Algorithms for the Hard Pre-Image Problem of String Kernels and the General Problem of String Prediction

Sébastien Giguère[1,2], **Amélie Rolland**[2], François Laviolette[2] Mario Marchand[2]

[1]Institute for Research in Immunology and Cancer
University of Montreal

[2]Department of Computer Science and Software Engineering
Laval University

ICML, 2015

Motivation

Word

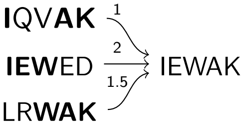
↓

declaring

Phonemes

c	o	f	f	e	e
↓	↓	↓	↓	↓	↓
k	c	f	-	i	-

Protein



Motivation

Word

↓

declaring

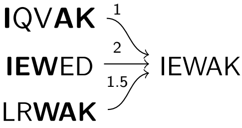
Phonemes

c o f f e e

↓ ↓ ↓ ↓ ↓ ↓

k c f - i -

Protein



Pre-image and general string prediction problem

Find the string $\mathbf{y} \in \mathcal{Y}$ of an input $\mathbf{x} \in \mathcal{X}$ maximizing a joint score function $F_{\mathbf{w}}$

$$h(\mathbf{x}) = \operatorname*{argmax}_{\mathbf{y} \in \mathcal{Y}} F_{\mathbf{w}}(\mathbf{x}, \mathbf{y}).$$

Structured prediction problem

- Dataset $\mathcal{S} = \{(\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_m, \mathbf{y}_m)\} \in \mathcal{X} \times \mathcal{Y}$
- Feature map $\boldsymbol{\phi}_{\mathcal{X}} : \mathcal{X} \to \mathcal{H}_{\mathcal{X}}$, $\boldsymbol{\phi}_{\mathcal{Y}} : \mathcal{Y} \to \mathcal{H}_{\mathcal{Y}}$
- $K_{\mathcal{X}}(\mathbf{x}, \mathbf{x}') = \langle \boldsymbol{\phi}_{\mathcal{X}}(\mathbf{x}), \boldsymbol{\phi}_{\mathcal{X}}(\mathbf{x}') \rangle \ \forall (\mathbf{x}, \mathbf{x}') \in \mathcal{X}^2$
- Linear operators $\mathbf{W} : \mathcal{H}_{\mathcal{X}} \to \mathcal{H}_{\mathcal{Y}}$

Structured prediction problem

- Dataset $\mathcal{S} = \{(\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_m, \mathbf{y}_m)\} \in \mathcal{X} \times \mathcal{Y}$
- Feature map $\boldsymbol{\phi}_{\mathcal{X}} : \mathcal{X} \to \mathcal{H}_{\mathcal{X}}$, $\boldsymbol{\phi}_{\mathcal{Y}} : \mathcal{Y} \to \mathcal{H}_{\mathcal{Y}}$
- $K_{\mathcal{X}}(\mathbf{x}, \mathbf{x}') = \langle \boldsymbol{\phi}_{\mathcal{X}}(\mathbf{x}), \boldsymbol{\phi}_{\mathcal{X}}(\mathbf{x}') \rangle \; \forall (\mathbf{x}, \mathbf{x}') \in \mathcal{X}^2$
- Linear operators $\mathbf{W} : \mathcal{H}_{\mathcal{X}} \to \mathcal{H}_{\mathcal{Y}}$

Pre-image problem

The predicted output $\mathbf{y}^{\mathbf{W}}(\mathbf{x})$ of $\mathbf{W}$ on input $\mathbf{x} \in \mathcal{X}$ is given by solving

$$\mathbf{y}^{\mathbf{W}}(\mathbf{x}) = \operatorname*{argmin}_{\mathbf{y} \in \mathcal{Y}} \left\| \boldsymbol{\phi}_{\mathcal{Y}}(\mathbf{y}) - \mathbf{W} \boldsymbol{\phi}_{\mathcal{X}}(\mathbf{x}) \right\| , \tag{1}$$

where $\| \cdot \|$ denotes the L_2 norm in $\mathcal{H}_{\mathcal{Y}}$ [Cortes et al., 2007].

Pre-image problem

Let $\mathcal{H}_\mathcal{Y}$ be the feature space of the n-gram kernel with $\mathcal{A} = [\mathrm{A}, \mathrm{B}]$ and $n = 2$.

$$\mathbf{W}\phi_\mathcal{X}(\mathbf{x}) = \begin{array}{c} \\ [\end{array} \begin{array}{cccc} \mathrm{AA} & \mathrm{AB} & \mathrm{BA} & \mathrm{BB} \\ 1.1 & 1 & 0 & 0 \end{array} \begin{array}{c} \\ ] \end{array}$$

Pre-image problem

Let $\mathcal{H}_{\mathcal{Y}}$ be the feature space of the n-gram kernel with $\mathcal{A} = [\mathrm{A}, \mathrm{B}]$ and $n = 2$.

	AA	AB	BA	BB
$\mathbf{W}\phi_{\mathcal{X}}(\mathbf{x}) =$	1.1	1	0	0
$\phi_{\mathcal{Y}}(\mathrm{AAA}) =$	2	0	0	0
$\phi_{\mathcal{Y}}(\mathrm{AAB}) =$	1	1	0	0
$\phi_{\mathcal{Y}}(\mathrm{ABA}) =$	0	1	1	0
...				

Pre-image problem

Let $\mathcal{H}_\mathcal{Y}$ be the feature space of the n-gram kernel with $\mathcal{A} = [\mathsf{A}, \mathsf{B}]$ and $n = 2$.

