

Colonial Competition Algorithm for Predicting Information Diffusion in Social Networks

Farzad Peyravi *

Assistant Professor, Faculty of Engineering,
Department of Computer Engineering, Islamic Azad
University, Marvdasht Branch, Marvdasht, Iran.

Abstract

In the modern world, social network analysis has various applications including social network management, market trend analysis, identifying influential individuals, supporters, and enhancing the performance of descriptive systems. One of the existing issues in the field of information diffusion is examining the extent of dissemination in such systems. In large-scale networks, computational complexity has increased, and one of the methods to cope with this problem is using evolutionary algorithms. In this paper, a new method is proposed to increase the accuracy of predicting information diffusion in social networks by combining the colonial competition algorithm and the centrality algorithm. In the proposed method, the centrality of nodes is first calculated for all nodes in the network. Then, based on the distance between users in the social network and the speed of information dissemination among users using the colonial competition algorithm, the optimal path for dissemination is determined, and information diffusion prediction is made accordingly. In the colonial competition algorithm, the optimal path for information dissemination in the network is identified based on how the colonies move towards the colonizer. In this paper, four different datasets were examined. The results show that the proposed method leads to better results. The algorithm's performance is evaluated based on accuracy, NMI, and error criteria, showing desirable results.

Keywords: social networks, node centrality, optimization, colonial competition algorithm

Received: 19/January/2024

Accepted: 19/May/2024

ISSN: 2980-8936

* Corresponding Author: fpeyravi@miau.ac.ir

الگوریتم رقابت استعماری در پیش‌بینی انتشار اطلاعات در شبکه‌های اجتماعی

استادیار دانشکده فنی و مهندسی، گروه مهندسی کامپیوتر، دانشگاه آزاد اسلامی،
واحد مرودشت، مرودشت، ایران.

فرزاد پیروی *

چکیده

در دنیای نوین امروزی، تحلیل شبکه‌های اجتماعی در بسیاری از کاربردها از جمله مدیریت شبکه اجتماعی، تحلیل گرایش بازار، شناسایی افراد تأثیرگذار، حامیان و ارتقای کارایی سامانه‌های توصیفگر قابل استفاده است. یکی از مسائل موجود در زمینه انتشار اطلاعات، بررسی میزان انتشار در چنین سیستم‌هایی است. در شبکه‌های با ابعاد بالا، پیچیدگی محاسباتی افزایش یافته و یکی از روش‌های مقابله با این مشکل، استفاده از الگوریتم‌های تکاملی است.

در این مقاله با ترکیب الگوریتم رقابت استعماری و الگوریتم مرکزیت گره، یک روش جدید برای افزایش دقت پیش‌بینی انتشار اطلاعات در شبکه‌های اجتماعی ارائه شده است. در روش پیشنهادی، ابتدا میزان مرکزیت گره‌ها برای تمام گره‌های شبکه محاسبه می‌شود. سپس، بر اساس فاصله بین کاربران شبکه اجتماعی و سرعت انتشار اطلاعات بین کاربران با استفاده از الگوریتم رقابت استعماری، مسیر بهینه برای انتشار تعیین شده و بر اساس آن، پیش‌بینی انتشار اطلاعات صورت می‌گیرد. در الگوریتم رقابت استعماری بر اساس نحوه حرکت مستعمره‌ها به سمت استعمارگر، مسیر بهینه برای انتشار اطلاعات در شبکه شناسایی می‌شود. در این مقاله، چهار مجموعه داده مختلف مورد بررسی قرار گرفت. بررسی نتایج نشان می‌دهد که روش پیشنهادی منجر به نتایج بهتری شده است. عملکرد الگوریتم با توجه به معیاری‌های دقت و NMI و خطا ارزیابی شده است و نتایج مطلوبی را نشان می‌دهد.

کلیدواژه‌ها: شبکه‌های اجتماعی، مرکزیت گره، بهینه‌سازی، الگوریتم رقابت استعماری

مقدمه

شبکه‌های اجتماعی با بهره‌گیری از به‌روزترین تکنولوژی‌ها و فناوری‌های جدید وب، راهکاری جدید جهت نمایش ارتباطات و تعاملات میان افراد به شمار می‌روند. این شبکه‌ها بر اساس فناوری‌هایی مبتنی بر سرویس وب هستند که تبادل داده بین اشخاص، سازمان‌ها و اجتماعات را محقق می‌کنند و مشخصه بارز آن‌ها خلق و مبادله محتوایی است که توسط کاربران تولید می‌شود. از مشخصه‌های مهم شبکه‌های اجتماعی می‌توان به غیرمتمرکز بودن، پویایی، پراکندگی جغرافیایی و سرعت انتشار اشاره کرد. کاربردهای این شبکه‌ها در شکل‌گیری فضای مباحثه و گفتگو، نقد و آگاهی از نظرات مختلف، آگاهی از رخدادهای اخبار رسمی و غیررسمی، کمپین‌های مختلف سیاسی و اجتماعی و تبلیغاتی، بهبود ارتباطات دوستانه، دستیابی به شناخت بیشتر از خود و دیگران و سرگرمی است (Jin et al., 2013).

شبکه‌های اجتماعی از تجمع تعدادی از عوامل و روابط میان آن‌ها تشکیل می‌شوند که برای نمایش و محاسبات اولیه بر روی آن‌ها، می‌توان نگاشتی از شبکه‌ها به گراف‌ها در نظر گرفت. در تئوری گراف، هر عامل یک گره و هر رابطه میان آن‌ها نشان‌دهنده یک یال است. شبکه‌های اجتماعی پایه را با استفاده از یک گراف جهت‌دار $G = (U, F)$ نشان می‌دهیم که در آن، U مجموعه کاربران و F مجموعه لینک‌های دوستی است. اطلاعات شبکه‌های اجتماعی توسط یک ماتریس $S \in \mathbb{R}$ نشان داده می‌شود. هر کاربر u دارای یک مجموعه F_u^+ از همسایه‌های مستقیم است که به او اعتماد دارند و یا آن را دنبال می‌کنند. برخی اوقات، u مورد اعتماد یا دنبال‌کننده به وسیله یک مجموعه F_u^- از کاربران است. هدایت و وزن‌دهی ارتباط اجتماعی کاربر u به کاربر v (به عنوان مثال، اعتماد/شناخت/دنبال کردن) توسط یک مقدار مثبت $S_{u,v} \in [0, 1]$ نشان داده می‌شود. این مقدار می‌تواند از بازخورد صریح کاربر u درباره کاربر v (مثلاً به کمک رأی‌گیری) و یا بازخورد ضمنی (به عنوان نمونه، درجه‌ای از تعامل یا ارتباط) بین آن‌ها به وجود آید (Qi et al., 2012).

در حال حاضر، شبکه‌های اجتماعی فرصتی مناسب را جهت بهبود دقت سیستم‌های توصیه‌گر فراهم کرده‌اند. استفاده از اطلاعات آن‌ها در سیستم‌های توصیه‌گر مبتنی بر محتوا، مسئله اعتماد بین فردی و گروهی را در سیستم‌های توصیه‌گر بهبود داده است. درواقع، شبکه‌های اجتماعی، دقت سیستم توصیه‌گر سنتی را با استفاده از اعتماد و سلیقه اجتماعی موجود بین کاربران افزایش می‌دهد. یکی از عوامل مهم در شبکه‌های اجتماعی، نحوه انتشار اطلاعات به منظور انجام رتبه‌بندی مناسب جهت یک آیتم خاص برای کاربران است. نحوه انجام انتشار در ارائه نتایج مطلوب به کاربران می‌تواند از اهمیت بالایی برخوردار باشد. استفاده از افراد بانفوذ در شبکه در انتشار اطلاعات جهت رتبه‌بندی یک آیتم، از جمله فاکتورهای مهم به شمار می‌رود؛ بنابراین، چالش پیش رو، روش مناسب برای تشخیص این افراد است. مثلاً یکی از معیارهای تشخیص این افراد در گراف‌های جهت‌دار اجتماعی، بررسی تعداد یال‌های ورودی به یک گره (کاربر) است که درواقع نشان‌دهنده تعداد افرادی است که آن شخص را دنبال کرده‌اند. در نتیجه، هرچه تعداد دنبال‌کننده‌ها برای یک فرد بیشتر باشد، فرد از اعتماد و نفوذ بیشتری در شبکه اجتماعی برخوردار خواهد بود (Meo et al., 2012; Katarya, 2018).

