

Multiobjective optimization in bioinformatics and computational biology

Julia Handl, Douglas B. Kell and Joshua Knowles

March 1, 2006

Abstract

This paper reviews the application of multiobjective optimization in the fields of bioinformatics and computational biology. A survey of existing work, organized by application area, forms the main body of the review, following an introduction to the key concepts in multiobjective optimization. An original contribution of the review is the identification of five distinct ‘contexts’ giving rise to multiobjective optimization problems: these are used to explain the reasons behind the use of multiobjective optimization in each application area and also point the way to potential future uses of the technique.

Part I — Key concepts

1 Introduction

Numerous problems encountered in bioinformatics and computational biology can be formulated as optimization problems, and thus lend themselves to the application of powerful heuristic search techniques [26, 122]. Traditionally, these problems have been stated as single-objective optimization problems, where the set of decision variables is optimized with respect to a single ‘goal’ to be maximized or minimized. This type of problem formulation has enjoyed enormous and varied success, and will continue to be useful for many problems in years to come. However, for certain problems, it is actually too restrictive because it is not always possible to properly formulate problems in terms of a single, well-defined goal that can serve as the basis for optimization.

The framework of *multiobjective optimization* offers greater flexibility in problem-solving. It can enable problems with a number of different, incommensurable, and often conflicting goals to be handled by *separate* objective functions, which are optimized simultaneously. Multiobjective optimization has been gaining recognition rapidly over the past few years through suc-

cessful applications across the science and engineering disciplines. Two key factors are responsible for much of the burgeoning interest in its use:

1. The recognition that the criteria and cost functions present in many real-world scenarios are conflicting and that their relative importance is often not known *a priori*. Multiobjective optimization allows one to explore the entire space of possible trade-off solutions, and to use subsequent *decision making* to single out preferred solutions.
2. Developments in the field of evolutionary computation. While *exact methods* for multiobjective optimization have been available for several decades [163], their applicability is restricted to certain types of optimization problems [95]. The last ten years have seen the development of powerful heuristics for multiobjective optimization that can deal with complicating factors (such as large numbers of variables, non-linearities or stochasticity) that are met in many real-world optimization problems [49, 95].

Bioinformatics and computational biology are two of the most exciting research areas in which heuristic optimization finds application, but multiobjective optimization approaches are only starting to be explored here. In this paper, we aim to outline the potential scope of methods for multiobjective optimization in biological applications, and to provide a review of existing work.

The remainder of this paper is structured as follows. Section 2 provides the background information relevant for an understanding of all subsequent sections. In particular, it provides definitions of bioinformatics, computational biology and single- and multiobjective optimization. In Section 3, we proceed to identify five distinct contexts in which multiple objectives may arise, or be used, in solving an optimization problem. Sections 4 to 8 contain the bulk of the survey material, organized by application area. References to the earlier

categorization by context are made in these sections in order to achieve our principal aim in this review: to unravel the variety of motivations behind the uses of multiobjective optimization in biological applications. Section 9 discusses our findings and issues arising from the survey, while Section 10 concludes the review.

2 Background

This section starts with definitions of the fields of bioinformatics and computational biology and a discussion of the subtle distinctions between the two. We then discuss the single- and multiobjective formulations of optimization problems.

2.1 Bioinformatics and computational biology

The National Institute of Health (NIH) defines bioinformatics and computational biology as follows [82].

Definition 1: Bioinformatics is the “research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral and health data, including those to acquire, store, organize, archive, analyze, or visualize such data.”

Definition 2: Computational biology is “the development and application of data-analytical and theoretical methods, mathematical modelling and computational simulation techniques to the study of biological, behavioral, and social systems.”

The above definitions accentuate a subtle difference between the fields of bioinformatics and computational biology. According to the NIH, the field of bioinformatics is concerned predominantly with the processing and storage of data, that is, it attempts to find the information hidden in the observational data gathered from *in vivo* or *in vitro* biological systems. In contrast, the field of computational biology generates novel data by means of *in silico* simulations that attempt to model and predict the behaviour of real biological systems. In practice, the differentiation between these two areas of research is, of course, rather fuzzy, and it is difficult to put some problems firmly in one category or the other. This fuzziness can be highlighted using the example of microarray data analysis. Typically, the data are analyzed using clustering methods or algorithms for the

inference of gene regulatory networks, which, according to the above definitions, should therefore be considered as bioinformatics techniques. However, while being a useful tool for data analysis, the inference of gene regulatory networks also creates mathematical models, which can, subsequently, be used for simulations and predictions, an application that falls clearly within the realm of computational biology.

Due to the difficulties of clearly separating the two fields, our review will not distinguish between applications appertaining to bioinformatics and/or computational biology. Also, in order to provide an overview as extensive as possible, our review will include problems that are of clear relevance in biology, even in cases where their multiobjective formulation has not yet been specifically explored within a biological application. Finally, our review will also include ‘borderline cases’, such as applications within the medical sciences.

2.2 Single-objective optimization

Single-objective optimization involves the optimization of a set of decision variables with respect to a given cost function defined over them.

Mathematically, a general (unconstrained) single-objective optimization problem (SOP) can be defined as:

$$\begin{aligned} \text{minimize } z &= f(\mathbf{x}) \\ \text{with } \mathbf{x} &= (x_1, x_2, \dots, x_n) \in X, \end{aligned}$$

where \mathbf{x} is an n -dimensional decision vector or solution, and X is the decision space, i.e. the set of all expressible solutions. The objective function f maps X into \mathbb{R} , and $z = f(\mathbf{x})$ is called the objective value. The image of X in objective space is the set of all attainable objective values, Z . A minimal objective value z exists, which is obtained by one or several solutions.¹

2.3 Multiobjective optimization

Multiobjective optimization involves the simultaneous optimization of a set of objective functions of the decision variables. Typically, the objective functions may estimate very different aspects of the solutions, aspects that are, therefore, *incommensurable* and often (partially or wholly) in conflict.

A general (unconstrained) multiobjective optimization problem (MOP) can be defined as:

$$\begin{aligned} \text{‘minimize’ } \mathbf{z} &= \mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})) \\ \text{with } \mathbf{x} &= (x_1, x_2, \dots, x_n) \in X, \end{aligned}$$

¹N.B. we assume minimization without loss of generality: any objective which is to be maximized can be replaced by its negative or reciprocal value.

where \mathbf{x} is an n -dimensional decision vector or solution, and X is the decision space, i.e. the set of all expressible solutions. The vector objective function $\mathbf{f}(\mathbf{x})$ maps X into \mathbb{R}^m , where $m \geq 2$ is the number of objectives. The vector $\mathbf{z} = \mathbf{f}(\mathbf{x})$ is an objective vector or point. The image of X in objective space is the set of all attainable points, Z .

The term ‘minimize’ appears above in quotation marks because its meaning is not yet defined. Alternative minimization problems exist, including: ordering the objectives and considering lexicographic optimization (e.g. as used in Olympic games medal tables, where a primary ranking is established based on the number of gold medals and draws in this ordering are resolved based on the number of silver medals, and, finally, based on the number of bronze medals); optimizing the worst objective function; a combination of these two; or others (see [47]).

However, by far the most frequent understanding of ‘minimize’, above, is in the sense of Pareto optimality. The *Pareto optimal set* of solutions consists of all those that it is impossible to improve in any objective without a simultaneous worsening in some other objective:

$$X^* = \{\mathbf{x}^* \in X \mid \nexists \mathbf{x} \in X, \mathbf{f}(\mathbf{x}) \leq \mathbf{f}(\mathbf{x}^*)\}, \text{ where}$$

$$\mathbf{f}(\mathbf{x}^1) \leq \mathbf{f}(\mathbf{x}^2) \text{ iff } \forall i \in 1..m, f_i(\mathbf{x}^1) \leq f_i(\mathbf{x}^2) \wedge$$

$$\exists j \in 1..m, f_j(\mathbf{x}^1) < f_j(\mathbf{x}^2).$$

The points in objective space corresponding to the Pareto optima are termed *nondominated* and form the *Pareto front*.

In most cases, the Pareto optimal set contains more than one element because there exist different trade-off solutions to the problem, which offer different compromises of the objectives. Thus, in practice, solving a multiobjective problem often means that a human decision-maker (DM) is involved who then chooses a solution that is Pareto optimal (ideally). Methods of decision-making (before, after or interactively during search) have been extensively investigated in a branch of management science/operations research known as multi-criterion decision making (MCDM) [46, 12] and may include advanced methods of visualization, e.g. [173].

The *a posteriori* and interactive methods of decision-making are usually thought to be more effective as the decision-maker is helped by ‘seeing’ what trade-off solutions are possible. The realization of this has led to a burgeoning of methods for generating the whole Pareto set, or an approximation to it. Increasingly in the literature, the general term ‘multiobjective optimization’ is used to mean generation of a Pareto set (or an approximation to it) though more strictly this was originally referred to as multiple objective pro-

gramming (see [163]).

Sometimes the optimization problem of interest may not admit an exact approach, or the Pareto set may be exponentially large, so it is not possible to enumerate the Pareto set completely. In such cases, an approximation to the Pareto set, defined as follows, is sought instead:

$$A^* = \{\mathbf{x}^* \in A \subseteq X \mid \nexists \mathbf{x} \in A, \mathbf{f}(\mathbf{x}) \leq \mathbf{f}(\mathbf{x}^*)\}.$$

We can see from this definition that if $A = X$ then it is the (true) Pareto set, but with $A \subset X$ it need not be. Approximation sets like this are partially ordered in terms of their quality (see [77]).

Several distinct types of method for generating good approximations to the Pareto set have been developed, e.g. see [95, 139]. Evolutionary algorithm approaches have become particularly popular and good overviews of these can be found in [24, 27, 41, 56]. For the use of all these techniques in specific application domains, see e.g. [49, 91, 117].

Note, in this review, we do not exclude any method of multiobjective optimization, exact or otherwise: to do so would reduce our outlook unnecessarily. However, approaches that generate the whole Pareto set/front (or an approximation to it) are used the most frequently in the application areas in the literature we consider here. Therefore, when using the term ‘multiobjective optimization’ this is what we mean, unless stated otherwise.

3 Five distinct contexts giving rise to multiobjective optimization problems

In this review paper, our principal aim is to unravel the different *reasons* underlying the need for multiobjective optimization in biological applications. So, rather than focusing on the optimization *techniques* used in specific biological problem domains, our survey will refer instead to a categorization (set forth below) that is based on the different types of contexts in which multiple objectives may arise or be usefully exploited. We hope that this perspective will help the reader to identify commonalities in the different biological problems considered, and help to assess the potential of multiobjective techniques in biological applications not covered by this review.

Note that the categorization introduced in this section is not used as the main structure for the remainder of the manuscript, however: for reasons of clarity, Section 4 to Section 8 remain arranged by biological problem domain. A summary of the classification of

these problems with respect to our categories is only provided at the end of this review, in Section 9.

3.1 Standard MOO

As a first category, we identify the ‘standard’ context of multiobjective optimization, where all objectives are clear, measurable goals that we would genuinely like to optimize.

Mathematically, this setting can be described as follows.

$$\begin{aligned} \text{‘minimize’ } \mathbf{z} = \mathbf{f}(\mathbf{x}) &= (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})) \\ \text{with } \mathbf{x} &= (x_1, x_2, \dots, x_n) \in X. \end{aligned}$$

We assume that all important criteria have been included as objectives. Hence, while we may be unsure about the relative importance of these objectives, we are certain that our ‘ideal’ solution will be Pareto optimal with respect to the objectives chosen. Thus, assuming an approach that generates a Pareto front (approximation) a decision maker may use the front to learn about the conflicts between the objectives, as well as the space of possible solutions, and may subsequently use preference information to select a single solution from this front.

Examples of this type of problem setting include typical engineering design problems where a trade-off between cost and quality has to be met (see Section 8).

3.2 Counterbalance for bias

The second category we identify are those optimization problems where multiobjective optimization serves as a tool to counter-balance the effect of a bias. In this setting, we typically have one primary performance criterion but a straightforward single-objective optimization is not possible due to a lack of objectivity of the primary objective: it is biased with respect to some measurable solution property. However, provided that each bias is at least monotonic in a measurable function $g(x)$, then optimizing $g(x)$ can be used to remove the effect of the bias (the precise form of the bias is not known, so it cannot simply be subtracted).

Mathematically, this setting can be described as follows, assuming just one primary objective:

$$f(\mathbf{x}) = f'(\mathbf{x}) + m(g(\mathbf{x})),$$

where f' is an ideal (i.e. unknown), unbiased measure of the primary objective, $m(g(\mathbf{x}))$ is a bias term where m is an unknown but monotone function of a measurable

function g , and f is the measurable but biased sum of the two.

We would like to minimize $f'(\mathbf{x})$ as follows:

$$\text{minimize } f'(\mathbf{x}) = f(\mathbf{x}) - m(g(\mathbf{x})),$$

but since m is unknown we cannot formulate the problem in this way. However, we may formulate the problem instead as:

$$\begin{aligned} \text{‘minimize’ } & (f(\mathbf{x}), -(g(\mathbf{x}))), \\ \text{with } \mathbf{x} &= (x_1, x_2, \dots, x_n) \in X, \end{aligned}$$

in terms of two measurable objectives.

Hence, the framework of multiobjective optimization is used as a means of introducing an additional objective, g , to counter-balance the bias of the primary objective.²

The set of Pareto optimal solutions will certainly contain the correct answer since each Pareto optimum is the best value of $f(x)$, given a fixed value of $g(x)$. In this scenario, selection of the best solution does not usually depend on preferences, but on the estimation of the biases. In some applications, the biases may be estimated using random control data, and this may help to identify the best solution in the Pareto front.

Examples of this type of problem include unsupervised feature selection, and sequence and structure alignment problems (see Section 4.2.3 and 6).

3.3 Multiple source integration

The third category we identify uses multiobjective optimization to integrate noisy data from multiple sources. Hence, in this setting, multiobjective optimization is used as an alternative to an *a posteriori* integration technique. The problems where this approach is used are originally single-objective. However, multiple noisy views of the data need to be integrated, as their combined use may yield better results than the use of data from a single information source.

