

The Characteristic of Logistics Network Specific Coding in Genetic Algorithms

Tzong-Heng Chi

Department of Information Management
Aletheia University
chi@ieee.org

Cheng-Li Lin

Department of Information Management
Aletheia University
aj003428@gmail.com

***Abstract**-Logistics network related problems are usually associated with geographical locations, but most of evolutionary computing heuristics such as genetic algorithms (GA) in solving them have not given appropriate labeling for locations. By our experiments, it can introduce fatal failure sometimes. However, once we took the linkage information into consideration, we found the linkage learning genetic algorithm (LLGA) is a more stable solution method possessing the independence of coding schemes for facility location problems than the simple genetic algorithm. Except for what mentioned above and many parameter-setting experiments, we also address the way to improve the performance of the LLGA with more but limited interpretation points. All of these can be good guidelines for users interested in applying the evolutionary computing to solve logistics network related problems.*

Keywords: Linkage Learning, Genetic Algorithm, Logistics Networks.

1. Introduction

Ever since Holland proposed the genetic algorithm (GA) to tackle combinatorial problems [6], the GA has been one of the most efficient solution methods for various challenging network design problems in supply chains. These problems include: vehicle routing problems, scheduling problems, bus network optimization, set covering problems, and location-allocation problems [17]. However, when the GA is applied to solve logistics network design problems, we usually assign each candidate location, modeled as a node in a graph, an integer number. Then the permutation encoding is applied further as a part of the design for the GA. In GA design, this almost has been the standard way to treat mentioned problems. Yet for the location or logistics network related problems, the simple GA (SGA) with

integer coding scheme for nodes can cost more memory and time by our experiments. Moreover, we found that if nodes are not assigned with numbers probably, the GA may fail and converge to the local optimal no matter with the integer or binary encodings. Therefore, our question is: given a logistics network related problem, can we find a GA to solve it stably independent to the modeling scheme of problems?

In fact, there have been many variants of the GA with miscellaneous designs. There is a good idea behind them is the tight linkage which has been thought evolutionarily advantageous [5]. We further found research in the past has shown that the GA with good ability in learning genetic linkage and exploiting good building-block (BB) linkage can solve bounded hard problems quickly, accurately, and reliably [1]. In contrast to the abundant studies about the concept of building block, the topic of linkage was less noticed. Therefore, having a good insight about the genetic linkage could be the key to make GA success. Fortunately, it seems that research in the linkage learning has been increasingly recent rich. Yu and Goldberg [15] have even sorted out linkage learning methods into implicit [5], explicit [10], probabilistic [13][7], and deterministic [16] ones. More detailed classification and survey can refer to [2]. One of them is called linkage learning GA (LLGA) which was first introduced by Harik [4]. About the applications of the LLGA to show its practicability, there are protein structure prediction problems and the design of withdrawal weighted SAW filters [3][14]. Except for these, most of related research is more focused on the theoretical studies. That being the case, we further like to know can the LLGA benefit other real-world problems such as the logistic network design problem. It was expected to have a better insight when applying the LLGA because of its inherent visual process such as shown in the [1]. One more thing we need to consider: can the LLGA be a stable solution method for the logistic network related problem often encountered in the supply

chain design? If so, what are strong and weak points in using it? Also, can we understand the logistics network related problem more by the usage of the LLGA?

For the following section, it introduces the coding problem when the logistics network related problem is considered. Section 3 gives a brief introduction to the LLGA. After that, section 4 shows our results and analyses for many experiments. Finally, we provide our conclusions for this research.

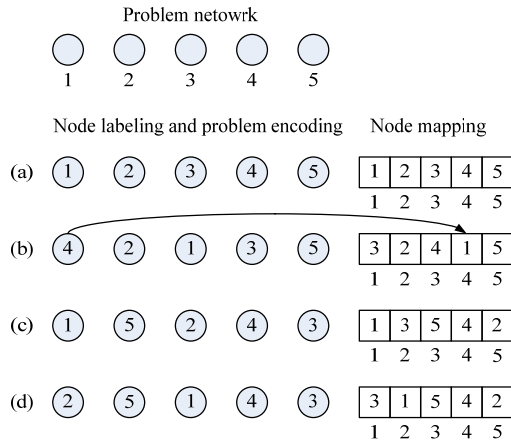


Figure 1. Four different encodings with different labeling for the same problem.

