SFFS-MR: A Floating Search Strategy for GRNs Inference

Fabrício M. Lopes^{1,2}, David C. Martins-Jr³,
Junior Barrera⁴, and Roberto M. Cesar-Jr^{2,5}

¹ Federal University of Technology - Paraná, Brazil
`fabricio@utfpr.edu.br`

² Institute of Mathematics and Statistics, University of São Paulo, Brazil
`fabriciolopes,cesar@vision.ime.usp.br`

³ Federal University of ABC, Brazil
`david.martins@ufabc.edu.br`

⁴ Faculty of Philosophy, Sciences and Letters of Ribeirão Preto,
University of São Paulo, Brazil

`jb@ime.usp.br`

⁵ CTBE, Brazil

Abstract. An important problem in the bioinformatics field is the inference of gene regulatory networks (GRN) from temporal expression profiles. In general, the main limitations faced by GRN inference methods is the small number of samples with huge dimensionalities and the noisy nature of the expression measurements. In face of these limitations, alternatives are needed to get better accuracy on the GRNs inference problem. In this context, this work addresses this problem by presenting an alternative feature selection method that applies prior knowledge on its search strategy, called SFFS-MR. The proposed search strategy is based on SFFS algorithm, with the inclusion of multiple roots at the beginning of the search, which are defined by the best and worst single results of the SFS algorithm. In this way, the search space traversed is guided by these roots in order to find the predictor genes for a given target gene, specially to better identify genes presenting intrinsically multivariate prediction, without worsening the asymptotical computational cost of the SFFS. Experimental results show that the SFFS-MR provides a better inference accuracy than SFS and SFFS, maintaining a similar robustness of the SFS and SFFS methods. In addition, the SFFS-MR was able to achieve 60% of accuracy on network recovery after only 20 observations from a state space of size 2^{20} , thus presenting very good results.

Keywords: SFS, SFFS, feature selection, inference, gene networks, pattern recognition, systems biology, bioinformatics.

1 Introduction

Systems biology is the study of live organisms viewed as integrated and interacting networks of genes, proteins and biochemical reactions. It has been an emergent

field of study in bioinformatics since the advent of high-throughput technologies for extraction of gene expressions (mRNA abundances or transcripts), such as DNA microarrays [29] or SAGE [35], and more recently RNA-Seq [36]. These high-throughput techniques together with computational methods, make possible to analyze thousands of transcripts simultaneously. Consequently, the volume of transcriptome data available from a multitude of species during the last ten years increased dramatically. In this context, a big challenge that researchers need to face is the large number of variables (thousands) for just a few experiments available (dozens). In order to infer relationships among those variables, it is needed a great effort in developing novel computational and statistical techniques that are able to alleviate the intrinsic error estimation committed in the presence of small number of samples with huge dimensionalities.

A commonly used approach to infer gene regulatory networks (GRN) is the use of feature selection techniques [18, 4, 13, 23, 21, 26, 11, 3, 38, 9]. A feature selection method is composed by two main parts: a search algorithm and a criterion function. As far as the search algorithms are concerned, there are two main categories: the optimal and sub-optimal algorithms. The optimal algorithms (including exhaustive and branch-and-bound searches) return the best feature subspace, but their computational costs are very high to be applied in general, especially for high dimensionality problems such as GRN inference focused in this paper. The sub-optimal algorithms do not guarantee that the solution is optimal, but some of them present a reasonable cost-benefit between computational cost and quality of the solution. In this work, we explore the exhaustive search (optimal), the Sequential Forward Selection (SFS - sub-optimal) and the Sequential Forward Floating Selection (SFFS - sub-optimal with excellent cost-benefit) [24].

The reason why efficient search algorithms such as SFS and SFFS do not always reach the optimal solution is due to the *nesting effect* in which a feature not present in the optimal solution may be included in the partial solution of the algorithm and never be discarded, which leads to a sub-optimal solution [32]. Such effect can be explained by the fact that two features grouped in a pair may perform a very nice prediction of the class (or value) of the target object, although their individual predictions about the target are bad. Such pair of features can be even better than two other features grouped that perform well individually. This phenomenon is called synergy or intrinsically multivariate prediction [1, 22].

The main contribution of this work is the proposal of a new search strategy algorithm that uses the very efficient SFFS starting from some initial features (denominated here as “roots”). Such roots contain not only the best individual features, but also the worst ones. This strategy tries to identify those features that are synergetic (or intrinsically multivariate predictive) in predicting the targets without worsening the asymptotical computational cost of the SFFS. The focus of the experiments presented here is given on the GRN inference application.