یکی از مسائل موجود در شبکه‌های اجتماعی، انتشار اطلاعات و بررسی میزان انتشار در چنین سیستم‌هایی است. به عبارت دیگر، به منظور رتبه‌بندی یک آیتم خاص جهت انتشار اطلاعات در شبکه، بهتر است بررسی کنیم مقدار شعاع پیشروی را برابر با چه مقداری قرار دهیم تا علاوه بر کاهش هزینه محاسباتی، در رتبه‌بندی دقیق و افزایش پوشش آیتم‌ها بیشترین کارایی را داشته باشیم. همچنین، در صورتی که در این شعاع خاص برای آیتم موردنظر هیچ‌گونه مقداری جهت رتبه‌بندی موجود نباشد، مقدار رتبه‌بندی برای آیتم موردنظر به چه صورت تعیین خواهد شد؟ آیا در زمان انتشار اطلاعات جهت رتبه‌بندی، بهتر است عمل انتشار برای تمام دوستان انجام شود یا کسانی که کاربر آن‌ها

را دنبال می‌کند و بیشترین میزان اعتماد را در بین سایرین به آن‌ها دارد؟ لازم به ذکر است، در صورتی که مقدار شعاع پیشروی را بیش از یک مقدار خاص قرار دهیم، به دلیل بررسی‌های بیشتر به منظور رتبه‌بندی، هزینه محاسباتی و پوشش افزایش پیدا کرده اما دقت کاهش خواهد یافت. درضمن، با کم کردن این شعاع جهت انتشار اطلاعات، هزینه محاسبات و میزان پوشش کاهش پیدا کرده و دقت افزایش می‌یابد چراکه دقت با پوشش رابطه عکس دارد. بر این اساس، تعیین یک مقدار پیشروی مناسب جهت انجام این انتشار که هزینه محاسباتی را کاهش داده و دقت را افزایش دهد، می‌تواند به عنوان یک چالش اساسی در انتشار اطلاعات به منظور رتبه‌بندی مناسب یک آیتم خاص در نظر گرفته شود.

الگوریتم‌های بسیاری در زمینه انتشار اطلاعات استفاده شده است. در یک پژوهش (Park et al., 2015) برای هر جفت از دوستان، شباهت رتبه آن‌ها با استفاده از یک مجموعه از احتمالات شرطی اندازه‌گیری می‌شود و توزیع رتبه‌بندی یک کاربر از کاربر دیگر به دست می‌آید و توسط آیتم‌های رتبه‌بندی شده مشترک بین کاربران محاسبه می‌شود. وقتی که یک کاربر می‌خواهد عمل رتبه‌بندی را برای یک آیتم خاص انجام دهد، ابتدا با توجه به آیتم موردنظر، یک درخت انتشار را بر اساس سایر کاربران ایجاد کرده و سپس یک پرس‌وجوی رتبه‌بندی را برای دوستان مستقیم خود در شبکه اجتماعی ارسال می‌کند. به محض دریافت یک پرس‌وجو برای یک آیتم، یک کاربر اگر از قبل به آیتم رتبه داده باشد، آن مقدار رتبه‌بندی را برمی‌گرداند و در غیر این صورت، آن پرس‌وجو را به دوستان مستقیم خود ارسال می‌کند.

مدل‌های انتشار اطلاعاتی ارائه شده است که در آن اطلاعات کدگذاری شده به عنوان یک پیام در گره منبع تولید شده و ممکن است در طول مسیر دلخواه به گره هدف در شبکه هدایت شوند. در چنین مدلی، برخی از نویسنده‌ها مانند نیومن^۱ (۲۰۰۵) و نوح و رایجر^۲ (۲۰۰۴) روش پیاده‌روی تصادفی بر روی شبکه‌های اجتماعی برای محاسبه مقدار مرکزیت را پیشنهاد داده‌اند.

روشی که در این تحقیق (Alahakoon et al., 2011) ارائه شده است به این صورت است که نویسنده معیار جدید مرکزیت یال به نام k -path را معرفی می‌کند. به طور کلی، سیستم، توپولوژی کل شبکه را ندارد و از کوتاه‌ترین مسیر برای مسیریابی پیام آگاه نیست. هر گره، تنها از گره‌هایی که در کتاب آدرس آن ظاهر می‌شود، آگاه است. گره با توجه به معیارهای خاص به یک یا چند مخاطب خود پیام می‌فرستد و به این امید که ادامه انتشار بیشتر خواهد شد. برای شبیه‌سازی انتشار پیام از پیاده‌روی تصادفی استفاده کرده‌اند. روش شبیه‌سازی انتشار پیام از طریق پیاده‌روی تصادفی می‌تواند حاکی از آن باشد که یک پیام می‌تواند بیش از یک بار از طریق یک یال عبور کند. در چنین مواردی، رتبه یال‌هایی که چندین بار از آن‌ها عبور داده می‌شود، به طور نامتناسبی زیاد خواهد بود (Moradi & Rostami, 2015; Girvan & Newman, 2002).

مرکزیت گره

مرکزیت یک گره در شبکه اجتماعی نشان‌دهنده اهمیت و نفوذ آن گره در شبکه است. گره‌هایی که در مرکز شبکه قرار دارند، تأثیرگذاری بیشتری دارند. مرکزیت یک معیار است که میزان نقش آفرینی یک گره را در شبکه کمی‌سازی می‌کند. میزان نفوذ یک گره در شبکه، معمولاً برحسب مرکزیت آن گره بیان می‌شود. به طور کلی، مرکزیت بیشتر، سبب داشتن ارتباطات بیشتر و کسب موقعیت مطلوب‌تر برای آن گره می‌شود که نهایتاً آن گره را قدرتمندتر می‌کند.

برای ارزیابی مرکزیت گره‌ها در شبکه اجتماعی، معیارهای رتبه^۱، نزدیکی^۲، بینابینی^۳ و بردار ویژه^۴ از پرستفاده‌ترین موارد بوده‌اند (Newman, 2005; Ferrara et al., 2012).

معیار مرکزیت رتبه، یکی از معیارها یا شاخص‌های شبکه‌ای است که در تحلیل ساختار اکثر شبکه‌ها و موقعیت‌های افراد در شبکه مفید است. معیار مرکزیت رتبه، به تعداد پیوندهای واردشده یا خارج‌شده از یک گره در یک شبکه اشاره دارد. این معیار، به‌طور ساده با شمارش تعداد ارتباطات نگه‌داری‌شده توسط هر فرد در شبکه، اندازه‌گیری می‌شود. در یک گراف، این کار با شمارش تعداد پیوندهای یک گره خاص تحقق می‌یابد. به‌طور کلی، مرکزیت رتبه، محاسبه میزان پیوندهایی است که فرد با دیگر افراد در شبکه دارد. رویکرد مرکزیت رتبه به این نکته تأکید دارد که افرادی که ارتباطات بیشتری دارند، به احتمال زیاد قوی‌تر و بانفوذتر هستند زیرا آن‌ها به‌صورت مستقیم بر سایر افراد در شبکه تأثیر می‌گذارند (Yan et al., 2013).

