

Формулировка и решение задачи оптимизации, сочетающей классификацию и регрессию, для оценки энергии связывания белка и маленьких молекул*

Грачева А. С., Соавтор И. О., Фамилия И. О.

gracheva.as@phystech.edu

¹ФИВТ МФТИ

При разработке лекарства возникает задача поиска маленьких молекул - лигандов, наиболее сильно взаимодействующих с исследуемым белком, а значит являющихся основными кандидатами в лекарства. Можно генерировать несколько возможных положений лиганда и классифицировать их как нативные и не нативные, но качество предсказания может быть повышено, если использовать экспериментальные данные о свободной энергии связывания молекул и решать одновременно задачи регрессии и классификации. В статье будут рассмотрены эксперименты с алгоритмом, использующим эту идею.

1 Введение

Предсказание наиболее выгодной ориентации и положения молекул по отношению друг к другу для образования устойчивого комплекса из белка и лиганда, или молекулярный докинг - задача, важная для ускорения процесса разработки новых лекарств. Есть два метода её решения: pose prediction - среди нескольких сгенерированных положений лиганда в белке определить наиболее близкое к реальному и scoring - предсказать аффинность (свободную энергию связывания) для комплексов различных белков с лигандами. При этом положение с наименьшей энергией связывания будет соответствовать нативной конформации. Первая задача решена в работе [1] с помощью оптимизации скоринговой функции, учитывающей всевозможные комбинации различных пар атомов и расстояния между ними. Раскладывая эту функцию по базису, авторы представляют её как вектор структурных коэффициентов и сводят задачу к модифицированной SVM-классификации.

Наше предположение заключается в том, что если использовать экспериментальные данные об аффинностях, то можно улучшить качество классификации. В эксперименте, описанном в данной статье, мы проверим эту гипотезу, а также постараемся решать оптимизационную задачу максимально эффективно вычислительно, чтобы иметь возможность использовать как можно больше доступных экспериментальных данных.

2 Название раздела

Данный документ демонстрирует оформление статьи, подаваемой в электронную систему подачи статей <http://jmla.org/papers> для публикации в журнале «Машинное обучение и анализ данных». Более подробные инструкции по стилевому файлу `jmla.sty` и использованию издательской системы \LaTeX 2_ε находятся в документе `authors-guide.pdf`. Работу над статьёй удобно начинать с правки \TeX -файла данного документа.

*Работа выполнена при финансовой поддержке РФФИ, проект №00-00-00000. Научный руководитель: Стрижов В. В. Задачу поставил: Эксперт И. О. Консультант: Консультант И. О.

2.1 Название параграфа.

Нет ограничений на количество разделов и параграфов в статье. Разделы и параграфы не нумеруются.

2.2 Теоретическую часть работы

желательно структурировать с помощью окружений Def, Axiom, Hypothesis, Problem, Lemma, Theorem, Corollary, State, Example, Remark.

Определение **Definition 1.** Математический текст хорошо структурирован, если в нём выделены определения, теоремы, утверждения, примеры, и т. д., а неформальные рассуждения (мотивации, интерпретации) вынесены в отдельные параграфы.

Утверждение **Statement 1.** Мотивации и интерпретации наиболее важны для понимания сути работы.

Теорема **Theorem 1.** Не менее 90% коллег, заинтересовавшихся Вашей статьёй, прочитают в ней не более 10% текста.

Доказательство. Причём это будут именно те разделы, которые не содержат формул. ■

Замечание **Remark 1.** Выше показано применение окружений Def, Theorem, State, Remark, Proof.

3 Некоторые формулы

Образец формулы: $f(x_i, \alpha^\gamma)$.

