An Experimental Comparison of Hierarchical Bayes and True Path Rule Ensembles for Protein Function Prediction

Matteo Re and Giorgio Valentini

DSI, Dipartimento di Scienze dell' Informazione,
Università degli Studi di Milano,
Via Comelico 39, 20135 Milano, Italia
{re,valentini}@dsi.unimi.it

Abstract. The computational genome-wide annotation of gene functions requires the prediction of hierarchically structured functional classes and can be formalized as a multiclass, multilabel, multipath hierarchical classification problem, characterized by very unbalanced classes. We recently proposed two hierarchical protein function prediction methods: the Hierarchical Bayes (HBAYES) and True Path Rule (TPR) ensemble methods, both able to reconcile the prediction of component classifiers trained locally at each term of the ontology and to control the overall precision-recall trade-off. In this contribution, we focus on the experimental comparison of the HBAYES and TPR hierarchical gene function prediction methods and their cost-sensitive variants, using the model organism *S. cerevisiae* and the FunCat taxonomy. The results show that cost-sensitive variants of these methods achieve comparable results, and significantly outperform both FLAT and their non cost-sensitive hierarchical counterparts.

1 Introduction

The hierarchical prediction of protein function annotations, such as terms in the Gene Ontology (GO), is a complex computational problem, characterized by several items: the number of functional classes is large, and a gene may belong to multiple classes; functional classes are structured according to a hierarchy; classes are usually unbalanced, with more negative than positive examples [1]. The simplest approach makes predictions for each term independently and, consequently, the predictor may assign to a single protein a set of terms that are inconsistent with one another. A possible solution for this problem is to train a classifier for each term of the reference ontology, to produce a set of prediction at each term and, finally, to reconcile the predictions by tacking into account the structure of the ontology. Many recent published works clearly demonstrated that this approach ensures an increment in precision, but this comes at expenses of the overall recall [2, 3].

Different research lines have been proposed for the hierarchical prediction of gene functions, ranging from structured-output methods, based on the joint

N. El Gayar, J. Kittler, and F. Roli (Eds.): MCS 2010, LNCS 5997, pp. 294–303, 2010.

kernelization of both input variables and output labels [4, 5], to ensemble methods, where different classifiers are trained to learn each class, and then combined to take into account the hierarchical relationships between functional classes [6, 3, 7].

Our work goes along this latter line of research. Our main contribution to this research area is represented by two methods, the HBAYES and TPR hierarchical ensemble-based gene function predictors [8, 16]. Both the methods are based on the concept of per-term predictions "reconciliation" which exploits information derived from the hierarchical relationships between the terms composing the considered functional ontology [6]. In this approach the first step is constituted by the prediction of protein functions (that is, the Functional Catalogue (FunCat) [10] or the Gene Ontology (GO) [11] terms) on a per-term basis. The obtained predictions are then combined in a post processing stage. The combination step can be realized using many methods and depends on the individual predictions performed by the component classifiers at each term of the ontology.

As observed in [6], many reconciliation methods yield reconciled probabilities with significantly lower precision than the original, unreconciled estimates. This problem can be solved by introducing one or more parameters able to modulate the overall precision but this approach is often associated with a corresponding loss in sensitivity that decrease the practical relevance of the prediction method. In order to ensure the applicability of the HBAYES and TPR ensemble methods in real world problems we recently proposed variants able to control the overall precision-recall trade-off. In this contribution we compare the hierarchical gene function prediction performances of the HBAYES and TPR ensemble methods, both in their respective "vanilla" and cost-sensitive versions, in order to highlight differences in their ability to reconcile base learners predictions and to preserve the overall precision and recall.

2 Methods

2.1 Concepts and Notation

Genome-wide gene function prediction can be modeled as a hierarchical, multiclass and multilabel classification problem. Indeed a gene/gene product $\boldsymbol{x}$ can be assigned to one or more functional classes of the set $\Omega = \{\omega_1, \omega_2, \ldots, \omega_m\}$. The assignments can be coded through a vector of multilabels $\mathbf{y} =< y_1, y_2, \ldots, y_m > \in \{0,1\}^m$, by which if $\boldsymbol{x}$ belongs to class ω_i, then $y_i = 1$, otherwise $y_i = 0$, where the variable $i, 1 \leq i \leq m$, refers to the indices corresponding to the m classes belonging to the set Ω.