Next section (Section 2) will introduce a brief background on the network inference problem. In Section 3, the feature selection problem is discussed in

more detail, including a short description of the SFS and SFFS techniques. Section 4 discusses the intrinsically multivariate prediction issue and how it can affect the greedy feature selection algorithms in such way that the achieved solution be relatively far away from the optimal. Section 5 describes our proposed feature selection method (SFFS-MR). Section 6 shows some experimental results. Finally, Section 7 concludes the work, discussing future perspectives.

2 GRN Inference

The inference of GRNs from temporal expression data is a great challenge. One of the main reasons for that is the usual limitation of the data itself in which the number of samples is not enough to infer relationships among elements present in the biological system with a reasonable confidence [15]. Due to this fact, there are several approaches proposed for modeling and identification of GRNs. The main approaches for modeling of GRNs are Boolean Networks, Differential Equations and Bayesian Networks. As Boolean Networks (and its stochastic version: Probabilistic Boolean Networks (PBN) [30]) are more suitable in situations with limited data samples, here we concentrate the attention in the PBN model.

Considering feature selection approaches for identification of GRNs, there are three types of criterion functions frequently used. The first one is the correlation between two features, in which there is an edge between two genes if the correlation between their expression profiles are larger than some predefined threshold [33]. Such method considers only 1-to-1 relationships, being suitable to identify co-regulation between genes, functional modules and clusters. Nevertheless, it ignores the fact that the expression of a given target gene may be regulated by a group of genes with multivariate interaction.

Another class of criterion functions refers to those based on the Bayesian error estimation of the predictors in classifying the target expression. A broadly used criterion to infer GRNs is the coefficient of non-linear determination (CoD) [13, 7, 12]. With this measurement, it is possible to capture N-to-1 (multivariate) relationships.

The information theory based (entropy, mutual information) criterion functions are also commonly employed for inference of GRNs. There are several works that use such metrics in substitution to correlation to infer 1-to-1 relationships in GRNs [4, 25, 21, 11, 26]. However, it is possible to employ these metrics for inference of multivariate relationships [18, 3, 38]. Basically, the difference between the entropy based and the Bayesian error based criteria is that the second relies on the minimum conditional probability distributions of the target given the subset of predictors, while the former relies on the uniformity of these conditional probability distributions as a whole (larger uniformity leads to higher entropy, which in turn leads to smaller mutual information).

The literature related to modeling and inference of GRNs is vast and continues to grow quickly, which reflects the importance of this research field. Some reviews on this topic can be found in [5, 16, 34, 28, 17, 14].

3 Feature Selection

A feature selection method consists in selecting a subset of features that makes a good representation, classification or prediction of states (or values) of the objects in study. Generally, it is composed by two main parts: a search strategy and a criterion function that attributes a quality value to the feature subsets. Due to the nesting effect (see Section 4), the search for the best subsets generally requires the investigation of the entire space of possible subsets (exhaustive search), although depending on some constraints of the criterion function adopted (e.g. monotonical or U-shaped), it is possible to obtain the best subset by looking for a restricted space of subsets employing “branch-and-bound” techniques [31,27].

As exhaustive search is computationally impractical for most “real world” tasks, and especially for inference of GRNs which involves data with thousands of features (genes), it is clear the existence of a trade-off between optimality and computational cost. Next we introduce two classical heuristics for feature selection.

3.1 Sequential Forward Selection (SFS)

The SFS algorithm starts with the empty set and adds the best feature found to this set according to the criterion function adopted. In the next step, it adds a second feature that, jointly with the feature already included, composes the best feature pair. This process continues until it reaches a stop condition, commonly based on a fixed dimension (number of features of the subset to be returned), or based on the criterion function value variation (it stops if the criterion value does not improve significantly from the previous to the next step). There is a variant of this algorithm, the Sequential Backward Selection (SBS), which starts with the complete set and successively removes the less relevant features according to the criterion function until the stop condition be satisfied [24].

3.2 Sequential Forward Floating Selection (SFFS)

The SFS and SBS search methods present an undesirable drawback known as nesting effect. This effect happens because the discarded features in the top-down approach are not inserted anymore, or the inserted features in the bottom-up approach are never discarded. Section 4 presents the reason why this phenomenon occurs, and the potential problem leaded by it in some important gene inference situations.

