

The Predictor Set Inference in Gene Regulatory Network Using Gravitational Search Algorithm and Mean Conditional Entropy Fitness Measure

M. Jafari¹, B. Ghavami^{2*}, V. Sattari Naeini²

¹MSc, Computer Engineering Department, Faculty of Engineering, Shahid Bahonar University of Kerman, Kerman, Iran

²Assistant Professor, Computer Engineering Department, Faculty of Engineering, Shahid Bahonar University of Kerman, Kerman, Iran

Receipt in the online submission system: 16 September 2016, received in revised form: 18 May 2017, accepted: 18 May 2017

Abstract

The inference of Gene Regulatory Network (GRN) using gene expression data is significantly important in order to understand gene dependencies, regulatory functions among genes, biological processes, way of process occurrence and avoiding some unplanned processes (disease). The accurate inference of GRN needs the accurate inference of predictor set. Generally, the main limitations of the predictor set inference are the small number of samples, the large number of genes and also the possibility influence of noise in gene expression data. Hence, providing efficient methods to infer predictor set with high reliability is a serious need. In this paper, an efficient method is proposed to infer predictor set using Gravitational Search Algorithm (GSA). A GSA is used for each target gene to infer the predictor subset of the gene. In a population, a mass represents a predictor subset of the associated gene. The initial population per target gene is generated by Pearson Correlation Coefficient (PCC). In order to guide the GSA, Mean Conditional Entropy (MCE) is used as the assessment criterion. Experimental results show that the proposed method has a good ability to infer the predictor set with high reliability. In addition, we also compared the proposed algorithm with a recent similar method based on genetic algorithm. Comparison results reveal the advantage of the proposed algorithm on biological datasets with small data volumes and large network scales.

Key words: *Gene Regulatory Network (GRN); Genomics; Mean Conditional Entropy (MCE); Gravitational Search Algorithm (GSA); Pearson Correlation Coefficient*

*Corresponding author

Address: Computer Engineering Department, Faculty of Engineering, Shahid Bahonar University of Kerman, Kerman, Iran

Tel: +98-34-32534531

Fax: +98-34-32534531

E-mail: ghavami@uk.ac.ir

استنتاج مجموعه پیش‌بینی کننده در شبکه تنظیم کننده ژن مبتنی بر الگوریتم جستجوی گرانشی و تابع برازندگی میانگین آنتروپی شرطی

مینا جعفری^۱، بهنام قوامی^{۲*}، وحید ستاری نایینی^۲

^۱ کارشناسی ارشد هوش مصنوعی، گروه مهندسی کامپیوتر، دانشکده فنی مهندسی، دانشگاه شهید باهنر، کرمان
^۲ استادیار، گروه مهندسی کامپیوتر، دانشکده فنی مهندسی، دانشگاه شهید باهنر، کرمان

تاریخ ثبت در سامانه: ۱۳۹۵/۶/۲۶، بازنگری: ۱۳۹۶/۲/۲۸، پذیرش قطعی: ۱۳۹۶/۲/۲۸

چکیده

استنتاج شبکه تنظیم کننده ژن (GRN) با استفاده از داده‌های بیان ژن، برای درک وابستگی و نحوه تنظیم ژن‌ها، درک فرآیندهای زیست‌شناسی، نحوه رخداد فرآیندها و همچنین جلوگیری از وقوع برخی فرآیندهای ناخواسته (بیماری)، حائز اهمیت است. ساخت صحیح GRN، نیازمند استنتاج صحیح مجموعه پیش‌بینی کننده است. به طور کلی، مهم‌ترین محدودیت برای استنتاج صحیح مجموعه پیش‌بینی کننده، حجم عظیم ژن‌ها، کم بودن تعداد نمونه‌ها و امکان نفوذ نویز در داده‌های بیان ژن است؛ بنابراین، ارائه روش‌هایی کارا برای استنتاج پیش‌بینی کننده‌ها با قابلیت اطمینان بالا، یک نیاز جدی است. در این مقاله، با استفاده از الگوریتم جستجوی گرانشی (GSA)، یک روش کارا برای استنتاج مجموعه پیش‌بینی کننده ارائه شده است. به‌ازای هر ژن هدف، یک الگوریتم GSA برای استنتاج زیرمجموعه پیش‌بینی کننده آن ژن استفاده شده است. در هر جمعیت، یک جرم نشان دهنده زیرمجموعه پیش‌بینی کننده مرتبط با آن ژن هدف است. جمعیت اولیه به‌ازای هر ژن هدف، براساس ضریب همبستگی پیرسون تولید می‌شود. برای هدایت الگوریتم GSA، از معیار ارزیابی میانگین آنتروپی شرطی (MCE) استفاده شده است. نتایج تجربی حاصل از اعمال این روش روی داده‌های زیست‌شناسی نشان می‌دهد که، روش پیشنهادی دقت بالایی برای استنتاج مجموعه پیش‌بینی کننده دارد. به‌علاوه، نتایج روی داده‌های زیست‌شناسی با مقیاس کوچک و بزرگ نشان می‌دهند که، میزان دقت روش پیشنهادی برای استنتاج GRN بیشتر از روش‌های مشابه است.

کلیدواژه‌ها: شبکه تنظیم کننده ژن (GRN)، ژنومیک، میانگین آنتروپی شرطی (MCE)، الگوریتم جستجوی گرانشی (GSA)، ضریب همبستگی پیرسون.

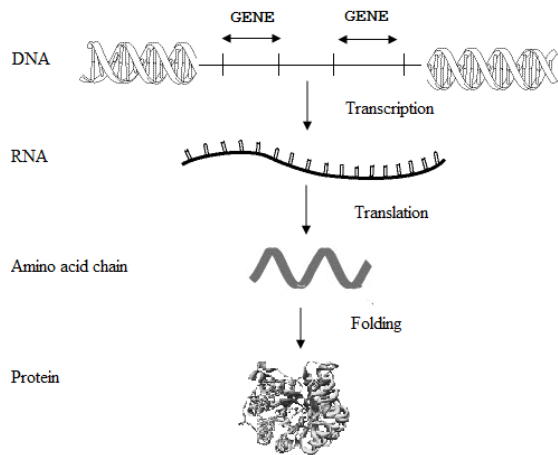
*نویسنده مسئول

نشانی: گروه مهندسی کامپیوتر، دانشکده فنی مهندسی، دانشگاه شهید باهنر، کرمان، کرمان، ایران

تلفن: +۹۸ (۳۴) ۳۲۵۳۴۵۳۱

دورنگار: +۹۸ (۳۴) ۳۲۵۳۴۵۳۱

پست الکترونیکی: ghavami@uk.ac.ir



شکل ۱- فرآیند بیان ژن

تعامل و ارتباط بین ژن‌ها، در قالب یک شبکه پیچیده و به هم پیوسته با عنوان شبکه تنظیم‌کننده ژن (GRN) مدل می‌شود، که می‌تواند مدلی بسیار مؤثر برای شناسایی روابط بین ژن‌ها باشد و نیز، نحوه تأثیر آنها بر یکدیگر را به‌وضوح نشان دهد. در GRN، پیش‌بینی‌کننده^۱ یک ژن، ژنی است که در تنظیم آن نقش دارد؛ براین اساس، زیرمجموعه پیش‌بینی‌کننده^۲ هر ژن به مجموعه ژن‌هایی گفته می‌شود، که به‌طور مستقیم در تنظیم آن ژن شرکت دارند. مجموعه پیش‌بینی‌کننده^۳ همچنین شامل تمام زیرمجموعه پیش‌بینی‌کننده‌های موجود در یک GRN است که این مجموعه، نشان‌دهنده توپولوژی شبکه است [۱۲، ۱۳]. شکل ۲، مفهوم زیرمجموعه پیش‌بینی‌کننده و مجموعه پیش‌بینی‌کننده در GRN را نشان می‌دهد. در این شکل، سه پیش‌بینی‌کننده برای هر ژن در نظر گرفته شده است.

