فصلنامه سیستم‌های پردازشی و ارتباطی چندرسانه‌ای هوشمند
Intelligent Multimedia Processing and Communication Systems (IMPCS)

Fall 2022, 3 (3), 29-39
DOR: 20.1001.1.27832570.1400.2.2.6.9

Received: 29 June 2022
Accepted: 13 Aug 2022

مقاله پژوهشی

Presenting a model for Multi-layer Dynamic Social Networks to discover Influential Groups based on a combination of Developing Frog-Leaping Algorithm and C-means Clustering

Lida Naderlou[1], Mohammad Tahghighi Sharabyan[2]*

1. MSc, Artificial Intelligence, Roozbeh Institute of Higher Education, Zanjan, Iran.
lidanaderlou@gmail.com
2. Assistant professor, Faculty of Electrical and Computer Engineering, Zanjan Branch, Islamic Azad University, Zanjan, Iran.*(Corresponding Author)* mtahghighi@yahoo.com

Abstract

Introduction: The current research examines a more complex social network called a multi-layered social network. Recently, the concept of the multilayer network has emerged from the area of complex networks, under the domain of complex systems. In the field of big data, simple and multi-layered social networks can be found everywhere and in all fields. The estimation of the importance of each node in these two types of networks is not the same, and weighting the nodes is very necessary to control the network. For this purpose, the relationship between the characteristics of the nodes and the relationship with the network structure should be examined. To find the degree of each node in the system function, parameters like reliability, controllability, and power should be considered. In this paper, a model for dynamic multi-layer social networks to discover influential groups, based on the combination of evolutionary frog jump algorithm and C-means clustering, has been presented.

Method: Once data are collected, they were cleaned and normalized so that the desired data leads to the identification of effective individuals and groups. The decision matrix is constructed based on these data. Using this matrix, identification, and clustering (based on fuzzy clustering) were conducted, and the importance of the groups was also determined to determine influential people and groups in social networks. The Jumping frog algorithm was used to improve the detection of influential parameters.

Results: In the evaluation and simulation of the clustering part, the proposed method was compared with the K-means method and the balance value of the method in cluster selection was 5. It should be noted that the proposed method showed better improvement compared to the compared methods. Also, the evaluation of the accuracy criterion of the proposed method has improved by 3.3 compared to similar methods and recorded an improvement of 3.3 compared to the basic M-ALCD method.

Discussion: In this paper, a multi-layer local community detection model is proposed, which is based on structure and feature information. This model can use the information of the characteristics of the nodes and the information of the strength of similarity that is revealed by social exchanges and improves the accuracy of community detection in Improve multilayer networks. Due to its modularity and computational efficiency, this algorithm performs better on most data sets, unlike the classic multi-layer and global community detection algorithms.

Keywords: Multi-layer Dynamic, Social Networks, Influential Groups, Developing Frog-Leaping Algorithm, C-means clustering.

فصلنامه سیستم‌های پردازشی و ارتباطی چندرسانه‌ای هوشمند
Intelligent Multimedia Processing and Communication Systems (IMPCS)
انجمن علمی تجارت الکترونیکی ایران

ارائه مدلی برای شبکه‌های اجتماعی چندلایهٔ پویا برای کشف گروه‌های تأثیرگذار، مبتنی‌بر ترکیب الگوریتم تکاملی جهش قورباغه و خوشه‌بندی C-means

تاریخ دریافت: ۱۴۰۱/۰۴/۰۸
تاریخ پذیرش: ۱۴۰۱/۰۵/۲۲

سال سوم، پاییز ۱۴۰۱
شمارهٔسوم ، صص: ۲۹-۳۹

لیدا ندرلو[۱]، محمد تحقیقی شربیان[۲*]

۱. دانشجوی کارشناسی ارشد ،گرایش هوش مصنوعی، مؤسسه آموزش عالی روزبه، زنجان، ایران.
lidanaderlou@gmail.com
۲. استادیار، دانشکده مهندسی برق و کامپیوتر، واحد زنجان، دانشگاه آزاد اسلامی، زنجان، ایران.(نویسنده مسئول)
mtahghighi@yahoo.com

چکیده: امروزه علم و فن‌آوری با آهنگی شتابناک در حال رشد است و شبکه‌های اجتماعی پیچیده به بخشی ضروری از زندگی تبدیل‌شده‌اند، آن-گونه‌که بحث جدایی مردم از شبکه‌های درهم پیچیده‌ای که مبتنی‌بر نیازهای اساسی زندگی است بحث ناگزیری در زندگی روزمره و عرصه دانش است. در پژوهش پیش رو مدلی برای شبکه‌های اجتماعی چندلایه‌ای پویا برای کشف گروه‌های تأثیرگذار، مبتنی‌بر ترکیبب الگوریتم تکاملی جهش قورباغه و خوشه‌بندی C-means ارائه‌شده.است بدین ترتیب که پس از جمع‌آوری داده‌ها به پاکسازی و نرمالسازی آن‌ها پرداخته شد تا داده‌های مطلوب منجر به شناسایی افراد و گروه‌های مؤثر شود که در ادامه‌ی کارماتریس تصمیم شکل‌گرفت و از روی آن شناسایی و خوشه‌بندی(مبتنی‌بر خوشه‌بندی فازی) انجام‌شد و اهمیت گروه‌ها نیز مشخص‌گردید. برای دستیابی به افراد و گروه‌های تأثیرگذار در شبکه‌های اجتماعی، از الگوریتم قورباغه جهنده برای بهبود تشخیص پارامترهای تأثیرگذار استفاده‌شد، که باعث بهبود اهمیت گره‌ها شده‌است. در ارزیابی و شبیه‌سازی بخش خوشه‌بندی، روش پیشنهادی با روش K-means مقایسه و نتیجه‌ی مقدار تعادل روش در انتخاب خوشه برابر ۵ شد. گفتنی است که روش پیشنهادی به نسبت روش‌های مورد مقایسه بهبودی مناسب‌تری را نشان‌داد. همچنین ارزیابی معیار صحت روش پیشنهادی به نسبت روشهای همسان بهبود ۳,۳ داشته و نسبت به روش پایه M-ALCD بهبود ۳,۸ را به ثبت رسانده‌است.

واژه‌های کلیدی: شبکه‌های اجتماعی چندلایهٔ پویا، گروه‌های تأثیرگذار، الگوریتم تکاملی جهش قورباغه و خوشه‌بندی C-means.

۱. مقدمه

یک شبکه اجتماعی ساختاری است که با مشارکت گروهی از افراد در عرصه فعالیت‌های اجتماعی شکل‌می‌گیرد و میان این فعالان گره‌های هم‌بستگی ایجادمی‌شود. شبکه اجتماعی، روش‌های زیادی برای تحلیل ساختار ویژگی‌های اجتماعی با الگوهایی که براساس تئوری هر شبکه اجتماعی تولیدمی‌شود، ایجادمی‌نماید[۱]. پژوهش حاضر شبکه اجتماعی پیچیده‌تری را به‌نام شبکه اجتماعی چندلایه بررسی‌می‌کند. اخیراً مفهوم شبکه چندلایه از ناحیه شبکه‌های پیچیده، زیردامنه حوزه سیستم‌های پیچیده پدیدآمده‌است و مانند یک زمین حاصلخیز، بستر مناسبی برای تحقیقات جدید است. در زمینه داده‌های عظیم، شبکه-های اجتماعی ساده و چندلایه هرجا و در هر زمینه‌ای یافت‌می‌شوند. تخمین درجه اهمیت هر گره در این شبکه که مشابه هم نیستند و وزن‌دهی به گره‌ها برای کنترل شبکه بسیار ضروری است. برای این منظور باید ارتباط میان ویژگی‌های گره‌ها و ارتباط با ساختار شبکه را بررسی‌نمود. برای یافتن درجه هر گره در تابع سیستم نیزمواردی مانند قابلیت اطمینان، کنترل و توان باید لحاظ‌شوند[۲].

