Protein Secondary Structure Prediction
For
A Hybrid Kernel Machine

http://www.jordat.fr/germaine/
Overview

A Hybrid Kernel Machine for Protein Secondary Structure Prediction

- Experimental results
- Hybrid and modular architecture: M-SVM + HMM
- Improving the generalization performance of SSPRO
- M-SVMs for protein secondary structure prediction
  - M-SVMs and the SVM inductive principle
  - Several training algorithms
  - One single architecture
  - Multi-class support vector machines (M-SVMs)
associated with the highest output.

discriminant function: \( f \)

\[(h, x) \int_{\mathbb{R}^n \times \mathbb{X}} dp_{\{h \neq (x)f\}} = (f)H = (\eta)H\]

Find in \( \mathbb{R} \) a function associated with the lowest expected risk (generalization error).

Goal: For a given pattern, find its category

\( [\eta] = \eta \) from \( \mathbb{X} \) into \( \mathbb{R} \)

\( D \) is a family of vector-valued functions on \( \mathbb{X} \)

Learning set: observations \( (x, y) \)

Joint distribution: prediction on \( \mathbb{X} \) fixed but unknown

Input space: \( \mathbb{X} \)

Hypotheses: Empirical data characterizing a joint probability distribution

Multi-class pattern recognition

A Hybrid Kernel Machine for Protein Secondary Structure Prediction

4, 2003
\[ \langle (z)^x \Phi, (z)^y \Phi \rangle = \langle (z)^x (1)^x \Phi \rangle \quad x, y \in (z)^x (1)^x \Phi \]

: \text{the kernel associated with } \mathcal{H}, \text{is related to through RHKS (reproducing kernel Hilbert spaces) where}

\[ \gamma q + \langle (x)\Phi, \gamma q \rangle = \gamma q + (x)^y q = (x)^y q, \quad \{0, \ldots , 1\} \subseteq \gamma A, \quad x \in xA \]

: Functions computed by the architecture are defined by

\textbf{Multi-class Support Vector Machines}

A Hybrid Kernel Machine for Protein Secondary Structure Prediction
\[ \forall \eta \in (x, y) \left( \sum_{\mu=1}^{\infty} \eta \mu \right) = (x, y) \in \{ 0, \ldots, 1 \} \in \forall \eta \]

Represent theorems established that training amounts to finding values of the coefficients \( \eta \) in:

**Dual Formulation**

\[ \left\{ \left( (x, y) \sum_{\mu=1}^{\infty} \mu \right) + \frac{1}{2} \left\| y \right\| \right\} \min_{\eta} \]

**Problem 1**

**Primal Formulation**

Let \( w \) be the training set

\[ \{ (x, y) \in \{ 0, \ldots, 1 \} \} = \sum_{\mu=1}^{\infty} \]

**Training Algorithm of M-SVMs: General Principle**
Problem 3 (M-SVM2 (Güermeur 02))

\[
\left\{ \sum_{i=1}^{n} \sum_{m} c_{i} + \sum_{i=1}^{n} \frac{z_{i}}{y} \right\}_{\gamma \in \mathbb{R}}
\]

\[
(\emptyset \supset (\exists x) \mathcal{C} \neq \emptyset \supset 1), (\mathcal{W} \supset i \supset 1) \quad \text{s.t.} \quad 0 \leq \gamma \leq 1
\]

Problem 2 (M-SVM1 (Vapnik & Blanz 98, Weston & Watkins 98, ...)

Training Algorithms of M-SVMs : Primal Formulation

A Hybrid Kernel Machine for Protein Secondary Structure Prediction
\[(\mathcal{O} \supset (\exists x) \mathcal{C} \neq \emptyset \supset I), (\emptyset \supset (\exists x) \mathcal{C} \neq \emptyset \supset I) \quad \text{primal formulation} \]

Problem 5: M-SVM (Linear SVM) with Lasso 01

\[(\mathcal{O} \supset \emptyset \supset I), (\emptyset \supset \emptyset \supset I) \quad \text{primal formulation} \]

Problem 6: M-SVM (Linear SVM) with Lasso 01

Training algorithms of M-SVM: primal formulation

A Hybrid Kernel Machine for Protein Secondary Structure Prediction

A Hybrid Kernel Machine for Protein Secondary Structure Prediction
 Behaviour of a M-SVM

A Hybrid Kernel Machine for Protein Secondary Structure Prediction
Figure 2 - Hyperplanes and support vectors of a linear M-SVM