Mathematically, this setting can be described by a set of objective functions:

$$\begin{aligned} f_1(\mathbf{x}) &= f'_1(\mathbf{x}) + \bar{n}_1 \\ f_2(\mathbf{x}) &= f'_2(\mathbf{x}) + \bar{n}_2 \\ &\vdots \\ f_m(\mathbf{x}) &= f'_m(\mathbf{x}) + \bar{n}_m \end{aligned}$$

where each function value of each objective function f_i is equal to an ideal function value f'_i with some unknown random noise \bar{n}_i on it, for $i \in 1..m$. In some

²N.B. the equations above can be generalized to more than one primary objective, where necessary.

cases, the f' are all identical, i.e. the ‘views’ of the data arise from the *same* types of measurement but e.g. taken at different times. By formulating the problem as,

$$\begin{aligned} \text{‘minimize’ } \mathbf{z} = \mathbf{f}(\mathbf{x}) &= (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})) \\ \text{with } \mathbf{x} &= (x_1, x_2, \dots, x_n) \in X, \end{aligned}$$

and finding the Pareto optima, the impact of the noise may be reduced, if it is reasonably uncorrelated with the solution space X . Nonetheless, note that it is not guaranteed that the desired solution will be among the Pareto optima.

Examples of this type of problem are the inference of phylogenetic trees and data clustering with several dissimilarity matrices (see Section 5.1 and 4.2.1).

3.4 Performance approximation by proxies

Category four comprises those applications in which the ‘real’, underlying objective of the problem, $f'(\mathbf{x}, \mathbf{y})$, is a function of both the solution \mathbf{x} and some ‘hidden’ variables \mathbf{y} that are not available during optimization. For example, in training a supervised classifier, \mathbf{y} refers to the generalization ability of the classifier on *future* data (which may be estimated using a test set *after* the optimization, providing the classifier is not trained using these examples).

Since the function f' is not suitable for use in the optimization process (because \mathbf{y} is unavailable), it needs to be replaced by ‘proxy’ objectives $f_i(\mathbf{x})$, which are functions of \mathbf{x} only. Often, such ‘proxy’ objectives only capture certain aspects of a good solution, and different proxies are complementary with respect to each other. In such a scenario, we can often assume that the desired solution will score relatively highly under all of the ‘proxy’ objectives, and the simultaneous optimization of all proxies therefore seems useful. Due to the use of ‘proxy’ objectives, the desired solution cannot be guaranteed to be among the associated set of Pareto optima.

Examples of this type of problem include supervised classifier training (as explained above), data clustering and protein structure prediction (see Section 4.1, 4.2.1 and 7).

Note, the difference between this context and that of standard MOO (as introduced above) may seem unclear to some readers. However, the distinction is clear: in the case of standard MOO, the objective functions have primacy, i.e. it is they that define the Pareto set, e.g.: the concept of a best car does not exist *per se* but given a search space a set of ‘best cars’ is *induced* by the objectives chosen. In contrast, in the context

of proxy objectives, it is the solution that has primacy, and the objectives are only a means of orienting the search in order to discover this solution, e.g.: the real structure of a protein exists, and we may try and find it by employing a number of different energy/cost functions.

3.5 Multi-objectivization

The fifth and final category we identify is where multi-objective optimization may be used solely as a way to obtain improved search ‘guidance’ in what is essentially a single-objective problem.³ Assuming a single objective that is measurable, a problem may still be difficult because of its search landscape. There are at least two difficulties in search landscapes that can potentially be reduced by ‘multi-objectivization’: (i) where a problem involves frustration (or epistasis), which causes excessive local optima in the search landscape; (ii) where the search landscape contains regions offering little or no objective function gradient. In the first case, decomposition of the primary objective into several different functions (each function either defined over all of the variables or a subset of them) may help to separate out the conflicting aspects of the problem, thus reducing the number of local optima ‘seen’ by a search algorithm [101]. In the second case, the use of extra ‘helper objectives’ in addition to the primary objective may provide helpful guidance in the flat regions of the landscape [87, 101].

Multi-objectivization may potentially be achieved by any reformulation of the problem that respects the following relation [101]:

$$\forall \mathbf{x}^{opt} \in X \exists \mathbf{x}^* \in X, \mathbf{x}^* = \mathbf{x}^{opt},$$

where \mathbf{x}^{opt} is an optimal solution to the original single objective problem and \mathbf{x}^* is a Pareto optimum of the multi-objectivized problem. This ensures that at least one of the true Pareto optimal solutions will be optimal with respect to the original primary objective and will correspond to the best solution.

An example of this type of problem is structure identification from powder diffraction data (see Section 7.1).

³NB: there is no reason why multi-objectivization cannot also be generalized to the case where the original problem is multi-objective.

Part II — Survey

4 Classification

A large number of problems typically encountered in bioinformatics are classification problems. We will now briefly introduce the main areas of classification, namely unsupervised, supervised and semi-supervised techniques, and will discuss the applicability of multi-objective approaches in these different areas.

4.1 Supervised classification

Supervised classification techniques require the presence of training data, that is, a (sufficiently large) set of data samples for which the correct classification is known. This set of data patterns paired with their known correct class memberships is then used to train a classifier (such as neural networks, support vector machines or decision trees) to learn and correctly predict the class memberships of these data items in the hope that the trained classifier can subsequently be applied to the classification of new unlabelled data [120]. Thus, the overall aim in supervised classification is to obtain a classifier with good *generalization properties*, i.e. a classifier that performs well on new previously unseen test data. Evidently, this goal of supervised classification cannot be measured objectively during the training of a classifier, so only ‘proxy’ objectives can be used to try and assess its expected generalization performance. For example, the accuracy of a classifier, i.e. the fraction of correctly classified data items, may be computed on the training data or, alternatively, on a second or further test data set. However, dependent on the application area tackled, other expectations regarding desirable properties of a classifier may exist, some of which we discuss in the following subsections.

Given a definition of quality by means of a set of ‘proxy’ objectives, we can then embark upon the optimization of a classifier with respect to the criteria chosen. In particular, the performance of a supervised classifier is influenced by two components, namely the set of input features and the classification model used, both of which can be optimized. These two different optimization problems are also referred to as supervised feature selection and model selection. Dependent on the number of objectives chosen, the corresponding optimization problem will be of a single- or a multiobjective nature. Here, we consider a number of examples from the literature, where a multiobjective formulation has proven useful.

4.1.1 Receiver operating characteristics (ROC) curve

When considering the performance of binary classifiers (e.g. for the distinction between tumor and healthy tissue), sensitivity and specificity are often seen as more informative measures of the classification performance than the overall classification accuracy: sensitivity captures the fraction of true positives for all positive conditions, whereas specificity captures the fraction of true negatives for all negative conditions. Especially as far as diagnostic systems are concerned, a description on this level may be more desirable to a practitioner than a description in terms of overall classification accuracy, as an explicit optimization of these two aspects of classification performance allows one to take into account one’s own knowledge regarding the relative costs of these two different types of classification errors.

The sensitivity and specificity of a classifier are always conflicting, and, for a given classifier, the trade-off between the two can be represented in the form of a receiver operating characteristics (ROC) curve [119]. Traditionally, this trade-off curve has not been explicitly optimized and, instead, an *a priori* weighting of the two objectives has been used during training. For the classifier obtained for a given data set, a ROC curve can then be generated by varying one or more of the parameters of the classifier, and plotting the impact on the values of sensitivity and specificity. A practitioner could then pick any point on the ROC curve that corresponds to the desired sensitivity or specificity. However, a ROC curve obtained in this way is unlikely to be optimal in the Pareto optimal sense, and more favourable trade-offs between sensitivity and specificity can be obtained through the direct use of Pareto optimization [105]. The principle of ROC curves can be extended to account for multi-class problems, and recent work has illustrated the optimization of the resulting ROC surfaces using multiobjective evolutionary optimization [51].

4.1.2 Partial classification / rule mining

In recent years, there has been increasing interest in methods of partial classification (also referred to as nugget mining or classification rule mining), of which the aim is to identify and describe interesting subsets of the data only. The motives underlying this kind of approach are that, in many data-mining applications such as the analysis of gene expression data [156], it may be (small) subsets of the data (which show exceptional and/or unexpected behaviour) that are of real interest. Classification rule mining therefore aims to identify crisp or fuzzy rules that describe the class memberships of individual subgroups of the data.

A fundamental difference between partial and traditional classification is the way in which performance can be evaluated. In particular, accuracy over the entire data set is not usually considered the most relevant performance measure in partial classification, and measures operating on subsets of the data are introduced instead. Among the most established performance measures in partial classification are the measures of coverage and confidence [11], which reflect the principles of sensitivity and specificity on a local basis. Coverage gives the proportion of total members of a class correctly described by a given rule (i.e. maximizing coverage means minimizing the number of false negatives), whereas confidence gives the proportion of the patterns to which a given rule has been correctly applied (i.e. maximizing confidence means minimizing the number of false positives). Evidently, coverage and confidence are conflicting criteria, and multiobjective evolutionary algorithms have therefore been applied to their optimization in rule mining [38, 39]. This is a flexible alternative to the use of fixed thresholds on coverage and/or confidence during or after the optimization process [25].

4.1.3 Balancing model accuracy and complexity

We have seen above that multiobjective optimization may provide a more comprehensive way of estimating classification performance in some specific supervised classification tasks. A different and persistent problem in supervised classification is the trade-off between model performance and model complexity: if sufficiently trained, many types of classifiers can obtain a very high classification accuracy on the training data. However, this usually entails the use (or identification) of a complex classification model (a large neural network, a large decision tree, etc.), which may yield low generalization properties and will be more difficult to interpret. In general, simple models are therefore preferred during classification (this is also known as Occam’s Razor [151, 168]) in order to avoid overtraining, and this principle needs to be integrated into the optimization process. Traditional approaches to avoid overtraining include cross-validation, early stopping in the training of neural networks [142] or the pruning of decision trees [50]. In contrast to these, multiobjective optimization provides a more general and flexible framework to integrate model complexity into the optimization process [58]. The benefits of integrating model complexity as a second and/or third objective have been previously demonstrated with respect to fuzzy rule mining [84, 85], learning classifier systems [113], decision trees [98], support vector ma-

chines [83], genetic programming (GP) [13, 37, 129] and artificial neural networks [2, 92, 172, 174]. A specific application in bioinformatics has been the use of multiobjective genetic programming for the identification of quantitative structure-activity relationships (QSAR) [129]. In this work, model complexity was measured using a number of different aspects including the total number of terms, the number of non-linear terms and a knowledge-based measure of the chemical interpretability of the descriptors used. The results obtained were shown to be at least as good as those obtained by traditional single-objective approaches and allowed the practitioner to explicitly explore the trade-off between statistical performance and model complexity.

4.1.4 Supervised feature selection

The problem of supervised feature selection is another example in which the trade-off between classification accuracy and model complexity is relevant. In a supervised scenario, larger feature sets will often result in a higher classification accuracy on the training data — yet, these solutions may often be prone to overfitting and may have a low generalization performance. The simultaneous maximization of the accuracy of a classifier and minimization of the cardinality of the feature space is therefore desirable in order to explore the trade-off between these two objectives: given two feature sets of different cardinality that result in the same classification accuracy, the smaller of the two is expected to result in a better generalization performance [78]. Multiobjective formulations of the supervised feature selection problem have been proposed, which directly capture this intuition [44, 104, 110, 111, 131, 154]. For example, Deb and Reddy used a multiobjective evolutionary algorithm to minimize three objectives: (i) feature set size, (ii) the number of mis-classifications on the training data (using leave-one-out cross-validation) and (iii) the number of mis-classifications on the testing data. The algorithm was applied to gene expression data and the results obtained revealed that certain subsets of features were present in many of the feature subsets in the Pareto front. They concluded that an analysis of the ensemble of feature sets present in the Pareto front may yield novel insight regarding the relative importance of individual features (see Figure 1 for a representative result obtained on the Leukemia data set [64]).

As an aside, a simpler and very common approach to supervised feature selection is the selection of variables based on their discriminatory power with respect to the target classes. This can, for example, be established using statistical t-tests [161], but for high-

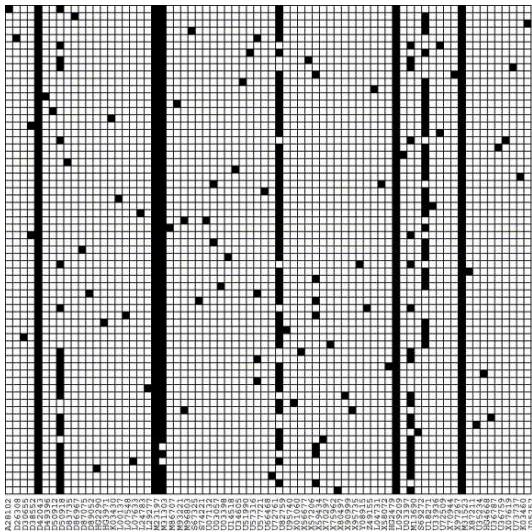


Figure 1: Representative set of solutions found by Deb and Reddy for the Leukemia data set [44]. Each row indicates an eight-gene classifier capable of making 100 per cent correct classification. The participating genes are marked in solid boxes (reprinted from [44] with permission from the American Chemical Society).

dimensional data, as are typically encountered in post-genomic data analysis [74], the approach inherently suffers from multiple testing issues (i.e. a large number of false positives will result due to random correlations). Alternative objectives regarding the properties of interesting genes may exist, such as small intra-cluster dispersion, large inter-cluster dispersion, or absolute differences between expression levels. The use of multiple, uncorrelated criteria for gene selection may help to reduce multiple testing issues. In recent work by Hero *et al.* [80, 79], the use of a multiobjective formulation has therefore been suggested as a formalized way for the screening of genes in the presence of more than one criterion.