In both the *Gene Ontology (GO)* and *FunCat* taxonomies the functional classes are structured according to a hierarchy and can be represented by a directed graph, where nodes correspond to classes, and arcs to relationships between classes. Hence the node corresponding to the class ω_i can be simply denoted by i. We represent the set of children nodes of i by child(i), and the

set of its parents by $\mathrm{par}(i)$. Moreover $y_{child(i)}$ denotes the labels of the children classes of node i and analogously $y_{par(i)}$ denotes the labels of the parent classes of i. Note that in FunCat only one parent is permitted, since the overall hierarchy is a tree forest, while in the GO, more parents are allowed, because the relationships are structured according to a directed acyclic graph.

Hierarchical ensemble methods train a calibrated classifier at each node of the taxonomy T. This is used to derive estimates $\widehat{p}_i(\boldsymbol{x})$ of the probabilities $p_i(\boldsymbol{x}) = \Pr\big(V_i = 1 \mid V_{\mathrm{par}(i)} = 1, \boldsymbol{x}\big)$ for all $\boldsymbol{x}$ and i, where $(V_1, \ldots, V_N) \in \{0,1\}^N$ is the vector random variable modeling the multilabel of a gene $\boldsymbol{x}$ and $\mathrm{par}(i)$ is the unique parent of node i in T. In order to enforce that only multilabels $\boldsymbol{V}$ that respect T should have nonzero probability, the base learner at node i is only trained on the subset of the training set including all examples $(\boldsymbol{x}, \boldsymbol{y})$ such that $y_{\mathrm{par}(i)} = 1$.

All the experiments presented in this work are performed using FunCat as reference functional ontology.

2.2 The Hierarchical Bayes (HBAYES) Ensemble

The HBAYES ensemble method is a general technique for solving hierarchical classification problems on generic taxonomies [12, 13]. In the evaluation phase, HBAYES predicts the Bayes-optimal multilabel $\widehat{\boldsymbol{y}} \in \{0,1\}^N$ for a gene $\boldsymbol{x}$ based on the estimates $\widehat{p}_i(\boldsymbol{x})$ for $i = 1, \ldots, N$. Namely, $\widehat{\boldsymbol{y}} = \mathrm{argmin}_{\boldsymbol{y}}\, \mathbb{E}\big[\ell_H(\boldsymbol{y}, \boldsymbol{V}) \mid \boldsymbol{x}\big]$, where the expectation is w.r.t. the distribution of $\boldsymbol{V}$. Here $\ell_H(\boldsymbol{y}, \boldsymbol{V})$ denotes the H-loss [12, 13], measuring a notion of discrepancy between the multilabels $\boldsymbol{y}$ and $\boldsymbol{V}$. Given fixed *cost coefficients* $c_1, \ldots, c_N > 0$, $\ell_H(\widehat{\boldsymbol{y}}, \boldsymbol{v})$ is computed as follows: all paths in the taxonomy T from the root 0 down to each leaf are examined and, whenever a node $i \in \{1, \ldots, N\}$ is encountered such that $\widehat{y}_i \neq y_i$, then c_i is added to the loss, while all the other loss contributions from the subtree rooted at i are discarded. As shown in [13], $\widehat{\boldsymbol{y}}$ can be computed via a simple bottom-up message-passing procedure whose only parameters are the probabilities $\widehat{p}_i(\boldsymbol{x})$. In the rest of the paper if there is no ambiguity we denote $\widehat{p}_i(\boldsymbol{x})$ simply by $\widehat{p}_i$.

