

Tuning Selection Pressure in Tournament Selection

Huayang Xie and Mengjie Zhang

Abstract—Selection pressure controls the selection of individuals from the current population to produce a new population in the next generation. It gives individuals of higher quality a higher probability of being used to create the next generation so that Evolutionary Algorithms (EAs) can focus on promising regions in the search space. An evolutionary learning process is dynamic and requires different selection pressures at different learning stages in order to speed up convergence or avoid local optima. Therefore, it desires selection mechanisms being able to automatically tune selection pressure during evolution. Tournament selection is a popular selection method in EAs. This paper focuses on tournament selection and shows that standard tournament selection is unaware of the dynamics in the evolutionary process thus is unable to tune selection pressure automatically. This paper then presents a novel approach which integrates the knowledge of the Fitness Rank Distribution (FRD) of a population into tournament selection. Through mathematical modelling, simulations and experimental study, this paper shows that the new approach is effective and using the knowledge of FRD is a promising way to modify the standard tournament selection method for tuning the selection pressure dynamically and automatically along evolution.

Index Terms—Tournament Selection, Selection Pressure, Tuning Strategy

I. INTRODUCTION

Evolutionary Algorithms (EAs) are inspired by biological evolution such as reproduction, mutation, recombination, natural selection and survival of the fittest, that is, the Darwinian natural selection theory. An instance of EAs can be abstracted as searching solutions by applying genetic operators to populations of individuals iteratively and often in a parallel manner. Therefore, there are many factors that can affect the evolutionary search performance of an instance of EAs for given problems. These factors include the size of a population, the representation of individuals in a population, the fitness evaluation of individuals, the selection mechanisms for reproduction and for survival, the genetic operations for modifying individuals, and many more. Amongst these factors, selection mechanisms play an extremely important role.

A selection mechanism consists of a selection scheme and a selection pressure control strategy. The latter is critical in designing a selection mechanism and has been widely studied in EAs [1], [2], [3], [4], [5], [6], [7]. According to the configuration of selection pressure, the search in EAs can have two extremes. One extreme, when there is no selection pressure, is completely stochastic so that the search acts just like the Monte Carlo method [8], randomly sampling the space of feasible solutions. The other extreme, when the selection pressure is very high, is minimally stochastic so that the search

acts like a local hill-climbing search method. It is clear that in general the drawback of the former extreme is its inefficiency and the drawback of the latter extreme is its possible confinement to local optima or “*premature convergence*”. Therefore, an effective and efficient evolutionary search algorithm must balance between these two extremes. In order to obtain the balanced situation, selection pressure, the key element in the selection mechanism, must be properly tuned so that the stochastic elements are maintained at an optimal level.

A common view of tuning selection pressure is that the selection pressure should be adapted during an EA run. For instance, the selection pressure should be weak at first to allow for more exploration and then stronger towards the end as an EA converges [9]. However, the evolutionary learning process itself is actually more dynamic than that. At some stages, it requires a fast convergence rate (i.e., high selection pressure) to find a solution quickly; at other stages, it requires a slow convergence rate (i.e., low selection pressure) to avoid being confined to a local optimum or converging prematurely. More importantly, these stages seem to appear in a mixed order rather than ones are always predecessors of the others. Although selection has been studied for several decades, due to the existence of the dynamic requirements, tuning selection pressure is still difficult, not as easy to parameterise as other factors (i.e., population size) [9], and remains an important open problem in EAs’ research.

Since the 1970s, there have been many selection schemes developed in EAs for selecting parents. Commonly used selection schemes include fitness proportionate selection [10], ranking selection [11], and tournament selection [12].

Standard tournament selection randomly draws/samples k individuals with replacement from the current population of size N into a tournament of size k and selects the one with the best fitness as a parent from the tournament into the mating pool [13], [14]. In general, selection pressure in tournament selection can be easily changed by using different tournament sizes; the larger the tournament size, the higher the selection pressure. Drawing individuals with replacement into a tournament makes the population remain unchanged, which in turn allows tournament selection to easily support parallel architectures. Selecting the winner involves simply ranking individuals partially (as the best one is only concerned) in a tournament of size k , thus the time complexity of a *single* tournament is $O(k)$. If the total number of tournaments required to generate the entire next generation is N^1 , the time complexity of tournament selection is $O(kN)$. The ease of changing selection pressure, the support of parallel architectures, and the linear time complexity have made tournament selection very

Huayang Xie and Mengjie Zhang are with the School of Engineering and Computer Science, Victoria University of Wellington, New Zealand (e-mail: {hxie, mengjie}@ecs.vuw.ac.nz)

¹This assumption commonly holds in EAs when two-offspring crossover operators and one-offspring mutation operators are used, and the population size remains fixed during evolution.

attractive in EAs, especially in Genetic Algorithms (GAs) and Genetic Programming (GP) [1], [2], [15], [16], [17], [18]. This paper, therefore, focuses on tournament selection to investigate the selection pressure tuning problem.

Although tuning selection pressure in tournament selection seems to fall in the parameter setting research field [19] as the selection pressure can be influenced by changing the tournament size parameter in tournament selection, performing this tuning is not as straightforward as being expected. During evolution, typically after a few generations, some individuals in the population would have the same or very similar fitness values. These individuals with the same/similar fitness values can be considered a "group", and a population can be considered having different groups of individuals. When this happens, the selection pressure between these "groups" increases, resulting in "better" groups dominating the next population and possibly causing premature convergence. We refer this as the *high between-group selection pressure* issue. This issue is actually part of the general dynamic issue in EAs. In other words, the tournament size itself is not always adequate for controlling the selection pressure during evolution.

A. Goals

To avoid the above issue (or improve the situation), this paper aims to develop a new selection method to dynamically tune selection pressure along evolution. To achieve this goal, we will firstly analyse the standard tournament selection mechanism via mathematical modelling and simulations to reveal why this mechanism is not aware of the evolution dynamics and can not automatically adjust the selection pressure during evolution. According to the findings of the theoretical analysis and simulations, we will develop a new selection method based on population clustering to address how the issue can be resolved. We will further analyse the method and compare it with the standard tournament selection method via mathematical modelling and simulations, and finally via empirically experiments on three typically data sets using GP, a common evolutionary algorithm. We expect the mathematical modelling, simulation analysis and the empirical experiments to reveal whether and/or how the proposed new method can resolve the above issue, whether and/or how the new selection method can automatically tune/adjust selection pressure during evolution, and whether and/or how the new method outperforms the standard tournament selection mechanism on different kinds of problems.

B. Organisation

Section II introduces background. Section III investigates the awareness of evolution dynamics of standard tournament selection. Sections IV and V present and analyse a novel approach to solving the selection pressure tuning problem. Section VI further investigates the effectiveness of the approach via experiments. Section VII concludes this paper.

II. LITERATURE REVIEW

This section gives a review of tournament selection alternatives, selection pressure measurements and selection behaviour modelling for standard tournament selection.

A. Alternative tournament selections for controlling selection pressure

Since tournament sizes are integer numbers, the selection pressure is controlled only at a coarse level originally. In order to control selection pressure at a fine level, Goldberg and Deb developed the probabilistic tournament selection [1]. In the form of tournament selection, an extra probability p is introduced. When conducting a tournament between two individuals, the individual with higher fitness value can be selected as a parent with the probability p , while the other has the probability $1 - p$. By setting p between 0.5 and 1, it is possible to control the selection pressure continuously between the random selection and the tournament selection with a tournament size of two. Later, Julstrom and Robinson introduced a weighted k-tournament which extends the probability tournament selection from two contestants to k contestants [20]. The weighted k-tournament assigns fixed probabilities to the ranks of its k contestants, and selects one to be a parent according to those probabilities. Recently, Hingee and Hutter [21] showed that every probabilistic tournament is equivalent to a unique polynomial ranking selection scheme.

Huber and Schell argued that the probabilistic tournament selection enabled a fine scaling of selection pressure but at an expense of increased complexity and reduced efficiency [22]. They introduced a mixed size tournament selection to fine-control the selection pressure at the cost of a minimal increase of the complexity and with almost no loss of efficiency. The mixed size tournament selection is implemented by introducing tournaments of varying size during a parent selection phase.

Filipović *et al.* [23] investigated a fine-grained tournament selection method for a simple plant location problem in GAs. They argued that standard tournament selection does not allow precise setting of the balance between exploration and exploitation [2]. In their fine grained tournament selection method, the tournament size is not fixed but close to a pre-set value. They claimed that the fine grained tournament selection makes the ratio between exploration and exploitation able to be set precisely, and that the method solves the simple plant location problem successfully.