Behaviour of a M-SVM
\[ |\{ \zeta > (\xi)_{(\xi)\xi} \nabla \} \mid \forall m \in \{((\nu x)_{(\nu)\xi x})_1^m\} = \frac{\nu}{m} = (\eta)^{m \eta}\nabla \]

Definition 2 (Empirical margin risk) (Breiman & al. 99) The empirical risk with

\[ \left\{ \left( x \right)_{\eta}^{\xi} \frac{\eta}{\max I - \left( x \right)_{\eta}^{\xi}} \right\} \frac{2}{I} = \left( x \right)_{\eta}^{\xi} \nabla \]

function satisfying \( \varphi \) is the function from \( \chi \) into \( \mathbb{R} \). Let \( \eta \) be a function from \( \chi \). Its canonical

\[ \text{otherwise} \quad x \leq |x|, \left\{ x \right\} ^{\eta} = (x)_{\eta}^{\xi} \nabla \]

For \( \nu \in (0, 1] \), let \( \nu : \mathbb{R} \rightarrow [\nabla, -] \) be the piecewise-linear squashing function denoted as

Empirical margin risk and uniform convergence result: multi-class case
\[
\left( (w_m x) \right)^{\infty_1}_{\infty_1} \mathcal{H} \mathcal{D} / \mathcal{D} \right) \mathcal{N}^{w_m X \mathcal{H} \mathcal{D} / \mathcal{D}}_{\mathcal{H} \mathcal{D} / \mathcal{D}} = (w_2 \mathcal{H} \mathcal{D} / \mathcal{D})^{\infty_1}_{\infty_1} \mathcal{N}
\]

\[
\left( \left( \frac{w_m x}{m} \right)^{\infty_1}_{\infty_1} \mathcal{H} \mathcal{D} / \mathcal{D} \right) \mathcal{N}^{w_m X \mathcal{H} \mathcal{D} / \mathcal{D}}_{\mathcal{H} \mathcal{D} / \mathcal{D}} = \left( (x) \mathcal{H} \mathcal{D} / \mathcal{D} \right)^{\infty_1}_{\infty_1} \mathcal{N}^{w_m X \mathcal{H} \mathcal{D} / \mathcal{D}}_{\mathcal{H} \mathcal{D} / \mathcal{D}} \mathcal{A} \mathcal{H} \mathcal{D} / \mathcal{D} \mathcal{A}
\]

where

\[
\{ \mathcal{H} \in \mathcal{D} / (\mathcal{D} \mathcal{H} \mathcal{D} / \mathcal{D}) \}^{\infty_1}_{\infty_1} \mathcal{N}^{w_m X \mathcal{H} \mathcal{D} / \mathcal{D}}_{\mathcal{H} \mathcal{D} / \mathcal{D}} = \mathcal{D} \mathcal{H} \mathcal{D} / \mathcal{D}
\]

bounded above by

\[
\text{risk } R(\mathcal{H}) \text{ of a function } \mathcal{H} \text{ computed by a numerical } \mathcal{D} / \mathcal{D} \text{-class discriminant model } \mathcal{H}
\]

is risk R(\mathcal{H}) of a function \( \mathcal{H} \) computed by a numerical \( \mathcal{D} / \mathcal{D} \)-class discriminant model \( \mathcal{H} \) is

indpendently from \( \mathcal{D} \). With probability at least \( 1 - \delta \), for every value of \( \mathcal{H} \) in \( (0, 1] \), the

Theorem 1 (Eliseeff & Helmbold, 1999): Let \( \mathcal{H} \) be a m-sample of examples drawn

Empirical margin risk and uniform convergence result: multi-class case

A Hybrid Kernel Machine for Protein Secondary Structure Prediction

2003
Then the graph dimension of $\mathcal{H}$ is defined to be the VC dimension of the space $\mathcal{S}(\mathcal{H})$.

