

Research Article

Viewing the Problem from Different Angles: A New Diversity Measure Based on Angular Distances

Henrik Berg^{1,2}

¹ Østfold University College, Faculty of Computer Sciences, 1757 Halden, Norway

² Norwegian Defense Research Establishment (FFI), Maritime Systems Division, 3191 Horten, Norway

Correspondence should be addressed to Henrik Berg, henrikbe@ifi.uio.no

Received 19 May 2009; Accepted 1 December 2009

Recommended by Natalio Krasnogor

It is commonly believed that diversity is crucial for an evolutionary system to succeed, especially when the problem to be solved contains local optima from which the population cannot easily escape. There exist numerous methods to measure population diversity, but none of these have been shown to be consistently useful. In this paper, a new diversity measure is introduced, and it is shown that high diversity according to this new measure generally leads to a more successful overall evolution in most of the cases considered.

Copyright © 2009 Henrik Berg. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

The concept of diversity has been studied extensively in the evolutionary computation literature [1–5]. It is generally believed that diversity is beneficial in evolutionary algorithms, in that some form of diversity is needed to efficiently explore the given search space. There are many different ways to measure diversity, the most common ones being various forms of *genotypic diversity* and *fitness-based diversity*. Genotypic diversity is a measure of the amount of different genetic material available. It is fairly obvious that an evolution cannot efficiently proceed unless the individuals in the population contain different genes that can be recombined to form new individuals. However, in many cases, the mapping from genotype to phenotype, or behavior, is a many-to-one-mapping, in that several different gene combinations may encode the same solution to the given problem. Therefore, it has been claimed that measuring diversity based on phenotypes rather than genotypes may be more useful [2]. The most common approach is to measure diversity based on fitness values, for example by counting the number of different fitness values found in a population [6] or computing some sort of fitness entropy [7].

When applying evolutionary algorithms to machine learning problems [8], another possibility is to measure diversity based on the subsets of training examples solved

by the different individuals in the population [5, 9]. This approach has received far less attention in the literature than the diversity measures mentioned above. This paper investigates the usefulness of this approach to measuring diversity, by analyzing the correlations between fitness and various forms of diversity in a number of regression problems. Additionally, a new training example-based diversity measure, *Angular Phenotype Diversity*, is introduced and investigated experimentally. Experiments conducted show that this new diversity measure is potentially very useful in the cases considered.

The rest of the paper is organized as follows. In Section 2, the concepts of premature convergence and various diversity measures are reviewed. Section 3 reviews the use of training examples to measure diversity, and introduces a new diversity measure. Sections 4 and 5 describe our experimental investigation of our new diversity measure. Finally, conclusions and directions for further research are given in Section 6.

2. Premature Convergence, Local Optima, and Diversity

One of the key features of all evolutionary systems is the maintaining of some kind of population. A major advantage of maintaining a population of potential solutions is that

compared to simple local search methods, an evolutionary system can follow multiple paths through the search landscape and pursue solutions in multiple directions. This makes evolutionary systems remarkably robust, even when applied to multimodal and noisy optimization problems [10–12].

However, in many problems, known as deceptive multimodal problems, there exist several rather good but suboptimal solutions from which it is difficult to develop better solutions. If such a solution is found, and if at that point of evolution no better solutions are known, selection pressure will increase the number of individuals representing this solution at each generation. At some point, this suboptimal solution may take over the entire population, leading to what is known as premature convergence. Once a population has prematurely converged onto a suboptimal solution, the evolutionary system will perform no better than ordinary local search, and given a deceptive problem, it is unlikely that any further improvement can be found.

2.1. Avoiding Premature Convergence. The most important factors determining the probability of premature convergence include the following.

- (i) *The fitness landscape of the problem* [13]. If the fitness landscape contains local optima, premature convergence may occur. The probability of premature convergence depends among other things on the number of local optima, their positions, and the topology of the surrounding landscape.
- (ii) *The selection pressure of the algorithm employed.* Selection schemes with high-selection pressure drive the population into local optima faster [14, 15]. Therefore, employing selection schemes with low-selection pressure (like traditional proportional selection [16] or (μ, λ) -selection with λ not much higher than μ [17]) may reduce the risk of premature convergence. However, low-selection pressure also leads to slower overall evolution.
- (iii) *The population size.* Larger populations obviously converge slower to a local optimum, giving the algorithm more time to find better solutions. However, larger populations also require more computational resources. Choosing the best population size to use for a given problem is a very difficult task, which has received much attention in the literature [18–21].
- (iv) *The population diversity.* If the entire population converges into a single suboptimal solution, the diversity of the population is obviously zero. Striving for diverse populations may reduce the risk of premature convergence [4].

The first of these factors is usually an inherent property of the problem at hand, and can therefore not be addressed by the user of the evolutionary system.

The other three can be influenced by the user in various ways. However, since increased population size and decreased selection pressure both lead to slower evolution,

maintaining the diversity of the population is usually regarded as the most important technique available to avoid premature convergence.

2.2. Diversity. There is no single and universal definition of the concept of diversity within the field of evolutionary computation, rather a number of different diversity measures have been proposed. These can be divided into two distinct classes.

- (i) *Genotypical* or structural diversity, measuring the syntactical differences between the individuals in the population. Examples include counting the number of syntactically distinct individuals [22], computing the Euclidean or Hamming distance between the genotypes of the individuals [3, 23], and computing the Edit distances between the individuals [24], that is, counting the number of editing steps required to go from one individual to another.
- (ii) *Phenotypical* or behavioral diversity, measuring the semantical differences between the individuals in the population. In situations such as Genetic Programming [22], with a many-to-one mapping between genotypes and phenotypes, Burke et al. [2] suggest that phenotypical diversity is more useful than genotypical diversity. Phenotypical diversity usually implies genotypical diversity, but the opposite is not necessarily true, since many genetically distinct individuals can all encode the same behavior in different ways. Another advantage of phenotypical diversity is that it is more general, since it does not depend on the particular encoding scheme used. For a given problem, a given phenotypical diversity measure can be used regardless of whether the individuals are encoded as bit-strings, vectors of real values, syntax trees, or some other representation.