	AA	AB	BA	BB
$\mathbf{W}\phi_\mathcal{X}(\mathbf{x}) =$	1.1	1	0	0
$\phi_\mathcal{Y}(\mathbf{AAB}) =$	1	1	0	0

Pre-image problem

Let $\mathcal{H}_\mathcal{Y}$ be the feature space of the n-gram kernel with $\mathcal{A} = [\mathrm{A}, \mathrm{B}]$ and $n = 2$.

$$\mathbf{W}\phi_\mathcal{X}(\mathbf{x}) = \begin{array}{c} \\ [\end{array} \begin{array}{cccc} \mathrm{AA} & \mathrm{AB} & \mathrm{BA} & \mathrm{BB} \\ \mathbf{1.1} & 1 & 0 & 0 \end{array} \begin{array}{c} \\ ] \end{array}$$

Pre-image problem

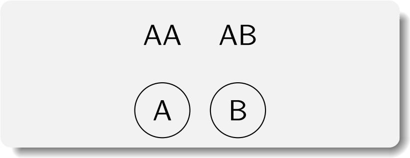
Let $\mathcal{H}_\mathcal{Y}$ be the feature space of the n-gram kernel with $\mathcal{A} = [\mathrm{A}, \mathrm{B}]$ and $n = 2$.

$$\mathbf{W}\boldsymbol{\phi}_\mathcal{X}(\mathbf{x}) = \begin{array}{c} \\ [\end{array} \begin{array}{cccc} \mathrm{AA} & \mathrm{AB} & \mathrm{BA} & \mathrm{BB} \\ 1 & 1 & 0 & 0 \end{array} \begin{array}{c} \\ ] \end{array}$$

Pre-image problem

Let $\mathcal{H}_\mathcal{Y}$ be the feature space of the n-gram kernel with $\mathcal{A} = [\mathrm{A}, \mathrm{B}]$ and $n = 2$.

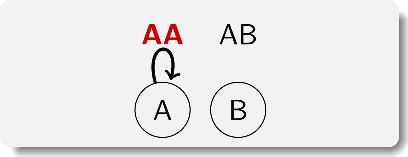
$$\mathbf{W}\phi_\mathcal{X}(\mathbf{x}) = \begin{matrix} & \mathrm{AA} & \mathrm{AB} & \mathrm{BA} & \mathrm{BB} & \\ [& 1 & 1 & 0 & 0 &] \end{matrix}$$



Pre-image problem

Let $\mathcal{H}_\mathcal{Y}$ be the feature space of the n-gram kernel with $\mathcal{A} = [\text{A}, \text{B}]$ and $n = 2$.

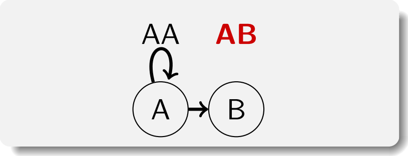
$$\mathbf{W}\phi_\mathcal{X}(\mathbf{x}) = \begin{array}{c} \\ \left[\right. \end{array} \begin{array}{cccc} \text{AA} & \text{AB} & \text{BA} & \text{BB} \\ 1 & 1 & 0 & 0 \end{array} \begin{array}{c} \\ \left. \right] \end{array}$$



Pre-image problem

Let $\mathcal{H}_\mathcal{Y}$ be the feature space of the n-gram kernel with $\mathcal{A} = [\text{A}, \text{B}]$ and $n = 2$.

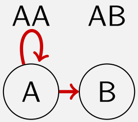
$$\mathbf{W}\phi_\mathcal{X}(\mathbf{x}) = \begin{array}{c} \\ \left[\right. \end{array} \begin{array}{cccc} \text{AA} & \text{AB} & \text{BA} & \text{BB} \\ 1 & 1 & 0 & 0 \end{array} \begin{array}{c} \\ \left. \right] \end{array}$$



Pre-image problem

Let $\mathcal{H}_{\mathcal{Y}}$ be the feature space of the n-gram kernel with $\mathcal{A} = [\mathrm{A}, \mathrm{B}]$ and $n = 2$.

	AA	AB	BA	BB
$\mathbf{W}\phi_{\mathcal{X}}(\mathbf{x}) = [$	1	1	0	0



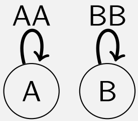
$\mathcal{O}(|\mathbf{y}|)$

$\mathbf{y} = \mathbf{AAB}$

Pre-image problem

Let $\mathcal{H}_\mathcal{Y}$ be the feature space of the n-gram kernel with $\mathcal{A} = [\mathrm{A}, \mathrm{B}]$ and $n = 2$.

$$\mathbf{W}\phi_\mathcal{X}(\mathbf{x}) = \begin{array}{c} \\ [\end{array} \begin{array}{cccc} \mathrm{AA} & \mathrm{AB} & \mathrm{BA} & \mathrm{BB} \\ 1 & 0 & 0 & 1 \end{array} \begin{array}{c} \\ ] \end{array}$$



Normalized pre-image problem

Given $\widehat{\phi}_{\mathcal{Y}}(\mathbf{y}) \stackrel{\text{def}}{=} \frac{\phi_{\mathcal{Y}}(\mathbf{y})}{\|\phi_{\mathcal{Y}}(\mathbf{y})\|}$, the pre-image problem becomes

$$\mathbf{y}^{\mathbf{w}}(\mathbf{x}) = \underset{\mathbf{y}\in\mathcal{Y}}{\operatorname{argmax}} \left\langle \widehat{\phi}_{\mathcal{Y}}(\mathbf{y}), \mathbf{W}\phi_{\mathcal{X}}(\mathbf{x}) \right\rangle . \tag{2}$$

Normalized pre-image problem

Given $\widehat{\phi}_{\mathcal{Y}}(\mathbf{y}) \stackrel{\text{def}}{=} \frac{\phi_{\mathcal{Y}}(\mathbf{y})}{\|\phi_{\mathcal{Y}}(\mathbf{y})\|}$, the pre-image problem becomes

$$\mathbf{y}^{\mathbf{w}}(\mathbf{x}) = \operatorname*{argmax}_{\mathbf{y}\in\mathcal{Y}} \left\langle \widehat{\phi}_{\mathcal{Y}}(\mathbf{y}), \mathbf{W}\phi_{\mathcal{X}}(\mathbf{x}) \right\rangle . \tag{2}$$

Vector-valued ridge regression

$$\mathbf{W}\phi_{\mathcal{X}}(x) = \sum_{i=1}^{m}\sum_{j=1}^{m} \widehat{\phi}_{\mathcal{Y}}(y_i) A_{i,j} K_{\mathcal{X}}(x_j, x) \tag{3}$$

$$\mathbf{y}^{\mathbf{w}}(\mathbf{x}) = \operatorname*{argmax}_{\mathbf{y}\in\mathcal{Y}} \sum_{i=1}^{m}\sum_{j=1}^{m} \frac{K_{\mathcal{Y}}(\mathbf{y}_i, \mathbf{y})}{\sqrt{K_{\mathcal{Y}}(\mathbf{y}_i, \mathbf{y}_i) K_{\mathcal{Y}}(\mathbf{y}, \mathbf{y})}} A_{i,j} K_{\mathcal{X}}(\mathbf{x}_j, \mathbf{x}) \tag{4}$$

String prediction for classification and regression

Given an alphabet $\mathcal{A}$ and $\mathcal{S} = \{(\mathbf{y}_1, r_1), \ldots, (\mathbf{y}_m, r_m)\} \in \mathcal{A}^* \times \mathbb{R}$.

