Deriving Evaluation Metrics for Applicability of Genetic Algorithms to Optimization Problems

Hsin-yi Jiang, Carl K. Chang
Department of Computer Science, Iowa State University
hsinyij@cs.iastate.edu, chang@cs.iastate.edu

ABSTRACT

This paper aims to identify the missing links from theory of Genetic Algorithms (GAs) to application of GAs.

Categories and Subject Descriptors: J.0 [Computer Applications]: General

General Terms: Algorithms, Measurement, Performance.

Keywords: Applicability, Evaluation Metric.

1. INTRODUCTION

Genetic Algorithms (GAs) are widely applied to searching optimal solutions in many problem domains. For instance, GAs are used to explore ways to intelligently match employees to tasks with respect to factors gathered in the initial development of a project [1, 2, 3, 4], to search the test aims for software structure testing [5, 6, 7, 8, 9, 10], to solve multicast routing problems for networks [11, 12], to reduce data dimension and extract data features for an analysis method in Agricultural Engineering [13], and so on.

With the tremendous growth of GA applications as well as the fact that *theory and practice do not talk to each other* [14, 15], to evaluate the applicability of GAs becomes an important research challenge for researchers in the application level. An example for that is, through the literature review, it is discovered that researchers reported success in applying GAs to their problems by comparing the results derived from GAs and other methods. Although GAs outperforms other methods, it is not proved that GAs guide the solutions to the global optima in their work. For most of the real world optimization problems, the optimal solutions are unknown. To determine whether or not to adopt the obtained solution is a difficult decision. In order to help the researchers in application level understand the solutions derived by GAs better, a formalized measurement to estimate the applicability of GAs to real world optimization problems is necessary.

The following section provides a brief literature review on how researchers evaluated the applicability (hardness) of GAs theoretically in the past and identifies some missing links between theory of GAs and application of GAs.

2. LITERATURE SURVEY AND FINDINGS

Researchers have investigated the foundation of GAs since GAs are invented and studied by John Holland [16]. In 1975,

Holland proposed the notion of schemas to formalize the informal notion of "building blocks" [17, 18]. Later, Bethke used discrete Walsh functions to analyze the fitness functions of GAs, and the term *deception* problem was coined by Goldberg [19, 20, 21]. Moreover, Horn suggested the categories of fitness landscapes which are hard for GAs [22]. All of the aforementioned work are explored to estimate the applicability of GAs to optimization problems. However, those are not investigated for practical use. There are missing links from theory of GAs to application of GAs.

- **Missing Link 1:** The fitness functions explored by researchers in theory (e.g., the fitness functions formulated for deception problems) may not be appropriate in practice.

- **Missing Link 2:** The transformations of fitness functions to specific formulas (e.g., walsh functions) are difficult.

- **Missing Link 3:** Most of the methods only analyze the behavior of GAs in a flat population (i.e., a population with individuals distributed uniformly). In that case, the methods cannot successfully capture the behavior of GAs when GAs start to converge.

There are a few researchers who have adopted Markov Chains to analyze the behavior of GAs [14, 23, 24, 25, 26, 27, 28, 29]. In our opinion, transition matrices of Markov Chains capture the essential properties of GAs well. However, in practice, there is a missing link for this approach.

- **Missing Link 4:** The computation time of the transition matrix with respect to an optimization problem is more than the computation time of all the feasible solutions. Therefore, this approach is impractical and all of the findings, such as the convergence rate, and the expected waiting time (i.e., the first hitting time), derived by Markov Chain analyses cannot be applied directly to application of GAs.

In general, the GA theory developed thus far shows that it is difficult to fully capture the behavior of GAs, especially in finite time with different types of fitness functions. To the best of our knowledge, *none of the existing work provided appropriate evaluation metrics for applicability of GAs to real world problems*. Without that, the researchers cannot determine whether or not GAs are applicable to their problems in application domains. Hence, the development of the evaluation metric is crucial.

3. CONCLUSION AND FUTURE WORK

The missing links from theory to application of GAs are identified. Based on the literature review, it is critical to develop the evaluation metric for applicability of GAs to real world applications. Our future research will be focused on the development of the evaluation metrics by extracting only the essential features of transition matrices of Markov Chains with respect to application of GAs. With those measurements, researchers are able to determine whether or not GAs are applicable to certain optimization problems.

4. REFERENCES

[1] C. K. Chang and M. J. Christensen, "A net practice for software project management," *IEEE Software*, vol. 16, no. 6, pp. 80–88, 1999.

[2] C. K. Chang, M. J. Christensen, and T. Zhang, "Genetic algorithms for project management," *Annals of Software Engineering 11*, pp. 107–139, 2001.

[3] C. Chao, "Spmnet: A new methodology for software management," Ph.D. dissertation, Univ. of Illinois at Chicago, 1995.