Sokolov and Whitley developed an unbiased tournament selection in order to reduce selection pressure by ensuring every individual in a population be sampled in tournaments [18]. They believed that the uniformly random sampling process is a bias present in standard tournament selection. The bias can cause a potential for better individuals not to be selected for recombination. Therefore, they developed the unbiased tournament selection that "lines up two different permutations of the population and performs a pairwise comparison" with a constraint, which forces compared individuals to be distinct. As a consequence, every individual is sampled at least once. A tournament size 2 was used to test the unbiased

tournament selection on three problems, one with permutation-based solution representation and two under bit encoding. Although the advantage of a generational genetic algorithm using the unbiased tournament selection varied for different population sizes on the three problems, the authors concluded that the impact of the bias is significant, and the unbiased tournament selection provides better performance than other selection methods, including standard tournament selection, a rank based selection and fitness proportionate selection.

B. Selection pressure measurements

Selection pressure controls the selection of individuals from the current population to produce a new population in the next generation. It gives individuals of higher quality a higher probability of being used to create the next generation so that EAs can focus on promising regions in the search space [2].

In tournament selection, the mating pool consists of tournament winners. The average fitness in the mating pool is usually higher than that in the population. The fitness difference between the mating pool and the population reflects the selection pressure, which is expected to improve the fitness of each subsequent generation [3].

In biology, the effectiveness of selection pressure can be measured in terms of differential survival and reproduction, and consequently in change in the frequency of alleles in a population. In EAs, there are several measurements for selection pressure in different contexts, including *takeover time*, *selection intensity*, *loss of diversity*, *reproduction rate*, and *selection probability distribution*.

Takeover time is defined as the number of generations required to completely fill a population with just copies of the best individual in the initial generation when only selection and copy operators are used [1]. For a given fixed-sized population, the longer the takeover time, the lower the selection pressure. Goldberg and Deb [1] estimated the takeover time for standard tournament selection as

$$\frac{1}{\ln k} (\ln N + \ln(\ln N)) \quad (1)$$

where N is the population size and k is the tournament size (these two notations will be used through out this paper unless otherwise noted). The approximation improves when $N \rightarrow \infty$. However, this measure is static and constrained and therefore does not reflect the selection behaviour dynamics from generation to generation in EAs.

Selection intensity was firstly introduced in the context of population genetics to obtain a normalised and dimensionless measure [24], and, later was adopted and applied to GAs [25]. Blickle and Thiele [2], [26] measured it using the expected change of the average fitness of the population. As the measurement is dependent of the fitness distribution in the initial generation, they assumed the fitness distribution followed the normalised Gaussian distribution and introduced an integral equation for modelling selection intensity in standard tournament selection.

For their model, analytical evaluation can be done only for small tournament sizes and numerical integration is needed for large tournament sizes. The model is not valid in the case

of discrete fitness distributions. In addition to these limitations, the assumption that the fitness distribution followed the normalised Gaussian distribution is not valid in general [27]. Furthermore, because the actual fitness values are ignored but the relative rankings are used in tournament selection, the model is of limited use.

Loss of diversity is defined as the proportion of individuals in a population that are not selected during a parent selection phase [2], [26]. Blickle and Thiele [2], [26] estimated the loss of diversity in the standard tournament selection as:

$$k^{-\frac{1}{k-1}} - k^{-\frac{k}{k-1}} \quad (2)$$

However, Motoki [4] pointed out that Blickle and Thiele's estimation of the loss of diversity in tournament selection does not follow their definition, and indeed their estimation is of loss of *fitness* diversity. Motoki recalculated the loss of *individual* diversity in a *wholly diverse* population, i.e., every individual has a distinct fitness value, on the assumption that the worst individual is ranked 1st, as:

$$\frac{1}{N} \sum_{j=1}^N (1 - P(W_j))^N \quad (3)$$

where $P(W_j) = \frac{j^k - (j-1)^k}{N^k}$ is the probability that an individual of rank j is selected in a tournament.

“Reproduction rate” is defined as the ratio of the number of individuals with a certain fitness f after and before selection [2], [26]. A reasonable selection method should favour good individuals by giving them a high ratio and penalise bad individuals by giving a low ratio. Branke *et al.* [28] introduced a similar measure which is the expected number of selections of an individual. It is calculated by multiplying the total number of tournaments conducted in a parent selection phase by the selection probability of the individual in a single tournament. They also provided a model to calculate the measure for a single individual of rank j in standard tournament selection in a wholly diverse population on the assumption that the worst individual is ranked 1st, as:

$$N \times P(W_j) = N \frac{j^k - (j-1)^k}{N^k} \quad (4)$$

This measure is termed *selection frequency* in this paper hereafter as “reproduction” has another meaning in GP and this new term can better reflect its real meanings in this situation.

Selection probability distribution of a population at a generation is defined as consisting of the probabilities of each individual in the population being selected at least once in a parent selection phase [29]. Although tournaments indeed can be implemented in a parallel manner, in [29] they are assumed to be conducted sequentially so that the number of tournaments conducted reflects the progress of generating the next generation. As a result, the selection probability distribution can be illustrated in a three dimensional graph, where the x-axis shows every individual in the population ranked by fitness (the worst individual is ranked 1st), the y-axis shows the number of tournaments conducted in the selection phase (from 1 to N), and the z-axis is the selection probability which shows how likely a given individual marked on x-axis can be

selected at least once after a given number of tournaments marked on y-axis. The selection probability is calculated by Equation 9, which is to be described in the next sub section. The measure somehow provides a full picture of the selection behaviour over the population during a parent selection phase. Figure 1 shows the selection probability distribution measure for standard tournament selection of tournament size 4 on a wholly diverse population of size 40.

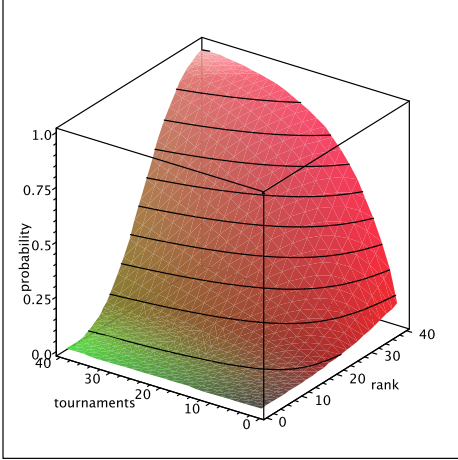


Fig. 1. An example of the selection probability distribution measure.

C. Sampling and Selection Behaviour Modelling

Based on the concept of takeover time [1], Bäck [30] compared several selection schemes, including tournament selection. He presented the selection probability of an individual of rank j in one tournament for a minimisation task (the best individual is ranked 1st), with an implicit assumption that the population is wholly diverse as:

$$N^{-k}((N - j + 1)^k - (N - j)^k) \quad (5)$$

In order to model the expected fitness distribution after performing tournament selection in a population with a more general form, Bickel and Thiele [2] extended the selection probability model in [30] to describe the selection probability of individuals with the same fitness. They defined the worst individual to be ranked 1st and introduced the *cumulative fitness distribution*, $S(f_j)$, which denotes the number of individuals with fitness value f_j or worse. They then calculated the selection probability of individuals with rank j as:

$$\left(\frac{S(f_j)}{N}\right)^k - \left(\frac{S(f_{j-1})}{N}\right)^k \quad (6)$$

In order to show the computational savings in backward-chaining evolutionary algorithms, Poli and Langdon [31] calculated the probability that one individual is not sampled in one tournament as $1 - \frac{1}{N}$, then consequently the expected number of individuals not sampled in any tournament as:

$$N \left(\frac{N}{N-1}\right)^{-ky} \quad (7)$$

where y is the total number of tournaments required to form an entire new generation.

In order to illustrate that selection pressure is insensitive to population size in standard tournament selection in a population with a more general situation (i.e., some individuals have the same fitness value and therefore have the same rank), Xie *et al.* [29] presented a sampling probability model that any individual p is sampled at least once in $y \in \{1, \dots, N\}$ tournaments as:

$$1 - \left(\left(\frac{N-1}{N}\right)^N\right)^{\frac{y}{N}k} \quad (8)$$

and a selection probability model that an individual p of rank j is selected at least once in $y \in \{1, \dots, N\}$ tournaments as:

$$1 - \left(1 - \frac{\left(\frac{\sum_{i=1}^j |S_i|}{N}\right)^k - \left(\frac{\sum_{i=1}^{j-1} |S_i|}{N}\right)^k}{|S_j|}\right)^y \quad (9)$$

where $|S_j|$ is the number of individuals of the same rank j and the worst individual is ranked 1st.

In the literature, a variety of selection pressure measurements have been developed; and many mathematical models have been introduced for *standard* tournament selection. However, few of researches analysed the awareness of evolution dynamics in tournament selection. In the next section, we will utilise some of the selection pressure measurements and mathematical models to investigate whether standard tournament selection is aware of the evolution dynamics through simulations.