Examples of phenotypical diversity measures include counting the number of distinct fitness values [6], computing the entropy of the set of fitness values [7], and computing the average distance between the individuals in the population using some distance metrics, for example, absolute distance between fitness values [25].

For a detailed description of these and other diversity measures and a thorough analysis of how they affect fitness during evolution, see Burke et al. [2, 26].

3. Diversity Based on Training Examples

Within the fields of Ecology and Evolutionary Biology, it is generally agreed that the more diverse an ecosystem, the higher its chances of survival through various environmental changes, and the better it will be able to evolve to adapt to new environments [27]. Diversity within these fields is usually taken as the varying abilities of the species to perform different tasks, like finding different kinds of food, utilizing different kinds of shelters, or resisting different kinds of diseases. In nature, this form of diversity is a

direct consequence of natural selection and environmental limitations. For example, since the supply of each kind of food will always be limited, species specializing in one or a few kinds of food will usually be able to eat more of these kinds of food than more general species. Consequently, different species usually evolve to find and exploit different kinds of food. Within the field of Ecology, this process is commonly known as niching.

One can define a similar diversity measure for evolutionary computation systems where fitness is calculated based on a set of training examples. Examples of such systems include Genetic Programming [22], the ADATE system for automatic programming [28], Genetic Algorithms-based Classifier Systems [29], and Evolution Strategies [17] used for solving regression problems [30], for example, evolution of Neural Networks [31]. If each training example is regarded as a kind of food, and each individual in the population has a certain ability to eat each kind of food, diversity can be defined in the same way as the ecological diversity described above, that is, as the variety of the abilities of the individuals to solve different training examples. More formally, one can define the *Euclidean Phenotype Distance* between two individuals i and j as the Euclidean distance between their fitness vectors:

$$\text{dist}_{\text{EP}}(i, j) = \|F(i) - F(j)\| = \sqrt{\sum_{c=1}^N (F(i)_c - F(j)_c)^2}, \quad (1)$$

where $F(k) = [F(k)_1, \dots, F(k)_N]$ is the fitness vector of the individual k , that is, the vector of fitness values achieved by the individual when solving each of the N training examples. In some cases, this fitness is a Boolean value (0 or 1), in these cases the Euclidean Phenotype Distance between two individuals would be equivalent to the square root of the Hamming distance between their fitness vectors. Generally, an evolutionary system might allow fitness values of any ordered type [32]; however, in this paper we restrict ourselves to real fitness values.

The Euclidean Phenotype Distance can be used to define the diversity of a population by computing the average distance between all pairs of individuals, we call this diversity measure *Euclidean Phenotype Diversity*:

$$\text{EPD}(P) = \frac{1}{\binom{|P|}{2}} \sum_{i, j \in P, i \neq j} \text{dist}_{\text{EP}}(i, j). \quad (2)$$

Burke et al. [2] do not consider this kind of diversity in their investigation, and we have only been able to find a few references discussing it. Gustafson et al. [33] investigate the number of unique behaviors in an evolving population. McQuesten [9] introduces a diversity measure based on distances between behavior vectors, but does not show or discuss its usability. Curran and O’Riordan [5] compute diversity using the Euclidean Phenotype Distance measure as defined in (1) above. None of these references correlate the measured diversities with the fitness values achieved by the evolving populations. Zenobi and Cunningham [34] study the effect of Euclidean Phenotype Diversity in ensembles of

classifiers [35], and conclude that ensembles with high diversity perform better than less diverse ensembles. However, their experiments use hill climbing rather than evolutionary methods, therefore their results are not necessarily transferable to evolutionary domains.

3.1. The Effects on Evolution. We expect diversity in the ability of the individuals to solve different subsets of training examples to be advantageous in many evolutionary domains, mostly due to the following two reasons.

Firstly, in many applications, the training examples have varying degrees of difficulty. For example, in the field of automatic programming [28], a common approach is to present the system with a series of training examples of increasing difficulty, just like what can be found in regular introductory text books. In cases like this, a typical evolutionary run will evolve solutions for the easy examples first, while individuals able to solve the more difficult examples will not appear until near the end of the evolution, if at all. If most individuals in the population are able to solve only the easy examples, high diversity could indicate that additionally some individuals exist that can solve some of the harder examples only. Although these individuals are not able to also solve the simpler examples, they should usually be capable of evolving the ability to do so, since these simpler examples are per definition easier to solve.

In other words, in cases in which the training examples differ in difficulty, high diversity indicates the existence of some individuals capable of solving some of the more difficult examples, which in turn is likely to simplify the remaining evolution.

Secondly, according to the Building Block Hypothesis [16], evolutionary systems work by accumulating and combining promising partial solutions, or building blocks. Individuals containing promising building blocks usually have better fitness than other individuals, as a result such promising building blocks tend to multiply and thus receive more computational resources as the evolution proceeds. Moreover, in systems using crossover, like traditional GA and GP, promising building blocks from two or more individuals can be recombined in the hope of creating an even better solution.

If we regard the ability to solve the different training examples as building blocks, a population with high diversity obviously contains a larger number of different building blocks than a population with lower diversity. In cases in which two individuals able to solve different training examples can with reasonably high probability be combined into one individual able to solve most of the training examples solved by each of its two parents, we should expect high diversity to be of significant importance.

3.2. A Diversity Measure Based on Angular Distances. Consider the situation depicted in Figure 1(a). The two axes represent two different training examples denoted by x and y , and the points a and b represent the performance of two different individuals with respect to these two training examples. Assuming that fitness is to be minimized, a

represents an individual performing quite well on training example x , but not quite so good on training example y . In contrast, individual b performs well on example y but not so good on example x . This is similar to what occurs in nature, in that different species specialize on different kinds of food, and in some lucky cases, evolution may be able to combine a and b into an individual able to perform quite well on both training examples x and y , by combining the best from both individuals.

However, now consider the situation depicted in Figure 1(b). In this situation, we have an individual a' , performing reasonably well on both training examples, and an individual b' , clearly inferior to a' on both examples. In this case, we should not expect evolution to be able to combine a' and b' into offspring performing better than both of them, even though the Euclidean distance d' between a' and b' is about the same as the distance d between a and b in Figure 1(a).