[4] Y. Ge, "Capability-based software project scheduling with system dynamics and heuristic search," Master's thesis, Iowa State University, 2004.

[5] D. Berndt, J. Fisher, L. Johnson, J. Pinglikar, and A. Watkins, "Breeding software test cases with genetic algorithms," in *Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03)*, 2003, pp. 338–347.

[6] D. J. Berndt and A. Watkins, "Investigating the performance of genetic algorithm-based software test case generation," in *Proceedings of the Eighth IEEE International Symposium on High Assurance Systems Engineering (HASE'04)*, 2004, pp. 261–262.

[7] L. Briand, Y. Labiche, and M. Shousha, "Performance stress testing of real-time systems using genetic algorithms," Carleton University, Tech. Rep. SCE 03-23, 2004.

[8] L. C. Briand, Y. Labiche, and M. Shousha, "Stress testing real-time systems with genetic algorithms," in *Proceedings of the 2005 conference on Genetic and evolutionary computation (GECCO'05)*, 2005, pp. 1021–1028.

[9] J. Clark, J. J. Dolado, M. Harman, R. M. Hierons, B. Jones, M. Lumkin, B. Mitchell, S. Mancoridis, K. Rees, M. Roper, and M. Shepperd, "Reformulating software engineering as a search problem," *IEE Proceedings - software*, vol. 150, no. 3, pp. 161–175, 2003.

[10] F. E. Vieira, R. Menezes, and M. Braga, "Using genetic algorithms to generate test plans for functionality testing," in *Proceedings of the 44th annual southeast regional conference*, 2006, pp. 140–145.

[11] J. H. Siregar, Y. Zhang, and H. Takagi, "Optimal multicast routing using genetic algorithm for wdm optical networks," *IEICE Transactions on Communications*, vol. E88-B, no. 1, pp. 219–226, 2005.

[12] L. S. Randaccio and L. Atzori, "A genetic algorithms based approach for group multicast routing," *Journal of Networks*, vol. 1, no. 4, pp. 1–9, 2006.

[13] H. Yao and L. Tian, "A genetic-algorithm-based selective principal component analysis (ga-spca) method for high-dimensional data feature extraction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 6, pp. 1469–1478, 2003.

[14] H. Jiang, C. K. Chang, D. Zhu, and S. Cheng, "A foundational study on the applicability of genetic algorithm to software engineering problems," in *Proceedings of the 2007 Congress on Evolutionary Computation (CEC'07)*, 2007, pp. 2210–2219.

[15] M. S. et al., "Bridging the gap between theory and practice," *MIT Press Journals - Evolutionary Computation*, vol. 15, no. 4, pp. iii–v, 2007.

[16] J. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. The MIT Press, 1992.

[17] D. A. Coley, *An Introduction to Genetic Algorithms for Scientists and Engineers*. New Jersey: World Scientific Publishing Co. Pte. Ltd., 1999.

[18] M. Mitchell, *An Introduction to Genetic Algorithm*. Cambridge, MA: MIT Press, 1996.

[19] A. D. Bethke, "Genetic algorithm as function optimizers," Ph.D. dissertation, Univ. of Michigan, 1980.

[20] S. Forrest and M. Mitchell, "What makes a problem hard for a genetic algorithm? some anomalous results and their explanation," *Machine Learning*, vol. 13, pp. 285–319, 1993.

[21] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Professional, 1989.

[22] J. Horn and D. E. Goldberg, *Genetic Algorithm Difficulty and the Modality of Fitness Landscapes*, ser. Foundations of Genetic Algorithms 3. San Francisco, CA: Morgan Kaufmann, 1995.

[23] J. Suzuki, "A markov chain analysis on simple genetic algorithms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 25, no. 4, pp. 655–659, 1995.

[24] ——, "A further result on the markov chain model of genetic algorithms and its application to a simulated annealing-like strategy," *IEEE Transactions on Systems, Man and Cybernetics - Part B: Cybernetics*, vol. 28, no. 1, pp. 95–102, 1998.

[25] M. D. Vose and G. E. Liepins, "Punctuated equilibria in genetic search," *Complex Systems*, vol. 5, no. 1, pp. 31–44, 1991.

[26] D. B. Fogel, *Evolutionary Computation*. New York: IEEE Press, 1995.

[27] G. Rudolph, "Convergence analysis of canonical genetic algorithms," *IEEE Transactions on Neural Networks*, vol. 5, no. 1, pp. 96–101, 1994.

[28] K. A. DeJong, W. M. Spears, and D. F. Gordon, "Using markov chains to analyze gafos," in *Foundations of Genetic Algorithms 3*, 1995, pp. 115–137.

[29] L. Ding and J. Yu, "Some theoretical results about the computation time of evolutionary algorithms," in *Proceedings of the 2005 conference on Genetic and evolutionary computation (GECCO'05)*, 2005, pp. 1409–1415.