III. ANALYSIS OF AWARENESS OF EVOLUTION DYNAMICS OF STANDARD TOURNAMENT SELECTION

This section analyses why the standard tournament selection mechanism is not aware of evolution dynamics and can not automatically adjust the selection pressure during evolution.

A. Methodology

We assume that population sizes are of a fixed size during evolution and the individual with the worst fitness is ranked 1st. Since a population can be seen as a collection of different sized bags which consists of individuals with equal fitness, and each “fitness bag” is associated with a distinct fitness rank, we characterise a population by the number of distinct fitness ranks and the size of each corresponding fitness bag, which we term *fitness rank distribution* (FRD).

We use four populations with four different FRDs, namely *uniform*, *reversed quadratic*, *random*, and *quadratic*, in our simulations. The four FRDs are designed to mimic the four stages of evolution (but they do not necessarily model all the real situations happening in a true run of evolution). The uniform FRD represents the initialisation stage, where each fitness bag has a roughly equal number of individuals. A typical case of the uniform fitness rank distribution can be found in a wholly diverse population. The reversed quadratic FRD represents the early evolving stage, where commonly very few individuals have good fitness values. The random FRD represents the middle stage of evolution, where better

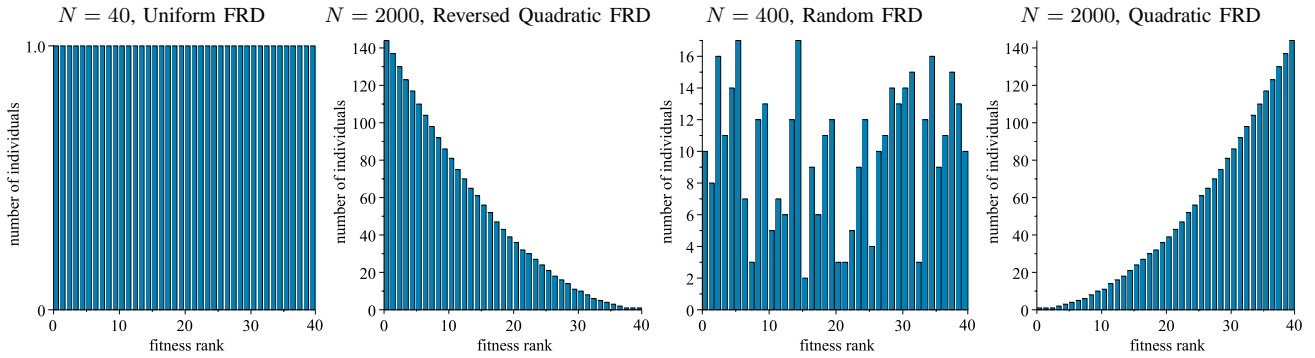


Fig. 2. Four populations with different fitness rank distributions.

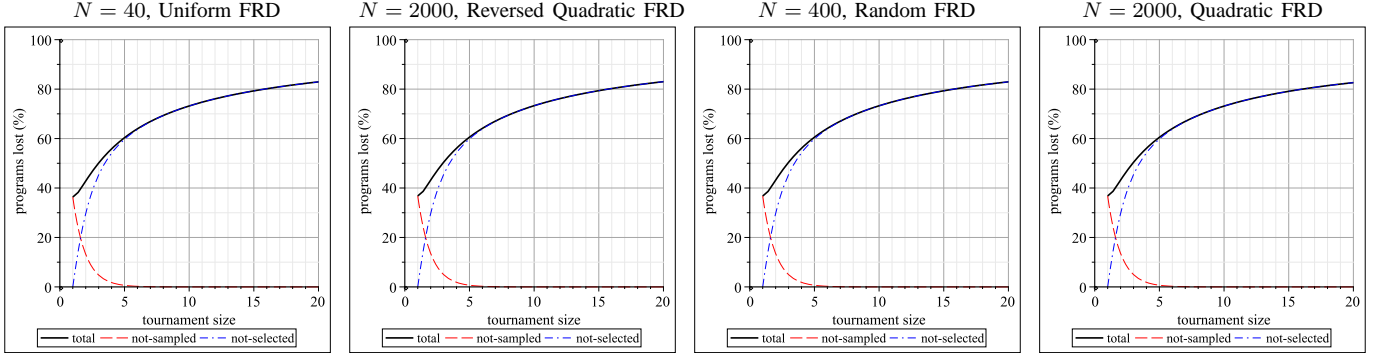


Fig. 3. Loss of individual diversity in the standard tournament selection scheme on four populations with different FRDs. Note that the tournament size is discrete but the plots show curves to aid interpretation.

and worse individuals are possibly randomly distributed. The quadratic FRD represents the later stage of evolution, where a large number of individuals have converged to better fitness values.

We follow the standard breeding operations, that is, two-offspring crossover operations and one-offspring mutation operations, so that the total number of tournaments is the same as the population size N at the end of generating all individuals in the next generation. In order to make the results of the selection behaviour analysis easily understandable, we assume that tournaments are conducted sequentially. We choose only the loss of individual diversity, the selection frequency, and the selection probability distribution measures for the selection behaviour analysis and ignore the takeover time and the selection intensity due to their limitations.

Since the impact of population size on selection behaviour is unclear, we test several different commonly used population sizes, ranging from small to large. This paper illustrates only the results for three population sizes, namely 40, 400, and 2000, for the uniform FRD, the random FRD, and the reversed quadratic and quadratic FRDs respectively. Note that although the populations with different FRDs are of different sizes, we design the number of distinct fitness ranks to be the same value (i.e. 40) for easy visualisation and comparison purposes (see Figure 2). We also studied and analysed other population sizes with different numbers of distinct fitness ranks (100, 500 and 1000), and obtained similar results (so these results are not shown in the paper).

B. Simulation Results and Analysis

From [29], the probability of an event W_j that an individual $p \in S_j$ is selected from a tournament is:

$$P(W_j) = \frac{\left(\sum_{i=1}^j |S_i|\right)^k - \left(\sum_{i=1}^{j-1} |S_i|\right)^k}{|S_j|} \quad (10)$$

We calculate the total loss of individual diversity using Equation 3 in which $P(W_j)$ is replaced by Equation 10. We also split the total loss of individual diversity into two parts. One part is from the fraction of the population that is *not sampled* at all during the selection phase. We calculate it also using Equation 3 by replacing $1 - P(W_j)$ with $\left(\frac{N-1}{N}\right)^k$, which is the probability that an individual has not been sampled in a tournament of size k . The other part is from the fraction of population that is sampled but *not selected*. We calculate it by taking the difference between the total loss of individual diversity and the contribution from not-sampled individuals.

Figure 3 shows the three loss of individual diversity measures, namely the *total* loss of individual diversity and the contributions from *not-sampled* and *not-selected* individuals for standard tournament selection on the four populations with different FRDs. Overall there were no noticeable differences for the three loss of individual diversity measures on the four different populations with different FRDs between the two selection schemes. The loss of individual diversity measure in standard tournament selection depends almost entirely on the tournament size, and is almost independent of the FRD. This might be because that: 1) the loss of individual diversity measure is not an adequate measure; and/or 2) standard

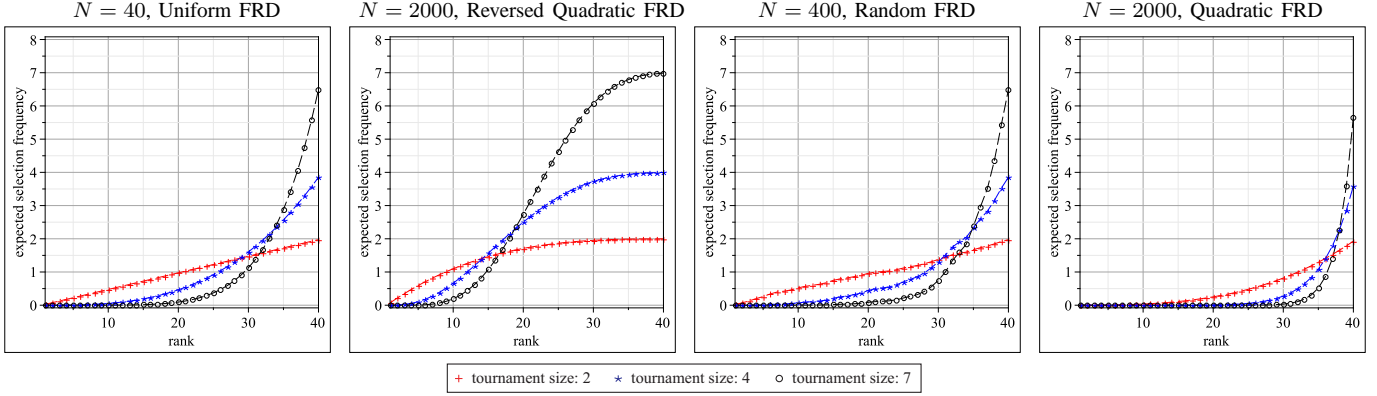


Fig. 4. Selection frequency in the standard tournament selection scheme on four populations with different FRDs.

tournament selection itself blinds the differences between these FRDs.