There is an obvious difference between the two situations in Figure 1: although the distances between the individuals are the same, the angle θ between a and b is significantly larger than the angle θ' between a' and b' , where boldface notation denotes points regarded as vectors from the origin. Regarding our discussion in the previous section, these angles seem to be more important for evolution than the Euclidean distances between the individuals, since as shown in Figure 1, a high distance between two point is not necessarily a sign of any kind of specialization; it can also be a result of one individual being clearly inferior to the other on most or all of the training examples. In light of this, we introduce a new diversity measure, the *Angular Phenotype Diversity*, defined as the average angle between the fitness vectors of all pairs of individuals in the population P :

$$\text{APD}(P) = \frac{1}{\binom{|P|}{2}} \sum_{i,j \in P, i \neq j} \cos^{-1} \frac{i \cdot j}{\|i\| \cdot \|j\|}. \quad (3)$$

The use of angular distances to measure or promote diversity in evolutionary systems is not new (see, e.g., [36]); however, we are not aware of any previous attempts to measure diversity based on the angular distances between fitness vectors like this. We expect this kind of diversity to be more important during evolution than the Euclidean Phenotype Diversity as defined in (2) above. This expectation will be verified experimentally in Sections 4 and 5.

4. Experiments

In order to investigate the effects of Euclidean and Angular Phenotype Diversity on the ability of an evolutionary system to evolve highly fit individuals, and to compare them with previously investigated diversity measures, we have conducted a series of experiments. We use a (μ, λ) Evolution Strategy [37] with Uniform Crossover to evolve standard feed-forward neural networks [31] in a series of nonlinear regression problems. The neural networks are encoded as vectors of N weights; each individual in the population consists of one such vector together with a vector of N corresponding mutation standard deviations. These

are used for mutations as described in [37] to modify each weight by a normally distributed random amount with the given standard deviation. In each generation, the μ parent individuals are used to create λ offspring individuals, by repeatedly selecting two uniformly random individuals from the parent populations, recombining them using uniform crossover, and mutating the resulting individual. Finally, the μ fittest of the λ offspring individuals are selected for the next generation.

The regression problems consist of sets S of (\mathbf{x}, y) pairs, where $\mathbf{x} \in \mathbb{R}^n$ and $y \in \mathbb{R}$. The task of the system was to evolve a neural network that could approximate these sets. That is, a neural network N with a minimal error value

$$\text{error}(N) = \sum_{(\mathbf{x}, y) \in S} (N(\mathbf{x}) - y)^2 \quad (4)$$

is sought, where $N(\mathbf{x})$ is the result of feeding the network N with the input values \mathbf{x} .

In order to be able to compute Angular and Euclidean Phenotype Diversity, we need to define the fitness vector $F(N)$ of a neural network N . To do this, we have to assume some arbitrary but fixed ordering of the training examples S , that is,

$$S = ((\mathbf{x}_1, y_1), \dots, (\mathbf{x}_{|S|}, y_{|S|})), \quad (5)$$

where $|S|$ denotes the number of training examples.

The fitness vector $F(N)$ can now be defined as

$$\begin{aligned} F(N) &= (F(N)_1, \dots, F(N)_{|S|}) \\ &= ((N(\mathbf{x}_1) - y_1)^2, \dots, (N(\mathbf{x}_{|S|}) - y_{|S|})^2). \end{aligned} \quad (6)$$

As is common in regression problems [30], we use the squared error value as a measure of the ability of the network to solve a particular training example. The lower this value is, the better the network is at solving the example.

The fitness of N is simply the sum of the components of the fitness vector:

$$\text{Fitness}(N) = \sum_{i=1}^{|S|} F(N)_i = \text{error}(N). \quad (7)$$

That is, the fitness of the network is the sum of the squared error values of the network. The objective of our Evolution Strategies is to minimize this fitness function.

When solving real-world machine learning problems, it may often be disadvantageous to evolve optimal solutions based on a limited set of training data, due to the problem of overfitting. However, we have chosen to ignore this issue in our experiments, since our focus here is on how well the evolutionary system is able to search for good solutions, rather than on actually evolving good and general solutions to the machine learning problems considered.

4.1. The Training Input Sets. In our experiments, we used real-world data from six regression data sets. These sets were the following.

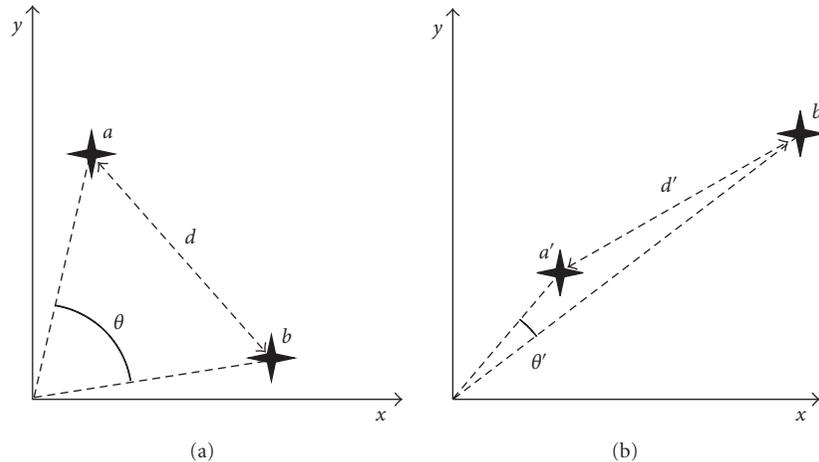


FIGURE 1: Different ways of measuring the distances between two individuals. x and y represent two different training examples on which fitness is to be minimized. The Euclidean distance is about the same in both (a) and (b), but the angular distance is much higher in situation (a) than in situation (b).

