

Cascading Discriminant and Generative Models for Protein Secondary Structure Prediction

Fabienne Thomarat, Fabien Lauer, and Yann Guermeur

LORIA – CNRS, INRIA, Université de Lorraine

Campus Scientifique, BP 239

54506 Vandœuvre-lès-Nancy Cedex, France

{Fabienne.Thomarat,Fabien.Lauer,Yann.Guermeur}@loria.fr

Abstract. Most of the state-of-the-art methods for protein secondary structure prediction are complex combinations of discriminant models. They apply a local approach of the prediction which is known to induce a limit on the expected prediction accuracy. A priori, the use of generative models should make it possible to overcome this limitation. However, among the numerous hidden Markov models which have been dedicated to this task over more than two decades, none has come close to providing comparable performance. A major reason for this phenomenon is provided by the nature of the relevant information. Indeed, it is well known that irrespective of the model implemented, the prediction should benefit significantly from the availability of evolutionary information. Currently, this knowledge is embedded in position-specific scoring matrices which cannot be processed easily with hidden Markov models. With this observation at hand, the next significant advance should come from making the best of the two approaches, i.e., using a generative model on top of discriminant models. This article introduces the first hybrid architecture of this kind with state-of-the-art performance. The conjunction of the two levels of treatment makes it possible to optimize the recognition rate both at the residue level and at the segment level.

Keywords: protein secondary structure prediction, discriminant models, class membership probabilities, hidden Markov models.

1 Introduction

With the multiplication of genome sequencing projects, the number of known protein sequences is growing exponentially. Knowing their (three-dimensional/tertiary) structure is a key in understanding their detailed function. Unfortunately, the experimental methods available to determine the structure, x-ray crystallography and nuclear magnetic resonance (NMR), are highly labor-intensive and do not ensure the production of the desired result (e.g., some proteins simply do not crystallize). As a consequence, the gap between the number of known protein sequences and the number of known protein structures is widening rapidly. To bridge this gap, one must resort to empirical inference. The prediction of protein structure from amino acid sequence, i.e., *ab initio*, is thus

a hot topic in molecular biology. Due to its intrinsic difficulty, it is ordinarily tackled through a divide and conquer approach in which a critical first step is the prediction of the secondary structure, the local, regular structure defined by hydrogen bonds. Considered from the point of view of pattern recognition, this prediction is a three-category discrimination task consisting in assigning a conformational state α -helix (H), β -strand (E) or aperiodic/coil (C), to each residue (amino acid) of a sequence.

For almost half a century, many methods have been developed for protein secondary structure prediction. Since the pioneering work of Qian and Sejnowski [1], state-of-the-art methods are machine learning ones [2–5]. Furthermore, a majority of them shares the original architecture implemented by Qian and Sejnowski. Two sets of classifiers are used in cascade. The classifiers of the first set, named sequence-to-structure, take in input the content of a window sliding on the sequence, or the coding of a multiple alignment, to produce an initial prediction. Those of the second set, named structure-to-structure, take in input the content of a second window sliding on the initial predictions. The structure-to-structure classifiers act both as ensemble methods (combiners) and filters of the initial predictions. The goal of filtering is to increase the biological plausibility of the prediction by making use of the fact that the conformational states of consecutive residues are correlated. Other specifications can be incorporated in the combiners, such as the requirement to output indices of confidence in the prediction or, even better, class posterior probability estimates. Currently, the recognition rate of the best cascades is roughly 80%, depending on the details of the experimental protocol (see [5] for a survey). However, it is commonly admitted that their prediction accuracy faces a strong limiting factor: the fact that local information is not enough to specify utterly the structure. This limitation is only partly overcome by using recurrent neural networks [2]. A natural alternative consists in using generative models. The first hidden Markov model (HMM) [6] dedicated to protein secondary structure prediction was presented in [7]. Since then, new models have regularly been introduced, with the focus being laid on the derivation of an appropriate topology [8, 9]. However, their recognition rate has never exceeded 75% so far [9]. The main reason that can be put forward to explain this disappointing behavior rests in the fact that they are not well-suited to exploit evolutionary information under its standard form, i.e., a profile of multiple alignment, or more precisely a position-specific scoring matrix (PSSM) produced by PSI-BLAST [10]. A generative model that appears more promising to process PSSMs is the dynamic Bayesian network (DBN) [11]. However, the assessment of its potential is still in its infancy [5].