برای تشخیص گره مهم (تأثیرگذار) در شبکه‌های اجتماعی ساده روش‌های بسیار زیادی پیشنهادشده و درمقابل نتیجه تحقیقات اخیر در شناسایی توزیع کنندگان مؤثر در شبکه‌های اجتماعی چندلایه پویا، بحث گسترده‌ای را برانگیخته و موضوع داغ این روزها می‌باشد. موضوع اصلی برای ارائه راه‌حل درست شناسایی گره‌های مهم در شبکه‌های چندلایه، درنظرگرفتن پارامتری برای آن گره تأثیرگذار است. درواقع، شناسایی گره تأثیرگذار (مهم) می‌تواند با استفاده از الگوریتم‌های تکاملی الهام‌گرفته از طبیعت، انجام‌پذیرد. هدف اصلی کار، بیشینه‌سازی تقدم کاربران در شبکه‌های اجتماعی می‌باشد[۳]. لذا در این پژوهش برای محاسبه مقدمات کاربران و تأثیرشان در شبکه و درنتیجه بیشینه‌سازی و تشخیص آن‌ها از ترکیب الگوریتم-های تکاملی و داده‌کاوی استفاده‌می‌شود تا به این سؤال پاسخ‌دهد که آیا ارائه مدلی برای شبکه‌های اجتماعی چندلایه‌ای پویا برای کشف گروه‌های تأثیرگذار مبتنی‌بر ترکیب الگوریتم تکاملی جهش قورباغه و خوشه‌بندی C-means امکان‌پذیر است؟

۲ . پیشینه پژوهش

Lei و همکارانش سال ۲۰۱۹ پژوهشی با عنوان «مدل تشخیص جامعه شبکه چندلایه برپایه ویژگی‌ها و شدت تعامل اجتماعی» ارائه‌داده‌اند و معتقدند که روابط پیچیده و پیوندهای بین کاربران مختلف، به‌دست-آوردن اطلاعات در کل یک سیستم را دشوارتر می‌کند، حتی با وجود ابزارهای تجمیع قدرتمند ، بیشتر شرکت‌ها نمی‌توانند هزینه‌های انسانی و مالی مربوطه را بپردازند. با این حال، روش‌های تشخیص شبکه چندلایه‌ی موجود برای تشخیص جامعه مبتنی‌بر ویژگی مناسب است بنابراین ، این مقاله یک مدل تشخیص جامعه محلی چندلایه ارائه‌می‌دهد که براساس اطلاعات ویژگی و ساختار است. این مدل می-تواند به‌نحوی مؤثر از اطلاعات ویژگی گره و اطلاعات قدرت شباهت آشکارشده توسط مبادلات اجتماعی برای بهبود دقت تشخیص جامعه در شبکه‌های چندلایه استفاده‌کند. این الگوریتم برخلاف ریتم‌های کلاسیک تشخیص چندلایه و جهانی جامعه به‌دلیل مدولار بودن و کارایی محاسباتی در بیشتر مجموعه‌های داده قوی است. این پژوهش یک الگوریتم تشخیص جامعه چندلایه مبتنی‌بر ویژگی (M-ALCD) را پیشنهادمی‌کند که می‌تواند از مزیت کامل ساختار و پارتیشن شبکه اطلاعات ویژگی استفاده‌کند، تا شبکه‌های اجتماعی چندلایه را ارتقا بخشد. الگوریتم M-ALCD می‌تواند با موفقیت شبکه‌هایی با اتصال‌های پراکنده و سطوح بالای نویز که نمونه‌ای از مجموعه داده‌های واقعی هستند را مدیریت‌کند. این الگوریتم نتایج دقیقی را به-دست‌می‌آورد و می‌تواند برای تشخیص مؤثرتر ساختار جامعه فازی یک شبکه استفاده‌شود؛ با این حال، در برخی موارد، نتایج آزمایش‌ها از نتایج به‌دست‌آمده از الگوریتم کلاسیک M-ALCD ضعیف‌تر و از نتایج حاصل از الگوریتم‌های تک‌لایهٔ سنتی LCD و لووین بهتر بودند. از نظر زمان ، زمان اجرای الگوریتم M-ALCD نسبت به سه الگوریتم دیگر در مجموعه داده حاوی ویژگی‌های رفتاری بهتر بود. اما برای مجموعه داده‌ها با لایه‌های بسیار زیاد، مانند مجموعه داده خطوط هوایی، الگوریتم M-ALCD کندتر از الگوریتم M-ALCD است. محققان نام-برده امیدوارند در آینده الگوریتم آن‌ها برای نشان‌دادن سطح بالاتری از مقاومت هنگام استفاده از مجموعه داده‌های چندلایه با تعاملات قوی و تکامل پویا تقویت‌شود [۴].

Jain و همکارانش سال ۲۰۱۹ پژوهش «تشخیص رهبر (سرخوشه) با استفاده از الگوریتم بهینه‌سازی نهنگ در شبکه اجتماعی آنلاین» را ارائه‌دادند. در این مقاله، آن‌ها یک شبکه اجتماعی جدید مبتنی‌بر الگوریتم الهام‌گرفته از طبیعت نهنگ را همراه با توابع بهینه‌سازی معیار استاندارد مختلف برای شناسایی رهبران فکری برتر در شبکه‌های اجتماعی پیشنهاد دادند و درابتدا، تابع هدف برای هر کاربر را با استفاده از فاصله و مرکزیت محاسبه و سپس، الگوریتم پارتیشن اجتماعی را برای تعیین جوامع در هر دو مجموعه داده اجراکردند. به-علاوه، تکنیک بهینه‌سازی نهنگ را با بهینه‌سازی‌های مختلف به-کارگرفتند تا رهبر فکری برتر را در حوزه محلی و جهانی تخمین‌بزنند. دارایی اصلی الگوریتم پیشنهادی این است که با افزایش تعداد کاربران، دقت و کارآیی الگوریتم نیز افزایش‌می‌یابد زیرا اطلاعات بیشتری در مورد موقعیت برداری کاربران دیگر انباشته‌شده‌است. محققان مذکور همچنین کد کامل و مجموعه داده‌های الگوریتم پیشنهادی را در لینک GitHub ارئه‌کرده‌اند. رویکرد پیشنهادی شامل ابعاد متعددی است. اولین مورد، بررسی یکی از الگوریتم های الهام‌گرفته از طبیعت است که منجر به کارایی و دقت بهتر می‌شود. مورد دوم، بررسی ویژگی‌های منحصر به فرد کاربران از قبیل وضعیت اجتماعی، فرهنگ، سابقه مالی، دانش فنی، تجربه، ارزش اعتماد، رابطه، هویت جهانی و غیره برای یافتن نتایج صحیح‌تر و دقیق‌تر است. در پایان، متذکرشده‌اند که

علاقه‌مند به پیاده‌سازی رویکرد پیشنهادی خود روی مجموعه داده شبکه اجتماعی پویای بلادرنگ هستند که رفتارهای پویا و کاربران را در طول زمان رصدمی‌کند و تلاش می‌کنند تا مقوله اعتماد را همراه با الگوریتم پیشنهادی خود برای پرورش جنبه‌های بیشتر در تشخیص رهبر فکری در شبکه اجتماعی درنظربگیرند[۵].

Lei و همکارانش سال ۲۰۱۹ پژوهشی با عنوان « تشخیص جوامع همپوشان شبکه‌های اجتماعی براساس روش ترکیبی C- means» ارائه‌دادند. بیشتر الگوریتم‌های تشخیص ساختار جامعه سنتی می‌توانند تمایز بین جوامع را بسیار خوب حل‌کنند. با این حال، بسیاری از این الگوریتم‌ها در تولید نتایج پارتیشن‌بندی واقعی، زمانی که جوامع همپوشان در شبکه‌های پیچیده وجوددارند، ناتوان‌اند. توجه داشته‌باشید که همپوشانی اجتماعات در شبکه‌های پیچیده یک ویژگی مشترک در دنیای واقعی است. در این مقاله برای حل این مشکل، روش خوشه‌بندی ترکیبی برای کشف همپوشانی که معمولاً در شبکه‌های پیچیده وجوددارد، طراحی شده‌است. روش پیشنهادی درجه عضویت فازی هر شیء متعلق به یک جامعه مشخص را با معرفی مجموعه‌های فازی به‌دست‌می‌آورد که برای درک و توصیف ساختار عمیق شبکه‌های همپوشان بسیار مفید است. به‌علاوه، برای به‌دست‌آوردن ساختارهای جوامع فازی مناسب‌تر، یک معیار شباهت طراحی‌کردند که هم اطلاعات جهانی و هم اطلاعات محلی را برای ایجاد رابطه معنی‌دار بین اشیا در شبکه‌های پیچیده فراهم‌می‌سازد. کار آینده به طراحی معیارهای شباهت جدید برای کشف ساختارهای جوامع هندسی متغیر اختصاص‌خواهدیافت و نیز به بررسی امکان طراحی الگوریتم‌های خوشه‌بندی پایدار برای تشخیص خودکار تعداد جوامع معنی‌دار در شبکه‌های پیچیده می‌پردازند[۶]

Farooq و همکارانش سال ۲۰۱۸ پژوهش «تکنیکی برای شناسایی بازیکنان کلیدی که به بهبود کسب‌وکار با استفاده از تحلیل شبکه اجتماعی چندلایه کمک‌می‌کند» ارائه‌دادند. هر روز تعدادی از مردم به شبکه‌های اجتماعی می‌پیوندند. تشخیص گره‌های مؤثر در چنین شبکه‌ای آسان نیست. بنابراین مقاله یادشده طرحی را برای شناسایی گره‌های تأثیرگذار برپایه معیارهای شبکه پیشنهادمی‌دهد شامل: درجه، مرکزیت فیمابینی، مرکزیت نزدیکی، مرکزیت بردار ویژه، رتبه صفحه و خوشه‌بندی. روش پیشنهادی با یک مورد تست آزمایش‌شده‌است. نتایج تجربی و آنالیز کمی جامع نشان‌می‌دهد که این طرح پیشنهادی راهی مؤثرتر برای شناسایی گره‌های تأثیرگذار در یک OSNs بزرگ است[۷].

