

Выпуклая релаксация некоторых задач квадратичного дискретного программирования

А. М. Катруца¹

¹Московский физико-технический институт (государственный университет)

Работа посвящена исследованию методов выпуклых релаксаций в задачах дискретного квадратичного программирования. Такие задачи возникают в различных областях науки и техники. Данная работа рассматривает применение выпуклых релаксаций в задаче предсказания боковой цепи белка, которая возникает в структурной биологии, и в задаче выбора признаков. В общем случае эти задачи являются NP-полными, однако выпуклые релаксации позволяют быстро получить приближённое, а иногда и точное, решение.

Задача выбора признаков в форме задачи квадратичного программирования поставлена в [1], где предлагается алгоритм отбора признаков без учёта точности построенной затем модели, а только на основании корреляции между признаками и целевым вектором (линейный член) и совместной полезности каждой пары признаков. Для оценки этой полезности используется совместная информация и корреляция Пирсона. При решении задачи квадратичного программирования используется PCA для понижения размерности пространства.

Задача предсказания боковой цепи белка возникает при решении проблемы предсказания третичной структуры белка. Ранее для решения этой задачи использовались следующие методы: генетический алгоритм [2], имитация отжига [3], методы из теории графов [4] и другие. Ключевыми элементами при построении алгоритма решения этой задачи являются функция энергии, которая оптимизируется, и метод поиска оптимальной конфигурации атомов, минимизирующий выбранную функцию энергии. Из-за физических требований к функции энергии квадратичная форма в оптимизационной задаче не является положительно определённой, поэтому необходимо рассмотреть возможные выпуклые релаксации.

В данной работе рассматриваются следующие способы аппроксимации исходной дискретной невыпуклой задачи квадратичного программирования: полуопределённая релаксация с различными ограничениями на неизвестный вектор (нормы l_1 , l_2 и l_∞), сдвиг спектра исходной матрицы квадратичной формы с аналогичными ограничениями на неизвестный вектор. Проведено сравнение точности рассматриваемых методов с предложенными ранее как в задаче выбора признаков, так и в задаче прогнозирования боковой цепи белка. Сделан вывод о целесообразности использования выпуклых релаксаций в задачах подобного типа.

Литература

- [1] Rodriguez-Lujan I. et al. Quadratic programming feature selection //The Journal of Machine Learning Research. – 2010. – Т. 11. – С. 1491-1516.
- [2] Pedersen J. T., Moult J. Genetic algorithms for protein structure prediction //Current Opinion in Structural Biology. – 1996. – Т. 6. – №. 2. – С. 227-231.
- [3] Hwang J. K., Liao W. F. Side-chain prediction by neural networks and simulated annealing optimization //Protein engineering. – 1995. – Т. 8. – №. 4. – С. 363-370.
- [4] Krivov G. G., Shapovalov M. V., Dunbrack R. L. Improved prediction of protein side-chain conformations with SCWRL4 //Proteins: Structure, Function, and Bioinformatics. – 2009. – Т. 77. – №. 4. – С. 778-795.