All these observations suggest the assessment of hybrid architectures cascading discriminant and generative models so as to combine the advantages of both approaches. This idea was popularized twenty years ago in the field of speech processing (see for instance [12]), and introduced more recently in bioinformatics, precisely for protein secondary structure prediction [13–15]. In short, discriminant models are used to compute class posterior probability estimates from which the emission probabilities of HMMs are derived, by application of Bayes'

formula. This approach widens the context used for the prediction, and makes it possible to incorporate some pieces of information provided by the biologist, such as syntactic rules. So far, the best prediction method based on this kind of hybrid architecture was YASPIN [15], whose recognition rate is only slightly superior to 77%. In this article, we introduce a new hybrid architecture for protein secondary structure prediction. It is obtained by post-processing the outputs of the prediction method we have developed during the last few years, MSVMpred2 [16, 17], with an “inhomogeneous HMM” (IHMM) [18]. It exhibits state-of-the-art prediction accuracy both at the residue level and at the segment level. The gain of roughly 5% in recognition rate compared to YASPIN is partly due to the recent availability of a very large data set of proteins with known structure and low sequence identity: CM4675.

The organization of the paper is as follows. Section 2 summarizes the main characteristics of MSVMpred2. Section 3 introduces the whole hybrid architecture, and focuses on the features of the upper part of the hierarchy, i.e., the specification and implementation of the IHMM. Experimental results are reported in Section 4. At last, we draw conclusions and outline our ongoing research in Section 5.

2 MSVMpred2

MSVMpred2, the lower part of the hierarchy of treatments, is a cascade of discriminant models implementing the architecture introduced by Qian and Sejnowski. Its main specificities can be summarized as follows. First, the sequence-to-structure prediction is performed by dedicated classifiers. Second, the combiners at the structure-to-structure level are chosen so as to satisfy two requirements: they must output class posterior probability estimates and cover a wide range in terms of capacity. Third, capacity control at this level is implemented through a convex combination of the combiners (with the consequence that the global outputs of the cascade are also class posterior probability estimates). The topology of MSVMpred2 is depicted in Figure 1.

2.1 Sequence-to-Structure Prediction

We first characterize the descriptions (vectors of predictors) $x \in \mathcal{X}$ processed at this initial level of the prediction. The predictors are derived from PSSMs produced by PSI-BLAST. The sliding window is centered on the residue of interest. The description x_i processed at this level to predict the conformational state of the i^{th} residue in the data set is thus obtained by appending rows of the PSSM associated with the sequence to which it belongs. Let n_1 be the integer such that $2n_1 + 1$ is the size of the sliding window. Then, the indices of these rows range from $i' - n_1$ to $i' + n_1$, where i' is the index of the residue of interest in its sequence. Since a PSSM has 20 columns, one per amino acid, this corresponds to $20(2n_1 + 1)$ predictors. More precisely, $\mathcal{X} \subset \mathbb{Z}^{20(2n_1+1)}$.

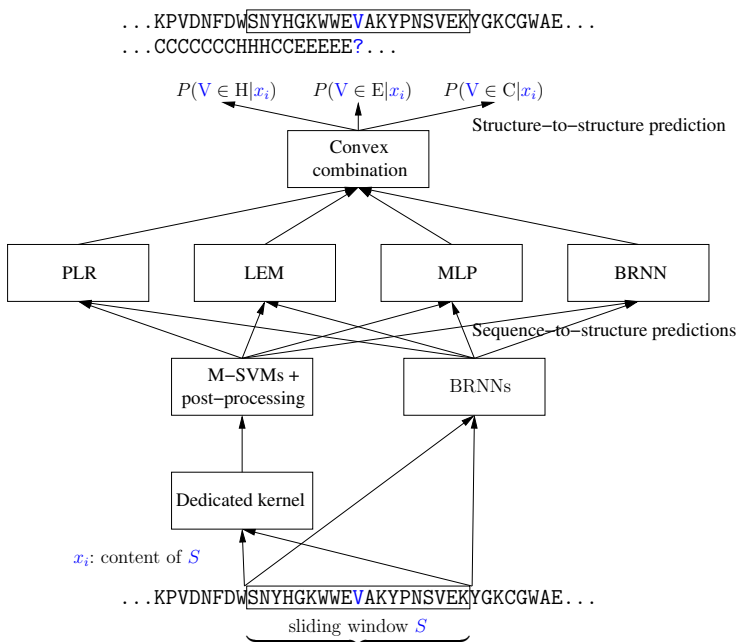


Fig. 1. Topology of MSVMpred2. The method computes estimates of the class posterior probabilities for the residue at the center of the sliding window S , here a valine (V).