2.3 Cost-Sensitive Variant of HBAYES

A simple cost-sensitive variant, HBAYES-CS, of HBAYES, described in [8] is suitable for learning datasets whose multilabels are sparse. This variant introduces a parameter α that is used to trade-off the cost of false positive (FP) and false negative (FN) mistakes. We start from an equivalent reformulation of the HBAYES prediction rule:

$$\widehat{y}_i = \operatorname*{argmin}_{y \in \{0,1\}} \left(c_i^- \widehat{p}_i (1 - y) + c_i^+ (1 - \widehat{p}_i) y + \widehat{p}_i \{y = 1\} \sum_{j \in \mathrm{child}(i)} H_j \right) \qquad (1)$$

where $H_j = c_j^- \widehat{p}_j (1 - \widehat{y}_j) + c_j^+ (1 - \widehat{p}_j) \widehat{y}_j + \sum_{k \in \mathrm{child}(j)} H_k$ is recursively defined over the nodes j in the subtree rooted at i with each $\widehat{y}_j$ set according to (1),

and $\{ A \}$ is the indicator function of event A. Furthermore, $c_i^- = c_i^+ = c_i/2$ are the costs associated to a FN (resp., FP) mistake. In order to vary the relative costs of FP and FN, in [8] we introduce a factor $\alpha \geq 0$ such that $c_i^- = \alpha c_i^+$ while keeping $c_i^+ + c_i^- = 2c_i$. Then (1) can be rewritten as

$$\widehat{y}_i = 1 \Longleftrightarrow \widehat{p}_i \left(2c_i - \sum_{j \in \text{child(i)}} H_j \right) \geq \frac{2c_i}{1+\alpha} \ . \tag{2}$$

This is the rule used by HBAYES-CS in our experiments.

2.4 The True Path Rule (TPR) Ensemble

The True Path Rule (TPR) ensemble method [16] not only explicitly takes into account the hierarchical relationships between functional classes, but is also directly inspired by the *true path rule* that can be summarized as follows [14]: "An annotation for a class in the hierarchy is automatically transferred to its ancestors, while genes unannotated for a class cannot be annotated for its descendants". According to this rule, that governs the annotations of both GO and FunCat taxonomies, the proposed ensemble method is characterized by a two-way asymmetric flow of information that traverses the graph-structured ensemble: positive predictions for a node influence in a recursive way its ancestors, while negative predictions influence its offsprings. The resulting ensemble embeds the functional relationships between functional classes that characterize the hierarchical taxonomy. In other words, if a gene is annotated with a specific functional term (functional class), then it is annotated with all the "parent" classes, and with all its ancestors in a recursive way.

The base classifiers estimate local probabilities $\bar{p}_i(\boldsymbol{x})$ that a given example $\boldsymbol{x}$ belongs to class ω_i, but the core of the algorithm is represented by the evaluation phase, where the ensemble provides an estimate of the "consensus" global probability $p_i(\boldsymbol{x})$. Let us consider the set $\phi_i(x)$ of the children of node i for which we have a positive prediction for a given example $\boldsymbol{x}$:

$$\phi_i(\boldsymbol{x}) = \{ j | j \in \text{child}(i), \widehat{y}_j(\boldsymbol{x}) = 1 \} \tag{3}$$

The global consensus probability $\widehat{p}_i(\boldsymbol{x})$ of the ensemble depends both on the local prediction $\bar{p}_i(x)$ and on the prediction of the nodes belonging to $\phi_i(x)$:

$$\widehat{p}_i(\boldsymbol{x}) = \frac{1}{1 + |\phi_i(\boldsymbol{x})|} \left(\bar{p}_i(\boldsymbol{x}) + \sum_{j \in \phi_i(\boldsymbol{x})} \widehat{p}_j(\boldsymbol{x}) \right) \tag{4}$$

The decision $\widehat{y}_i(x)$ at node/class i is set to 1 if $\widehat{p}_i(\boldsymbol{x}) > t$, and to 0 otherwise (a natural choice for t is 0.5). Note that the restriction to nodes belonging to $\phi_i(\boldsymbol{x})$ in the summation of eq. 4 depends on the true path rule: indeed only children nodes for which we have a positive prediction can influence their parent. In the

leaf nodes the sum of eq. 4 disappears and eq. 4 reduces to $\widehat{p}_i(\boldsymbol{x}) = \bar{p}_i(\boldsymbol{x})$. In this way positive predictions propagate from bottom to top. On the contrary if for a given node $\widehat{y}_i(\boldsymbol{x}) = 0$, then this decision is propagated to its subtree.