استنتاج GRN با استفاده از داده‌های بیان ژن، که به مهندسی معکوس^۴ نیز معروف است؛ به‌عنوان یکی از چالش‌برانگیزترین مسائل پژوهشی در زمینه بیوانفورماتیک مطرح شده است [۱۴]. بزرگ‌ترین چالش برای استنتاج روابط بین ژن‌ها، حجم عظیم ژن‌ها، کم بودن تعداد نمونه‌ها و عدم دقت کافی در اندازه‌گیری داده‌های بیان ژن (نویزی بودن داده‌ها) است [۱۵، ۱۶].

تاکنون، روش‌های مختلفی برای استنتاج GRN معرفی شده است [۱۷]. در حوزه استنتاج GRN با استفاده از روش‌های انتخاب ویژگی، لویز و همکارانش [۱۸] یک روش جستجوی

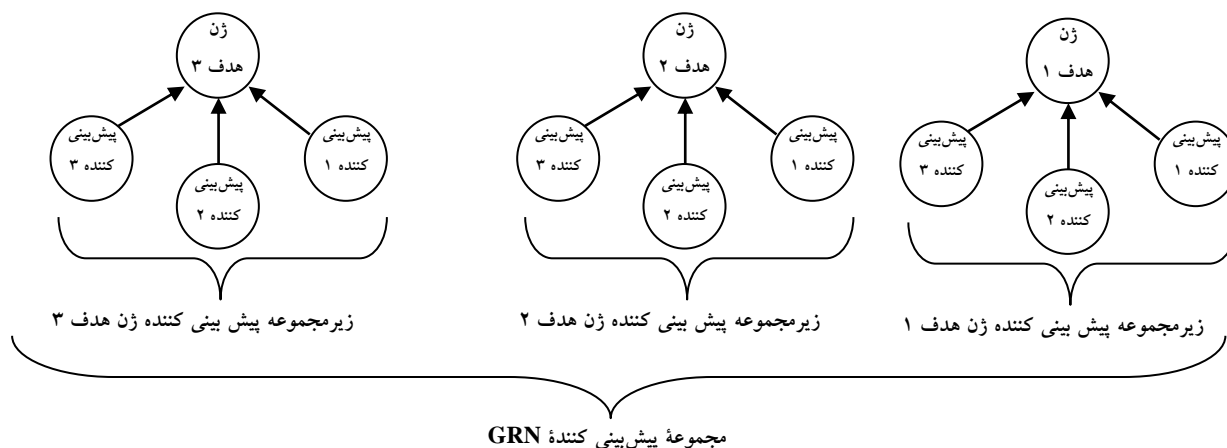
۱- مقدمه

مولکول بزرگی به نام پروتئین^۱ مسئول تمام عملکردهای یک سلول است. پروتئین توسط ژن سنتز می‌شود و می‌تواند به‌عنوان عامل رونویسی تنظیم بیان ژن دیگر عمل کند. بنابراین، یک موجود زنده را می‌توان مانند یک شبکه پیچیده و به هم پیوسته از مولکول‌های متصل‌شده توسط واکنش‌های بیوشیمیایی در نظر گرفت [۱]. مکانیسم تنظیمی حاصل از این واکنش‌ها، به شکل یک سیستم پیچیده از ارسال و دریافت سیگنال‌ها در نظر گرفته می‌شود، که می‌تواند برای شناسایی مکانیسم‌های کنترل سلول و روابط بین نهادهای مختلف زیست‌شناسی بررسی شود. درک روابط بین ژن‌ها و نحوه تنظیم ژن‌ها از طریق انتقال سیگنال، هدفی مهم در سیستم‌های زیست‌شناسی است [۲].

هر پروتئین از زنجیره اسیدآمینه ساخته شده است، که توسط ژن متناظر با آن تعیین می‌شود. تمام اطلاعات ژنتیکی هر شخص، در DNA آن شخص کدگذاری شده است. برای ساخت پروتئین، ابتدا RNA از روی DNA رونویسی^۲ می‌شود و در مرحله بعد، RNA به اسیدآمینه ترجمه^۳ می‌شود؛ سپس زنجیره‌ای از اسیدآمینه‌ها به‌صورت سه‌بعدی در می‌آیند، که به این عمل تاشدگی^۴ می‌گویند. این حالت سه‌بعدی اسیدآمینه‌ها، همان پروتئین است. این فرآیند، بیان ژن^۵ نام دارد. شکل ۱، نشان‌دهنده این بخش از فرآیند بیان ژن است. به‌طور خلاصه، ترتیب اسیدآمینه‌ها در یک پروتئین توسط ژن مربوط به آن تعیین می‌شود. در واقع، ژن‌ها آرایش پروتئین‌های بدن را تعیین می‌کنند. تغییر در ژن‌ها، به تغییر در پروتئین‌ها منجر می‌شود و با توجه به اینکه هر شخص اطلاعات ژنتیک منحصر به فرد دارد؛ در نتیجه این تنوع ژنتیک، عامل پاسخ متفاوت هر فرد به یک داروی یکسان است [۳-۶].

بعد از توسعه تکنولوژی‌هایی برای استخراج عبارت‌های بیان ژن، مانند DNA Microarray [۷]، SAGE [۸] و RNA Sequencing [۹]، اندازه‌گیری ده‌ها هزار عبارت ژنی به‌صورت هم‌زمان ممکن شد. با افزایش دسترسی به داده‌های بیان ژن، توجه محققین به روابط بین ژن‌ها و نحوه تنظیم آنها معطوف شد [۱۰، ۱۱].

^۱ Protein^۲ Transcription^۳ Translation^۴ Folding^۵ Gene Expression^۶ Predictor^۷ Predictor Subset^۸ Predictor Set^۹ Reverse Engineering



شکل ۲- زیرمجموعه پیش‌بینی کننده و مجموعه پیش‌بینی کننده در GRN

ژن‌های پیش‌بینی کننده در هر کروموزوم نگهداری می‌شوند. جیمنز و همکارانش [۲۳] با استفاده از یک الگوریتم ژنتیک برای هر ژن به‌طور مجزا، GRN را استنتاج کرده‌اند؛ در نتیجه، یک جمعیت اولیه برای هر ژن در نظر گرفته شده است. به کمک انتخاب ویژگی اطلاعات متقابل^۵ (MI)، یک سری از پیش‌بینی کننده‌های مرتبط با هر ژن هدف انتخاب می‌شوند و جمعیت اولیه هر ژن را تشکیل می‌دهند. این عمل، برای تمام ژن‌های شبکه تکرار می‌شود و در مراحل بعد، عملیات ترکیب و جهش انجام می‌شوند. معیار اطلاعاتی آکائیکه نیز به‌عنوان تابع برازندگی در نظر گرفته شده است.