Two kinds of classifiers are implemented at this level: multi-class support vector machines (M-SVMs) [19] and bidirectional recurrent neural networks (BRNNs) [20]. The kernel of the M-SVMs is an elliptic Gaussian kernel function applying a weighting on the predictors as a function of their position in the window. This weighting is learned by application of the principle of multi-class kernel target alignment [17]. The BRNNs are recurrent neural networks exploiting a context from both sides of the sequence processed. This makes them especially well-suited for the task at hand. Indeed, they obtain the highest prediction accuracy among all the neural networks assessed so far in protein secondary structure prediction [2, 14, 21]. Contrary to the BRNNs, the M-SVMs do not output class posterior probability estimates. In order to introduce homogeneity among the outputs of the different base classifiers, and more precisely ensure that they all belong to the probability simplex, the outputs of the M-SVMs are post-processed by the polytomous (multinomial) logistic regression (PLR) model [22].

2.2 Structure-to-Structure Prediction

We start with the characterization of the descriptions $z \in \mathcal{Z}$ processed at this level. Let N be the number of classifiers available to perform the sequence-to-structure prediction. The function computed by the j^{th} of these classifiers (after the post-processing in the case of an M-SVM) is denoted $h^{(j)} = \left(h_k^{(j)} \right)_{1 \leq k \leq 3}$.

The second sliding window, of size $2n_2 + 1$, is also centered on the residue of interest. As a consequence, the description z_i processed by the combiners to estimate the probabilities associated with the i^{th} residue in the data set is:

$$z_i = \left(h_k^{(j)}(x_{i+t}) \right)_{1 \leq j \leq N, 1 \leq k \leq 3, -n_2 \leq t \leq n_2} \in U_2^{(2n_2+1)N},$$

where U_2 is the unit 2-simplex.

The four discriminant models used as structure-to-structure classifiers are the PLR, the linear ensemble method (LEM) [23], the multi-layer perceptron (MLP) [24] and the BRNN. They have been listed in order of increasing capacity [25]. Indeed, the PLR and the LEM are linear separators. An MLP using a softmax activation function for the output units and the cross-entropy loss (a sufficient condition for its outputs to be class posterior probability estimates) is an extension of the PLR obtained by adding a hidden layer. The boundaries it computes are nonlinear in its input space. At last, the BRNN can be seen roughly as an MLP operating on an extended description space. The availability of classifiers of different capacities for the second level of the cascade is an important feature of MSVMpred2. It makes it possible to cope with one of the main limiting factors to the performance of modular architectures: overfitting. The capacity control is implemented by the convex combination combining the four structure-to-structure classifiers. The behavior of this combination is predictable: it assigns high weights to the combiners of low complexity when the training set size is small (and the combiners of higher complexity tend to overfit the training set). On the contrary, due to the complexity of the problem, the latter combiners are favored when this size is large (see [17] for an illustration of the phenomenon).

3 Hybrid Prediction Method

In the field of biological sequence processing, the rationale for post-processing the outputs of discriminant models with generative models is two-fold: widening the context exploited for the prediction and incorporating high-level knowledge on the task of interest (mainly in the topology of the generative models). The generative model selected here to meet these goals is an IHMM with three states, one for each of the three conformational states. The advantage of this model compared to the standard HMM rests in the fact that its state transition probabilities are time dependent. This makes it possible to exploit a more suitable model of state durations, a necessary condition to get a high prediction accuracy at the conformational segment level. The global topology of the hierarchy is depicted in Figure 2.

For a given protein sequence, the final prediction is thus obtained by means of the dynamic programming algorithm computing the single best sequence of states (path), i.e., the variant of Viterbi's algorithm dedicated to the IHMM [18]. It must be borne in mind that this calls for a slight adaptation of the formulas, since MSVMpred2 provides estimates of the class posterior probabilities, rather