4.1.5 Ensemble learning

In the previous subsections we have explored different ways in which multiobjective optimization can be used to obtain a set of classifiers, which correspond to different trade-offs between the objectives. The aim in most of these applications was to provide an increased degree of choice to the practitioner who will eventually need to decide on a single classifier to be used. However, given the diversity present in the set of classifiers generated by these multiobjective approaches, an attractive idea is its use as the input for an ensemble method.

There is a general agreement in the machine learn-

ing community that the integration of several classifiers by means of ensemble techniques may prevent overfitting, and increase both classification accuracy and robustness of the individual approaches. Generally, such a strategy may increase the level of confidence in the results returned [9, 78], and it has therefore been used in a variety of biochemical applications [15, 22, 53]. Two different types of ensemble schemes can be distinguished in the literature. On one hand, different types of voting strategies can be used to combine the output of multiple different classifiers trained on the same data set. Alternatively, methods such as bagging [17], boosting [143] or random forests [18] can be used, which enhance the performance of a given classification method by combining the output of diverse classifiers of the same type. The latter approaches differ in the way diversity between the classifiers is obtained: this may be done using training on different bootstrapping data sets, a reweighting of the input data, or the introduction of noise.

In [1], the use of multiobjective optimization was explored as a tool to ensure diversity in the classifier ensemble. In particular, two different formulations were proposed, both of which used the classification accuracy on two different data sets as an individual objective. In the first approach these two data sets were obtained through a split of the available training data, while the second approach used a random perturbation of the original training data in order to obtain the second data set. The former of these approaches was found to be superior and to produce results comparable to those obtained by alternative ensemble techniques such as negative correlation learning [111]. An alternative multiobjective formulation, which explicitly optimizes classification accuracy and diversity of the members of the ensemble, was suggested in [23].

4.2 Unsupervised classification

In the previous section, we have reviewed applications of multiobjective optimization in supervised classification. In this section, we will now shift our focus to unsupervised classification, a field of research where the potential of multiobjective approaches has been less extensively explored.

Unsupervised classification works in the absence of any training data as such, i.e. without knowledge of the class memberships of individual samples. It therefore relies on the presence of distinct structure in the data, and it must be hoped that a distance measure or a reduced feature space can be identified under which related data items cluster together in data space. The overall aim in unsupervised classification is to identify interesting patterns in the data. This concept of

interestingness is even harder to quantify than that of ‘generalization performance’ in supervised classification, and equally calls for the use of ‘proxy’ objectives. In contrast to the supervised scenario, no predominant criterion like classification accuracy exists, and formulations of the unsupervised classification problem vary widely in their choice of objectives [86]. The appeal of the use of multiobjective approaches to the task of unsupervised classification predominantly arises from this variety of existing objectives. Many of these capture very different aspects of solution quality, and their simultaneous optimization may, therefore, be expected to yield a more robust performance.

In the following we will highlight a number of examples from the literature, where a multiobjective formulation of unsupervised classification problems has already been explored or at least seems promising.

4.2.1 Clustering

Clustering is the partitioning of data into subgroups, and it is one of the fundamental tasks met in unsupervised classification. Many different formulations of the clustering problem exist, the best known of which are based on the criterion of intra-cluster variance [86]. It is well-known that none of the existing clustering criteria can capture all of the different aspects that humans perceive as properties of a good clustering, such as the compactness of clusters, spatial separation between them and the compliance with local density distributions [74]. One possibility to reduce the problem of the failure of a clustering algorithm in scenarios where the clustering criterion employed is inappropriate, is the use of ensemble techniques to integrate the results of a variety of different clustering methods [164, 166]. An alternative to this *a posteriori* integration of different clustering results, is the direct optimization of a partitioning with respect to a number of complementary clustering criteria (see Figure 2). Recent work has shown that such multiobjective approaches to clustering can indeed result in an improved and robust performance across data exhibiting a range of different data properties, and may be superior to some *a posteriori* integration approaches [68, 70]. This work has also illustrated that good clustering solutions tend to give rise to distinct ‘knees’ in the Pareto front, and may be automatically identified through a comparison to random control data [68, 70].

The use of multiobjective optimization for clustering has also been proposed for situations in which multiple dissimilarity matrices are available [34, 54]. Multiple (and potentially conflicting) dissimilarity matrices may arise as a consequence of the availability of multiple sources of data that should be integrated into a

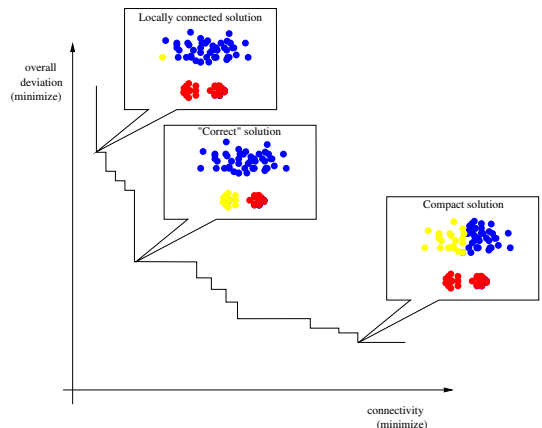


Figure 2: Illustration of the multiobjective optimization of two clustering objectives. The clustering solution in the centre, which is the most intuitive to a human observer, corresponds to a trade-off between the two clustering objectives used [70].

single clustering. Such data may be tackled through an *a priori* fusion of the data and the use of a standard clustering algorithm, the use of ensemble techniques for the *a posteriori* fusion of the different partitionings obtained, or the selection of a primary clustering objective and the definition of all others as the constraints in a constrained optimization problem. Dale and Dale [34] and Ferligoj and Batagelj [54] argue that a direct multiobjective approach may be more appropriate than any of the above methods, as it provides more information and choice to the decision maker.

Multiobjective optimization has also been used as a tool to counter-balance the bias of certain clustering criteria [112]: criteria such as the intra-cluster variance automatically decrease for a higher number of clusters, and are therefore unsuitable for the direct use as the only objective in a clustering algorithm if the number of clusters is a free variable. A simultaneous minimization of the variance and the number of clusters will counterbalance this bias and result in a trade-off curve containing at most one clustering solution for every number of clusters. Note that the trade-off curves obtained in this application are very similar to the curves employed in the Gap Statistic [165], a statistically motivated approach to the determination of the number of clusters. The design of a method for automated solution selection (in the style of the ad hoc approach formalized in the Gap Statistic) should, therefore, be straightforward.

4.2.2 Cluster validation

Cluster validation can be seen as a post-processing phase to clustering. It involves the assessment of the quality of a given clustering solution, with the aim of confirming that real structure in the data has been identified, or of performing model selection: in many applications, we will be able to choose from a number of partitions of the same data (generated by different algorithms or corresponding to different numbers of clusters), and model selection serves to reduce this set to a single solution. In the absence of external knowledge about the correct cluster memberships (which are required for the use of external techniques of cluster validation [74]), cluster validation relies on the use of internal validation techniques [74]. The literature contains a range of internal validation techniques, which are usually based on the linear or non-linear combination of several clustering criteria (such as cluster compactness and cluster separation). The use of these techniques will usually yield a complete ranking of all clustering solutions, which may, however, (due to conflicts between the individual measures) be inconsistent with the ranking returned by some of the individual criteria. Here, multiobjective optimization provides a principled approach to obtain a partial ranking between clustering solutions, which is consistent with all of the clustering criteria considered [74]. Preliminary results regarding the performance of a multiobjective approach to cluster validation are given in [69].

4.2.3 Unsupervised feature selection

The majority of existing clustering algorithms operate on the original input feature space, and do not include any mechanisms for feature selection. In many applications, this may be detrimental to the clustering task, as the input data may contain many noisy or irrelevant variables, which will hide the structure in the data. It is therefore important to develop algorithms that can reduce the set of input variables to those that contain clear cluster structures and may, therefore, be interesting to analyze. This can either be done through the development of specialized clustering algorithms, which explicitly search through feature subspaces (such as biclustering algorithms, see below), or the development of algorithms for feature selection that can be used as pre-processing methods for the subsequent application of any clustering method.

While the subject of supervised feature selection has been thoroughly explored in the literature, little work exists on the topic of unsupervised feature selection. This is due to the difficulty of the formalization of criteria for the objective assessment of the quality of different feature subspaces. One particular problem is

the comparison of feature subspaces of different cardinality, as existing measures are usually biased towards small or large feature subspaces [45]. Multiobjective optimization has recently been introduced as a potential solution to this problem, as it allows one to optimize one of these objectives and to counterbalance its bias through the simultaneous minimization or maximization of feature cardinality [73, 99, 123].

4.2.4 Biclustering

The concept of a bicluster has recently been introduced as a generalization of a cluster: a bicluster consists of a group of data items that are homogeneous (show similar patterns) within a subspace of the original feature space only. Biclustering algorithms [114] aim to identify such cluster structures in subspaces and are of increasing importance in the analysis of gene expression data. The aim in biclustering is the identification of maximally sized biclusters of a pre-defined low intra-cluster inhomogeneity score. The choice of this threshold is required, as the criterion of intra-cluster inhomogeneity is biased towards bicluster size and will decrease for decreasing cluster sizes. Evidently, the selection of a fixed threshold on the maximally allowed degree of inhomogeneity may have a strong impact on the final clustering result, and a more flexible optimization approach may be advantageous. One possibility would be the formulation of biclustering as a multi-objective optimization problem, as alluded to in [14]: “the optimization goal is to find one or several biclusters that are optimal with respect to their homogeneity and their size. These two objectives are usually competing”. In this way, the simultaneous minimization of intra-cluster homogeneity and maximization of bicluster size may be used as a means to counter-balance the bias present while avoiding the need for a fixed threshold parameter. The resulting Pareto front and sets of (partially) hierarchical solutions may additionally serve to learn more about the structure intrinsic to the data set.

4.2.5 Association rule mining

Association rule mining is the unsupervised equivalent to classification rule mining, thus it is not pattern-class relationships that are sought, but relationships between patterns in the data space. The quality of association rules is inherently difficult to assess, and a range of different objectives have been introduced in the literature [57]. These include measures related to confidence and coverage in classification rule mining, but also comprise methods assessing the complexity of the rules found. As several of these measures

are complementary and conflicting, their multiobjective optimization has been proposed by a number of authors [60, 97]. For example in [97], a five-objective formulation of the problem was suggested and the resulting multiobjective evolutionary algorithm was employed for the identification of an optimal set of association rules on a gene expression data set.

4.2.6 Multidimensional scaling

Next to clustering, rule discovery and feature selection, another important problem in unsupervised classification is the projection of a data set to lower-dimensional subspaces. Usually, a projection to a two- or three-dimensional subspace will be used, with the aim of obtaining a visualization of the data set that is interpretable by a human observer.

Multidimensional scaling (MDS, [31]) is an example of such a visualization technique. Given information about the dissimilarities between data items, MDS provides an embedding of these data into a multidimensional space of specified dimensionality, such that distances between data items are preserved, but the actual positions of individual data items are meaningless. In metric multidimensional scaling, the multidimensional space is required to be Euclidean, and an explicit functional connection between the dissimilarities and the distances needs to be defined. In methods of non-metric multidimensional scaling, both of these requirements are relaxed. A range of different methods for metric and non-metric multidimensional scaling exist, which differ in the loss function (optimization criterion) used.

Brusco [20] suggests that a multiobjective formulation of the multidimensional scaling problem may be advantageous for two different reasons. A first advantage would be the opportunity to consider a number of different loss functions simultaneously. This can be seen as an extension of Brusco and Stahl's earlier work on the multiobjective optimization of asymmetric unidimensional seriation problems [21]. Alternatively, the multiobjective framework could be used for multidimensional scaling in the presence of multiple dissimilarity matrices, where the same loss function is optimized with respect to the individual dissimilarity matrices.

4.3 Semi-supervised classification

In the above sections, we have seen that multiobjective optimization can be a useful tool both in unsupervised and supervised classification. In certain classification scenarios it can be advantageous to combine the advantages of unsupervised and supervised classification techniques, that is, to exploit both previ-

ous knowledge of class labels and the underlying data structure: semi-supervised approaches aim to do this. Through the combined use of labelled and unlabelled data it becomes possible to give a degree of external guidance to the classification algorithm, while still permitting intrinsic data structures to be taken into account. This is considered particularly useful when dealing with data sets consisting of a large number of unlabelled data items and relatively few labelled ones, and, more generally, in the case of very limited prior knowledge. For example, in cases where the classes within a particular data set are only partially known, additional ones may be identified by taking the data distribution into account. Also, due to the combination of two fundamentally different sources of information, semi-supervised approaches may be expected to be more robust than both unsupervised and supervised approaches, and may be less sensitive towards both annotation errors and the occlusion of data structures due to noise.

Data sets with the above properties are frequently encountered in application domains where the categorization of individual data items is accompanied by high computational, analytical or experimental costs. Initially introduced in the field of information retrieval, semi-supervised methods have now seen first application in post-genomic data analysis, in particular for gene functional classification [103, 108], protein classification [171] and the prediction of patient survival from gene expression data [8].

The different approaches to semi-supervised classification differ primarily in the emphasis paid to unlabelled and labelled data. Semi-supervised classifiers (such as transductive support vector machines [93]) are based on traditional classifiers, but aim to shift decision boundaries into areas of low density (as measured across the unlabelled data). In contrast, semi-supervised clustering methods build upon standard clustering methods, but modify these to direct the search towards clusters consistent with the class labels given. Frequently, this is done by means of constraints [65] or the adaptation of the clustering objective or distance function, which is transformed into a linear or non-linear combination between an unsupervised and supervised objective component [76, 158]. Given the need for the integration of these two information sources, as well as a lack of knowledge about the correct weighting and the respective correctness of the two, their simultaneous optimization within a multiobjective framework seems natural. In preliminary work on this subject [71, 72], promising results have been obtained. In particular, multiobjective formulations of both semi-supervised clustering and semi-supervised feature selection were shown to outperform

single-objective approaches based on the linear and non-linear combination of internal and external objective components.

5 Inverse problems

Inverse modelling problems arise in all those applications where the data generated by a (biological) process or system can be measured, and where we are aiming to infer the original system from the observed data [96]. The challenges typically encountered in these applications include noisy data, the integration of several types of data and the underdetermination of the inference problem at hand. In the following, we will review how multiobjective optimization can serve as a tool to tackle some of these issues.