$\{\mathcal{H} \in \mathcal{F}(\eta) \mid (\eta) \mathcal{S} \text{ is a graph space of } \mathcal{H} \text{ and the graph space of } \mathcal{H} \text{ is } (\eta) \mathcal{S} \}$

For any $\eta \in \mathcal{F}(\mathcal{H})$, the graph $g$ of functions on a set $X$ taking their values in a countable set. For any $h \in \mathcal{H}$, the graph $g$ of

Definition 4 (Graph dimension (Dudley 87, Natarajan 89)) Let $\mathcal{H}$ be a set of

size is finite, or to infinity otherwise.

which maps a value $y$ to the size of the largest set $\mathcal{F}$-shattered by functions of $\mathcal{H}$, if this

dimension function $\mathcal{H}$ of the set $\mathcal{H}$ is a function from the positive real numbers to the integers

The vector $v_i$ is then said to witness the $\mathcal{F}$-shattering of $\mathcal{H}$ by $\mathcal{H}$. The $\mathcal{F}$-shattering

$(m \geq i \geq 1) \land \eta \geq \eta^\mathcal{H}(\gamma - (\eta)^v)$

vector $v_i$ such that $\mathcal{H} \in \mathcal{V}(\eta)$ satisfying there is a function $\eta \in \mathcal{V}(\eta)$ for each binary

said to be $\mathcal{F}$-shattered by $\mathcal{H}$ if there is a vector $v_i$ of each binary

Definition 3 (Fat-Shattering dimension (Kearns & Schapire 90)) Let $\mathcal{H}$ be a set

Extended notions of VC dimension

A Hybrid Kernel Machine for Protein Secondary Structure Prediction
$M$-fat-shatter dimension: extension of the fat-shatter dimension to the
multivariable case and scale-sensitive version of the graph dimension

Defintion 5 ($M$-fat-shatter dimension) Let $\mathcal{X}$ be a set of functions on a set $\mathcal{C}$

The couple \((\mathcal{C}, \nu)\) is then said to witness the $M$-fat-shattering of some by $\mathcal{C}$.
The

\( (m \geq \nu \geq 1) \wedge \exists \, \nu (\nu - (x)^{\nu}) \)

by functions of $\mathcal{X}$, if this size is finite, or to infinity otherwise.

numbers to the integers which maps a value $\nu$ to the size of the largest set $M$-fat-shattered

$M$-fat-shatter dimension $\mathcal{M}$-fat of the set $\mathcal{X}$ is a function from the positive real

function $\mathcal{H} \in (\mathcal{D} \supset \mathcal{Y} \supset 1)$, such that for each binary vector $\mathcal{w}$

there is a $\mathcal{u} \in \mathcal{W}$ and a vector $w \in \mathcal{W}$ such that for each binary vector $\mathcal{v}$

taking their values in $\mathbb{R}^d$. For $\nu$ a subset $\nu$, a subset $\nu$, a subset $\nu$, a subset $\nu$, a subset $\nu$, a subset $\nu$. A Hybrid Kernel Machine for Protein Secondary Structure Prediction
Main Difficulty: \( W\)-Fact cannot be bounded in terms of \( W\)-Fact.

\[
\exists \varepsilon \left( \forall x \in A \right) \sum_{\Theta} \leq \left( \forall x \right) \text{positive value of } \varepsilon
\]

The following bound holds true for all the different components of the functions. Then the following bound holds true for \( \Theta \geq \kappa \geq 1 \), \( \left( \forall x \right) \text{be the sets of real-valued functions } h \text{ corresponding to } x \text{ into } \mathbb{R} \). Let \( \Theta \) \( \text{be a set of vector-valued functions } h \text{ from a set } \) where

\[
\left( \forall x \right) \text{positive value of } \varepsilon
\]

Theorem 2 Let \( \Theta \) be a set of functions from \( x \) into \( \mathbb{R} \). For every value of \( \Theta \in \text{multi-class case} \)

Generalization of Sauer's Lemma: multi-class case
\[
\left( \left( \frac{\mathcal{E}}{\mathcal{M}(x) \Phi} \right) \right) \mathcal{O} = (\mathcal{E})^{\mathbb{M}, \mathbb{N}, \mathbb{Q}} \{ \mathcal{O}, \cdots, \mathcal{I} \} \ni \mathcal{A}
\]

Suppose that is included in a ball of radius \((x) \Phi\) and that the vectors \(w\) satisfy

\[
\{q - \gamma q + \langle (x) \Phi, \mathcal{I}n - \gamma m \rangle \} \text{ minimized over } \gamma \mathcal{N} = (x) \gamma \mathcal{N}
\]

