DOR: 20.1001.1.2538161.1403.9.2.3.4

DOI: 10.22034/csj.2024.203193

تابستان ۱۴۰۳، دوره ۹، شماره ۲، پیاپی ۳۳

دریافت مقاله: ۱۴۰۳/۰۲/۱۵

پذیرش مقاله: ۱۴۰۳/۰۴/۰۹

نوع مقاله: پژوهشی

بیشینه‌سازی انتشار و کمینه‌سازی هزینه به صورت هم‌زمان در شبکه‌های اجتماعی مبتنی بر جدول درهم‌ساز

محسن قنبری قمصری

دانشکده برق و کامپیوتر- دانشگاه کاشان-ایران

پست الکترونیکی: mohsenghq@gmail.com

فرشته دهقانی*

استادیار، دانشکده برق و کامپیوتر- دانشگاه کاشان-ایران

پست الکترونیکی: fdehghni@kashanu.ac.ir

سید مهدی وحیدی پور

استادیار، دانشکده برق و کامپیوتر- دانشگاه کاشان-ایران

پست الکترونیکی: vahidipour@kashanu.ac.ir

چکیده

با توسعهٔ شبکه‌های اجتماعی، چالش‌های بیشتری در تحلیل آنها ایجاد شده است که نیازمند روش‌های دقیق‌تر و سریع‌تر هستند. یکی از آن چالش‌ها، یافتن افراد تأثیرگذار در شبکه‌های اجتماعی است که کاربردهای متنوعی مانند بازاریابی، انتشار اطلاعات و پیشگیری از بیماری‌ها دارد. بیشتر روش‌های پیشین این چالش را به عنوان یک مسئله تک هدفه دنبال می‌کردند؛ به حداکثر رساندن میزان انتشار در شبکه. در این راستا، تعداد مشخصی از افراد به عنوان افراد تاثیرگذار (یا اعضای بذر) انتخاب می‌شوند و میزان انتشار منجر از انتخاب آنها در شبکه ارزیابی می‌شود. در حالی که اهداف دیگری مانند هزینه انتخاب هر بذر نیز اهمیت ویژه‌ای دارد؛ به‌خصوص در امور تبلیغاتی و راهبردهای مدیریتی. در این مقاله حداکثر رساندن انتشار به‌عنوان یک مسئلهٔ تک هدفه و همچنین کمینه‌کردن هزینه انتخاب بذرها با توجه به حداکثر رساندن انتشار به عنوان یک مسئلهٔ دو هدفه در نظر گرفته شده است. نوآوری مقاله، استفاده از توابع درهم‌ساز در جهت افزایش سرعت اجرای چالش یافتن اعضای بذر است. در این راستا، الگورتیم H-GA (ترکیب الگوریتم ژنتیک و تابع درهم‌ساز) برای حل مسئلهٔ تک هدفه و الگوریتم H-NSGA-II (ترکیب روش NSGA-II و تابع درهم‌ساز) برای حل مسئلهٔ دو هدفه پیشنهاد می‌شود. در الگوریتم‌های پیشنهادی، استفاده از توابع درهم‌ساز باعث می‌شود تا از تکرار محاسبات ارزیابی جلوگیری شود. بدین‌ترتیب، ضمن حفظ دقت حل مساله یافتن اعضای بذر، بهبود قابل توجهی در زمان اجرا ایجاد می‌شود؛ بهبود زمان اجرا به طور متوسط ۲۱/۸٪ در تمامی آزمایش‌ها.

واژه‌های کلیدی: شبکه‌های پیچیده، بیشینه‌سازی انتشار، کمینه کردن هزینه انتخاب بذر، بهینه‌سازی چند هدفه، الگوریتم ژنتیک

* نویسندهٔ مسئول

۱- مقدمه

شـبکه‌های اجتماعی شـامل مجموعـه‌ای از افراد یا سـازمان‌ها و ارتباطات بین آن‌ها می‌باشـند. این شبکه‌ها دارای سـاختار نامنظـم و پیچیده‌ای هسـتند و ارتباطات در آن بر اسـاس نوع شـبکه می‌تواند به صورت دوستی، همکاری، روابط علمی، دنبال کردن و غیره باشـند. ماهیت شـبکه‌های اجتماعی به گونه‌ای اسـت که می‌توانند افراد در مدت زمان کوتاهی به یکدیگر متصل شـوند و سـبب شکل‌گیری ارتباطات جدید و در نتیجه تاثیرگذاری افراد بر یکدیگر شود. از طرف دیگر، حجم بزرگی از داده‌ها به دلیل گسـترش و انتشار این شبکه‌ها تولید می‌شود که تجزیه و تحلیل ایـن داده‌ها می‌تواند به افـراد در درک ویژگی‌های سـاختاری و رفتاری شـبکه‌های اجتماعی کمک کند و به برخی از مسـائل اقتصادی، اجتماعی و جامعه‌شـناختی پاسـخ دهد. از این‌رو، تحلیل شبکه‌های اجتماعی به عنوان یکـی از زیرمجموعه‌های علـوم کامپیوتر در سـال‌های اخیر محسوب می‌شـود. از موضوعات چالشی در تحلیل شـبکه‌های اجتماعی می‌توان به نمونه‌گیری شـبکه[1] [۱]، پارتیشن‌بندی شبکه[2] [۲] و تشخیص ناهنجاری[3] [۳] اشاره کرد.

یکی از موضوعات گسترده و پرکاربرد در زمینه تحلیل شبکه‌ها، بیشینه‌سازی انتشار[4] است. فرض کنید می‌خواهید تبلیغ یـک محصول را در یک شـبکه اجتماعـی به اطلاع حداکثر افراد حاضر در شبکه برسانید (انتشار حداکثری). مهمترین چالش انتخاب یک زیرمجموعه از افراد تأثیرگذار در شبکه است که فرایند انتشـار تبلیغ را شروع کنند. این افراد انتشـار را آغاز می‌کنند و بقیه افراد آن را گسـترش می‌دهند. برای گسترش انتشار روی شبکه می‌توان فرض کرد که این انتشار توسـط مدل مشخصی در شبکه پیش می‌رود؛ بر اسـاس مدل، انتشار خبر در شبکه شبیه‌سازی می‌شـود. با اسـتفاده از مجموعه اولیه افراد (یا مجموعه بذر[5]) و انتخاب یک مدل انتشار، میزان تأثیرگذاری مجموعه بذر انتخاب شـده ارزیابی می‌شود. بنابراین مسئلۀ اصلی در بیشینه‌سازی انتشار، انتخاب مجموعه‌ای از افراد (بذر) اسـت که میزان انتشار در شـبکه را به حداکثر برسانند و بالاترین تأثیر را در انتشار داشته باشند. در نظریۀ گراف، تأثیر یک معیار است که نشـان می‌دهد یک گره چه مقدار می‌تواند وضعیت دیگر گره‌های شبکه را تغییر دهد.

مطالعه بیشینه‌سازی انتشار معمولاً بر مبنای مدل‌های انتشـار و الگوریتم‌های شناسایی افراد تأثیرگذار صورت می‌گیرد. بنابراین، در صورتی که یک شرکت نیاز به تبلیغ یک محصول یا ایده جدید داشته باشد، یکی از راهکارهای موثرتر برای آغاز این فرآیند، شناسـایی گروه کوچکی از افراد تأثیرگذار مانند ستارگان سینما، ورزشکاران مشهور یا هنرمندان معروف به عنوان گروه اولیه است تا محصول بتواند از طریق تأثیر آن‌ها به گسترش بیشتری دست یابد. این در حالی است که شرکت‌های تبلیغاتی باید هزینه‌هایی را برای جذب این افراد به‌عنوان بذرهای اصلی پرداخت کنند؛ بدیهی است که هزینه هر فرد به ویژگی‌های تأثیرگذاری‌اش بسـتگی داشـته باشـد. از طرفی کاهش این هزینه برای شـرکت‌ها اهمیت دارد. بنابراین، روشی که هزینه حداقلی را برای انتخـاب گره‌های بذر به منظور به‌دسـت آوردن بیشترین تأثیر انجام دهد، یک مسئله عملی و مهم است [٤].

در این مقاله، ابتدا مسـئله بیشینه‌سـازی انتشـار به صورت یک مسئله بهینه‌سازی تک هدفه در نظر گرفته شده است. در ادامه با توجه به اهمیت رویکرد کاهش هزینه در راهبردهای مدیریتی، بیشینه‌سازی انتشار به همراه کاهش هزینه به‌عنوان یک مسئله بهینه‌سازی دو هدفه مورد توجه قرار گرفته است. همان‌گونه که مشخص است این دو هدف با یکدیگر متعارض[6] هسـتند و افزایـش یکی باعث کاهش دیگری می‌شود و برعکس. به منظور در نظر گرفتن هزینه گراف از درجه گره به عنوان پارامتر استفاده می‌شود.