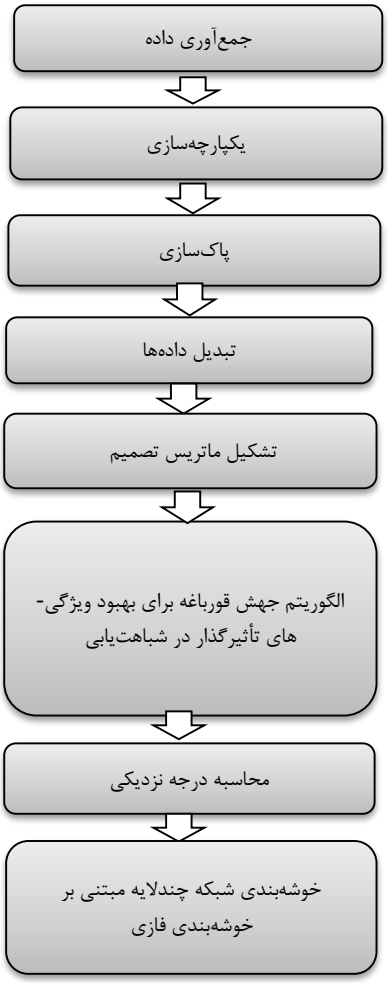
Mitral **و** همکارانش سال ۲۰۱۹ پژوهشی به‌نام «دسته‌بندی افراد تأثیرگذار در شبکه‌های اجتماعی چندلایه» ارائه‌دادند. امروزه، رسانه‌های اجتماعی یکی از حالت‌های رایج تعامل و انتشار اطلاعات هستند؛ معمولاً منبع اصلی انتشار اطلاعات نیز برخی از نهادها هستند که مؤلفه‌های تأثیرگذار نامیده‌می‌شوند. تأثیرگذار یک نهاد یا فرد است که توانایی تأثیرگذاری بر دیگران به‌دلیل رابطه با مخاطبانش را دارد. در این مقاله، پژوهشگران روشی برای طبقه‌بندی مؤلفه‌های تأثیرگذار در شبکه‌های اجتماعی چندلایه پیشنهادمی‌کنند. یک شبکه اجتماعی چندلایه شبیه به یک شبکه اجتماعی تک‌لایه است که چندین ویژگی از یک گره را دربرگرفته و آن‌ها را در چند لایه مدل‌سازی‌می‌کند. روش پیشنهادی ترکیبی از تکنیک‌های یادگیری ماشین (SVM، شبکه‌های عصبی و غیره) با معیارهای مرکزیت است. آن‌ها الگوریتم پیشنهادی را بر روی برخی شبکه‌های واقعی برای تأیید اثربخشی رویکرد خود در سیستم‌های چند لایه آزمایش‌کردند[۸].

Al-garadi و همکارانش سال ۲۰۱۶ پژوهشی با عنوان «پخش‌کننده‌های مؤثر در تعاملات چندلایه شبکه‌های اجتماعی آنلاین» ارائه‌دادند. این مطالعه براساس تحلیل اولیه نتیجه‌گیری کرده‌است که بهبود دقت شناسایی پخش‌کننده‌های تأثیرگذار نه تنها مبتنی‌بر بهبود الگوریتم‌های شناسایی بلکه مبتنی‌بر توسعه یک توپولوژی شبکه است که انتشار اطلاعات را نشان‌می‌دهد. به‌علاوه، تعامل چندلایه بین کاربران و فرآیند گسترش باید به‌دقت بررسی‌شود. دراینجا، پیچیدگی شبکه باید برای درک و پیش‌بینی پویایی انتشار به‌دقت در OSNs مورد توجه قرارگیرد. نتایج حاضل کار نشان‌داده‌است که الگوریتم شناسایی پخش‌کننده واحدی وجودندارد که در تمام انواع شبکه‌های توپولوژیکی به‌خوبی عمل‌کند (یعنی در هر نوع شبکه‌ای بسته به شرایط آن یک الگوریتم خاص نیاز است). لازم است دریابیم که چگونه داده‌های شبکه استخراج‌می‌شوند و چگونه کاربران درون شبکه در تعامل هستند تا بهترین الگوریتم ممکن شناسایی شود. نتایج به‌دست‌آمده نشان داد که نمایش توپولوژیکی از OSNs که هم فعل و انفعالات متقابل و هم وزن تعامل را درنظرمی‌گیرد نتایجی ارائه‌کرده‌است که قابل‌اعتمادترند[۹].

Liu و همکارانش سال ۲۰۱۷ پژوهشی با عنوان « تعبیه شبکه چندلایه» ارائه‌دادند و با گسترش گراف‌کاوی استاندارد به شبکه چندلایه، سه روش «تجمیع شبکه»، «تجمیع نتایج» و «تحلیل مشترک لایه» را برای نگاشت یک شبکه چندلایه به یک فضای برداری پیوسته پیشنهادکردند که بدون استفاده از فعل و انفعالات میان لایه‌ها، تجمیع شبکه و تجمیع نتایج، روش تعبیه شبکه استاندارد را بر روی گراف ادغام‌شده یا هر لایه برای یافتن یک فضای برداری برای شبکه چندلایه اِعمال‌می‌کند؛ از سوی دیگر، برای درنظرگرفتن تأثیر متقابل بین لایه‌ها، «تحلیل مشترک لایه» روش تعبیه شبکه تک‌لایه را به شبکه چندلایه گسترش‌می‌دهد. با معرفی احتمال انتقال لینک براساس فاصله اطلاعات، این روش نه تنها از پیمایش تصادفی مرتبه اول و دوم برای عبور بر روی یک لایه استفاده‌می‌کند، بلکه توانایی عبور بین لایه‌ها با استفاده از تعاملات را نیز دارد. نویسندگان مقاله مذکور با این ارزیابی، ثابت‌کردند که در مقایسه با روش‌های پیش‌بینی لینک منظم، «تحلیل مشترک لایه» بهترین عملکرد را در بیشتر مجموعه داده‌ها به‌دست‌آورده‌است، درحالی‌که «تجمیع شبکه» و «تجمیع نتایج» عملکرد بهتری نسبت به روش‌های پیش‌بینی لینک دارند[۱۰].

۳. روش پیشنهادی

در این بخش، برای یک شبکه‌ی چندلایهٔ واقعی که دارای ارتباطات پراکنده، اختلالات زیاد و نویزهای قابل‌توجه است، ویژگی‌های تمام گره‌های شبکه و قدرت شباهت اجتماعی به‌طور مشترک درنظرگرفته‌شد. این روش اولین مدل تشخیص جامعهٔ محلی براساس ویژگی‌های چندلایه و ساختار جامعه (AM-ALCD) را پیشنهادمی‌دهد. مدلی که توسعه یافتهٔ تشخیص جامعهٔ منطقه‌ای است، به یک معیار جدید از شدت شباهت جامعه متکی است که به‌نوبه خود بر اساس شباهت ساختار جامعه است.



شکل۱: بلوک دیاگرام روش پیشنهادی

برای توصیف ویژگی‌های گروه و قدرت شباهت اجتماعی یک معیار متفاوت تعریف‌می‌شود. به‌کمک استفادهٔ مؤثر از این معیار جدید، یک الگوریتم برای تشخیص جوامع محلی پیشنهادمی‌شود. روش پیشنهادی، شباهت گره را با استفاده از ویژگی‌های گره زندگی واقعی خاصی تعیین‌می‌کند. برای به‌دست‌آوردن سطح شباهت یکپارچه از ارتباط مابین گره‌ها از الگوریتم جهش قورباغه و خوشه‌بندی C-Means استفاده‌می‌شود.