Predictor

Real-valued prediction function h (KRR, SVR, SVM)

$$h(\mathbf{y}) = \sum_{i=1}^{m} \alpha_i K_{\mathcal{Y}}(\mathbf{y}_i, \mathbf{y}), \tag{5}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)$ is obtained by minimizing an objective function and where $K_{\mathcal{Y}}$ is assumed to be a string kernel.

String prediction for classification and regression

Given an alphabet $\mathcal{A}$ and $\mathcal{S} = \{(\mathbf{y}_1, r_1), \ldots, (\mathbf{y}_m, r_m)\} \in \mathcal{A}^* \times \mathbb{R}$.

Predictor

Real-valued prediction function h (KRR, SVR, SVM)

$$h(\mathbf{y}) = \sum_{i=1}^{m} \alpha_i K_{\mathcal{Y}}(\mathbf{y}_i, \mathbf{y}), \tag{5}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)$ is obtained by minimizing an objective function and where $K_{\mathcal{Y}}$ is assumed to be a string kernel.

Maximization

Find string $\mathbf{y}$ maximizing $h(\mathbf{y})$

$$\mathbf{y}^h = \underset{\mathbf{y} \in \mathcal{A}^*}{\operatorname{argmax}}\ h(\mathbf{y}). \tag{6}$$

Normalization

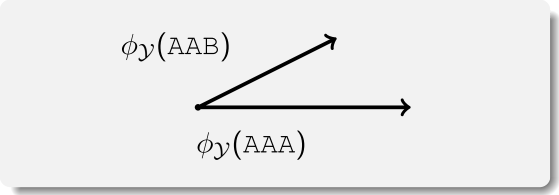
$\mathbf{y}$	$\mathbf{y}'$	$K_{\mathcal{Y}}(\mathbf{y},\mathbf{y}')$	$\frac{K_{\mathcal{Y}}(\mathbf{y},\mathbf{y}')}{\sqrt{K_{\mathcal{Y}}(\mathbf{y},\mathbf{y})K_{\mathcal{Y}}(\mathbf{y}',\mathbf{y}')}}$
AAB	AAB	5	**1**
AAA	AAB	**6**	0.89

Table : Example using the 1-gram kernel where $K_{\mathcal{Y}}(\mathbf{y},\mathbf{y}') > K_{\mathcal{Y}}(\mathbf{y}',\mathbf{y}')$, $||\phi_{\mathcal{Y}}(\texttt{AAB})|| = \sqrt{5}$, $||\phi_{\mathcal{Y}}(\texttt{AAA})|| = \sqrt{9}$.

Normalization

$\mathbf{y}$	$\mathbf{y}'$	$K_{\mathcal{Y}}(\mathbf{y}, \mathbf{y}')$	$\frac{K_{\mathcal{Y}}(\mathbf{y},\mathbf{y}')}{\sqrt{K_{\mathcal{Y}}(\mathbf{y},\mathbf{y})K_{\mathcal{Y}}(\mathbf{y}',\mathbf{y}')}}$
AAB	AAB	5	**1**
AAA	AAB	**6**	0.89

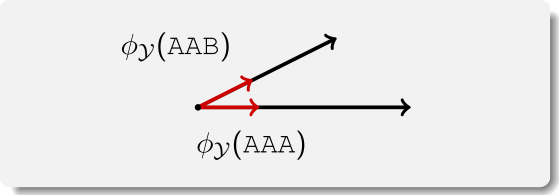
Table : Example using the 1-gram kernel where $K_{\mathcal{Y}}(\mathbf{y}, \mathbf{y}') > K_{\mathcal{Y}}(\mathbf{y}', \mathbf{y}')$, $||\phi_{\mathcal{Y}}(\texttt{AAB})|| = \sqrt{5}$, $||\phi_{\mathcal{Y}}(\texttt{AAA})|| = \sqrt{9}$.



Normalization

$\mathbf{y}$	$\mathbf{y}'$	$K_{\mathcal{Y}}(\mathbf{y}, \mathbf{y}')$	$\frac{K_{\mathcal{Y}}(\mathbf{y},\mathbf{y}')}{\sqrt{K_{\mathcal{Y}}(\mathbf{y},\mathbf{y})K_{\mathcal{Y}}(\mathbf{y}',\mathbf{y}')}}$
AAB	AAB	5	**1**
AAA	AAB	**6**	0.89

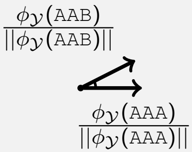
Table : Example using the 1-gram kernel where $K_{\mathcal{Y}}(\mathbf{y}, \mathbf{y}') > K_{\mathcal{Y}}(\mathbf{y}', \mathbf{y}')$, $||\phi_{\mathcal{Y}}(\texttt{AAB})|| = \sqrt{5}$, $||\phi_{\mathcal{Y}}(\texttt{AAA})|| = \sqrt{9}$.