- (i) breast-cancer-wisconsin: A set containing data on a number of breast cancer patients. Each case is recorded with 32 numerical attributes in addition to the target value, which was the time of recurrence. The original data set contains data from a total of 198 cases, many of these are nonrecurring and/or lack some data. In our experiments, we only used the 46 recurring cases with no missing data.
- (ii) concrete: This set gives the compressive strength of concrete for various ages and ratios of seven different ingredients. 1030 instances are given, each of which contains 8 numerical input values in addition to the compression strength target value.
- (iii) forest fires: Data from 517 different forest fires in Portugal. Each instance contains 12 numerical attributes, such as geographical coordinates and various meteorological data. The target value is the total burned area. This data set turned out to be the hardest one, requiring more computational resources than any of the other data sets.
- (iv) cars: A set containing fuel consumption data for 392 different cars. Each car is recorded with 7 numerical attributes in addition to the target value (miles per gallon). The original data set contained data for 406 cars; however, some of these lack data the target value or the horsepower attribute, and were therefore removed from the set.
- (v) NO₂: Measurements of NO₂ concentrations at Alnabru in Oslo, Norway. Each of the 500 measurements is recorded together with 7 numerical attributes containing information about traffic volume, various meteorological data, time and date.
- (vi) boston_corrected: Data about house prices in Boston, coupled with various geographical and demographical data. The 16 instances containing censored observations as reported in [38] were removed from

the data set. The resulting data set contained 490 instances, each with 17 numerical values in addition to the target value.

The first three datasets were downloaded from the UCI Machine Learning Repository [39]; the other three were downloaded from the StatLib dataset archive [40].

4.2. Parameters Used in the Experiments. Although it is also possible to evolve the topology of neural networks by evolutionary approaches [41], for simplicity we have chosen to fix the topology in advance for each data set, and just use the Evolution Strategy to evolve the weights of the nodes. For each data set, neural networks with one hidden layer containing a fixed number of hidden nodes were evolved. In order to find a suitable number of hidden nodes and number of generations to use for each data set, a number of preliminary runs were conducted. These preliminary evolutions were all run for 20,000 generations, and the number of hidden nodes was set to 10, 20, ..., 90, or 100. For each data set, the lowest number of hidden nodes yielding the best or near-best results was selected, and the number of generations needed to reach this result was also recorded. The results are given in Table 1, these were the parameters used in the subsequent experiments. Since these parameters are the results of a simple informal analysis of a low number of evolutions, we do not expect them to be optimal in any way. However, our concern here is not really to run as optimal evolutions as possible, but rather to investigate the impact of various diversity measures on these problems. For this purpose, it is sufficient to investigate evolutions yielding reasonably near-optimum solutions.

All hidden nodes used the tanh function as activation function. Since the target values contain any ranges of y -values, we simply used the identity function as activation function in the output node.

Each of these sets of experiments were conducted using a (100, 700) Evolution Strategy, that is, using a parent size

TABLE 1: Number of hidden nodes and generations used for the various data sets.

Data set	Number of hidden nodes	Number of generations
Breast-cancer-wisconsin	40	10,000
Concrete	40	5,000
Forestfires	100	14,000
Cars	40	10,000
NO2	40	5,000
Boston_corrected	70	10,000

of $\mu = 100$ and, following the recommendation in [37], an offspring size of $\lambda = 7$, $\mu = 700$.

After each experiment, the evolution was evaluated based on the fitness of the best individual ever achieved during the run. The resulting fitness value, together with the different diversity measures of the population at each generation, were used as described in Section 4.3 below to compute at each generation the correlation between the different diversity measures and the best achieved fitness.

The diversity measures used were Euclidean and Angular Phenotype Diversity as defined in (2) and (3). Additionally, in order to compare these measures with some more commonly used diversity measures, we also measured Fitness Diversity and Genotype Diversity. The former consists of summing the absolute distances between the fitness values occurring in a population [25], whereas in the latter the Euclidean distances between the genotypes in the population are summed [3].

4.3. Computing the Correlation between Fitness and Diversity. To compute the correlation between two variables, a commonly used measure is Spearman’s Rank Correlation Coefficient [42], commonly known simply as Spearman Correlation. The Spearman Correlation is a value between -1 and 1 that indicates the degree to which two variables relate monotonically to each others. No assumption is made about the kind of relationship between the variables, all that is needed is a total ordering of their respective domains.

A Spearman Correlation of 1 means that the relationship between the two variables is completely monotonic, and that if one of them increases, the other one increases as well. Conversely, a Spearman Correlation of -1 means that as one of the variables increases, the other one always decreases. A Spearman Correlation of 0 means that if one of the variable increases, the other variable is just as likely to increase as to decrease, and vice versa. All other Spearman Correlation values between -1 and 1 denote varying degrees of correspondence between the values of the two variables.

Given two series of values, the Spearman Correlation between them can be computed using the simple formula

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}, \quad (8)$$

where d_i is the distance in rank between the i th values of the two series and n is the number of values.

Like in [2], we ranked the fitness and diversity values by their actual values, giving low ranks to the best runs and the runs with low diversity. Consequently, if good individuals tend to occur in runs with high diversity, the Spearman correlation will be negative.

For each data set, 100 runs were conducted. The results of these runs were used to compute the Spearman Correlations for the different diversity measures at each generation. To compute the correlations, the 100 runs were ranked by the fitness of the best individual found during each run. Therefore, the computed correlation values give the correlation between diversity at each generation and the best fitness ever achieved during the run. This is in contrast to Burke et al. [2], who compute the correlation between diversity at each generation and fitness achieved at the same generation, not considering the effect of diversity on the final results of the evolution.

5. Results

5.1. The Evolution of Fitness and Angular Phenotype Diversity. After running the 100 runs for each data set, we investigated how fitness and Angular Phenotype diversity evolved during the runs. Figure 2 shows the best fitness and Angular Phenotype diversity at each generation for the first generations of the evolutions for each data set. The fitness values and diversities were averaged over all the 100 runs performed for each data set.

The scales vary; this is due to different numbers of generations and various differences among the data sets. Note in particular the scale of the Angular Phenotype Diversity in the forestfires experiments, where diversity levels are significantly lower than in all the other experiments.