در روش پیشنهادی ابتدا مدل گراف شبکه چندلایه توصیف-می‌شود. سپس مدل AM-ALCD برای تشخیص جامعه محلی در یک شبکه چندلایه پیشنهادمی‌شود. درپایان، چارچوب الگوریتمِ مبتنی‌بر گره‌های seed ورودی ارائه‌می‌شود. بلوک دیاگرام روش پیشنهادی در شکل نمایش‌داده‌شده‌است.

۱،۳. شباهت ویژگی در میان گره‌ها در لایهٔ یکسان

ویژگی‌ها یا خصوصیات گره‌های اصلی یا مرکزی در هر لایه، برای توصیف ویژگی‌های گره بسیار مهم است. بنابراین، علاوه‌بر شباهت اجتماعی مبتنی‌بر ساختار، شباهت میان گره‌ها در لایه یکسان که از شباهت مبتنی‌بر ویژگی می‌آید نیز باید درنظرگرفته‌شود[۴]. طبقه‌بندی گره ب اساس ویژگی را می‌توان با دسته‌بندی گره‌های دارای خصوصیات یکسان در دستهٔ یکسان به‌دست‌آورد.

فرض‌کنید $G_{\mathcal{L}} = (V_{\mathcal{L}}.E_{\mathcal{L}}.V.\mathcal{L})$ مدل شبکه مبتنی‌بر ویژگی را نشان‌دهد که در آن $V_{\mathcal{L}}$ بیانگر مجموعه گره‌ها در لایه‌ی L است، V متعلق به مجموعه $\{V_1.V_2.\ldots..V_n\}$ و n بیانگر تعداد گره‌های موجود در شبکه است و $E_{\mathcal{L}}$ نشانگر یال‌هایی است که گره‌های موجود در لایه‌ی L را به هم متصل‌می‌کند. فرض‌کنید $S_{\mathcal{L}} = (S_1.S_2.\ldots.S_n)$ مجموعه ویژگی‌های گره را نشان‌می‌دهد که در آن S_L مجموعه ویژگی‌های موجود در لایه‌ی L است و S_1 ویژگی گره‌ی V_1 را نشان‌می‌دهد. یک شبکه با L لایه را مدنظر قراردهید که در آن هر لایه دارای n ویژگی است که می‌توان با N کلمه کلیدی متناظر $\{X_1.X_2.\ldots..X_n\}$ نمایش-داد. آنگاه مجموعه ویژگی‌ها برای گره‌ی V_i را می‌توان توسط بردار $S_i = (W_{i1}.W_{i2}.\ldots.W_{in})$ نشان‌داد. عنصر W_{ir} درصورتی ۱ است که کلمه کلیدی X_r در گره‌ی V_i ظاهرشود؛ درغیراین‌صورت برابر صفر است[۱۷]. برای گره V_i در لایه X، مجموعه ویژگی زیرسیستم همسایه آن C_{im} را می‌توان به شرح زیر محاسبه‌کرد:

$$S_x C_{im} \frac{1}{n}(\sum_{j=1}^{n} W_{j1}.\sum_{j=1}^{n} W_{j2}.\ldots.\sum_{j=1}^{n} W_{jn}) \quad (۱)$$

که در آن N تعداد همسایگان موجود در C_{im} است. شباهت ویژگی $S_{imX}(S_i.S_{cim})$ بین گره V_i و مجموعه همسایه آن C_{im} در لایه X را می‌توان با کسینوس زاویه مشمولِ بین دو بردار متناظر محاسبه‌کرد [۴، ۱۱]:

$$S_{xim}(S_i.S_{Cim}) = \frac{\sum_{j=1}^{N} s_{ij} \times S_{Cim}}{\sqrt{\left[\sum_{j=1}^{N} s_{ij}\right]^2 \left[\sum_{j=1}^{N} s_{Cim}j\right]}} \quad (۲)$$

در اینجا، S_{ij} و S_{cim} به‌ترتیب بیانگر j امین عناصر بردارهای ویژگی V_i و C_{im} در لایه X هستند، یعنی W_{ij} و $\frac{1}{n}\sum_{k=1}^{n} W_{kj}$ یک $S_{xim}(S_i.S_{Cim})$ با مقدار بالاتر، میزان شباهت ویژگی بیشتری را بین گره‌ها و احتمال بیشتری از تعلق گره به زیرسیستم همسایه آن یعنی C_{im} نشان‌می‌دهد. اگر دو یا چند ویژگی وجود داشته‌باشند که بیشترین تکرار را دارند، از خصوصیات گره برای محاسبه شباهت میان گره و هریک از زیرسیستم‌های همسایه متناظر آن استفاده‌می‌شود. ویژگی زیرسیستم همسایه که برای آن $S_{xim}(S_i.S_{Cim})$ بیشترین مقدار را داشته‌باشد به-عنوان ویژگی گره seed یعنی V_0 انتخاب‌می‌شود.

۲،۳. قدرت تعامل اجتماعی در میان گره‌ها در لایه‌ای یکسان

در روش پیشنهادی، میزان شباهت از نظر اطلاعات اجتماعی استفاده‌می‌شود تا به تعیین آن گره نهایی که باید در طی بسط اجتماعی به یک جامعه اضافه‌شود، کمک‌کند. یعنی زمانی بسط و توسعه جامعه خاتمه‌پیدامی‌کند که میزان شباهت مابین گره فعلی و گره اصلی زیرسیستم همسایه با همان ویژگی، کمتر از میزان شباهت میان گره فعلی و گره اصلی یک زیرسیستم همسایه با یک ویژگی متفاوت باشد.

$$S_x tr(S_i.S_{Cim}) = \frac{\sum_{i\in\Gamma i\cap\Gamma cim}^{N} S_{ij}\times S_{Cim}}{\sqrt{\sum_{i\in\Gamma i}\frac{1}{D(t)}}\sqrt{\sum_{i\in\Gamma cim}\frac{1}{D(t)}}} \qquad (۳)$$

در اینجا، $S_x tr(S_i.S_{Cim})$ بیانگر میزان شباهت میان گره i و زیرسیستم همسایه آن در لایه X است؛ Γi و Γcim مجموعه گره‌هایی را نشان‌می‌دهند که به‌ترتیب گره i به آن اشاره‌می‌کند و مجموعه گره‌هایی که زیرسیستم C_{im} اشاره‌می‌کند؛ و D(t) درجه گره را نشان‌می‌دهد.

۳،۳. میزان شباهت ویژگی و شباهت تعامل اجتماعی در یک شبکه چندلایه

بحث فوق روی تعریف میزان شباهت ویژگی گره و شباهت تعامل اجتماعی در یک شبکه تک‌لایه تمرکزمی‌کند. بااین حال، روابط انسانی در شبکه‌های اجتماعی چندوجهی و شبکه‌های اجتماعی دارای چندین لایه هستند. دراینجا، لایه‌های مختلف یک شبکه اجتماعی درنظرگرفته‌شده و چگونگی اندازه‌گیری شباهت‌های تعامل اجتماعی و ویژگی در یک شبکه چندلایه بررسی‌می‌شود. ویژگی‌های یک مجموعه از همسایگان را می‌توان از نظر ویژگی‌های تمام گره‌های موجود در سیستم نشان‌داد. برای گره V_i، فرض‌کنید C_{xyim} مجموعه همسایگان آن را در یک شبکه چندلایه نشان‌می‌دهد؛ پس مجموعه ویژگی‌های C_{xyim} را می‌توان به شرح زیر محاسبه‌کرد:

$$S_{xycim} = \frac{1}{n}(\sum_{ix.iy\in\mathcal{L}}\sum_{i.j=1} W_{i.j2}\cdot\ldots\cdot\sum_{ix.iy\in\mathcal{L}}\sum_{i.j=1} W_{i.jn}) \qquad (۴)$$

که در آن N بیانگر تعداد گره‌ها در مجموعه C_{xyim} است و X شباهت میان گره‌ی V_i در لایه‌ی X و مجموعه همسایگان آن در لایه Y را نشان می‌دهد.شباهت ویژگی مابین گره V_i در لایه X و گره J_i در لایه Y را می‌توان با کسینوس زاویه مشمول میان دو بردار مربوطه اندازه‌گیری‌کرد:

$$S_{xy}im(S_i.S_{Cim}) = \frac{\sum_{j=1}^{N} S_{ij}\sum_{ix.iy\in\mathcal{L}} S_{Cim}}{\sqrt{[\sum_{j=1}^{N} S_{ij}]^2[\sum_{ix.iy\in\mathcal{L}}^{N} S_{Cim}j]}} \qquad (۵)$$