شوکت و کبیر [۲۴] با استفاده از الگوریتم تکاملی قدرت پرتو^۶ SPEA2+ و یک مدل خطی متغیر در زمان، GRN را استنتاج کردند. در این روش، پارامترهای مدل در هر کروموزوم کدگذاری شده‌اند. توابع برازندگی، براساس میانگین مربعات خطا^۷ و کمینه کردن تعداد روابط بین ژن‌ها می‌باشند. لی و همکارانش [۲۵] با ارائه یک مدل ترکیبی از دو الگوریتم ژنتیک و بهینه‌سازی ازدحام ذرات^۸، به جستجوی بهترین شبکه پرداخته‌اند؛ همچنین در این مقاله، یک راهکار تقسیم و غلبه برای مقابله با شبکه‌های بزرگ استفاده شده است. بطور کلی، تمام روش‌های ذکر شده در یک یا چند مورد زیر ضعف دارند:

- کم بودن قابلیت اطمینان شبکه‌های GRN ساخته شده
- عدم قابلیت توسعه‌پذیری برای داده‌های بزرگ
- گیر کردن در بهینه محلی و سرعت همگرایی کم برای مدل‌های ارائه شده در حوزه الگوریتم‌های تکاملی

شناور برای استنتاج GRN ارائه کرده‌اند. این روش با استفاده از یک روش کلاسیک، انتخاب ویژگی با عنوان انتخاب به جلو شناور متوالی^۱ (SFFS) و با انتخاب ریشه‌های متعدد در آغاز، به جستجوی پیش‌بینی کننده‌ها می‌پردازد و همچنین معیار ضریب تشخیص^۲ (CoD) برای ارزیابی پیش‌بینی کننده‌ها استفاده شده است. لویز و همکارانش [۱۹] با استفاده از روش انتخاب ویژگی SFFS، زیرمجموعه پیش‌بینی کننده را به ازای هر ژن هدف مبتنی بر معیار ارزیابی میانگین آنتروپی شرطی^۳ (MCE) انتخاب کرده‌اند. این روش

فضای جستجو را با استفاده از یک روش مقیاس آزاد برای کاهش فضا، جستجو می‌کند. همچنین، در مرجع [۲۰] علاوه بر داده‌های بیان ژن، از سایر داده‌های زیست‌شناسی نیز برای افزایش دقت ساخت GRN استفاده شده است. در این روش، با ترکیب روش انتخاب ویژگی SFFS و محاسبه MCE، به‌عنوان معیار ارزیابی، با استفاده از داده‌های بیان ژن و محاسبه معیار امتیاز زیست‌شناسی^۴ (BS) با استفاده از سایر داده‌های زیست‌شناسی، روابط بین ژن‌ها به‌خوبی نتیجه شده است. در [۲۱] نیز، روشی برای استنتاج مجموعه پیش‌بینی کننده با استفاده از بهره اطلاعات، ارائه شده است. همچنین در این روش، قیدی برای انتخاب پیش‌بینی کننده‌های مرتبط‌تر اعمال شده است.

در حوزه استنتاج GRN با استفاده از الگوریتم‌های تکاملی، مندوزا و همکارانش [۲۲] با استفاده از الگوریتم ژنتیک، GRN را استنتاج کرده‌اند. در این روش از معیار آنتروپی Tsallis، برای ارزیابی کروموزوم‌ها استفاده شده است. همچنین، شماره

^۵ Mutual Information

^۶ Strength Pareto Evolutionary Algorithm 2

^۷ Mean Square Error (MSE)

^۸ Particle Swarm Optimization (PSO)

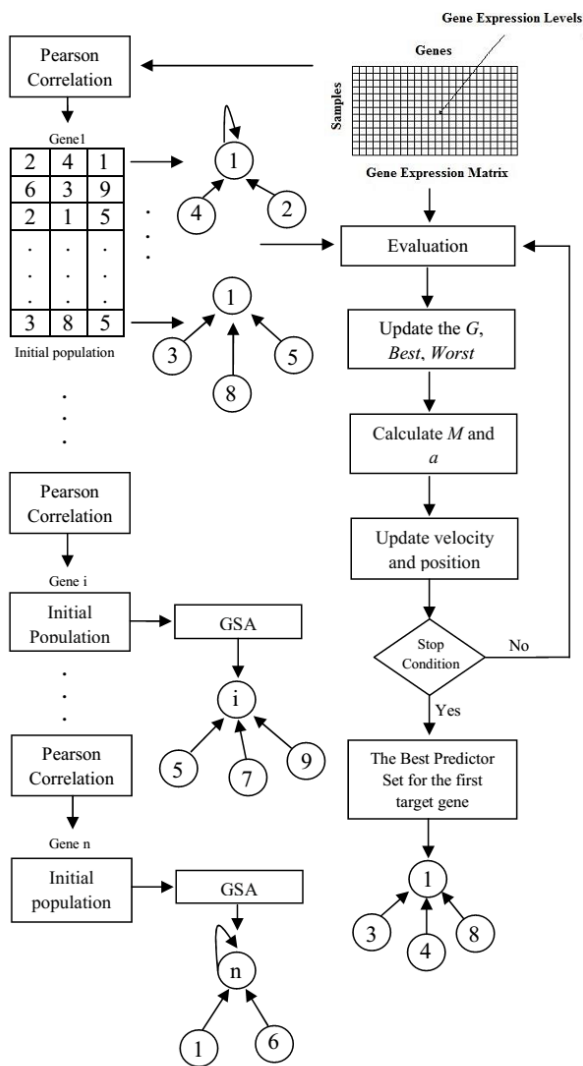
^۱ Sequential Floating Forward Selection

^۲ Coefficient of Determination

^۳ Mean Conditional Entropy

^۴ Biological Score

ضریب همبستگی پیرسون به‌ازای هر ژن هدف و تک‌تک ژن‌های موجود در داده‌های بیان ژن محاسبه می‌شود. بعد از مرتب‌سازی مقادیر این ضرایب، ژن‌ها با همبستگی بیشتر نسبت به ژن هدف (مقدار ضریب همبستگی بیشتر) انتخاب می‌شوند و جمعیت اولیه آن ژن هدف را تشکیل می‌دهند. ارزیابی هر جرم، با استفاده از MCE انجام می‌شود. محاسبه MCE نیز با استفاده از داده‌های بیان ژن انجام می‌شود. بعد از اجرای مراحل GSA، چنانچه شرط خاتمه برقرار باشد، بهترین زیرمجموعه پیش‌بینی‌کننده جستجو شده برای آن ژن هدف، به‌عنوان خروجی برگردانده می‌شود. بهترین زیرمجموعه پیش‌بینی‌کننده‌ها، توپولوژی GRN را تشکیل می‌دهند. تمام قسمت‌های روش پیشنهادی در ادامه با جزئیات توضیح داده می‌شوند.



شکل ۳- شمای روش پیشنهادی برای استنتاج GRN

در این مقاله، برای رفع برخی از مشکلات ذکر شده، یک روش جدید براساس الگوریتم جستجوی گرانشی^۱ (GSA) برای استنتاج مجموعه پیش‌بینی‌کننده ارائه شده است. به‌ازای هر ژن هدف، یک GSA برای یافتن بهترین زیرمجموعه پیش‌بینی‌کننده آن ژن هدف استفاده شده است. جمعیت اولیه برای هر ژن هدف، برای انتخاب بهتر پیش‌بینی‌کننده‌ها و پایداری بیشتر پاسخ‌ها براساس ضریب همبستگی پیرسون^۲ تولید می‌شود. برای ارزیابی جرم‌ها، از تابع میانگین آنتروپی شرطی^۳ (MCE) استفاده شده است. نتایج حاصل از مقایسه راهکار پیشنهادی و روش ارائه شده توسط جیمنز و همکارانش [۲۳] و نیز مندوزا و همکارانش [۲۲] روی داده‌های زیست‌شناسی، نشان می‌دهد که روش پیشنهادی می‌تواند مجموعه پیش‌بینی‌کننده را در GRN، با قابلیت اطمینان بالاتری استنتاج کند.

ادامه این مقاله به این صورت سازمان‌دهی شده است؛ بخش ۲، روش پیشنهادی استنتاج مجموعه پیش‌بینی‌کننده را معرفی می‌کند؛ نتایج حاصل از اعمال روش ارائه شده، در بخش ۳ بیان می‌شوند؛ درنهایت، بخش ۴ به نتیجه‌گیری مقاله می‌پردازد.