Although the scales vary, we see some clear trends in these graphs. In all cases, the evolution of diversity over time is about the same: initially, a slight increase in diversity is observed, followed by a steady decrease until the run is terminated. The initial increase in Angular Phenotype Diversity might indicate that at the start of evolution, there is a short phase in which some sort of specialization occurs, in that individuals evolve to specialize in different subsets of the training examples. The decrease in diversity for the remainder of the evolution is then probably a result of a gradual convergence towards one of these specialized individuals, which at the same time is improved and generalized by mutations and recombinations with other individuals.

5.2. The Correlation between Fitness and Diversity. The correlation between the different forms of diversity at each generation and best fitness achieved during evolution is plotted against generation number in Figure 3 for the different data sets used. In all of our cases, the correlation at the start of the evolution is near zero for all diversity measures, indicating that the diversity of the initial random population has little or no effect on the eventual fitness achieved.

In general, high diversity means that a high number of different genes, traits, or other kinds of potential building

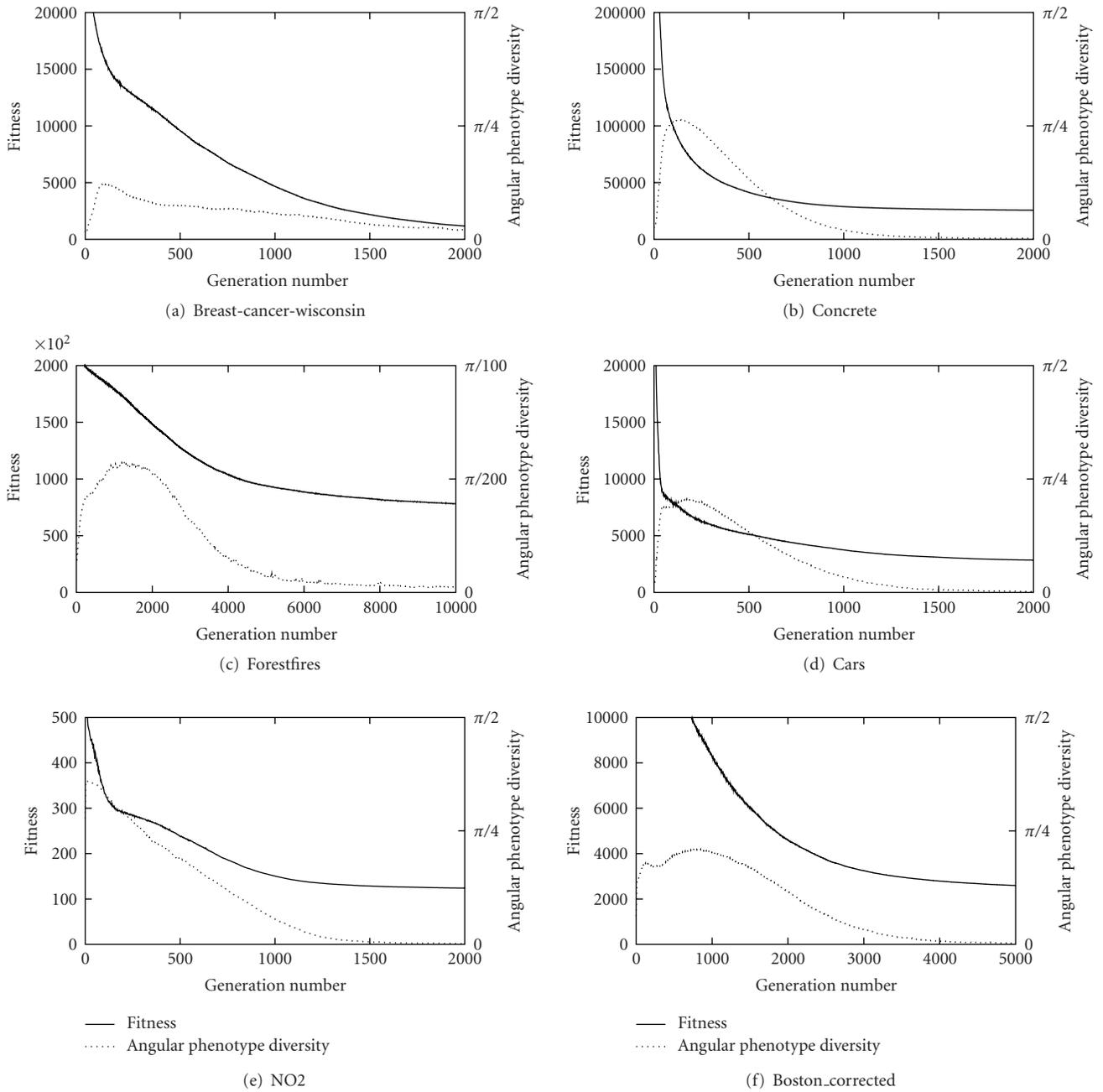


FIGURE 2: Averaged evolution of fitness and angular phenotype diversity during the experiments with the six data sets.

blocks are present in the population. However, this high number of different potential building blocks will only be useful for further evolution if a sufficient number of the building blocks are actually useful in some sense, in that they contain some parts of a solution to the problem which can later be recombined into better and more general solutions. Since each run is initialized with a random population, the number of useful building blocks in the initial population is likely to be very low. Only after a few generations, useful building blocks are likely to have evolved, and until then the impact of diversity is very low.