۴،۳. شباهت تعامل اجتماعی میان گره‌ها در لایه‌های مختلف

در روش پیشنهادی، از شباهت تعامل اجتماعی در یک شبکه چندلایه برای تعیین گره پایانی که باید طی بسط و گسترش جامعه به آن اضافه‌شود، استفاده‌می‌شود، یعنی روند بسط یا توسعه جامعه زمانی خاتمه‌می‌یابد که میزان شباهت میان گره فعلی و گره اصلی زیرسیستم همسایه با ویژگی یکسان کمتر از میزان شباهت بین گره فعلی و گره اصلی یک زیرسیستم همسایه با یک ویژگی متفاوت‌باشد.

$$S_{xy}tr(S_i.S_{Cim}) = \frac{\sum_{ix.iy\in\mathcal{L}}\sum_{ix.iy\in\mathcal{L}}\frac{1}{D(t)}}{\sqrt{\sum_{t\in\Gamma i}\frac{1}{D(t)}}\sqrt{\sum_{ix.iy\in\mathcal{L}}\sum_{t\in\Gamma cim}\frac{1}{D(t)}}} \qquad (۶)$$

در اینجا، $S_{xy}tr(S_i.S_{Cim})$ بیانگر میزان شباهت بین گره i در لایه‌ی X و زیرسیستم همسایه آن در تمام لایه‌های دیگر است؛ Γi و Γcim بیانگر مجموعه‌ی گره‌هایی است که گره i به آن اشاره‌می‌کند و مجموعه گره‌هایی که به زیرسیستم C_{xyim} اشاره‌می‌کنند و D(t) درجه گره را نشان‌می‌دهد.

۵،۳. الگوریتم جهش قورباغه و بهبود ویژگی‌های تأثیرگذار در شباهت‌یابی

الگوریتم‌های Memetic دسته خاصی از روش‌های جستجوی فرااکتشافی هستند که از مدل‌های منطبق‌برسیستم‌های طبیعی نتیجه‌می‌شوند. الگوریتم جهش قورباغه(SFLA) نیز یک الگوریتم جستجوی سراسری است، که الگوهای رفتاری تکاملی گروهی از قورباغه‌ها را زمانی که به‌دنبال یافتن محلی با بیشترین مقدار غذا هستند، تقلیدمی‌کند. نتایج بهینه‌سازی ثابت‌می‌کند که الگوریتمSFLA در مقایسه با الگوریتم‌های فراابتکاری مدرن و همچنین روش‌های مرسوم بسیار بهتر عمل‌می‌کند و نتایج حاصل نشان‌می‌دهد این الگوریتم نسبت به سایر الگوریتم‌ها در توزیع بار بهینه‌تر است[۱۲]. از الگوریتم جهش قورباغه با انعطاف و قدرت آن در تعیین میزان تأثیر و اهمیت ویژگی استفاده‌شده‌است.

۶،۳. خوشه‌بندی شبکه

برای این عمل از C-Means فازی استفاده‌می‌شود. FCM از متداول‌ترین الگوریتم‌های خوشه‌بندی است. خوشه‌بندی FCM براساس همان مراکز خوشه‌ای تعریف و از طریق تنظیم مکرر مکان‌هایشان و حداقل‌سازی تابع هدف به‌عنوان الگوریتم K-means (KM) ساخته‌شده‌است[۱۳]. الگوریتم FCM اهداف بُعد–L را به مجموعه‌های فازی تقسیم‌می‌کند. N بیانگر تعداد اهداف، و C بیانگر تعداد مجموعه‌ها می‌باشند. به‌منظور خوشه‌بندی داده‌ها، تابع هدف از طریق به‌روزرسانی مکرر عضویت اهداف و مراکز خوشه به حداقل می‌رسد. تابع هدف به‌صورت زیر می‌باشد:

$$J(U.V) = \sum_{i=1}^{N}\sum_{c=1}^{C} u_{ic}^{m} d^2(x_i.v_c) \qquad (۷)$$

درجهٔ عضویت U_{ic} مطابق با شرایط محدودیت است:

$$\sum_{c=1}^{C} u_{ic} = 1 \quad .\forall\, i \qquad (۸)$$

بنابراین، فرمول‌های محاسبهٔ درجهٔ عضویت و مرکز خوشه به‌صورت زیر می‌باشند:

$$u_{ic} = \frac{1}{\sum_{c=1}^{C}\left(\frac{d(x_i.v_c)}{d(x_i.v_c)}\right)^{m-1}} \qquad (۹)$$

$$v_c = \frac{\sum_{i=1}^{N} u_{ic}^{m} x_i}{\sum_{i=1}^{N} u_{ic}^{m}} \qquad (۱۰)$$

عملکرد روش FCM برپایه روش K-Means است که یکی از روش‌های خوشه‌بندی است.

دو هدف اصلی در این الگوریتم دنبال‌می‌شود، هدف نخست به‌دست‌آوردن گره‌ای به‌عنوان مراکز خوشه‌ها است، که این گره‌ها درواقع همان مقدار میانگین گره‌های متعلق به هر خوشه هستند و دوم، نسبت‌دادن هر گره به یک خوشه به‌طوری که گره کمترین فاصله تا

مرکز آن خوشه را داشته‌باشد. فضای مسئله‌ای n بعدی را درنظربگیرید، گره‌ها در این فضا به‌صورت $D=\{x_i|i=1,2, \dots ,n\}$ قابل‌تعریف است در K-means هریک از خوشه‌ها با یک گره واحد که همان مرکز یا مقدار میانگین خوشه است، نمایش‌داده‌می‌شوند. مجموعه $C=\{c_j|j=1,2, \dots ,k\}$, معرف مرکز خوشه‌هاست. برداری نیز به‌نام M برای ذخیره شماره خوشه منتسب به هر یک از گره‌ها درنظرگرفته‌می‌شود، که در آن هر m_i شماره خوشه برای داده X_i است. در الگوریتم k-means معیار پیش‌فرض برای اندازه‌گیری شباهت داده‌ها فاصله اقلیدسی است و الگوریتم به‌دنبال حداقل‌کردن مجموع توان دو فاصله اقلیدسی بین هر X_i و C_j منتسب آن است[۱۳]. تابع زیر به‌عنوان تابع هدف مطرح است.

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| X_i^{(j)} - c_j \right\|^2 \qquad (۱۱)$$

که $\left\| X_i^{(j)} - c_j \right\|^2$ معیار فاصله بین نقاط و cj مرکز خوشه j ام است.

۵.نتایج و تحلیل شبیه‌سازی

در این قسمت روش پیشنهادی آزمایش و نتایج مراحل مختلف ارزیابی تجزیه و تحلیل شد. برای شبیه‌سازی روش ارائه‌شده در فصل قبل نیز از دو نرم‌افزار Matlab و Weka استفاده‌شده‌است.

۱،۵.مجموعه داده

برای ارزیابی روش پیشنهادی مجموعه داده‌های wiki4HE با لینک https://archive.ics.uci.edu/ml/datasets/wiki4he استفاده شده‌است. این مجموعه دارای ۹۱۳ نمونه است و در هر نمونه ویژگی‌های مختلف فردی و اجتماعی از یک شبکه اجتماعی بیانگر میزان استفاده از ویکی‌پدیا است. همراه ویژگی‌های فردی، ویژگی‌های میزان علاقه، اشتراک‌گذاری، میزان سودمندی، مؤثربودن، تصویر، میزان تعامل و میزان مرتبط بودن در آن قراردارد. در این مجموعه داده، تعداد ۵۳ ویژگی از دسته‌های نامبرده قراردارد.

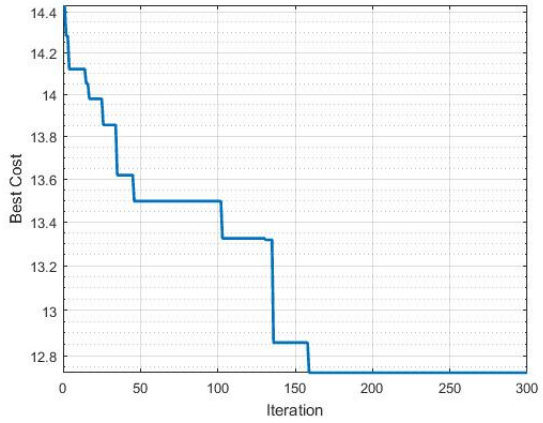
۲،۵. ارزیابی

در روش پیشنهادی AM-ALCD از الگوریتم جهش قورباغه برای بهبود کشف شباهت و کمک به خوشه‌بندی نهایی در استخراج تأثیر گره‌ها، نزدیکی عوامل تأثیرگذار به یکدیگر، وزن‌دهی این عوامل، انتخاب پارامترهای مناسب مشخص‌سازی پارامترهای با اهمیت بیشتر، کاهش زمان نتیجه‌گیری و افزایش کیفیت استفاده‌می‌شود. پارامترهای درنظرگرفته‌شده برای این الگوریتم در جدول ۱ نشان داده‌شده‌است.