5.0.1 Time series prediction

Time series prediction is commonly treated as a supervised classification problem, but may also be framed as an inverse modelling problem. In time series prediction, we are usually provided with a sequence of measurements describing the state of a system at regular intervals in time, and are aiming to extrapolate from this data to correctly predict future trends. Often, the process of extrapolation will rely on the inference of an appropriate model that correctly describes the observed data, and which we can use to model future behaviours of the system at hand. Evidently, the aim in time series prediction is the reliable prediction of this future behaviour, but in practice the reliability of the model can only be measured with respect to past observations. Once again, during optimization we therefore rely on the use of ‘proxy’ objectives that attempt to capture desirable aspects of the model. A straightforward objective in time series prediction is the root mean square error between the observed and the predicted data, but a range of other possible objectives exist, such as the relative absolute error (which uses a normalization based on random control data), or an objective assessing the prediction of directional trends [6], which may be of primary importance in certain application domains. In [55], a multiobjective evolutionary approach was proposed, which optimized neural networks with respect to two different error functions, and experimental results suggested a substantial performance advantage of the neural network models obtained by this training process.

5.1 Phylogenetic inference

Phylogenetic tree inference is a more specific example of an inverse modelling problem encountered in the

biochemical domain. Phylogenetic trees are used to describe and visualize the evolutionary relationships between species that are believed to have a common ancestor, and are an important tool in evolutionary biology [132]. Phylogenetic trees can be manually constructed based on prior knowledge, but we may also attempt to automatically infer them from sequential or structural information that is believed to reflect the relevant evolutionary similarities.

Existing approaches for phylogenetic tree inference can be classed into three major groups, namely distance matrix methods, maximum parsimony methods, and maximum likelihood methods [29, 52, 140]. All three types of approaches are usually based on the formulation of the problem as a single-objective optimization problem. These traditional approaches to phylogenetic tree inference do not take the existence of multiple data sets from different sources into account. While such data sets may often be noisy and partially conflicting, they can usually be assumed to complement each other and to be more informative in combination than on their own. It is therefore of high importance to develop methods that combine these data from different information sources, and, currently, this is most commonly done prior to or after the actual phylogenetic tree inference [40]. In very recent work, it has been suggested that multiobjective optimization may provide an alternative tool to integrate and trade-off such conflicting data during the inference process, and that such an approach may in fact be more robust and informative than the *a priori* or *a posteriori* integration currently used in the literature [133].

5.2 Gene regulatory networks

As a result of ongoing sequencing efforts, researchers working in the biological sciences today have access to the complete genomes of an increasing number of organisms. However, the functionality of many of these genes is still unknown, and an improved understanding of these organisms at the molecular level requires fundamentally more knowledge about the activation patterns of the individual genes. Specifically, we need to be able to know and to predict, at what level, at what time, under which conditions and at which location, specific genes are expressed in a given organism. The mechanism at the bottom of these patterns of gene expression is a complex interplay of interactions between DNA, RNA, proteins and metabolites (also referred to as a genetic regulatory system), which, at an abstract, simplified level, can be modelled as a network of inhibitory and stimulatory interactions between genes — a *gene regulatory network* (GRN).

The actual inference task involves parameterizing

a chosen network model in order to fit it to the given data, that is, finding a set of parameters with which the observed data can be perfectly or approximately reproduced. The identification of such a consistent model is, generally, achieved through the minimization of the root mean square error between the model output and the real data. However, this optimization task is typically highly underdetermined, i.e. there may exist (infinitely) many different GRNs that are consistent with the observed data, and the integration of additional information may be necessary to differentiate between these.

As discussed in several previous sections, multiobjective optimization can serve as a tool both for the flexible integration of topological constraints and prior knowledge, as well as the integration of data from several sources. Some preliminary work [102, 157, 159, 175] has exemplified its use for GRN inference and the results obtained seem highly promising. An example of its use for the integration of several data sets is given in [102], where the gene regulatory network controlling the flowering time in the rice *Oryza sativa* is optimized with respect to three different data sets: the first of these described RNA concentration levels under long-day conditions, the second one gave the equivalent results obtained under short-day conditions, and the third data set consisted of the actual flowering dates. In order to obtain a robust network model, the root mean square error with respect to all three of these data sets was simultaneously optimized.

Algorithms for the inference of gene regulatory networks are closely related to those for other inverse modelling tasks such as the inference of protein networks and metabolic pathways from experimental data. Clearly, similar problems arise in all of these modelling scenarios and the multiobjective approaches discussed above will be equally applicable in all of them. The relevance of multiple objectives is evident when considering tools such as the software tool YANA [150], which is aimed at the steady state analysis of metabolic networks. An effective steady state analysis may frequently require to take into account additional objectives, which may serve for the integration of available experimental data (such as enzyme expression levels observed in gene-chip experiments) or of prior biological knowledge (for example regarding enzyme activities or metabolic constraints). YANA currently allows the user to do this through the flexible integration of such objectives by means of user-specified weights.

6 Sequence and structure alignment

In this section, we will highlight applications related to the assessment of sequential and structural similarities of DNA and RNA macromolecules, as well as proteins. Tools for the assessment of similarity and the identification of related sequences or structures are amongst the most important tool in bioinformatics, as they may serve for the functional categorization of novel genes or proteins of unknown function. This is because sequential and structural similarities may provide evidence of evolutionary relationships between entities and may indicate shared or related functional properties and/or whether evolution is convergent or divergent.

6.1 Sequence alignment

Sequence alignment aims to arrange two or more DNA, RNA or protein sequences in a way that highlights their similarities. For this purpose, the sequences are set down one upon the other and are padded with gaps, so that the number of columns containing identical or related nucleotides is maximized [125].

Existing methods for sequence alignment are usually based on a single-objective optimization. Specifically, the computation of an alignment requires the definition of a substitution matrix and the gap penalties. The substitution matrix is of size $|C| \times |C|$, where $|C|$ is the size of the alphabet (which is four for RNA and DNA strands, and 21 for proteins). The values within the substitution matrix indicate the rewards and penalties respectively for the alignment of any two characters in the alphabet. Often, affine gap penalties are used, that is, the cost of inserting a sequence of k letters grows as $g(k) = A + Bk$, for positive constants A and B [124]. The substitution matrix and the gap scores are then linearly combined to provide an overall score assessing the quality of a given sequence alignment.

In the light of the difficulty of deriving suitable substitution matrices and gap penalties, a multiobjective approach to sequence alignment has been proposed in [138]. It is based on a modified form of dynamic programming and, instead of a scalar scoring function, the $\left(\frac{C \times (C+1)}{2} + 1\right)$ -dimensional vector containing all substitution values and gap penalties is optimized. The same authors have also investigated a bi-objective approach to sequence alignment, in which only the gap penalties are treated as a separate objective [137]. While a multiobjective formulation may be too expensive to use in certain practical applications, it may certainly prove a useful tool for a thorough analysis of the trade-offs inherent to the problem, and for

the use in the design of novel and improved substitution matrices.

An important distinction in sequence alignment is that between global and local alignments. Global alignment aims to align the sequences given along their entire range. In contrast, local alignment only tries to identify subregions of the sequences in which their patterns concur. As far as local sequence alignment is concerned, a further trade-off can be observed between the length of the patterns compared and the quality scores obtained: evidently, the number of adverse substitutions or of gap penalties tends to increase for longer alignments, i.e. there is a bias causing the preference of short alignments. In recent research [175], it has therefore been suggested that a simultaneous maximization of the score and the length may be useful. The potential of this approach has been illustrated using a multiobjective evolutionary approach for the identification of short interspersed repetitive element in the DNA sequence of *Tripa-nosoma cruzi*: the method obtained all of the solutions identified by alternative single-objective approaches and discovered additional efficient trade-offs between the two objectives used [175].

Local sequence alignment is also an important tool for the identification of sequence motifs, and this problem has been previously approached from a multiobjective perspective. In [128], sequence motifs for a wide range of specificities and sensitivities were identified by a simple enumeration strategy named EMOTIF. It was observed that the shape of the resulting Pareto front (see Figure 3) provides information about the structure of the set of sequences identified (such as clusters of sequences), and the motifs identified resulted in new hypotheses about the function of 172 proteins of unknown function in the yeast genome.

In [136], the identification of binding site motifs was formulated as a semi-supervised classification problem, in which the classification accuracy and the similarity to experimentally determined hypothetical motifs were simultaneously optimized. The results obtained were shown to outperform alternative unsupervised computational approaches.

In a similar vein, Cotik *et al.* [28] used a multiobjective evolutionary algorithm for the discovery of binding site motifs in promoters. The analysis of prokaryotic promoters reveals the presence of two-well conserved submotifs separated by a variable distance. However, the variability of both this distance and the individual submotifs hampers the design of algorithms for reliable promoter detection. Cotik *et al.* formulated the problem as a three-dimensional problem in which the match to each of the two different submotifs and the distance between them were simultaneously optimized.

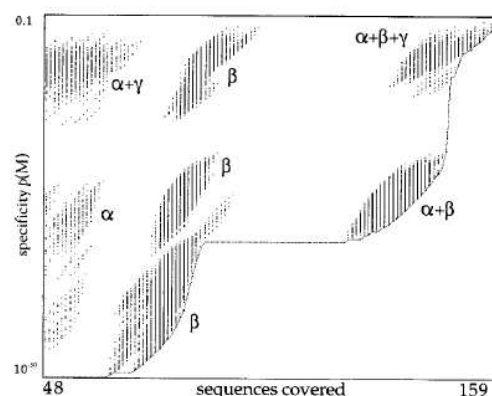


Figure 3: Enumeration of tubolin motifs by EMOTIF. Each motif is plotted as a dot in the figure where the horizontal axis gives the coverage of the motif, and the vertical axis plots its specificity. The curve towards the bottom right is the Pareto optimal curve, which represents the most specific motif at each level of sensitivity [128] (reprinted from [128]; awaiting permission from the National Academy of Sciences.)

6.2 Structure alignment

When sequence comparisons between pairs of RNAs or proteins fail to reveal significant relationships, structural comparisons between them may serve as a fallback mechanism. Due to the high evolutionary pressure on the structure, which determines the function of these macromolecules, structural similarities are preserved to a much higher degree than are sequence similarities and may, consequently, still be identified in the absence of any apparent sequence similarities.

Analogously to the case in sequence alignment, structural alignment can be performed both on a global or a local level. Global sequence alignment serves to provide information on the evolutionary distance between macromolecular structures, usually to predict functional relationships or evolutionary relationships between molecules. In contrast, alignments on a local basis are more specifically aimed at the identification of shared functionally active regions, so called pharmacophores. In applications involving local structure alignment or pharmacophore discovery, multiobjective approaches have been applied by a number of researchers [10, 30, 75, 88]. In [75], the problem was formulated as a bi-objective optimization problem in which the length of the alignment and the geometric fit had to be optimized. In [30], a three-objective formulation was used, which considered the conformational energy, as well as scores related to the volume and the features used. In other work, it has been suggested that

“optimal structure alignment requires a multiobjective optimization procedure, which is concerned with the minimization of a vector” [88], describing the length of the alignment, the geometric fit and the number of gaps introduced. Finally, local structure alignment for a set of molecules with respect to a given target molecule was formulated as a three-objective problem in [10], where the coverage of a given substructure, the size of the substructure, and the average deviation of the aligned molecules were optimized. In this application, the Pareto optimal set was obtained by a complete enumeration of the search space.

An application related to pharmacophore identification is the search for median molecules that are representative for a given set of macromolecules [89]. The search for such median molecules, which will share properties with all of the target molecules, is an important technique in computer-aided molecular design [145]. Traditional approaches to the evolution of median molecules use a single-objective approach based on the sum of the distances to all target molecules or the distance to an average description molecule. As pointed out by Brown *et al.* [19], such an approach may be suboptimal, as the the objective value may be dominated by one of the target molecules, and the resulting solution may be more representative of the corresponding molecule than of any of the other molecules. They therefore suggest a multiobjective evolutionary approach to the problem, in which the distance to each target molecule is treated as a separate objective [19]. The advantages of the algorithm proposed were demonstrated in an application involving the evolution of potential medians for sets of two target molecules (see Figure 4).

7 Structure prediction and design

In this section, we continue to focus on optimization tasks related to the structure of macromolecules, in particular the task of structure prediction and design. As mentioned previously, the functional properties of macromolecules derive from their three-dimensional shape (tertiary structure), which, in turn, is, predominantly, determined by their sequence of bases or amino acids (primary structure): specifically, the native tertiary structure of a macromolecule is assumed to correspond to its lowest free-energy conformation. In theory, this direct relation between sequence and structure, which has been known for several decades, allows for *in silico* structure/function prediction (for a given sequence), as well as the design of new RNAs and proteins (for a given structure/function).

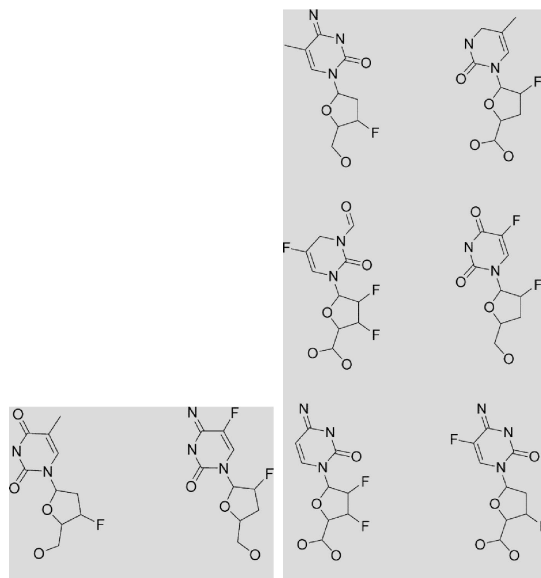


Figure 4: Left: Two similar molecules used as the target molecules. Right: A six-member representative, diverse subset of the evolved median molecules (reprinted from [19] with permission from Elsevier).