After the first few generations, the correlations of the different diversity measures evolve differently. Comparing the four different forms of diversity considered, we note that for all our data sets, Angular Phenotype Diversity seems to be more beneficial for the evolution of fit individuals than any of the other diversity measures. This is particularly true for the breast-cancer-wisconsin and forestfires data sets, especially in the latter, Angular Phenotype Diversity correlates very strongly with best fitness achieved by the evolution. In the cars and NO2 data sets, Angular Phenotype Diversity is also more beneficial for evolving fit individuals than any of the other diversity measures investigated, although

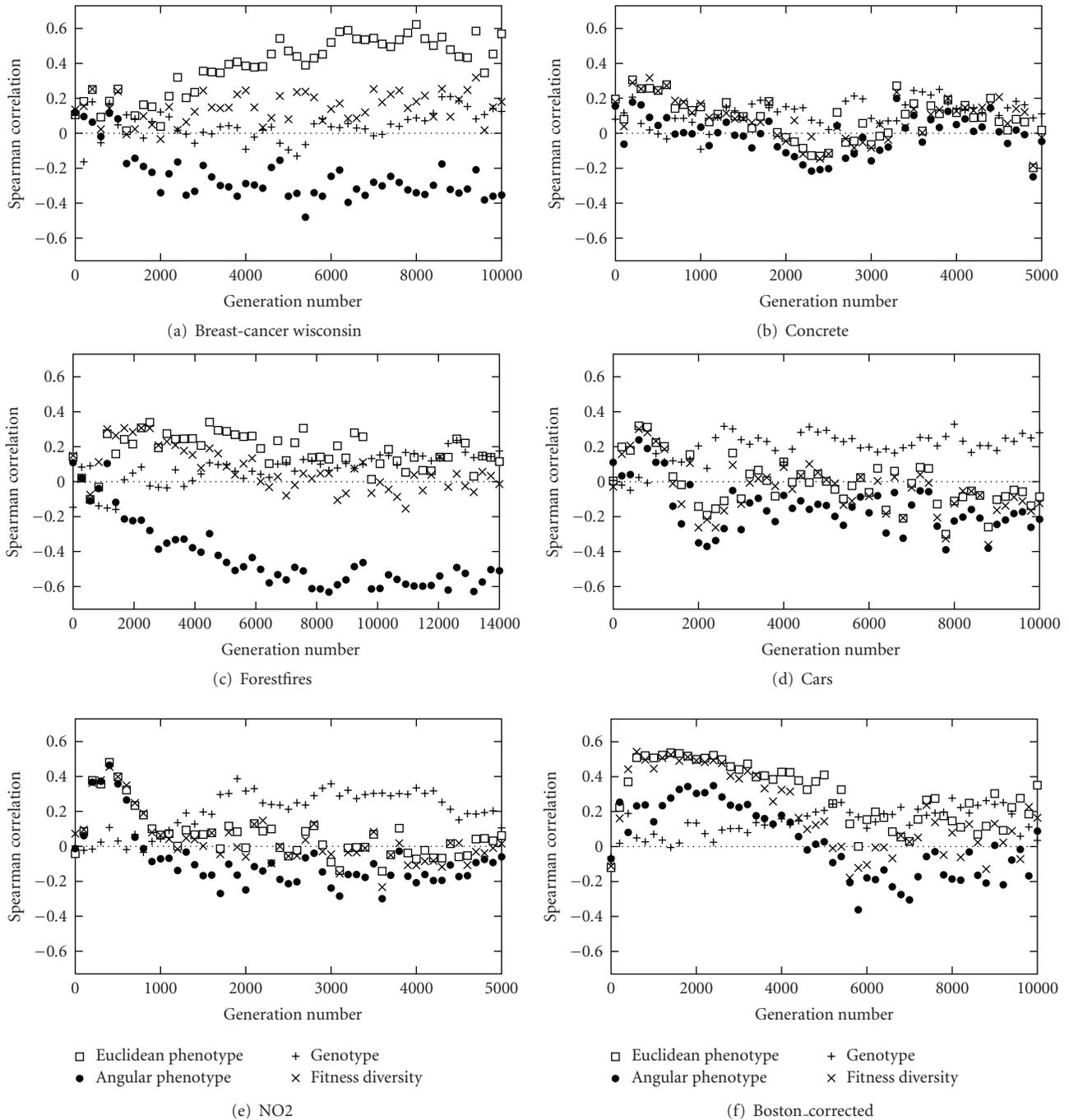


FIGURE 3: Spearman correlation between fitness and various forms of diversity during evolution using the six different data sets. Negative correlations indicate that diversity is beneficial for evolution of good fitness; positive correlation indicates that diversity is harmful.

correlations here are quite weak. During the final half of the evolution when solving the boston_corrected data set, Angular Phenotype Diversity is somewhat more beneficial than the other diversity measures considered, but on this data set, the correlation between Angular Phenotype Diversity and fitness was positive in the first half of the evolution, indicating that high diversity is only useful in the final half of the evolution. Finally, in the concrete data set, none of the diversity measures seem to be significant; most of them show

a slight positive correlation with fitness, indicating that high diversity decreases the chances of evolving fit individuals. But even here, Angular Phenotype Diversity seems more useful than the other diversity measures.

By comparing Figure 3 with Figure 2, we note that the data set yielding the strongest negative correlations between fitness and Angular Phenotype Diversity, namely, the forestfires data set, also gave rise to the lowest Angular Phenotype Diversity values of all the data sets considered

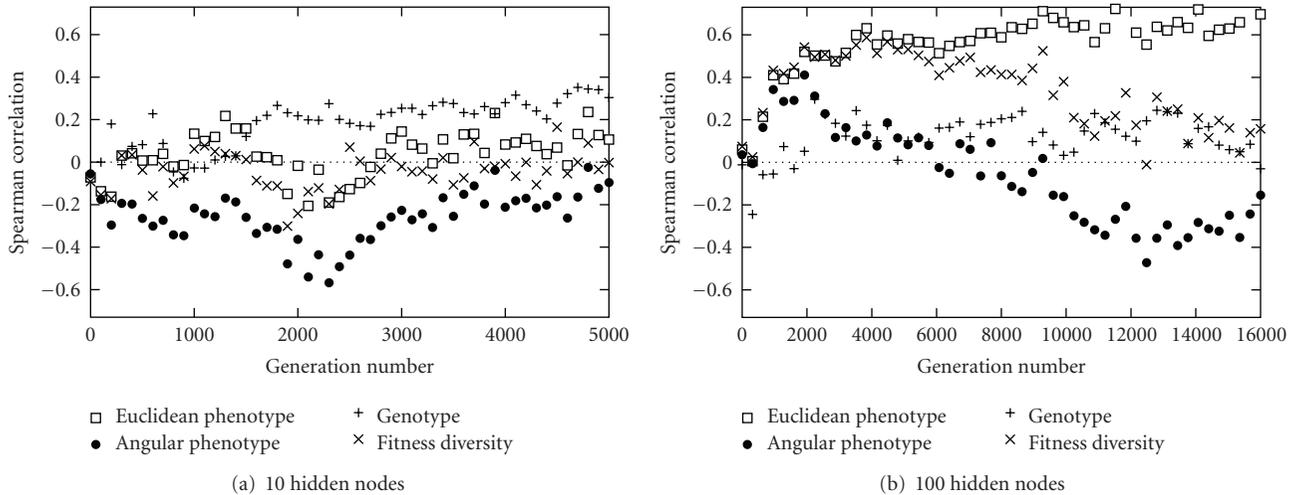


FIGURE 4: Spearman correlation between fitness and various forms of diversity during evolution using the breast-cancer-wisconsin data set and different numbers of hidden nodes. Negative correlations indicate that diversity is beneficial for evolution of good fitness, positive correlation indicates that diversity is harmful.