۲- روش پیشنهادی برای استنتاج مجموعه پیش‌بینی‌کننده در GRN

شمای روش پیشنهادی برای استنتاج مجموعه پیش‌بینی‌کننده، در شکل ۳ نشان داده شده است. در این روش، برای یافتن بهترین زیرمجموعه پیش‌بینی‌کننده هر ژن هدف از الگوریتم GSA استفاده شده است، که نسبت به الگوریتم GA به تنظیم پارامترهای کمتری نیاز دارد و همچنین، سرعت همگرایی بهتری دارد [۲۶، ۲۷]. در روش پیشنهادی، یک GSA به‌ازای هر ژن هدف برای یافتن بهترین زیرمجموعه پیش‌بینی‌کننده آن ژن اجرا می‌شود. اگر n تعداد ژن‌های هدف باشد؛ پس در این روش، n الگوریتم GSA اجرا می‌شود. هر جرم^۴ در GSA، نشان‌دهنده یک زیرمجموعه پیش‌بینی‌کننده برای آن ژن هدف است. پیش‌بینی‌کننده‌ها برای هر ژن هدف، با استفاده از روش ضریب همبستگی پیرسون انتخاب می‌شوند. در حقیقت، با استفاده از داده‌های بیان ژن، مقدار

^۱Gravitational Search Algorithm

^۲Pearson Correlation Coefficient

^۳ Mean Conditional Entropy

^۴Mass

۲-۱- الگوریتم جستجوی گرانشی

در الگوریتم GSA [۲۶، ۲۷]، یافتن پاسخ بهینه، براساس قوانین گرانشی و حرکت انجام می‌شود. تمام اشیا، یکدیگر را توسط نیروی جاذبه جذب می‌کنند. این نیرو، باعث جنبش سراسری تمام اشیا به سمت اشیا با جرم سنگین‌تر می‌شود. در GSA، مجموعه‌ای از جرم‌ها مسئول تشکیل جمعیت اولیه هستند. هر جرم دارای چهار مشخصه موقعیت، جرم اینرسی، جرم گرانشی فعال و جرم گرانشی غیرفعال است. جرم گرانشی فعال، معیاری از میزان شدت نیروی گرانشی حول یک جسم و جرم اینرسی، معیاری از مقاومت جسم در مقابل حرکت است. مقدار دو مشخصه جرم گرانشی و اینرسی، با توجه به برآزندگی هر جرم تعیین می‌شوند. موقعیت هر جرم، نقطه‌ای در فضا است که جوابی از مسئله می‌باشد. اگر سیستمی متشکل از N جرم (عامل) در نظر گرفته شود، موقعیت عامل i ام در D بعد به صورت رابطه زیر است:

$$X_i = (x_i^1, \dots, x_i^d, \dots, x_i^D) \quad (1)$$

نیروی وارده بر عامل i ام از سمت عامل j ام در زمان t ، به صورت رابطه ۲ تعریف می‌شود [۲۷]:

$$F_{ij}^d(t) = G(t) \frac{M_{pi}(t) * M_{aj}(t)}{R_{ij}(t) + \epsilon} (x_j^d(t) - x_i^d(t)) \quad (2)$$

به طوری که، M_{aj} برابر با جرم گرانشی فعال مرتبط با عامل j ام و M_{pi} نیز برابر با جرم گرانشی غیرفعال عامل i ام است. $G(t)$ ، ثابت گرانشی در زمان t ، ϵ یک مقدار ثابت و R_{ij} نیز برابر با فاصله اقلیدسی بین دو عامل i و j است، که از طریق رابطه ۳ به دست می‌آید [۲۷]:

$$R_{ij}(t) = \|X_i(t), X_j(t)\|_2 \quad (3)$$

برای اعمال خاصیت اتفاق (تصادف)، کل نیروی وارده به عامل i ام در d بعد، به صورت رابطه ۴ محاسبه می‌شود [۲۷]:

$$F_i^d(t) = \sum_{j=1, j \neq i}^N \text{rand}_j F_{ij}^d(t) \quad (4)$$

که rand_j ، یک عدد تصادفی را در بازه $[0,1]$ تولید می‌کند. طبق قانون حرکت، شتاب عامل i ام در زمان t و در جهت d ، به صورت رابطه ۵ محاسبه می‌شود [۲۷]:

$$a_i^d(t) = \frac{F_i^d(t)}{M_{ii}(t)} \quad (5)$$

در رابطه ۵، M_{ii} جرم اینرسی عامل i ام است. موقعیت و سرعت هر عامل در زمان $t+1$ به ترتیب توسط رابطه‌های ۶ و ۷ بیان می‌شود [۲۷]:

$$v_i^d(t+1) = \text{rand}_i * v_i^d(t) + a_i^d(t) \quad (6)$$

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) \quad (7)$$

در رابطه ۶، rand_i یک متغیر تصادفی یکنواخت در بازه $[0,1]$ است. در ابتدا، یک مقدار اولیه (G_0) به ثابت گرانشی G داده می‌شود، که این مقدار با گذر زمان برای کنترل دقت جستجو، کاهش خواهد یافت. بنابراین G ، طبق رابطه ۸ تابعی از زمان و مقدار اولیه‌اش خواهد بود [۲۷]:

$$G(t) = G(G_0, t) \quad (8)$$

اگر M_i برابر با جرم اینرسی یا گرانشی باشد، مقدار مربوط به این جرم‌ها، براساس مقدار تابع ارزیابی توسط روابط زیر محاسبه می‌شود [۲۷]:

$$m_i(t) = \frac{\text{fit}_i(t) - \text{worst}(t)}{\text{best}(t) - \text{worst}(t)} \quad (9)$$

$$M_i(t) = \frac{m_i(t)}{\sum_{j=1}^N m_j(t)} \quad (10)$$

$$M_{ii} = M_{ai} = M_{pi} = M_i \quad (11)$$

در رابطه ۹، $\text{best}(t)$ و $\text{worst}(t)$ به ترتیب بدترین و بهترین مقدار تابع ارزیابی با توجه به نوع مسئله (بیشینه یا کمینه) هستند و fit_i نیز برابر با مقدار تابع ارزیابی مربوط به عامل i ام در زمان t است [۲۷].

۲-۲- نحوه نمایش جرم‌ها

هر جرم در الگوریتم GSA بیانگر یک راه حل از مسئله است، که این راه حل برای مسئله استنتاج پیش‌بینی‌کننده‌ها در GRN را می‌توان یک گراف از وابستگی‌های ژن هدف و سایر ژن‌ها در مجموعه داده‌های بیان ژن دانست؛ بنابراین، هر جرم باید شامل اطلاعاتی درباره توپولوژی ژن هدف باشد. در روش پیشنهادی، هر جرم از مجموعه‌ای با اندازه k که شامل تعدادی شاخص است، تشکیل شده است. هر شاخص، بیانگر اندیس ژن پیش‌بینی‌کننده است. در این مقاله، k برابر با تعداد پیش‌بینی‌کننده‌ها فرض شده است.

شکل ۴، مثالی از نحوه کدگذاری جرم‌ها و مفهوم آن را نشان می‌دهد. در شکل ۴-الف، بهترین جرم ژن‌های هدف ۱، ۲، ۳، ۴ و ۵ نشان داده شده است. همان‌طور که مشخص است، ژن‌های

که \bar{X} و \bar{Y} طبق رابطه ۱۳ تعریف می‌شوند و X_i و Y_i مقادیر متغیرها هستند (زوج داده‌ها برای بررسی میزان همبستگی) [۲۸]:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i \quad (13)$$

در این روش، ابتدا ژن‌ها براساس مقادیر ضریب همبستگی پیروسون، به‌ازای هر ژن هدف، مرتب می‌شوند (مقدار بیشتر این ضریب، همبستگی بیشتر را نشان می‌دهد)؛ سپس، ژن‌ها با بیشترین میزان همبستگی، به‌عنوان پیش‌بینی‌کننده‌های یک ژن هدف انتخاب می‌شوند و جمعیت اولیه ژن را تشکیل می‌دهند (مجموعه‌ای از زیرمجموعه پیش‌بینی‌کننده‌های ژن هدف).