Corresponding to the \(k\)-th component of the \(\gamma \mathcal{N}\) functions

\[
\gamma \mathcal{N} \mathcal{V}
\]

by a \(\mathcal{M}\)-SVM. For \(\gamma \mathcal{N} \mathcal{V}\) be the set of real-valued functions \(\gamma \mathcal{N} \mathcal{V}\) functions \(\gamma \mathcal{N} \mathcal{V}\) computed

\[
\text{Theorem 4. Let } \mathcal{N} \text{ be the set of vector-valued functions } \gamma \mathcal{N} \mathcal{V}, \text{ computed}
\]

Dependence of the capacity on the control term
Theorem 4–6 Let \( \mathcal{H} \) be a set of functions computed by a \( \text{M-SVM} \). Suppose that the biases satisfy the following bounds hold true:

\[
(m', \epsilon, ' m) \mathcal{H} \mathcal{V} \epsilon, ' (m', \mathcal{H} \mathcal{V} \epsilon, ' m) \leq (m', \mathcal{H} \mathcal{V} \epsilon, ' m) \mathcal{H} \mathcal{V} \epsilon, ' (m', \mathcal{H} \mathcal{V} \epsilon, ' m)
\]

Theorem 5 Let \( \mathcal{H} \) be a set of functions from \( X \) into \( \mathbb{R} \). For every value of \( \epsilon \) and \( \gamma \), simple pathways to bound the covering numbers:

\[
(m', \epsilon, ' m) \mathcal{H} \mathcal{V} \epsilon, ' (m', \mathcal{H} \mathcal{V} \epsilon, ' m) \leq (m', \mathcal{H} \mathcal{V} \epsilon, ' m) \mathcal{H} \mathcal{V} \epsilon, ' (m', \mathcal{H} \mathcal{V} \epsilon, ' m)
\]
\[
\left( \frac{1 + (u) \log u}{w} + 1 \right) \log u \geq \left( \log u \right)^2
\]

for all \( n \in \mathbb{N} \).

\[ \frac{-1}{1} \left( \left[ 1 + (u) \log u \right] \right) \log u \geq \left( \log u \right)^2 \]

Theorem 7 (Maurey) \( \mathcal{L} \in \mathcal{L} \subseteq \mathcal{H} \).

\[ ((\mathcal{H} \mathcal{L}) \mathcal{L})^u e = (\mathcal{L})^u e \quad \sup_{\|x\| \leq \epsilon} = \|x\| \]

Definition 7 (Entropy numbers of a bounded linear operator) \( \mathcal{L} \in \mathcal{L} \subseteq \mathcal{H} \).

\[ \{ u \geq (d, H; e, N) / 0 < e \} \quad \mathcal{L} = (H)^u e \]

Definition 6 (Entropy numbers) \( \mathcal{L} \in \mathcal{L} \subseteq \mathcal{H} \).

Covariant numbers and entropy numbers

A Hybrid Kernel Machine for Protein Secondary Structure Prediction
\[ u \supset (w, \gamma)_{0, \infty} \rightarrow \alpha \subseteq (w, L) u \gamma \]

Then

\[ \left[ \langle (w, \Phi) \gamma, m \rangle \cdots \langle (w, \Phi) \gamma, m \rangle \cdots \langle (\eta, \Phi) \gamma, m \rangle \cdots \langle (\eta, \Phi) \gamma, m \rangle \right] \leftarrow m \]

\[ \left[ w, S \right] \leftarrow 2 \]

Theorem 8 Let \( \gamma \) be the linear operator given by \( S^m \) with \( m \gamma \leq 2 \) satisfies \( \gamma \) defined as above, with the additional constraint that each

Covering numbers and entropy numbers
All these multi-class SVMs are equivalent

The same set of primal variables generates solutions for the three problems

<table>
<thead>
<tr>
<th>Solution</th>
<th>Objective function</th>
<th>$p_0$</th>
<th>$\sum_i \alpha_i \theta_i \sum_j \sum_k \theta_i | \tau_k - \hat{y}<em>{ik} |</em>\delta$ = $(\hat{y}', m')$</th>
<th>$\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(p_{0}^{(1)} g(1) \sum_{(1)} \sum_{(1)} \sum_{(1)} q_{1}^{(1)} m)$</td>
<td>$I_{O}$</td>
<td>$p_0 = \hat{y} m = \gamma \sum_{(1)}$</td>
<td>$(\hat{y}', m')$</td>
<td>$\delta$</td>
</tr>
<tr>
<td>$(I + \theta)^{1} v(1 + \theta)^{1} z_{1}^{(1)} q_{1}^{(1)} m)$</td>
<td>$I_{O}(1 + \theta)$</td>
<td>-</td>
<td>$(\hat{y}', m')$</td>
<td>$\delta$</td>
</tr>
<tr>
<td>$(z^{(1)} v_{1}^{(1)} q_{1}^{(1)} m)$</td>
<td>$I_{O}$</td>
<td>-</td>
<td>$(\hat{y}', m')$</td>
<td>$\delta$</td>
</tr>
</tbody>
</table>