Based on Equation 4, we calculate the expected selection frequency of each individual in the selection phase for each of the four populations with different FRDs using the probability model of an individual being selected in a tournament (Equation 10). Figure 4 shows the selection frequency in standard tournament selection on the four populations with different FRDs. Instead of plotting the expected selection frequency for every individual, we plot it only for an individual in each of the 40 unique fitness ranks so that plots in different-sized populations have the same scale and it is easy to identify what fitness ranks may be lost. We chose three different tournament sizes (2, 4, and 7) commonly used in the literature to illustrate how tournament size affects the expected selection frequency.

From the figure, overall the standard tournament selection scheme favours better-ranked individuals for all tournament sizes, and the selection pressure is biased towards better individuals as the tournament size increases. Furthermore, skewed FRDs (reversed quadratic and quadratic) aggravate selection bias quite significantly. For the reversed quadratic FRD, there are more individuals of worse-ranked fitness that received selection preference, indicating that the search still wanders around without paying sufficient attention to the small number of outstanding individuals. Ideally, in this situation, a good selection mechanism should focus on the small number of good individuals to speed up evolution. For the quadratic FRD, the selection frequencies are strongly biased towards individuals with better ranks, indicating that the population diversity is quickly lost, the convergence speeds up, and the search may be confined to local optima. Ideally, in this situation, a good selection scheme should slow down the convergence. Unfortunately, the simulation results show that standard tournament selection does not know the dynamic requests thus is unable to change selection pressure accordingly to meet the expectations.

For the random FRD, we expect to see differences when comparing with the uniform FRD. However, the selection frequency shapes look very similar. Ideally, in this situation, a good selection mechanism should be able to adjust the selection pressure distinguishably according to the changes in the fitness rank distribution. For instance, it should give

a relatively higher selection preference to an individual in a fitness bag with a smaller size in order to increase the chance of propagating this genetic material and a relatively lower selection preference to an individual in another fitness bag with a larger size in order to reduce the chance of propagating the same or similar materials. However, only slight fluctuations and differences can be found in the random FRD under very close inspection. This implies that standard tournament selection may tolerate the difference between the uniform and random FRDs, and therefore sometimes take long time to converge. To interpret this finding, we provide the following analysis.

Assume μ is the average number of individuals for each S_j . In the uniform FRD, for all $j \in \{1, \dots, |S|\}$ where $|S|$ is the number of distinct fitness bags in the populations, $|S_j| = \mu$. While in the random FRD, it has

$$\frac{\sum_{i=1}^j |S_i|}{j} \approx \mu \quad (11)$$

and the approximation becomes more precise when j is close to $|S|$. As the selection frequency for an individual p of rank j is $N \times P(W_j)$, we simplify $P(W_j)$ for the uniform FRD as:

$$\begin{aligned} P(W_j) &= \frac{\left(\frac{j\mu}{|S|\mu}\right)^k - \left(\frac{(j-1)\mu}{|S|\mu}\right)^k}{\mu} \\ &= \frac{1}{\mu|S|^k} (j^k - (j-1)^k) \end{aligned} \quad (12)$$

For the random FRD, the $P(W_j)$ is:

$$\begin{aligned} P(W_j) &\approx \frac{\left(\frac{j\mu}{|S|\mu}\right)^k - \left(\frac{(j-1)\mu}{|S|\mu}\right)^k}{|S_j|} \\ &= \frac{1}{|S_j||S|^k} (j^k - (j-1)^k) \end{aligned} \quad (13)$$

From Equation 12, in the uniform FRD, the selection frequency for an individual of rank j will be just

$$\frac{1}{|S|^{k-1}} (j^k - (j-1)^k) \quad (14)$$

which is independent of the actual number of individuals of the same rank.

From Equation 13, the selection frequency of an individual of rank j in the random FRD is approximately:

$$\begin{aligned} & \frac{1}{|S_j||S|^k} (j^k - (j-1)^k) \times |S|\mu \\ &= \frac{\mu}{|S_j|} \times \frac{1}{|S|^{k-1}} (j^k - (j-1)^k) \end{aligned} \quad (15)$$

which differs from that (Equation 14) in the uniform FRD by a factor of $\frac{\mu}{|S_j|}$. For a random FRD, $\frac{\mu}{|S_j|}$ could be small. Therefore, only slight fluctuations and differences can be found in Figure 4 under very close inspection when comparing the random FRD with the uniform FRD.

We also calculate the selection probability distribution based on Equation 9. The simulations of the selection probability distribution are consistent with that of the selection frequency so the figures are omitted in this paper.

In summary, through the simulation analysis, this section showed that the standard tournament selection method is unaware of the evolution dynamics and is unable to apply different selection pressures at different learning stages accordingly to meet the expectations.

IV. A NEW APPROACH TO TUNING SELECTION PRESSURE

The standard tournament selection method has the high between-group selection pressure issue and cannot adjust selection bias in response to the FRD of a population. To address these issues, we need to modify standard tournament selection to become aware of the dynamics along evolution and to be able to adjust selection pressure accordingly. In generational EAs, the population at each generation can be seen as an abstract carrier holding the evolution dynamics. The number of fitness bags and the size of each bag, namely the FRD of a population, can reflect the dynamic evolutionary process, especially the degree of convergence of the population. Therefore, this section proposes an automatic selection pressure tuning strategy for tournament selection which uses the knowledge of the population FRD.

A. The approach

Figure 5 gives an overview of the proposed approach and shows the relationships between the major components: population clustering and clustering tournament selection. Other standard components are not detailed in the figure.

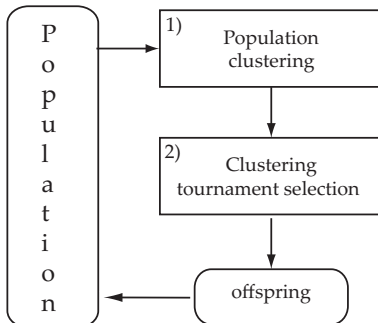


Fig. 5. Overview and relationship between the major components.

In the approach, the first component is population clustering. Populations can be partitioned into a set of clusters according to some criteria, including fitness values, individual structures and context, and each cluster is then assigned a distinct fitness value. In this paper, we cluster a population based on fitness values so that each fitness bag in a population becomes a cluster of the population.

The second component is a new tournament selection method called *clustering tournament selection*. Instead of sampling individuals as tournament candidates, the clusters are treated as the tournament candidates in the clustering tournament selection method: the best fitness cluster wins the tournament, and an individual in the cluster is randomly selected as a parent to participate in the recombination process. For a population S (of size N), which has been clustered into a set of $|S|$ clusters based on fitness values, the clustering tournament selection algorithm is as follows:

- 1: **for** $y = 1$ to N **do**
- 2: Sample k clusters from the $|S|$ clusters with replacement
- 3: Select the winning cluster from the tournament using fitness values
- 4: Return an individual randomly chosen from the winning cluster
- 5: **end for**

Therefore, this clustering tournament selection mechanism gets aware of the evolution dynamics by integrating the knowledge of the population FRD. We expect that the clustering tournament selection can automatically adjust selection pressure along evolution accordingly.

B. Modelling selection behaviour of the clustering tournament selection

Lemma 1. Let S_j be the cluster of individuals of rank j in the population. The probability of the event D that an individual $p \in S_j$ is sampled at least once in a tournament of size k is

$$P(D) = 1 - \left(1 - \frac{1}{|S||S_j|}\right)^k \quad (16)$$

Proof: In contrast to standard tournament selection, the sampling behaviour in clustering tournament selection is influenced by the number clusters and the actual size of a given cluster instead of the population size. It is clear that each cluster has the same probability $1/|S|$ to be sampled, and individuals in a cluster also have equal probability of being sampled, $1/|S_j|$. Therefore, the probability that p is sampled is $\frac{1}{|S||S_j|}$. The probability that p is never sampled into a tournament of size k is $\left(1 - \frac{1}{|S||S_j|}\right)^k$. Thus, we obtain Equation 16. ■