Normalization

$\mathbf{y}$	$\mathbf{y}'$	$K_{\mathcal{Y}}(\mathbf{y}, \mathbf{y}')$	$\frac{K_{\mathcal{Y}}(\mathbf{y},\mathbf{y}')}{\sqrt{K_{\mathcal{Y}}(\mathbf{y},\mathbf{y})K_{\mathcal{Y}}(\mathbf{y}',\mathbf{y}')}}$
AAB	AAB	5	**1**
AAA	AAB	**6**	0.89

Table : Example using the 1-gram kernel where $K_{\mathcal{Y}}(\mathbf{y}, \mathbf{y}') > K_{\mathcal{Y}}(\mathbf{y}', \mathbf{y}')$, $||\phi_{\mathcal{Y}}(\texttt{AAB})|| = \sqrt{5}$, $||\phi_{\mathcal{Y}}(\texttt{AAA})|| = \sqrt{9}$.



String prediction for classification and regression

Normalized problem

Consequently, we are interested at solving

$$\mathbf{y}^{\widehat{h}} = \operatorname*{argmax}_{\mathbf{y} \in \mathcal{A}^*} \widehat{h}(\mathbf{y}) \,, \tag{7}$$

where

$$\widehat{h}(\mathbf{y}) = \sum_{i=1}^{m} \alpha_i \frac{K_{\mathcal{Y}}(\mathbf{y}_i, \mathbf{y})}{\sqrt{K_{\mathcal{Y}}(\mathbf{y}_i, \mathbf{y}_i) K_{\mathcal{Y}}(\mathbf{y}, \mathbf{y})}} \,. \tag{8}$$

String prediction for classification and regression

Normalized problem

Consequently, we are interested at solving

$$\mathbf{y}^{\widehat{h}} = \underset{\mathbf{y}\in\mathcal{A}^*}{\operatorname{argmax}}\ \widehat{h}(\mathbf{y})\,, \tag{7}$$

where

$$\widehat{h}(\mathbf{y}) = \sum_{i=1}^{m} \alpha_i \frac{K_{\mathcal{Y}}(\mathbf{y}_i, \mathbf{y})}{\sqrt{K_{\mathcal{Y}}(\mathbf{y}_i, \mathbf{y}_i) K_{\mathcal{Y}}(\mathbf{y}, \mathbf{y})}}\,. \tag{8}$$

String prediction

Unified problem

Both normalized problems reduce to solving

$$\mathbf{y}^\star = \operatorname*{argmax}_{\mathbf{y}\in\mathcal{A}^*} \frac{1}{\sqrt{K_\mathcal{Y}(\mathbf{y},\mathbf{y})}} \sum_{i=1}^{m} \beta_i K_\mathcal{Y}(\mathbf{y}_i,\mathbf{y}) \,. \qquad (9)$$

- Structured prediction : $\beta_i = \beta_i(\mathbf{x}) = \sum_{j=1}^{m} \frac{A_{i,j} K_\mathcal{X}(\mathbf{x}_j,\mathbf{x})}{\sqrt{K_\mathcal{Y}(\mathbf{y}_i,\mathbf{y}_i)}}$
- Predictor maximization : $\beta_i = \frac{\alpha_i}{\sqrt{K_\mathcal{Y}(\mathbf{y}_i,\mathbf{y}_i)}}$

String kernels

Substrings (n-gram)

2-gram

ABCDE : AB BC CD DE

String kernels

Substrings (*n*-gram)

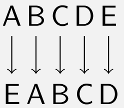
2-gram

ABCDE : AB BC CD DE

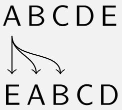
Position

Hamming

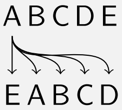
Weighted Degree



Oligo



N-Gram



String kernels

Substrings (*n*-gram)

2-gram

ABCDE : AB BC CD DE

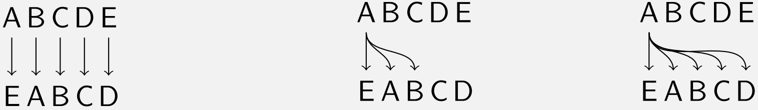
Position

Hamming

Weighted Degree

Oligo

N-Gram



Similarity

- Amino acids
- Phonemes

Generic String kernel

GS kernel [Giguère et al., 2013]

$$GS(\mathbf{y}, \mathbf{y}', n, \sigma_p, \sigma_c) \stackrel{\text{def}}{=}$$

$$\sum_{l=1}^{n} \sum_{i=0}^{|\mathbf{y}|-l} \sum_{j=0}^{|\mathbf{y}'|-l} \underbrace{\exp\left(\frac{-(i-j)^2}{2\sigma_p^2}\right)}_{\text{position similarity}} \underbrace{\exp\left(\frac{-\|\boldsymbol{\psi}^l(y_{i+1},..,y_{i+l}) - \boldsymbol{\psi}^l(y'_{j+1},..,y'_{j+l})\|^2}{2\sigma_c^2}\right)}_{l\text{-gram similarity}},$$

where $\boldsymbol{\psi}^l : \mathcal{A}^l \rightarrow \mathbb{R}^{ld}$ maps each letter of the l-gram to a real-valued vector containing d components.

Generic String kernel

GS kernel [Giguère et al., 2013]

$$GS(\mathbf{y}, \mathbf{y}', n, \sigma_p, \sigma_c) \stackrel{\text{def}}{=}$$

$$\sum_{l=1}^{n} \sum_{i=0}^{|\mathbf{y}|-l} \sum_{j=0}^{|\mathbf{y}'|-l} \underbrace{\exp\left(\frac{-(i-j)^2}{2\sigma_p^2}\right)}_{\text{position similarity}} \underbrace{\exp\left(\frac{-\|\boldsymbol{\psi}^l(y_{i+1},..,y_{i+l}) - \boldsymbol{\psi}^l(y'_{j+1},..,y'_{j+l})\|^2}{2\sigma_c^2}\right)}_{l\text{-gram similarity}},$$

where $\boldsymbol{\psi}^l : \mathcal{A}^l \to \mathbb{R}^{ld}$ maps each letter of the l-gram to a real-valued vector containing d components.

String kernels generalized by the GS

Hamming distance, Dirac delta, Oligo, RBF, Blended Spectrum, Blended Spectrum RBF, Weighted degree, Weighted degree RBF.

