

Вероятностный подход для задачи предсказания биологической активности ядерных рецепторов

Володин Сергей Евгеньевич

Московский физико-технический институт

Курс: Численные методы обучения по прецедентам
(практика, В. В. Стрижов)/Группа 374, осень 2016

Цель исследования

Решается задача предсказания биологической активности молекул протеинов (лиганд) с рецепторами. По признакам лиганда необходимо оценить вероятность связывания этой молекулы с одним или несколькими клеточными рецепторами и построить бинарный классификатор

Проблема

Экспертные знания в области биохимии и фармакологии дают основания пред- полагать, что факты связывания одних и тех же молекул с различными рецепторами не независимы.

Задача

Необходимо построить модель, учитывающую схожесть рецепторов. Модель должна предсказывать вероятность взаимодействия.

Картинка с молекулами (жду)

- 1 Olexandr Isayev Sherif Farag Stephen J. Capuzzi, Regina Politi and Alexander Tropsha. Qsar modeling of tox21 challenge stress response and nuclear receptor signaling toxicity assays.
- 2 Geoff Holmes Eibe Frank Jesse Read, Bernhard Pfahringer. Classifier chains for multi-label classification.
- 3 Eyke H.0 Krzysztof Dembczynski, Weiwei Cheng. Bayes optimal multilabel classification via probabilistic classifier chains. 2010.

Постановка задачи

Задана выборка $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\} = \mathcal{L} \sqcup \mathcal{T}$. $\mathbf{x}_i \in \mathbb{R}^n$. $\mathbf{y}_i \in \{0, 1, \square\}^I$,
 \square — пропуск в данных.

\mathbf{X}, \mathbf{Y} — случайные величины, между классами есть зависимости.

Модель классификации: функция $f: \mathbf{W} \times \mathbf{X} \times \mathbf{Y} \rightarrow [0, 1]$,

$$f(\mathbf{w}, \mathbf{x}, \mathbf{y}) = P(\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x}; \mathbf{w})$$

Функция потерь — логарифм правдоподобия

$$Q(f | \mathbf{w}, \mathcal{Z}) = - \sum_{(\mathbf{x}, \mathbf{y}) \in \mathcal{Z}} \log f(\mathbf{w}, \mathbf{x}, \mathbf{y}) P(\mathbf{X} = \mathbf{x})$$

Требуется минимизировать Q :

$$\mathbf{w}^* = \arg \min_{\mathbf{w} \in \mathbf{W}} Q(f | \mathbf{w}, \mathcal{L})$$

Для выбора конкретной модели используется AUC.

Распишем вероятность, ...

Слайд появится после T , пока тут пусто

Цель эксперимента

Сравнение различных моделей по критерию AUC для различных классов.

Показать характерные графики ROC-кривой для разных моделей и классов (несколько слайдов).

Показать графики в зависимости от гиперпараметров

Вычислительный эксперимент

Результаты эксперимента

Таблица модели × рецепторы, сравнивающая их по AUC.

- 1 Предложена модель для предсказания взаимодействия, учитывающая зависимости между классами
- 2 Проведено сравнение модели с другими по критерию AUC
- 3 PCC лучше BR для классов a, b, c по критерию AUC.