Lemma 2. Let S_j be the cluster of individuals of rank j in the population, the probability of the event E_j that an individual $p \in S_j$ is selected in a single tournament is

$$P(E_j) = \frac{(j)^k - (j-1)^k}{|S|^k \times |S_j|} \quad (17)$$

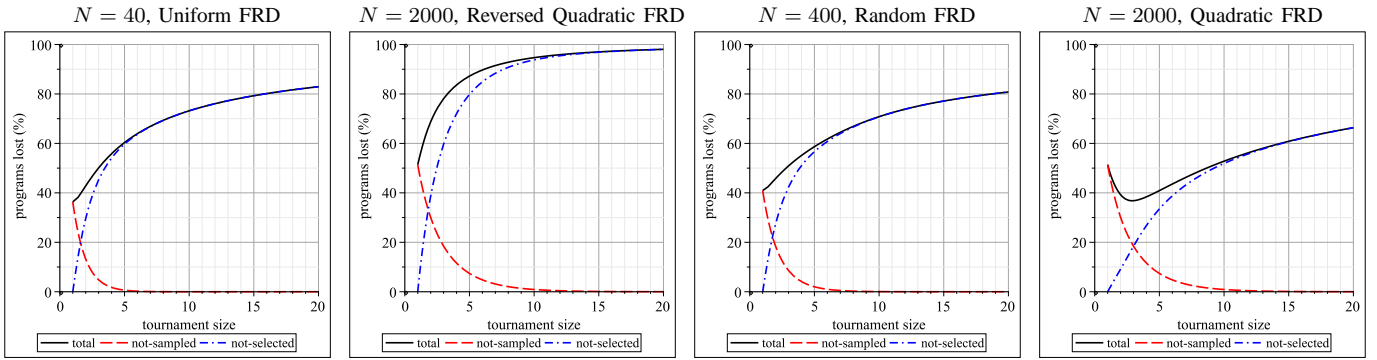


Fig. 6. Loss of individual diversity in the clustering tournament selection scheme on four different FRDs. Note that tournament size is discrete but the plots show curves to aid interpretation.

Proof: According to the algorithm, the number of tournament candidates is effectively reduced from the whole population size N to the number of clusters $|S|$. The probability that a cluster ranked j wins a tournament is simply:

$$\frac{(j)^k - (j-1)^k}{|S|^k} \quad (18)$$

Since all individuals in the winning cluster have the same probability to be chosen as a parent, we divide Equation (18) by the size of the j th cluster $|S_j|$ and obtain Equation (17). ■

V. ANALYSES OF CLUSTERING TOURNAMENT SELECTION VIA SIMULATIONS

We follow the method explained in Section III and used Equations 16 and 17 to calculate the three loss of individual diversity measures, the selection frequency and the selection probability distribution. The next subsections analyse the simulation results to investigate the clustering tournament selection in detail.

A. The loss of individual diversity analysis

Figure 6 illustrates the loss of individual diversity of the clustering tournament selection on four populations with different FRDs.

In the clustering tournament selection, for the uniform FRD, the three loss of individual diversity measures are identical to those of the standard tournament selection (see Figure 3). This is because each cluster contains the same number of individuals (in this case the number is one) so that the clustering tournament selection is effectively acting the same as standard tournament selection.

For the reversed quadratic FRD, the total loss of individual diversity is considerably higher compared with that of standard tournament selection and compared with those for other FRDs. We expect that the lost individuals are mainly the worse-ranked individuals. By ignoring most of the worse-ranked individuals at this stage, the search will be able to concentrate on the promising region so that the evolution will speed up to save unnecessary cost. The next subsection will verify the expectation when analysing the selection frequency.

For the quadratic FRD, the total loss of individual diversity in the clustering tournament selection is greater than that in the standard one when the tournament size is one, but is considerably lower for other tournament sizes. The reduction quickly reaches by about 20% ($60\% - 40\% = 20\%$) when the tournament size increases to five. Also we observed that when the tournament size is 3, the total loss of individual diversity becomes the lowest. The figure indicates that the individual diversity is maintained in a better manner than that in standard tournament selection. This is we expected for this type of FRD, as it may slow down the population convergence to avoid the confinement to local optima.

For the random FRD, there are some slight differences when comparing with that in the standard tournament selection. The total loss of individual diversity is about 5% higher than that in standard tournament selection for small tournament sizes 1 and 2, but gradually becomes lower for large tournament sizes.

Although the analysis of the loss of individual diversity showed a different selection behaviour in the clustering tournament selection, it only provided information at a coarse level and sometimes at a limited level, especially for the random FRD. Therefore, we need to further investigate the different selection behaviour in the clustering tournament selection using other measures.

B. The selection frequency and the selection probability distribution analyses

The simulation results of the selection frequency and the selection probability distribution are consistent again in the clustering tournament selection. Therefore, we only reported the selection frequency of the clustering tournament selection on the four populations with different FRDs (see Figure 7). Recall that the tournament size 3 provides the lowest total loss of individual diversity for the quadratic FRD, therefore in addition to the commonly three tournament sizes (2, 4, and 7), the tournament size 3 is added in this analysis and its impact is presented in a dash line in the figure.

The selection frequency trends on the uniform FRD in the clustering tournament selection are identical to those of the standard tournament selection for the reason given in Section V-A.

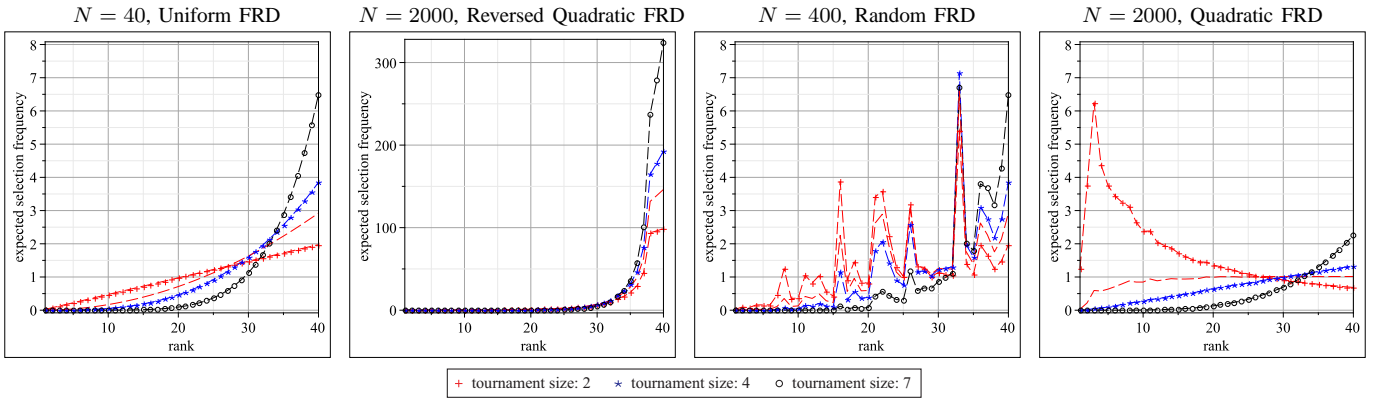


Fig. 7. Selection frequency of the clustering tournament selection scheme on four populations with different FRDs. Note that the extra dash line represents tournament size 3.

The other three FRDs reveal significant differences when compared with the standard tournament selection (see Figure 4).

For the reversed quadratic FRD (representing the early stage of evolution), most of the low fitness ranks have very low selection frequencies so that they are effectively discarded. This observation supports our expectation in the analysis of the loss of individual diversity in the previous sub section that the lost individuals are mainly those worse-ranked ones, and meets the desiderata of a good selection scheme that can pay attention to small number of outstanding individuals to speed up evolution other than wandering around in unpromising regions.

For the random FRD (representing the middle stage of evolution), the selection frequency trends are very ragged instead of the smooth trends we usually saw in standard tournament selections. There is some interesting selection behaviour here. For instance, for the tournament size 4, the expected selection frequency for an individual of rank 33 is above 7, while the expected selection frequencies for individuals of better ranks are much lower; even one of the best-ranked individuals in the population is below 4. From Figure 2, we can see that $|S_{33}|$ is only 3 while $|S_j|_{j>33}$ are much higher. The results show that apart from being governed by the tournament size, the clustering tournament selection is aware of the random changes in the FRD and can adjust the selection pressure automatically. It gives a relatively high selection preference to an individual in a fitness bag with a smaller size to increase the chance of propagating its genetic material. It then gives relatively low selection preferences to other better individuals in fitness bags with larger sizes to restrict their propagation. This kind of selection behaviour is unique to the clustering tournament selection and appears to again meet the desiderata expectation of a good selection scheme that can adjust the selection pressure distinguishably.

For the quadratic FRD (representing a converged stage of evolution), the clustering tournament selection significantly reduces the selection frequency of better-ranked individuals, while increasing the frequency of middle-ranked individuals. Therefore, the clustering tournament selection can reduce the chance that groups of better-ranked individuals dominate the

next generation and it is able to maintain the population diversity better than the standard one.

Note that for the quadratic FRD, tournament size 2 resulted in a strong bias to worse-ranked individuals, especially the third-ranked ones²; this may be undesirable. On the other hand, tournament size 3 provided almost even selection frequencies on all fitness ranks. This observation may explain why tournament size 3 provided the lowest total loss of individual diversity.