Образец выключной формулы без номера:

$$y(x, \alpha) = \begin{cases} -1, & \text{если } f(x, \alpha) < 0; \\ +1, & \text{если } f(x, \alpha) \geq 0. \end{cases}$$

Образец выключной формулы с номером:

$$y(x, \alpha) = \begin{cases} -1, & \text{если } f(x, \alpha) < 0; \\ +1, & \text{если } f(x, \alpha) \geq 0. \end{cases} \quad (1)$$

Образец выключной формулы, разбитой на две строки с помощью окружения align:

$$R'_N(F) = \frac{1}{N} \sum_{i=1}^N \left(P(+1 | x_i) C(+1, F(x_i)) + \right. \\ \left. + P(-1 | x_i) C(-1, F(x_i)) \right). \quad (2)$$

Образцы ссылок: формулы (1) и (2).

4 Пример иллюстрации

Рисунки вставляются командой `\includegraphics`, желательно с выравниванием по ширине колонки: `[width=\linewidth]`.

Практически все популярные пакеты рисуют графики с подписями, которые трудно читать на бумаге и на слайдах из-за малого размера шрифта. Шрифт на графиках (подписи осей и цифры на осях) должны быть такого же размера, что и основной текст.

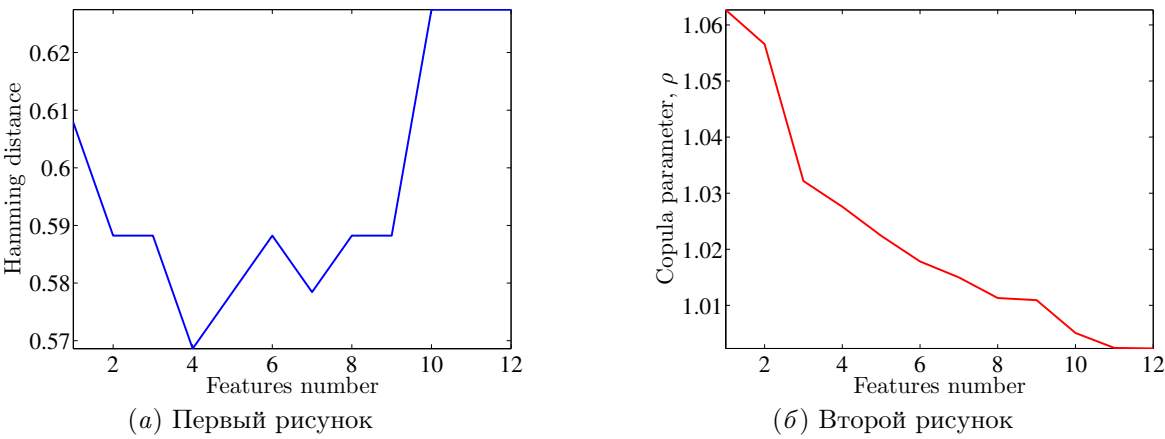


Рис. Figure 1 Подпись должна размещаться под рисунком.

Таблица Table 1 Подпись размещается над таблицей.

Задача	CCEL	boosting
Cancer	3.46 ± 0.37 (3.16)	4.14 ± 1.48
German	25.78 ± 0.65 (1.74)	29.48 ± 0.93
Hepatitis	18.38 ± 1.43 (2.87)	19.90 ± 1.80

При значительном количестве рисунков рекомендуется группировать их в одном окружении {figure}, как это сделано на рис. 1.

5 Пример таблицы

Подпись делается *над таблицей*, см. таблицу 1.

6 Заключение

Желательно, чтобы этот раздел был, причём он не должен дословно повторять аннотацию. Обычно здесь отмечают, каких результатов удалось добиться, какие проблемы остались открытыми.

ЛитератураReferences

ЛитератураReferences

[1] Maria Kadukova, Sergei Grudinin Maria Kadukova, Sergei Grudinin Convex-PL: a novel knowledge-based potential for protein-ligand interactions deduced from structural databases using convex optimizationConvex-PL: a novel knowledge-based potential for protein-ligand interactions deduced from structural databases using convex optimization // Journal of ComputerAided Molecular Design, Springer Verlag Journal of ComputerAided Molecular Design, Springer Verlag, 2017, 31 (10), pp.943-958.