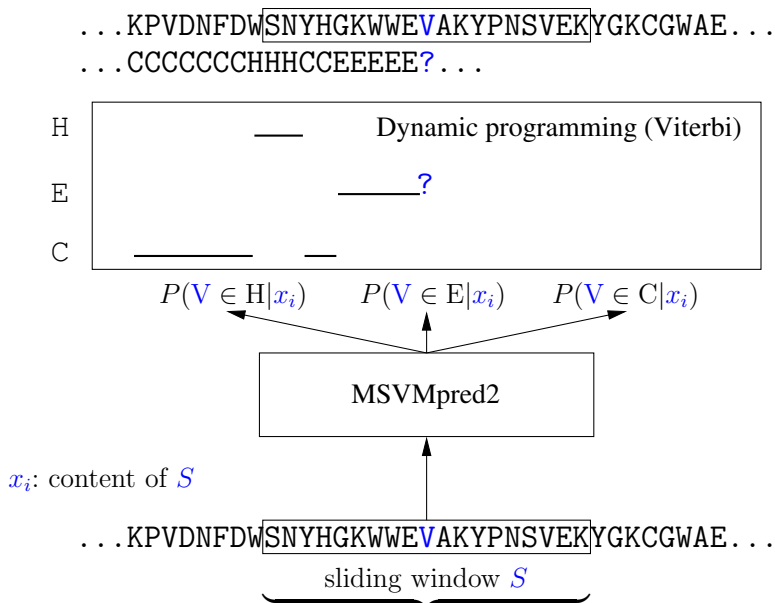


Fig. 2. Topology of the hybrid prediction method. The context available to perform the prediction exceeds that resulting from the combination of the two sliding windows of MSVMpred2.

than emission probabilities. Since the IHMM has exactly one state per conformational state, and the conformational state of each residue in the training set is known, the best state sequence is known for all the sequences of the training set. As a consequence, applying the maximum likelihood principle to derive the initial state distribution and the transition probabilities boils down to computing the corresponding frequencies on the training set.

Implementing a hybrid approach of the prediction is fully relevant only if the quality of the probability estimates computed by the discriminant models is high enough for the generative model to exploit them efficiently. Our hybrid architecture incorporates an optional treatment specifically introduced to address this issue: a basic post-processing of the outputs of MSVMpred2. This post-processing aims at constraining the final prediction so as to keep it in a vicinity of that of MSVMpred2. For each residue, the vector of probability estimates is replaced with a vector that is close to the binary coding of the predicted category. Precisely, given a small positive value ε , the highest of the class membership probability estimates is replaced with $1 - 2\varepsilon$, the two other estimates being replaced with ε . In this setting, the influence of the Viterbi algorithm on the path selected vanishes when ε goes to zero.

4 Experimental Results

4.1 Protein Data Sets

Our prediction method was assessed on two data sets. The first one is the well-known CB513 data set, fully described in [26], whose 513 sequences are made up of 84119 residues. The second one is the newly assembled CM4675 data set. It contains 4675 sequences, for a total of 851523 residues. The corresponding maximum pairwise percentage identity is 20%, i.e., it is low enough to meet the standard requirements of *ab initio* secondary structure prediction.

To generate the PSSMs, the version 2.2.25 of the BLAST package was used. Choosing BLAST in place of the more recent BLAST+ offers the facility to extract more precise PSSMs. Three iterations were performed against the NCBI nr database. The E-value inclusion threshold was set to 0.005 and the default scoring matrix (BLOSUM62) was used. The nr database, downloaded in May 2012, was filtered by pflt [27] to remove low complexity regions, transmembrane spans and coiled coil regions. The initial secondary structure assignment was performed by the DSSP program [28], with the reduction from 8 to 3 conformational states following the CASP method, i.e., H+G \rightarrow H (α -helix), E+B \rightarrow E (β -strand), and all the other states in C (coil).

4.2 Experimental Protocol

The configuration chosen for MSVMpred2 includes the four main models of M-SVMs: the models of Weston and Watkins [29], Crammer and Singer [30], Lee, Lin, and Wahba [31], and the M-SVM² [32]. At the sequence-to-structure level, they are used in parallel with four BRNNs. The programs implementing the different M-SVMs are those of MSVMpack [33], while the 1D-BRNN package is used for the BRNNs. The sizes of the first and second sliding windows are respectively 13 and 15 ($n_1 = 6$ and $n_2 = 7$).