In order to amenize this problem, the SFFS [24] was adopted. The SFFS algorithm tries to avoid the nesting effect by allowing to insert and to exclude features on a subset in a floating way, i.e., without defining the number of insertions or exclusions. In this algorithm, SFS and SBS are successively applied. A schematic flowchart of the SFFS algorithm is presented in Figure 1.

The aforementioned algorithm is computationally efficient and usually returns a solution very close to the optimal, presenting excellent cost-benefit. There are also adaptive and generalized floating methods that try to improve the SFFS

Fig. 1. Simplified flowchart of the SFFS algorithm [19]. K refers to the size of the current solution subset while d refers to the size of the final solution subset.

results at the expense of a significant increase on the computational cost. Nevertheless, they still can not avoid the nesting effect completely [32].

4 Intrinsically Multivariate Prediction

A set of predictor features is considered intrinsically multivariate predictive with regard to a target feature if the target behavior is strongly predicted by the whole set of predictors, but poorly predicted by any of its proper subsets. Formally, a set of features $\mathbf{X}$ is intrinsically multivariate predictive for the target feature Y with respect to λ and δ , for $0 \leq \lambda, \delta \leq 1$ and $\lambda < \delta$, if

$$\max_{\mathbf{Z} \subsetneq \mathbf{X}} \mathcal{F}_Y(\mathbf{Z}) \leq \lambda \wedge \mathcal{F}_Y(\mathbf{X}) \geq \delta \quad (1)$$

where $\mathcal{F}$ is a criterion function that varies from 0 to 1 (0 meaning absence of prediction and 1 meaning full prediction) [22]. Generally, λ has small value (usually smaller than 0.2) and δ has a high value (usually larger than 0.6). For a pair predictors-target $(\mathbf{X}, Y)$, the largest δ for which the prediction is intrinsically multivariate is $\delta = \mathcal{F}_Y(\mathbf{X})$. In this way, it is possible to define a score of intrinsically multivariate prediction (IMP score) through the maximum value of $\delta - \lambda$. Thus, the IMP score is given by

$$I_Y(\mathbf{X}) = \mathcal{F}_Y(\mathbf{X}) - \max_{\mathbf{Z} \subsetneq \mathbf{X}} \mathcal{F}_Y(\mathbf{Z}) \quad (2)$$

The concept of intrinsically multivariate prediction is related to the nesting effect that occurs when a greedy feature selection algorithm like SFS or other sub-optimal heuristics are applied. Next we present an example that clarifies this concept. Suppose two Boolean features $X_1 = x_1 \in \{0, 1\}$, $X_2 = x_2 \in \{0, 1\}$ and another Boolean feature $Y = y \in \{0, 1\}$ considered as target. Also suppose the

joint probability distributions (JPD) $P(x_1, x_2, y) \forall \{x_1, x_2, y\} \in \{0, 1\}^3$ given in Table 1. Considering the nonlinear Coefficient of Determination (CoD) as criterion function defined as $CoD_Y(\mathbf{X}) = \frac{\varepsilon_Y - \varepsilon_Y(\mathbf{X})}{\varepsilon_Y}$, where ε_Y is the error obtained by classifying Y in the absence of other observations (prior error) and $\varepsilon_Y(\mathbf{X})$ is the error obtained by classifying Y based on the observation of the feature set $\mathbf{X}$ [8]. Such pair predictors-target has $CoD_Y(X_1, X_2) = \frac{0.5 - 0.2}{0.5} = 0.6$. On the other hand, if we consider X_1 and X_2 individually, both $CoD_Y(X_1)$ and $CoD_Y(X_2)$ are zero, since $P(X_1 = x_1, Y = y) = P(X_2 = x_2, Y = y) = 0.25 \forall x_1 \in \{0, 1\}, x_2 \in \{0, 1\}, y \in \{0, 1\}$, which implies $CoD_Y(X_1) = CoD_Y(X_2) = \frac{0.5 - 0.5}{0.5} = 0$. The IMP score in this case is $I_Y(X_1, X_2) = 0.6 - 0 = 0.6$, which is considered high (X_1, X_2 and Y form an IMP set).

Table 1. Example of a joint probability distribution (JPD) between the target Y and two predictors X_1 and X_2 in which such features form an intrinsically multivariate predictive set ($I_Y(\mathbf{X}) = 0.6$ for CoD as criterion function).