2.5 The Weighted TPR (TPR-W) Method

In the TPR algorithm there is no way to explicitly balance the local prediction $\bar{p}_i(\boldsymbol{x})$ at node i with the positive predictions coming from its offsprings (eq. 4). By balancing the local predictions with the positive predictions coming from the ensemble, we can explicitly modulate the interplay between local and descendant predictors. To this end we introduced a *weight* w, $0 \leq w \leq 1$, such that if $w = 1$ the decision at node i depends only by the local predictor, otherwise the prediction is shared proportionally to w and $1-w$ between respectively the local parent predictor and the set of its children [15]:

$$\widehat{p}_i(\boldsymbol{x}) = w \cdot \bar{p}_i(\boldsymbol{x}) + \frac{1-w}{|\phi_i(\boldsymbol{x})|} \sum_{j \in \phi_i(\boldsymbol{x})} \widehat{p}_j(\boldsymbol{x}) \tag{5}$$

3 Experimental Setup

In order to compare the capabilities of the HBAYES and TPR methods in hierarchical gene function prediction, we predicted the functions of genes of the unicellular eukaryote *S. cerevisiae* at genome and ontology-wide level using the *FunCat* taxonomy [10], and the data sets described below.

Data sets: In our experiments we used 7 bio molecular data sets, whose characteristics are summarized in Tab. 1. In order to get a not too small set of positive examples for training, for each data set we selected only the FunCat-annotated genes and the classes with at least 20 positive examples. As negative examples we selected for each node/class all genes not annotated to that node/class, but annotated to its parent class. From the data sets we also removed uninformative features (e.g., features with the same value for all the available examples).

Table 1. Data sets

Data set	Description	n. examples	n. feat.	n.classes
Pfam-1	protein domain binary data from *Pfam*	3529	4950	211
Pfam-2	protein domain log E data from *Pfam*	3529	5724	211
Phylo	phylogenetic data	2445	24	187
Expr	gene expression data	4532	250	230
PPI-BG	PPI data from *BioGRID*	4531	5367	232
PPI-VM	PPI data from von Mering experiments	2338	2559	177
SP-sim	Sequence pairwise similarity data	3527	6349	211

Cross validated comparison of ensemble methods: For each ensemble we used gaussian SVMs as base learners. The probabilistic output of the SVMs composing the ensembles has been computed using the sigmoid fitting proposed in [17]. Given a set $\widehat{p}_1, \ldots, \widehat{p}_N$ of trained estimates, we compared on these estimates the results of the HBAYES and TPR ensembles with their corresponding cost-sensitive versions: HBAYES-CS and TPR-W. Both the cost factor α for HBAYES-CS and the w parameter in TPR-W ensembles have been set by internal cross-validation of the F-measure with training data. The threshold t of TPR ensembles has been set to 0.5 in all the experiments. The performance of the ensembles have been compared using external 5-fold cross-validation techniques.

Performances evaluation: Considering the unbalance between positive and negative examples, we could adopt the classical F-score to jointly take into account the precision and recall of the ensemble for each class of the hierarchy. Nevertheless, the classical precision and recall measures, conceived for unstructured classification problems, appear to be inadequate to fully address the hierarchical nature of functional annotation. To this end we used the *Hierarchical F-measure*. This measure is based on the estimation of how much the predicted classification paths correspond to the correct paths. More precisely, given a general taxonomy G representing the graph of the functional classes, for a given gene/gene product x consider the graph $P(x) \subset G$ of the predicted classes and the graph $C(x)$ of the correct classes associated to x, and let be $l(P)$ the set of the leaves (nodes without children) of the graph P. Given a leaf $p \in P(x)$, let be $\uparrow p$ the set of ancestors of the node p that belong to $P(x)$, and given a leaf $c \in C(x)$, let be $\uparrow c$ the set of ancestors of the node c that belong to $C(x)$. Starting from the definitions of Hierarchical Precision (HP), Hierarchical Recall (HR) and Hierarchical F-score (HF) provided in [18], it is easy to demonstrate that in the case of the FunCat taxonomy, since it is structured as a tree, we can simplify HP, HR and HF as follows:

$$HP = \frac{1}{|l(P(x))|} \sum_{p \in l(P(x))} \frac{|C(x) \cap \uparrow p|}{|\uparrow p|}$$

$$HR = \frac{1}{|l(C(x))|} \sum_{c \in l(C(x))} \frac{|\uparrow c \cap P(x)|}{|\uparrow c|}$$

$$HF = \frac{2 \cdot HP \cdot HR}{HP + HR} \tag{6}$$

The hierarchical F-measure expresses the correctness of the structured prediction of the functional classes, taking into account also partially correct paths in the overall hierarchical taxonomy, thus providing in a synthetic way the effectiveness of the hierarchical prediction.