2. Logistics-Specific Coding Problems

Facility location decisions play a critical role in the strategic design of logistics networks [8][12]. Operation research practitioners have developed a number of mathematical programming models to represent a wide range of location problems. Unfortunately, the resulting models can be extremely difficult to solve to optimality [12]. Thus, it seems being worthy to reexamine facility location problems with a visual way such as provided by the LLGA. The implementation of the LLGA is more complicated than the SGA, but the LLGA can provide more concrete evidence to show the evolutionary course than the SGA.

The p-median, p-center, location set covering, and maximum covering location problems have been four core models for the facility location problem [12]. The p-median problem asks for selected p facilities can minimize the total (average) distances (or costs) for supplying customer demands. As the p-center problem, it is asked for selected p facilities can minimize the maximum distance between any customer and its nearest center. The location set covering problem locates the least number of centers to serve all customers.

The maximum covering location problem seeks the maximum coverage with a given number of centers.

To explain it in more detail, we first give an one-dimensional example as shown in Figure 1. All five nodes of that problem network are arranged along the line. If we like to solve them with the SGA, the first step is to encode the problem into the chromosome by labeling nodes with integer numbers. We may label nodes in different ways. It is supposed that we get the four different encodings as shown in the Figure 1. In the Figure 1 (b), it means that the node on first physical location is labeled number 4. That is, it is encoded into the fourth position in the chromosome array and adjacent to the physical node 4 and 5. This adjacent relationship is incorrect. Worse, it can be the noise to mislead the solution method such as SGA possibly. If the guess about the genetic linkage was right, then the adjacent relationship between the first node and second node is disrupted. We may see the longer convergence time. In fact, its effect to the performance of the SGA is overlooked all the time.

We give a very simple metric here. The total coding offset (TCO) is the sum of differences the labeled integer between adjacent nodes, that is, $\sum_1^k (|i - j| - 1)$, where k is the length (or number of nodes) of the coding and i and j are labels of adjacent nodes. Therefore, the TCO is 0, 3, 6, 7 for (a) to (d) separately. This is the relative measurement. For this simple example, we can also calculate the absolute offset easily by subtracting the labeled integers from their absolute positions. Then, we get 0, 6, 7, and 8 separately. If these TCOs can represent the genetic linkage, they show gene loci should be capable of being changed to get better evolutionary performance. However, by our experiments which will be introduced in the section 4, we found the coding offset is roughly in proportion to the convergence time. The key point is that we can label nodes with right number only if we know the answer. It is impossible for real-world problems.

Finally, following the Harik's theoretical works [4], the characteristics of logistics network specific coding could be the clue coming from the real-world problems to show the requirement of genetic linkage. Almost all models of facility location problems are mirrors of the real geography no matter adopting which kind of distance metric. Moreover, for the global logistics networks, the amount of nodes is quite huge in a big geographical 3-dimensional space. It is even not easy to label them well with all directions. If

solution methods can promise us the independence of coding schemes, things will be getting easier.

3. Linkage Learning GA (LLGA)

In general, the GA is referred to as a stochastic artificial intelligent technique whose solution search process mimics natural evolutionary phenomena: genetic inheritance and Darwinian strife for survival [9]. The GA works by discovering, emphasizing, and recombining good “building blocks” of solutions in a highly parallel fashion [6]. That is, good solutions tend to be made up of good building blocks - combinations of bit values that confer higher fitness on the strings in which they are present. Also, it has been assumed the structural representation would finally evolve tightly linked representations with dedicated designed genetic operators [5]. Harik hence proposed the LLGA which can learn genetic linkage in the evolutionary process with correspondence to the Holland’s call for the evolution of tight linkage [4][1]. The key point is ability to identify or separate building blocks. After that, we may avoid the disruption of building blocks and link two or more good building blocks together to get higher fitness. However, the forces of selection usually are strong to choose deceptive or worse building blocks then the evolution by mutation can be too slow and fail. In another word, the evolution is possible to be successful only by pair-wise recombination with optimal or better building blocks if the chromosome contains all associated allele values for each locus and the diversity is reserved.