To assess the accuracy of our prediction method, we implemented a distinct experimental protocol for each of the data sets. The reason for this distinction was to take into account the difference in size of the two sets. For CB513, the protocol was basically the 7-fold cross-validation procedure already implemented in [16, 17] (with distinct training subsets for the sequence-to-structure level and the structure-to-structure level). At each step of the procedure, the values of the parameters of the IHMM that had to be inferred were derived using the whole training set. As for CM4675, it was simply split into the following independent subsets: a training set for the kernel of the M-SVMs (500 sequences, 98400 residues), a training set for the sequence-to-structure classifiers (2000 sequences, 369865 residues), a training set for the post-processing of the M-SVMs with a PLR (300 sequences, 52353 residues), a training set for the structure-to-structure classifiers (1000 sequences, 178244 residues), a training set for the convex combination (200 sequences, 34252 residues), and a test set (675 sequences, 118409 residues). Once more, the missing values of the parameters of the IHMM were derived using globally all the training subsets (4000 sequences, 733114 residues).

It can be inferred from the introduction that a secondary structure prediction method must fulfill different requirements in order to be useful for the biologist. Thus, several standard measures giving complementary indications must be used to assess the prediction accuracy [34]. We consider the three most popular ones: the recognition rate Q_3 , Pearson-Matthews correlation coefficients $C_{\alpha/\beta/coil}$, and the segment overlap measure (Sov) in its most recent version (Sov'99).

4.3 Results

The experimental results obtained with MSVMpred2 and the two variants of the hybrid model (with and without post-processing of the outputs of MSVMpred2) are reported in Table 1.

Table 1. Prediction accuracy of MSVMpred2 and the hybrid model on CB513 and CM4675. Results in the last row were obtained with the optional treatment of the outputs of MSVMpred2 described in Section 3.

Method	CB513					CM4675				
	Q_3 (%)	Sov	C_α	C_β	C_{coil}	Q_3 (%)	Sov	C_α	C_β	C_{coil}
MSVMpred2	78.3	74.4	0.74	0.64	0.60	81.8	78.9	0.79	0.73	0.65
Hybrid	77.3	73.1	0.74	0.64	0.57	80.8	77.5	0.78	0.71	0.63
Hybrid ($\varepsilon = 0.01$)	78.3	75.5	0.74	0.64	0.60	81.8	80.0	0.79	0.73	0.65

In applying the two sample proportion test (the one for large samples), one can notice that even when using CB513, the superiority of MSVMpred2 over YASPIN appears statistically significant with confidence exceeding 0.95. Of course, such a statement is to be tempered since the figures available for YASPIN correspond to a different set of protein sequences. The recognition rate is always significantly above 80% when CM4675 is used. This highlights a fact already noticed in [5]: the complexity of the problem calls for the development of complex modular prediction methods such as ours. The feasibility of their implementation increases with the growth of the protein data sets available. The hybrid method is only superior to MSVMpred2 when it implements the post-processing described in Section 3. In that case, the gain is emphasized, as expected, by means of the Sov. This measure increases by the same amount (1.1 point) on both data sets.

The value of ε for which the results of the last row of Table 1 were obtained is a favorable one. We now present a short study of the prediction accuracy as a function of this parameter. Figures 3 and 4 illustrate the main phenomena observed.

The gain in Sov induced by the introduction of the IHMM can be obtained for ε varying in a relatively large interval (the precise boundaries depend on the data set chosen). In addition, to ensure that this gain is not balanced by a decrease of the Q_3 , it suffices to choose a small enough value. This implies that the

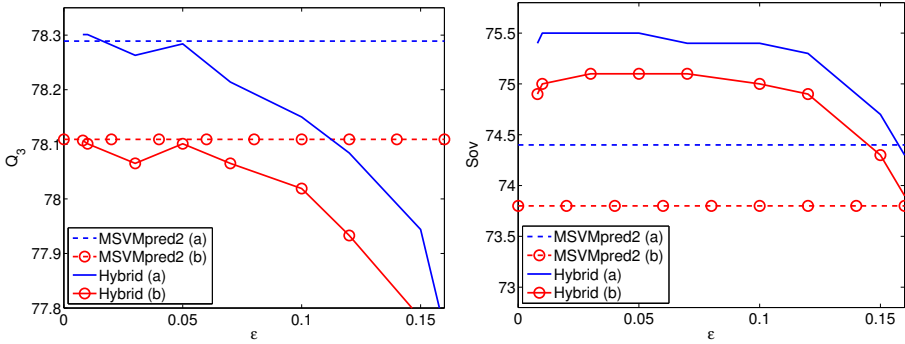


Fig. 3. Prediction accuracy on CB513 in terms of Q_3 (left) and Sov (right) as a function of the value of ε for MSVMpred2 (alone) and the hybrid model. In both cases, two variants of MSVMpred2 are considered: the one specified in Section 4.2 (a) and a simplified one including a single structure-to-structure classifier (b).