7.1 Protein structure prediction

In protein structure prediction, two fundamentally different approaches can be identified; those based on comparative modelling and those based on *de novo* modelling. Comparative modelling approaches predict structure based on that of homologous proteins: they are therefore only applicable if there are proteins with a high sequence similarity and known structure. In contrast to this, *de novo* modelling approaches can be applied to any protein sequence, but are, currently, less effective. The inherent difficulty of *de novo* protein structure prediction arises from two different issues: (1) the intricacy of formulating an energy function that realistically models the different local and global interactions contributing to protein folding, and (2) the size of the space of possible conformations, which cannot be explored exhaustively. Progress in *de novo* protein structure prediction therefore crucially relies on progress both in the design of appropriate (i.e. more accurate) energy functions and the development of specialized efficient sampling methods.

Traditionally, empirical energy functions consist of a sum of the different energetic components contributing to the process of the folding of the macromolecule. Only recently, researchers have aimed to identify the optimal weighting between these components using regression on a set of training data [167]. While this approach may lead to promising results, it

is unclear whether such a fixed linear combination can provide optimal discrimination for all types of macromolecules and in all regions of the search space. Multiobjective optimization may therefore be a more principled approach, and its use in protein structure prediction has recently been suggested by a number of authors [33, 36, 149].

Schulze-Kremer [149] suggested the decomposition of the energy function into a nine-dimensional vector. Among others, the torsion energy, the van der Waals energy, the electrostatic energy and a penalty energy term promoting compact folding patterns were taken into account and optimized using a multiobjective evolutionary algorithm. In [33, 36], a simpler two-objective formulation based on the CHARMM energy potential was proposed. In particular, local and non-local interactions were taken into account by separate objectives. Promising results were obtained in a comparison to other algorithms across five different proteins [33].

A conceptually different approach has been suggested in [134]. Here, a weighted-sum approach was used for the integration of different objectives in the prediction of protein structure from powder diffraction diagrams. Importantly, the objectives used related to fundamentally different types of information sources: one involved the minimization of the difference between the calculated and the measured diffraction patterns, while the other was based on the minimization of the potential energy of the system. In [134], the problem was optimized for a single weighting only (using single-objective simulated annealing), but in subsequent work, a more rigorous exploration of different weights was proposed [106]. As discussed in [106], the attractiveness of this approach derives from the properties of the search landscapes created by the two individual objectives. The objective based on the diffraction is assumed to have a distinct global minimum, which can be unambiguously identified; unfortunately, the search for this minimum is hampered by the presence of multiple local minima and a small basin of attraction. In contrast, the objective based on free energy is assumed to have multiple local minima, which cannot be reliably differentiated, but have large basins of attraction; one of these minima can be expected to coincide with the global minimum in the first objective. The combination of both two objectives may therefore serve as a way of facilitating the search problem by increasing the basin of attraction surrounding the optimal solution.

7.2 Protein and ligand docking

Protein-protein docking is closely related to protein structure prediction, but it aims to predict the proper-

ties of the complexes formed by two or more proteins. Specifically, for any given set of proteins, it aims to determine whether these proteins can bind to each other *in vivo*, identify the resulting spatial configuration, and assess the strength of the interaction. In contrast to protein-protein docking, ligand docking usually refers to the study of complexes formed by proteins in combination with smaller molecules.

Methods for protein and ligand docking use empirical energy functions very similar to those used in protein structure prediction and, thus, share the same motivation for a multiobjective approach. The use of multiobjective evolutionary algorithms for ligand docking was proposed in [109], suggesting that the multiobjective approach may be beneficial with respect to the exploration of the optimal balance between different components of the scoring function (which may result in the development of improved energy functions), and the exploration of the trade-offs between different energy functions.

7.3 Directed evolution

Directed evolution refers to the iterative production, evaluation and selection of macromolecules in an *in vitro* environment. The aim of directed evolution is to evolve molecules with one or a number of desired functional properties by a process inspired by the process of selective breeding. As discussed previously, the function of a macromolecule is determined by its three-dimensional structure. While this can, in theory, be predicted from the sequence of a macromolecule, such rational design methods remain very limited in practice. Directed evolution (purportedly) avoids the need for efficient structure prediction methods through the evaluation of the performance of the macromolecules *in vitro* [7].

Recently, the success of techniques of directed evolution has been demonstrated by a number of researchers [3, 144, 162]. These studies have considered different types of macromolecules and objectives, but have been limited to a single objective. However, it may be argued that, in many applications, the aim of directed evolution may be better described as a multiobjective optimization problem, for example if we aim to optimize the stability and the reaction rate of a given enzyme. A multiobjective evolutionary approach to the directed evolution of DNA sequences has been recently explored in [155]. The aim in this work was to obtain DNA sequences suitable for use in DNA computing, which results in a number of different objectives. Specifically, the authors identified four different groups of objectives: those aimed at (i) the prevention of undesired reactions, (ii) controlling the secondary

structure, (iii) controlling the chemical characteristics of the molecule, and (iv) restricting the use of particular bases. Using a multiobjective evolutionary algorithm, Shin *et al.* [155] implemented a six-objective version of the problem within a DNA sequence design system. The results obtained were compared to those obtained using traditional, single-objective formulations of the problem and, overall, the experiments indicated an increased robustness of the sequences generated using multiobjective optimization.

7.4 Peptide identification

Peptide identification aims at the identification of a protein based on its chromatogram returned by mass spectrometry. This can either be done through a comparison of the chromatogram to those available for a list of reference proteins, or in a *de novo* approach, where the sequence of the protein is predicted based on the properties of the chromatogram. *De novo* peptide identification relies on the availability of a quantitative model or scoring function that assesses the likelihood of the correspondence between a given peptide sequence and an experimentally observed chromatogram.

Malard *et al.* [115] suggested a multiobjective approach to *de novo* peptide identification. The advantages of such an approach are the availability of flexible constraint handling and the possibility to simultaneously optimize several (and possibly conflicting) scoring functions. This simultaneous optimization can be very useful due to the different simplifications and assumptions underlying different scoring functions. An exploration of the trade-offs between them may help to understand their relative strengths and weaknesses and to obtain a more robust overall performance. A proof-of-concept is presented in [115], but no comparisons to other methods are presented.

8 System optimization and experimental design

In the previous section we have seen that multiobjective optimization can be useful in the design of improved macromolecules. It may also be used to investigate the degree of optimality of naturally occurring biochemical systems or to design optimal biochemical processes, and these ideas have been explored in a number of papers.

8.1 Study of the optimality of biochemical systems

In [169], an experimentally derived kinetic model was optimized in order to study the trade-off between maximizing ethanol production in the yeast *Saccharomyces cerevisiae* while minimizing each of the internal metabolite concentrations. The optimization method employed was a linear programming approach, which was made publicly available [170].

The optimality of the heat shock response in cells was studied in [141]. Here, the term heat shock refers to the unfolding or misfolding of proteins, which is caused by sudden increases in temperature, and is counteracted *inter alia* by an increased production of chaperones and proteases, which repair or degrade the damaged proteins. The costs related to this repair process, and the costs related to the presence of damaged proteins can be seen as conflicting objectives, and Samad *et al.* used a model of Differential Algebraic Equations and numerical optimization methods to identify the corresponding Pareto front. The results indicated that the heat response exercised by cells is close to optimal in the Pareto sense.

8.2 Optimization of biochemical processes

Optimal protocols for polymerization processes have been studied by a number of authors [32, 43, 59, 121]. Polymerization is the reaction process that joins single molecules into polymer chains, and is of fundamental importance in the chemical and biochemical industries. Optimal polymerization is subject to a range of different conflicting objectives, including the polymerization degree and the reaction time, and it is thus naturally suited to a multiobjective approach. The problem has been approached using a number of different objectives and optimization algorithms.

Three other examples of the use of multiobjective optimization are applications related to the beer fermentation process [5], the citric acid fermentation of *Aspergillus niger* [116] and the production of gluconic acid [67]. The beer fermentation process was formulated as an eight-objective optimization problem comprising criteria such as the final ethanol concentration, the spoiling risk and the process time. Three of these were converted into fixed constraints, and the remaining five objectives were optimized using a multiobjective evolutionary algorithm. *Aspergillus niger* is a fungus used in the production of chemical compounds such as citric and gluconic acid. During fermentation, it produces a number of enzymes. In [116], catalase and protease were selected as the enzymes of interest, and

their final concentrations were optimized using a differential evolutionary algorithm. Gluconic acid is a carboxylic acid used as a food additive and in cleaning products. Its production was optimized with respect to the production rate, the final gluconic acid and the final substrate concentration. Further examples of the use of multiobjective optimization techniques for the optimization of biochemical process include investigations regarding optimal liver function [152] and the production of oil in the yeast *Yarrowia lipolytica* [126].

8.3 Experimental design

Combinatorial library design refers to the optimization of the collection of compounds to be used in a screening test for the identification of compounds that interact with a target enzyme or receptor. Despite the development of high-throughput screening tests, the proportion of potentially interesting compounds than can be subjected to the screening test in practice remains negligibly small. For this reason, a careful design of combinatorial libraries remains of fundamental importance in the field of drug design.

It has been appreciated by a number of researchers that effective combinatorial drug design involves the optimization of a number of conflicting design criteria, and may best be tackled using methods of multiobjective optimization. In [4], a multiobjective version of simulated annealing was used to optimize a range of eight different objectives, including the diversity, activity and selectivity of the compounds selected. In [61, 62, 63], a multiobjective evolutionary algorithm was used and applied for two and for five different objectives. The bi-objective problem was based on the maximization of diversity and the minimization of the root mean square deviation to a lead molecule. In the five-objective formulation three additional objectives were integrated, namely, the frequencies of the occurrence of rotatable bonds, hydrogen bond donors and hydrogen bond acceptors.

Single-nucleotide polymorphisms (SNPs) are sequences of the genome at which members of a species differ in a single nucleotide. SNPs are thought to have a major impact on the predisposition towards diseases and, therefore, play a major role in biomedical and pharmaceutical research. Due to constraints in terms of costs and throughput, researchers will typically only be able to use a fraction of the SNPs available for use in genomics studies. Analogously to the case of combinatorial library design, they will need to optimize the subset chosen with respect to a number of conflicting criteria. These include the reliability of the chosen SNPs, biochemical factors, as well as intellectual property costs and the expected

effectiveness of the subset for identifying disease loci. In [81], the problem was formulated as a bi-objective optimization problem in which the cost and the expected effectiveness of the SNPs were simultaneously optimized. A multiobjective evolutionary approach to the problem was developed, and its use was illustrated on a real-world data set.

The multiobjective nature of the problem of oligonucleotide-design for the use in microarray and PCR experiments, is pointed out in [147, 148]. The oligonucleotides used as the query sequences in these types of experiments are required to minimize the likelihood of undesired cross-reactions caused through secondary structure formation, cross-hybridization with non-target sequences or local complementarity with other query sequences. In [147, 148], this problem was tackled by a multi-level filtering approach, but a genuine Pareto-based approach as used in Probe design [107] may be considered preferable.

A final example of a multiobjective approach to experimental design is multiplex PCR assay design. In multiplex PCR, oligonucleotides need to be assigned to different tubes in order to group those that are compatible (have no risk of cross-hybridization) and separate those that are not compatible. Usually, a number of conflicting objectives arise during this design including the minimization of the number of tubes, the maximization of tube size uniformity and the maximization of the coverage of all nucleotides to be considered [135].

8.4 Instrument optimization

In this last subsection of the survey, we will consider instrument optimization, which is another important application domain of multiobjective algorithms.

A number of papers in the literature have explored multiobjective approaches to the planning of radiation therapy [48, 66, 146]. For example, in [146] the number, orientation and weights of the beams in intensity-modulated radiation therapy were optimized with respect to four different objectives. These considered the variation around the prescription dose, the dose values for normal tissue and for organs at risk, and the total number of beams used. A simpler bi-objective version of the problem was studied in [48].

In some instrument configuration problems, the evaluation of a given solution may not be possible without the actual use of the instrument, and the optimization therefore requires the use of a so-called closed loop procedure. Here, closed-loop refers to the fact that the optimization algorithm directly interacts with the analytical instrument, which is involved in the evaluation of every single candidate solution. An example of such an application has been discussed in [130].

In this work, a mass spectrometer was optimized subject to three different objectives: specifically, the signal to noise ratio, the processing time and the number of peaks were optimized using a multiobjective evolutionary algorithm. Due to the involvement of actual instruments, closed-loop optimization scenarios are usually very costly in terms of running time and material and are therefore classed into the category of expensive optimization problems. Specialized algorithms for single-objective expensive optimization problem exist [94, 90], which attempt to restrict the number of evaluations needed, and equivalent approaches have recently been proposed for multiobjective optimization [127, 100].

Part III — Synthesis

9 Discussion and outlook

In the previous sections, we have seen that multiobjective optimization has wide applications in computational biology and bioinformatics. The performance gains and flexibility afforded by multiobjective optimization have been illustrated in a range of initial studies, but in many of these problem domains the full potential of multiobjective approaches in comparison to the current state-of-the-art techniques remains to be explored. We hope that this review will serve as a starting point for computer science researchers interested in exploring the use of multiobjective optimization in the field of biology, and will also improve the awareness about these techniques within the biological communities.

9.1 Classification of problems

The common denominator between all of the applications reviewed in this paper is that *some form* of multiobjective optimization is potentially useful or has already been demonstrated. The reasons underlying the use of multiobjective optimization differ widely across these application domains, and we believe this aspect to be more interesting and revealing than a distinction between the specific techniques used. In Section 3, we have put forward five different motivations for the use of multiobjective optimization, and have referred to these different contexts along the way. We will now attempt to summarise this classification of the different application areas. Note that several of the problems considered can potentially fall into more than one category, dependent on the specific viewpoint taken. In Table 1, we have only indicated those categorizations that correspond to views taken in the literature reviewed within the scope of this paper, but this cate-

gorization is not final and, for some of these problems, additional multiobjective viewpoints may be proposed in the future.