In summary, the analysis results showed that in addition to the usual selection preference for better individuals governed by tournament size, the clustering tournament selection tends to give additional selection preference to individuals in small sized clusters. Furthermore, when most individuals of the population are of worse fitness ranks and evolution encounters a danger of missing good individuals, it tends to increase selection bias to better individuals, hoping to quickly drive the population to promising regions. When the population tends to converge to local optima and evolution encounters a danger of losing genetic material, it tends to decrease selection bias to better individuals, hoping to keep the population diverse. Therefore, the clustering tournament selection is an selection mechanism that can automatically adjust the selection pressure for the dynamic evolutionary process.

VI. ANALYSES OF CLUSTERING TOURNAMENT SELECTION VIA EXPERIMENTS

The simulations above suggest that the clustering tournament selection can be aware of the dynamics in evolution, and adjust the selection pressure accordingly. This section further analyses the effectiveness of the clustering tournament selection empirically through experiments. The instance of EAs used in the experiments is the commonly used tree-based generational GP [32].

A. Data Sets

The experiments involve three different problem domains with different difficulties: an Even- n -Parity problem (EvePar),

²This is because the ranks 1 to 3 have the same smallest number of individuals (Figure 2).

a Symbolic Regression problem (SymReg), and a Binary Classification problem (BinCla). We chose these three type of problems in particular because they have received considerable attention as examples in the literature of GP.

1) *EvePar*: An even- n -parity problem has an input of a string of n Boolean values. It outputs *true* if there are an even number of true's, and otherwise *false*. The most characteristic aspect of this problem is the requirement to use all inputs in an optimal solution and a random solution could lead to a score of 50% accuracy [33]. Furthermore, optimal solutions could be dense in the search space as an optimal solution generally does not require a specific order of the n inputs presented. EvePar considers the case of $n = 6$. Therefore, there are 2^6 combinations of unique 6-bit length strings as fitness cases.

2) *SymReg*: SymReg is shown in Equation 19 and visualised in Figure 8. We generated 100 fitness cases by choosing 100 values for x from $[-5,5]$ with equal steps.

$$f(x) = \exp(1 - x) \times \sin(2\pi x) + 50\sin(x) \quad (19)$$

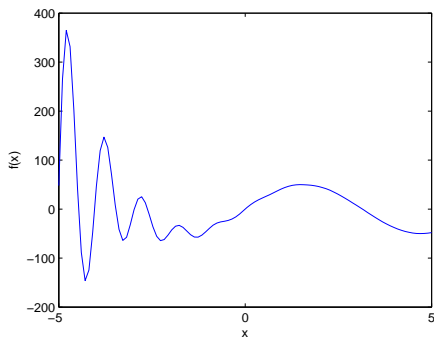


Fig. 8. The symbolic regression problem.

3) *BinCla*: BinCla involves determining whether examples represent a *malignant* or a *benign* breast cancer. The dataset is the Wisconsin Diagnostic Breast Cancer dataset chosen from the UCI Machine Learning repository [34]. BinCla consists of 569 data examples, where 357 are benign and 212 are malignant. It has 10 numeric measures (see Table I) computed from a digitised image of a fine needle aspirate of a breast mass and are designed to describe characteristics of the cell nuclei present in the image. The mean, standard error, and “worst” of these measures are computed, resulting in 30 features [34]. The whole original data set is split randomly and equally into a training data set, a validation data set, and a test data set with class labellings being evenly distributed across the three data sets for each individual GP run.

B. Function sets and terminal sets

The function set used for EvePar consists of the standard Boolean operators $\{and, or, not\}$ and *if* function. The *if* function takes three arguments and returns its second argument if the first argument is *true*, and otherwise returns its third argument. In order to increase the problem difficulty, we do not include the *xor* function in the function set.

The function set used for SymReg includes the standard arithmetic binary operators $\{+, -, *, /\}$ and unary operators $\{abs, sin, exp\}$. The $/$ function returns zero if it is given invalid arguments.

The function set used for BinCla includes the standard arithmetic binary operators $\{+, -, *, /\}$. We hypothesised that convergence might be quicker if using only the four arithmetic operators, and more functions might lead to better results. Therefore, the function set also includes unary operators $\{abs, sqrt, sin\}$ and *if* function. The *sqrt* function automatically converts a negative argument to a positive one before operating on it. The *if* function takes three arguments and returns its second argument if the first argument is positive, and returns its third argument otherwise. The *if* function allows a program to contain a different expression in different regions of the feature space, and allows discontinuous programs, rather than insisting on smooth functions.

The terminal set for EvePar consists of n Boolean variables. The terminal set for SymReg and BinCla includes a single variable x and 30 terminals, respectively. Real valued constants in the range $[-5.0, 5.0]$ are also included in the terminal sets for SymReg and BinCla. The probability mass assigned to the whole range of constants when constructing programs is set to 5%.

TABLE I
TEN FEATURES IN THE DATASET OF BINCLA

a	radius	f	compactness
b	texture	g	concavity
c	perimeter	h	concave points
d	area	i	symmetry
e	smoothness	j	fractal dimension

C. Fitness function

For even- n -parity problems, the standard fitness function counts the number of wrong outputs (misses) for the 2^n combinations of n -bit strings and treats zero misses as the best raw fitness [32]. There is an issue with this fitness function: the worst program according to this fitness function is the one that has 2^n misses. However, this program actually captures most of the structure of the problem and can be easily converted to a program of zero misses by adding a *not* function node to the root of the program. Therefore, programs with a very large number of misses are, in a sense, just as good as programs with very few misses.

In this paper, we used a new fitness function for EvePar:

$$fitness = \begin{cases} m & , \text{ if } m < 2^{n-1} \\ 2^n - m & , \text{ otherwise} \end{cases} \quad (20)$$

where m is the number of misses.

The fitness function in SymReg is the root-mean-square (RMS) error of the outputs of a program relative to the expected outputs. Because neither class is weighted over the other, the fitness function for BinCla is the classification error rate on the training data set (the fraction of fitness cases that are incorrectly classified by a program as a proportion of the total number of fitness cases in the training data set). A

program classifies the fitness case as *benign* if the output of the program is positive, and *malignant* otherwise. Note that class imbalance design in fitness function for BinCla is beyond the scope of this paper. All three problems have an ideal fitness of zero.

D. Parameter setting and configurations

The genetic parameters are the same for all three problems. The ramped half-and-half method is used to create new programs and the maximum depth of creation is four (counted from zero). To prevent code bloat, the maximum size of a program is set to 50 nodes during evolution based on some initial experimental results. The standard subtree crossover and mutation operators are used [32]. The crossover rate, the mutation rate, and the copy rate are 85%, 10% and 5% respectively. The best individual in the current generation is explicitly copied into the next generation, ensuring that the population does not lose its previous best solution³. The population size is 500. A run is terminated when the number of generations reaches the pre-defined maximum of 101 (including the initial generation), or the problem has been solved (there is a program with a fitness of zero on the training data set), or the error rate on the validation set starts increasing (for BinCla). Four tournament sizes 2, 3, 4, and 7 are used.

We ran experiments comparing two GP systems using the standard and the clustering tournament selection methods respectively for each of the three problems. In each experiment, we repeated the whole evolutionary process 500 times independently. In each pair of the 500 runs, an initial population is generated randomly and is provided to both GP systems in order to reduce the performance variance caused by different initial populations.

E. Impact on population diversity analysis

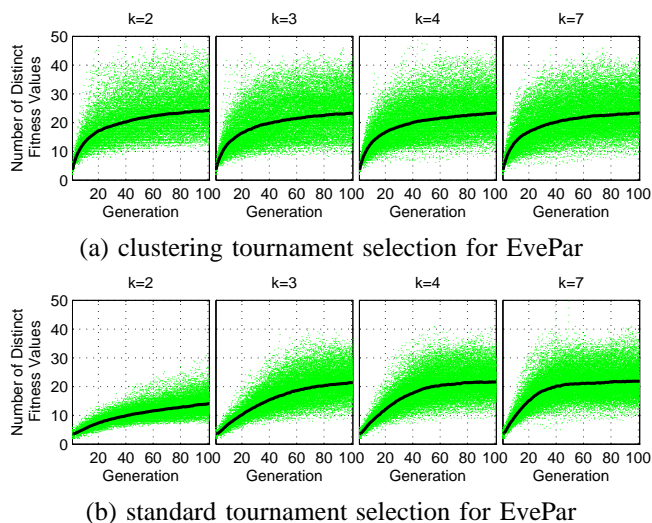


Fig. 9. Comparison of population diversity maintenance between the clustering tournament selection and the standard tournament selection for EvePar for four tournament sizes.

³This is referred to as elitism [35].