مرکزیت نزدیکی، فاصله یک فرد با کلیه افراد دیگر در شبکه را می‌سنجد. هر چقدر یک فرد به دیگران نزدیک‌تر باشد، آن فرد برگزیده‌تر و مشهورتر است. افراد با نمرات نزدیکی بالا، به دلیل اینکه میانجی‌های کمتری بین آن‌ها وجود دارد، اطلاعات را خیلی سریع‌تر از دیگران دریافت می‌کنند و یا به دیگران انتقال می‌دهند. در معیار مرکزیت نزدیکی، ارزیابی از طریق قضاوت نزدیکی یک فرد به افراد دیگر صورت می‌گیرد. این نوع مرکزیت از طریق طول مسیرها یا گام‌های موردنیاز برای یک فرد تا به دیگر افراد در شبکه برسد، اندازه‌گیری می‌شود. افرادی که قادرند به دیگر افراد شبکه با طول مسیر کوتاه‌تری برسند، قدرت و نفوذ بیشتری در شبکه دارند (Jin et al., 2013; Hansen et al., 2010).

معیار مرکزیت بینابینی، شاخصی است که مسیر دقیق‌تری را برای اندازه‌گیری مرکزیت یک فرد مشخص می‌کند. این فاکتور، مرکزیت را با بررسی وسعتی که در آن فرد خاص بین دیگر افراد در شبکه قرار می‌گیرد، اندازه‌گیری می‌کند. معیار مرکزیت بینابینی، موقعیت یک فرد را درون یک شبکه بر حسب توانایی آن برای ایجاد ارتباط با سایر زوج‌ها یا گروه‌ها در شبکه شناسایی می‌کند. به‌طور کلی، فردی با بالاترین مرکزیت بینابینی دارای موقعیت مطلوب و مستحکمی در شبکه است و تأثیر بسیار زیادی بر شبکه می‌گذارد. نقاط با مرکزیت بینابینی بالا، نقاط واسطه‌ای هستند که راه‌های ارتباطی نقاط دیگر از آن‌ها می‌گذرد. این نقاط دارای قدرت ایزوله‌سازی یا افزایش ارتباطات هستند (Jiang et al., 2013; Zhou et al., 2015).

مرکزیت بردار ویژه بر اساس این ایده پیشنهاد شده است که مرکزیت یک گره خاص نمی‌تواند مجزا از مرکزیت دیگر گره‌هایی که با آن متصل شده است، تخمین زده شود. نمره مرکزیت به یک گره بر اساس این اصل است که ارتباط به گره‌های با نمره بالا در نمره آن گره خاص تأثیر دارد. یک نقطه با مرکزیت بردار ویژه بیشتر، دارای همسایگان مرکزی بیشتری است. به عبارت دیگر، مرکزیت بردار ویژه سبب قدرت بیشتر می‌شود.

هر کدام از این معیارها دارای مزایا و معایبی هستند. سادگی و پیچیدگی محاسباتی کم از مزایای معیار مرکزیت رتبه است. از طرف دیگر، این معیار برای محاسبه مرکزیت گره، فقط از ساختار محلی گره‌ها استفاده کرده و ساختار جهانی شبکه در نظر گرفته نمی‌شود؛ بنابراین، ممکن است با وجود اینکه یک گره مرکزیت رتبه بالایی داشته باشد، در موقعیت مناسبی برای انتقال سریع اطلاعات نباشد. معیار مرکزیت نزدیکی از ساختار جهانی شبکه برای محاسبه معیار مرکزیت استفاده نموده و نسبت به معیار رتبه با دقت بیشتری مرکزیت را محاسبه می‌کند اما دارای این محدودیت

1. degree centrality
2. closeness centrality
3. betweenness centrality
4. eigenvector centrality

است که اگر گراف دارای یک گره باشد که با هیچ کدام از گره‌های دیگر ارتباط نداشته باشد، مقدار مرکزیت نزدیکی برای کل شبکه صفر خواهد شد. با وجود مزایا و معایب هر کدام از این معیارها، ارائه یک معیار مرکزیت مناسب که هم از لحاظ پیچیدگی محاسباتی کارا بوده و هم اطلاعات مناسبی را از ساختار شبکه ارائه دهد، یکی از چالش‌های این حوزه محسوب می‌شود. برای نمونه، معیار مرکزیت رتبه دارای کارایی بالایی بوده ولی نشان‌دهنده اطلاعات کافی در مورد ساختار شبکه نیست. از طرف دیگر، معیارهایی مانند نزدیکی و بینابینی، از ساختار جهانی شبکه استفاده کرده و به صورت دقیق‌تر شبکه را مورد تحلیل قرار می‌دهند اما به علت پیچیدگی محاسباتی بالا، برای شبکه‌های با مقیاس بزرگ قابل استفاده نیستند (Pellegrina, 2023). دو شاخص پیچیدگی محاسباتی و کیفیت، اغلب در تعارض با یکدیگر بوده و بهبود یکی از آن دو باعث تنزل یافتن دیگری می‌شود. یکی از معیارهای مرکزیت که به تازگی ارائه شده، معیار مرکزیت لاپلاسین^۱ (Qi et al., 2012) است. این معیار که برای محاسبه مرکزیت گره‌ها در شبکه‌های وزن‌دار ارائه شده است، هم از ساختار محلی گره‌ها و هم از ساختار جهانی شبکه برای تحلیل ساختار شبکه استفاده می‌کند. در این معیار، نه فقط همسایگان مستقیم یک گره بلکه اهمیت و نفوذ آن همسایگان نیز مدنظر گرفته می‌شود. $G(V, E, W)$ به عنوان یک گراف وزن‌دار دارای مجموعه رئوس $V(G) = \{v_1, v_2, \dots, v_n\}$ و مجموعه یال‌های E که هر یال $e(v_i, v_j)$ به وسیله وزن w_{ij} به هم وصل شده‌اند، در نظر گرفته می‌شود. دو ماتریس W و X وجود دارد که در اینجا $X_i = \sum_{j=1}^n w_{i,j} = \sum_{u \in N(v_i)} w_{v_i, u}$ نشان‌دهنده مجموع وزن‌های گره v_i است. $N(v_i)$ مجموعه همسایه‌های گره v_i را نشان می‌دهد. لازم به ذکر است، اگر بین دو گره v_i و v_j یالی وجود نداشته باشد، آنگاه $w_{i,j} = 0$. انرژی لاپلاسین برای گراف G به صورت رابطه ۱ محاسبه می‌شود:

$$E_L(G) = \sum_{i=1}^n X_i + 2 \sum_{i < j} w_{i,j}^2 \quad (1)$$

معیار مرکزیت لاپلاسین برای گره v_i به صورت رابطه ۲ محاسبه می‌شود.

$$C_L(v_i, G) = \frac{(\Delta E)_i}{E_L(G)} = \frac{E_L(G) - E_L(G_i)}{E_L(G)} \quad (2)$$

که در اینجا، G_i نشان‌دهنده گراف G پس از حذف گره v_i است.