9.2 Theoretical considerations

We hope that this review has also highlighted that the different applications considered do not only differ in the reasons underlying the use of multiobjective optimization, but may also differ in the way the final set of solutions is treated. In certain applications, such as instrument optimization or clustering, the decision maker will be given the approximation set obtained with the main aim of identifying a single preferred solution. In such applications, we may be able to obtain additional valuable information from the shape of the Pareto front obtained, but we are ultimately interested in obtaining a single solution. In other applications, in contrast, it is the entire set of Pareto optimal solutions that is of interest and that will be selected for future use. Examples of this are the use of the concept of Pareto optimality for gene filtering (see Section 4.1.4), or its use for the generation of diverse classifiers for the integration in an ensemble classifier (see Section 4.1.5).

These differences are interesting, as they affect the conclusions that can be drawn about the quality of the solutions obtained, when comparing the solutions of a single- and a multiobjective optimization problem based on the same objectives. The principles of Pareto optimality guarantee that the optima of the single-objective problems will also form part of the Pareto front for the corresponding multiobjective problem. Hence, the Pareto set is a superset of the set of optima of the individual single-objective problems, and is therefore guaranteed to contain solutions at least as good as those defined by the single-objective problems. Here, the availability of additional solutions (corresponding to trade-offs between the individual objectives) is usually seen as an advantage of multiobjective optimization, but is difficult to quantify, as all of the solutions in the Pareto front (including the solutions to the single-objective problems) are incomparable with respect to each other.

Note, however, that this situation changes, as we move towards applications, in which sets of solutions are to be identified. If a set of N solutions is selected based on single-objective optimization, this will typically be done by selecting the N solutions scoring highest under an individual objective (e.g. a t-test in gene filtering). Note that only the best solution out of the resulting set is guaranteed to be Pareto optimal, whereas the solutions ranked 2 to N are not guaranteed to lie on the Pareto front. In contrast, for a number of solutions N smaller than the number of Pareto optima

Table 1: Categorization of the main applications discussed. Some categories, such as “Multi-objectivization” are currently under-represented, but we believe that further applications of this type will emerge as the use of multiobjective optimization propagates in the biological domain. The highest number of entries can currently be observed for the category “Proxy”, which contains all those applications in which a ‘gold standard’ (e.g. the best possible classifier over ‘future’ data, the true structure of a protein, the true network of regulatory relationships or the true evolutionary relationship between sequences/structures) exists, which we would like to reach, but do not have direct access to during the optimization process.

Problem	Standard MOO	Bias	Multiple data sources	Proxies	Multi-objectivization
ROC curves	—	—	—	✓	—
Rule mining	—	—	—	✓	—
Accuracy vs. complexity	—	—	—	✓	—
Supervised feature selection	—	—	—	✓	—
Ensemble learning	—	—	—	✓	—
Clustering	—	✓	✓	✓	—
Cluster validation	—	—	—	✓	—
Unsupervised feature selection	—	✓	—	—	—
Biclustering	—	✓	—	—	—
Association rule mining	—	—	—	✓	—
Multidimensional scaling	—	—	✓	✓	—
Semi-supervision	—	—	✓	—	—
Time series prediction	—	—	—	✓	—
Phylogenetic trees	—	—	✓	—	—
Gene regulatory networks	—	—	✓	✓	—
Sequence alignment	—	✓	—	✓	—
Structure alignment	—	✓	—	✓	—
Structure prediction	—	—	—	✓	✓
Ligand docking	—	—	—	✓	—
Directed evolution	✓	—	—	—	—
Peptide identification	—	—	—	✓	—
Biochemical systems	✓	—	—	—	—
Biochemical processes	✓	—	—	—	—
Experimental design	✓	—	—	—	—
Instrument optimization	✓	—	—	—	—

P , a multiobjective approach will allow one to select N Pareto-optimal solutions. Hence, in a comparison of such sets of solutions, a large fraction of the solutions identified by a single-objective approach would be expected to be dominated by those identified by the multiobjective approach. So, in this setting a multiobjective approach may have an even clearer advantage than when only a single solution is sought.

Evidently, in either case, some assumptions are necessary for these theoretical advantages to hold. Firstly, we must assume the use of exact optimization techniques, that is, both the single- and the multiobjective optimization technique used must be guaranteed to find the optimal solutions. Secondly, the ranking returned by the objectives in combination with the principle of Pareto optimality must be assumed to reflect the true quality of the solutions. As mentioned previously, if either noisy or vague ‘proxy’ objectives

are used, the best solutions cannot be guaranteed to lie in the Pareto front, and, consequently, no guarantees on the relative performance of the single- and multiobjective optimization techniques can be given.

Despite these limitations in practice, much of the literature surveyed in this paper has explored the use of heuristic techniques for multiobjective optimization, including its application in the presence of noise and/or when relying on ‘proxy’ objectives. The results obtained in most of this work have been very encouraging and would seem to indicate that the advantages of multiobjective optimization do hold in practice.

9.3 Evaluation function development

We have mentioned previously that the computational complexity of multiobjective optimization may prove prohibitive in some practical settings, such as se-

quence alignment or protein structure prediction for large molecules. However, we believe that despite such limitations regarding its potential application within standard processing or search tools, multiobjective optimization may prove a valuable tool in the design and development of improved and efficient software for these bioinformatics tasks. In particular, it will allow us to learn about the trade-offs in a problem and to identify recurring patterns in the Pareto fronts, the knowledge and understanding of which may help in the formulation of novel and better single-objective problem formulations.

9.4 Visualization and solution identification

The successful application of multiobjective optimization in biological and other applications requires the development of advanced methods for the visualization of the Pareto front and for the support of the decision maker in selecting solutions from the Pareto front.

Evidently, straightforward visualizations of the Pareto front are only possible in two or three dimensions, and a representation of the solutions obtained and the relationships between them becomes much more intricate for higher dimensions. To date, only few methods for effective visualization have been introduced that can deal with the truly multi-dimensional case (one of the main examples is a parallel axis plot [153]), and visualization remains a major topic for future research.

The identification of promising solutions from Pareto front approximations has been investigated in several recent works [16, 35, 42, 118, 160]. However, these papers have generally dealt with the reduction of the size of the approximation set in the absence of additional knowledge about user preferences, and this is done to guide or focus the search towards the (potentially) more important areas. An alternative approach is to first obtain the most complete Pareto front approximation set possible, and then to, *a posteriori*, reduce this set to a single solution. This approach has been investigated for several unsupervised classification tasks [68, 70, 73], but remains to be explored in many other application domains.

10 Conclusion

This paper has outlined the wide applicability of multiobjective optimization in biological problem domains, and, where available, has illustrated its potential with references to existing results from the literature. Rather than differentiating between differences in the optimization techniques used, we have opted to

emphasize differences in the reasons underlying the attractiveness of multiobjective approaches in different problem domains. We hope that this viewpoint will help to provide additional insight into the advantages afforded by multiobjective optimization with regard to the applications listed and/or additional problems encountered in the field.

Acknowledgments

JH gratefully acknowledges support of a doctoral scholarship from the German Academic Exchange Service and the Gottlieb Daimler- and Karl Benz-Foundation, Germany. JK is supported by a David Phillips Fellowship from the Biotechnology and Biological Sciences Research Council (BBSRC), UK.

References

- [1] H. A. Abbass. Pareto neuro-evolution: Constructing ensemble of neural networks using multi-objective optimization. In *Proceedings of the IEEE Conference on Evolutionary Computation*, pages 2074–2080. IEEE Press, Anaheim, CA, 2003.
- [2] H. A. Abbass. Speeding up backpropagation using multiobjective evolutionary algorithms. *Neural computation*, 15:2705–2726, 2003.
- [3] M. M. Altamirano and J. M. Blackburn, C. Aguayo, and A. R. Fersht. Directed evolution of new catalytic activity using the alpha/beta-barrel scaffold. *Nature*, 403:617–622, 2000.
- [4] D. K. Agrafiotis. Multiobjective optimization of combinatorial libraries. *IBM Journal of Research and Development*, 45(3/4):545–566, 2001.
- [5] B. Andres-Toro, J. M. Giron-Sierra, P. Fernandez-Blanco, J. A. Lopez-Orozco, and E. Besada-Portas. Multiobjective optimization and multiobjective control of the beer fermentation process with the use of evolutionary algorithms. *Journal of Zhejiang University SCIENCE*, 5(4):378–389, 2004.
- [6] J. Armstrong and F. Collopy. Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting*, 8(1):69–80, 1992.
- [7] F. H. Arnold. When blind is better: protein design by evolution. *Nature Biotechnology*, 16:617–618, 1998.
- [8] E. Bair and R. Tibshirani. Semi-supervised methods to predict patient survival from gene expression data. *PLoS Biol.*, 2(4):0511–0521, 2004.
- [9] E. Bauer and R. Kohavi. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine Learning*, 36(1-2):105–139, 1999.

- [10] D. Baum. Multiple semi-flexible 3D superposition of drug-sized molecules. In *Proceedings of the First International Symposium on Computational Life Sciences*, pages 198–207. Springer-Verlag, Berlin, Germany, 2005.
- [11] R. Bayardo and R. Agrawal. Mining the most interesting rules. In *Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 145–154, 1999.
- [12] V. Belton and T. J. Stewart. *Multiple Criteria Decision Analysis: An Integrated Approach*. Springer-Verlag, Berlin, Germany, 2002.
- [13] S. Bleuler, M. Brack, L. Thiele, and E. Zitzler. Multi-objective genetic programming: Reducing bloat using SPEA2. In *Proceedings of the IEEE Congress on Evolutionary Computation*, pages 536–543. IEEE Press, Anaheim, CA, 2001.
- [14] S. Bleuler, A. Prelic, and E. Zitzler. An EA framework for biclustering of gene expression data. In *Proceedings of the IEEE Congress on Evolutionary Computation*, pages 166–173. IEEE Press, Anaheim, CA, 2004.
- [15] P. E. Blower and K. P. Cross. Decision tree methods in pharmaceutical research. *Curr. Top. Med. Chem.*, 6(1):31–9, 2006.
- [16] J. Branke, K. Deb, H. Dierolf, and M. Osswald. Finding knees in multi-objective optimization. In *Proceedings of the Eighth International Conference on Parallel Problem Solving from Nature*, Birmingham, UK, 2004. Springer. To appear.
- [17] L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.
- [18] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
- [19] N. Brown, B. McKay, F. Gilardoni, and J. Gasteiger. A graph-based genetic algorithm and its application to the multiobjective evolution of median molecules. *J. Chem. Inf. Comput. Sci.*, 44:1079–1087, 2004.
- [20] M. Brusco. Multiobjective multidimensional scaling. In *DIMACS Workshop on Algorithms for Multidimensional Scaling*. DIMACS Center, Rutgers University, Piscataway, 2001. <http://dimacs.rutgers.edu/Workshops/>.
- [21] M. Brusco and S. Stahl. An interactive multiobjective programming approach to combinatorial data analysis. *Psychometrika*, 66(1):5–24, 2001.
- [22] O. Camoglu, T. Can, A. K. Singh, and Y. F. Wang. Decision tree based information integration for automated protein classification. *J. Bioinform. Comput. Biol.*, 3(3):717–742, 2005.
- [23] A. Chandra and X. Yao. Divace: Diverse and accurate ensemble learning algorithm. In *Proceedings of the Fifth International Conference on Intelligent Data Engineering and Automated Learning*, pages 619–625. Springer-Verlag, Berlin, Germany, 2004.
- [24] C. A. Coello Coello, David A. Van Veldhuizen, and Gary B. Lamont. *Evolutionary algorithms for Solving Multi-Objective Problems*. Kluwer Academic Publishers, New York, NY, 2002.
- [25] F. Coenen, P. Leng, and L. Zhang. Threshold tuning for improved classification association rule mining. In *Proc PAKDD*. Springer, 2005.
- [26] J. Cohen. Bioinformatics — an introduction for computer scientists. *ACM Computing Surveys*, 36(2):122–158, 2004.
- [27] David W. Corne, Kalyanmoy Deb, Peter J. Fleming, and Joshua D. Knowles. The good of the many outweighs the good of the one: evolutionary multi-objective optimization. *Connections. The Newsletter of the IEEE Neural Networks Society*, 1(1):9–13, 2003.
- [28] V. Cotik, R. R. Zaliz, and I. Zwir. A hybrid promoter analysis methodology for prokaryotic genomes. *Fuzzy Sets and Systems*, 2005.
- [29] C. Cotta and P. Moscato. Inferring phylogenetic trees using evolutionary algorithms. In *Proceedings of the 7th International Conference on Parallel Problem Solving from Nature*, pages 720–729. Springer-Verlag, Berlin, Germany, 2002.
- [30] S. J. Cottrell, V. J. Gillet, R. Taylor, and D. J. Wilton. Generation of multiple pharmacophore hypotheses using multiobjective optimisation techniques. *Journal of Computer-Aided Molecular Design*, 18:665–682, 2004.
- [31] T. F. Cox and M. A. A. Cox. *Multidimensional Scaling*. Chapman and Hall, New York, NY, 2001.
- [32] S. Curteanu, F. Leon, and D. Galea. Alternatives for multiobjective optimization of a polymerization process. *Journal of Applied Polymer Science*, 2006. To appear.
- [33] V. Cutello, G. Narzisi, and G. Nicosia. A multi-objective evolutionary approach to the protein structure prediction problem. *J. R. Soc. Interface*, 3(6):139–151, 2006.
- [34] M. B. Dale and P. T. Dale. Classification with multiple dissimilarity matrices. *Coenoses*, 9(1):1–13, 1994.
- [35] I. Das. On characterizing the ‘knee’ of the Pareto curve based on normal-boundary intersection. *Structural Optimization*, 18:107–115, 1999.