۲-۴- تابع هدف

برای ارزیابی هر زیرمجموعه پیش‌بینی‌کننده (هر جرم)، تابع هدف میانگین آنروپی شرطی (MCE) در نظر گرفته شده است. مقدار MCE به‌ازای ژن هدف Y و مجموعه پیش‌بینی‌کننده X ، به‌صورت رابطه زیر محاسبه می‌شود [۲۹]:

$$H(Y|X) = \sum_{x \in X} P(x) H(Y|x) \quad (14)$$

مقدار $H(Y|x)$ از طریق رابطه ۱۵ محاسبه می‌شود [۲۹]:

$$H(Y|x) = - \sum_{y \in Y} P(y|x) \log P(y|x) \quad (15)$$

در رابطه ۱۵، $P(y|x)$ برابر با احتمال شرطی است. مقدار کم H ، نشان‌دهنده مقدار اطلاعات بیشتر به‌دست آمده Y با مشاهده X است.

برای جریمه کردن در مواردی که تعداد نمونه‌ها اندک و تعداد ژن‌ها بسیار زیاد است، راهکاری شامل نمونه‌های مشاهده نشده در محاسبه معیار ارزیابی استفاده می‌شود. MCE با این نوع از جریمه، به‌صورت زیر بیان می‌شود [۲۹]:

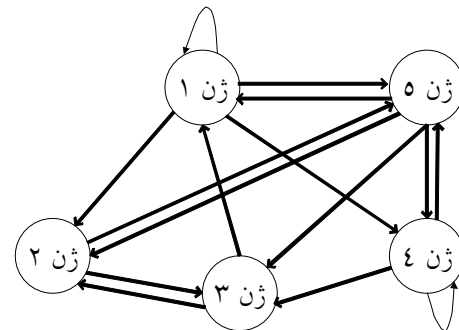
$$H(Y|X) = \frac{\alpha(M - N)H(Y)}{\alpha M + s} + \frac{\sum_{i=1}^N (f_i + \alpha) H(Y|X = x_i)}{\alpha M + s} \quad (16)$$

که M برابر با تعداد نمونه‌های ممکن بردار X و N تعداد نمونه‌های مشاهده‌شده است؛ بنابراین تعداد نمونه‌های مشاهده نشده برابر با $M - N$ است. f_i ، تعداد مشاهدات x_i و s نیز تعداد نمونه‌ها است. جرم احتمال برای موارد غیرمشاهده شده توسط α بیان می‌شود.

۱، ۵ و ۳، پیش‌بینی‌کننده ژن هدف ۲ هستند و ژن هدف ۳، با ژن‌های ۲، ۴ و ۵ پیش‌بینی می‌شود؛ به‌همین صورت برای سایر ژن‌های هدف، بهترین پیش‌بینی‌کننده‌ها مشخص شده‌اند. در شکل ۴-ب نیز این روابط نشان داده شده است.

ژن ۱	ژن ۳	ژن ۵	ژن ۱
ژن ۲	ژن ۳	ژن ۵	ژن ۱
ژن ۳	ژن ۵	ژن ۴	ژن ۲
ژن ۴	ژن ۵	ژن ۱	ژن ۴
ژن ۵	ژن ۱	ژن ۴	ژن ۲

(الف)



(ب)

شکل ۴- نحوه کد کردن یک جرم در GSA

۲-۳- جمعیت اولیه

هر جرم، معادل با یک زیرمجموعه پیش‌بینی‌کننده مرتبط با ژن هدف در نظر گرفته شده است. برای انتخاب پیش‌بینی‌کننده‌های مرتبط‌تر، به‌جای تولید تصادفی جمعیت اولیه، از روش ضریب همبستگی پیروسون استفاده شده است. این عمل باعث می‌شود که زیرمجموعه پیش‌بینی‌کننده‌ها، امیدوارانه‌تر و مطمئن‌تر انتخاب شوند و همچنین، پاسخ‌ها پایدارتر باشند. ضریب همبستگی پیروسون، یک ضریب همبستگی خطی است. اگر دو متغیر با یکدیگر همبستگی مثبت داشته باشند، آنگاه این ضریب به عدد یک (+۱) نزدیک‌تر است. چنانچه دو متغیر با یکدیگر همبستگی منفی داشته باشند، این ضریب به عدد منفی یک (-۱) نزدیک‌تر است و اگر این دو متغیر هیچ همبستگی به یکدیگر نداشته باشند، این ضریب برابر با صفر است. رابطه ۱۲، نشان‌دهنده این ضریب است [۲۸]:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \cdot \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (12)$$

MATLAB و روی ماشینی با پردازنده اینتل 2.50 GHz Core i5 با 4GB RAM پیاده‌سازی شده است. برای بررسی روش ارائه‌شده، دو شبکه از رقابت Dream3 با تعداد ژن‌های مختلف انتخاب شده است. شبکه Dream100 ۱۰۰ ژن، نمونه و شامل ۱۶۶ رابطه در شبکه واقعی دارد و شبکه دوم، شبکه SOS DNA Repair است که ۹ ژن و ۹ نمونه دارد [۳۰].

بعد از استنتاج مجموعه پیش‌بینی‌کننده، برای ارزیابی میزان دقت روش ارائه‌شده، میزان شباهت شبکه به دست آمده با شبکه Gold Standard مقایسه می‌شود. شبکه Gold Standard شامل روابط بین ژن‌ها در شبکه واقعی است. در این مقاله با استفاده از معیارهای شباهت PPV، Sensitivity و Similarity، که براساس تعداد روابط درست و نادرست بین ژن‌ها عمل می‌کنند، روش ارائه‌شده را ارزیابی می‌کنیم. این معیارها، به ترتیب در رابطه‌های ۲۰، ۲۱ و ۲۲ ارائه شده‌اند [۳۱]:

$$PPV = \frac{TP}{TP+FP} \quad (20)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (21)$$

$$Similarity = \sqrt{PPV \cdot Sensitivity} \quad (22)$$

که در این روابط، مقدار TP نشان‌دهنده تعداد روابطی (یال) است که در شبکه Gold Standard حضور دارند و در شبکه ساخته‌شده نیز به درستی نتیجه شده‌اند. مقدار FP نیز بیانگر تعداد روابطی است که در شبکه Gold Standard حضور ندارند؛ اما این روابط در شبکه ساخته‌شده به دست آمده‌اند. به طور مشابه، مقدار FN، تعداد روابطی است که در شبکه Gold Standard حضور دارند؛ اما در شبکه ساخته‌شده استنتاج نشده‌اند. این روابط، در جدول ۱ نشان داده شده‌اند.

جدول ۱- ماتریس سردرگمی

وضعیت یال در شبکه واقعی	استنتاج شده	استنتاج نشده
حضور یال	TP	FN
عدم حضور یال	FP	TN

معیار Similarity، میزان شباهت شبکه حاصل و شبکه Gold Standard را نشان می‌دهد. هر چه مقدار این معیار نزدیک به ۱ باشد، میزان شباهت این دو شبکه و در نتیجه دقت روش ارائه‌شده، بیشتر است.

