

FUNCTIONAL SITE PREDICTION BY EXPLOITING CORRELATIONS BETWEEN LABELS OF INTERACTING RESIDUES

Предсказание функциональных участков белков
по зависимостям в типах взаимосвязи
аминокислотных остатков

17

Андрей Белобородов

ЗАЧЕМ?

- Белки важны
- Белков много
- Структура определяет функциональность
...но нетривиально



Задача: предсказать по структуре белка какие аминокислотные остатки будут функциональны в нём.



Что?

Белок состоит из соединённых в цепочку аминокислот. Каждая аминокислота-кирпичик может быть

функциональной – участвовать в реакциях с другими веществами

или нефункциональной – служить каркасом белка.

При построении модели учитываются всевозможные параметры остатков.

- Положение в группе остатков
- Физико-химические свойства
- Электростатические свойства

Подходы

- Шаблоны - поиск 3D паттернов в структуре белка. Паттерны строятся экспертом.
- Остатки – обучение компьютерных моделей распознавать функциональные остатки, на основе тренировочных выборок.
- Предложенный способ относится ко 2ой категории с элементами первой
 - Нигде не учитывается взаимодействие между аминокислотными остатками

СЕТЬ ВЗАИМОДЕЙСТВИЙ ОСТАТКОВ

Residue Interaction Network (RIN)

- Белок, содержащий n аминокислотных остатков

$$A = \{a_1, a_2, \dots, a_n\}$$

- Информация о контакте остатков

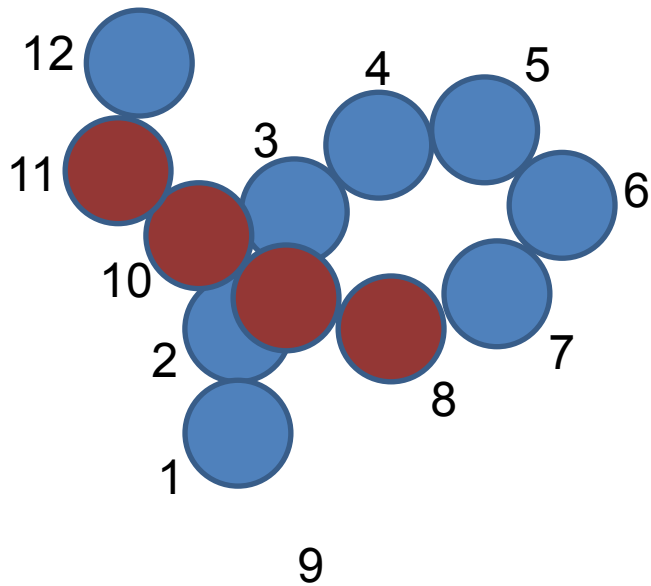
$$e_{ij} = \begin{cases} 1 & \text{if } a_i \text{ and } a_j \text{ are in contact} \\ 0 & \text{otherwise.} \end{cases}$$

- Признак функциональности остатка

$$y_i = \begin{cases} +1 & \text{if } a_i \text{ is a functional residue} \\ -1 & \text{otherwise} \end{cases}$$

- Информация о взаимосвязи (схожести) остатков

$$w_{ij} = \begin{cases} +1 & y_i = y_j \text{ and } e_{ij} = 1 \\ -1 & y_i \neq y_j \text{ and } e_{ij} = 1 \\ 0 & \text{otherwise.} \end{cases}$$