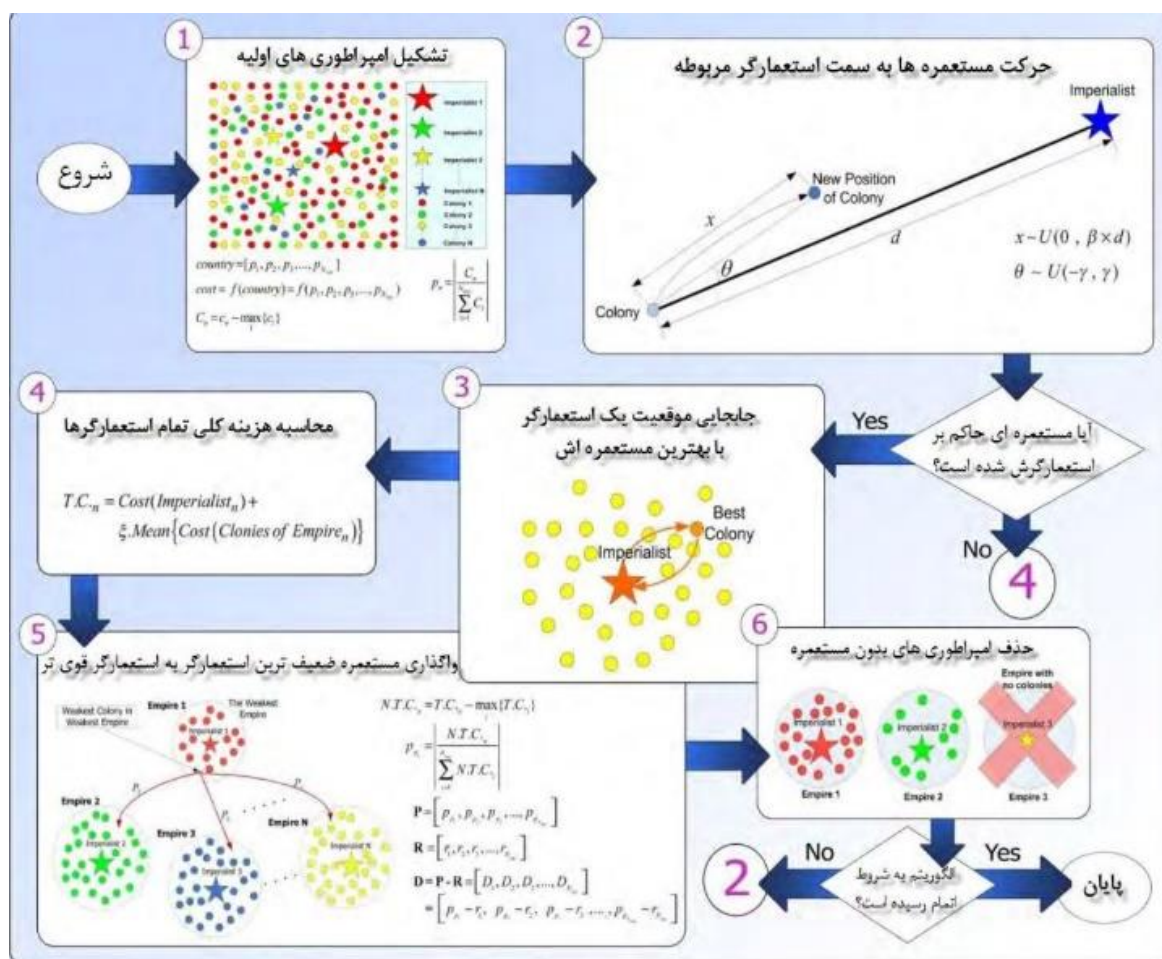
در مقایسه با سایر روش‌های محاسبه معیار مرکزیت (مانند رتبه، نزدیکی و بینابینی)، معیار مرکزیت لاپلاسین معیاری است که بین دو حالت مختلف محاسبه معیار مرکزیت بر اساس ساختار محلی و ساختار جهانی قرار دارد (Li et al., 2016). برای نمونه، روش رتبه از ساختار محلی و روش بینابینی از ساختار جهانی گراف برای محاسبه معیار مرکزیت استفاده می‌کنند. از لحاظ پیچیدگی محاسباتی نیز معیار لاپلاسین دارای پیچیدگی محاسباتی بیشتر از معیار رتبه و کمتر از معیار بینابینی است. برای یک گراف وزن‌دار با n گره و m یال، معیار رتبه دارای پیچیدگی محاسباتی $O(m)$ و معیار بینابینی و نزدیکی دارای پیچیدگی محاسباتی $O(n^3)$ هستند. معیار مرکزیت لاپلاسین نیز دارای پیچیدگی محاسباتی $O(n \cdot \Delta^2)$ بوده که Δ حداکثر درجه گره‌ها را نشان می‌دهد؛ بنابراین، این معیار برای گراف‌هایی با تعداد گره‌های بسیار زیاد نیز قابل استفاده است.

الگوریتم رقابت استعماری^۲

در الگوریتم‌های بهینه‌سازی، علی‌رغم توجه به تکامل زیستی انسان و سایر موجودات (الگوریتم‌های ژنتیک و غیره)، توجه چندانی به تکامل اجتماعی و تاریخی به عنوان پیچیده‌ترین و موفق‌ترین حالت تکامل نشده است (Li & Li, 2016).

1. laplacian centrality
2. imperialist competitive algorithm

(2010; Dorigo et al., 2006). در این مقاله از الگوریتم الهام گرفته‌شده از تکامل اجتماعی انسان، برای بهینه‌سازی استفاده شده است. الگوریتم رقابت استعماری با الهام‌گیری از یک فرایند اجتماعی سیاسی، نسبت به روش‌های مطرح‌شده دارای توانایی بالایی بوده و تا حد بسیار زیادی نیز سریع است. شکل ۱، تصویر کلی الگوریتم را نشان می‌دهد. الگوریتم همانند سایر روش‌های بهینه‌سازی تکاملی، با تعدادی جمعیت اولیه شروع می‌شود (Cai et al., 2010). هر عنصر جمعیت، یک کشور نامیده می‌شود. کشورها به دو دسته مستعمره^۱ و استعمارگر^۲ تقسیم می‌شوند. هر استعمارگر، متناسب با قدرت خود تعدادی از کشورهای مستعمره را به سلطه خود درآورده و آن‌ها را کنترل می‌کند. سیاست جذب و رقابت استعماری، هسته اصلی این الگوریتم را تشکیل می‌دهد. مطابق سیاست جذب که به صورت تاریخی توسط کشورهای استعمارگری همچون فرانسه و انگلیس، در مستعمراتشان اعمال می‌شد، کشورهای استعمارگر با استفاده از روش‌هایی همچون احداث مدارس به زبان خود، سعی در استعمار کشور با از میان بردن زبان کشور مستعمره و فرهنگ و رسوم آن داشتند. در ارائه این الگوریتم، این سیاست با حرکت دادن مستعمرات یک امپراتوری، مطابق یک رابطه خاص صورت می‌پذیرد (Hosseini & Khaled, 2014).



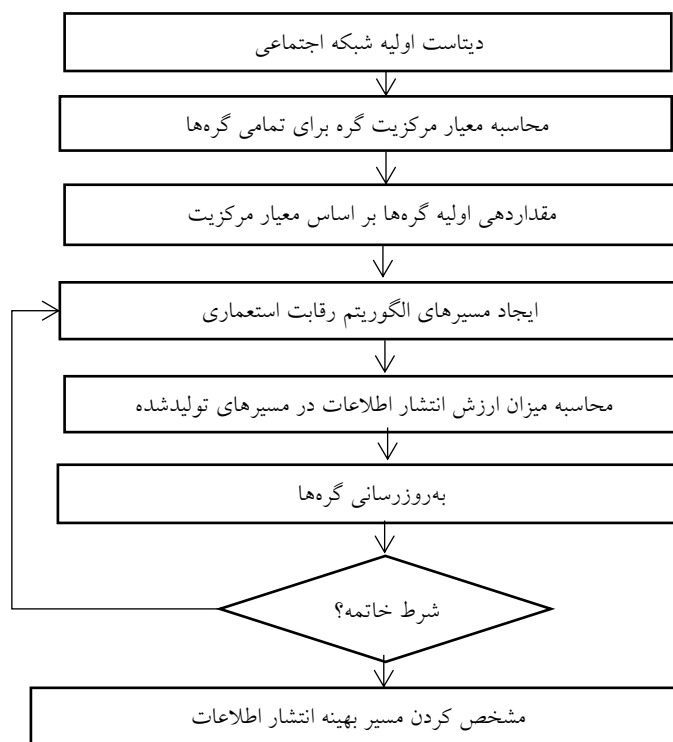
شکل ۱. نمای کلی الگوریتم رقابت استعماری

اگر در حین حرکت، یک مستعمره نسبت به استعمارگر به موقعیت بهتری برسد، جای آن دو با هم عوض می‌شود. قدرت کل یک امپراتوری به صورت مجموع قدرت کشور استعمارگر به اضافه درصدی از قدرت میانگین مستعمرات

آن تعریف می‌شود. رقابت استعماری، بخش مهم دیگری از این الگوریتم را تشکیل می‌دهد. در طی رقابت استعماری، امپراتوری‌های ضعیف، به تدریج قدرت خود را از دست داده و به مرور زمان با تضعیف شدن از بین می‌روند. رقابت استعماری باعث می‌شود که در انتها فقط یک امپراتوری وجود داشته باشد. این حالت زمانی است که الگوریتم رقابت استعماری با رسیدن به نقطه بهینه تابع هدف، متوقف می‌شود.