- [36] R. O. Day, J. B. Zydallis, G. B. Lamont, and R. Pachter. Solving the protein structure prediction problem through a multiobjective genetic algorithm. *Nanotechnology*, 2:32–35, 2002.
- [37] E. D. De Jong, R. A. Watson, and J. B. Pollack. Reducing bloat and promoting diversity using multi-objective methods. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 11–18. Morgan Kaufmann Publishers, San Francisco, CA, 2001.
- [38] B. de la Iglesia, M. S. Philpott, A. J. Bagnall, and V. J. Rayward-Smith. Data mining using multi-objective evolutionary algorithms. In *Proceedings of the IEEE Congress on Evolutionary Computation*, pages 1552–1559. IEEE Press, Anaheim, CA, 2003.
- [39] B. de la Iglesia, A. P. Reynolds, and V. J. Rayward-Smith. Developments on a multi-objective meta-heuristic (MOMH) algorithm for finding interesting sets of classification rules. In *Proceedings of the Third International Conference on Evolutionary Multi-Criterion Optimization*, pages 826–840. Springer-Verlag, Berlin, Germany, 2005.
- [40] A. de Queiroz, M. J. Donoghue, and J. Kim. Separate versus combined analysis of phylogenetic evidence. *Ann. Rev. Eco. Syst.*, 26:657–681, 1995.
- [41] K. Deb. *Multi-Objective Optimization using Evolutionary Algorithms*. John Wiley & Sons, Chichester, UK, 2001.
- [42] K. Deb. *Multi-objective evolutionary algorithms: Introducing bias among Pareto-optimal solutions*, pages 263–292. Springer-Verlag, London, UK, 2003.
- [43] K. Deb, K. Mitra, R. Dewri, and S. Majumdar. Towards a better understanding of the epoxy polymerization process using multi-objective evolutionary computation. *Chemical Engineering Science*, 59(20):4261–4277, 2004.
- [44] K. Deb and A. Raji Reddy. Reliable classification of two-class cancer data using evolutionary algorithms. *Biosystems*, 72(1–2):111–129, 2003.
- [45] J. G. Dy and C. E. Brodley. Feature selection for unsupervised learning. *Journal of Machine Learning Research*, 5(5):845–889, 2004.
- [46] James S. Dyer, Peter C. Fishburn, Ralph E. Steuer, Jyrki Wallenius, and Stanley Zionts. Multiple criteria decision making, multiattribute utility theory: The next ten years. *Management Science*, 38(5):645–654, 1992.
- [47] M. Ehrgott. *Multicriteria Optimization*. Number 491 in Lecture Notes in Economics and Mathematical Systems. Springer-Verlag, Berlin, Germany, 2000.
- [48] M. Ehrgott and R. Johnston. Optimisation of beam directions in intensity modulated radiation therapy planning. *OR Spectrum*, 25(2):251–264, 2003.
- [49] Matthias Ehrgott and Xavier Gandibleux. A survey and annotated bibliography of multiobjective combinatorial optimization. *OR Spectrum*, 22(4):425–460, 2000.
- [50] F. Esposito, D. Malerba, and G. Semeraro. A comparative analysis of methods for pruning decision trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5):476–491, 1997.
- [51] R. M. Everson and J. E. Fieldsend. Multi-class ROC analysis from a multi-objective optimisation perspective. *Pattern Recognition Letters, Special Number on ROC Analysis in Pattern Recognition*, 2006. To appear.
- [52] D. Felsenstein. *Inferring Phylogenies*. Sinauer Associates, Inc, Sunderland, MA, 2004.
- [53] K. Y. Feng, Y. D. Cai, and K. C. Chou. Boosting classifier for predicting protein domain structural class. *Biochem. Biophys. Res. Commun.*, 334(1):213–217, 2005.
- [54] A. Ferligoj and V. Batagelj. Direct multicriterion clustering. *Journal of classification*, 9:43–61, 1992.
- [55] J. E. Fieldsend and S. Singh. Pareto evolutionary neural networks. *IEEE Transactions on Neural Networks*, 16(2):338–354, 2005.
- [56] Carlos M. Fonseca and Peter J. Fleming. Genetic algorithms for multiobjective optimization: formulation, discussion and generalization. In *Proceedings of the Fifth International Conference on Genetic Algorithms*, pages 416–423. Morgan Kaufman Publishers, San Francisco, CA, 1993.
- [57] A. Freitas. On rule interestingness measures. *Knowledge-based systems journal*, 12(5-6):309–315, 1999.
- [58] A. A. Freitas. A critical review of multi-objective optimization in data-mining: a position paper. *SIGKDD Explorations*, 6(2):77–86, 2004.
- [59] S. Garg and S. K. Gupta. Multiobjective optimization of a free radical bulk polymerization reactor using genetic algorithm. *Marcromol. Theory Simul.*, 8:46–53, 1999.
- [60] A. Ghosh and B. Nath. Multi-objective rule mining using genetic algorithms. *Information Sciences*, 163:123–133, 2004.
- [61] V. J. Gillet. Designing combinatorial libraries optimized on multiple objectives. *Methods in Molecular Biology*, 275:335–54, 2004.

- [62] V. J. Gillet, W. Khatib, P. Willett, P. J. Fleming, and D. V. S. Green. Combinatorial library design using a multiobjective genetic algorithm. *Journal of Chemical Information and Computer Science*, 42:375–385, 2002.
- [63] V. J. Gillet, P. Willett, P. J. Fleming, and D. V. S. Green. Designing focused libraries using MoSELECT. *J. Mol. Graph. Model.*, 20(6):491–8, 2002.
- [64] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, 286:531–537, 1999.
- [65] A. D. Gordon. A survey of constrained classification. *Computational Statistics & Data Analysis*, 21:17–29, 1996.
- [66] O. C. L. Haas, K. J. Burnham, and J. A. Mills. Optimization of beam orientation in radiotherapy using planar geometry. *Phys. Med. Biol.*, 43:2179–2193, 1998.
- [67] H. Halsall-Whitney, D. Taylor, and J. Thibault. Multicriteria optimization of gluconic acid production using net flow. *Bioprocess. Biosyst. Eng.*, 25:299–307, 2003.
- [68] J. Handl and J. Knowles. Exploiting the trade-off: the benefits of multiple objectives in data clustering. In *Proceedings of the Third International Conference on Evolutionary Multicriterion Optimization*, pages 547–560. Springer-Verlag, Berlin, Germany, 2005.
- [69] J. Handl and J. Knowles. *Multiobjective clustering and cluster validation*, chapter 2, pages 21–48. Studies in Computational Intelligence, Springer-Verlag, Berlin, Germany, 2005.
- [70] J. Handl and J. Knowles. An evolutionary approach to multiobjective clustering. *IEEE Transactions on evolutionary computation (conditionally accepted)*. To appear., 2006.
- [71] J. Handl and J. Knowles. On semi-supervised clustering via multiobjective optimization. Technical Report TR-COMPSYSBIO-2006-02, Manchester Interdisciplinary Biocentre, University of Manchester, UK, 2006. Available from <http://dbk.ch.umist.ac.uk/handl/publications.html>.
- [72] J. Handl and J. Knowles. Semi-supervised feature selection via multiobjective optimization. Technical Report TR-COMPSYSBIO-2006-03, Manchester Interdisciplinary Biocentre, University of Manchester, UK, 2006. Available from <http://dbk.ch.umist.ac.uk/handl/publications.html>.
- [73] J. Handl and J. Knowles. Feature subset selection in unsupervised learning via multiobjective optimization. *International Journal on Computational Intelligence Research*, 2006. To appear. Available from <http://dbk.ch.umist.ac.uk/handl/publications.html>.
- [74] J. Handl, J. Knowles, and D. B. Kell. Computational cluster validation in post-genomic data analysis. *Bioinformatics*, 21(15):3201–3212, 2005.
- [75] S. Handschuh, M. Wagener, and J. Gasteiger. Superposition of three-dimensional chemical structures allowing for conformational flexibility by a hybrid method. *Journal of Chemical Information and Computer Sciences*, 38:220–232, 1998.
- [76] D. Hanisch, A. Zien, R. Zimmer, and T. Lengauer. Co-clustering of biological networks and gene expression data. *Bioinformatics*, 18:145–154, 2002.
- [77] Michael Pilegaard Hansen and Andrzej Jaszkievicz. Evaluating the quality of approximations to the non-dominated set. Technical Report IMM-REP-1998-7, Technical University of Denmark, 1998.
- [78] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: data mining, inference and prediction*. Springer-Verlag, Berlin, Germany, 2001.
- [79] A. Hero and G. Fleury. Pareto-optimal methods for gene analysis. *Journal of VLSI Signal Processing*, 2003.
- [80] A. O. Hero, G. Fleury, A. J. Mears, and A. Swaroop. Multicriteria gene screening for analysis of differential expression with DNA microarrays. *Journal on Applied Signal Processing*, 1:43–52, 2004.
- [81] R. M. Hubley, E. Zitzler, and J. C. Roach. Evolutionary algorithms for the selection of single nucleotide polymorphisms. *BMC Bioinformatics*, 4(30), 2003.
- [82] M. Huerta, F. Haseltine, Y. Liu, G. Downing, and B. Seto. NIH working definition of bioinformatics and computational biology, 2000.
- [83] Ch. Igel. Multiobjective model selection for support vector machines. In *Proceedings of the Third International Conference on Evolutionary Multi-Criterion Optimization*, pages 534–546. Springer-Verlag, Berlin, Germany, 2005.
- [84] H. Ishibuchi and T. Yamamoto. Effects of three-objective genetic rule selection on the generalization ability of fuzzy rule-based systems. In *Proceedings of the second International Conference on Evolutionary Multi-Criterion Optimization*, pages 608–622. Springer-Verlag, Berlin, Germany, 2003.
- [85] H. Ishibuchi and T. Yamamoto. Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining. *Fuzzy Sets and Systems*, 141(1):59–88, 2004.

- [86] A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: a review. *ACM Computing Surveys*, 31(3):264–323, 1999.
- [87] M. T. Jensen. Guiding single-objective optimization using multi-objective methods. In *Applications of Evolutionary Computation*, pages 268–279. Springer-Verlag, Berlin, Germany, 2003.
- [88] Y. Jia, T. G. Dewey, I. N. Shindyalov, and P. E. Bourne. A new scoring function and associated statistical significance for structure alignment by CE. *Journal of computational biology*, 11(5):787–799, 2004.
- [89] X. Jiang, A. Muenger, and H. Bunke. On median graphs: Properties, algorithms, and applications. *IEEE Trans. Patt. Anal. Mach. Intell.*, 23:1144–1151, 2001.
- [90] Y. Jin. A comprehensive survey of fitness approximation in evolutionary computation. *Soft computing*, 9(1):3–12, 2005.
- [91] Y. Jin, editor. *Multi-Objective Machine Learning*, volume 16 of *Studies in Computational Intelligence*. Springer-Verlag, Berlin, Germany, 2006. (In press).
- [92] Y. Jin, T. Okabe, and B. Sendhoff. Neural network regularization and ensembling using multi-objective evolutionary algorithms. In *Proceedings of the IEEE Congress on Evolutionary Computation*, pages 1–8. IEEE Press, Anaheim, CA, 2004.
- [93] T. Joachims. Transductive inference for text classification using support vector machines. In *Proceedings of 16th International Conference on Machine Learning*, pages 200–209. Morgan Kaufmann Publishers, San Francisco, CA, 1999.
- [94] D. Jones, M. Schonlau, and W. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 12(4):455–492, 1998.
- [95] D. F. Jones, S. K. Mirrazavi, and M. Tamiz. Multi-objective meta-heuristics: An overview of the current state-of-the-art. *European Journal of Operational Research*, 137(1):1–9, 2002.
- [96] D. B. Kell. Metabolomics, modelling and machine learning in systems biology — towards an understanding of the languages of cells. The 2005 Theodor Buecher Lecture. *FEBS Journal*, 273:873–894, 2006.
- [97] M. Khabzaoui, C. Dhaenens, and E.-G. Talbi. A multicriteria genetic algorithm to analyze microarray data. In *Proceedings of the IEEE Congress on Evolutionary Computation*, volume 2, pages 1874–1881, 2004.
- [98] D. Kim. Structural risk minimization on decision trees using an evolutionary multiobjective optimization. In *Proceedings of the Seventh European Conference on Genetic Programming*, pages 338–348. Springer-Verlag, Berlin, Germany, 2004.
- [99] Y. Kim, W. N. Street, and F. Menczer. Evolutionary model selection in unsupervised learning. *Intelligent Data Analysis*, 6(6):531–556, 2002.
- [100] J. Knowles. Parego: A hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. *IEEE Transactions on Evolutionary Computation*, 10(1):50–66, 2006.
- [101] J. D. Knowles, R. A. Watson, and D. W. Corne. Reducing local optima in single-objective problems by multi-objectivization. In *Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization*, pages 269–283. Springer-Verlag, Berlin, Germany, 2001.
- [102] P. Koduru, S. Das, S. Welch, and J. L. Roe. A multi-objective GA-simplex hybrid approach for gene regulatory network models. In *Proceedings of the IEEE Congress on Evolutionary Computation*, pages 2084–2090. IEEE Press, Anaheim, CA, 2004.
- [103] M.-A. Krogel and T. Scheffer. Effectiveness of information extraction, multi-relational, and semi-supervised learning for predicting functional properties of genes. In *Proceedings of the Third IEEE International Conference on Data Mining*, pages 569–573. IEEE Press, Anaheim, CA, 2003.
- [104] V. Kumar and P. Rockett. Multiobjective genetic algorithm partitioning for hierarchical learning of high-dimensional pattern spaces: a learning-follows-decomposition strategy. *IEEE Transactions on Neural Networks*, 9(5), 1998.
- [105] M. A. Kupinski and M. A. Anastasio. Multiobjective genetic optimization of diagnostic classifiers with implications for generating receiver operating characteristic curves. *IEEE Transactions on Medical Imaging*, 18(8):675–685, 1999.
- [106] O. J. Lanning, S. Habershon, K. D. Harris, R. L. Johnston, B. M. Kariuki, E. Tedesco, and G. W. Turner. Definition of guiding function in global optimization: a hybrid approach combining energy and r-factor in structure solution from powder diffraction data. *Phys. Lett.*, 317:296–303, 2000.
- [107] I.-H. Lee, S. Kim, and B. T. Zhang. Multi-objective evolutionary probe design based on thermodynamic criteria for HPV detection. In *Proceedings of the Eighth Pacific Rim International Conference on Artificial Intelligence*, pages 742–750. Springer-Verlag, Berlin, Germany, 2004.
- [108] T. Li, S. Zhu, Q. Li, and M. Ogihara. Gene functional classification by semi-supervised learning from heterogeneous data. In *Proceedings of the Symposium on Applied Computing*, pages 78–82. ACM Press, New York, NY, 2003.