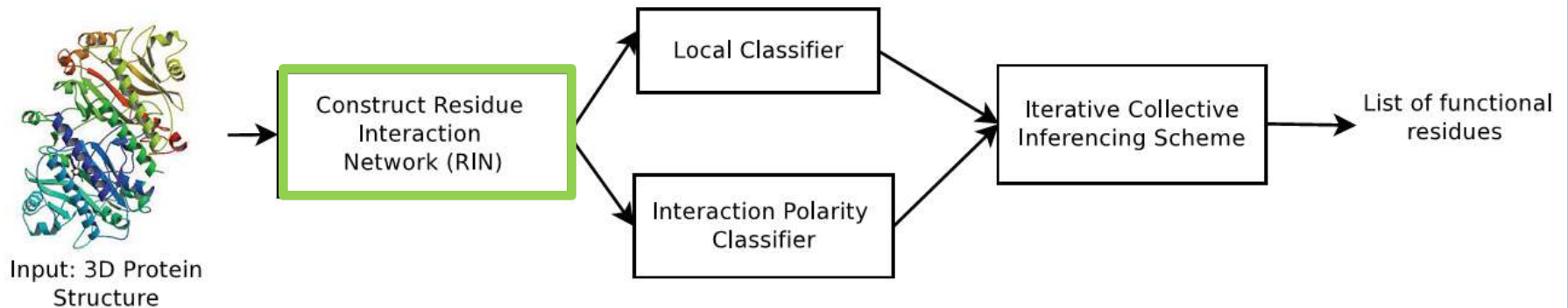


W	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1	0	0	0	0	0	0	0	0	0	0
2	1	1	1	0	0	0	0	0	-1	-1	0	0
3	0	1	1	1	0	0	0	0	-1	-1	0	0
4	0	0	1	1	1	0	0	0	0	0	0	0
5	0	0	0	1	1	1	0	0	0	0	0	0
6	0	0	0	0	1	1	1	0	0	0	0	0
7	0	0	0	0	0	1	1	-1	0	0	0	0
8	0	0	0	0	0	0	-1	1	1	0	0	0
9	0	-1	-1	0	0	0	0	1	1	1	0	0
10	0	-1	-1	0	0	0	0	0	1	1	1	0
11	0	0	0	0	0	0	0	0	0	1	1	-1
12	0	0	0	0	0	0	0	0	0	0	-1	1

Y -1 -1 -1 -1 -1 -1 -1 1 1 1 1 -1

E	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1	0	0	0	0	0	0	0	0	0	0
2	1	1	1	0	0	0	0	0	1	1	0	0
3	0	1	1	1	0	0	0	0	1	1	0	0
4	0	0	1	1	1	0	0	0	0	0	0	0
5	0	0	0	1	1	1	0	0	0	0	0	0
6	0	0	0	0	1	1	1	0	0	0	0	0
7	0	0	0	0	0	1	1	1	0	0	0	0
8	0	0	0	0	0	0	1	1	1	0	0	0
9	0	1	1	0	0	0	0	1	1	1	0	0
10	0	1	1	0	0	0	0	0	1	1	1	0
11	0	0	0	0	0	0	0	0	0	1	1	1
12	0	0	0	0	0	0	0	0	0	0	1	1

ПЛАН ДЕЙСТВИЙ

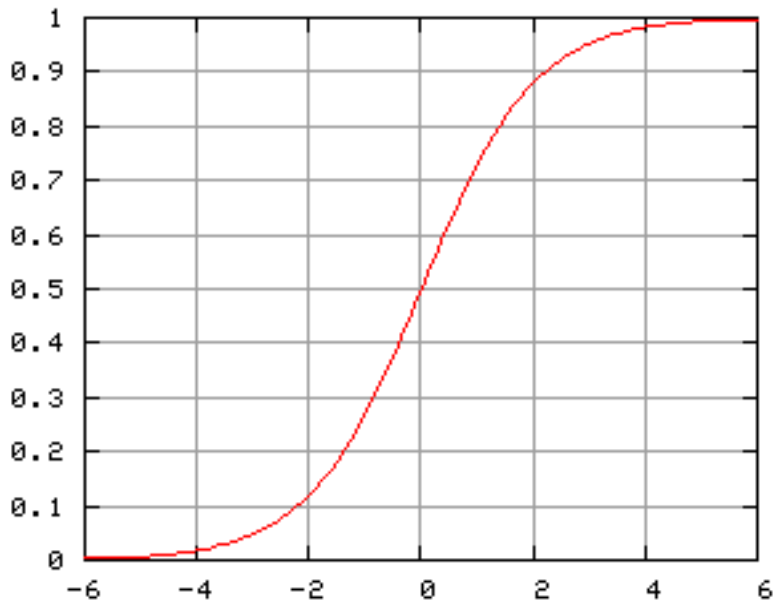


Input: Protein structure \mathcal{S}

Output: Label vector $\tilde{\mathbf{y}}$

- 1: $\mathcal{G} \leftarrow \text{formRIN}(\mathcal{S})$
- 2: $\vartheta \leftarrow \text{applyInteractionPolarityClassifier}(\mathbf{X}_{ij}; \theta)$
- 3: $\varphi \leftarrow \text{applyLocalClassifier}(\mathbf{X}; \rho)$
- 4: $\tilde{\mathbf{y}} \leftarrow \text{applyCollectiveInference}(\varphi, \vartheta, \mathcal{G}, \alpha, \beta, \gamma)$