هر چه مقدار H کمتر باشد، بهتر است؛ بنابراین این مسئله، یک مسئله کمینه‌سازی است. بهترین و بدترین مقدار تابع برازندگی، توسط روابط ۱۷ و ۱۸ به دست می‌آید:

$$\text{best}(t) = \min \text{fit}_j(t) \quad j \in \{1, \dots, N\} \quad (17)$$

$$\text{worst}(t) = \max \text{fit}_j(t) \quad j \in \{1, \dots, N\} \quad (18)$$

مقادیر بیان ژن به صورت پیوسته هستند و از طرفی روش پیشنهادی براساس MCE، به داده‌های گسسته نیاز دارد؛ بنابراین، قبل از استفاده از روش پیشنهادی، یک مرحله گسسته‌سازی روی داده‌های بیان ژن نیاز است. قبل از گسسته‌سازی نیز، یک مرحله نرمال‌سازی نیاز است. در این مقاله، از انتقال نرمال برای نرمال‌سازی استفاده شده است. اگر $g(t)$ به طور فرضی یک ویژگی باشد، فرآیند نرمال‌سازی به صورت رابطه ۱۹ بیان می‌شود [۲۹]:

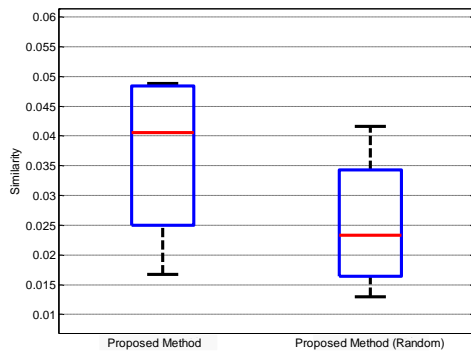
$$\eta(g(t)) = \frac{g(t) - E(g(t))}{\sigma(g(t))} \quad (19)$$

که $E(g(t))$ ، امید ریاضی $g(t)$ و $\sigma(g(t))$ نیز انحراف معیار است. فرآیند گسسته‌سازی مبتنی بر روش ارائه‌شده در مرجع [۲۹] است. ویژگی نرمال‌شده، از حالت پیوسته به m حالت گسسته $(0, 1, \dots, m-1)$ نگاشت می‌شود. در واقع با در نظر گرفتن یک حد آستانه، g به m بازه با اندازه مساوی تقسیم می‌شود؛ در نهایت، داده‌های واقعی (پیوسته) به این m حالت گسسته، نگاشت می‌شوند.

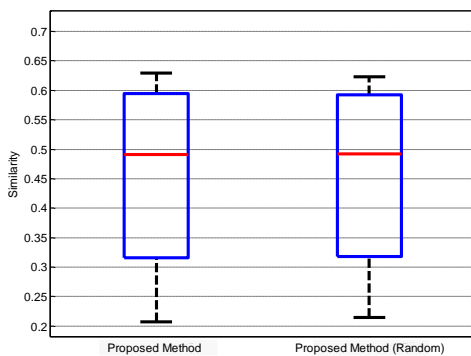
۳- یافته‌ها و بحث

در این بخش، برای بررسی مؤثر بودن تولید جمعیت اولیه با استفاده از ضریب همبستگی پیرسون در مقایسه با تولید جمعیت اولیه به طور تصادفی، ابتدا آزمایشات به دست آمده از این دو روش، تجزیه و تحلیل می‌شوند. در ادامه، نتایج حاصل از روش پیشنهادی با روش ارائه‌شده توسط جیمنز و همکارانش [۲۳] و نیز مندوزا و همکارانش [۲۲]، با استفاده از دو بانک داده مختلف مقایسه می‌شود. در هر دو روش، مجموعه پیش‌بینی‌کننده با استفاده از الگوریتم ژنتیک نتیجه شده است. در مدل جیمنز و همکارانش، یک الگوریتم ژنتیک به‌ازای هر ژن هدف اجرا می‌شود و تابع اطلاعات آکائیکه نیز، به‌عنوان تابع هدف در نظر گرفته شده است. در مدل ارائه‌شده توسط مندوزا و همکارانش، یک شبکه GRN در هر کروموزوم کدگذاری شده است. همچنین، تابع آنتروپی Tsallis به‌عنوان تابع برازندگی فرض شده است. روش پیشنهادی در نرم‌افزار

مشاهده نمی‌شود؛ اما این تفاوت به خوبی در شکل ۷ مشاهده می‌شود. می‌توان نتیجه گرفت که استنتاج GRN با روش پیشنهادی، یعنی تولید جمعیت اولیه با استفاده از ضریب همبستگی پیرسون، می‌تواند کیفیت بالاتری از تولید تصادفی داشته باشد.



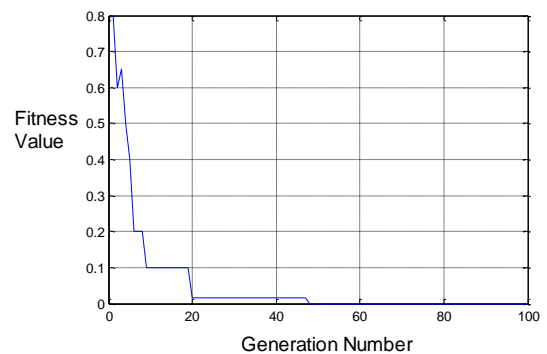
شکل ۷- مقایسه روش پیشنهادی با تولید جمعیت اولیه با استفاده از ضریب همبستگی پیرسون و تولید جمعیت اولیه به‌طور تصادفی، با استفاده از معیار Similarity و بانک داده Dream100



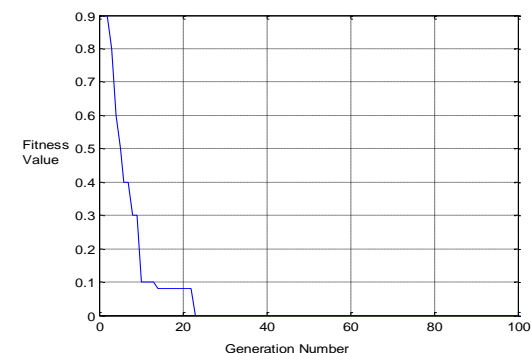
شکل ۸- مقایسه روش پیشنهادی با تولید جمعیت اولیه با استفاده از ضریب همبستگی پیرسون و تولید جمعیت اولیه به‌طور تصادفی، با استفاده از معیار Similarity و بانک داده SOS DNA Repair

شکل ۹، روش پیشنهادی و روش‌ها بارائه شده توسط جیمز و همکارانش [۲۳] و مندوزا و همکارانش [۲۲]، را با استفاده از معیار ارزیابی Similarity و بانک داده Dream100 مقایسه می‌کند. همان‌طور که در شکل مشاهده می‌شود، روش پیشنهادی به مقدار بیشتری Similarity رسیده است. در شکل‌های ۱۰ و ۱۱، نیز روش پیشنهادی و روش‌های ارائه شده توسط جیمز و همکارانش و مندوزا و همکارانش به ترتیب با استفاده از معیارهای ارزیابی Sensitivity و PPV مقایسه می‌شوند. بانک داده استفاده شده در این شکل‌ها، Dream100 است. براساس نتایج، روش پیشنهادی، دقت بیشتری در استنتاج GRN دارد.

برای هر آزمایش، اندازه جمعیت و تعداد تکرارهای GSA به ترتیب ۱۰۰ و ۱۰۰۰ در نظر گرفته شده است. به علاوه، برای در نظر گرفتن اثرات تصادفی GSA، روش پیشنهادی، ۲۰ بار برای هر آزمایش اجرا می‌شود؛ در نتیجه، نتایج هر آزمایش، حاصل میانگین ۲۰ بار اجرای روش پیشنهادی هستند. شکل‌های ۵ و ۶، نشان‌دهنده سرعت همگرایی الگوریتم برای ژن اول در دو بانک داده مختلف در ۱۰۰ نسل هستند.