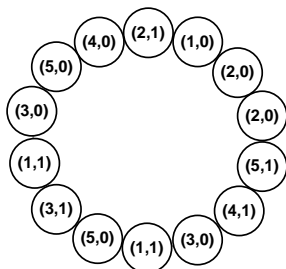


Figure 2. The circular chromosome called EPE-2 used in the LLGA with 5 genes where 1 and 2 are coding genes and 3, 4, and 5 are non-coding genes.

Because that Harik picked the two-point crossover up to define and measure linkages among genes, he naturally collocated genes as circular chromosome representation [4]. He further introduced probabilistic expression (PE) where each chromosome represents not one solution, but

a probability distribution over the range of possible solutions. For the PE, each chromosome will contain a fully-over-specified set of exons spaced by introns. Exons are coding or functional genes. The complement coding genes are used to conserve the diversity and actual genes will be expressed by the fitness function. Introns are instead non-coding genes do not participate in the fitness function but used to implement the probabilistic scheme. They facilitate the propagation of building blocks [14]. There is an interpretation point which is randomly chose to determine the canonical form with clockwise manner for each chromosome.

Once encoding is done, the LLGA has two phases: selection and exchange. Usually the tournament selection is applied to pick the better one into the mating pool. The exchange is more complicated because of the circular forms of chromosomes. The exchange operator is defined on pairs of chromosomes. The segment randomly cut from donating chromosome is injected into the recipient chromosome on the grafting point. After the removal of duplications, the recipient chromosome becomes a new child chromosome stepping into the next generation.

Leaving additional copies of the unexpressed allele in a chromosome increases the chance that allele being transferred in a subsequent exchange. This extended probabilistic expression, called EPE-n where n is the number of copies of genes, enhances the LLGA maintaining diversity for a longer time. It can actively resist an allele’s convergence through the propagation of its alternatives with increased probability. Thus, the building blocks can survive with lower possibility of disruption under higher selection rate. It is a good compromise between fast convergence and fast exploration. In this paper, we follow and use EPE-2. For more details, readers can refer to [4][1].

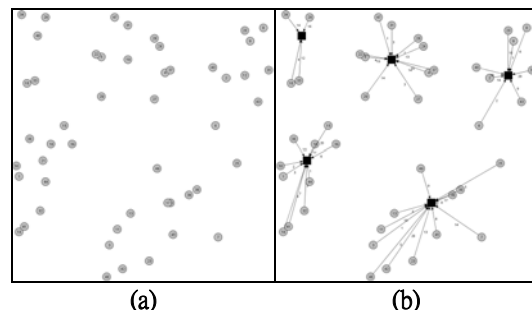


Figure 3. (a) The graph of the pmedcap1-1 problem. (b) One of its optimal solutions known so far.

4. Experiments

Our experiments will examine the effects of SGA and LLGA for different facility location problems including the p-median and p-center. Due to the space limitation, we may not provide too many details about testing problems here. However, to look after fairness, effectiveness, and time limitation, we refer some problems and known results of pmedcap1 which is available in the public OR-library [11] known as the benchmark. The testing problem set pmedcap1 includes 20 problems with two scales: 50 and 120 nodes. We assign the number for each of them one by one. Next, to tune the LLGA well, other experiments are made to find good parameter settings. For every test, statistics are made based on at least 25 runs and most 100 runs of the test. Without extra explanation, the population size is 100 or 300 and the GA stops at the 100th or 1000th generation for each testing run. If the SGA is used, the crossover rate is 0.8 and the mutation rate is 0.08. As the LLGA, the crossover rate is 1 and the tournament size is 4. Also, there is only one interpretation point for the LLGA if without additional comments. Finally, the length of non-coding genes is 1000.

Table 1. the number of optimal solutions for binary and permutation coding.