РЕГУЛЯРИЗОВАННАЯ L1 ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ



$$x = a + x'$$

a – параметр, склоняющий классификатор в ту или иную сторону

Логистическая функция: $f(x) = \frac{1}{1 + e^{-x}}$

ЛОКАЛЬНЫЙ КЛАССИФИКАТОР

Local classifier

- x_i - вектор (набор) известных параметров остатка a_i
- Вероятность того, что остаток окажется функциональным на основе его параметров

$$\varphi_i = \Pr(\hat{y}_i = +1 \mid \mathbf{x}_i; \rho) = \frac{1}{1 + \exp(-\rho^T \mathbf{x}_i)}$$

- Порог после которого считать, что остаток функционален

$$\hat{y}_i = \begin{cases} +1 & \text{if } \varphi_i \geq \beta \\ -1 & \text{otherwise.} \end{cases}$$

КЛАССИФИКАТОР ТИПОВ ВЗАИМОСВЯЗЕЙ

Interaction Polarity Classifier

- Конкатенация параметров остатков

$$\mathbf{x}_{ij} = \begin{cases} \text{cat}(\mathbf{x}_i, \mathbf{x}_j) & \text{if } i < j \text{ and } e_{ij} = 1 \\ \text{cat}(\mathbf{x}_j, \mathbf{x}_i) & \text{if } i > j \text{ and } e_{ij} = 1 \end{cases}$$

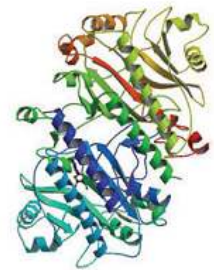
- Вероятность того, что два остатка окажутся одинаковыми на основе их параметров

$$\vartheta_{ij} = \Pr(\hat{w}_{ij} = +1 | \mathbf{x}_{ij}; \theta) = \frac{1}{1 + \exp(-\theta^T \mathbf{x}_{ij})}$$

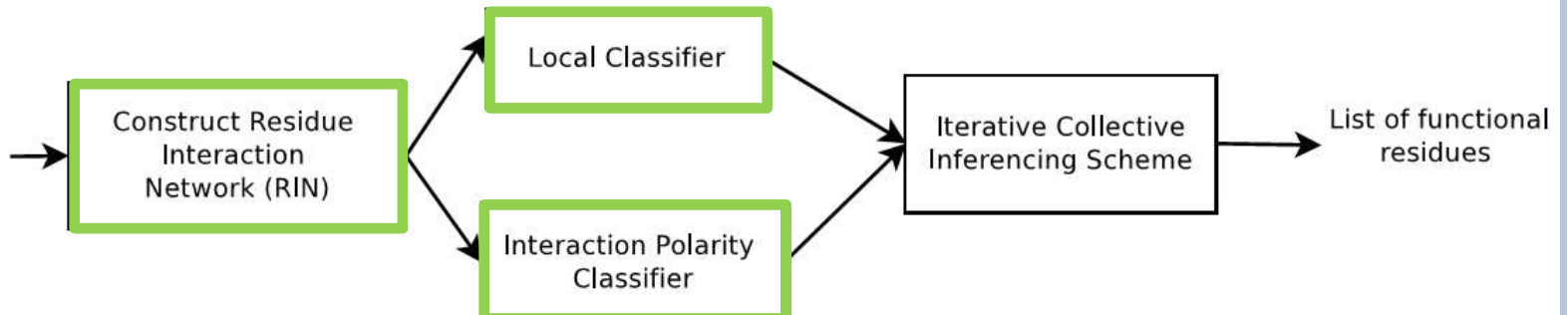
- Порог после которого считать, что связь есть

$$\hat{w}_{ij} = \begin{cases} +1 & \text{if } \vartheta_{ij} \geq \alpha \\ -1 & \text{otherwise.} \end{cases}$$

ПОСЛЕДНИЙ РЫВОК



Input: 3D Protein Structure



ИТЕРАТИВНАЯ НАХОЖДЕНИЕ КОНСЕНСУСА

Iterative collective inference scheme

- Степень подтверждения статуса остатка от соседей

$$q_i = \frac{\sum_{\forall j: e_{ij}=1} \max(0, \hat{y}_j \times \widehat{W}_{ij})}{\sum_j e_{ij}}$$

- Порог, после которого принимается решение переобозначить вектор

$$\tilde{y}_i = \begin{cases} +1 & \text{if } q_i > \max(0.5, \gamma) \\ -1 & \text{if } q_i < 1 - \max(0.5, \gamma) \end{cases}$$

$$w_{ij} = \begin{cases} +1 & y_i = y_j \text{ and } e_{ij} = 1 \\ -1 & y_i \neq y_j \text{ and } e_{ij} = 1 \\ 0 & \text{otherwise.} \end{cases}$$

$$y_i = \begin{cases} +1 & \text{if } a_i \text{ is a functional residue} \\ -1 & \text{otherwise} \end{cases}$$