برای حل مسئلۀ تک هدفه بیشینه‌سازی انتشار می‌توان

1-Network sampling
2-Network partitioning
3-Anomaly detection
4-Influence Maximization
5-Seed
6-Trade-off

از روش‌های جستجو مانند الگوریتم ژنتیک استفاده کرد. به‌طوری که در هر تکرار الگوریتم ژنتیک، مجموعه‌های مختلفی از گره‌های بذر تولید و ارزیابی شوند. بیشترین زمان مورد نیاز به ارزیابی باز می‌گردد که در آن گسترش انتشار توسط مدل انتشار شبیه‌سازی می‌شود تا میزان انتشار برای مجموعه بذر تخمین زده شود. نکته بسیار مهم آن است که مجموعه‌های بذر تکراری در طول تکرارهای الگوریتم ژنتیک تولید می‌شوند که ارزیابی مجدد آنها باعث اتلاف وقت و منابع محاسباتی خواهد شد؛ در حالی که می‌توان از محاسبه مجدد خودداری کرد. بنابراین در این مقاله، با استفاده از توابع درهم‌ساز[۷] و با رویکرد ذخیره‌سازی اطلاعات، نتایج ارزیابی مجموعه‌های بذر ذخیره می‌شود تا در صورت تولید مجدد این مجموعه بذر توسط تکرارهای بعدی الگوریتم مورد استفاده قرار گیرد. این روش پیشنهادی با نام (H-GA) معرفی شده است که در آن الگوریتم ژنتیک و تابع درهم‌ساز با یکدیگر همکاری می‌کنند تا سرعت اجرای جستجو افزایش یابد. آزمایش‌ها نشان می‌دهد که حل مسئلهٔ تک هدفه بیشینه‌سازی انتشار با H-GA به طور متوسط ۳۰/۶ ٪ سریع‌تر است.

همچنین ایده استفاده از تابع درهم‌ساز در بهینه‌سازی چند هدفه نیز در این مقاله پیشنهاد شده است. الگوریتم NSGA-II [۵] و توابع در هم ساز ترکیب شده و روش جدید H-NSGA-II پیشنهاد شده است. آزمایش‌ها نشان می‌دهد که روش پیشنهادی در حل مسئلهٔ بهینه‌سازی دو هدفه معرفی شده در این مقاله، سرعت بهتری دارد (به طور متوسط متوسط ۱۳ ٪ بهبود سرعت). با افزایش تعداد بذرها در الگوریتم H-NSGA-II سربار زمانی جستجو در جدول درهم‌ساز بیشتر می‌شود و همزمان تعداد رخدادهای یک کروموزوم (به علت حفظ تنوع در نسل‌های مختلف الگوریتم) کاهش می‌یابد. در نتیجه در مسئلهٔ دوهدفه نسبت به مسئلهٔ تک هدفه بهبود سرعت کمتری مشاهده شد.

ادامه این مقاله به شرح زیر سازماندهی شده است: بخش دوم شامل مفاهیم استفاده شده در این مقاله است. در بخش سوم، شرحی از کارهای مرتبط در زمینه بیشینه‌سازی انتشار در شبکه‌های اجتماعی بررسی می‌شود. روش پیشنهادی در بخش چهارم شرح داده شده است. در بخش پنجم، آزمایش‌های انجام شده قرار دارد. در نهایت در قسمت ششم نتایج گزارش شده است.

مفاهیم پایه

در این بخش بیشینه‌سازی انتشار و مدل انتشار "Independent Cascade" به عنوان یک مدل انتشار رایج مرور می‌شود. سپس دو الگوریتم ژنتیک و NSGA-II که در این مقاله استفاده شده‌اند بررسی می‌شوند.

بیشینه سازی انتشار

بیشینه‌سازی انتشار، یک مسئلهٔ بهینه‌سازی در حوزه گراف و شبکه‌های اجتماعی است که هدف آن تعیین مجموعه‌ای از گره‌های موثر در یک گراف است، به گونه‌ای که اگر انتشار از این گره‌ها شروع شود، در نهایت تعداد بیشینه‌ای از گره‌ها فعال شوند. در این مسئله، هر گره می‌تواند دو وضعیت فعال یا غیرفعال داشته باشد. گره‌های غیرفعال گره‌هایی هستند تحت تأثیر همسایه‌های خود قرار نگرفته‌اند، در حالی که گره‌های فعال تحت تأثیر یکی از همسایه‌های خود تغییر وضعیت داده‌اند. به طور خاص، هدف این مسئله یافتن یک زیرمجموعه کوچک از گره‌ها با اندازه k از یک شبکه است، به نحوی که تعداد کل گره‌های تحت تأثیر این زیرمجموعه بیشینه شود.

اگر گراف $G(V,E)$ را داشته باشیم که V مجموعه گره‌های گراف و E مجموعه یال‌ها باشد، هدف پیدا کردن مجموعه S است که $S \subseteq V$ و $|S| = k$ باشد و k تعداد گره‌هایی است که به دنبال آن هستیم تا $S = \underset{S}{\operatorname{argmax}} f(S)$ برقرار شود و f یک مدل انتشار است. مجموعه S را مجموعه گره‌های بذر[۸] هم می‌گویند. در روند یافتن گره‌های مدنظر ابتدا مجموعه بذر فعال می‌شوند و طبق مدل انتشار

7-Hash based-Genetic Algorithm

8-Seed

تلاش می‌کنند تا گره‌های همسایه خود را فعال کنند، هر گرهی که فعال شد نیز می‌تواند همسایه‌های خود را فعال کند و این کار تا زمانی که گره جدیدی فعال نشود ادامه پیدا می‌کند. در ادامه یک مدل انتشار رایج شرح داده می‌شود.

مدل Independent Cascade

در حوزه نظریهٔ گراف، از مدل‌های انتشار به منظور مدل‌سازی روند انتشار اطلاعات یا تأثیرگذاری گره‌ها از طریق شبکه استفاده می‌شود. یکی از این مدل‌ها که با عنوان Independent Cascade یا IC شناخته می‌شود، در شکل ۱ توضیح داده شده است. ورودی این مدل شامل گراف G، مجموعه بذر A_0، تعداد تکرارهای شبیه‌سازی مونت-کارلو (MC) و احتمال‌های تأثیر گره i بر گره j که با λ_{ij} نشان داده می‌شوند، می‌باشد. خروجی این مدل، میانگین تعداد گره‌های فعال شده است.

در این مدل، شبکه اجتماعی به عنوان یک گراف مدل‌سازی می‌شود که در آن گره‌ها نمایانگر افراد و یال‌ها نمایانگر روابط آنها می‌باشند. برای شروع، یک مجموعه بذر A_0 (شامل k گره فعال) در نظر گرفته می‌شود. مجموعه A نیز نمایانگر گره‌هایی است که تا به حال فعال شده‌اند و در ابتدا با A_0 برابر است. هر گره فعال شده تنها یک بار مجاز است تا گره‌های مجاور خود را فعال سازد. این گره‌ها در مجموعه B نگهداری می‌شوند و در ابتدای هر بار اجرای حلقه به‌روز رسانی می‌شود. فعال‌سازی با احتمال از پیش تعیین شده λ_{ij} اتفاق می‌افتد، که یا از دانش قبلی استخراج می‌شود یا به صورت یکسان برای همه یال‌ها در نظر گرفته می‌شود. گره‌های فعال شده جدید در هر مرحله در مجموعه $nextB$ نگهداری می‌شوند و به دلیل این که این گره‌های جدید مجاز به فعال‌سازی همسایه‌های خود هستند، به مجموعه B اضافه می‌شوند. سپس، تمام گره‌های فعال شده جدید به مجموعه A اضافه می‌شوند. این مراحل تکرار می‌شوند تا زمانی که مجموعه B خالی نباشد، به این معنا که گره‌های فعال شده جدیدی وجود داشته باشد. به دلیل تصادفی بودن این روش، الگوریتم بر اساس تعداد تکرارهای MC که توسط کاربر مشخص می‌شود، تکرار می‌شود و در نهایت میانگین تعداد گره‌های فعال شده برگردانده می‌شود.

```
1   spread = 0
2   repeat:
3       τ ← False    τ: Has The Propagation Ended?
4       A ← A0       A: Active Nodes
5       B ← A        B: the set of nodes activated
6       while not τ do
7           nextB ← ∅;
8           for each n ∈ B do
9               for each neighbour m of n , where m ∉ A, do
10                  with probability λij, add m to nextB
11          B ←nextB
12          A ←A ∪ B
13          if B is empty then
14              τ← True
15      Spread. Append(Len(A))
16  until MC
17  return mean(Spread)
```

شکل ۱: الگوریتم Independent cascade model

الگوریتم ژنتیک

الگوریتم ژنتیک از مفاهیم و تکنیک‌های زیست‌شناسی مانند وراثت، جهش و انتقال ویژگی‌ها در نسل‌های بعد برای بهینه‌سازی و یا پیش‌بینی الگوها استفاده می‌کند. این الگوریتم با استفاده از یک جمعیت از کروموزوم‌ها، که هر کدام یک پاسخ ممکن به مسئله را نمایان می‌کنند، به دنبال بهینه‌سازی یا پیدا کردن الگوی بهتر می‌گردد.

در ابتدا، چندین کروموزوم تصادفی به عنوان جمعیت اولیه تولید می‌شود. سپس با استفاده از عملگرهای مختلف الگوریتم ژنتیک، این کروموزوم‌ها تغییر می‌کنند تا به پاسخ‌های بهتر و بهینه‌تری برای مسئله دست پیدا کند. ورودی این الگوریتم یک مسئلهٔ بهینه‌سازی است و خروجی آن بهترین پاسخ یافت شده به این مسئله است.