4 Results

4.1 Hierarchical F-Measure Results

As explained in the experimental set-up (Sect. 3), the hierarchical F-measure is a more appropriate performance metric for the hierarchical classification of gene functions. We compared the performances of the considered "vanilla" and cost-sensitive probability reconciliation methods using the Hierarchical F-measure and gaussian SVMs as component classifiers. The results collected in this test are reported in Tab. 2. FLAT ensembles corresponds to predictions directly made by the base learners without any "reconciliation" of the local predictions. According to the Wilcoxon signed-ranks test [19], both HBAYES-CS and TPR-W outperform at 0.01 significance level FLAT, HBAYES and TPR ensemble. No significant difference between HBAYES-CS and TPR-W can be detected (*p-value* $\simeq 0.24$).

Table 2. Hierarchical F-measures comparison between "vanilla" and cost-sensitive hierarchical methods. TPR: True Path Rule hierarchical ensembles; HB-CS: Hierarchical Bayesian bottom-up Cost Sensitive ensembles; TPR-W True Path Rule weighted hierarchical ensembles. Base learners: gaussian SVMs

Data set	FLAT	HBAYES	TPR	HB-CS	TPR-W
Pfam-1	0.1624	0.3359	0.3113	0.4518	0.4188
Pfam-2	0.0402	0.0476	0.1929	0.2030	0.1892
Phylo	0.1196	0.0694	0.2557	0.2682	0.2780
Expr	0.1153	0.0639	0.2390	0.2555	0.2638
PPI-BG	0.0836	0.0847	0.2709	0.2920	0.2315
PPI-VM	0.1720	0.3468	0.3983	0.4329	0.4381
SP-sim	0.1432	0.3246	0.2502	0.4542	0.4501
Average	0.1194	0.1818	0.2732	0.3368	0.3242

4.2 Tuning Precision and Recall in HBAYES-CS and TPR-W Ensembles

In order to compare the capabilities of both the HBAYES and TPR ensemble methods in the modulation of the precision-recall trade-off we tested their performances by varying the values of *alpha* and the parent weight w hyper parameters. Results of this test are reported in (Fig. 1).

The precision/recall characteristics of HBAYES-CS ensemble can be tuned via a single global parameter, the cost factor $\alpha = c_i^- / c_i^+$ (Sect. 2). By setting $\alpha = 1$ we obtain the original version of the hierarchical Bayesian ensemble and by incrementing α we introduce progressively lower costs for positive predictions, thus encouraging the ensemble to make positive predictions. Indeed, by increasing the cost factor, the recall of the ensemble tends to increase (Fig. 1 (a) and (c)). The behavior of the precision is more complex: it tends to increase and then to