METHOD

Case when $\sigma_p = 0$

Pre-image problem when $\sigma_p = 0$

In that case, $K_{\mathcal{Y}}(\mathbf{y}, \mathbf{y}) = c$ for all strings of length ℓ

$$\mathbf{y}^\star = \operatorname*{argmax}_{\mathbf{y} \in \mathcal{A}^\ell} \frac{1}{\sqrt{c}} \sum_{i=1}^{m} \beta_i K_{\mathcal{Y}}(\mathbf{y}_i, \mathbf{y}) \,. \tag{10}$$

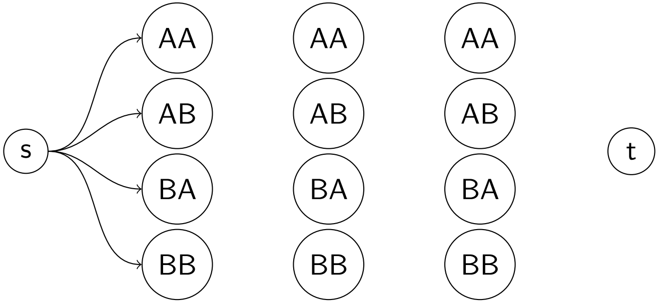
Case when $\sigma_p = 0$

$$n = 2, \mathcal{A}^n = \{\mathrm{AA}, \mathrm{AB}, \mathrm{BA}, \mathrm{BB}\}, \ell = 4$$

Case when $\sigma_p = 0$

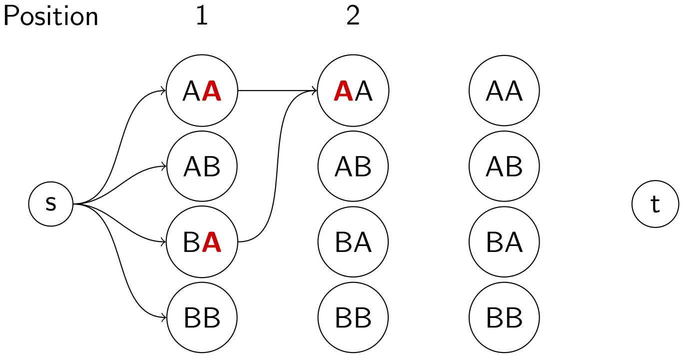
$$n = 2, \mathcal{A}^n = \{AA, AB, BA, BB\}, \ell = 4$$

Position 1



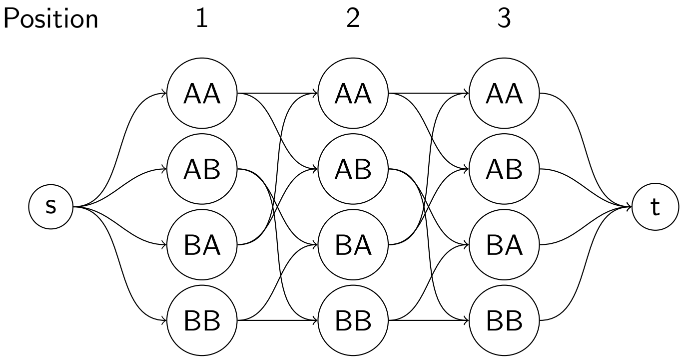
Case when $\sigma_p = 0$

$$n = 2, \mathcal{A}^n = \{AA, AB, BA, BB\}, \ell = 4$$



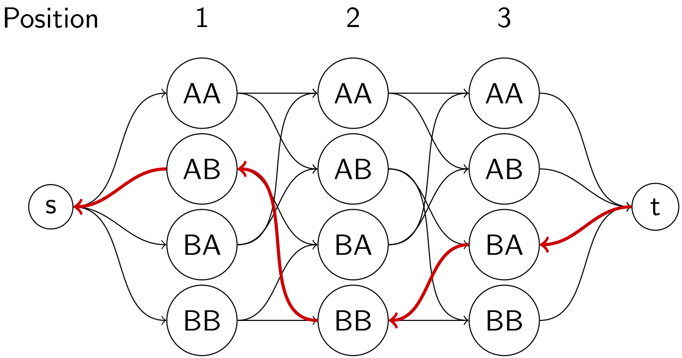
Case when $\sigma_p = 0$

$$n = 2, \mathcal{A}^n = \{\text{AA}, \text{AB}, \text{BA}, \text{BB}\}, \ell = 4$$



Case when $\sigma_p = 0$

$$n = 2, \mathcal{A}^n = \{AA, AB, BA, BB\}, \ell = 4$$

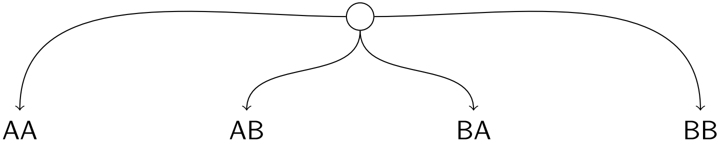


$\mathbf{y}^\star = ABBA$

$\mathcal{O}(\ell|\mathcal{A}|^n)$

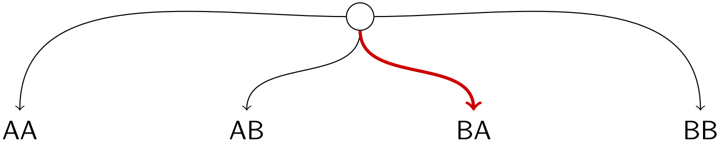
Branch and bound

$$n = 2, \mathcal{A}^n = \{\mathrm{AA}, \mathrm{AB}, \mathrm{BA}, \mathrm{BB}\}, \ell = 4$$



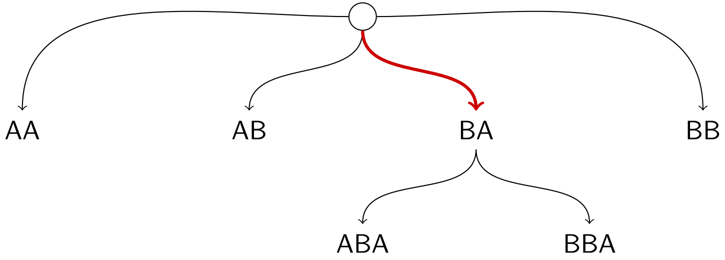
Branch and bound

$$n = 2, \mathcal{A}^n = \{\mathrm{AA}, \mathrm{AB}, \mathrm{BA}, \mathrm{BB}\}, \ell = 4$$