$X_1 = x_1$	0	0	1	1
$X_2 = x_2$	0	1	0	1
$P(X_1 = x_1, X_2 = x_2, Y = 0)$	0.2	0.05	0.05	0.2
$P(X_1 = x_1, X_2 = x_2, Y = 1)$	0.05	0.2	0.2	0.05

It is important to notice that, in the example given above, Y is given by a stochastic exclusive-or (XOR), i.e., $\argmax_{y \in Y} P(y|x_1, x_2) = 0$ if $x_1 = x_2$ or $\argmax_{y \in Y} P(y|x_1, x_2) = 1$ if $x_1 \neq x_2$. According to [22], in the case of 2 binary predictors, there are 8 logics that can produce high IMP score: XOR, NXOR (negated XOR), AND, OR, NOR, NAND, $x_1 \wedge x_2$ and $x_1 \vee x_2$. However, there are other properties besides prediction logic that can jointly originate IMP sets: predictive power (defined as $1 - \varepsilon_Y(\mathbf{X})$), covariance between predictors and probability distribution of each isolated predictor (marginal probabilities).

Most feature selection heuristics discard features that perform bad individual prediction about the target to compose the initial solutions. Due to the intrinsically multivariate prediction phenomenon, such heuristics tend to reach local minima that sometimes are far from the optimal solution. Next section describes the main contribution of the paper: a feature selection strategy that applies SFFS starting not only looking for good features, but also searching for bad individual features which can form intrinsically multivariate predictive sets.

5 SFFS with Multiple Roots (SFFS-MR)

The inference of GRNs is one of the most challenging problems of Systems Biology in these days, mainly because of the intrinsic error estimation due to small number of samples with huge dimensionalities and the presence of genes that have an intrinsically multivariate prediction. In this context, we propose an alternative search strategy for GRNs inference problem, called SFFS-MR, which

extends the SFFS method by including not only the best individual features, but also the worst ones. Algorithm 1 presents the specification of the SFFS-MR algorithm.

Algorithm 1. SFFS-MR (Δ, d, irb, irw)

```

1: var exelist, bestset, newsubset
2: var vector ibroots[irb], iwroots[irw]
3: var float bestvalue, newvalue
4: var integer k  $\leftarrow$  1
5: bestvalue  $\leftarrow$  SFS(ibroots, iwroots, k)
6: bestset  $\leftarrow$  ibroots[1]
7: for i = 1 to irb do
8:   exelist.append(ibroots[i])
9: end for
10: for i = 1 to irw do
11:   exelist.append(iwroots[i])
12: end for
13: while exelist is not empty do
14:   newsubset  $\leftarrow$  exelist.removefirst
15:   k  $\leftarrow$  newsubset.cardinality
16:   if k < d then
17:     newvalue  $\leftarrow$  SFS(newsubset,  $\emptyset$ , k + 1)
18:     if newvalue < bestvalue and (bestvalue - newvalue) >  $\Delta$  then
19:       newvalue  $\leftarrow$  SBS(newsubset)
20:       bestvalue  $\leftarrow$  newvalue
21:       bestset  $\leftarrow$  newsubset
22:     end if
23:     exelist.append(newsubset)
24:   end if
25: end while
26: return bestset

```

The Algorithm 1 starts by applying the SFS (Section 3.1) in order to discover the *irb* best and *irw* worst individual features ($k = 1$), which are ranked according to the adopted criterion function. The variable *bestvalue* represents the criterion function value achieved by the best feature *bestset*[1]. These individual subsets are appended in an execution list (*exelist*). In the *while* loop, the first subset in the execution list will be removed and its cardinality will be tested. If its cardinality has not reached the limit, the SFS will increment its cardinality, i.e., to include a new feature that jointly with the features already present in *newsubset* composes the best feature subset with cardinality $k + 1$. The variable *newvalue* represents the criterion function value achieved by *newsubset*. If *newsubset* has a better criterion function value (lower or higher, depending on the adopted criterion function), and the gain is more than Δ , then a conditional exclusion is performed, which is represented by calling SBS function, and the *bestvalue* and *bestset* are updated by *newvalue* and *newsubset* respectively. At the end, *newsubset* will be stored in the execution list for a new attempt to

extend. In summary, the SFFS-MR differs from SFFS (Section 3.2) because of the exploration of multiple roots. If irb and irw (number of initial roots) are small constant values compared to the total number of variables (order of thousands for the application focused here), its asymptotical computational cost is not worse than SFFS.