$\sum_i \alpha_i \theta_i \sum_j \sum_k \theta_i \| \tau_k - \hat{y}_{ik} \|_\delta$ = $(\hat{y}', m')$

$\delta$ = Add. cost.

<table>
<thead>
<tr>
<th>$p_0 = \hat{y} m = \gamma \sum_{(1)}$</th>
<th>$\sum_i \alpha_i \theta_i \sum_j \sum_k \theta_i | \tau_k - \hat{y}<em>{ik} |</em>\delta$ = $(\hat{y}', m')$</th>
<th>$\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>Germann &amp; al. 69.98</td>
<td>66</td>
</tr>
<tr>
<td>Breidenbier &amp; al.</td>
<td>66.98</td>
<td>66</td>
</tr>
<tr>
<td>Westen &amp; Walkins</td>
<td>86.98</td>
<td>86</td>
</tr>
<tr>
<td>Vapnik &amp; Bains</td>
<td>86.98</td>
<td>86</td>
</tr>
</tbody>
</table>

Add. cost.

Objective function

Constraints on the hyperplanes

A Hybrid Kernel Machine for Protein Secondary Structure Prediction

April 14, 2003
Basic notions about proteins

- Proteins: macromolecules made up of amino acids
- 20 amino acids, each of them represented by a letter (A, R, N, D, C, E, ...)

Hierarchical description of the conformation

- Primary structure (sequence of amino acids) ⇐⇒ sequencing
- Secondary structure (sequence of structural elements) ⇐⇒ circular dichroism
- Tertiary structure (three-dimensional structure) ⇐⇒ X-ray, NMR

Importance of structure prediction

- Knowing the structure is a prerequisite to gain a thorough understanding of the function
- Difficulty to determine the structure experimentally
Modular and hierarchical approach of the prediction

\[ \ldots \text{...VKPVDNFDSNYHGKWWEVAKYPNSVEKYGKGWAE...} \]

<table>
<thead>
<tr>
<th>H</th>
<th>Dynamic Programming (IHMM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>?</td>
</tr>
<tr>
<td>C</td>
<td></td>
</tr>
</tbody>
</table>

\[ \frac{P(V \text{ in } H | S)}{P(V \text{ in } E | S)} \quad \frac{P(V \text{ in } C | S)}{ \text{(If necessary)}} \]

Post-processing

Filtering neural nets

Combiner

MLRC, M-SVM...

NN 1

NN 2

NN 3

SOPMA

\[(\text{GOR IV}) \quad M-SVM\]

SIMPA 96

\[\ldots \text{VKPVDNFDSNYHGKWWEVAKYPNSVEKYGKGWAE...} \]

Content of the sliding window \(W\)
stacked generalization

- Training of the M-SVM + post-processing (perception with softmax units) + HHMM
- Homology with the training set of the BRNNs
- New database of 1096 protein sequences (235551 amino acids) exhibiting no

Experimental protocol

Improve performance by implementing a superior ensemble method = M-SVMs

Goal

- SSPROZ: unweighted average on the outputs of the BRNNs
- Experts: II "Bihitensional Recurrent Neural Networks (BRNNs)
- Data: profiles of alignments derived from PSI-BLAST

SSPROZ (Pollastri et al., 2002)

Collaboration with the Bioinformatics group of Prof. P. Baldi at the UCI

Combination of the BRNNs of SSPROZ

A Hybrid Kernel Machine for Protein Secondary Structure Prediction

\[ \text{April 14, 2003} \]
\[
\left\{ \begin{array}{l}
\text{(1)} \sum_{\text{max}} \frac{I + ((\exists \mathcal{S}) \Delta \mathcal{E}, (\exists \mathcal{S}) \cdot \mathcal{E})}{\mathcal{E}} - ((\exists \mathcal{S}) \Delta \mathcal{E}, (\exists \mathcal{S}) \cdot \mathcal{E}) \\
\text{min}
\end{array} \right\} - \int \sum_{\mathcal{E}} \frac{u}{I} = (\mathfrak{g})_{\alpha} \mathcal{S}
\]