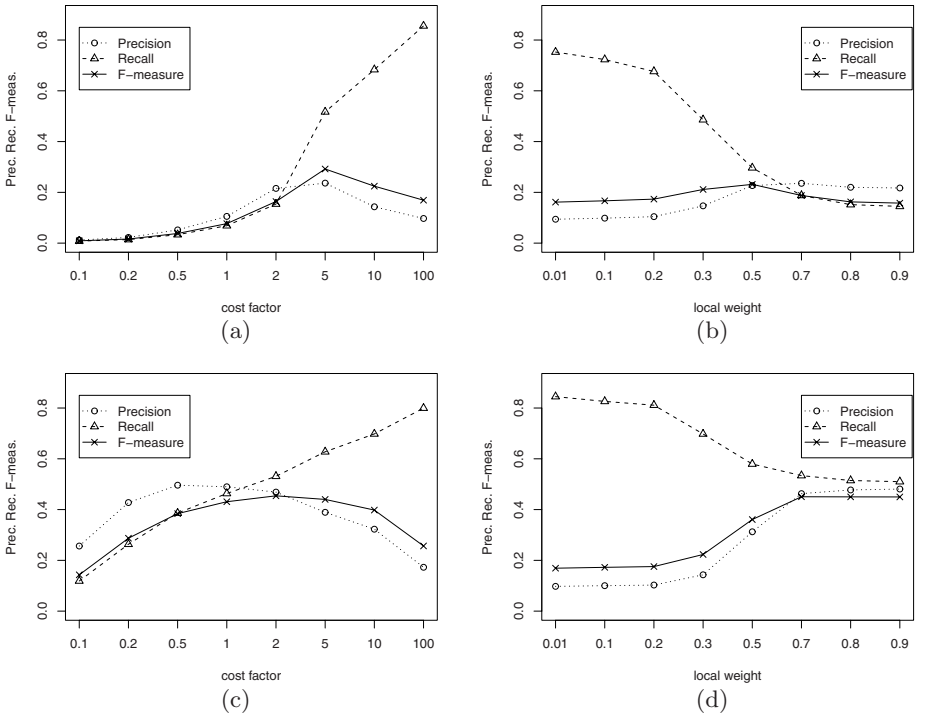

Fig. 1. Precision, Recall and F-measure as a function of the parent weight in TPR-W ensembles and of the global α parameter in HBAYES-CS. PPI BioGRID data: (a) HBAYES-CS (b) TPR-W; Pairwise sequence similarity data: (c) HBAYES-CS (d) TPR-W.

decrease after achieving a maximum. Quite interestingly, the maximum of the hierarchical F-measure is achieved for values of α between 2 and 5 not only for the two data sets reported in Figure 1, but also for all the considered data sets (data not shown).

As seen for HBAYES-CS also the TPR-W ensembles is capable of tuning precision and recall rates, through a single global parameter: the weight w (eq. 5). Fig. 1 (graphs (b) and (d)) shows the hierarchical precision, recall and F-measure as functions of the parameter w. For small values of w (w can vary from 0 to 1) the weight of the decision of the parent local predictor is small, and the ensemble decision depends mainly by the positive predictions of the offsprings nodes (classifiers): as a consequence we obtain a higher hierarchical recall for the TPR-W ensemble. On the contrary higher values of w correspond to a higher weight of the parent predictor, with a resulting higher precision. The opposite trends of precision and recall are quite clear in graphs (b) and (d) of Fig. 1. The best F-score is achieved for "middle" or relatively high values of the w parameter: in practice in most of the analyzed data sets the best F-measure is achieved for w between 0.5 and 0.8, but if we need higher recall rates (at the expense of the precision) we can choose lower w values, and higher values of w

are needed if precision is our first aim. Comparable results were obtained for all the considered data sets (data not shown).

5 Conclusions

In this work we compared the performances of two recently proposed hierarchical gene function prediction methods, the HBAYES and TPR-W ensemble systems. Looking at the results summarized in Tab. 1 it is clear that the usage of hierarchical prediction methods results in a consistent increment in performances if compared with methods that does not take into account the structure of the reference ontology. Also the application of hierarchical methods unable to finely modulate the precision-recall trade-off is suboptimal because the hierarchical F-measure is consistently lowered by the loss in sensitivity associated to the increment in precision due to the reconciliation of the local predictions.

With respect to the comparison of the ability of the cost-sensitive variants of both the HBAYES and TPR methods in the modulation of the overall precision-recall trade-off, the observed results do not allow to define a clear winner. Indeed the performances of the compared methods are quite similar, even if small differences can be found in the performances achieved in the evaluation of different datasets. Both methods share the same top-down strategy to set negative nodes belonging to the subtree rooted at a node predicted as negative by the ensemble, but they pursue very different strategies in the bottom-up computation of the ensemble probabilities. Indeed HBAYES approximates the Bayesian-optimal predictor w.r.t. the H-loss, while TPR is based on a heuristic borrowed from the true path rule. Interestingly enough, these different methods lead to similar results, and their cost-sensitive counterparts significantly outperform FLAT and "vanilla" hierarchical ensembles. The hierarchical algorithms are general enough to be applied to problems other then gene function prediction. Indeed, the cost-sensitive HBAYES and TPR methods, even if conceived for gene function problems, can be applied to other hierarchical classification problems where the descendant classes can be interpreted as parts or subclasses of their corresponding ancestors.