به منظور ساخت جمعیت اولیه، ابتدا تابع CreatInitialPopulation فراخوانی می‌شود تا چندین کروموزوم تصادفی تولید شود. سپس با فراخوانی تابع () Evaluate، هر کروموزوم ارزیابی و برازندگی[9] آن محاسبه

9-Fitness

می‌شـــود تا میـــزان کارایـــی و بهترینی آن تعیین شـــود. در مرحلـــه بعد، با تعداد نســـل‌های موردنظر که به عنوان Number of Generation مشخص می‌شود، جمعیت جدید از روی جمعیـــت قبلی با اســـتفاده از عملگرهای الگوریتم ژنتیک ساخته می‌شود.

الگوریتم ژنتیـــک از تکنیک‌های زیست‌شناســـی مانند وراثت، جهش زیست‌شناســـی، اصـــول انتخابی داروین و انتقال ویژگی در نســـل‌های بعد بـــرای یافتن فرمول بهینه جهت پیش‌بینی یا تطبیق الگو اســـتفاده می‌شود. الگوریتم ژنتیک در شکل ۲ نشـــان داده شده است. به طورکلی ابتدا چند پاســـخ تصادفی بـــا نام کروموزوم[10] تولید می‌شـــود ســـپس با اســـتفاده از عملگرهای الگوریتم ژنتیک آن‌ها را تغییر می‌دهیم تا به پاسخ‌های بهتری تبدیل شوند. ورودی الگوریتم یک مســـئلهٔ بهینه‌ســـازی و خروجی آن بهترین پاسخ پیدا شده است. ابتدا جمعیت اولیه‌ای از کروموزوم‌ها به‌صـــورت تصادفی ســـاخته می‌شـــود (فراخوانـــی تابع CreatInitialPopulation()). جمعیت اولیه با فراخوانی تابع Evaluate() ارزیابی می‌شـــود (یا به عبارتی fitness آن‌ها محاسبه می‌شـــود)؛ میزان خوب بودن آن تعیین می‌شود. ســـپس به تعـــداد نســـل Number of Generation که به دلخواه تعیین می‌شود، جمعیت جدید از روی جمعیت قبلی ساخته می‌شود.

در فرآیند ســـاخت جمعیت جدیـــد در الگوریتم ژنتیک، مراحـــل ۵ تا ۱۰ شـــکل ۲ انجـــام می‌شـــوند. در مرحله ۵، کروموزوم‌های والدین انتخاب می‌شوند. سپس در مرحله ٦، با اســـتفاده از ترکیب دو والد (Crossover())، دو فرزند جدید برای جمعیت بعدی ایجاد می‌شـــوند. با اعمال عملگر جهش، برخی از فرزنـــدان جدید تغییرات اندکی می‌پذیرند. فرزنـــدان جدید در جمعیت جدید قـــرار می‌گیرند و دوباره ارزیابی می‌شوند. در نهایت با توجه به اصل نخبه‌گرایی در الگوریتم ژنتیک، کروموزوم‌هایی که در نسل قبلی عملکرد خوبی داشته‌اند، بدون هیچ تغییری در جمعیت جدید حفظ می‌شوند (مرحله ۱۱).

10-Chromosome

```
1   InitializationPopulation ()
2       CreateInitialPopulation ()
3       Evaluate ()
4   repeat
5       Select Parents ()
6       Crossover ()
7       Mutation ()
8       Replacement ()
9       Evaluate ()
10      Elitism ()
11  until Number of Generations
12  return Best Solution
```

شکل ۲: الگوریتم ژنتیک

الگوریتم NSGA-II

در مسایل بهینه‌سازی چندهدفه، در اغلب موارد به علت متعارض بودن یا غیر قابل مقایســـه بودن اهداف، یک راه حل تکی برای مســـائل چند هدفه به گونه‌ای که برای تمام توابع هدف بهینه باشد، وجود ندارد. از این‌رو مفهوم نقاط بهینه پارتو[11] و مجموعه پارتو[12] تعریف می‌شود.

به مجموعه تمام بردارهای هدف بهینه پارتو، مجموعه پارتو گفته می‌شـــود. با توجه به تعریـــف پارتو، می‌توان جواب‌های شـــدنی مســـئله چند هدفه را بر حسب مغلوب شـــدن یا نشـــدن، با مجموعـــه از جواب‌هـــا از بهترین تا بدترین مجموعه پاسخ دســـته‌بندی کرد. برای این منظور در ادامه مفهوم پاسخ نامغلوب[13] توضیح داده می‌شود. اگر $x^1 \in X$ و $x^2 \in X$ (X مجموعه جواب‌های شدنی است) دو راه حل برای مســـئلهٔ چند هدفه باشند، آن‌گاه x^1 بر x^2 غلبه دارد، اگر در همه توابع هدف مقدار بهتر و یا مساوی در x^1 داشته باشند. به همین ترتیب، در صورتی که مقدار تابع هـــدف در x^1 و x^2 در تمام اهداف نســـبت به یکدیگر برتری نداشته باشند، دو پاسخ نامغلوب بر یکدیگر هستند.

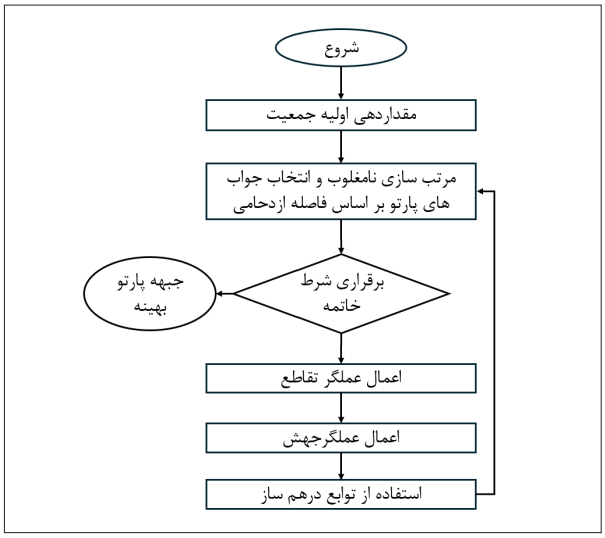
NSGA-II یک روش قدرتمند برای حل مسائل بهینه‌سازی چند هدفه اســـت. این روش از الگوریتم ژنتیک برای تولید فرزندان نســـل بعدی استفاده می‌کند. ســـپس از الگوریتم مرتب‌ســـازی نامغلوب و فاصلـــه ازدحامی بـــرای یافتن پاسخ‌های دور بعدی الگوریتم اســـتفاده می‌کند. الگوریتم مرتب‌سازی نامغلوب به منظور رتبه‌دهی به تمام جواب‌های

11-Optimal Pareto front
12-Pareto set
13-Non-dominated solutions

مجموعه شدنی[14] پیدا شده در مسائل بهینه‌سازی چند هدفه از خوب تا بد ارائه شده است.

در این روش پاسخ‌ها به منظور رتبه‌بندی دو به دو با هم مقایسه می‌شوند. به گونه‌ای که عضو مغلوب به مجموعه پاسخ‌های غالب اضافه می‌شود و به تعداد بار مغلوب شدن آن‌ها یکی اضافه می‌گردد. بنابراین بعد از این مقایسه، پاسخ‌هایی که توسط هیچ پاسخی مغلوب نشده‌اند، پاسخ‌های نامغلوب و یا نقاط بهینه جبهه پارتو[15] با رتبه یک هستند. با حذف جواب‌های رتبه یک از مجموعه جواب‌ها و کم کردن تعداد بار مغلوب شدن از عناصری که توسط آن‌ها غلبه شده‌اند، پاسخ‌های با رتبه دو حاصل می‌شوند. این کار تا زمانی ادامه پیدا کند که هیچ پاسخ شدنی باقی نماند.

پس از مرتب‌سازی بر اساس نامغلوب بودن، پاسخ‌هایی با رتبه یکسان در هر رتبه به هم دیگر غلبه نمی‌کنند و در نتیجه با یکدیگر قابل مقایسه نیستند. به همین علت به یک روش مقایسه دیگری برای انتخاب از میان پاسخ‌هایی با رتبه یکسان نیاز است. یکی از این ملاک‌ها، فاصله ازدحامی است که در الگوریتم NSGA-II نیز استفاده شده است. در این روش، پاسخ‌هایی که در مناطق کم ازدحام‌تری قرار دارند و در نتیجه فاصله ازدحامی بیشتری دارند، بر دیگر پاسخ‌ها ارجحیت دارند تا تنوع در پاسخ‌ها حفظ شود. برای محاسبه فاصله ازدحامی، پاسخ‌ها در هر رتبه بر اساس هر تابع هدف مرتب می‌شوند. سپس به ازای هر تابع هدف، تفاوت مقدار تابع هدف هر نقطه نسبت به دو همسایه چپ و راست آن محاسبه و با هم جمع می‌شوند. لازم به ذکر است که پاسخ‌های اول و آخر به علت مشخص کردن محدوده جواب، همیشه پاسخ‌های مطلوبی هستند و در مجموعه پاسخ‌های انتخاب شده همیشه وجود دارند. در نهایت پس از اجرای الگوریتم، در هر رتبه پاسخ‌ها بر اساس فاصله ازدحامی مرتب می‌شوند. این روند در شکل ۳ نشان داده شده است.