So, coefficients (Rost \& al., 94; Zemla \& al., 99)

\[\left( P_{\text{pred}} - P_{\text{qo}} \right) \sum_{\mathcal{E}} \frac{u}{I} = \varphi \theta \]

Root mean square deviation (R.M.S.D.)

\[
\frac{(\varphi \theta + \varphi u)(\varphi n + \varphi u)(\varphi \theta + \varphi d)(\varphi n + \varphi d)}{\varphi \theta \varphi n - \varphi u \varphi d} = \gamma \mathcal{C}
\]

Pearson's/Mattheew's correlation coefficients

\( \theta \) : Recognition rate at the residue level

Different measures of prediction accuracy
Table 1 - Combination of the II BRNNs of SSpro2 with two-class and multi-class SVMs

<table>
<thead>
<tr>
<th></th>
<th>So°</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>69.0</td>
<td>67.6</td>
<td>68.9</td>
<td>68.5</td>
<td>68.1</td>
<td>66.1</td>
<td>69.6</td>
</tr>
<tr>
<td>68.6</td>
<td>69.0</td>
<td>66.9</td>
<td>67.3</td>
<td>69.0</td>
<td>67.1</td>
<td>67.6</td>
</tr>
<tr>
<td>72.0</td>
<td>72.2</td>
<td>72.4</td>
<td>72.0</td>
<td>72.2</td>
<td>72.2</td>
<td>72.4</td>
</tr>
<tr>
<td>72.9</td>
<td>72.9</td>
<td>73.6</td>
<td>72.0</td>
<td>72.2</td>
<td>72.2</td>
<td>72.9</td>
</tr>
<tr>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
</tr>
<tr>
<td>0.63</td>
<td>0.63</td>
<td>0.63</td>
<td>0.63</td>
<td>0.63</td>
<td>0.63</td>
<td>0.63</td>
</tr>
<tr>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
</tr>
<tr>
<td>79.5</td>
<td>79.6</td>
<td>79.6</td>
<td>79.6</td>
<td>79.6</td>
<td>79.6</td>
<td>79.6</td>
</tr>
<tr>
<td>87.8</td>
<td>87.9</td>
<td>87.8</td>
<td>87.3</td>
<td>87.6</td>
<td>87.7</td>
<td>87.8</td>
</tr>
<tr>
<td>98.8</td>
<td>96.9</td>
<td>98.8</td>
<td>98.8</td>
<td>97.7</td>
<td>98.7</td>
<td>98.8</td>
</tr>
<tr>
<td>77.08</td>
<td>77.12</td>
<td>77.09</td>
<td>77.10</td>
<td>77.14</td>
<td>76.91</td>
<td>76.94</td>
</tr>
<tr>
<td>M-SVM2</td>
<td>M-SVM2</td>
<td>SVM1a'</td>
<td>M-LDP</td>
<td>M-LDP</td>
<td>M-LDP</td>
<td>M-LDP</td>
</tr>
<tr>
<td>77.03</td>
<td>77.03</td>
<td>77.03</td>
<td>77.03</td>
<td>77.03</td>
<td>77.03</td>
<td>77.03</td>
</tr>
</tbody>
</table>

Relative prediction accuracy of combiners
<table>
<thead>
<tr>
<th></th>
<th>68.7</th>
<th>68.6</th>
<th>68.8</th>
<th>SVM + g + e</th>
<th>M-MLP</th>
</tr>
</thead>
<tbody>
<tr>
<td>69.4</td>
<td>68.7</td>
<td>69.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>74.8</td>
<td>74.8</td>
<td>74.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>72.3</td>
<td>72.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.63</td>
<td>0.63</td>
<td>0.63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.73</td>
<td>0.73</td>
<td>0.72</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>79.5</td>
<td>79.5</td>
<td>79.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>87.8</td>
<td>87.8</td>
<td>87.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>87.1</td>
<td>87.1</td>
<td>87.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>77.26</td>
<td>77.06</td>
<td>77.02</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2 - Combination of the II BRNNS of SSpros and PSI-BLAST derived profiles

Taking the PSI-BLAST profiles into account improves performance

A Hybrid Kernel Machine for Protein Secondary Structure Prediction
the window are not taken into account.