during evolution. Similarly, the breast-cancer-wisconsin data set, yielding the second strongest negative correlations between fitness and Angular Phenotype Diversity, gave rise to the second lowest Angular Phenotype Diversity values during evolution. When using the data sets yielding weak negative or even positive correlations between fitness and Angular Phenotype Diversity, most notably the boston_corrected, NO2 and concrete data sets, the Angular Phenotype Diversity levels were significantly higher. In other words, in cases in which Angular Phenotype Diversity is on average low, more diversity is beneficial for the evolution of good solutions, whereas in cases in which Angular Phenotype Diversity is on average higher, more diversity does not help, and may even be harmful.

5.3. About the Robustness of the Results. As mentioned in Section 4.2 above, values for the parameters used in the experiments were selected based on a small number of preliminary runs. In particular, for each data set the number of hidden nodes and the number of generations to use was selected by running one evolution for each of a few different numbers of hidden nodes, and selecting the lowest number of hidden nodes and the lowest number of generations that yielded the best or approximately best results. In order to investigate the robustness of the results with respect to these choices, we repeated the experiments twice for the breast-cancer-wisconsin data set, using 10 and 100 hidden nodes, respectively. The number of hidden nodes affects the convergence rate in that with a higher number of hidden nodes, more generations are needed before the population converges at a near-optimal solution. Therefore, we also changed the number of generations in these experiments to 5,000 and 16,000, respectively.

The results of these experiments are given in Figure 4. Comparing these results with those given in Figure 3(a), where the same data set was solved using 40 hidden nodes,

we note that our result that Angular Phenotype Diversity is more useful than the other kinds of diversity still holds, except in the first part of the evolutions with 100 hidden nodes, where the correlation between fitness and Angular Phenotype Diversity is positive, indicating that diversity is harmful for evolution. Genotype Diversity seems to be better, that is, less harmful, during this period. However, after the first 5000 generations, Angular Phenotype Diversity seems to be less harmful than the other forms of diversity considered, and after about 10,000 generations, it is the only form of diversity that correlates negatively with fitness, indicating that high Angular Phenotype Diversity in this phase of the search increases the probability of the system to evolve good solutions.

6. Conclusions and Further Work

In this paper, we have introduced a new diversity measure, the Angular Phenotype Diversity, based on angular distances between the fitness vectors of the individuals in the population. Comparisons with other diversity measures have been made by repeatedly running a number of regression problems and comparing the Spearman Correlation between achieved fitness and the different diversity measures, and in most of the experiments, Angular Phenotype Diversity turned out to have a stronger correlation with fitness than the other diversity measures considered. We draw the conclusion that our new diversity measure is potentially very useful for the domain considered, in the sense that the amount of Angular Phenotype Diversity in a population has a significant impact on the probability of finding good fitness values during the remainder of the evolution.

However, we have so far only considered evolution of neural networks to solve regression problems. Directions for future research may include investigating Angular Phenotype Diversity in other applications like Genetic Algorithms based

Classifier Systems [29], Genetic Programming [22], and other evolutionary systems for automatic programming, for example, the ADATE system [28]. Additionally, if Angular Phenotype Diversity turns out to be useful in other domains and other applications, diversity maintenance strategies (e.g., sharing [43–45]) aimed at maximizing Angular Phenotype Diversity should be investigated.