شکل ۵- سرعت همگرایی الگوریتم در ۱۰۰ نسل و بانک داده Dream 100



شکل ۶- سرعت همگرایی الگوریتم در ۱۰۰ نسل و بانک داده SOS DNA Repair

برای بررسی نقش ضریب همبستگی پیرسون در تولید پیش‌بینی‌کننده‌های اولیه، ابتدا جمعیت اولیه در روش پیشنهادی با استفاده از ضریب همبستگی پیرسون تولید می‌شود و سپس، جمعیت اولیه به‌طور تصادفی تولید می‌شود. شکل ۷، نتایج به دست آمده از دو روش با استفاده از بانک داده Dream100 را نشان می‌دهد. همان‌طور که در شکل مشاهده می‌شود، تولید جمعیت اولیه با استفاده از ضریب همبستگی پیرسون، عملکرد بهتری داشته است. شکل ۸، نیز نتایج حاصل از بانک داده SOS DNA Repair را نشان می‌دهد. در این شکل، با توجه به این‌که بانک داده استفاده شده بسیار کوچک است، تفاوت چندانی در تولید جمعیت اولیه به‌طور تصادفی یا با استفاده از ضریب همبستگی پیرسون،

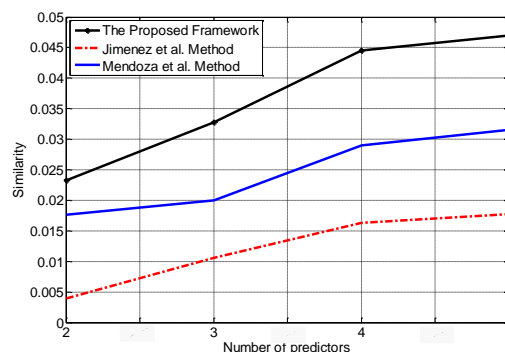
جدول ۲- مقایسه روش پیشنهادی و روش‌های ارائه شده توسط جیمنز و همکارانش و مندوزا و همکارانش با استفاده از معیارهای ارزیابی Similarity, Sensitivity, PPV, TP, FP و FN و بانک داده SOS DNA Repair (این نتایج، حاصل میانگین ۲۰ بار اجرای روش پیشنهادی می‌باشند)

تعداد پیش‌بینی‌کننده‌ها	معیار ارزیابی	نام روش		
		روش پیشنهادی	جیمنز و همکاران	مندوزا و همکاران
۴	Similarity	۰/۲۰۷۵	۰/۲۰۴۱	۰/۲۰۴۱
	Sensitivity	۰/۱۲۷۱	۰/۱۲۵۰	۰/۱۲۵۰
	PPV	۰/۳۳۸۹	۰/۳۳۳۳	۰/۳۳۳۳
	TP	۶/۱	۶	۶
	FP	۱۱/۹	۱۲	۱۲
	FN	۴۱/۹	۴۲	۴۲
۴	Similarity	۰/۴۲۳۶	۰/۳۱۲۵	۰/۴۸۱۵
	Sensitivity	۰/۳۱۷۷	۰/۲۳۴۴	۰/۲۷۰۸
	PPV	۰/۵۶۴۸	۰/۴۱۶۷	۰/۳۶۱۱
	TP	۱۵/۲۵	۱۱/۲۵	۱۳
	FP	۱۱/۷۵	۱۵/۷۵	۱۴
	FN	۳۲/۷۵	۳۶/۷۵	۳۴/۲۵
۴	Similarity	۰/۵۵۹۳	۰/۳۹۸۱	۰/۵۱۱۱
	Sensitivity	۰/۴۸۴۴	۰/۳۴۴۸	۰/۳۸۳۳
	PPV	۰/۶۴۵۸	۰/۴۵۹۷	۰/۴۴۲۶
	TP	۲۳/۲۵	۱۶/۵۵	۱۸/۳
	FP	۱۲/۷۵	۱۹/۴۵	۱۸/۷۵
	FN	۲۴/۷۵	۳۱/۴۵	۲۸/۷۰
۴	Similarity	۰/۶۲۹۴	۰/۴۷۰۱	۰/۵۶۶۷
	Sensitivity	۰/۶۰۹۴	۰/۴۵۵۲	۰/۵۳۱۳
	PPV	۰/۶۵	۰/۴۸۵۶	۰/۵۴۸۷
	TP	۲۹/۲۵	۲۱/۸۵	۲۵/۶
	FP	۱۵/۷۵	۲۳/۱۵	۲۰/۸۴
	FN	۱۸/۷۵	۲۶/۱۵	۲۱/۲۵

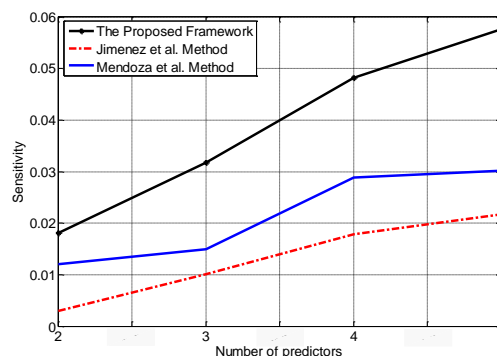
براساس نتایج مشاهده شده، روش پیشنهادی، دقت و کارایی بیشتری دارد. برخی از تفاوت‌های این روش، که به بهبود دقت و کارایی آن نسبت به روش‌های دیگر منجر شده است، در زیر ارائه شده‌اند:

- سرعت همگرایی بهتر الگوریتم GSA [۲۶، ۲۷] و توانایی حل مسائل بهینه‌سازی مختلف با کارایی بالا [۳۲، ۳۳]
- تولید جمعیت اولیه با استفاده از ضریب همبستگی پیرسون
- استفاده از میانگین آنروپی شرطی به‌عنوان تابع برازندگی

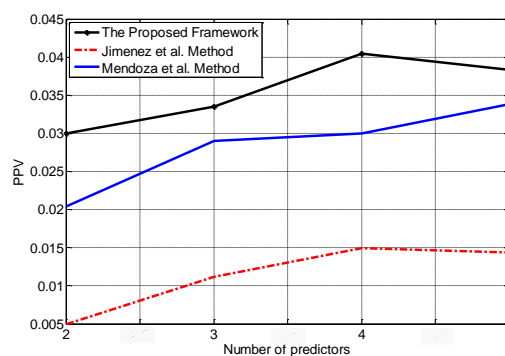
در جدول ۲، نیز روش پیشنهادی و روش‌های ارائه شده توسط جیمنز و همکارانش و مندوزا و همکارانش، با استفاده از بانک داده SOS DNA Repair مقایسه می‌شوند. در این جدول، معیارهای ارزیابی برای مقایسه این دو روش Similarity, Sensitivity, PPV, TP, FP و FN در نظر گرفته شده‌اند. همان‌طور که در جدول مشاهده می‌شود، روش پیشنهادی در این بانک داده نیز به نتایج برتری دست یافته است و مقدار معیار Similarity، تقریباً ۳۳٪ افزایش یافته است.



شکل ۹- مقایسه روش پیشنهادی و روش‌های ارائه شده توسط جیمنز و همکارانش و مندوزا و همکارانش با استفاده از معیار ارزیابی Similarity و بانک داده Dream100



شکل ۱۰- مقایسه روش پیشنهادی و روش‌های ارائه شده توسط جیمنز و همکارانش و مندوزا و همکارانش با استفاده از معیار ارزیابی Sensitivity و بانک داده Dream100



شکل ۱۱- مقایسه روش پیشنهادی و روش‌های ارائه شده توسط جیمنز و همکارانش و مندوزا و همکارانش با استفاده از معیار ارزیابی PPV و بانک داده Dream100

۴- نتیجه‌گیری

استنتاج شبکه تنظیم‌کننده ژن (GRN) با استفاده از داده‌های بیان ژن، از مسائل مهم در زمینه سیستم‌های زیست‌شناسی است. استنتاج صحیح GRN، نیازمند استنتاج صحیح زیرمجموعه پیش‌بینی‌کننده هر ژن هدف است. در این مقاله برای استنتاج مجموعه پیش‌بینی‌کننده، از الگوریتم جستجوی گرانشی (GSA) با تابع برازندگی میانگین آنتروپی شرطی (MCE) استفاده شده است. یکی از معایب استفاده از الگوریتم‌های تکاملی، ناپایداری پاسخ‌ها است. در این مقاله برای حل این مشکل، استفاده از ضریب همبستگی پیرسون جایگزین روش تصادفی در تولید جمعیت اولیه به‌جای روش می‌شود. نتایج حاصل با استفاده از داده‌های زیست‌شناسی، نشان می‌دهند که روش ارائه‌شده، دقت بالایی دارد. روش ارائه‌شده می‌تواند به‌راحتی موازی شود؛ به‌این صورت که، می‌توان هر الگوریتم GSA (یکی برای هر ژن) را مستقل از دیگری پردازش کرد. به‌عبارت دیگر، زیرمجموعه پیش‌بینی‌کننده هر ژن هدف، به‌طور مستقل و هم‌زمان با ژن هدف دیگر استنتاج شود؛ بنابراین، هر هسته پردازش، مسئول به‌دست آوردن بهترین زیرمجموعه پیش‌بینی‌کننده برای یک ژن هدف است. در این صورت، شبکه‌های واقعی با وجود هزاران ژن، می‌توانند به‌سرعت استنتاج شوند.