جدول ۱: پارامترهای استفاده شده در الگوریتم جهش قورباغه

ردیف	نام پارامتر	توضیحات	مقدار
۱	nVar	تعداد متغیرهای تصمیم‌گیری	برابر با تعداد ویژگی
۲	VarMin	حد پایین مقادیر متغییر	-۱۰
۳	VarMax	حد بالای مقادیر متغییر	۱۰
۴	MaxIt	حداکثر تکرار الگوریتم	۳۰۰
۵	nPop	تعداد قورباغه	۵۰
۶	nPopMemeplex	سایز Memeplex	۱۰
۷	nMemeplex	تعداد Memeplex	۵
۸	Alpha	مقدار آلفا	۳
۹	Beta	مقدار بتا	۵
۱۰	Sigma	مقدار زیگما	۲

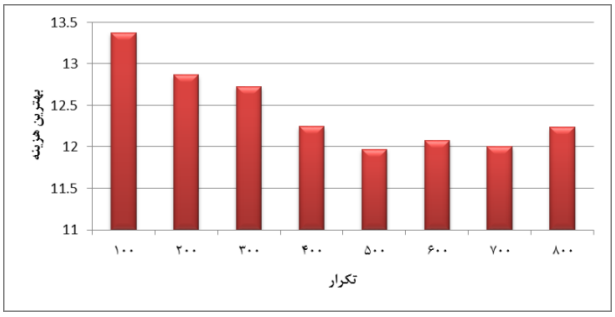
مقادیر ثابت استفاده‌شده در الگوریتم از تکرارهای مکرر به‌دست‌آمده و نشان‌دهنده بهترین خروجی با این مقادیراست. نمونه‌ای از نتیجه اجرا این الگوریتم در شکل ۲ نشان‌داده‌شده‌است.



شکل ۲: نتیجه اجرای الگوریتم جهش قورباغه

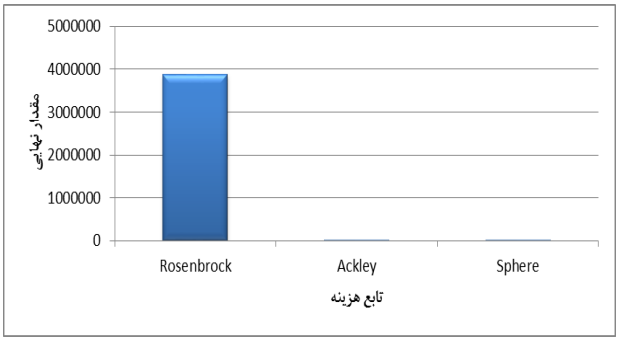
در این قسمت برای دستیابی به افراد و گروه‌های تأثیرگذار در شبکه‌های اجتماعی، از الگوریتم قورباغه جهنده برای بهبود تشخیص پارامترهای تأثیرگذار استفاده‌شده‌است که باعث بهبود اهمیت گره‌ها شود و همان‌گونه که شکل۱ نشان‌می‌دهد در شروع اجرای الگوریتم، مقدار بهترین هزینه بالا بوده و الگوریتم سعی در به‌حداقل‌رسانی و میل آن به سمت عدد صفر دارد که با پیشروی در تعداد تکرار به این امر نزدیک‌می‌شود. انتخاب حداکثر تعداد تکرار بسیار حائز اهمیت است، زیرا در تعداد کم، نتیجه کار بهترین هزینه نامطلوب و در تعداد زیاد، نتیجه افزایش زمان پاسخگویی خواهدبود.

طبق شکل ۳، نتایج اجراهای مختلف و مقدار بهترین هزینه در حال کاهش است. این کاهش تا تکرار ۵۰۰ ادامه‌داشته و افزایش تعداد تکرارها نیز۱۰۰ واحد در هر مرحله بوده‌است. این کاهش مقدار در دورهای۴۰۰ و ۵۰۰ اندک بوده و بعد از آن افزایش داشته‌است. ازاین‌رو، مقدار ۵۰۰ به‌عنوان حداکثر تعداد تکرار الگوریتم جهش قورباغه در روش پیشنهادی استفاده‌شده‌است.



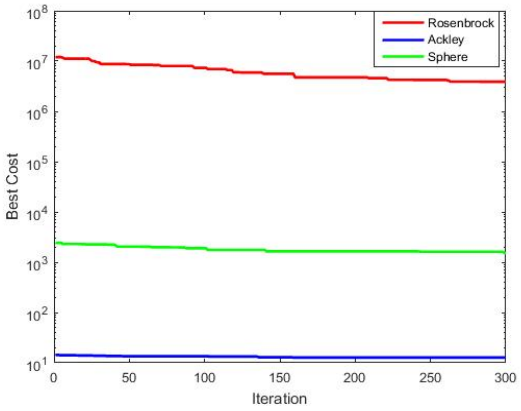
شکل ۳:مقایسه مقدار بهترین هزینه در اجراهای مختلف الگوریتم جهش قورباغه

در الگوریتم جهش قورباغه با توجه به مسئله، سه تابع هزینه مختلف انتخابی وجوددارد شامل: Rosenbrock ، Ackley و Sphere می‌باشند. این سه تابع به‌عنوان محاسبه بهترین هزینه در الگوریتم در نظرگرفته می‌شوند که الگوریتم را با شرایط یکسان و تعداد تکرار ۳۰۰ با هر سه تابع اجراکنند. نتایج بهترین هزینه نهایی در شکل ۴ دیده‌می‌شود.



شکل ۴: مقایسه توابع قابل استفاده در الگوریتم جهش قورباغه از نظر مقدار بهترین هزینه نهایی

مطابق شکل ۴، تابع Rosenbrock بدترین مقدار و تابع Ackley بهترین مقدار را از خود در انتهای اجرا و بعد از ۳۰۰ تکرار نشان‌داده‌است. نمودار تغییر مقدار تابع هزینه در سه تابع مورد مقایسه در شکل ۵ دیده‌می‌شود.



شکل ۵: مقایسه مقدار تابع هزینه درتکرارهای مختلف الگوریتم جهش قورباغه با سه تابع قابل استفاده

همانگونه که در شکل ۵ دیده‌می‌شود تابع Rosenbrock بدترین مقدار و تابع Ackley بهترین مقدار را از خود در انتهای اجرا و بعد از ۳۰۰ تکرار نشان‌داده‌است. با توجه به این نتایج در روش پیشنهادی در بخش انتخاب مسیر تابع Ackley استفاده‌می‌شود.

برای ارزیابی خوشه‌بندی روش AM-ALCD ارائه‌شده در فصل قبل ابتدا با استفاده از مجموعه داده معرفی‌شده FCM را برروی آن اجرامی‌کنیم. گام‌های جمع‌آوری داده و پیش‌پردازش در مجموعه داده اعمال‌شده‌است. نتایج روش پیشنهادی را با الگوریتم K-means مقایسه می‌کنیم. نتایج الگوریتم K-means و نتایج روش پیشنهادی به‌ترتیب در جدول ۳ و ۴ آمده‌است.

جدول ۲: تایج روش K-means

	شماره خوشه	۰	۱	۲	۳	۴	۵	۶
K=2	تعداد عضو	۶۱۴	۲۹۹					
	درصد عضو	۶۷	۳۳					
K=3	تعداد عضو	۴۱۴	۲۵۰	۲۴۹				
	درصد عضو	۴۵	۲۷	۲۷				
K=4	تعداد عضو	۳۰۱	۲۰۳	۲۲۷	۱۸۲			
	درصد عضو	۳۳	۲۲	۲۵	۲۰			
K=5	تعداد عضو	۲۵۵	۱۹۴	۱۸۶	۱۵۳	۱۲۵		
	درصد عضو	۲۸	۲۱	۲۰	۱۷	۱۴		
K=6	تعداد عضو	۲۰۵	۱۶۱	۱۹۸	۱۷۲	۷۵	۱۰۲	
	درصد عضو	۲۲	۱۸	۲۲	۱۹	۸	۱۱	
K=7	تعداد عضو	۱۸۰	۱۶۰	۱۷۱	۱۴۷	۷۵	۱۰۰	۸۰
	درصد عضو	۲۰	۱۸	۱۹	۱۶	۸	۱۱	۵۹

طبق جدول ۲، روش K-means در تعداد K برابر با ۲ دارای عدم‌تعادل بوده و خوشه یک ۶۷ درصد اعضا را به خود اختصاص‌داده‌است با افزایش مقدار K این روند متعادل‌تر می‌شود. در تعداد ۵ این مقدار به یک تعادل نسبی قابل‌قبول می‌رسد.

