results of Bayesian Networks learning for each individual agent, in order to learn rational strategies to obtain optimal gain in a competitive multiagent environment.

2. BACKGROUND

To explore the hidden dependencies among the variables, Bayesian Networks are used to analyze genotype data derived from simulated evolutionary processes and provide a graphical model to describe connections among genes. There are a number of models available for data analysis such as artificial neural networks, decision trees, factor analysis, BNs, and so on. Yet BNs have distinct advantages as analytical methods which can discern hidden relationships among variables. Two main approaches, constraint-based and score-based, have been used to learn BN structure. However, both suit either sparse structures or dense structures. Firstly, a hybrid algorithm, called “the E-algorithm” [4, 3], is introduced to complement the benefits and limitations in both approaches for BN structure learning. Testing the E-algorithm against a standardized benchmark dataset ALARM, suggests valid and accurate results. Bayesian Network ANALysis (BANANA) is then developed which incorporates the

E-algorithm to analyze the genetic data from ALGAE. The resulting BN topological structure with conditional probability distributions reveals the principles of how survivors adapt during evolution, producing an optimal genetic profile for evolutionary fitness [5].

3. ALGAE

In order to answer these questions, we simulate an artificial life environment as a simple ecology system, ALGAE. ALGAE describes two species which compete for resources to survive, where each individual is represented as an artificial 32-bit chromosome. Each chromosome is a combination of individual genetic characteristics, which determine the survival ability in the environment. (see Table 1 for details)

Table 1: 32-bit chromosome descriptors

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

4. BANANA

The simulation ALGAE records the generational evolution over lengthy time frames. With the record of this evolutionary process, we next examine how these remaining species genetic descriptors correlate to ensure successful survival (see Figure 1).

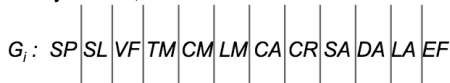


Figure 1: 32-bit chromosome

In Figure 2, we use BANANA to learn a graphical representation of the relations among the gene descriptors. The E-algorithm within BANANA is a hybrid BN structure learning algorithm which combines constraint-based and MDL score-based methods. This graphical representation reveals these 12 descriptors with 26 connections among them, showing that: a) Defense ability (DA) is the major factor in survival (with 8 arcs/constraints); b) Energy lost in fighting (LA) is the secondary 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) are key constraints/factors.

We therefore see that combat occupies a central role, and there are different levels of importance of each gene, 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, while attack skills or energy status affect 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

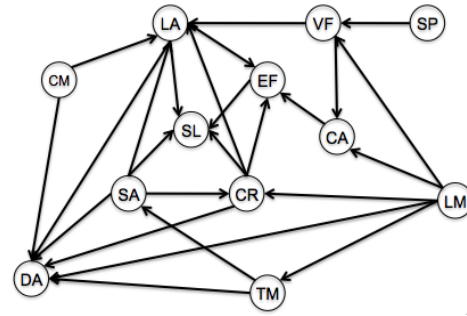


Figure 2: Bayesian Networks Representation for ALGAE

the factors and how these connections influence the way the whole diagram works. The BN shows the reality of why these species could continue to live and thrive.

5. CONCLUSION AND FUTURE WORK

This causal relationships are revealed by Bayesian analysis to individuals survival ability in ALGAE. Applying this knowledge to individuals can accelerate the solution search process, while increasing individual survivability. "Re-attack with choice" refers to the situation where each individual performs in the environment to compete for survival with learned knowledge from their ancestors. This knowledge describes the critical factors and how these factors influence the survival outcomes. Initial research on Bayesian networks gives promise to the prediction ability in an uncertain environment. Applying these constraints which exist among various factors in the multiagent environment, this will give an analytical choice based on known effects, rather than mere random exploration.

6. ACKNOWLEDGMENTS

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

7. REFERENCES

- [1] J. C. Bongard and G. S. Hornby. Guarding against premature convergence while accelerating evolutionary search. In *GECCO*, pages 111–118, 2010.
- [2] D. Heckerman, D. Geiger, and D. M. Chickering. Learning bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, 20(3):197–243, September 1995.
- [3] J. Ji, C. Liu, J. Yan, and N. Zhong. Bayesian networks structure learning and its application to personalized recommendation in a b2c portal. In *WI '04*, pages 179–184, Washington, DC, USA, 2004. IEEE Computer Society.
- [4] J. Yan. Bayesian network structure learning. Thesis, Beijing University of Technology, July 2003.
- [5] L. J. Yan and N. Cercone. Bayesian network modeling for evolutionary genetic structures. *Comput. Math. Appl.*, 59:2541–2551, April 2010.