The possibility of insertions/deletions, the nature of the substitutions and the position in

\[ \left( \frac{\epsilon \theta \gamma}{\epsilon x^t x^q u - \epsilon \gamma u} - (1 + u\gamma) \right) \] 
\[ \text{dx} \in \left( \frac{\epsilon \theta \gamma}{\epsilon \|x - x\|} - \right) \] 
\[ \text{dx} \in \left( L x, x \right) \]

**Kernel function computed by the kernel**

\[ \text{canonical code of the amino acid in position i} \] 
\[ \text{(binary vector)} \]

\[ (i + u\gamma = |M|) \] 
\[ \gamma \in (1 + u\gamma) \{ 0, 1 \} \] 
\[ \in \{ u x, \cdots, x, \cdots, u x \} = x \]

**Canonical code of the content of the sliding window**

M-SVM as basic classifier - Shortcomings of standard kernels

A Hybrid Kernel Machine for Protein Secondary Structure Prediction

47
Taking into account the substitutions
Stochastic Gradient Descent

Training Algorithm

\[
\begin{cases}
\text{otherwise} & \frac{\alpha}{1 - \alpha} = (x^T x)^{-1} \\
\text{if } C = (x) & \frac{\alpha}{1 - \alpha} = (x^T x)^{-1} \\
\end{cases}
\]

Pic. 5 - Optimal choice of the representatives of the categories

Kernel target alignment: multi-class case (Variant 02)

M-SVM as basic classifier - Optimization of the Kernel

A Hybrid Kernel Machine for Protein Secondary Structure Prediction
Influence of the position in the window

\[
\left( \frac{202}{\left( \langle \langle x^I \rangle \langle x^I \rangle \rangle - 1 \rangle \frac{1}{2} \theta \sum_{u=\frac{r}{2}}^{\frac{r}{2}} - \right) dx \exp = (x^I x)^{\theta} y}
\]

Analytical expression

Dedicated Kernel

A Hybrid Kernel Machine for Protein Secondary Structure Prediction
TABLE 3—Prediction accuracy on a set of 1096 non-homologous protein sequences

<table>
<thead>
<tr>
<th></th>
<th>0.52</th>
<th>0.50</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.51</td>
<td>0.50</td>
<td>0.59</td>
<td></td>
</tr>
<tr>
<td>0.54</td>
<td>0.53</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>0.69</td>
<td>0.68</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>M-SVM</td>
<td>0.48</td>
<td>0.58</td>
<td></td>
</tr>
</tbody>
</table>

Prediction accuracy of a M-SVM without post-processing
Implementation of N-Best Algorithms

Best sequence of states obtained with a variation of Viterbi's algorithm

Prediction

\[
\begin{align*}
\mathbb{P}(\text{transition probability including a transition model}) & \quad (\mathbb{P} \geq p \geq 1, 0 \leq \gamma \leq 1) \\
\{\mathbb{P}\} & = \mathcal{A} \\
\text{parameters} & \quad \forall, \quad \mathcal{B} & = \mathcal{A} \\
\text{One state for each structural state} & \\
\text{Underlying Hidden Markov Model: HMM (Ramshaw & Wilson 92)}
\end{align*}
\]

Post-processing of the posterior probability estimates

A Hybrid Kernel Machine for Protein Secondary Structure Predictions
post-processing become similar

annealing-like algorithm so that the corresponding frequencies before and after

Transition probabilities adapted with a Monte-Carlo and simulated

Training algorithm

Top of the hierarchy: higher-level treatments

A Hybrid Kernel Machine for Protein Secondary Structure Prediction
Lengths of the conformational segments, observed and predicted
As basic classifiers, should prove superior to standard conjunctionist architectures

- Little difference in generalization performance
- As ensemble methods, provide good generalization performance

M-SVMs for secondary structure prediction

- Possibility to develop new machines
- New justifications of the training algorithms of all the M-SVMS proposed so far
- New pathway to bound the risk of multi-class discriminant models

Theory of M-SVMS

Conclusions
Global optimization of all the components of the hierarchical architecture

Fusion of additional knowledge sources provided by the biologists

Additional work on the design and (efficient) implementation of specific kernels

M-SVMs for secondary structure prediction / biological sequence processing

Specification of optimization methods devoted to the new machines

(Boucheron & al. 99, Bartlett & al. 02, Bougaut 02)

Comparison with studies involving data dependent capacity measures

Theory and Implementation of M-SVMs

Work in progress