روش تحقیق

در شکل ۲، فلوچارت روش پیشنهادی برای پیش‌بینی انتشار اطلاعات در شبکه‌های اجتماعی نشان داده شده است.



شکل ۲. فلوچارت روش پیشنهادی

هر دیتاست شبکه اجتماعی، یک ماتریس مربعی بوده که اندازه آن وابسته به اندازه گره‌های شبکه است. هر خانه از این ماتریس نشان‌دهنده این است که وضعیت ارتباط بین دو گره خاص در شبکه به چه صورت است. مرحله دوم از روش پیشنهادی، محاسبه معیار مرکزیت گره است که بر اساس الگوریتم مرکزیت لاپلاسین محاسبه می‌شود.

شکل‌دهی امپراتوری‌های اولیه

در ابتدا آرایه‌ای از متغیرهای مسئله را که باید بهینه شوند، ایجاد می‌کنیم. متغیرها در الگوریتم کشور هستند. هدف، پیش‌بینی انتشار اطلاعات بوده و هر گره به‌عنوان یک کشور در نظر گرفته می‌شود. مسئله بهینه‌سازی یک کشور، یک آرایه یک‌بعدی است. برای شروع الگوریتم باید کشورهای اولیه الگوریتم ایجاد شوند. برای بهبود مسئله به جای اینکه ماتریس اولیه کشورها به‌صورت تصادفی تشکیل شود، هزینه هر کشور برابر معکوس مرکزیت آن گره (کشور) در نظر گرفته می‌شود.

هزینه یک کشور با ارزیابی تابع f در متغیرهای $(p_1, p_2, p_3, \dots, p_{N_{\text{var}}})$ یافت می‌شود؛ بنابراین:

$$cost_i = f(country_i) = f(p_1, p_2, p_3, \dots, p_{N_{var}}) \quad (3)$$

برای شروع الگوریتم، تعداد $N_{country}$ کشور اولیه را ایجاد و N_{imp} تا از بهترین کشورها (کشورهای دارای کمترین مقدار تابع هزینه) را به عنوان استعمارگر انتخاب می‌کنیم. باقیمانده N_{col} تا از کشورها، مستعمراتی را تشکیل می‌دهند که هر کدام به یک امپراتوری تعلق دارند. برای تقسیم مستعمرات اولیه بین استعمارگرها، به هر استعمارگر، تعدادی از مستعمرات را که این تعداد متناسب با قدرت آن است، می‌دهیم. برای انجام این کار با داشتن هزینه همه استعمارگرها، هزینه نرمالیزه آن‌ها را به صورت زیر در نظر می‌گیریم.

$$C_n = \max_i \{c_i\} - c_n \quad (4)$$

که در آن C_n ، هزینه استعمارگر \ln ، $\max_i \{c_i\}$ بیشترین هزینه میان استعمارگرها و C_n ، هزینه نرمالیزه شده این استعمارگر خاص است. هر استعمارگر که دارای هزینه بیشتری باشد (استعمارگر ضعیف‌تری باشد)، هزینه نرمالیزه کمتری خواهد داشت. با داشتن هزینه نرمالیزه، قدرت نسبی نرمالیزه هر استعمارگر به صورت زیر محاسبه شده و بر مبنای آن، کشورهای مستعمره بین استعمارگرها تقسیم می‌شوند.

$$p_n = \frac{C_n}{\sum_{i=1}^{N_{imp}} C_i} \quad (5)$$

از دیدگاه دیگر، قدرت نرمالیزه شده یک استعمارگر، نسبت مستعمراتی است که توسط آن استعمارگر اداره می‌شود؛ بنابراین، تعداد اولیه مستعمرات یک استعمارگر برابر خواهد بود با:

$$N.C_n = \text{round}\{p_n . (N_{col})\} \quad (6)$$

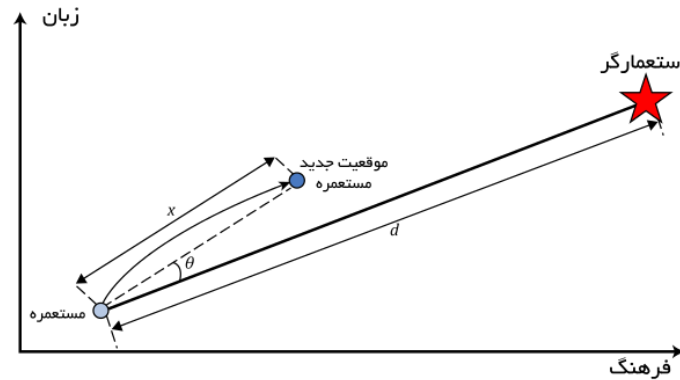
که در آن $N.C_n$ ، تعداد اولیه مستعمرات یک امپراتوری و N_{col} نیز تعداد کل کشورهای مستعمره موجود در جمعیت کشورهای اولیه است. round نیز تابعی است که نزدیک‌ترین عدد صحیح به یک عدد اعشاری را می‌دهد. با در نظر گرفتن $N.C_n$ برای هر امپراتوری، این تعداد از کشورهای مستعمره اولیه را به صورت تصادفی انتخاب کرده و به استعمارگر \ln می‌دهیم. با داشتن حالت اولیه تمام امپراتوری‌ها، الگوریتم رقابت استعماری شروع می‌شود. روند تکامل در یک حلقه قرار دارد که تا برآورده شدن یک شرط توقف، ادامه می‌یابد.

سیاست همگون‌سازی (جذب) با هدف استعمار بیشتر انجام می‌گیرد. این بخش از فرایند استعمار در الگوریتم بهینه‌سازی، به صورت حرکت مستعمرات به سمت کشور استعمارگر مدل شده است.

کشور استعمارگر، کشور مستعمره را در راستای محورهای فرهنگ و زبان به سمت خود جذب می‌کند. همان‌گونه که در شکل ۳ نشان داده شده است، کشور مستعمره به اندازه x واحد در جهت خط واصل مستعمره به استعمارگر، حرکت کرده و به موقعیت جدید کشانده می‌شود. در این شکل، فاصله میان استعمارگر و مستعمره با d نشان داده شده است. x نیز عددی تصادفی با توزیع یکنواخت است؛ یعنی برای x داریم:

$$x \sim U(0, \beta \times d) \quad (7)$$

که در آن، β عددی بزرگ‌تر از ۱ و نزدیک به ۲ است. یک انتخاب مناسب می‌تواند $\beta = 2$ باشد. وجود ضریب $\beta > 1$ باعث می‌شود کشور مستعمره در حین حرکت به سمت کشور استعمارگر، از جهت‌های مختلف به آن نزدیک شود.