References

- [1] B. Wyns, P. De Bruyne, and L. Boullart, "Characterizing diversity in genetic programming," in *Proceedings of the 9th European Conference on Genetic Programming*, vol. 3905 of *Lecture Notes in Computer Science*, pp. 250–259, 2006.
- [2] E. K. Burke, S. Gustafson, and G. Kendall, "Diversity in genetic programming: an analysis of measures and correlation with fitness," *IEEE Transactions on Evolutionary Computation*, vol. 8, no. 1, pp. 47–62, 2004.
- [3] A. Toffolo and E. Benini, "Genetic diversity as an objective in multi-objective evolutionary algorithms," *Evolutionary Computation*, vol. 11, no. 2, pp. 151–167, 2003.
- [4] A. Ekárt and S. Z. Németh, "Maintaining the diversity of genetic programs," in *Proceedings of the 5th European Conference on Genetic Programming*, pp. 162–171, 2002.
- [5] D. Curran and C. O'Riordan, "Increasing population diversity through cultural learning," *Adaptive Behavior*, vol. 14, no. 4, pp. 315–338, 2006.
- [6] J. P. Rosca, "Genetic programming exploratory power and the discovery of functions," in *Proceedings of the 4th Annual Conference on Evolutionary Programming*, pp. 719–736, 1995.
- [7] J. P. Rosca, "Entropy-driven adaptive representation," in *Proceedings of the Workshop on Genetic Programming: From Theory to Real-World Applications*, pp. 23–32, 1995.
- [8] T. M. Mitchell, *Machine Learning*, McGraw-Hill, Singapore, 1997.
- [9] P. H. McQuesten, *Cultural enhancement of neuroevolution*, Ph.D. dissertation, Artificial Intelligence Laboratory, The University of Texas, Austin, Tex, USA, 2002.
- [10] J. He and X. Yao, "From an individual to a population: an analysis of the first hitting time of population-based evolutionary algorithms," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 5, pp. 495–511, 2002.
- [11] V. Nissen and J. Propach, "On the robustness of population-based versus point-based optimization in the presence of noise," *IEEE Transactions on Evolutionary Computation*, vol. 2, no. 3, pp. 107–119, 1998.
- [12] D. V. Arnold and H.-G. Beyer, "On the benefits of populations for noisy optimization," *Evolutionary Computation*, vol. 11, no. 2, pp. 111–127, 2003.
- [13] T. Smith, P. Husbands, P. Layzell, and M. O'Shea, "Fitness landscapes and evolvability," *Evolutionary Computation*, vol. 10, no. 1, pp. 1–34, 2002.
- [14] T. Bäck, "Selective pressure in evolutionary algorithms: a characterization of selection mechanisms," in *Proceedings of the 1st IEEE Conference on Evolutionary Computation*, pp. 57–62, 1994.
- [15] T. Blickle and L. Thiele, "A comparison of selection schemes used in evolutionary algorithms," *Evolutionary Computation*, vol. 4, no. 4, pp. 361–394, 1996.
- [16] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley Longman, Boston, Mass, USA, 1989.
- [17] H.-P. Schwefel, *Evolution and Optimum Seeking*, John Wiley & Sons, New York, NY, USA, 1995.
- [18] J. T. Alander, "On optimal population size of genetic algorithms," in *Proceedings of Computer Systems and Software Engineering (CompEuro '92)*, pp. 65–70, The Hague, The Netherlands, May 1992.
- [19] D. E. Goldberg, "Genetic algorithms, noise, and the sizing of populations," *Complex Systems*, vol. 6, pp. 333–362, 1992.
- [20] T. Jansen, K. A. De Jong, and I. Wegener, "On the choice of the offspring population size in evolutionary algorithms," *Evolutionary Computation*, vol. 13, no. 4, pp. 413–440, 2005.
- [21] M. Fuchs, "Large populations are not always the best choice in genetic programming," in *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO '99)*, pp. 1033–1038, 1999.
- [22] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, MIT Press, Cambridge, Mass, USA, 1992.
- [23] H. Shimodaira, "Dcga: a diversity control oriented genetic algorithm," in *Proceedings of the 9th International Conference on Tools with Artificial Intelligence*, pp. 367–374, 1997.
- [24] U.-M. O'Reilly, "Using a distance metric on genetic programs to understand genetic operators," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, vol. 5, pp. 4092–4097, 1997.
- [25] M. Hutter and S. Legg, "Fitness uniform optimization," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 5, pp. 568–589, 2006.
- [26] E. Burke, S. Gustafson, and G. Kendall, "A survey and analysis of diversity measures in genetic programming," in *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 716–723, 2002.
- [27] A. E. Magurran, *Ecological Diversity and Its Measurement*, Princeton University Press, Princeton, NJ, USA, 1988.
- [28] R. Olsson, "Inductive functional programming using incremental program transformation," *Artificial Intelligence*, vol. 74, no. 1, pp. 55–81, 1995.
- [29] L. B. Booker, D. E. Goldberg, and J. H. Holland, "Classifier systems and genetic algorithms," *Artificial Intelligence*, vol. 40, no. 1–3, pp. 235–282, 1989.
- [30] N. R. Draper and H. Smith, *Applied Regression Analysis*, John Wiley & Sons, New York, NY, USA, 3rd edition, 1998.
- [31] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice-Hall, Upper Saddle River, NJ, USA, 2nd edition, 1999.
- [32] R. Olsson, "The art of writing specifications for the adate automatic programming system," in *Proceedings of the 3rd Annual Conference on Genetic Programming*, pp. 278–283, 1998.
- [33] S. Gustafson, E. K. Burke, and G. Kendall, "Sampling of unique structures and behaviours in genetic programming," in *Proceedings of the 7th European Conference on Genetic Programming*, vol. 3003 of *Lecture Notes in Computer Science*, pp. 279–288, 2004.
- [34] G. Zenobi and P. Cunningham, "Using diversity in preparing ensembles of classifiers based on different feature subsets to minimize generalization error," in *Proceedings of the 12th European Conference on Machine Learning (ECML '01)*, pp. 576–587, Freiburg, Germany, September 2001.
- [35] T. G. Dietterich, "Ensemble methods in machine learning," in *Proceedings of the 1st International Workshop on Multiple Classifier Systems*, pp. 1–15, 2000.
- [36] O. Takahashi and S. Kobayashi, "An angular distance dependent alternation model for real-coded genetic algorithms," in *Proceedings of the Congress on Evolutionary Computation (CEC '04)*, vol. 2, pp. 2159–2165, 2004.

- [37] T. Bäck, *Evolutionary Algorithms in Theory and Practice*, Oxford University Press, Oxford, UK, 1996.
- [38] O. W. Gilley and R. K. Pace, "On the harrison and rubinfeld data," *Journal of Environmental Economics and Management*, vol. 31, no. 3, pp. 403–405, 1996.
- [39] A. Asuncion and D. Newman, "UCI machine learning repository," 2007, <http://archive.ics.uci.edu/ml>.
- [40] P. Vlachos, "The StatLib data set repository," 2009, <http://lib.stat.cmu.edu/datasets>.
- [41] X. Yao, "Evolving artificial neural networks," *Proceedings of the IEEE*, vol. 87, no. 9, pp. 1423–1447, 1999.
- [42] C. Spearman, "The proof and measurement of association between two things," *The American Journal of Psychology*, vol. 15, no. 1, pp. 72–101, 1904.
- [43] B. Sareni and L. Krähenbühl, "Fitness sharing and niching methods revisited," *IEEE Transactions on Evolutionary Computation*, vol. 2, no. 3, pp. 97–106, 1998.
- [44] P. Darwen and X. Yao, "Every niching method has its niche: fitness sharing and implicit sharing compared," in *Parallel Problem Solving from Nature*, vol. 1141 of *Lecture Notes in Computer Science*, pp. 398–407, Springer, Berlin, Germany, 1996.
- [45] C. D. Rosin and R. K. Belew, "New methods for competitive coevolution," *Evolutionary Computation*, vol. 5, no. 1, pp. 1–29, 1997.