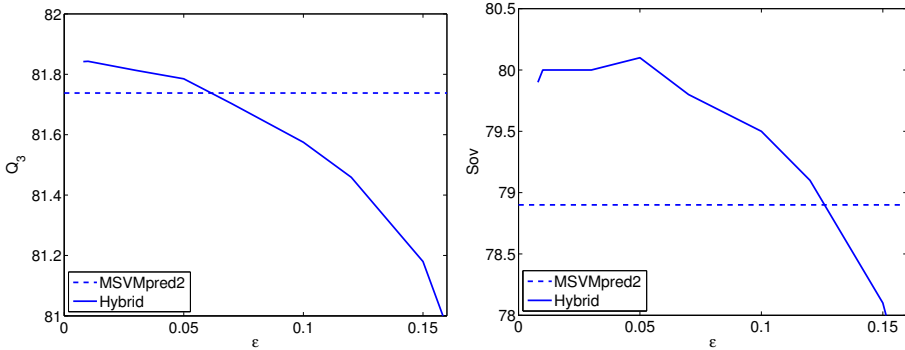


Fig. 4. Prediction accuracy on CM4675 in terms of Q_3 (left) and Sov (right) as a function of the value of ε for MSVMpred2 (alone) and the hybrid model

selection of an appropriate value for ε does not raise particular difficulties. Figure 3 also displays the results obtained with a simplified variant of MSVMpred2. By focusing on the difference between these results and those obtained with the standard MSVMpred2, we also see that the performance of the hybrid model is positively correlated with the prediction accuracy of the classifier providing the class membership probability estimates.

To sum up, these experiments support the thesis that hybrid models can be used to increase the performance of secondary structure prediction methods (at least at the level of the structural elements), while directly benefiting from improvements of the latter. The negative aspect is that we have not yet been able to make full use of the values of the class posterior probability estimates provided by MSVMpred2.

5 Conclusions and Ongoing Research

This article has introduced a new method for protein secondary structure prediction. This method, based on a hierarchical architecture obtained by cascading MSVMpred2 and an IHMM, is the first hybrid model exhibiting state-of-the-art performance. It takes benefit of the availability of a very large data set of proteins with known structure and low sequence identity: CM4675. So far, the main improvement resulting from introducing the generative model is an increase of the Sov'99 measure, i.e., an improvement of the prediction accuracy at the segment level. This should prove especially useful for the biologist using our method as an intermediate step of a tertiary structure prediction.

Obviously, there are many options one can think of to improve our method. A simple one consists in taking benefit of its flexibility to integrate knowledge sources and modules borrowed from the literature [4, 5]. The researcher in machine learning should be primarily interested in the following question: can we expect the class posterior probability estimates produced by MSVMpred2 to become accurate enough to be exploitable as is by the generative model? Answering this fundamental question is currently our main goal.

Acknowledgements. The authors would like to thank C. Magnan for providing them with the CM4675 data set and the 1D-BRNN package.

References

1. Qian, N., Sejnowski, T.J.: Predicting the secondary structure of globular proteins using neural network models. *Journal of Molecular Biology* 202, 865–884 (1988)
2. Pollastri, G., Przybylski, D., Rost, B., Baldi, P.: Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. *Proteins* 47, 228–235 (2002)
3. Cole, C., Barber, J.D., Barton, G.J.: The Jpred 3 secondary structure prediction server. *Nucleic Acids Research* 36, W197–W201 (2008)
4. Kountouris, P., Hirst, J.D.: Prediction of backbone dihedral angles and protein secondary structure using support vector machines. *BMC Bioinformatics* 10, 437 (2009)
5. Aydin, Z., Singh, A., Bilmes, J., Noble, W.S.: Learning sparse models for a dynamic Bayesian network classifier of protein secondary structure. *BMC Bioinformatics* 12, 154 (2011)
6. Rabiner, L.R.: A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77, 257–286 (1989)
7. Asai, K., Hayamizu, S., Handa, K.: Prediction of protein secondary structure by the hidden Markov model. *CABIOS* 9, 141–146 (1993)
8. Martin, J., Gibrat, J.-F., Rodolphe, F.: Analysis of an optimal hidden Markov model for secondary structure prediction. *BMC Structural Biology* 6, 25 (2006)
9. Won, K.-J., Hamelryck, T., Prügél-Bennett, A., Krogh, A.: An evolutionary method for learning HMM structure: prediction of protein secondary structure. *BMC Bioinformatics* 8, 357 (2007)

10. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J.: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25, 3389–3402 (1997)
11. Yao, X.-Q., Zhu, H., She, Z.-S.: A dynamic Bayesian network approach to protein secondary structure prediction. *BMC Bioinformatics* 9, 49 (2008)
12. Krogh, A., Riis, S.K.: Hidden neural networks. *Neural Computation* 11, 541–563 (1999)
13. Guermeur, Y.: Combining discriminant models with new multi-class SVMs. *Pattern Analysis and Applications* 5, 168–179 (2002)
14. Guermeur, Y., Pollastri, G., Elisseeff, A., Zelus, D., Paugam-Moisy, H., Baldi, P.: Combining protein secondary structure prediction models with ensemble methods of optimal complexity. *Neurocomputing* 56, 305–327 (2004)
15. Lin, K., Simossis, V.A., Taylor, W.R., Heringa, J.: A simple and fast secondary structure prediction method using hidden neural networks. *Bioinformatics* 21, 152–159 (2005)
16. Guermeur, Y., Thomarat, F.: Estimating the Class Posterior Probabilities in Protein Secondary Structure Prediction. In: Loog, M., Wessels, L., Reinders, M.J.T., de Ridder, D. (eds.) *PRIB 2011. LNCS (LNBI)*, vol. 7036, pp. 260–271. Springer, Heidelberg (2011)
17. Bonidal, R., Thomarat, F., Guermeur, Y.: Estimating the class posterior probabilities in biological sequence segmentation. In: *SMTDA 2012* (2012)
18. Ramesh, P., Wilpon, J.G.: Modeling state durations in hidden Markov models for automatic speech recognition. In: *ICASSP 1992*, pp. 381–384 (1992)
19. Guermeur, Y.: A generic model of multi-class support vector machine. *International Journal of Intelligent Information and Database Systems* (in press, 2012)
20. Baldi, P., Brunak, S., Frasconi, P., Soda, G., Pollastri, G.: Exploiting the past and the future in protein secondary structure prediction. *Bioinformatics* 15, 937–946 (1999)
21. Chen, J., Chaudhari, N.S.: Cascaded bidirectional recurrent neural networks for protein secondary structure prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 4, 572–582 (2007)
22. Hosmer, D.W., Lemeshow, S.: *Applied Logistic Regression*. Wiley, London (1989)
23. Guermeur, Y.: Combining multi-class SVMs with linear ensemble methods that estimate the class posterior probabilities. *Communications in Statistics* (submitted)
24. Anthony, M., Bartlett, P.L.: *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, Cambridge (1999)
25. Guermeur, Y.: VC theory of large margin multi-category classifiers. *Journal of Machine Learning Research* 8, 2551–2594 (2007)
26. Cuff, J.A., Barton, G.J.: Evaluation and improvement of multiple sequence methods for protein secondary structure prediction. *Proteins* 34, 508–519 (1999)
27. Jones, D.T., Swindells, M.B.: Getting the most from PSI-BLAST. *Trends in Biochemical Sciences* 27, 161–164 (2002)
28. Kabsch, W., Sander, C.: Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577–2637 (1983)
29. Weston, J., Watkins, C.: Multi-class support vector machines. Technical Report CSD-TR-98-04, Royal Holloway, University of London, Department of Computer Science (1998)

30. Crammer, K., Singer, Y.: On the algorithmic implementation of multiclass kernel-based vector machines. *Journal of Machine Learning Research* 2, 265–292 (2001)
31. Lee, Y., Lin, Y., Wahba, G.: Multicategory support vector machines: Theory and application to the classification of microarray data and satellite radiance data. *Journal of the American Statistical Association* 99, 67–81 (2004)
32. Guermeur, Y., Monfrini, E.: A quadratic loss multi-class SVM for which a radius-margin bound applies. *Informatica* 22, 73–96 (2011)
33. Lauer, F., Guermeur, Y.: MSVMpack: a multi-class support vector machine package. *Journal of Machine Learning Research* 12, 2293–2296 (2011)
34. Baldi, P., Brunak, S., Chauvin, Y., Andersen, C.A.F., Nielsen, H.: Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics* 16, 412–424 (2000)