شکل ۳. حرکت واقعی مستعمرات به سمت استعمارگر

با بررسی تاریخی پدیده همگون‌سازی، یک حقیقت آشکار در این زمینه این است که علی‌رغم اینکه کشورهای استعمارگر به‌طور جدی پیگیر سیاست جذب بودند اما وقایع به‌طور کامل مطابق سیاست اعمال‌شده آن‌ها پیش نمی‌رفت و انحرافات در نتیجه کار وجود داشت. در الگوریتم معرفی‌شده، این انحراف احتمالی با افزودن یک زاویه تصادفی به مسیر جذب مستعمرات انجام می‌گیرد. بدین منظور، در حرکت مستعمرات به سمت استعمارگر، زاویه تصادفی را نیز به جهت حرکت مستعمره اضافه می‌کنیم. شکل ۳ این حالت را نشان می‌دهد. بدین منظور، این بار به‌جای حرکت به اندازه x به سمت کشور استعمارگر و در جهت بردار واصل مستعمره به استعمارگر، به همان میزان ولی با انحراف θ در مسیر به حرکت خود ادامه می‌دهیم. θ را به‌صورت تصادفی و با توزیع یکنواخت در نظر می‌گیریم؛ بنابراین:

$$\theta \sim U(-y, y) \quad (8)$$

در این رابطه، y پارامتری دلخواه بوده که افزایش آن باعث افزایش جستجوی اطراف امپریالیست شده و کاهش آن نیز باعث می‌شود مستعمرات تا حد ممکن، به بردار واصل مستعمره به استعمارگر، نزدیک حرکت کنند. با در نظر گرفتن واحد رادیان برای θ ، عددی نزدیک به $\pi/4$ در اکثر پیاده‌سازی‌ها، انتخاب مناسبی بوده است.

ایجاد مسیرهای الگوریتم رقابت استعماری

در مرحله بعد بر اساس الگوریتم رقابت استعماری، تلاش می‌شود مسیر بهینه انتشار اطلاعات شناسایی گردد. برای این کار، هر گره در شبکه اجتماعی به‌عنوان یک گره در الگوریتم رقابت استعماری در نظر گرفته می‌شود. در این روش پیشنهادی، یک تابع اکتشافی خاص برای محاسبه میزان سودمند بودن گره بعدی ارائه شده است. در ارائه این تابع، هم میزان مناسب بودن آن گره و هم میزان افزونگی آن با گره‌های پیش‌تر انتخاب‌شده تأثیر داده شده است. بدین ترتیب، گره‌های نامرتب و دارای افزونگی، شانس کمتری برای انتخاب شدن خواهند داشت. تابع اکتشافی ارائه‌شده به‌صورت رابطه ۹ محاسبه می‌شود:

$$\eta(F_i, VF_k) = \left[FS(F_i) - \frac{1}{|VF_k|} \sum_{F_x \in VF_k} sim(F_i, F_x) \right] \quad (9)$$

که $FS(F_i)$ نشان‌دهنده امتیاز فیشر برای گره F_i ، $sim(F_i, F_x)$ نشان‌دهنده شباهت مابین گره F_i و F_x و $|VF_k|$ بیانگر تعداد گره‌های انتخاب‌شده توسط گره k تا این لحظه خواهد بود. همان‌طور که مشخص است، در قسمت ابتدایی تابع اکتشافی ارائه‌شده، میزان مناسب بودن گره در قسمت دوم آن میزان افزونگی آن با گره‌های پیش‌تر انتخاب‌شده در نظر گرفته می‌شود؛ بنابراین، ترکیب این دو قسمت با هم سبب انتخاب گره‌هایی با بیشترین ارتباط با کلاس هدف

و کمترین افزونگی می‌شود. لازم به ذکر است، به‌منظور نرمال‌سازی مقادیر امتیاز فیشر به‌دست‌آمده برای هر گره در بازه صفر و یک، از مقیاس‌گذاری غیرخطی استفاده می‌شود.

بررسی شرط خاتمه

در این قدم از الگوریتم، تعداد تکرارهای الگوریتم بهینه‌سازی بررسی می‌شود. اگر تعداد تکرارها به یک مقدار از پیش تعیین‌شده رسیده باشد، به قدم بعد می‌رود. در غیر این صورت، چرخه الگوریتم تکرار شده و قدم چهارم یعنی ایجاد مسیر اجرا می‌شود.

یافته‌ها

برای انجام آزمایش‌ها، از چهار مجموعه داده باشگاه کاراته (Zachary, 1997)، شبکه دلفین (Lusseau et al., 2003)، کتاب‌های سیاسی (Krebs, 2004) و فوتبال کالج آمریکا (Girvan & Newman, 2002) استفاده شده است.

جدول ۱. جریئات مجموعه داده

| توضیحات شبکه | تعداد لبه‌ها | تعداد نودها | شبکه |
|-----------------------|--------------|-------------|----------------|
| باشگاه کاراته زکریا | ۷۸ | ۳۴ | کاراته |
| شبکه اجتماعی دلفین | ۱۵۹ | ۶۲ | دلفین |
| کتاب‌های سیاسی آمریکا | ۴۴۱ | ۱۰۵ | کتاب‌های سیاسی |
| کالج فوتبال آمریکایی | ۶۱۳ | ۱۱۵ | فوتبال |

مقایسه و ارزیابی سیستم پیشنهادی

به‌منظور ارزیابی روش پیشنهادی، عملکرد الگوریتم پیشنهادی را با الگوریتم لی (Li et al., 2016) در مقاله پایه در نظر گرفتیم و الگوریتم ژنتیک در چهار مجموعه داده مقایسه شد.

جدول ۲. نتایج حاصل از الگوریتم‌های انتشار اطلاعات از نظر میزان دقت

| روش پیشنهادی | الگوریتم ژنتیک | مقاله پایه | شبکه |
|--------------|----------------|------------|----------------|
| ۰/۷۵ | ۰/۶۱ | ۰/۶۴ | کاراته |
| ۰/۸۱ | ۰/۷۳ | ۰/۷۶ | دلفین |
| ۰/۸۸ | ۰/۷۸ | ۰/۸۴ | کتاب‌های سیاسی |
| ۰/۸۹ | ۰/۷۵ | ۰/۸۱ | فوتبال |

جدول ۲، نتایج حاصل از الگوریتم‌های انتشار اطلاعات از نظر میزان دقت در ۱۰۰ اجرا است که به‌طور مستقل مقارنه‌ای اولیه شده‌اند. از تجزیه و تحلیل این جدول نتیجه می‌گیریم که استفاده از الگوریتم پیشنهادی منجر به نتایج بهتری شده است. عملکرد الگوریتم بر روی مجموعه داده آزمایش شده و بهبود مطلوبی داشته است. همان‌طور که در این جدول دیده می‌شود، روش پیشنهادی در میان سه الگوریتم مختلف، بهترین عملکرد را دارد. دلیل آن این است که در روش پیشنهادی، هم از معیار نفوذ و مرکزیت گره‌ها استفاده شده و هم با استفاده از الگوریتم بهینه‌سازی رقابت استعماری، مسیر بهینه انتشار اطلاعات شناسایی می‌شود. شرایط توقف بعد از اتمام بهینه‌سازی همه ستون‌های یک پوشش بررسی می‌شود که در نتیجه، یک پیمانه بهبودیافته خوب بعد از هر مجموعه تکرار داریم.

ارزیابی کیفیت

در این بخش، کیفیت جوامع ایجادشده توسط الگوریتم ارائه‌شده تجزیه و تحلیل می‌شود. برای ارزیابی کیفیت نتایج از معیاری به نام اطلاعات متقابل نرمال‌شده یا NMI که در الگوریتم لی (Li et al., 2016) معرفی شده و ریشه در نظریه اطلاعات دارد، استفاده می‌شود.