۵- مراجع

- [8] Velculescu, V.E., L. Zhang, B. Vogelstein and K. Kinzler, *Serial analysis of gene expression*. Science, 1995. **270**(5235): p. 484-487.
- [9] Ray, S.S., S. Halder, and D. Bhattacharyya, *HD-RNAS: hierarchical database of RNA structures*. 2009.
- [10] Sîrbu, A., M. Crane, and H.J. Ruskin, *EGIA—evolutionary optimisation of gene regulatory networks, an integrative approach*, in *Complex Networks V*. 2014, Springer. p. 217-229.
- [11] da Rocha Vicente, F.F. and F.M. Lopes, *SFFS-SW: A Feature Selection Algorithm Exploring the Small-World Properties of GNs*, in *Pattern Recognition in Bioinformatics*. 2014, Springer. p. 60-71.
- [12] Shmulevich, I. and E.R. Dougherty, *Genomic signal processing*. 2014: Princeton University Press.
- [13] Lin, P.-C.K. and S.P. Khatri, *Logic Synthesis for Genetic Diseases*. 2014: Springer. 93-95.
- [14] Marbach, D., J. Costello, R. Küffner, N. Vega, R. Prill, D. Camacho, K. Allison, M. Kellis, J. Collins and G. Stolovitzky, *Wisdom of crowds for robust gene network inference*. Nature methods, 2012. **9**(8): p. 796-804.
- [15] Dougherty, E.R., *The fundamental role of pattern recognition for gene-expression/microarray data in bioinformatics*. Pattern Recognition, 2005. **38**(12): p. 2226-2228.
- [16] Emmert-Streib, F., G. Glazko, and R. De Matos Simoes, *Statistical inference and reverse engineering of gene regulatory networks from observational expression data*. Frontiers in genetics, 2012. **3**: p. 8.
- [17] Chai, L.E., S.K. Loh, S.T. Low, M.S. Mohamad, S. Deris and Z. Zakaria, *A review on the computational approaches for gene regulatory network construction*. Computers in biology and medicine, 2014. **48**: p. 55-65.
- [18] Lopes, F.M., D.C. Martins Jr, J. Barrera and R.M. Cesar Jr, *SFFS-MR: a floating search strategy for GRNs inference*, in *Pattern Recognition in Bioinformatics*. 2010, Springer. p. 407-418.
- [19] Lopes, F.M., D.C. Martins, J. Barrera and R.M. Cesar, *A feature selection technique for inference of graphs from their known topological properties: Revealing scale-free gene regulatory networks*. Information Sciences, 2014. **272**: p. 1-15.
- [20] Lopes, F.M., S.S. Ray, R.F. Hashimoto and R.M. Cesar, *Entropic Biological Score: a cell cycle investigation for GRNs inference*. Gene, 2014. **541**(2): p. 129-137.
- [21] Jafari, M., B. Ghavami, and V.S. Naeni. *An information gain approach to infer gene regulatory networks*. in *2015 22nd Iranian Conference on Biomedical Engineering (ICBME)*. 2015. IEEE.
- [22] Mendoza, M.R., F.M. Lopes, and A.L.C. Bazzan. *Reverse engineering of grns: An evolutionary approach based on the tsallis entropy*. in *Proceedings of the 14th annual conference on*
- [1] Voet, D., J.G. Voet, and C.W. Pratt, *Fundamentals of Biochemistry: Life at the molecular level*. 2006.
- [2] Balkenhol, N., S. Cushman, A. Storfer and L. Waits, *Landscape Genetics: Concepts, Methods, Applications*. 2015: John Wiley & Sons.
- [3] Cox, M.M., *Lehninger principles of biochemistry*. 2013: Freeman.
- [4] Jones, B., *Gene expression: Layers of gene regulation*. Nature Reviews Genetics, 2015. **16**(3): p. 128-129.
- [5] Abrahams, E. and M. Silver, *The case for personalized medicine*. Journal of diabetes science and technology, 2009. **3**(4): p. 680-684.
- [6] Ginsburg, G.S. and H.F. Willard, *Genomic and personalized medicine: foundations and applications*. Translational research, 2009. **154**(6): p. 277-287.
- [7] Hori, M., J. Shibato, T. Nakamachi, R. Rakwal, T. Ogawa, S. Shioda and S. Numazawa, *Two-color Dye-swap DNA Microarray approach toward confident gene expression profiling in PMCAO mouse model for ischemia-related and PACAP38-influenced genes*. Genomics Data, 2015. **3**: p. 148-154.

- [28] Wang, J., *Pearson Correlation Coefficient*, in *Encyclopedia of Systems Biology*. 2013, Springer. p. 1671-1671.
- [29] Lopes, F.M., D.C. Martins, and R.M. Cesar, *Feature selection environment for genomic applications*. BMC Bioinformatics, 2009. **10**(1): p. 285.
- [30] Aghdam, R., M. Ganjali, and C. Eslahchi, *IPCA-CMI: an algorithm for inferring gene regulatory networks based on a combination of PCA-CMI and MIT score*. PloS one, 2014. **9**(4).
- [31] Dougherty, E.R., *Validation of gene regulatory networks: scientific and inferential*. Briefings in bioinformatics, 2011. **12**(3): p. 245-252.
- [32] Li, C. and J. Zhou, *Parameters identification of hydraulic turbine governing system using improved gravitational search algorithm*. Energy Conversion and Management, 2011. **52**(1): p. 374-381.
- [33] Hassanzadeh, H.R. and M. Rouhani. *A multi-objective gravitational search algorithm*. in *Computational Intelligence, Communication Systems and Networks (CICSyN), 2010 Second International Conference on*. 2010. IEEE.
- Genetic and evolutionary computation*. 2012. ACM.
- [23] Jimenez, R.D., D.C. Martins-Jr, and C.S. Santos, *One genetic algorithm per gene to infer gene networks from expression data*. Network Modeling Analysis in Health Informatics and Bioinformatics, 2015. **4**(1): p. 1-22.
- [24] Showkat, D. and M. Kabir. *Inference of genetic networks using multi-objective hybrid spea2+ from microarray data*. in *Cognitive Informatics & Cognitive Computing (ICCI* CC), 2013 12th IEEE International Conference on*. 2013. IEEE.
- [25] Lee, W.-P. and Y.-T. Hsiao, *Inferring gene regulatory networks using a hybrid GA-PSO approach with numerical constraints and network decomposition*. Information Sciences, 2012. **188**: p. 80-99.
- [26] Rashedi, E., H. Nezamabadi-Pour, and S. Saryazdi, *BGSA: binary gravitational search algorithm*. Natural Computing, 2010. **9**(3): p. 727-745.
- [27] Rashedi, E., H. Nezamabadi-Pour, and S. Saryazdi, *GSA: a gravitational search algorithm*. Information sciences, 2009. **179**(13): p. 2232-2248.