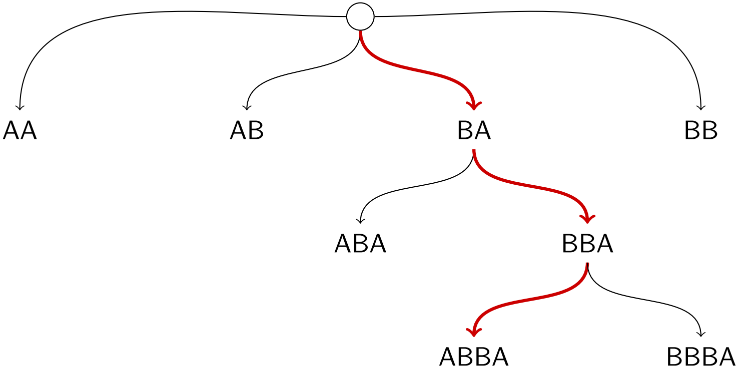
Branch and bound

$$n = 2, \mathcal{A}^n = \{\text{AA}, \text{AB}, \text{BA}, \text{BB}\}, \ell = 4$$



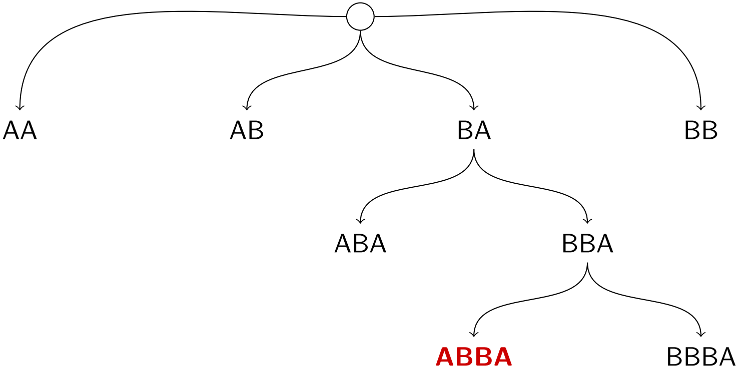
Branch and bound

$$n = 2, \mathcal{A}^n = \{\text{AA}, \text{AB}, \text{BA}, \text{BB}\}, \ell = 4$$



Branch and bound

$$n = 2, \mathcal{A}^n = \{\text{AA}, \text{AB}, \text{BA}, \text{BB}\}, \ell = 4$$



Case when $\sigma_p > 0$

- Bound value of partial solution $\mathbf{y}' = y'_1, .., y'_p$
- String of ℓ characters: $y_1, .., y_{(\ell-p)}, y'_1, .., y'_p$

$$F(\mathbf{y}', \ell) \geq \max_{\mathbf{y} \in \mathcal{A}^{\ell-p} \times \{\mathbf{y}'\}} \frac{1}{\sqrt{K_{\mathcal{Y}}(\mathbf{y}, \mathbf{y})}} \sum_{i=1}^{m} \beta_i K_{\mathcal{Y}}(\mathbf{y}_i, \mathbf{y}) \,. \tag{11}$$

Case when $\sigma_p > 0$

- Bound value of partial solution $\mathbf{y}' = y'_1, .., y'_p$
- String of ℓ characters: $y_1, .., y_{(\ell-p)}, y'_1, .., y'_p$

$$F(\mathbf{y}', \ell) \geq \max_{\mathbf{y} \in \mathcal{A}^{\ell-p} \times \{\mathbf{y}'\}} \frac{1}{\sqrt{K_{\mathcal{Y}}(\mathbf{y}, \mathbf{y})}} \sum_{i=1}^{m} \beta_i K_{\mathcal{Y}}(\mathbf{y}_i, \mathbf{y}) . \quad (11)$$

Case when $\sigma_p > 0$

- Bound value of partial solution $\mathbf{y}' = y'_1, .., y'_p$
- String of ℓ characters: $y_1, .., y_{(\ell-p)}, y'_1, .., y'_p$

$$F(\mathbf{y}', \ell) \geq \max_{\mathbf{y} \in \mathcal{A}^{\ell-p} \times \{\mathbf{y}'\}} \frac{1}{\sqrt{K_{\mathcal{Y}}(\mathbf{y}, \mathbf{y})}} \sum_{i=1}^{m} \beta_i K_{\mathcal{Y}}(\mathbf{y}_i, \mathbf{y}) \,. \tag{11}$$

Case when $\sigma_p > 0$

- Bound value of partial solution $\mathbf{y}' = y'_1, .., y'_p$
- String of ℓ characters: $y_1, .., y_{(\ell-p)}, y'_1, .., y'_p$

$$F(\mathbf{y}', \ell) \geq \max_{\mathbf{y} \in \mathcal{A}^{\ell-p} \times \{\mathbf{y}'\}} \frac{1}{\sqrt{K_{\mathcal{Y}}(\mathbf{y}, \mathbf{y})}} \sum_{i=1}^{m} \beta_i K_{\mathcal{Y}}(\mathbf{y}_i, \mathbf{y}) \,. \tag{11}$$

Problem decomposition

Let $F(\mathbf{y}', \ell) \stackrel{\text{def}}{=} \frac{1}{\sqrt{f(\mathbf{y}', \ell)}} g(\mathbf{y}', \ell)$, where

$$f(\mathbf{y}', \ell) \leq \min_{\mathbf{y} \in \mathcal{A}^{\ell-p} \times \{\mathbf{y}'\}} K_{\mathcal{Y}}(\mathbf{y}, \mathbf{y}), \tag{12}$$