در این معیار، فرض بر این است که با توجه به گراف G یک حقیقت ضمنی به منظور انتشار اطلاعات در دسترس است. A به عنوان یک ساختار مسیر انتشار اطلاعات درست در G است که G از C_A مسیر تشکیل شده است. یک الگوریتم انتشار اطلاعات α در نظر گرفته می‌شود. α را در G اجرا کرده و فرض کنیم که ساختار مسیر B متشکل از C_B مسیر است. ماتریس CM با ابعاد $C_A \times C_B$ تعریف می‌شود (که ماتریس سردرگمی^۱ نامیده می‌شود) به طوری که در هر سطر از CM متناظر با یک مسیر در α و هر ستون CM مربوط به یک مسیر در B است. عنصر عمومی CM_{ij} برابر است با تعداد عناصر مسیرهای واقعی i ام که در حال حاضر در مسیر j ام توسط الگوریتم قرار گرفته است. اطلاعات متقابل نرمال‌شده توسط این تعریف شروع شده و به این صورت بیان می‌شود:

$$NMI(A, B) = \frac{-2 \sum_{i=1}^{C_A} \sum_{j=1}^{C_B} N_{ij} \log \left(\frac{N_{ij} N}{N_i \cdot N_j} \right)}{\sum_{i=1}^{C_A} N_i \log \left(\frac{N_i}{N} \right) + \sum_{j=1}^{C_B} N_j \log \left(\frac{N_j}{N} \right)} \quad (10)$$

نتایج به دست آمده در جدول ۳ نشان داده شده است. مقدار NMI در محدوده ۰ تا ۱ است و اعداد بالاتر نمایانگر عملکرد بهتر الگوریتم است. از تجزیه و تحلیل این جدول نتیجه می‌گیریم که استفاده از الگوریتم پیشنهادی، منجر به نتایج بهتری نسبت به الگوریتم‌های مورد مقایسه شده است.

با توجه به نتایج به دست آمده بر اساس معیارهای متفاوت، همان‌طور که در جدول ۲ و ۳ نشان داده شده است، روش الگوریتم ژنتیک در مقایسه با سایر روش‌ها، دارای مقدار دقت پایین‌تر نسبت به سایر روش‌ها است. دلیل این امر این است که در الگوریتم ژنتیک از روش ماتریس فضای مسئله مدل‌سازی می‌شود و در دو روش دیگر این کار بر اساس روش گرافی مسئله مدل‌سازی شده است که دقت بالاتری دارد. همچنین در روش پیشنهادی که هم از مدل‌سازی گرافی و هم از روش بهینه‌سازی تکاملی استفاده شده، نتایج بهبود داده شده است.

جدول ۳. معیار NMI به دست آمده توسط هر الگوریتم برای هر یک از این مجموعه داده‌ها

| روش پیشنهادی | الگوریتم ژنتیک | مقاله اصلی | شبکه |
|--------------|----------------|------------|----------------|
| ۰/۷۷ | ۰/۶۹ | ۰/۷۲ | کاراته |
| ۰/۸۲ | ۰/۷۸ | ۰/۷۹ | دلفین |
| ۰/۸۸ | ۰/۸۱ | ۰/۸۶ | کتاب‌های سیاسی |
| ۰/۹۱ | ۰/۸۲ | ۰/۸۵ | فوتبال |

از دیگر معیارهایی که می‌توان روش پیشنهادی را با سایر روش‌ها مقایسه کرد، معیار MAE و $MAUE$ است. در معیار خطا، اختلاف مقادیر پیش‌بینی شده نسبت به مقدار واقعی آن‌ها محاسبه می‌شود. واضح است که هر چه خطاهای موردنظر برای یک سیستم انتشار اطلاعات دارای مقدار کمی باشند، نرخ‌های پیش‌بینی شده به نرخ‌های واقعی کاربر نزدیک‌تر بوده و کارایی سیستم افزایش خواهد یافت. از جمله این معیارها می‌توان به معیار میانگین خطای مطلق^۲ و

1. confusion matrix

2. Mean absolute error (MAE)

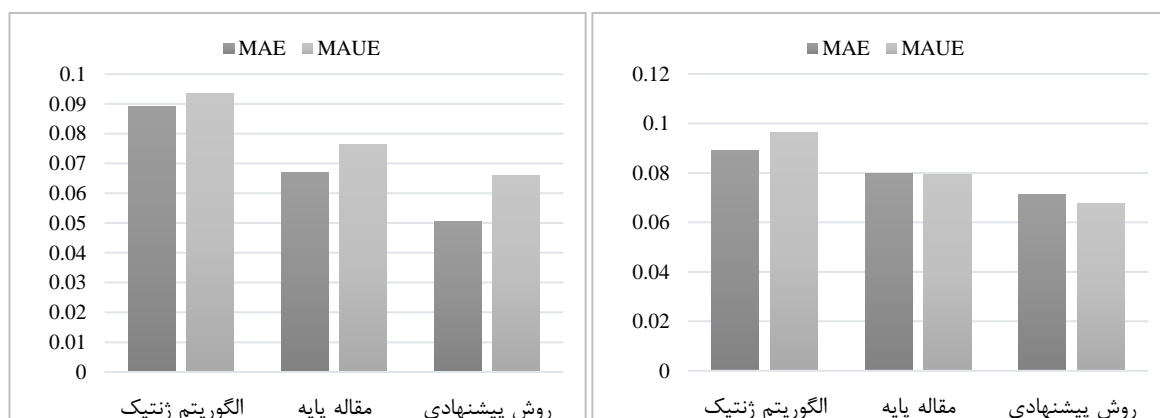
میانگین خطای مطلق کاربر^۱ جهت اندازه‌گیری دقت پیش‌بینی رتبه‌ها اشاره نمود که در ادامه روابط مربوط به هر یک را نشان خواهیم داد.

$$MAE = \frac{\sum_u \sum_i |r_{u,i} - p_{u,i}|}{N} \quad (۱۱)$$

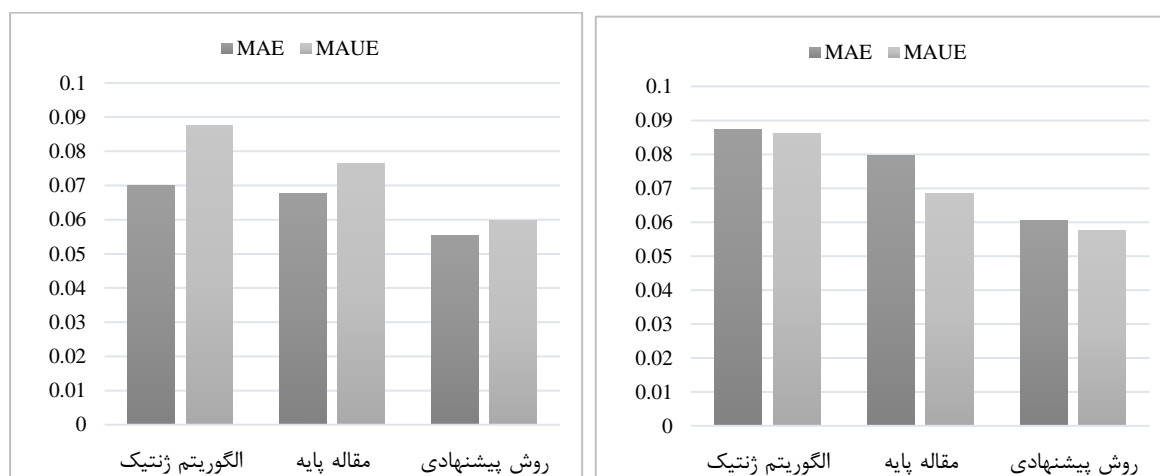
در رابطه فوق، MAE میانگین خطای مطلق، N نشان‌دهنده تعداد گره‌های موجود در شبکه، $r_{u,i}$ مقدار انتشار واقعی گره u از مسیر i و $p_{u,i}$ مقدار انتشار پیش‌بینی شده برای مسیر i توسط سیستم انتشار اطلاعات است.

$$MAUE = \frac{\sum_{u \in U} MAE_u}{N} \quad (۱۲)$$

همچنین در این رابطه، $MAUE$ میانگین خطای مطلق کاربر، N نشان‌دهنده تعداد گره‌های موجود در شبکه و MAE_u میانگین خطای مطلق برای گره u بوده که توسط رابطه ۱۲ برای هر گره u به صورت جدا محاسبه می‌گردد. در نمودار ۱ و ۲، نتایج عملکرد روش پیشنهادی بر روی مجموعه داده‌های مختلف بر اساس معیارهای MAE و $MAUE$ نشان داده شده است. روش پیشنهادی در مقایسه با سایر روش‌ها از MAE و $MAUE$ کمتری برخوردار است. این مسئله نشان‌دهنده این موضوع است که روش ارائه شده دارای خطای کمتری در پیش‌بینی انتشار اطلاعات بوده و توانسته مسیرهای انتشار اطلاعات را با دقت بیشتری نسبت به سایر روش‌ها پیش‌بینی کند.