شکل ۳- الگوریتم NSGA-II

در روش NSGA-II حفظ جمعیت متنوع بسیار مهم است. این امر کاوش در مناطق مختلف فضای جستجو را تضمین می‌کند و از همگرایی زودرس به یک بهینه محلی جلوگیری می‌کند. طراحی این الگوریتم به گونه است که این تنوع جمعیت را حفظ می‌کند و احتمال تکرار یک کروموزوم در آن کمتر است.

۲- کارهای مرتبط

در این قسمت در ابتدا کارهای پیشین مربوط به بیشینه‌سازی انتشار به صورت تک هدفه و در ادامه به بررسی تحقیقات انجام شده به صورت چند هدفه (با تمرکز بر روی کاهش هزینه و افزایش انتشار) بررسی می‌شود.

مسئله بیشینه‌سازی انتشار اولین بار در [۶, ۷] به عنوان یک مسئله برای یافتن مجموعه‌ای از مشتریان مفید در بازاریابی مطرح شد. برای مدل‌سازی این موضوع، آنها از حوزه‌های تصادفی مارکوف استفاده کردند. اگرچه روش ارائه شده می‌تواند مشکل بیشینه‌سازی انتشار را حل کند، اما نمی‌تواند تأثیر کاربران بر یکدیگر را به وضوح شناسایی کند. مقالات [۸, ۹] مسئله بیشینه‌سازی انتشار را به عنوان یک مسئله بهینه‌سازی گسسته برای اولین

14-Feasible set
15-Optimal Pareto front

بار مـــدل کردند. آنها ثابت کردند که به طورکلی، بیشـــینه ســـازی انتشار جزو مسائل NP-Hard طبقه بندی می‌شود، اما آنها دو مدل انتشـــار به نام‌هـــای Independent cas-cade mode و Linear threshold model و همچنیـــن یک الگوریتم حریصانه با حد تقریب $1-\frac{1}{e-\varepsilon}$ پیشـــنهاد دادند کـــه e و ε به ترتیب عدد اویلر و دقت تخمین هســـتند. این الگوریتم حریصانه هر چند دقت خوبی دارد اما مقیاس‌پذیر نیست. بعد از آن چند راهکار برای بهبود این الگوریتم ارائه شد که به عنوان مثال می‌توان از CELF[۱۰]،CELF++[۱۱]، TIM[۱۲]، BCT[۱۳]، RIS [۱٤] و MixedGreedy [۱٥] نام برد.

در برخـــی از کارهای قبلـــی، از روش‌های اکتشـــافی برای حل مســـئلهٔ بیشینه‌سازی انتشار استفاده شد. برخی از این مدل‌ها بر اســـاس معیارهـــای مرکزیت مانند درجه، گره‌های موثر را تعیین می‌کردند[۱٦, ۱۷]؛ آنها مســـتقل از مدل انتشـــار بودند. گروه دیگری از روش‌های اکتشافی به طور خاص برای یک مدل انتشار طراحی شده‌اند و کارایی مناسبی را در مدل پیشـــنهادی نشان دادند به عنوان مثال IRI[۱۸] برای مدل انتشار Independent cascade mode به خوبی کار می‌کند. خوبی روش‌های اکتشافی این بود که هم پیچیدگی زمانی را بهبود و هم مقیاس‌پذیری را افزایش دادند.

در [۱۹] از الگوریتـــم ژنتیک ســـاده برای حل مســـئلهٔ بیشینه‌سازی انتشار اســـتفاده شده و نشان داده است که با استفاده از یک عملگر ساده ژنتیکی، می‌توان یک راه حل تقریبـــی برای تأثیرگذاری در زمان اجرای بهتر نســـبت به استفاده از الگوریتم‌های حریصانه قبلی به دست آورد. در [۲۰]، با گســـترش الگوریتم ژنتیک قبلـــی، یک الگوریتم فرا ابتکاری برای انتخاب یک زیرمجموعه با طول ثابت پیشنهاد کردند. برای این پیشـــنهاد از اطلاعات مربوط به ســـاختار شبکه استفاده شده است. در [۲۱] از الگوریتم ژنتیک برای یافتـــن راه‌حل و یک الگوریتم حریصانه برای بهبود راه‌حل به‌دســـت آمده از طریق الگوریتم ژنتیک به منظور افزایش کارایی مسئلهٔ بیشینه‌سازی انتشـــار استفاده می‌شود. در [٤] بـــا هدف افزایش انتشـــار و کاهـــش هزینه به صورت هم‌زمان به مدل‌کردن شـــبکه‌های اجتماعی پرداخته است تا ویژگی‌های شـــبکه‌های دنیای واقعی را بهتر به تصویر بکشد. برای این منظور الگوریتم بهینه‌سازی ازدحام ذرات گسسته چند هدفه پیشنهاد شده است. این الگوریتم می‌تواند هـــم هزینه فردی و هم تأثیر فردی را در نظر بگیرد. دراین تحقیق از درجه گره‌ها به عنوان تشـــخیص معیاری برای هزینه استفاده شده است.

در تحقیق [۲۲] افزایش انتشـــار به همراه کاهش هزینه در شـــبکه‌های اجتماعی و به صـــورت دو هدفه و همزمان در نظر گرفته شـــده است. به منظور کاهش هزینه انتشار، کاهـــش تعداد دانه‌های بذر در نظر گرفته اســـت. برای این منظور، مجموعه‌ای از گره‌های بـــذر با اندازه‌های مختلف را شناسایی می‌شوند، به نحوی که بهترین تأثیر را داشته باشند. در نهایت با استفاده از دو مدل انتشار تأثیر متفاوت، مورد بررسی قرار گرفته شده است. نتایج نشان می‌دهد که رویکرد پیشنهادی تقریباً همیشه می‌تواند بهتر از اکتشافی عمل می‌کند. بیشینه‌سازی انتشار به همراه در نظر گرفتن هزینه در [۲۳] مدل شد و با استفاده از الگوریتم MODEA[۲۱] جواب‌های نزدیک به بهینه آن مورد بررســـی قرار گرفت. در این تحقیق هزینه فعال‌ســـازی گره بذر بر اساس تأثیر سراسری گره در نظر گرفته شد.

در [۲٤] افزایـــش انتشـــار به همراه کاهـــش هزینه در شـــبکه‌های اجتماعی و به صـــورت دو هدفه و همزمان در نظر گرفته شده است. هزینه در این مقاله برای همه گره‌های یکسان در گرفته شده است. در ادامه یک روش بهینه‌سازی تکاملی دیفرانســـیل چندهدفه گسسته با نام DMODE[۲۲] به همـــراه عملگرهای جهش، تقاطع و انتخاب به منظور بهبود

16-Cost Effective Lazy Forward
17-Two-phase Influence Maximization
18-Billion-scale Cost-award Targeted
19-Reverse Influence Sampling
20-Influence ranking influence estimation
21-Multi-objective Differential Evolutionary Algorithm
22-Discrete Multi-objective Differential Evolution optimization

اکتشاف[۲۳] و بهره‌برداری[۲۴] ارائه شده است.

در [۲۵]، بـه منظـور برقـراری توازن میان انتشـار اطلاعات و هزینه انتشار، یک الگوریتم جستجوی کلاغ چند هدفه (MOCSA[۲۵]) برای حل مسـئلهٔ بهینه‌سازی چندهدفه ارائه شده است. به منظور بهبود همگرایی الگوریتم، تنظیم پارامترها بر اساس راهبرد کنترل پویا و راهبرد راه رفتن تصادفی بر اساس سیاه‌چاله‌ها اتخاذ شده است و در نهایت شـش شـبکه اجتماعی واقعی برای آزمایش انتخاب و در مقایسه با سایر الگوریتم‌های پیشرفته مورد تجزیه‌وتحلیل قرار گرفته‌اند.

تمرکـز روش‌های قبلی اکثرا در یافتن بهترین جواب از نظر اهداف مسـئله مثل انتشار و هزینه است اما به مسئلهٔ زمـان اجرا کمتر پرداخته شـده و در مواردی که بهبودی حاصل شده در ازای از دست رفتن جواب‌های خوب است. در این مقاله روشـی ارائه شـده که زمان اجرا را بدون از دسـت دادن جواب‌های خوب به مقدار قابل توجهی کاهش دهد و در ادامه نتایج نشان داده که این روش در مسائل تک هدفه و چندهدفه در حالات مختلف کاربرد دارد.

۳- روش پیشنهادی

در این بخش روش پیشنهادی شرح داده می‌شود. برای این منظور در زیر بخش اول مدل مسـئلهٔ بهینه‌سازی تک هدفه (بیشینه کردن انتشار) و در ادامه الگوریتم پیشنهادی به منظور حل آن شرح داده می‌شود. همچنین در زیر بخش بعدی مسـئلهٔ بهینه‌سازی دو هدفه (بیشینه‌کردن انتشار و کمینه کردن هزینه) شرح داده می‌شود و در ادامه الگوریتم پیشنهادی به منظور حل شرح داده می‌شود.