The parameter d represents the maximum cardinality of the subset of predictors. A Δ of criterion function value variation is also included. Here the Δ value prevents that minor variations of the criterion function ($\leq \Delta$) causes the increase of the subset of predictors. The present paper adopted $d = 5$, $\Delta = 0.05$, $irb = 1$ and $irw = 5$.

6 Experimental Results

This section presents the experimental results obtained by considering a synthetic networks approach, which was adapted from [20]. The artificial gene networks (AGNs) were generated by considering the uniformly-random Erdős-Rényi (ER) [10] topology. The Probabilistic Boolean Networks (PBN) [30] approach was applied to generate the network dynamics, i.e., the temporal expression profiles.

For all experiments, the network model (ER) was applied with 20 vertices (genes). The average degree $\langle k \rangle$ per gene varied from 1 to 5, and the number of observed instants of time (signal size) varied from 5, 10, 15, 20 to 100 in steps of 20. For each gene g_i of the network, its value was given by a randomly selected function from 3 possible Boolean functions $\{f_1^{(i)}, f_2^{(i)}, f_3^{(i)}\}$, where the probabilities of each function be selected are given by $c_1^{(i)} = 0.95$, $c_2^{(i)} = 0.025$, $c_3^{(i)} = 0.025$, $i = 1, \dots, 20$.

In order to identify the networks, the simulated temporal expressions were submitted to the software described in [19] which implements feature selection methods for network inference, applying the SFS and SFFS as search strategies. The same method and parameters (default) were kept fixed during comparative analysis with SFFS-MR. We adopted as similarity measure between the AGN and the inferred network, the PPV (Positive Predictive Value, also known as accuracy or precision) and Sensitivity (or recall) measurements presented by [6], which are widely used to compare the results of the GRNs inference methods. The experimental results were obtained from 50 simulations of each signal size and $\langle k \rangle$ value.

The first experiment was performed in order to compare the three methods: SFS, SFFS and SFFS-MR with respect to the temporal expressions size. Fig. 2 presents these results, in which the PPV measure was calculated by taking into account the average results for all variations of average degree $\langle k \rangle$.

It is possible to notice that all methods have an increase on its performance by increasing the number of observations. However, the improvement of the SFFS-MR was more consistent, e.g., achieving 60% of similarity against 55%

Fig. 2. PPV measure obtained by SFS, SFFS and SFFS-MR applied to infer network edges from different sizes of temporal expression profiles (signal size). Similarity measure represents the mean over 50 executions.

Fig. 3. Similarity measure, $\text{SQRT}(\text{PPV} \cdot \text{Sensitivity})$, obtained by SFS, SFFS and SFFS-MR applied to infer network edges from different network complexities in terms of average degree $\langle k \rangle$. Similarity measure represents the mean over 50 executions.

(SFFS) and 51% (SFS) after only 20 observations and getting 75% against 57% (SFFS) and 54% (SFS) after 100 observations, even in the presence of some perturbations in the temporal signal, implied by the stochasticity in the application of transition functions.

The second experiment was performed in order to compare the robustness of the methods by increasing the complexity of the networks in terms of its average degree $\langle k \rangle$. The geometric mean between PPV and Sensitivity, presented in Fig. 3, was calculated by taking into account the average results for all variations of signal size. Fig. 3 shows that the three methods were very robust to increasing complexity of networks, presenting a soft decrease of similarity with the increase of average degree $\langle k \rangle$. In addition, the SFFS-MR showed slightly better results than the SFS and SFFS.

7 Conclusion

This work presents a floating search strategy for the inference of gene regulatory networks. Given the known limitations, our focus is to bring attention to the inclusion of prior knowledge on search methods, so that it occurs more efficiently. The proposed strategy is based on the assumption that some genes in biological organisms have an intrinsically multivariate prediction. The presented method exploits this property by the inclusion of multiple roots at the beginning of the search, which are defined by the best and worst single results of the SFS algorithm. In this context, the search space traversed by the SFFS-MR method is a little wider than SFS and SFFS, but does not worsen the asymptotical computational cost of the SFFS.

The experimental results show that the SFFS-MR provides a better inference accuracy (PPV) than SFS and SFFS, when considering small signal sizes with 15-20 time points and also with large ones, with 100 time points. In addition, the SFFS-MR was able to achieve 60% of accuracy on network recovery after only 20 observations from a state space of size 2^{20} , presenting very good results.