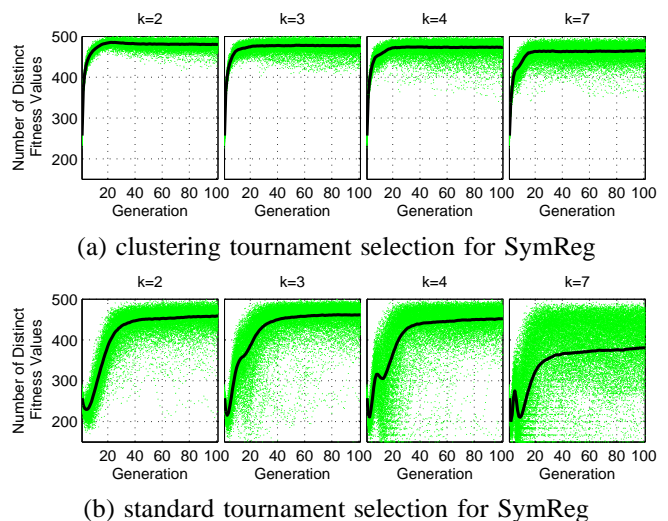


Fig. 10. Comparison of population diversity maintenance between the clustering tournament selection and the standard tournament selection for SymReg for four tournament sizes.

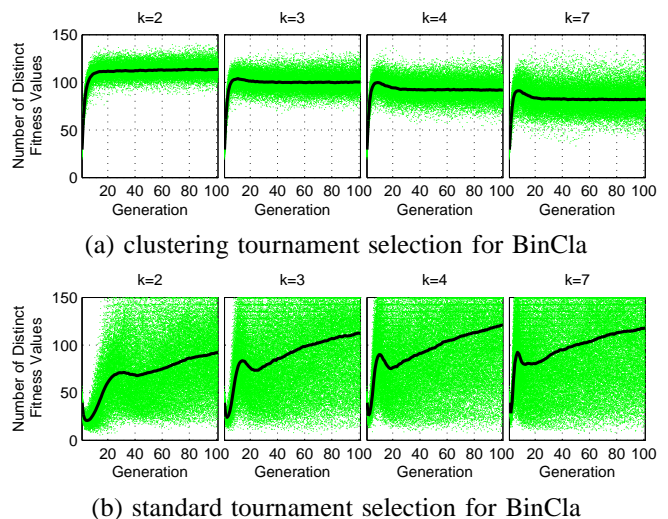


Fig. 11. Comparison of population diversity maintenance between the clustering tournament selection and the standard tournament selection for BinCla for four tournament sizes.

While it is possible to visualise the loss of individual diversity, the selection frequency, and the selection probability distribution of an arbitrary population in a run, it is very likely that the FRD at a generation in an experiment run using one selection method differs from that using another selection method. Therefore, comparing these measures on populations in different runs using different selection schemes could not provide much useful information. As a result, we used another measure — the number of distinct fitness values — in the experimental result analysis.

Figures 9, 10 and 11 compare the clustering tournament selection and the standard tournament selection in terms of population diversity measured by the number of distinct fitness values generation by generation using each of the four tournament sizes for EvePar, SymReg, and BinCla, respectively. The dark line in each chart represents the mean value over the 500 runs.

It is clear that the clustering tournament selection can quickly increase the population diversity to a certain level and maintain it stably. The four different tournament sizes have only small impact on the population diversity: for EvePar the four trends of the average numbers of distinct fitness values are almost identical, and for SymReg and BinCla there are only slight drops when the tournament size increases.

In contrast, standard tournament selection performs differently, especially for SymReg and BinCla (chart (b) in Figures 10 and 11). The population diversity fluctuates along evolution and has a larger variation in the 500 runs. It is also sensitive to tournament size. This comparison demonstrates the advantage of the clustering tournament selection in maintaining population diversity in terms of the number of distinct fitness values.

F. Overall GP performance analysis

Table II compares the performances of GP systems using the standard and the clustering tournament selections. The measure for EvePar is the average number of misses over the 500 runs. The measures for SymReg and BinCla are the averages of the RMS error and the classification error rate on test data over the 500 runs respectively. Therefore, the smaller the value, the better the performance. Note that the standard deviation is shown after the \pm sign.

TABLE II
PERFORMANCE COMPARISON BETWEEN THE CLUSTERING AND THE STANDARD TOURNAMENT SELECTION SCHEMES.

Tournament Selection		EvePar	SymReg	BinCla
Scheme	Size	Miss	RMS Error	Test Error (%)
clustering	2	14.2 \pm 6.5	47.6 \pm 5.9	7.4 \pm 2.3
	3	13.2 \pm 7.1	39.7 \pm 7.6	7.5 \pm 2.3
	4	13.1 \pm 6.9	36.8 \pm 7.9	7.7 \pm 2.5
	7	12.9 \pm 6.9	33.5 \pm 8.3	7.9 \pm 2.5
standard	2	20.7 \pm 3.8	48.2 \pm 5.2	9.2 \pm 2.9
	3	13.1 \pm 6.7	39.9 \pm 6.6	8.7 \pm 2.7
	4	11.4 \pm 7.2	37.6 \pm 8.3	8.7 \pm 2.7
	7	10.6 \pm 6.6	40.9 \pm 11.3	8.7 \pm 2.7

The results suggest that the GP system using the clustering tournament selection generally has advantages over the GP system using standard tournament selection. In order to provide statistically sound comparison results for the advantage of the clustering tournament selection, we calculated the confidence intervals at 99% levels (two-sided) for their differences in misses, in RMS errors, and in error rates for EvePar, SymReg and BinCla respectively.

We firstly calculated the difference of the measures between a pair of runs using the same initial population for each of the 500 pairs of runs, then used the formula

$$\bar{x} \pm Z \frac{s}{\sqrt{500}} \quad (21)$$

to calculate the confidence interval, where \bar{x} is the average difference over 500 values, s is the standard deviation, and Z is 2.58 for 99% confidence. If zero is not included in the confidence interval, then the difference is statistically significant. Since the smaller the measure, the better the performance, the clustering tournament selection is significantly better than the standard one when the confidence interval is less than zero.

Table III shows the results with significant differences (either better or worse) shown in bold.

TABLE III
CONFIDENCE INTERVALS AT 99% LEVEL FOR THE DIFFERENCES BETWEEN THE CLUSTERING AND THE STANDARD TOURNAMENT SELECTION SCHEMES.

Tournament size	EvePar	SymReg	BinCla
2	(-7.2, -5.7)	(-1.5, 0.3)	(-2.1, -1.3)
3	(-0.9, 1.1)	(-1.3, 0.9)	(-1.6, -0.8)
4	(0.6, 2.6)	(-2.0, 0.5)	(-1.4, -0.7)
7	(1.5, 3.3)	(-8.9, -5.9)	(-1.3, -0.5)

For BinCla (the hardest problem), the clustering tournament selection is consistently significantly better than the standard one for all four tournament sizes.

For SymReg, the clustering tournament selection is slightly better than the standard one using tournament sizes 2, 3, and 4, but significantly better for tournament size 7. A large tournament size represents a strong selection bias towards better individuals and therefore there is a great potential for losing diversity. The clustering tournament selection appears to be able to counteract this potential effectively.

For EvePar (the simplest problem), when the tournament size is 2, the clustering tournament selection is significantly better than the standard tournament selection. However, when the tournament size is 4 or 7, it is significantly worse than the standard tournament selection.

The performance reported here shows that when the selection pressure is adjusted according to the dynamics in evolution and the population diversity is well maintained by the clustering tournament selection, the overall GP search performance is improved in most problems, but not every case. Possible explanations for the exceptions include:

- Easy problems can be solved easily using high selection pressure so that it is not necessary to adjust the selection pressure.
- The way to cluster population (by fitness values in this case) may not work well for EvePar or possibly other Boolean problems because of the limited number of possible distinct fitness values.
- Although good parents may be selected, the probability of finding better offspring in a large offspring space is small so that the advantage of the clustering tournament selection cannot be properly illustrated.

Although in theory tournament size 3 was shown to have the lowest total loss of individual diversity for the quadratic FRD in the clustering tournament selection, the experimental results did not show that tournament size 3 is significantly better than others. This might be because the quadratic FRD will not appear if the clustering tournament selection is applied from the beginning of a GP search.

VII. CONCLUSIONS AND FUTURE WORK

Due to the dynamics in EAs, it is necessary to have a selection mechanism that can tune selection pressure along evolution. This paper used the loss of individual diversity, the selection frequency, and the selection probability distribution

on four simulated populations with different FRDs to analyse the selection behaviour in standard tournament selection. It showed that the standard tournament selection method is unaware of the dynamics in evolution, thus is unable to tune selection pressure accordingly during evolution.