۳-۱- بیشینه‌کردن انتشار (مسئله بهینه‌سازی تک هدفه)

فرض کنید گراف $G=(V,E)$ داده شـده اسـت که V مجموعه گره‌ها و E مجموعه یال‌هاست. همچنین k تعداد گره‌های بذر اسـت. برای حل مسئلهٔ بیشینه‌سازی انتشار از مدل IC استفاده خواهد شـد. رابطه شماره (۱) میانگین انتشـار را در گراف توسـط مجموعه بذر S توسط $f_1(S)$ نشان داده شده است.

$$f_1(S)=\frac{1}{M}\sum_{i=1}^{M} Influence_i(s) \qquad (۱)$$

در ایـن رابطـه $Influence_i(S)$ تعداد گره‌های فعال شـده توسـط مجموعه بذر S را در بار iام نشـان می‌دهد و M تعداد شبیه‌سازی‌های مونت-کارلو است.

برای حل این مسـئله بهینه‌سازی تک هدفه از الگوریتم ژنتیک اسـتفاده شـده اسـت. به این صورت که ابتدا چند کروموزوم مانند شکل ٤ تشـکیل می‌شود که هر ژن عدد یک گره مشـخص را تعیین می‌کنـد و تعداد ژن‌ها همان k اسـت. در شـکل ٤، مجموعه گره‌های ۱، ۳، ۷، ۸ و ۱۱ به عنوان ٥ گره بذر پیشنهاد شده است ($k=5$).

شکل ۴: نمونه یک کروموزوم ساخته شده

سـپس تمام کروموزوم‌ها باید ارزیابی شـوند که این کار توسط مدل انتشار IC صورت می‌گیرد. به این صورت که گره‌های موجود در کروموزوم در گراف فعال شـده و ارزش آن کروموزوم که همان تعداد گره‌های فعال شـده است با یک عدد باز می‌گردد. هر کروموزومی که بیشترین گره را فعال کرده باشـد شـانس بالاتری برای انتخاب به عنـوان والـد خواهد داشـت. همچنیـن از نخبه‌گرایی نیز استفاده شده تا دقت الگوریتم بالاتر برود.

در الگوریتم پیشـنهادی H-GA از عملگرهای تقاطع و جهش الگوریتم ژنتیک اسـتاندارد اسـتفاده شده است و پیشـنهاد این روش برای بهبود در زمان اجرا با استفاده از تابع درهم‌سـاز و حذف گره‌های کم اهمیت اسـت. به این صـورت که در هر بار اجـرای الگوریتم IC برای یک کروموزوم مقدار نهایی محاسـبه شده به همراه آرایه‌ای شـامل همان کروموزوم در یک جدول ذخیره می‌شـود و در مراحـل بعدی اگـر کروموزوم تکراری وارد شـد

23-Explore
24-Exploit
25-Multi-objective Crow Search Algorithm

بـــه جای اجـــرای کامل تابـــع IC فقط مقـــدار آن از جدول بازگردانده می‌شـــود. برای این کار از جدول در هم‌ســـاز با اندازه $3 * 10^7$ سطر شامل جفت‌های کلید–مقدار برای ذخیره‌ســـازی میانگین ارزیابی هر کروموزوم اســـتفاده می‌شود. هر کلید توســـط تابع درهم‌ساز به یک کد تبدیل می‌شود و در جایگاه همان کد در جدول قرار می‌گیرد در این حالت موقع جســـتجو نیاز نیســـت کل جدول جستجو شـــود و تنها کافی اســـت کلید را تبدیل به آدرس کرده و همان یک خانه بررســـی شـــود و اگر مقداری در آن بود برگردانده شود.

۲-۳- بیشینه کردن انتشار و کمینه کردن هزینه (مسئله بهینه‌سازی چند هدفه)

به منظور مدل‌سازی مسئله به صورت دو هدفه، علاوه بر هدف رابطه (۱)، هدف کمینه کردن هزینه نیز به صورت رابطه شـــماره (۲) در نظر گرفته شـــده اســـت. به منظور کمینه‌کردن هزینه لازم است تا اطلاعاتی در مورد گره‌های گراف وجود داشته باشد که تعیین کند برای اضافه کردن یک گره به مجموعه بذر چه هزینه‌ای باید پرداخته شود. به عنوان مثال، هزینه استخدام یک انسان برای بازاریابی یک محصول یا ایده جدید همیشـــه به میـــزان تأثیر او بر افراد بســـتگی دارد، که می‌تواند به عنوان درجه گره متناظر در نظر گرفته شـــود. یا حقوق کارکنان بـــا توجه به رتبه آنها پرداخت می‌شـــود. اگر افراد در یک رتبه باشند، اجرت آنها تقریباً یکسان خواهد بود. در غیر این صورت، دستمزد آنها می‌توانـــد کاملا متفاوت باشـــد [٤]. در این تحقیق از رابطه (۲) مشابه مرجع [٤] از درجه هر گره به این منظور استفاده شـــده است، با این فرض که گره‌های با درجه بالاتر هزینه بیشتری به منظور نشر اطلاعات دریافت خواهند کرد.

در رابطه بالا، $f_2(S)$ هزینه کلی فعال‌ســـازی تمام گره از مجموعه بذر می‌باشـــد و C_i هزینـــه انتخاب گره iام به عنوان گره بذر اســـت. در ایـــن رابطه k تعداد گره‌های بذر است و d_i درجه گره i ام است. پارامترهای ثابت $m,\ s,\ r$ دلخواه هستند و توســـط کاربر تعیین می‌شود که r برای تقسیم هزینه استفاده می‌شود (برای جلوگیری از تولید عدد بـــزرگ)، m و s به ترتیب بـــرای اندازه‌گیری تفاوت هزینه برای گره‌های مشـــابه و متفاوت استفاده می‌شوند. در این تحقیق مقادیر پارامترها برابر با $m = 1.1,\ s = 1.1$ و $r = 10$ انتخاب شده‌اند [۲٦, ۲۷].

در نهایت مســـئلهٔ بهینه‌سازی با دو هدف بیشینه‌کردن انتشـــار و کمینه‌کردن هزینه به صورت رابطه شماره (۳) بیان می‌شود.

$$\min\ (f_1(S), -f_2(S)) \quad (۳)$$
$$subject\ to\ |S| = k$$

در رابطـــه بـــالا، تابع هدف $f_2(S)$ بـــا علامت منفی در رابطه اســـتفاده شده است، تا با رابطه کمینه‌کردن سازگار شـــود. دو مزیت در نظر گرفتن مسئله به صورت دو هدفه دراین بخش نســـبت به تک هدفه ارائه شده بخش قبلی در این اســـت که راه‌حل‌های ارائه شده در مسئلهٔ بهینه‌سازی چنـــد هدفه می‌تواند طیف وســـیعی از انتخاب‌هـــا را برای تصمیم‌گیرندگان بر اساس موقعیت‌هایشان فراهم کند. دوم این است که مســـئلهٔ دوهدفه هم انتشار و هم هزینه را در نظر می‌گیرد، که می‌تواند به وضوح ویژگی‌های شبکه‌های دنیای واقعی را نسبت به تک هدفه نشان دهد.

در الگوریتـــم پیشـــنهادی H-NSGA-II از الگوریتـــم NSGA-II به عنوان پایه اســـتفاده شده است و برای بهبود در زمـــان اجرا مانند H-GA از تابع درهم‌ســـاز اســـتفاده می‌شـــود، به ایـــن ترتیب که این بار در هـــر دو تابع هدف مقادیر محاســـبه شده به صورت کلید–مقدار در دو جدول متفاوت ذخیره می‌شوند و درصورت تکرار یک کروموزوم در مراحـــل بعدی، مقـــدار آن از هر دو جـــدول فراخوانی می‌شـــود. از آنجایی که ماهیت مسئله به گونه‌ای است که تبدیل یک کروموزوم به کد یکتا در همه توابع هدف یکسان اســـت، این عمل فقط یک بار به ازای هر کروموزوم جدید انجام می‌شـــود و برای تمام توابع هدف استفاده می‌شود.

از آنجا که جدول‌ها بر اســـاس آدرس‌دهی ساخته شده‌اند جســـتجوی یک کلید از مرتبه زمانی $O(1)$ است و سربار زمانی اضافه ایجاد نمی‌کند.

۴- آزمایش‌ها

در ایـــن بخش ســـه آزمایش بـــرای بررســـی کارایی الگوریتم NSGA و مقایسه آن با الگوریتم ژنتیک در مسئلهٔ تک هدفه بررسی شده است. آزمایش اول نتایج اجرا برای بیشینه‌ســـازی انتشار در مســـئلهٔ تک هدفه است. آزمایش دوم برای تحلیل جواب‌های پارتو به دست آمده و مقایسه آن با نتایج آزمایش اول می‌باشد. در نهایت آزمایش سوم برای تاثیر جدول درهم‌سازی در زمان اجرا در ادامه آمده است. پارامترهای استفاده شده در همه آزمایش‌ها به شرح زیر است:

تابع IC: مقدار $\lambda_{ij} = 0.1$ و $MC = 50$ است.