Acknowledgments

The authors thank the reviewers for their comments. The authors gratefully acknowledge partial support by the PASCAL2 Network of Excellence under EC grant no. 216886. This publication only reflects the authors' views.

References

[1] Friedberg, I.: Automated protein function prediction-the genomic challenge. Brief. Bioinformatics 7, 225–242 (2006)
[2] Pena-Castillo, L., et al.: A critical assessment of Mus musculus gene function prediction using integrated genomic evidence. Genome Biology 9(S1) (2008)

[3] Guan, Y., Myers, C., Hess, D., Barutcuoglu, Z., Caudy, A., Troyanskaya, O.: Predicting gene function in a hierarchical context with an ensemble of classifiers. Genome Biology 9(S2) (2008)
[4] Sokolov, A., Ben-Hur, A.: A structured-outputs method for prediction of protein function. In: MLSB 2008, the Second International Workshop on Machine Learning in Systems Biology (2008)
[5] Astikainen, K., Holm, L., Pitkanen, E., Szedmak, S., Rousu, J.: Towards structured output prediction of enzyme function. BMC Proceedings 2(suppl. 4:S2) (2008)
[6] Obozinski, G., Lanckriet, G., Grant, C., Jordan, M.I., Noble, W.S.: Consistent probabilistic output for protein function prediction. Genome Biology 9(S6) (2008)
[7] Jiang, X., Nariai, N., Steffen, M., Kasif, S., Kolaczyk, E.: Integration of relational and hierarchical network information for protein function prediction. BMC Bioinformatics 9(350) (2008)
[8] Cesa-Bianchi, N., Valentini, G.: Hierarchical cost-sensitive algorithms for genome-wide gene function prediction. Journal of Machine Learning Research, W&C Proceedings (to appear)
[9] Valentini, G.: True path rule hierarchical ensembles. In: Benediktsson, J.A., Kittler, J., Roli, F. (eds.) MCS 2009. LNCS, vol. 5519, pp. 232–241. Springer, Heidelberg (2009)
[10] Ruepp, A., Zollner, A., Maier, D., Albermann, K., Hani, J., Mokrejs, M., Tetko, I., Guldener, U., Mannhaupt, G., Munsterkotter, M., Mewes, H.: The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. Nucleic Acids Research 32(18), 5539–5545 (2004)
[11] The Gene Ontology Consortium: Gene ontology: tool for the unification of biology. Nature Genet. 25, 25–29 (2000)
[12] Cesa-Bianchi, N., Gentile, C., Tironi, A., Zaniboni, L.: Incremental algorithms for hierarchical classification. In: Advances in Neural Information Processing Systems, vol. 17, pp. 233–240. MIT Press, Cambridge (2005)
[13] Cesa-Bianchi, N., Gentile, C., Zaniboni, L.: Hierarchical classification: Combining Bayes with SVM. In: Proc. of the 23rd Int. Conf. on Machine Learning, pp. 177–184. ACM Press, New York (2006)
[14] Gene Ontology Consortium: True path rule (2009), `http://www.geneontology.org/GO.usage.shtml#truePathRule`
[15] Valentini, G., Re, M.: Weighted True Path Rule: a multilabel hierarchical algorithm for gene function prediction. In: MLD-ECML 2009, 1st International Workshop on learning from Multi-Label Data, Bled, Slovenia, pp. 133–146 (2009)
[16] Valentini, G.: True Path Rule hierarchical ensembles for genome-wide gene function prediction. IEEE ACM Trans. on Comp. Biol. and Bioinformatics (in press)
[17] Lin, H., Lin, C., Weng, R.: A note on Platt's probabilistic outputs for support vector machines. Machine Learning 68, 267–276 (2007)
[18] Verspoor, K., Cohn, J., Mnizewski, S., Joslyn, C.: A categorization approach to automated ontological function annotation. Protein Science 15, 1544–1549 (2006)
[19] Demsar, J.: Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research 7, 1–30 (2006)