Coding scheme	Number of optimal solutions	
	Average fitness	# of opt. solution within 25 runs
Permutation	741.061	3
Binary	711.326	22

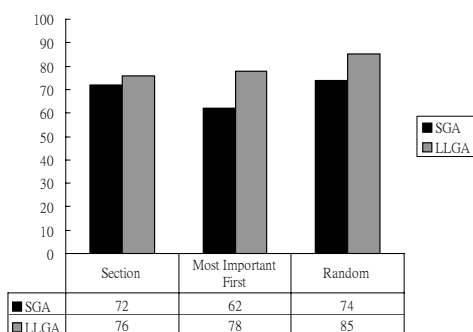


Figure 5. The comparison of 3 different binary encoding strategies for the 5-median problem. The y-axis represents the number of optimal solutions within 100 runs of tests.

Figure 3. is the graph model of the pmedcap1-1

problem. If $p=5$, the SGA will choose 5 sites from 50 nodes to be the median sites for serving customers. Both of the binary (with coding length 50) and permutation (with coding length 5) encodings are used for this 5-median problem separately. By the experimental results of Table 1, it is easy to find that binary coding is much better than the permutation coding. This result may show the optimal 5 sites are not relevant to each other too much. After all, all of them are supported by the neighboring nodes not appeared in the coding. And the SGA can only apply its own searching mechanisms. Instead, let the SGA consider all of locations simultaneously, the linkage information may affect the evolutionary process.

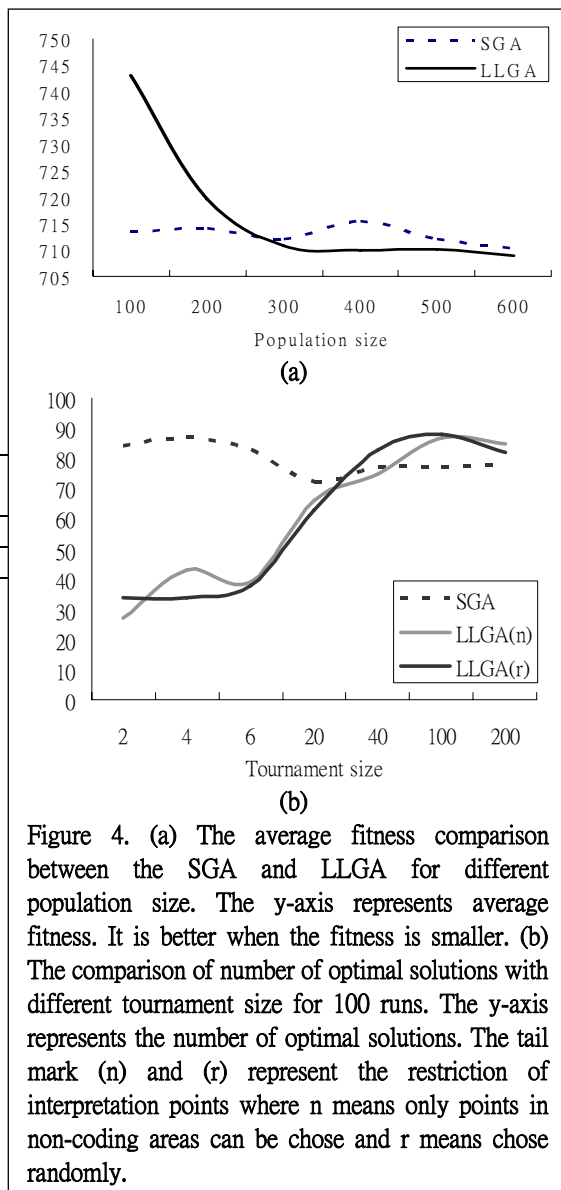


Figure 4. (a) The average fitness comparison between the SGA and LLGA for different population size. The y-axis represents average fitness. It is better when the fitness is smaller. (b) The comparison of number of optimal solutions with different tournament size for 100 runs. The y-axis represents the number of optimal solutions. The tail mark (n) and (r) represent the restriction of interpretation points where n means only points in non-coding areas can be chose and r means chose randomly.