الگوریتم ژنتیک: طبق جدول(۱)

الگوریتم NSGA-II: پارامتر اضافه ندارد و پارامترهای الگوریتم ژنتیک آن طبق جدول(۱) است.

گراف اســـتفاده شده شـــامل ۱۱۳۳ گره و ۱۰۹۰۲ یال است.

نتایج دو روش چندهدفهٔ دیگر[۲۸, ۲۹] نیز برای مقایسه با کار فعلی در شـــرایط مشـــابه جهت مقایســـه بهتر آمده اســـت. نویسندگان در [۲۸] با اســـتفاده از هوشمندسازی توابع تولید جمعیت اولیه و جهش نشان دادند که می‌توانند در مســـئلهٔ چندهدفه موفق عمل کننـــد، در ادامه این روش بـــا نـــام hyper-aware آمده اســـت. همچنیـــن در [۲۹] از IM–balanced اســـتفاده کرده و مســـئلهٔ چندهدفه را حل می‌کند و در نهایت از بین جبههٔ پارتو یک جواب را به عنوان بهتریـــن برمی‌گرداند که در آزمایش‌های این قســـمت فقط جبهه پارتو به‌دســـت آمده با این روش را مقایسه خواهیم کرد.

جدول ۱. پارامترهای الگوریتم ژنتیک

پارامتر	مقدار
جمعیت اولیه	۱۰۰
تعداد نسل‌ها	2500
احتمال جهش	۰,۲
نوع تقاطع	تک نقطه‌ای
نسبت Elitism	۰,۰۲

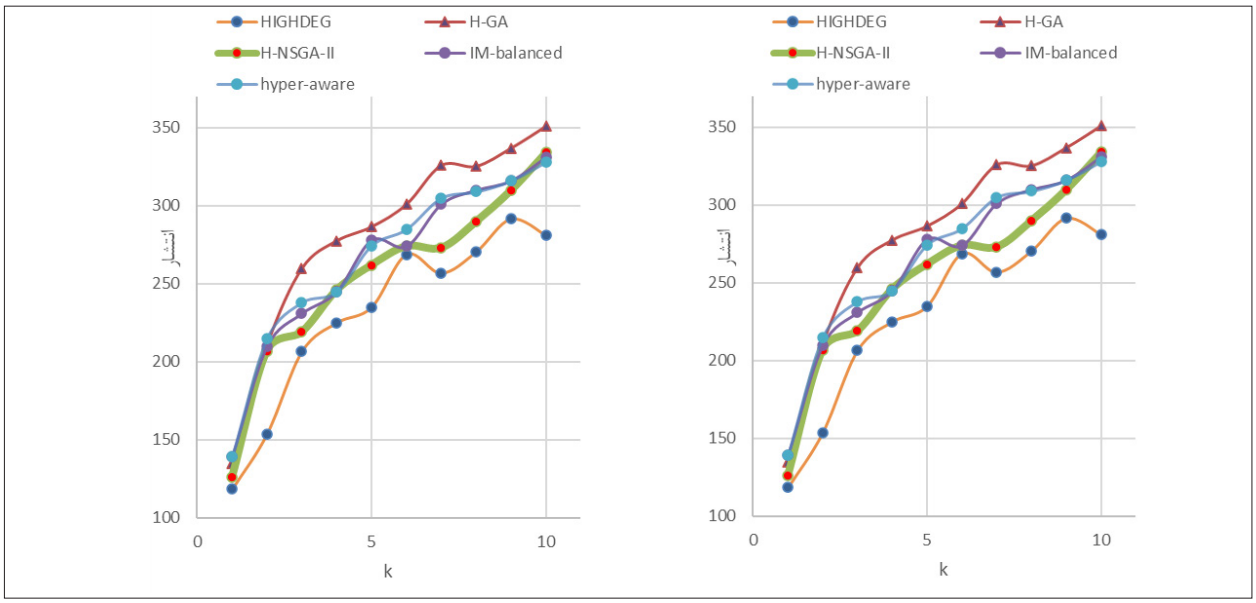
آزمایش اول: بیشینه‌سازی انتشار تک هدفه

در این آزمایش روش پیشنهادی با ٤ روش دیگر از جهت حداکثر میزان انتشار مقایسه شده است که HIGHDEG[۱۷] تنها گره‌های با بیشترین درجه را انتخاب می‌کند.

انتظار می‌رود با افزایش k میزان انتشـــار نیز بیشـــینه شود یعنی نمودار اکیدا صعودی باشد. با این که در شکل ٥ به طورکلی این افزایش قابل مشاهده است در برخی موارد استثنا هم وجود دارد. روش H-GA به علت تک‌هدفه بودن و در نظر گرفتن بیشینه کردن انتشار به تنهایی بهترین میزان انتشار را نسبت به بقیه روش‌ها دارد. الگوریتم H-NSGA–II در چندین k به خصوص برای $k = 10$ عملکردی مشـــابه hyper-aware و IM-balanced دارد و در بقیه kها اختلاف اندکی را دارا است. در هر حال با توجه به روش ارائه شده در این مقاله، می‌توان جدول درهم‌سازی را به الگوریتم‌های حل مســـائل بهینه‌سازی (هم تک هدفه و هم چند هدفه)، به منظور بهبود سرعت اجرا اضافه نمود.

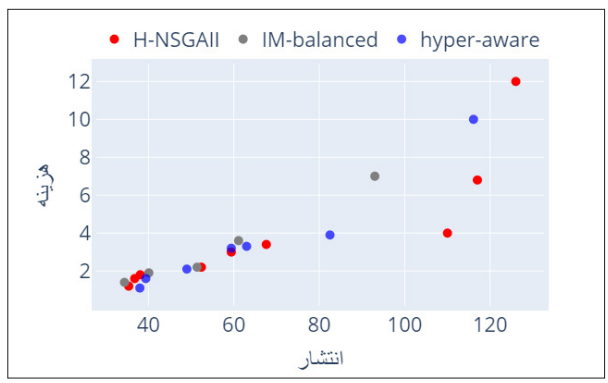
آزمایش دوم: جواب‌های جبهه پارتو در مسئله دو هدفه

این آزمایش جواب‌های به دست آمده که حاصل اجرای روش پیشنهادی و دو روش IM-balanced و hyper-aware برای $k = 1,2,3,4,5$ اســـت را نشـــان می‌دهـــد. نتایج از شـــکل‌های (٦ تا ۱۰) نشـــان‌دهنده کارایی الگوریتم است و همچنین تناقض میان افزایش انتشار و کاهش هزینه نشان داده شده است. از نظر مفهومی در دنیای واقعی نیز نتایج صحیح اســـت، زیرا در همه موارد انتشار بیشتر با هزینهٔ

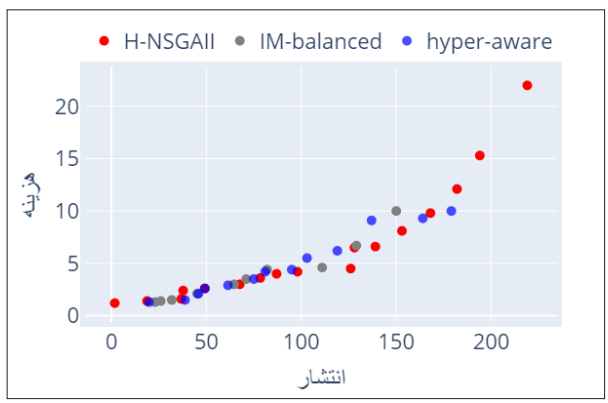


شکل ۵. مقایسه میزان انتشار با توجه به تعداد بذر

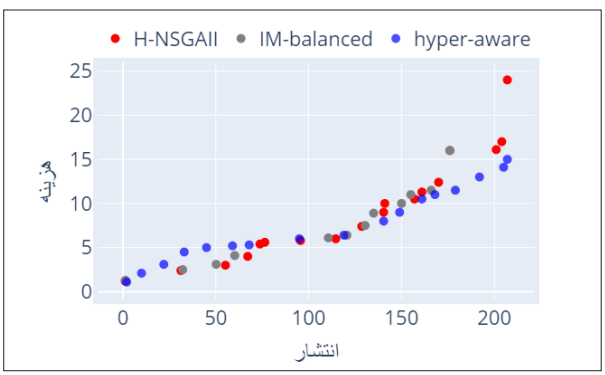
بیشـــتری همراه اســـت. لازم به ذکر است که مقادیر نشان داده شده در مسئلهٔ دوهدفه، بهترین جواب‌های جبهه پارتو در ۵۰ بار اجرای الگوریتم‌ها می‌باشند.