- [109] B. Liu, S. Mardikian, R. Jackson, and D. Westhead. Computational molecular modelling of protein structure and function. Technical report, University of Sheffield, Sheffield, UK, 2004.
- [110] J. Liu and H. Iba. Selecting informative genes using a multiobjective evolutionary algorithm. In *Proceedings of the World Congress on Computational Intelligence*, pages 297–302. IEEE Press, Anaheim, CA, 2002.
- [111] J. Liu and H. Iba. Prediction of tumor outcome based on gene expression data. *Wuhan University Journal of Natural Sciences*, 9(2):177–182, 2004.
- [112] Y. Liu, T. Oezyer, R. Alhajj, and K. Barker. Integrating multi-objective genetic algorithm and validity analysis for locating and ranking alternative clustering. *Informatica*, 29:33–40, 2005.
- [113] X. Llorca and D. E. Goldberg. Bounding the effect of noise in multiobjective learning classifier systems. *Evolutionary Computation*, 11(3):278–297, 2003.
- [114] S. Madeira and A. Oliveira. Biclustering algorithms for biological data analysis: a survey. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 1:24–45, 2004.
- [115] J. M. Malard, A. Heredia-Langner, D. J. Baxter, K. H. Jarman, and W. R. Cannon. Constrained de novo peptide identification via multi-objective optimization. In *Proceedings of the Third IEEE International Workshop on High Performance Computational Biology*, 2004. <http://www.hicomb.org/HiCOMB2004/>.
- [116] Ch. Mandal, R. D. Gudi, and G. K. Suraishkumar. Multi-objective optimization in asperigillus niger fermentation for selective product enhancement. *Bio-process. Biosyst. Eng.*, 28:149–164, 2005.
- [117] R.T. Marler and J.S. Arora. Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization*, 26(6):369–395, 2004.
- [118] C. A. Mattson, A. A. Mullur, and A. Messac. Smart pareto filter: obtaining a minimal representation of multiobjective design space. *Engineering Optimization*, 36:721–740, 2004.
- [119] C. E. Metz. Basic principles of roc analysis. *Semin Nucl Med*, 8(4):283–298, 1978.
- [120] T. Mitchell. *Machine Learning*. McGraw Hill, Hightstown, NJ, 1997.
- [121] K. Mitra, S. Majumdar, and S. Raha. Multiobjective dynamic optimization of epoxy polymerization process. *Computer and Chemical Engineering*, 28(12):2583–2594, 2004.
- [122] S. Mitra. Computational intelligence in bioinformatics. *T. Rough Sets*, pages 134–152, 2005.
- [123] M. Morita, R. Sabourin, F. Bortolozzi, and C. Y. Suen. Unsupervised feature selection using multi-objective genetic algorithms for handwritten word recognition. In *Proceedings of the Seventh International Conference on Document Analysis and Recognition*, pages 666–671. IEEE Press, New York, NY, 2003.
- [124] R. Mott. Local sequence alignments with monotonic gap penalties. *Bioinformatics*, 15(6):455–462, 1999.
- [125] D. W. Mount. *Bioinformatics: Sequence and Genome Analysis*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2004.
- [126] L. Muniglia, L. N. Kiss, C. Fonteix, and I. Marc. Multicriteria optimization of a single-cell oil production. *European Journal of Operational Research*, 153(2):360–369, 2003.
- [127] P. K. S. Nain and K. Deb. A computationally effective multi-objective search and optimization technique using coarse-to-fine grain modeling. Technical Report Kangal Report No. 2002005, IITK, Kanpur, India, 2005.
- [128] C. G. Nevill-Manning, T. D. Wu, and D. L. Brutlag. Highly specific protein sequence motifs for genome analysis. *Proc. Natl. Acad. Sci. USA*, 95:5865–5871, 1998.
- [129] O. Nicolotti, V. J. Gillet, P. J. Fleming, and D. V. S. Green. Multiobjective optimization in quantitative structure-activity relationships: deriving accurate and interpretable QSARs. *Journal of Medical Chemistry*, 45:5069–5080, 2002.
- [130] S. O’Hagan, W. B. Dunn, M. Brown, J. D. Knowles, and D. B. Kell. Closed-loop, multiobjective optimization of analytical instrumentation: gas chromatography/time-of-flight mass spectrometry of the metabolomes of human serum and of yeast fermentations. *Analytical Chemistry*, 77(1):290–303, 2005.
- [131] L. S. Oliveira and R. Sabourin. A methodology for feature selection using multiobjective genetic algorithms for handwritten digit string recognition. *International Journal of Patterns Recognition and Artificial Intelligence*, 17(6):903–929, 2003.
- [132] R. D. M. Page and E. C. Holmes. *Molecular evolution: a phylogenetic approach*. Blackwell Science, Oxford, UK, 1998.
- [133] L. Poladian and L. S. Jermin. Multi-objective evolutionary algorithms and phylogenetic inference with multiple data sets. *Soft Computing*, 10:359–368, 2006.
- [134] H. Putz, J. C. Schoen, and M. Jansen. Combined method for ab initio structure solution from powder diffraction data. *J. Appl. Cryst.*, 32:864–870, 1999.

- [135] J. Rachlin, Ch. Ding, Ch. Cantor, and S. Kasif. Multiplex: multi-objective multiplex PCR assay design. *Nucleic Acids Res.*, 33, 2005.
- [136] M. Rajapakse, B. Schmidt, and V. Brusic. Multi-objective evolutionary algorithm for discovering peptide binding motifs. In *Proceedings of the 4th European Workshop on Evolutionary Computation and Machine Learning in Bioinformatics*. Springer-Verlag, Berlin, Germany, 2006. To appear.
- [137] M. A. Roytberg, M. N. Semionenkov, and O. Yu. Tabolina. Sequence alignment without gap penalties. In *Proceedings of the International Conference on Bioinformatics of Genome Regulation and Structure*, 1998. <http://www.bionet.nsc.ru/bgrs/thesis/85/>.
- [138] M. A. Roytberg, M. N. Semionenkov, and O. Yu. Tabolina. Pareto-optimal alignment of biological sequences. *Biofizika*, 44(4):581–94, 1999.
- [139] S. Ruzika and M. M. Wiecek. Approximation methods in multiobjective programming. *Journal of Optimization Theory and Applications*, 126(3):473–501, 2005.
- [140] M. Salemi and A.-M. VanDamme. *Handbook of Phylogenetic Methods*. Cambridge University Press, Cambridge, UK, 2003.
- [141] H. El. Samad, M. Khamash, C. Homescu, and L. Petzold. Optimal performance of the heat-shock gene regulatory network. In *Proceedings 16th IFAC World Congress*. Elsevier Press, Oxford, UK, 2005.
- [142] W. S. Sarle. Stopped training and other remedies for overfitting. In *Proceedings of the 27th Symposium on the Interface of Computing Science and Statistics*, pages 352–360, 1995. <ftp://ftp.sas.com/pub/neural/inter95.ps.Z>.
- [143] R. E. Schapire. The strength of weak learnability. *Machine Learning*, 5(2):197–227, 1990.
- [144] C. Schmidt-Dannert, D. Umeno, and F. H. Arnold. Molecular breeding of carotenoid biosynthetic pathways. *Nature Biotechnology*, 18:750–753, 2000.
- [145] G. Schneider and U. Fechner. Computer-based de novo design of drug-like molecules. *Nature Reviews Drug Discovery*, 4(8):649–63, 2005.
- [146] E. Schreibmann, M. Lahanas, L. Xing, and D. Baltas. Multiobjective evolutionary optimization of the number of beams, their orientations and weights for intensity-modulated radiation therapy. *Physics in Medicine and Biology*, 49(5):747–770, 2004.
- [147] C. Schretter and M. C. Milinkovitch. Oligonucleotide design by multilevel optimization. Technical report, Unit of Evolutionary Genetics, Institute of Molecular Biology and Medicine, Free University of Brussels Gosselies, Belgium, 2005.
- [148] C. Schretter and M. C. Milinkovitch. Oligofactory: a visual tool for interactive oligonucleotide design. *Bioinformatics*, 22(1):115–116, 2006.
- [149] S. Schulze-Kremer. Application of evolutionary computation to protein folding with specialized operators. In G. B. Fogel and D. W. Corne, editors, *Evolutionary Computation in Bioinformatics*, chapter 8, pages 163–191. Morgan Kaufmann Publishers, Amsterdam, The Netherlands, 2003.
- [150] R. Schwarz, P. Musch, A. von Kamp, B. Engels, H. Schirmer, S. Schuster, and T. Dandekar. Yana — a software tool for analyzing flux modes, gene-expression and enzyme activities. *BMC Bioinformatics*, 6(135), 2005.
- [151] M. B. Seasholtz and B. R. Kowalski. The parsimony principle applied to multivariate calibration. *Analytica Chimica Acta*, 277:165–177, 1993.
- [152] N. S. Sharma, M. G. Ierapetritou, and M. L. Yarmush. Novel quantitative tools for engineering analysis of hepatocyte cultures in bioartificial liver systems. *Biotechnology and Bioengineering*, 92(3), 2005.
- [153] K. J. Shaw, A. L. Nortcliffe, M. Thompson, J. Love, P. J. Fleming, and C. M. Fonseca. Assessing the performance of multiobjective genetic algorithms for optimization of a batch process scheduling problem. In *Proceedings of the IEEE Congress on Evolutionary Computation*, pages 37–45, 1999.
- [154] S. Y. M. Shi, P. N. Suganthan, and K. Deb. Multi-class protein fold recognition using multi-objective evolutionary algorithms. In *Proceedings of the IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, pages 61–66, 2004.
- [155] S.-Y. Shin, I.-H. Lee, D. Kim, and B.-T. Zhang. Multiobjective evolutionary optimization of DNA sequences for reliable DNA computing. *IEEE Transactions on Evolutionary Computation*, 9(2):143–158, 2005.
- [156] L. A. Soinov, M. A. Krestyaninova, and A. Brazma. Towards reconstruction of gene networks from expression data by supervised learning. *Genome Biol.*, 4(1), 2003.
- [157] E. P. Someren, L. F. A. Wessels, E. Backer, and M. J. T. Reinders. Multi-criterion optimization for genetic network modeling. *Signal processing*, 83:763–775, 2003.
- [158] N. Speer, C. Spieth, and A. Zell. A memetic clustering algorithm for gene expression profiles and biological annotation. In *Proceedings of the IEEE Congress on Evolutionary Computation*, pages 1631–1638. IEEE Press, Anaheim, CA, 2004.

- [159] C. Spieth, F. Streichert, N. Speer, and A. Zell. Multi-objective model optimization for inferring gene regulatory networks. In *Proceedings of the Conference on Evolutionary Multi-Criterion Optimization*, pages 607–620. Springer-Verlag, Berlin, Germany, 2005.
- [160] G. Stehr, H. Graeb, and K. Antreich. Performance trade-off analysis of analog circuits by normal-boundary intersection. In *40th Design Automation Conference*, pages 958–963, Anaheim, CA, 2003. IEEE Press.
- [161] D. Stekel. *Microarray Bioinformatics*. Cambridge University Press, Cambridge, UK, 2003.
- [162] W. P. C. Stemmer. Rapid evolution of a protein in vitro by DNA shuffling. *Nature*, 370(6488):389–391, 1994.
- [163] R. Steuer. *Multiple criteria optimization, theory, computation and applications*. Wiley, New York, NY, 1986.
- [164] A. Strehl and J. Ghosh. Cluster ensembles — a knowledge reuse framework for combining multiple partitions. *Journal on Machine Learning Research*, 3:583–617, 2002.
- [165] R. Tibshirani, G. Walther, and T. Hastie. Estimating the number of clusters in a dataset via the Gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):411–423, 2001.
- [166] A. Topchy, A. K. Jain, and W. Punch. Clustering ensembles: Models of consensus and weak partitions. *Under submission to the IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(12):1866–1881, 2005.
- [167] J. Tsai, R. Bonneau, A. Morozov, B. Kuhlman, C. A. Rohl, and D. Baker. An improved protein decoy set for testing energy functions for protein structure prediction. *Proteins*, 52:76–87, 2003.
- [168] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, NY, 1995.
- [169] J. Vera, P. d. Atauri, M. Cascante, and N. V. Torres. Multicriteria optimization of biochemical systems by linear programming: application to production of ethanol by *saccharomyces cerevisiae*. *Biotechnology and Bioengineering*, 83(3):335–343, 2003.
- [170] J. Vera and N. V. Torres. Metmap: an integrated matlab package for analysis and optimization of metabolic systems. In *Silico Biology*, 4(2):97–108, 2004.
- [171] J. Weston, C. Leslie, D. Zhou, A. Elisseeff, and W. Noble. Semi-supervised protein classification using cluster kernels. *Bioinformatics*, 21(15):3241–3247, 2005.
- [172] S. Wiegand, Ch. Igel, and U. Handmann. Evolutionary multi-objective optimization of neural networks for face detection. *International Journal of Computational Intelligence and Applications*, 4(3):237–253, 2004.
- [173] H. Wright, K. Brodlie, and T. David. Navigating high-dimensional spaces to support design steering. In *Proceedings of the 11th IEEE Visualization Conference*, pages 291–296. IEEE Press, Anaheim, CA, 2000.
- [174] G. G. Yen and H. Lu H. Hierarchical rank-density genetic algorithm for radial basis function neural network design. *International Journal of Computational Intelligence and Applications*, 3(3):213–232, 2003.
- [175] I. Zwir, R. R. Zaliz, and E. H. Ruspini. Automated biological sequence description by genetic multiobjective generalized clustering. *Ann. N. Y. Acad. Sci.*, 980:65–82, 2002.