For what mentioned above, it is about the competition between binary and permutation encodings. For the binary coding itself, we examine three different encoding strategies: section, most important first, and random. Firstly, we surround the known optimal median sites with their true neighboring locations to form many sections. Secondly, we deliberately put the known optimal median sites in the front of the coding as the permutation encoding. Finally, we just randomly label the nodes and encode them into the chromosome. By the Figure 5, it tells us the sectional strategy is the best one. Again, the linkage among gene loci is important to the evolutionary performance. As the influence with different encodings to the LLGA, it is not significant as shown in Figure 5. For the p-center problem, our experiments show the similar results.

The stable performance of the LLGA should be deserved much notice. As assumed in the Harik's studies, the convergent process can be divided into both of gene locus and gene allele phases. This is also the reason why the LLGA has to slow down the convergence. That is, if we like to make the alleles converge to the global optimal, we have to make the gene loci in the right positions first. After the examination of the characteristic of logistics network specific coding in the SGA and LLGA. We conclude the LLGA is more stable and outperform the SGA.

To apply the LLGA better, we hence perform other experiments for different population sizes and tournament sizes. This can be critical because that the selection pressure is a key to the LLGA. As shown in the Figure 4, the performance of the LLGA is in proportion to the population size and tournament size. We also can see the LLGA outperforms the SGA when the population size exceeding 300 and the tournament size exceeding 50. This result should not be so surprise. But the other hand, to take care both of selection resistance and diversity preservation, the LLGA abandons the mutation. Moreover, the chromosome is only a distribution but not a deterministic mapping to the phenotype. That is, we think the population size and tournament size are two parts of the cost to maintain the diversity.

Next, let's go back to our main theme. What is the relationship between the coding scheme and the concept of building block? Does the building

block really exist for logistics network related problems? Further, is the gene locus indeed influential? From the Figure 5., we could boldly conclude it that the LLGA is independent of coding schemes because it takes the gene locus, or more correctly is that the gene linkage, into consideration. Figure 6 shows convergent process for 3-median problems with 20 locations. We can see the formation of building blocks at the lowest side. Finally, all of coding genes are put together as a long building block with assumed ordering if possible. This is more concrete than the imagination.

As mentioned, the LLGA must pay a lot of efforts to outperform the SGA. To gain a better control, Chen proposed the mechanism of promoters [1] to restrict the number and position of interpretation points. Same here! We also like to cool down the degree of uncertainty. Therefore, we modify the LLGA with neighborhood exploration and call it the nLLGA. We provide each chromosome more chances to sample other interpretations as candidate canonical chromosomes. We randomly choose different interpretation points n times where n is a parameter. The best one of them will be chose. This can be analogized to the tournament selection in its own class if that the circular chromosome represents a set of canonical chromosomes. By our experiments, it is indeed effective. We show the result as follows.

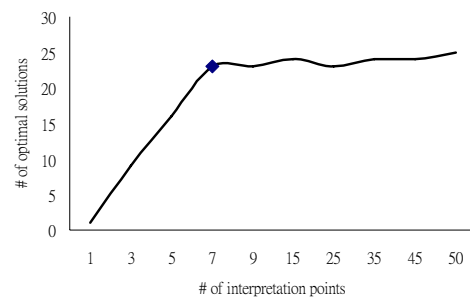


Figure 7. The comparison of the number of optimal solutions for different number of interpretation points.

If we randomly sample 50 interpretations from 150 possible starting points, the LLGA can reach optimum each run for the testing 5-median problem pmedcap1-1 with 50 locations. There are



Figure 6. The convergent process of the LLGA for 3-median problems with 30 locations.

50 genes with binary allele for 50 locations. The number 150 is because that we apply EPE-2 to the LLGA. Within it, there are 3 copies for each gene. However, the best number of sampling points can be the 7 for this testing problem. This means it is not appropriate for picking only one interpretation point in the original LLGA. As how much is the best for other problems, more studies are needed to be done.