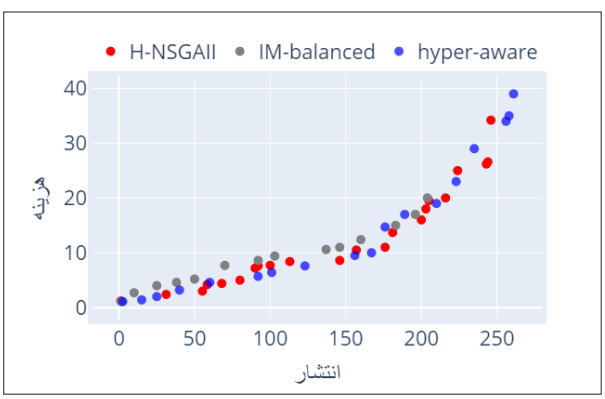
شکل ۶. انتشار در مقابل هزینه، k=1



شکل ۷. انتشار در مقابل هزینه، k=2

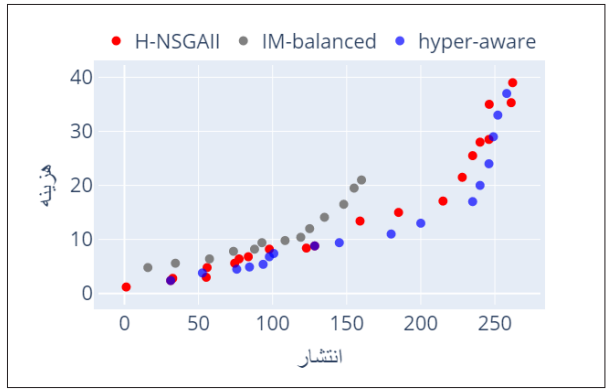


شکل ۸. انتشار در مقابل هزینه، k=3



شکل ۹. انتشار در مقابل هزینه، k=4

با توجـــه به این کـــه در آزمایش ۱، مســـئلهٔ تک هدفه اســـت و تمرکز الگوریتم ژنتیک روی بیشینه‌کردن انتشار به تنهایی است، میزان انتشـــار در مسئلهٔ تک هدفه بیشتر



شکل ۱۰. انتشار در مقابل هزینه، k=5

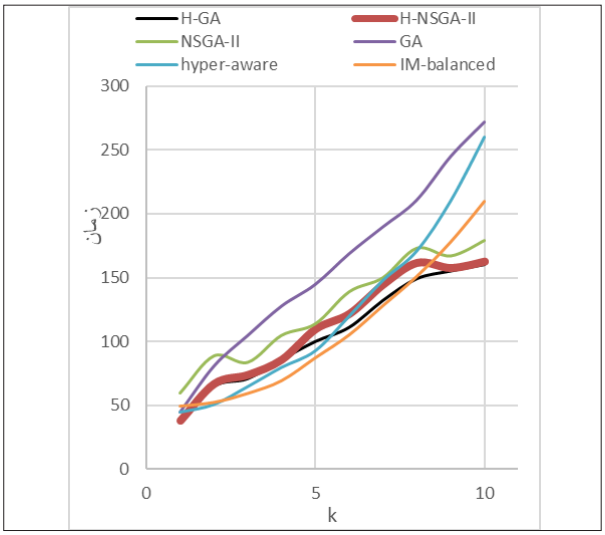
از میزان انتشـــار در مســـئلهٔ دو هدفه، با توجه به تناقض دو هدف، اســـت. بدین‌منظور مقایسه‌ای بین میزان انتشار الگوریتم H-GA و H-NSGA-II برای مقادیر مختلف k انجام شده که در جدول ۱ قابل مشاهده است.

جدول ۱. مقایسه میزان انتشار در شرایط یکسان برای مسئله تک هدفه (با استفاده از H-GA) و دو هدفه (با استفاده از H-NSGA-II)

k	انتشار(تک هدفه)	انتشار(دو هدفه)
۱	۱۳۵	۱۲۶
۲	۲۱۱	۲۱۹
۳	۲۶۰	۲۰۷
۴	۲۷۷.۶	۲۴۶
۵	۲۸۶.۶	۲۶۲
۶	۳۰۱	۲۷۴
۷	۳۲۶	۲۷۳
۸	۳۲۵.۴	۲۹۰
۹	۳۳۷	۳۱۰

آزمایش سوم: کمینه‌سازی زمان اجرا

در این آزمایش تاثیر اســـتفاده از جدول درهم‌ســـازی در زمـــان اجرای هـــر دو الگوریتـــم H-GA و H-NSGA-II در مقایســـه با روش‌های دیگر بررســـی شده است. نتایج زمان اجرای روش پیشـــنهادی برای $k = 1, 2, \dots, 10$ در شکل ۱۱ آمده اســـت. با افزایش k سربار زمانی جستجو در جدول درهم‌ســـازی بیشتر می‌شـــود و هم‌زمان تعداد رخدادهـــای یـــک کرومـــوزوم کاهـــش می‌یابد بـــه همین دلیل در مســـئلهٔ دو هدفه نســـبت به مسئلهٔ تک‌هدفه بهبود کمتری مشـــاهده می‌شـــود. دلیل اصلی‌تر آن این است که در دسته الگوریتم‌های NSGA تنوع را در نسل‌های مختلف حفظ می‌کننـــد و همین موضوع باعث می‌شـــود تکرار در کروموزوم‌هـــا به شـــدت کاهش یابد. امـــا همچنان کاهش زمان اجرا برای kهای مختلف مشهود است. همان‌گونه که دیده می‌شود، زمان اجرا با اضافه‌کردن جدول درهم‌سازی در روش H-NSGA-II در مقایســـه با روش‌هایی مانند IM-balance و hype-aware به‌خصوص برای kهای بزرگتر از ۸ بهبود یافته است.



شکل ۱۱. مقایسه زمان اجرای الگوریتم‌ها

جدول ۲. مقایسه زمان اجرای الگوریتم‌ها

k	H-NSGA-II	hyper-aware	IM-balanced
۱	۳۸	۴۵	۴۹
۲	۶۷	۵۱	۵۲
۳	۷۴	۶۵	۵۹
۴	۸۶	۸۰	۶۹
۵	۱۱۰	۹۳	۸۷
۶	۱۲۲	۱۲۰	۱۰۵
۷	۱۴۵	۱۴۸	۱۲۸
۸	۱۶۲	۱۷۱	۱۵۱
۹	۱۵۸	۲۱۰	۱۷۸
۱۰	۱۶۳	۲۶۰	۲۱۰

در جـــدول ۲، میـــزان زمـــان اجـــرای الگوریتم‌های H-NSGA–II، hyper–aware و IM-balanced بـــه ازای

بذرهای متفاوت مورد بررسی قرار گرفته شده است. همان‌گونه که دیده می‌شود زمان اجرای میان دو الگوریتم H-NSGA-II در مقایسه با الگوریتم hyper-aware تا $k = 8$ بسیار به هم نزدیک است و برای $k > 8$، زمان اجرای الگوریتم پیشنهادی به‌خصوص برای $k = 10$ با اختلاف زیاد از بقیه الگوریتم‌ها بهتر است، زیرا الگوریتم‌های IM-balanced و hyper-aware پیچیدگی نمایی برحسب k دارند که برای مقادیر بزرگ k مشکل‌ساز است.

نمودار آبی‌رنگ زمان اجرای الگوریتم ژنتیک بهبودیافته با جدول درهم‌سازی است که در مسئله تک هدفه استفاده شده است و برای مقایسه با الگوریتم پیشنهادی رسم شده است. همانطور که مشخص است با این که الگوریتم H-NSGA-II در هر ارزیابی باید ۲ هدف را محاسبه کند، باز هم میزان زیادی زمان اضافه برای اجرا نیاز ندارد و دلیل آن هم این است که ساختار الگوریتم H-NSGA-II به آن اجازه می‌دهد نرخ همگرایی بالاتری برای رسیدن به جواب بهینه داشته باشد.

به منظور مقایسه بهتر میزان بهبود زمان اجرا، مقادیر زمانی اجرای الگوریتم‌های H-NSGA-II و NS-GA-II در جدول ۳ به ازای تعداد بذرهای متفاوت نشان داده شده است. همچنین مدت زمان اجرای الگوریتم‌های H-GA و GA به توجه به تعداد بذر متفاوت در جدول ٤ مقایسه شده است. نتایج نشان می‌دهند که الگوریتم H-NSGA-II به حداکثر بهبود به ازای k=۱ از نظر زمان اجرا می‌رسد و به طور میانگین مدت زمان اجرای به ازای تعداد بذرهای از یک تا ده، ۱۳ درصد افزایش می‌دهد. همچنین الگوریتم H-GA در مقایسه با GA زمان اجرا را به طور میانگین تا ۳۰ درصد افزایش می‌دهد، درحالی که بهترین بهبود زمان اجرا را برای k=۱۰ به دست می‌آورد. با توجه به نتایج به دست آمده، بررسی تاثیر اضافه‌کردن جدول درهم‌سازی به الگوریتم‌های حل مسائل بهینه‌سازی توصیه می‌شود.

