Supplementary material of " A max-margin training for RNA secondary structure prediction integrated with the thermodynamic model "

Manato Akiyama[1] Kengo Sato[1,*] Yasubumi Sakakibara[1]

S1 Derivation of scoring functions for max-margin training

Let $NN(x,y)$ be a set of substructures into which a secondary structure $y \in \mathcal{Y}(x)$ can be decomposed according to the nearest neighbor model. A fetaure representation $\Phi(x,y)$ can also be decomposed into local feature representations $\phi(l)$ as follows:

$$\Phi(x,y) = \sum_{l \in NN(x,y)} \phi(l).$$

We define a l-closing base-pair as a base-pair that closes a substructure l, that is, the outermost base-pair of l. We denote by $y_{l-\text{closing}} = y_{ij}$ for the l-closing base-pair (i,j).

The loss function of the prediction $\hat{y}$ against the training data y (Eq. (8) in the main paper) can be transformed into the following using binary-valued variables:

$$\begin{aligned}\Delta(y,\hat{y}) =& \delta^{\text{FN}} \times (\text{\# of false negative base-pairs}) + \delta^{\text{FP}} \times (\text{\# of false positive base-pairs}) \\ =& \delta^{\text{FN}} \sum_{i<j} I(y_{ij}=1)I(\hat{y}_{ij}=0) + \delta^{\text{FP}} \sum_{i<j} I(y_{ij}=0)I(\hat{y}_{ij}=1) \\ =& \sum_{i<j} \left\{ \delta^{\text{FN}} y_{ij}(1-\hat{y}_{ij}) + \delta^{\text{FP}}(1-y_{ij})\hat{y}_{ij} \right\}.\end{aligned}$$

Here, $I(y_{ij}=1)I(\hat{y}_{ij}=0)=1$ if $\hat{y}_{ij}$ is a false negative and 0 otherwise, and $I(y_{ij}=0)I(\hat{y}_{ij}=1)=1$ if $\hat{y}_{ij}$ is a false positive and 0 otherwise. We also use the fact that $I(y_{ij}=1)=y_{ij}$ and $I(y_{ij}=0)=1-y_{ij}$.

Therefore, the first term of Eq. (7) in the main paper can be simplified into:

$$\begin{aligned}f(x,\hat{y}) + \Delta(y,\hat{y}) =& \lambda^\top \Phi(x,\hat{y}) + \sum_{i<j} \left\{ \delta^{\text{FN}} y_{ij}(1-\hat{y}_{ij}) + \delta^{\text{FP}}(1-y_{ij})\hat{y}_{ij} \right\} \\ =& \sum_{l \in NN(x,\hat{y})} \lambda^\top \phi(l) + \sum_{i<j} \left\{ \left[-\delta^{\text{FN}} y_{ij} + \delta^{\text{FP}}(1-y_{ij}) \right] \hat{y}_{ij} + \delta^{\text{FN}} y_{ij} \right\} \\ =& \sum_{l \in NN(x,\hat{y})} \lambda^\top \phi(l) + \sum_{i<j \text{ s.t. } \hat{y}_{ij}=1} \left[-\delta^{\text{FN}} y_{ij} + \delta^{\text{FP}}(1-y_{ij}) \right] + Const \\ =& \sum_{l \in NN(x,\hat{y})} \left[\lambda^\top \phi(l) - \delta^{\text{FN}} y_{l-\text{closing}} + \delta^{\text{FP}}(1-y_{l-\text{closing}}) \right] + Const \\ =& \sum_{l \in NN(x,\hat{y})} \left\{ \lambda^\top \phi(l) + \tau_l \right\} + Const,\end{aligned}$$

*to whom correspondence should be addressed
[1]Department of Biosciences and Informatics, Keio University, 3–14–1 Hiyoshi, Kohoku-ku, Yokohama 223–8522, Japan

where

$$\begin{aligned}\tau_l &= -\delta^{\mathrm{FN}} y_{l-\mathrm{closing}} + \delta^{\mathrm{FP}}(1 - y_{l-\mathrm{closing}}) \\ &= \begin{cases} -\delta^{\mathrm{FN}} & (\text{if } y_{l-\mathrm{closing}}{=}1) \\ +\delta^{\mathrm{FP}} & (\text{if } y_{l-\mathrm{closing}}{=}0) \end{cases} \\ Const &= \sum_{i<j} \delta^{\mathrm{FN}} y_{ij}\end{aligned}$$

The last equation indicates that we can calculate the first term of the objective function (7) by adding a penalty term τ_l, which depends on on the correct structure y, to the score of the substructure l.