The SFFS-MR has also proved to be robust, as SFS and SFFS, when submitted to the increasing complexity of the networks in terms of its average degree $\langle k \rangle$. The robustness is an important property for the inference methods, even in the presence of some perturbations in the temporal signal, implied by the stochasticity in the application of transition functions. Besides, the SFFS-MR showed slightly better results than the SFS and SFFS.

A possible extension of the present work is to apply the SFFS-MR in order to evaluate large-scale networks, as well as to compare this method to other network inference methods based on feature selection. Also, we plan to apply this technique to infer GRNs from real data. Another very important improvement in search methods for GRNs inference would be the inclusion of topology information, such as the small-world (WS) [37] and scale-free (BA) [2] in order to guide the search process for the correct topology inference of these networks.

Acknowledgments

This work was supported by FAPESP, CNPq and CAPES.

References

1. Anastassiou, D.: Computational analysis of the synergy among multiple interacting genes. *Molecular Systems Biology* 3(83) (2007)
2. Barabási, A.L., Albert, R.: Emergence of scaling in random networks. *Science* 286(5439), 509–512 (1999)
3. Barrera, J., Cesar Jr., R.M., Martins Jr., D.C., Vencio, R.Z.N., Merino, E.F., Yamamoto, M.M., Leonardi, F.G., Pereira, C.A.B., Portillo, H.A.: Methods of Microarray Data Analysis V. In: Constructing Probabilistic Genetic Networks of Plasmodium Falciparum, from Dynamical Expression Signals of the Intraerythrocytic Development Cycle, pp. 11–26. Springer, Heidelberg (2007)
4. Butte, A., Kohane, I.: Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. In: Proceedings of the Pacific Symposium on Biocomputing, pp. 418–429 (2000)
5. D’haeseleer, P., Liang, S., Somogyi, R.: Genetic network inference: from co-expression clustering to reverse engineering. *Bioinformatics* 16(8), 707–726 (2000)
6. Dougherty, E.R.: Validation of inference procedures for gene regulatory networks. *Current Genomics* 8(6), 351–359 (2007)
7. Dougherty, E.R., Brun, M., Trent, J.M., Bittner, M.L.: Conditioning-Based Modeling of Contextual Genomic Regulation. *IEEE/ACM TCBB* 6(2), 310–320 (2009), <http://doi.ieeecomputersociety.org/10.1109/TCBB.2007.70247>
8. Dougherty, E.R., Kim, S., Chen, Y.: Coefficient of determination in nonlinear signal processing. *Signal Processing* 80, 2219–2235 (2000)
9. Dougherty, J., Tabus, I., Astola, J.: Inference of gene regulatory networks based on a universal minimum description length. *EURASIP Journal on Bioinformatics and Systems Biology*, 1–11 (2008)
10. Erdős, P., Rényi, A.: On random graphs. *Publ. Math. Debrecen* 6, 290–297 (1959)
11. Faith, J., Hayete, B., Thaden, J., Mogno, I., Wierzbowski, J., Cottarel, G., Kasif, S., Collins, J., Gardner, T.: Large-scale mapping and validation of escherichia coli transcriptional regulation from a compendium of expression profiles. *PLoS Biology* 5(1), 259–265 (2007)
12. Ghaffari, N., Ivanov, I., Qian, X., Dougherty, E.R.: A CoD-based reduction algorithm for designing stationary control policies on Boolean networks. *Bioinformatics* 26(12), 1556–1563 (2010) doi: 10.1093/bioinformatics/btq225, <http://bioinformatics.oxfordjournals.org/cgi/content/abstract/26/12/1556>
13. Hashimoto, R.F., Kim, S., Shmulevich, I., Zhang, W., Bittner, M.L., Dougherty, E.R.: Growing genetic regulatory networks from seed genes. *Bioinformatics* 20(8), 1241–1247 (2004)
14. Hecker, M., Lambeck, S., Toepfer, S., van Someren, E., Guthke, R.: Gene regulatory network inference: Data integration in dynamic models - A review. *Biosystems* 96(1), 86–103 (2009)
15. Hovatta, I., Kimppa, K., Lehmußola, A., Pasanen, T., Saarela, J., Saarikko, I., Saharinen, J., Tiikkainen, P., Toivanen, T., Tolvanen, M., et al.: DNA microarray data analysis. In: CSC, 2nd edn., Scientific Computing Ltd. (2005)
16. de Jong, H.: Modeling and simulation of genetic regulatory systems: A literature review. *Journal of Computational Biology* 9(1), 67–103 (2002)
17. Karlebach, G., Shamir, R.: Modelling and analysis of gene regulatory networks. *Nat. Rev. Mol. Cell Biol.* 9(10), 770–780 (2008)
18. Liang, S., Fuhrman, S., Somogyi, R.: Reveal: a general reverse engineering algorithm for inference of genetic network architectures. In: Proceedings of the Pacific Symposium on Biocomputing, pp. 18–29 (1998)