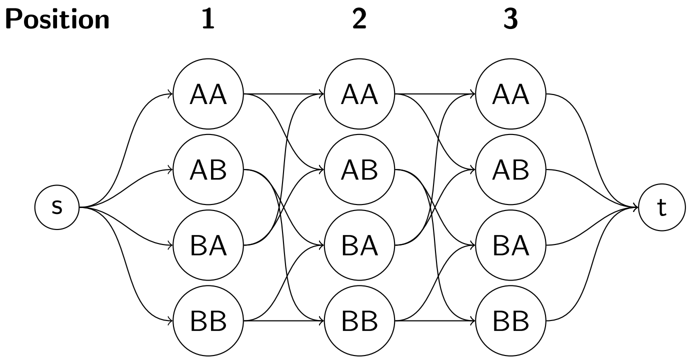
$$g(\mathbf{y}', \ell) \geq \max_{\mathbf{y} \in \mathcal{A}^{\ell-p} \times \{\mathbf{y}'\}} \sum_{i=1}^{m} \beta_i K_{\mathcal{Y}}(\mathbf{y}_i, \mathbf{y}) \,. \tag{13}$$

Upper bound g

$$g(\mathbf{y}', \ell) \geq \max_{\mathbf{y} \in \mathcal{A}^{\ell - p} \times \{\mathbf{y}'\}} \sum_{i=1}^{m} \beta_i K_{\mathcal{Y}}(\mathbf{y}_i, \mathbf{y})$$

Upper bound g

$$g(\mathbf{y}', \ell) \geq \max_{\mathbf{y} \in \mathcal{A}^{\ell-p} \times \{\mathbf{y}'\}} \sum_{i=1}^{m} \beta_i K_{\mathcal{Y}}(\mathbf{y}_i, \mathbf{y})$$



Upper bound g

$$g(\mathbf{y}', \ell) \geq \max_{\mathbf{y} \in \mathcal{A}^{\ell-p} \times \{\mathbf{y}'\}} \sum_{i=1}^{m} \beta_i K_{\mathcal{Y}}(\mathbf{y}_i, \mathbf{y})$$

Position	**1**	**2**	**3**
AA	AA	AA	AA
AB	AB	AB	AB
BA	BA	BA	BA
BB	BB	BB	BB

Upper bound g

$$g(\mathbf{y}', \ell) \geq \max_{\mathbf{y} \in \mathcal{A}^{\ell-p} \times \{\mathbf{y}'\}} \sum_{i=1}^{m} \beta_i K_{\mathcal{Y}}(\mathbf{y}_i, \mathbf{y})$$

Position	**1**	**2**	**3**
AA	0.5	3.2	5
AB	1	3	4.1
BA	2.5	4	3
BB	2	2	4

Upper bound g

$$g(\mathbf{y}', \ell) \geq \max_{\mathbf{y} \in \mathcal{A}^{\ell-p} \times \{\mathbf{y}'\}} \sum_{i=1}^{m} \beta_i K_{\mathcal{Y}}(\mathbf{y}_i, \mathbf{y})$$

Position	**1**	**2**	**3**
AA	0.5	3.2	5
AB	1	3	4.1
BA	2.5	4	3
BB	2	2	4

$\mathbf{y}' = \text{ABB}, \ \ell = 4$

Upper bound g

$$g(\mathbf{y}', \ell) \geq \max_{\mathbf{y} \in \mathcal{A}^{\ell - p} \times \{\mathbf{y}'\}} \sum_{i=1}^{m} \beta_i K_{\mathcal{Y}}(\mathbf{y}_i, \mathbf{y})$$

Position	**1**	**2**	**3**
AA	0.5	3.2	5
AB	1	**3**	4.1
BA	2.5	4	3
BB	2	2	4

$\mathbf{y}' =$ **AB**B, $\ell = 4$

Upper bound g

$$g(\mathbf{y}', \ell) \geq \max_{\mathbf{y} \in \mathcal{A}^{\ell-p} \times \{\mathbf{y}'\}} \sum_{i=1}^{m} \beta_i K_{\mathcal{Y}}(\mathbf{y}_i, \mathbf{y})$$

Position	**1**	**2**	**3**
AA	0.5	3.2	5
AB	1	3	4.1
BA	2.5	4	3
BB	2	2	4

$\mathbf{y}' = \text{A}\mathbf{BB}$, $\ell = 4$

Upper bound g

$$g(\mathbf{y}', \ell) \geq \max_{\mathbf{y} \in \mathcal{A}^{\ell-p} \times \{\mathbf{y}'\}} \sum_{i=1}^{m} \beta_i K_{\mathcal{Y}}(\mathbf{y}_i, \mathbf{y})$$

Position	**1**	**2**	**3**
AA	0.5	3.2	5
AB	1	3	4.1
BA	2.5	4	3
BB	2	2	4

$\mathbf{y}' = \text{ABB}$, $\ell = 4$

$\mathcal{O}(1)$

Lower bound f

$$f(\mathbf{y}', \ell) \leq \min_{\mathbf{y} \in \mathcal{A}^{\ell - p} \times \{\mathbf{y}'\}} K_{\mathcal{Y}}(\mathbf{y}, \mathbf{y})$$

Lower bound f

$$f(\mathbf{y}', \ell) \leq \min_{\mathbf{y} \in \mathcal{A}^{\ell - p} \times \{\mathbf{y}'\}} K_{\mathcal{Y}}(\mathbf{y}, \mathbf{y})$$

$\mathbf{y}' = \mathrm{AB}, \ell = 5$

? ? ? A B

? ? ? A B

Lower bound f

$$f(\mathbf{y}', \ell) \leq \min_{\mathbf{y} \in \mathcal{A}^{\ell-p} \times \{\mathbf{y}'\}} K_{\mathcal{Y}}(\mathbf{y}, \mathbf{y})$$

$\mathbf{y}' = \text{AB}, \ell = 5$

? ? ? **AB**

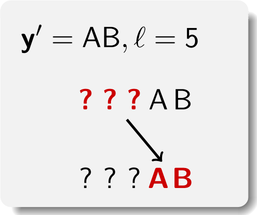
↓

? ? ? **AB**

$$GS(\mathbf{y}', \mathbf{y}', n, \sigma_p, \sigma_c)$$

Lower bound f

$$f(\mathbf{y}', \ell) \leq \min_{\mathbf{y} \in \mathcal{A}^{\ell-p} \times \{\mathbf{y}'\}} K_{\mathcal{Y}}(\mathbf{y}, \mathbf{y})$$



$$GS(\mathbf{y}', \mathbf{y}', n, \sigma_p, \sigma_c) + 2YY'(\mathbf{y}', \ell, n, \sigma_p, \sigma_c)$$