5. Conclusion

To our knowledge, there is none like us to address the characteristic of logistics network specific coding in genetic algorithms systemically as so. For this application domain, we also points out the importance of gene linkage often overlooked before. Excepting the demonstration of convergent process for the facility location problem, we might be the first one to apply the LLGA to the logistics network related problem, too. Meanwhile, once guidelines needed, interested users can refer our parameter-setting experiments for the LLGA including population size, tournament size, and the number of interpretation points which is never mentioned in other research. Also, while the global supply chain is booming, this exploratory research could be helpful to solve the increasing complex logistics network design problems.

References

- [1] Y.-P. Chen, *Extending the scalability of linkage learning genetic algorithms: Theory and practice*, Volume 190 of Studies in Fuzziness and Soft Computing, Springer, 2005.
- [2] Y.-p Chen, T.-L. Yu, K. Sastry, and D. E. Goldberg, "A survey of linkage learning techniques in genetic and evolutionary algorithms," *IlligAL Report No. 2007014*, University of Illinois at Urbana-Champaign, Urbana IL, 2007.
- [3] Karl R. Deerman, Gary B. Lamont, and Ruth Pachter, "Linkage-learning genetic algorithm application to the protein structure prediction problem," *Proceedings of the 2001 ACM symposium on Applied computing*, 333 – 339, 2001.
- [4] G. R. Harik. *Learning gene linkage to efficiently solve problems of bounded difficulty using genetic algorithms*. PhD thesis, University of Michigan, Ann Arbor, MI, 1997.
- [5] G. R. Harik and D. E. Goldberg, "Learning linkage," *Foundations of Genetic Algorithms 4*, 247-262, 1996.
- [6] J. H. Holland, *Adaption in natural and artificial systems*. University of Michigan Press, Ann Arbor, 1975.
- [7] P. Larranaga, & J. A. Lozano, (Eds.), *Estimation of distribution algorithms*, Kluwer Academic Publishers, 2002.
- [8] M. T. Melo, S. Nickel, F. Saldanha-da-Gama, "Facility location and supply chain management – A review," *European Journal of Operational Research*, doi:10.1016/j.ejor.2008.05.007, 2008.
- [9] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer, Berlin, 1999.
- [10] M. Munetomo & D. E. Goldberg, "Identifying linkage groups by nonlinearity/non-monotonicity detection," *Proceedings of the Genetic and Evolutionary Computation Conference 1999*, Volume 1, 433-440, 1999.
- [11] I.H. Osman and N. Christofides, "Capacitated clustering problems by hybrid simulated annealing and tabu search," *International Transactions in Operational Research*, Vol. 1, No. 3, 317-336, 1994.
- [12] S.H. Owen, M.S. Daskin, "Strategic facility location: A review," *European Journal of Operational Research*, 111, 423–447, 1998.
- [13] M. Pelikan, D. E. Goldberg, & F. G. Lobo, "A survey of optimization by building and using probabilistic models," *IlligAL Report No. 99018*, University of Illinois at Urbana-Champaign, Illinois Genetic Algorithms Laboratory, 1999.
- [14] V. Prabhu, B. S. Panwar, and Priyanka, "Linkage Learning Genetic Algorithm for the Design of Withdrawal Weighted SAW Filters," *Proceedings of IEEE Ultrasonics Symposium*, 357-360, 2002.
- [15] T.-L. Yu, K. Sastry, & D. E. Goldberg, "Linkage learning, overlapping building blocks, and systematic strategy for scalable recombination," *Proceedings of the Genetic and Evolutionary Computation Conference 2005 (GECCO 2005)*, 1217-1224, Jun. 2005.
- [16] T.-L. Yu, D. E. Goldberg, A. Yassine, & Y.-P. Chen, "Genetic algorithm design inspired by organizational theory: Pilot study of a dependency structure matrix driven genetic algorithm," *Proceedings of Artificial Neural Networks in Engineering 2003 (ANNIE 2003)*, 327-332, 2003.
- [17] Gengui Zhou, Hokey Min, and Mitsuo Gen, "A Genetic Algorithm Approach to the Bi-criteria Allocation of Customers to Warehouses." *International Journal of Production Economics*, Vol. 86, 35-45, 2003.