19. Lopes, F.M., Martins Jr., D.C., Cesar Jr., R.M.: Feature selection environment for genomic applications. *BMC Bioinformatics* 9(1), 451 (2008)
20. Lopes, F.M., Cesar Jr., R.M., Costa, L.d.F.: AGN simulation and validation model. In: Bazzan, A.L.C., Craven, M., Martins, N.F. (eds.) *BSB 2008. LNCS (LNBI)*, vol. 5167, pp. 169–173. Springer, Heidelberg (2008)
21. Margolin, A., Basso, K.N., Wiggins, C., Stolovitzky, G., Favera, R., Califano, A.: ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 7(suppl. 1), S7 (2006)
22. Martins Jr., D.C., Braga-Neto, U., Hashimoto, R.F., Dougherty, E.R., Bittner, M.L.: Intrinsically multivariate predictive genes. *IEEE Journal of Selected Topics in Signal Processing* 2(3), 424–439 (2008)
23. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE TPAMI* 27(8), 1226–1238 (2005)
24. Pudil, P., Novovičová, J., Kittler, J.: Floating search methods in feature-selection. *Pattern Recognition Letters* 15(11), 1119–1125 (1994)
25. Steuer, R., Kurths, J., Daub, C., Weise, J., Selbig, J.: The mutual information: detecting and evaluating dependencies between variables. *Bioinformatics* 18(Suppl. 2), 231–240 (2002)
26. Rao, A., Hero III, A., States, D., Engel, J.: Using directed information to build biologically relevant influence networks. In: *Proc. LSS Comput. Syst. Bioinform.*, pp. 145–156 (August 2007)
27. Ris, M., Martins Jr., D.C., Barrera, J.: U-curve: A branch-and-bound optimization algorithm for u-shaped cost functions on boolean lattices applied to the feature selection problem. *Pattern Recognition* 43(3), 557–568 (2010)
28. Schllit, T., Brazma, A.: Current approaches to gene regulatory network modelling. *BMC Bioinformatics* 8(suppl. 6), S9 (2007)
29. Shalon, D., Smith, S.J., Brown, P.O.: A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Research* 6(7), 639–645 (1996)
30. Shmulevich, I., Dougherty, E.R., Kim, S., Zhang, W.: Probabilistic boolean networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics* 18(2), 261–274 (2002)
31. Somol, P., Pudil, P., Kittler, J.: Fast branch & bound algorithms for optimal feature selection. *IEEE TPAMI* 26(7), 900–912 (2004)
32. Somol, P., Pudil, P., Novovičová, J., Paclík, P.: Adaptive floating search methods in feature selection. *Pattern Recognition Letters* 20, 1157–1163 (1999)
33. Stuart, J.M., Segal, E., Koller, D., Kim, S.K.: A gene-coexpression network for global discovery of conserved genetic modules. *Science* 302(5643), 249–255 (2003)
34. Styczynski, M.P., Stephanopoulos, G.: Overview of computational methods for the inference of gene regulatory networks. *Computers & Chemical Engineering* 29(3), 519–534 (2005)
35. Velculescu, V.E., Zhang, L., Vogelstein, B., Kinzler, K.W.: Serial Analysis of Gene Expression. *Science* 270(5235), 484–487 (1995)
36. Wang, Z., Gerstein, M., Snyder, M.: RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10(1), 57–63 (2009)
37. Watts, D.J., Strogatz, S.H.: Collective dynamics of small-world networks. *Nature* 393, 440–442 (1998)
38. Zhao, W., Serpedin, E., Dougherty, E.R.: Inferring connectivity of genetic regulatory networks using information-theoretic criteria. *IEEE/ACM TCBB* 5(2), 262–274 (2008)