This paper then presented a novel approach to the selection pressure tuning problem by integrating the population FRDs into tournament selection. Through both simulated and empirical analyses, this paper showed that the approach can tune selection pressure automatically and dynamically along evolution.

The simulations and experimental analyses in this paper provided additional insight into the control of the selection pressure in tournament selection and the outcomes are as follows:

- The high between-group selection pressure issue has a strong interaction with the FRD of a population. FRDs change generation by generation and can be seen as the analogue of the dynamics in evolution. Using the knowledge of FRD is a promising way to modify the standard tournament selection in order to tune the selection pressure dynamically and automatically.
- The clustering tournament selection is an automatic selection pressure tuning strategy and is worth further investigation. It can significantly improve GP search performance for relatively difficult problems, although may not be required for easy problems (i.e. EvePar).
- There are likely to be other, more effective population clustering methods other than merely using the fitness values, for instance by genotype, which may be suitable for solving EvePar [36]. Nonetheless, in light of the results presented in Sections V, VI-E, and VI-F, we hope that researchers will be encouraged to experiment with the simple population clustering method in the initial stages of the development of their alternative parent selection algorithms.

This paper investigated the research questions in the context of generational EAs under some assumptions. The empirical study was conducted only on tree-based generational GP systems. In the future, we will extend the empirical study to other generational EA paradigms, as well as investigate the selection pressure tuning problem for non-generational EAs.

ACKNOWLEDGEMENT

We would like to thank Dr Peter Andreae for his comments, discussions and suggestions for the early versions. This work was also supported in part by the Marsden Fund council from the government funding (VUW0806), administered by the Royal Society of New Zealand, and the Faculty Research Grant (FRG08-2367/88167) at Victoria University of Wellington.

REFERENCES

- [1] D. E. Goldberg and K. Deb, "A comparative analysis of selection schemes used in genetic algorithms," *Foundations of Genetic Algorithms*, pp. 69–93, 1991.
- [2] T. Blickle and L. Thiele, "A mathematical analysis of tournament selection," in *Proceedings of the Sixth International Conference on Genetic Algorithms*, 1995, pp. 9–16.
- [3] B. L. Miller and D. E. Goldberg, "Genetic algorithms, tournament selection, and the effects of noise," University of Illinois at Urbana-Champaign, Tech. Rep. 95006, July 1995.
- [4] T. Motoki, "Calculating the expected loss of diversity of selection schemes," *Evolutionary Computation*, vol. 10, no. 4, pp. 397–422, 2002.
- [5] M. Affenzeller, S. Wagner, and S. Winkler, "GA-selection revisited from an ES-driven point of view," in *Artificial Intelligence and Knowledge Engineering Applications: A Bioinspired Approach*, ser. Lecture Notes in Computer Science. Springer, 2005, vol. 3562, pp. 262–271.
- [6] H. Xie, M. Zhang, and P. Andreae, "An analysis of constructive crossover and selection pressure in genetic programming," in *Proceedings of Genetic and Evolutionary Computation Conference*, 2007, pp. 1739–1746.
- [7] S. Winkler, M. Affenzeller, and S. Wagner, "Offspring selection and its effects on genetic propagation in genetic programming based system identification," *Cybernetics and Systems*, vol. 2, pp. 549–554, 2008.
- [8] R. Rubinstein and D. Kroese, *Simulation and the Monte Carlo Method*, 2nd ed. John Wiley and Sons, 2007.
- [9] K. D. Jong, "Parameter setting in eas: a 30 year perspective," in *Parameter Setting in Evolutionary Algorithms*. Springer, 2007, pp. 1–18.
- [10] J. H. Holland, *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press, 1975.
- [11] J. J. Grefenstette and J. E. Baker, "How genetic algorithms work: A critical look at implicit parallelism," in *Proceedings of the 3rd International Conference on Genetic Algorithms*, J. D. Schaffer, Ed. Morgan Kaufmann Publishers, 1989, pp. 20–27.
- [12] A. Brindle, "Genetic algorithms for function optimisation," Ph.D. dissertation, Department of Computing Science, University of Alberta, 1981.
- [13] T. Stewart, "Extrema selection: accelerated evolution on neutral networks," in *Proceedings of the 2001 IEEE Congress on Evolutionary Computation*. IEEE Press, 2001, pp. 25–29.
- [14] R. Poli and W. B. Langdon, "Backward-chaining genetic programming," in *Proceedings of Genetic and Evolutionary Computation Conference*, H.-G. Beyer and et al, Eds., vol. 2. Washington DC, USA: ACM Press, 25-29 Jun. 2005, pp. 1777–1778.
- [15] B. L. Miller and D. E. Goldberg, "Genetic algorithms, selection schemes, and the varying effects of noise," *Evolutionary Computation*, vol. 4, no. 2, pp. 113–131, 1996.
- [16] K. Matsui, "New selection method to improve the population diversity in genetic algorithms," in *Proceedings of 1999 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 1999, pp. 625–630.
- [17] W. B. Langdon and R. Poli, *Foundations of Genetic Programming*. Berlin: Springer, 2002.
- [18] A. Sokolov and D. Whitley, "Unbiased tournament selection," in *Proceedings of Genetic and Evolutionary Computation Conference*. ACM Press, 2005, pp. 1131–1138.
- [19] F. G. Lobo, C. F. Lima, and Z. Michalewicz, Eds., *Parameter Setting in Evolutionary Algorithms*, ser. Studies in Computational Intelligence. Springer, 2007, vol. 54.
- [20] B. A. Julstrom and D. H. Robinson, "Simulating exponential normalization with weighted k-tournaments," in *Proceedings of the 2000 IEEE Congress on Evolutionary Computation*. IEEE Press, 2000, pp. 227–231.
- [21] K. Hingee and M. Hutter, "Equivalence of probabilistic tournament and polynomial ranking selection," in *Proceedings of IEEE Congress on Evolutionary Computation*, 2008, pp. 564–571.
- [22] R. Huber and T. Schell, "Mixed size tournament selection," *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, vol. 6, pp. 449–455, 2002.
- [23] V. Filipović, J. Kratica, D. Tošić, and I. Ljubić, "Fine grained tournament selection for the simple plant location problem," in *5th Online World Conference on Soft Computing Methods in Industrial Applications*, 2000, pp. 152–158.
- [24] M. Bulmer, *The Mathematical Theory of Quantitative Genetics*. Oxford, UK: Oxford University Press, 1980.
- [25] H. Muhlenbein and D. Schlierkamp-Voosen, "Predictive models for the breeder genetic algorithm, I: continuous parameter optimization," *Evolutionary Computation*, vol. 1, no. 1, pp. 25–49, 1993.
- [26] T. Blickle and L. Thiele, "A comparison of selection schemes used in evolutionary algorithms," *Evolutionary Computation*, vol. 4, no. 4, pp. 361–394, 1997.
- [27] E. Popovici and K. D. Jong, "Understanding EA dynamics via population fitness distributions," in *Proceedings of the Genetic and Evolutionary Computation Conference 2003*, 2003, pp. 1604–1605.

- [28] J. Branke, H. C. Andersen, and H. Schmeck, "Global selection methods for SIMD computers," in *Proceedings of the AISB96 Workshop on Evolutionary Computing*, 1996, pp. 6–17.
- [29] H. Xie, M. Zhang, and P. Andreae, "Another investigation on tournament selection: modelling and visualisation," in *Proceedings of Genetic and Evolutionary Computation Conference*, 2007, pp. 1468–1475.
- [30] T. Bäck, "Selective pressure in evolutionary algorithms: A characterization of selection mechanisms," in *Proceedings of the First IEEE Conference on Evolutionary Computation.*, 1994, pp. 57–62.
- [31] R. Poli and W. B. Langdon, "Backward-chaining evolutionary algorithms," *Artificial Intelligence*, vol. 170, no. 11, pp. 953–982, 2006.
- [32] J. R. Koza, *Genetic Programming — On the Programming of Computers by Means of Natural Selection*. Cambridge: MIT Press, 1992.
- [33] S. M. Gustafson, "An analysis of diversity in genetic programming," Ph.D. dissertation, University of Nottingham, 2004.
- [34] D. Newman, S. Hettich, C. Blake, and C. Merz, "UCI repository of machine learning databases," 1998. [Online]. Available: <http://archive.ics.uci.edu/ml>
- [35] R. Poli, N. F. McPhee, and L. Vanneschi, "Elitism reduces bloat in genetic programming," in *Proceedings of the 10th annual conference on Genetic and evolutionary computation*. ACM Press, 2008, pp. 1343–1344.
- [36] H. Xie, M. Zhang, and P. Andreae, "Automatic selection pressure control in genetic programming," in *Proceedings of the sixth International conference on Intelligent Systems Design and Applications*. IEEE Computer Society Press, 2006, pp. 435–440.