1: $\widehat{\mathbf{W}} \leftarrow \text{getPolarity}(\vartheta, \alpha)$

2: $\hat{\mathbf{y}} \leftarrow \text{getLabel}(\varphi, \beta)$

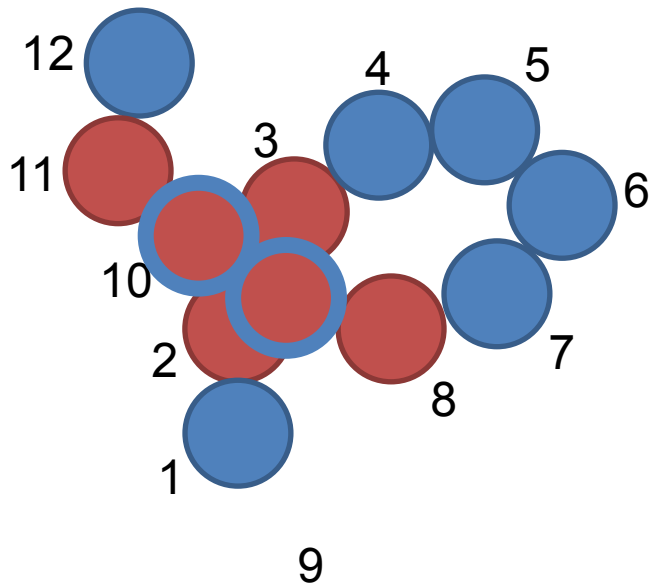
3: $\tilde{\mathbf{y}} \leftarrow \hat{\mathbf{y}}$

4: **repetir**

5: $\mathbf{q} \leftarrow \hat{\mathbf{y}}^T \widehat{\mathbf{W}} \mathbf{o}$

6: $\tilde{\mathbf{y}} \leftarrow \text{updateLabel}(\mathbf{q}, \hat{\mathbf{y}}, \varphi, \gamma)$

7: **hasta** \mathbf{q} stops showing any change



W	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1	0	0	0	0	0	0	0	0	0	0
2	1	1	1	0	0	0	0	0	-1	-1	0	0
3	0	1	1	1	0	0	0	0	-1	-1	0	0
4	0	0	1	1	1	0	0	0	0	0	0	0
5	0	0	0	1	1	1	0	0	0	0	0	0
6	0	0	0	0	1	1	1	0	0	0	0	0
7	0	0	0	0	0	1	1	-1	0	0	0	0
8	0	0	0	0	0	0	-1	1	-1	0	0	0
9	0	-1	-1	0	0	0	0	-1	1	1	0	0
10	0	-1	-1	0	0	0	0	0	1	1	-1	0
11	0	0	0	0	0	0	0	0	0	-1	1	-1
12	0	0	0	0	0	0	0	0	0	0	-1	1

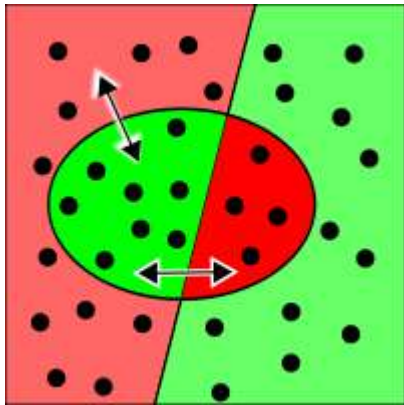
Y -1 -1 1 1 -1 -1 -1 1 -1 1 1 -1

E	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1	0	0	0	0	0	0	0	0	0	0
2	1	1	1	0	0	0	0	0	1	1	0	0
3	0	1	1	1	0	0	0	0	1	1	0	0
4	0	0	1	1	1	0	0	0	0	0	0	0
5	0	0	0	1	1	1	0	0	0	0	0	0
6	0	0	0	0	1	1	1	0	0	0	0	0
7	0	0	0	0	0	1	1	1	0	0	0	0
8	0	0	0	0	0	0	1	1	1	0	0	0
9	0	1	1	0	0	0	0	1	1	1	0	0
10	0	1	1	0	0	0	0	0	1	1	1	0
11	0	0	0	0	0	0	0	0	0	1	1	1
12	0	0	0	0	0	0	0	0	0	0	1	1