مطابق جدول ۳، روش پیشنهادی دارای نتایج نزدیک به روش K-means و قابل‌قبول است. همچنین میزان تعادل در تعداد ۵ خوشه در هر دو روش روی مجموعه داده پیشنهادی را می‌توان به‌عنوان تعداد خوشه‌ها پیشنهادداد.

جدول ۳: نتایج روش پیشنهادی

	شماره خوشه	۰
nC =2	تعداد عضو	۵۷
	درصد عضو	۵
nC =3	تعداد عضو	۳۸
	درصد عضو	۶۶
nC =4	تعداد عضو	۸۵
	درصد عضو	۴۱
nC =5	تعداد عضو	۵۸
	درصد عضو	۲۸
nC =6	تعداد عضو	۵۱
	درصد عضو	۲۴
nC =7	تعداد عضو	۵۱
	درصد عضو	۲۴

در FCM اجراشده در روش پیشنهادی مقادیر جدول ۴ به‌عنوان مقادیر پارامترها درنظرگرفته‌شده‌است.

جدول ۴: پارامترهای استفاده شده در روش

ردیف	پارامتر	مقدار
۱	M	۲
۲	تتا	۲
۳	حد آستانه اصطلاحات	۰,۰۰۰۰۱
۴	حداکثر تکرار	۱۰۰

در ادامه ارزیابی خوشه‌های به‌دست‌آمده از تعداد برابر ۵ با روش‌های طبقه‌بندی مختلف مقایسه شده تا قدرت و بهبود روش

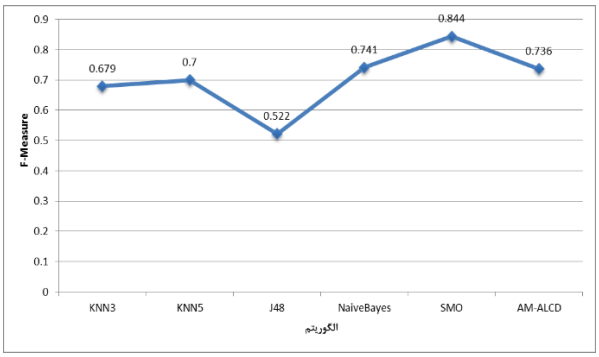
پیشنهادی نسبت به عمده روش‌های شناخته‌شده و پرکاربرد به‌دست‌آید. برای ارزیابی از معیارهای جدول ۵ استفاده‌شده‌است[۱۵،۱۶]:

جدول ۵: معیارهای ارزیابی روش پیشنهادی

معیار	توضیحات
صحت	نزدیکی توافق بین مقدار میانگین حاصل از تعداد زیادی از نتایج آزمون و مقدار مرجع پذیرفته‌شده به صحت «درستی میانگین» نیز گفته می‌شود که از رابطه-۱۶ به‌دست‌می‌آید. $Accuracy = \frac{TN+TP}{TN+FN+TP+FP}$ (رابطه-۱۶)
TP Rate	بیانگر تعداد رکوردهایی است که دسته واقعی آن‌ها مثبت بوده و الگوریتم طبقه‌بندی نیز دسته به‌درستی مثبت تشخیص‌داده است.
FP Rate	بیانگر تعداد رکوردهایی است که دسته واقعی آن‌ها منفی بوده و الگوریتم طبقه‌بندی دسته، به اشتباه مثبت تشخیص‌داده است.
Recall	یک پارامتر عمومی که برای ارزیابی سودمند بودن الگوریتم پیشنهادی به‌کارمی‌رود و از رابطه-۱۷ به‌دست‌می‌آید. $Recall = \frac{TP}{FN+TP}$ (رابطه-۱۷)
Precision	یک پارامتر عمومی که برای اندازه گیری مفید بودن الگوریتم پیشنهادی استفاده‌شده و از رابطه-۱۸ به‌دست می‌آید. $Precision = \frac{TP}{FP+TP}$ (رابطه-۱۸)
F-Measure	این معیار از محاسبه میانگین همساز بین دو معیار سودمندبودن و مفیدبودن است و از رابطه -۱۹ به‌دست می‌آید. $F-Measure = 2 * \frac{Precision * Recall}{Precision + Recall}$ (رابطه-۱۹)

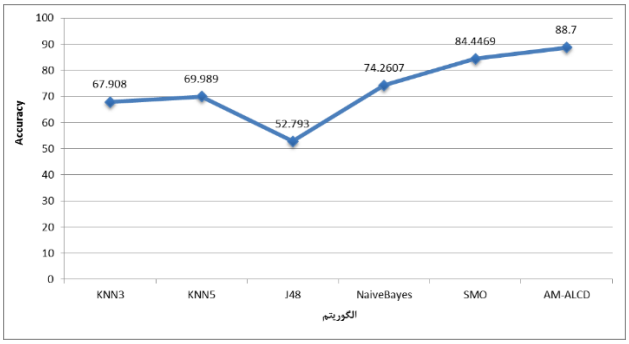
در کلیه آزمایش‌ها از تست صحت به روش K-Fold با K=10 استفاده‌شده‌است. در این نوع اعتبارسنجی داده‌ها به K زیرمجموعه افرازمی‌شوند. از این‌K زیرمجموعه، هربار یکی برای اعتبارسنجی و K-1 تای دیگر برای آموزش به‌کارمی‌روند. این روال K بار تکرارمی‌شود و همه داده‌ها دقیقاً یکبار برای آموزش و یکبار برای اعتبارسنجی به‌کارمی‌روند. درنهایت میانگین نتیجه این‌K بار اعتبارسنجی به‌عنوان یک تخمین نهایی برگزیده‌می‌شود. برای ارزیابی از الگوریتم‌های طبقه‌بندی K نزدیک‌ترین همسایه، J48 ، SMO، Decision Table ، الگوریتم مبتنی‌بر نظریه بیز و Bagging استفاده‌می‌شود[۱۴].

در شکل ۶ روش پیشنهادی و سایر روش‌ها از نظر معیار F-Measure با یکدیگرمقایسه‌شده‌اند.



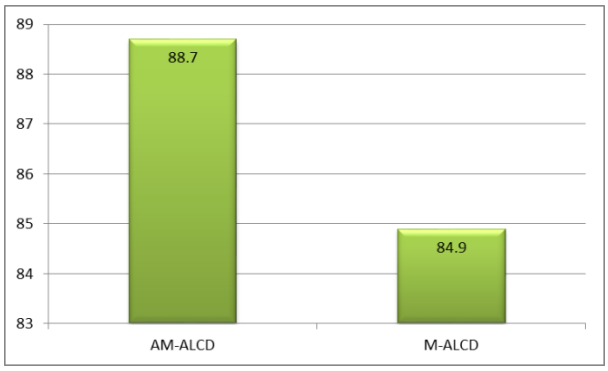
شکل ۶: ارزیابی روش پیشنهادی و سایر روش‌ها از نظر معیار F-Measure

طبق شکل ۶، در این معیار روش AM-ALCD به نسبت روش SMO و بیزین دارای مقدار کمتری است و دلیل آن بالابودن هر دو مقدار TP و FP است، یعنی با توجه به روش پیشنهادی هم درست‌ها و هم غلط‌ها را به‌درستی تشخیص‌می‌دهد و این امر سبب کاهش این معیار شده‌است چون این معیار از ترکیب Precision و Recall به‌دست‌می‌آید ولی روش SMO دارای مقدار FP بسیار کمی بوده که این اختلاف سبب افزایش در معیارF-Measure شده‌است. پایین‌ترین مقدار در این معیار نیز مربوط به روش Bagging است. در شکل ۷ روش‌های مختلف از نظر معیار صحت با یکدیگر مقایسه‌شدند.



