

ЭФФЕКТИВНАЯ ФОРМУЛИРОВКА ТЕРМОДИНАМИЧЕСКИХ ОГРАНИЧЕНИЙ В АНАЛИЗЕ БАЛАНСА ПОТОКОВ В ВИДЕ ЗАДАЧИ ЦЕЛОЧИСЛЕННОГО ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

Сафонов И.А. (ИТМО)

Научный руководитель – кандидат технических наук, доцент Сергушичев А.А. (ИТМО)

Введение. Метод баланса потоков [1] (англ. *Flux Balance Analysis*, сокр. *FBA*) является стандартным инструментом системной биологии для исследования стационарных режимов метаболизма и широко применяется, например, в задачах метаболической инженерии и системной медицины [2]. При этом классическая постановка *FBA* допускает решения с термодинамически невозможными циклами, которые способны исказить результаты. Добавление ациклических ограничений позволяет исключить циклы, однако делает задачу *NP*-трудной [3]. Для решения задач метода баланса потоков с ациклическими ограничениями используется целочисленное линейное программирование (англ. *Mixed-Integer Linear Programming*, сокр. *MILP*). Метод *Fast-SNP* [4] позволил улучшить стандартную *MILP*-формулировку [5] ациклических ограничений, ускорив оптимизацию на порядки за счет использования разреженных ограничений и сокращения числа целочисленных переменных. В настоящей работе мы предлагаем альтернативную *MILP*-формулировку ациклических ограничений на основе метаболических потенциалов, использующую эти оптимизации. Мы демонстрируем улучшение эффективности по сравнению с методом *Fast-SNP* на примере задачи вариации потока по реакциям и задачи поиска заблокированных реакций. Дополнительным преимуществом является то, что предлагаемая формулировка строится на порядки быстрее *MILP*-формулировки в методе *Fast-SNP* для моделей масштаба млекопитающих с десятками тысяч реакций и метаболитов, что позволяет эффективно решать задачу поиска заблокированных реакций для таких моделей.

Основная часть. Основной способ представления ациклических ограничений в виде *MILP*-формулировки, который используется в популярных пакетах *COBRA Toolbox* [6] и *COBRAPy* [7], основан на построении базиса пространства циклов. С помощью базиса пространства циклов ограничивается вектор энергий реакций, который используется в *MILP*-формулировке. Метод *Fast-SNP* [4] улучшает *MILP*-формулировку по сравнению со стандартным методом за счет того, что строит разреженную матрицу базиса пространства циклов с помощью алгоритма *Fast-SNP*. Также он учитывает возможные направления реакций в модели, что позволяет сократить число затронутых циклами реакций, и в результате сократить число целочисленных переменных в *MILP*-формулировке.

Мы предлагаем вместо представления циклов с помощью базиса ввести непрерывные переменные потенциалов для метаболитов, с помощью которых можно выразить вектор энергий реакций в пространстве метаболитов. Это позволяет строить ограничения напрямую по исходной разреженной стехиометрической матрице модели, избегая промежуточного построения базиса пространства циклов. Для сопоставимого сокращения числа целочисленных переменных мы вводим предварительный этап выделения подмножества реакций, которые могут участвовать в циклах. Будем называть такие реакции циклическими. Их можно находить с помощью решения серии задач линейного программирования (англ. *Linear Programming*, сокр. *LP*). Далее целочисленные ограничения накладываются только на циклические реакции, за счет чего циклы исключаются только там, где они допускаются стандартной формулировкой. В результате целочисленные переменные создаются ровно для тех же реакций, что и в методе *Fast-SNP*.

Важным результатом настоящей работы является методология сравнения эффективности предлагаемой *MILP*-формулировки с *MILP*-формулировкой в методе *Fast-SNP*. Для сравнения использовались модели из базы метаболических моделей *BiGG* [8]. В качестве *LP*- и *MILP*-решателей использовалась библиотека *IBM ILOG CPLEX* версии 22.1.1.

Было проведено стандартное сравнение эффективности на задаче вариации потока по реакциям с ациклическими ограничениями [5] (англ. *loopless Flux Variability Analysis*, сокр. *ll-FVA*). Предлагаемая формулировка метаболических потенциалов имеет такую же эффективность на моделях малого и среднего размера и в среднем на 30% эффективнее на модели клеток человека *RECON1*. Однако на моделях еще большего размера оба метода работают очень долго, что затрудняет их сравнение.

Для сравнения эффективности на моделях большого размера была использована задача поиска заблокированных реакций с учетом ациклических ограничений. На практике эта задача может использоваться для оценки качества метаболической модели. Для этой задачи мы предложили быстрый эвристический алгоритм на основе решения ряда *MILP*-задач. При использовании предлагаемой *MILP*-формулировки ациклических ограничений, этот алгоритм работает в два и более раз быстрее *MILP*-формулировки из метода *Fast-SNP* на крупнейших метаболических моделях *Mouse-GEM* и *Human-GEM*, содержащих более десяти тысяч реакций и метаболитов. Это достигается за счет более эффективного построения предлагаемой *MILP*-формулировки по сравнению с алгоритмом *Fast-SNP*, а также за счет более эффективной *MILP*-оптимизации в ходе алгоритма поиска заблокированных реакций.

Выводы. В настоящей работе мы предложили *MILP*-формулировку ациклических ограничений на основе метаболических потенциалов и продемонстрировали, что она имеет такую же или лучшую эффективность по сравнению с формулировкой из метода *Fast-SNP* на ряде задач метода баланса потоков с ациклическими ограничениями. Таким образом, эта формулировка может быть использована вместо метода *Fast-SNP* в роли *MILP*-формулировки по умолчанию для задач метода баланса потоков. В будущем планируется интеграция формулировки в пакет для метаболического моделирования *COBRAPy*.

Список использованных источников:

1. Orth J. D., Thiele I., Palsson B. Ø. What is flux balance analysis? // *Nat Biotechnol.* 2010. Т. 28. № 3. С. 245–248. doi: 10.1038/nbt.1614.
2. Judge A., Dodd M. S. Metabolism. // *Essays in Biochemistry.* 2020. Т. 64. № 4. С. 607–647. doi: 10.1042/EBC20190041.
3. Müller A. C., Bockmayr A. Fast thermodynamically constrained flux variability analysis. // *Bioinformatics.* 2013. Т. 29. № 7. С. 903–909. doi: 10.1093/bioinformatics/btt059.
4. Saa P. A., Nielsen L. K. Fast-SNP: a fast matrix pre-processing algorithm for efficient loopless flux optimization of metabolic models. // *Bioinformatics.* 2016. Т. 32. № 24. С. 3807–3814. doi: 10.1093/bioinformatics/btw555.
5. Schellenberger J., Lewis N. E., Palsson B. Ø. Elimination of Thermodynamically Infeasible Loops in Steady-State Metabolic Models. // *Biophysical Journal.* 2011. Т. 100. № 3. С. 544–553. doi: 10.1016/j.bpj.2010.12.3707.
6. Heirendt L. и др. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. // *Nat Protoc.* 2019. Т. 14. № 3. С. 639–702. doi: 10.1038/s41596-018-0098-2.
7. Ebrahim A. и др. COBRAPy: COntstraints-Based Reconstruction and Analysis for Python. // *BMC Syst Biol.* 2013. Т. 7. № 1. С. 74. doi: 10.1186/1752-0509-7-74.
8. Schellenberger J. и др. BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. // *BMC Bioinformatics.* 2010. Т. 11. № 1. С. 213. doi: 10.1186/1471-2105-11-213.