Q 0 1 1 0 0 0 0 0 1.75 .75 1 0

Q 0 1 1 0 0 0 0 0 1 1 1 1 0

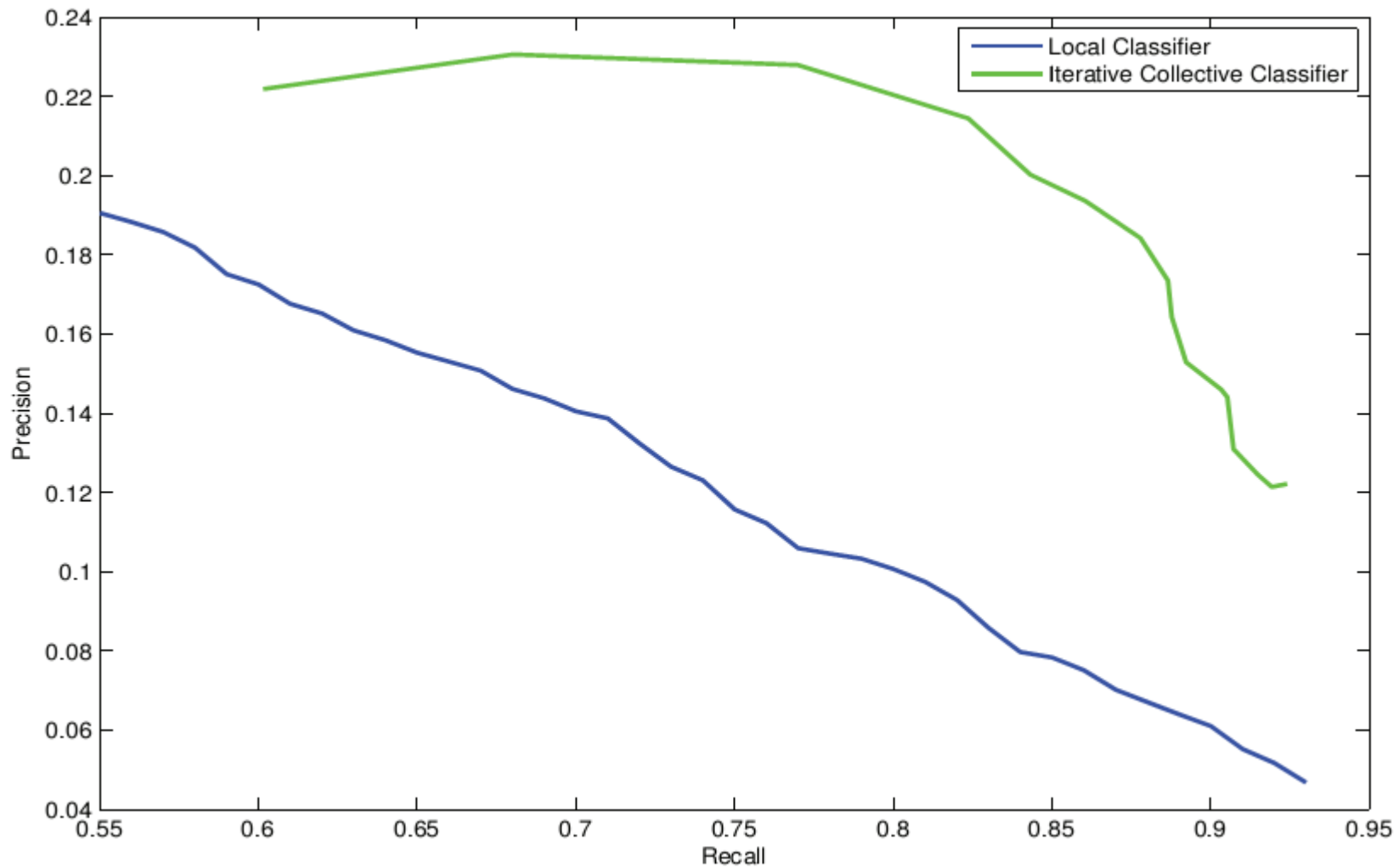
PR, HO HE PEKJAMA



$$\text{Precision} = \frac{TP}{TP+FP} = \frac{\text{Green shape}}{\text{Green shape} + \text{Red shape}}$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{\text{Green shape}}{\text{Green shape} + \text{Green shape}}$$

РЕЗУЛЬТАТЫ



СРАВНЕНИЕ С ДРУГИМИ МЕТОДАМИ

Method	Precision ₆₉ (%)	Recall ₁₈ (%)
Collective Inference	23.06	87.78
Discern [30]	18	69
CRF [30]	18	69
Local classifier	14.38	58

СПАСИБО ЗА ВНИМАНИЕ

Подходящий момент, чтобы задать вопрос