شکل ۷: ارزیابی روش پیشنهادی و سایر روش‌ها از نظر معیار صحت

مطابق شکل ۷، روش AM-ALCD به نسبت کلیه روش‌های مورد مقایسه دارای صحت بالاتری است. بعد از روش پیشنهادی با اختلاف ۳،۳ درصد روش SMO قراردارد و بیزین روش بعدی در این ترتیب است. روش‌های KNN با K برابر ۵ و K برابر ۳ نیز با اختلاف حدود ۲۰ درصدی در رتبه‌های بعدی قراردارند و سه روش J48، Decision Table و Bagging از جمله الگوریتم‌هایی هستند که به‌ترتیب پایین‌ترین مقدار صحت را نشان داده‌اند. ازآنجا که روش AM-ALCD مقدار بالاتری در صحت طبقه‌بندی داشته، تشخیص گروه مؤثر و وزن‌دهی مطلوب‌تری نیز نسبت به سایر روش‌ها خواهدداشت. ارزیابی نهایی مقایسه با روش پایه [۴] بوده که در شکل ۸ آمده‌است.



شکل ۸: مقایسه روش پیشنهادی و روش پایه از نظر معیار صحت

طبق شکل ۸، براساس یکسان‌سازی پارامترهای ورودی و ارزیابی موارد شبیه‌سازی قبل، روش پیشنهادی AM-ALCD نسبت به روش پایه M-ALCD، ۳,۸ درصد بهبود ثبت‌کرده‌است.

۶. نتیجه‌گیری و کارهای آتی

در این مقاله یک مدل تشخیص جامعه محلی چندلایه پیشنهادشده-است که براساس اطلاعات ساختار و ویژگی است. این مدل می‌تواند از اطلاعات ویژگی گره‌ها و اطلاعات قدرت شباهتی که توسط مبادلات اجتماعی آشکارمی‌شود، بهره‌برداری کرده و دقت تشخیص جامعه را در شبکه‌های چندلایه بهبودبخشد. این الگوریتم به‌دلیل ماژولاریتی و بهره‌وری محاسباتی خود، برخلاف الگوریتم‌های تشخیص جامعه سراسری و چندلایه کلاسیک، روی بیشتر مجموعه‌های داده، قدرتمند عمل‌می‌کند. این روش برای دستیابی به افراد و گروه‌های تأثیرگذار در شبکه‌های اجتماعی از الگوریتم قورباغه جهنده برای بهبود تشخیص پارامترهای تأثیرگذار استفاده‌کرده‌است. براساس روش پیشنهادی بعد از جمع‌آوری داده‌ها به پاکسازی و نرمال‌سازی آن پرداخته‌ایم تا با داده‌های مطلوب، موفق به شناسایی افراد و گروه‌های مؤثر شویم. با این هدف، ابتدا ماتریس تصمیم تشکیل‌شده و از روی آن به شناسایی و خوشه‌بندی پرداخته‌می‌شود و اهمیت گروه‌ها نیز در نظرگرفته‌می‌شود. اهمیت گروه‌ها با الگوریتم قورباغه جهنده بهبودمی‌یابد. روش ارائه‌شده مبتنی‌بر خوشه‌بندی فازی است. بعد از تشریح روش ارائه‌شده، شبیه‌سازی و ارزیابی انجام‌شد و پارامترهای لازم در الگوریتم‌ها سنجیده و با شبیه‌سازی، بهترین آن‌ها به‌دست‌آمد. در ارزیابی خوشه‌بندی، روش پیشنهادی با روش K-means مقایسه و مقدار ۵ خوشه برای روش پیشنهادی- به‌عنوان مقداری که روش در آن به-تعادل‌رسیده‌است- ارائه‌شد. سپس خوشه‌های به‌دست‌آمده مورد ارزیابی قرارگرفت و با روش‌های مختلف بررسی‌شد؛ در این بررسی‌ها روش پیشنهادی نسبت به روش‌های دیگر از خود بهبودی نشان‌داده‌است یعنی در معیار صحت، روش پیشنهادی نسبت به روش بردار پشتیبان تصمیم که نزدیکترین مقدار صحت را به روش پیشنهادی داشته، بهبود ۳,۳ درصدی داشته‌است. همچنین روش AM-ALCD نسبت به روش پایه M-ALCD بهبود ۳,۸ را به ثبت رسانده‌است.

برای بهبود و کارهای آینده با روش پیشنهادی، می‌توان به پیاده‌سازی کامل آن در دنیای واقعی اشاره‌کرد، به‌نحوی که نواقص روش پیشنهادی در دنیای واقعی کشف و مرتفع‌گردد. همچنین می‌توان از روش‌های تکاملی و اکتشافی دیگر به غیر از الگوریتم تکاملی قورباغه جهنده برای نشان‌دادن بهترین مقدار در تعداد خوشه‌ها استفاده‌کرد یا الگوریتم‌های ترکیبی در طبقه‌بندی و خوشه‌بندی را به-کارگرفت.

مراجع

[1] Dasgupta. S, Prakash. Ch, (2016), “Intelligent Detection of Influential Nodes in Networks”, IEEE, International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), pp. 2626-2628.

[2] Yang. Y, Xie. G, (2016),“Efficient identification of node importance in social networks”, Science Direct, Information Processing and Management, Vol. 52, pp. 911-922.

[3] [Zhou. J, Zhang. Y, Cheng. J, (2014) , ”Preference-based mining of top-K influential nodes in social networks “, Science Direct, Future Generation Computer System, Vol. 31, pp. 41-47.

[4] Xiaoming Li, Guagquan Xu, Litao Jiao, Yinan Zhou, Wei Yu.(2019).“ Multi-layer network community detection model based on attributes and social interaction intensity”. 0045-7906/© 2019 Elsevier Ltd. All rights reserved.

[5] Lokesh Jain , Rahul Katarya , Shelly Sachdeva.(2019). “Opinion Leader detection using Whale Optimization Algorithm in Online Social Network”, https://doi.org/10.1016/j.eswa.2019.113016.

[6] Yu Lei, Ying Zhou, Jiao Shi.(2019). “Overlapping Communities Detection of Social Network based on Hybrid C-means Clustering Algorithm ” https://doi.org/10.1016/j.scs.2019.101436.

[7] Aftab Farooq, Usman Akram, Gulraiz Javaid Joyia, and Chaudhry Naeem Akbar.(2018). “A Technique to Identify Key Players that Helps to Improve Businesses Using Multilayer Social Network Analysis” International Journal of Future Computer and Communication, Vol. 7, No. 4, December 2018

[8] Ruchi Mittal and M.P.S Bhatia,.(2019). “Classifying the Influential Individuals in Multi-Layer Social Networks” International Journal of Electronics, Communications, and Measurement Engineering Volume 8 Issue 1 January-June.

[9] Mohammed Ali Al-garadi, Kasturi Dewi Varathan, Sri Devi Ravana, Ejaz Ahmed, Victor Chang .(2016). “Identifying the influential spreadersin multilayer interactions ofonline socialnetworks” Published in Journal of Intelligent and fuzzy system.

[10] Weiyi Liu, Pin-Yu Chen, Sailung Yeung, Toyotaro Suzumura and Lingli Chen .(2017). “Principled Multilayer Network Embedding” arXiv:1709.03551v3 [cs.SI] 15 Sep 2017.

[11] Interdonato R, Tagarelli A, Ienco D, et al. Local community detection in multilayer networks. Data Min Knowl Discov 2017;31(5):14 4 4–79. https: //doi.org/10.1007/s10618- 017- 0525- y .

[12] T. Hoang Huynh, “A modified shuffled frog leaping algorithm for optimal tuning of multivariable PID controllers” IEEE international conference, 2008.

[13] Bezdek J C, Ehrlich R, Full W. FCM: The fuzzy c-means clustering algorithm[J]. Computers & Geosciences, 1984, 10(2-3): 191-203.
[14] Mogelmose, A.; Trivedi, M.M..(2012). Moeslund, T.B. Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey. IEEE Trans. Intell. Transp. Syst.2012, 13, 484–1497
[15] Rachmawanto, Eko Hari, Galang Rambu Anarqi, and Christy Atika Sari. "Handwriting Recognition Using Eccentricity and Metric Feature Extraction Based on K-Nearest Neighbors." In 2018 International Seminar on Application for Technology of Information and Communication, pp. 411-416. IEEE, 2018.
[16] Sa'di, Sadri, Amanj Maleki, Ramin Hashemi, Zahra Panbechi, and Kamal Chalabi. "Comparison of data mining algorithms in the diagnosis of type II diabetes" International Journal on Computational Science & Applications (IJCSA) 5, no. 5 (2015): 1-12.
[17] Chen D, Zhao W, Wang D, et al. Similarity-based local community detection for bipartite networks. Filomat 2018;32(5):1559–70 .