نمودار ۱. نتایج ارزیابی روش پیشنهادی با توجه به معیارهای ارزیابی MAE و $MAUE$ بر روی مجموعه داده کاراته و دلفین



نمودار ۲. نتایج ارزیابی روش پیشنهادی با توجه به معیارهای ارزیابی MAE و $MAUE$ بر روی مجموعه داده کتاب و فوتبال

بحث و نتیجه‌گیری

با گسترش شبکه‌های اجتماعی، نیاز به تحلیل ساختارها و رفتارهای آن‌ها به‌عنوان یکی از نیازمندی‌های اساسی شرکت‌های تجاری به وجود آمده است. تحلیل شبکه‌های اجتماعی در بسیاری از کاربردها از جمله مدیریت شبکه اجتماعی، تحلیل گرایش بازار، شناسایی افراد تأثیرگذار و حامیان و ارتقای کارایی سامانه‌های توصیفگر قابل استفاده است. یکی از مسائل موجود در زمینه انتشار اطلاعات، بررسی میزان انتشار در چنین سیستم‌هایی است. یکی از مشکلات راهکارهای قبلی این است که در شبکه‌های با ابعاد بالا، پیچیدگی محاسباتی افزایش یافته است. یکی از روش‌های مقابله با این مشکل، استفاده از الگوریتم‌های تکاملی است. در این مقاله با استفاده از ترکیب الگوریتم رقابت استعماری و معیار مرکزیت گره، یک روش جدید برای پیش‌بینی انتشار اطلاعات ارائه شده است. در روش پیشنهادی، ابتدا میزان مرکزیت گره‌ها برای تمام گره‌های شبکه محاسبه شده و سپس بر اساس فاصله بین کاربران شبکه اجتماعی و سرعت انتشار اطلاعات بین کاربران با استفاده از الگوریتم رقابت استعماری، مسیر بهینه برای انتشار تعیین شده و بر اساس آن به پیش‌بینی انتشار اطلاعات پرداخته می‌شود. در واقع، در الگوریتم رقابت استعماری بر اساس نحوه حرکت مستعمره‌ها به سمت استعمارگر، مسیر بهینه برای انتشار اطلاعات در شبکه شناسایی می‌شود. در این مقاله، چهار مجموعه داده مختلف مورد بررسی قرار گرفت. بررسی نتایج نشان می‌دهد روش پیشنهادی منجر به نتایج بهتری شده است. عملکرد الگوریتم با توجه به معیاری‌های دقت و NMI ارزیابی شده است که نتایج مطلوبی را نشان می‌دهد.

برای بهبود کارایی روش پیشنهادی، در ادامه راهکارهایی برای کارهای آتی پیشنهاد می‌گردد. می‌توان از سایر الگوریتم‌های تکاملی مانند الگوریتم کلونی مورچه‌ها، الگوریتم بهینه‌سازی فاخته و الگوریتم زنبور عسل استفاده کرد. با توجه به اینکه هر الگوریتم تکاملی دارای یک سری نقاط قوت و نقاط ضعف است، پیشنهاد می‌گردد در کارهای آتی از ترکیب دو یا چند الگوریتم تکاملی برای بهبود کارایی انتشار اطلاعات بهره گرفت. به‌عنوان مثال، می‌توان از ترکیب الگوریتم کلونی زنبور عسل و رقابت استعماری استفاده کرد که دارای هر دو فاکتور دقت و سرعت در کنار یکدیگر باشند.

منابع

- Alahakoon, T., Tripathi, R., Kourtellis, N., Simha, R., & Iamnitchi, A. (2011). K-path centrality: A new centrality measure in social networks. In *Proceedings of The 4th Workshop on Social Network Systems* (pp. 1-6).
- Cai, Y., Minjie, Z., & Hao, C. (2010). A hybrid chaotic quantum evolutionary algorithm. In *2010 IEEE International Conference on Intelligent Computing and Intelligent Systems* (Vol. 2, pp. 771-776). IEEE.
- De Meo, P., Ferrara, E., Fiumara, G., & Ricciardello, A. (2012). A novel measure of edge centrality in social networks. *Knowledge-Based Systems*, 30, 136-150.
- Dorigo, M. (2007). Ant colony optimization. *Scholarpedia*, 2(3), 1461.
- Girvan, M., & Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of The National Academy of Sciences*, 99(12), 7821-7826.
- Hansen, D., Shneiderman, B., & Smith, M. A. (2010). *Analyzing social media networks with NodeXL: Insights from a connected world*. Morgan Kaufmann.
- Hosseini, S., & Al Khaled, A. (2014). A survey on the imperialist competitive algorithm metaheuristic: implementation in engineering domain and directions for future research. *Applied Soft Computing*, 24, 1078-1094.
- Jiang, Y., Jia, C., & Yu, J. (2013). An efficient community detection method based on rank centrality. *Physica A: Statistical Mechanics And Its Applications*, 392(9), 2182-2194.
- Jin, L., Chen, Y., Wang, T., Hui, P., & Vasilakos, A. V. (2013). Understanding user behavior in online social networks: A survey. *IEEE Communications Magazine*, 51(9), 144-150.
- Katarya, R. (2018). Movie recommender system with metaheuristic artificial bee. *Neural Computing and Applications*, 30(6), 1983-1990.

- Krebs, V. (2004). Books about US politics. <http://www.orgnet.com>.
- Li, C. T., Shan, M. K., Jheng, S. H., & Chou, K. C. (2016). Exploiting concept drift to predict popularity of social multimedia in microblogs. *Information Sciences*, 339, 310-331.
- Li, Y., & Li, J. (2010). Swarm intelligence optimization algorithm based on orthogonal optimization. In *2010 Second International Conference on Computer Modeling and Simulation* (Vol. 4, pp. 12-16). IEEE.
- Lusseau, D., Schneider, K., Boisseau, O. J., Haase, P., Slooten, E., & Dawson, S. M. (2003). The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations: can geographic isolation explain this unique trait?. *Behavioral Ecology and Sociobiology*, 54, 396-405.
- Moradi, P., & Rostami, M. (2015). A graph theoretic approach for unsupervised feature selection. *Engineering Applications of Artificial Intelligence*, 44, 33-45.
- Newman, M. E. (2005). A measure of betweenness centrality based on random walks. *Social networks*, 27(1), 39-54.
- Noh, J. D., & Rieger, H. (2004). Random walks on complex networks. *Physical review letters*, 92(11), 118701.
- Park, S. J., Lim, Y. S., & Park, H. W. (2015). Comparing Twitter and YouTube networks in information diffusion: The case of the "Occupy Wall Street" movement. *Technological forecasting and social change*, 95, 208-217.
- Pellegrina, L. (2023). Efficient centrality maximization with rademacher averages. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 1872-1884).
- Qi, X., Fuller, E., Wu, Q., Wu, Y., & Zhang, C. Q. (2012). Laplacian centrality: A new centrality measure for weighted networks. *Information Sciences*, 194, 240-253.
- Yan, X., Zhai, L., & Fan, W. (2013). C-index: A weighted network node centrality measure for collaboration competence. *Journal of Informetrics*, 7(1), 223-239.
- Zachary, W. W. (1977). An information flow model for conflict and fission in small groups. *Journal of anthropological research*, 33(4), 452-473.
- Zhou, K., Martin, A., & Pan, Q. (2015). A similarity-based community detection method with multiple prototype representation. *Physica A: Statistical Mechanics and its Applications*, 438, 519-531.