Lower bound f

$$f(\mathbf{y}', \ell) \leq \min_{\mathbf{y} \in \mathcal{A}^{\ell - p} \times \{\mathbf{y}'\}} K_{\mathcal{Y}}(\mathbf{y}, \mathbf{y})$$

$\mathbf{y}' = \text{AB}, \ell = 5$

? ? ? A B

$\downarrow$

? ? ? A B

$$GS(\mathbf{y}', \mathbf{y}', n, \sigma_p, \sigma_c) + 2YY'(\mathbf{y}', \ell, n, \sigma_p, \sigma_c) + YY(\ell - |\mathbf{y}'|, n, \sigma_p, \sigma_c)$$

RESULTS

OCR results

 $\longrightarrow$ declaring

	0/1 risk	
	known $\lvert\mathbf{y}\rvert$	unknown $\lvert\mathbf{y}\rvert$
$\mathbf{y}^{\star}_{2\text{-gram}}$ $\text{Eulerian}_{2\text{-gram}}$		
$\mathbf{y}^{\star}_{3\text{-gram}}$ $\text{Eulerian}_{3\text{-gram}}$		

Table : Empirical results on the optical word recognition task.

OCR results

$\longrightarrow$ declaring

	0/1 risk	
	known $\lvert\mathbf{y}\rvert$	unknown $\lvert\mathbf{y}\rvert$
$\mathbf{y}^{\star}_{2\text{-gram}}$	**18.82** $\pm .72$	
$\text{Eulerian}_{2\text{-gram}}$	20.86 $\pm .67$	
$\mathbf{y}^{\star}_{3\text{-gram}}$	**6.18** ± 1.2	
$\text{Eulerian}_{3\text{-gram}}$	6.95 $\pm .80$	

Table : Empirical results on the optical word recognition task.

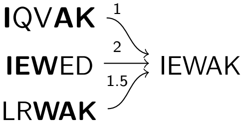
OCR results

 ⟶ declaring

	0/1 risk	
	known $\|\mathbf{y}\|$	unknown $\|\mathbf{y}\|$
$\mathbf{y}^{\star}_{2\text{-gram}}$	**18.82** $\pm .72$	**22.29** $\pm .85$
Eulerian$_{2\text{-gram}}$	20.86 $\pm .67$	64.11 ± 2.0
$\mathbf{y}^{\star}_{3\text{-gram}}$	**6.18** ± 1.2	**9.07** $\pm .65$
Eulerian$_{3\text{-gram}}$	6.95 $\pm .80$	34.76 ± 2.5

Table : Empirical results on the optical word recognition task.

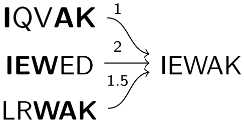
Peptide results



	BPPs		CAMPs	
Method	$\mathbf{y}$	$\|\|\phi_{\mathcal{Y}}(\mathbf{y})\|\|$	$\mathbf{y}$	$\|\|\phi_{\mathcal{Y}}(\mathbf{y})\|\|$
$h_{\sigma_p=\text{Small}}$				
$\widehat{h}_{\sigma_p=\text{Small}}$				
$h_{\sigma_p=\text{Large}}$				
$\widehat{h}_{\sigma_p=\text{Large}}$				

Table : Predicted peptides with highest bioactivity

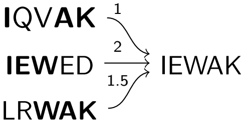
Peptide results



	BPPs		CAMPs	
Method	$\mathbf{y}$	$\|\|\phi_{\mathcal{Y}}(\mathbf{y})\|\|$	$\mathbf{y}$	$\|\|\phi_{\mathcal{Y}}(\mathbf{y})\|\|$
$h_{\sigma_p=\text{Small}}$	IEWAK	3.46	WWKWWKRLRRLFLLV	9.05
$\widehat{h}_{\sigma_p=\text{Small}}$	IEWAK	3.46	WWKWWKRLRRLFLLV	9.05
$h_{\sigma_p=\text{Large}}$				
$\widehat{h}_{\sigma_p=\text{Large}}$				

Table : Predicted peptides with highest bioactivity

Peptide results



	BPPs		CAMPs	
Method	$\mathbf{y}$	$\|\|\phi_{\mathcal{Y}}(\mathbf{y})\|\|$	$\mathbf{y}$	$\|\|\phi_{\mathcal{Y}}(\mathbf{y})\|\|$
$h_{\sigma_p=\text{Small}}$	IEWAK	3.46	WWKWWKRLRRLFLLV	9.05
$\widehat{h}_{\sigma_p=\text{Small}}$	IEWAK	3.46	WWKWWKRLRRLFLLV	9.05
$h_{\sigma_p=\text{Large}}$	**WA**K**WA**	**4.24**	**FK**KI**FK**KI**FK**KI**FK**F	**12.81**
$\widehat{h}_{\sigma_p=\text{Large}}$	VEWAK	3.46	WKKI**FK**KIWKFRV**FK**	9.59

Table : Predicted peptides with highest bioactivity

Summary

- General solution for string prediction
- Search for the exact pre-image instead of an approximation
- Avoid the bias of an unnormalized predictor

Use our code:
https://github.com/a-ro/preimage

References I

Cortes, C., Mohri, M., and Weston, J. (2007).
A general regression framework for learning string-to-string mappings.
In Bakır, G., Hofmann, T., Schölkopf, B., Smola, A. J., Taskar, B., and Vishwanathan, S., editors, *Predicting Structured Data*, chapter 8, pages 143–168. MIT Press, Cambridge, MA.

Giguère, S., Marchand, M., Laviolette, F., Drouin, A., and Corbeil, J. (2013).
Learning a peptide-protein binding affinity predictor with kernel ridge regression.
BMC Bioinformatics, 14.