

عنوان مقاله: (Multi-Feasibility Variable Selection) انتخاب متغیر چند-شدنی

علی فتحی^۱، محمد رشید^۱، شایان رنجبر زاده^۱ و مجتبی تفاق^{۱*}

^۱ آزمایشگاه بهینه‌سازی و کاربردها، دانشکده علوم ریاضی، دانشگاه صنعتی شریف، تهران، ایران

mtefagh@sharif.edu

Abstract

Background: In a large number of application areas, it is desirable to obtain sparse solutions. In the case of nonconvex problems, which appear in various circumstances, there is no certain tractable method to find an answer with the minimum number of nonzero elements, where it is usually approximated by a convex problem. In computational biology, one of the well-known problems is the *flux variability analysis*, FBA, in which we find a set of reaction fluxes in a cell consistent with constraint vectors with maximum cell growth.

Aim of the study: In this paper, first, we address five variations of the FBA problem designed in the !Optimizer 2021 competition to solve a more general nonconvex variation of the FBA by having many constraints and using sparse methods. In the first two problems, we seek to find the sparsest vectors, and in the next three problems, our goal is to find the matrix V satisfying some constraints with jointly sparse columns. Then, we provide methods and algorithms to overcome these problems, released by the Panda team as the winner solution in this competition.

Materials and Methods: The main algorithm is weighted ℓ_1 -norm minimization, in which we solve a weighted ℓ_1 -norm problem in each iteration, but our weight updating method is different from conventional methods. Using randomness is a new technique in the process of updating weights. We also extend this idea to find jointly sparse matrices. All materials can be found in the Codes and data availability section.

Results: Although the data size for this competition is significantly huge, our linear algorithm spends only a short time finding the near-optimal solution. A benchmark of our main algorithms, which outperforms other methods in speed and accuracy, is provided in the Results section of this paper.

Conclusions: By providing this fast algorithm for the multi-column variations of the FBA, it is now possible to include many different constraint data to solve an FBA problem simultaneously, which helps researchers accurately reconstruct new organisms' metabolic networks in a polynomial time.

Key words: Metabolic Networks Optimization, Flux Balance Analysis, Sparse Optimization, Linear Optimization, Separation Method.

چکیده

پیش‌زمینه: در بسیاری از کاربردها، مطلوب، پیدا کردن جوابی تنک برای مسئله مدنظر است. در حالت مسئله‌های غیرمحدب، که در بسیاری از حوزه‌ها ظاهر می‌شوند، راه‌حل کارای مشخصی برای پیدا کردن جوابی با کمینه عناصر ناصفر در دست نیست و برای اکثر آنها به سراغ تقریبی محدب از مسئله می‌رویم. در زیست‌شناسی محاسباتی، یکی از مسائل شناخته شده، آنالیز تغییر شار، FBA، است که در آن به دنبال پیدا کردن شار واکنش‌های داخل سلول هستیم به طوری که با قیدها سازگار بوده و رشد سلول را بیشینه کند.

هدف تحقیق: در این گزارش، به پنج نمونه از مسئله FBA، طراحی شده در مسابقه !Optimizer 2021 به منظور حل حالتی کلی و غیرمحدب از FBA اشاره می‌کنیم که طی آن قرار است به کمک روش‌های تنک، تعداد زیادی قید در بر گرفته شوند. در دو مسئله‌ی اول به دنبال پیدا کردن تنک‌ترین بردارهای شار، و در سه مسئله‌ی بعدی به دنبال پیدا کردن ماتریس شارهایی هستیم که علاوه بر ارضای قیود، دارای ستون‌هایی مشترک-تنک باشد. در ادامه، روش‌ها و الگوریتم‌هایی برای غلبه بر این مسائل ارائه می‌دهیم که توسط تیم برنده مسابقه، تیم پاندا، ارائه شده‌اند.

مواد و روش‌ها: الگوریتم اصلی استفاده شده، کمینه‌سازی نرم ℓ_1 وزن‌دار است که طی آن در هر دور از الگوریتم، مسئله‌ای نرم ℓ_1 وزن‌دار را حل کرده و وزن‌ها را برای استفاده در دور بعدی با روشی متفاوت با روش‌های معمول به‌هنگام می‌کنیم. استفاده از تصادفی‌سازی نیز یک تکنیک جدید در فرآیند به‌هنگام سازی وزن‌ها است. این ایده‌ها برای پیدا کردن ماتریس مشترک-تنک نیز به کار برده شده و به همراه سایر راه‌حل‌ها در بخش کدها و داده‌های این مقاله آورده شده‌اند.

نتایج: با وجود ابعاد بسیار بالای داده‌های مسابقه، الگوریتم‌های خطی ما زمان بسیار کمی را برای پیدا کردن جواب نزدیک به بهینه نیاز دارند. بررسی عملکرد الگوریتم‌ها، که در مقایسه با سایر روش‌ها سرعت و دقت بسیار بیشتری به نمایش گذاشته، در بخش نتایج در این مقاله آورده شده‌اند.

نتیجه‌گیری: با ارائه‌ی الگوریتمی سریع برای FBA چند ستونی، اکنون این امکان فراهم است تا در حل FBA قیود متنوعی را بتوان همزمان به کار برد، که به پژوهشگران این امکان را می‌دهد تا شبکه متابولیکی موجودات مختلف را در زمانی چندجمله‌ای بازسازی کنند.

کلید واژه‌ها: بهینه‌سازی شبکه‌های متابولیکی، آنالیز توازن شار، بهینه‌سازی تنک، بهینه‌سازی خطی، روش جداسازی.