جدول ۳: میزان بهبود در زمان اجرا به‌وسیله جدول درهم‌سازی در NSGA-II

k	NSGA-II	H-NSGA-II	بهبود(درصد)
۱	۶۰	۳۸	۳۶,۷
۲	۸۹	۶۷	۲۴,۷
۳	۸۴	۷۴	۱۱,۹
۴	۱۰۵	۸۶	۱۸,۱
۵	۱۱۴	۱۱۰	۳,۵
۶	۱۳۹	۱۲۲	۱۲,۲
۷	۱۵۰	۱۴۵	۳,۳
۸	۱۷۳	۱۶۲	۶,۴
۹	۱۶۷	۱۵۸	۵,۴
۱۰	۱۷۹	۱۶۳	۸,۹
			میانگین: ۱۳,۱

جدول ۴: میزان بهبود در زمان اجرا به‌وسیله جدول درهم‌سازی در الگوریتم ژنتیک

k	GA(ثانیه)	H-GA(ثانیه)	بهبود(درصد)
۱	۴۵	۳۷	۱۷,۸
۲	۸۱	۶۵	۱۹,۸
۳	۱۰۵	۷۱	۳۲,۴
۴	۱۲۸	۸۶	۳۲,۸
۵	۱۴۵	۱۰۰	۳۱,۰
۶	۱۶۹	۱۱۱	۳۴,۳
۷	۱۹۰	۱۳۲	۳۰,۵
۸	۲۱۱	۱۴۹	۲۹,۴
۹	۲۴۵	۱۵۵	۳۶,۷
۱۰	۲۷۲	۱۶۰	۴۱,۲
			میانگین: ۳۰,۶

۵- نتیجه‌گیری

با توجه به این که اکثر شبکه‌های دنیای واقعی اندازه بزرگی دارند تحلیل آن‌ها زمان‌بر است. این امر در مسائل چندهدفه به علت زمان اجرای بالاتر آن‌ها پررنگ‌تر می‌شود. در این مقاله در ابتدا مسئله بیشینه‌سازی انتشار به صورت تک هدفه و سپس با در نظر گرفتن کمینه‌کردن

هزینه انتخاب بذرها به صورت دو هدفه ارائه شـــده است. درادامه به منظـــور بهبود زمان اجـــرای یافتن جواب‌های نزدیک به بهینه از الگوریتم H-GA و H-NSGA-II به ترتیب در مدل تک هدفه و دو هدفه اســـتفاده شـــده است. در این الگوریتم‌ها به علت اســـتفاده از جدول درهم‌سازی سرعت اجـــرا برای مدل تک هدفه بـــرای بذرهای یک تا ده به طور میانگین ۳۰ درصـــد و برای مدل دو هدفه به طور میانگین حدود ۱۳ درصد (بدون از دســـت دادن دقت) افزایش داده شـــد. همچنین مشاهده شـــد که روش پیشنهادی نسبت به الگوریتم‌هـــای چندهدفه اخیر، از نظـــر زمانی برای مقادیر بزرگ k عملکرد بهتری دارد. برای کارهای آتی بررســـی اســـتفاده از توابع درهم ساز در الگوریتم‌های تکاملی دیگر و همچنین بررسی اهداف دیگر در کنار هدف بیشینه‌سازی انتشار پیشنهاد می‌شود.

منابع

[1] M. Granovetter, “Network sampling: Some first steps,” American journal of sociology, vol. 81, no. 6, pp. 1287-1303, 1976.

[2] L. A. Sanchis, “Multiple-way network partitioning,” IEEE Transactions on Computers, vol. 38, no. 1, pp. 62-81, 1989.

[3] V. Chandola, A. Banerjee, and V. Kumar, “Anomaly detection: A survey,” ACM computing surveys (CSUR), vol. 41, no. 3, pp. 1-58, 2009.

[4] J. Yang and J. Liu, “Influence maximization-cost minimization in social networks based on a multiobjective discrete particle swarm optimization algorithm,” IEEE Access, vol. 6, pp. 2320-2329, 2017.

[5] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, and A. Fast, “Nsga-ii,” IEEE transactions on evolutionary computation, vol. 6, no. 2, pp. 182-197, 2002.

[6] M. Richardson and P. Domingos, “Mining knowledge-sharing sites for viral marketing,” in Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, 2002, pp. 61-70.

[7] P. Domingos and M. Richardson, “Mining the network value of customers,” in Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, 2001, pp. 57-66.

[8] D. Kempe, J. Kleinberg, and É. Tardos, “Maximizing the spread of influence through a social network,” in Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, 2003, pp. 137-146.

[9] D. Kempe, J. Kleinberg, and É. Tardos, “Influential nodes in a diffusion model for social networks,” in Automata, Languages and Programming: 32nd International Colloquium, ICALP 2005, Lisbon, Portugal, July 11-15, 2005. Proceedings 32, 2005: Springer, pp. 1127-1138.

[10] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, “Cost-effective outbreak detection in networks,” in Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, 2007, pp. 420-429.

[11] A. Goyal, W. Lu, and L. V. Lakshmanan, “Celf++ optimizing the greedy algorithm for influence maximization in social networks,” in Proceedings of the 20th international conference companion on World wide web, 2011, pp. 47-48.

[12] Y. Tang, X. Xiao, and Y. Shi, “Influence maximization: Near-optimal time complexity meets practical efficiency,” in Proceedings of the 2014 ACM SIGMOD international conference on Management of data, 2014, pp. 75-86.

[13] H. T. Nguyen, M. T. Thai, and T. N. Dinh, “A billion-scale approximation algorithm for maximizing benefit in viral marketing,” IEEE/ACM Transactions On Networking, vol. 25, no. 4, pp. 2419-2429, 2017.

[14] C. Borgs, M. Brautbar, J. Chayes, and B. Lucier, “Maximizing social influence in nearly optimal time,” in Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms, 2014: SIAM, pp. 946-957.

[15] J. Lv, J. Guo, and H. Ren, “Efficient greedy algorithms for influence maximization in social networks,” Journal of Information Processing Systems, vol. 10, no. 3, pp. 471-482, 2014.

[16] L. C. Freeman, “Centrality in social networks: Conceptual clarification,” Social network: critical concepts in sociology. Londres: Routledge, vol. 1, pp. 238-263, 2002.

[17] W. Chen, Y. Wang, and S. Yang, “Efficient influence maximization in social networks,” in Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, 2009, pp. 199-208.

[18] K. Jung, W. Heo, and W. Chen, “Irie: Scalable and robust influence maximization in social networks,” in 2012 IEEE 12th international conference on data mining, 2012: IEEE, pp. 918-923.

[19] D. Bucur and G. Iacca, “Influence maximization in social networks with genetic algorithms,” in Applications of Evolutionary Computation: 19th European Conference, EvoApplications 2016, Porto, Portugal, March 30--April 1, 2016, Proceedings, Part I 19, 2016: Springer, pp. 379-392.

[20] P. Krömer and J. Nowaková, “Guided genetic algorithm for the influence maximization problem,” in Computing and Combinatorics: 23rd International Conference, COCOON 2017, Hong Kong, China, August 3-5, 2017, Proceedings 23, 2017: Springer, pp. 630-641.

[21] C.-W. Tsai, Y.-C. Yang, and M.-C. Chiang, “A genetic newgreedy algorithm for influence maximization in social network,” in 2015 IEEE International Conference on Systems, Man, and Cybernetics, 2015: IEEE, pp. 2549-2554.

[22] D. Bucur, G. Iacca, A. Marcelli, G. Squillero, and A.

Tonda, “Multi-objective evolutionary algorithms for influence maximization in social networks,” in Applications of Evolutionary Computation: 20th European Conference, EvoApplications 2017, Amsterdam, The Netherlands, April 19-21, 2017, Proceedings, Part I 20, 2017: Springer, pp. 221-233.

[23] J.-b. Guo, F.-z. Chen, and M.-q. Li, “A multi-objective optimization approach for influence maximization in social networks,” in Proceeding of the 24th International Conference on Industrial Engineering and Engineering Management 2018, 2019: Springer, pp. 706-715.

[24] P.-L. Lu, L. Zhang, J.-X. Tang, J.-M. Lan, H.-Y. Zhu, and S.-H. Song, “Solving the Influence Maximization-Cost Minimization Problem in Social Networks by Using a Multi-Objective Differential Evolution Algorithm,” Journal of Computers, vol. 34, no. 5, pp. 285-303, 2023.

[25] P. Wang and R. Zhang, “A multi-objective crow search algorithm for influence maximization in social networks,” Electronics, vol. 12, no. 8, p. 1790, 2023.

[26] L. Zhang, Y. Liu, F. Cheng, J. Qiu, and X. Zhang, “A local-global influence indicator based constrained evolutionary algorithm for budgeted influence maximization in social networks,” IEEE Transactions on Network Science and Engineering, vol. 8, no. 2, pp. 1557-1570, 2021.

[27] J. Yang and J. Liu, “Influence Maximization-Cost Minimization in Social Networks Based on a Multiobjective Discrete Particle Swarm Optimization Algorithm,” IEEE Access, vol. 6, pp. 2320-2329, 2018, doi: 10.1109/ACCESS.2017.2782814.

[28] S. Genetti, E. Ribaga, E. Cunegatti, Q. F. Lotito, and G. Iacca, “Influence Maximization in Hypergraphs using Multi-Objective Evolutionary Algorithms,” arXiv preprint arXiv:2405.10187, 2024.

[29] X. Fu, R. R. Bhatt, S. Basu, and A. Pavan, “Multi-objective submodular optimization with approximate oracles and influence maximization,” in 2021 IEEE International Conference on Big Data (Big Data), 2021: IEEE, pp. 328-334.