APJ 5(4):1-14, Winter 2025

Arman Process Journal (APJ)

Arman Process Journal (APJ)

Homepage: https://www.armanprocessjournal.ir

Link Prediction System in Social Networks Based on Meta-Heuristic Algorithm

M. Ahmadi Shakib[1], D. Karimzadgan Moghadam[2*], MR. Sanaei[3]

[1] Department of Management and Accounting, Qazvin Branch, Islamic Azad University, Qazvin, Iran
[2] Department of Computer Science, Payam Noor University, Tehran, Iran
[3] Department of Management and Accounting, Qazvin Branch, Islamic Azad University, Qazvin, Iran

ABSTRACT

RESEARCH PAPER

Received: 7 July 2024
Accepted: 18 November 2024

KEYWORDS:
Link prediction,
Meta-heuristic algorithms,
Data preprocessing,
Big data problems,
Computer social networks

[1] **Corresponding author:**
d.karimzadgan.pnu.ac@gmail.com.

Social networks are primarily represented and analyzed in the form of graphs with a large number of vertices and edges, structured as an adjacency matrix. The edges indicate relationships between individuals and act as connections between the vertices. The structural characteristics of each network are determined by the features of the edges and vertices within it. In this research, conducted on various types of social network data from the Stanford University database, a preprocessing method was employed using a competitive colonial algorithm for feature selection with the highest merit (lowest cost). To evaluate the impact of feature selection on the final output, experiments were conducted both with and without feature selection operations using various algorithms commonly used in this field. Valid metrics such as accuracy, precision, sensitivity, and recall were independently measured on the output results with an average of 10 program executions. The comparison of results between scenarios with and without feature selection showed a significant impact on all metrics of the final outcome. Many features in the datasets were either unused or contained minimal information. Not removing these features not only increased the computational burden but also affected the accuracy of the output results due to time-consuming execution.

نشریه تخصصی آرمان پردازش، دوره ۵، شماره ۴، زمستان ۱۴۰۳

فصلنامه تخصصی آرمان پردازش
(APJ)
Homepage: www.armanprocessjournal.ir

فصلنامه تخصصی فناوری اطلاعات و ارتباطات
شماره مجوز: ۸۷۰۹۰

سیستم پیش‌بینی لینک در شبکه‌های اجتماعی بر اساس الگوریتم فراابتکاری

سیده معصومه احمدی شکیب[۱]، داود کریم زادگان مقدم[۲*]، محمدرضا ثنائی[۳]

گروه مدیریت، دانشکده مدیریت و حسابداری، واحد قزوین، دانشگاه آزاد اسلامی، قزوین، ایران
دانشکده علوم کامپیوتر، دانشگاه پیام نور، تهران، ایران
گروه مدیریت، دانشکده مدیریت و حسابداری، واحد قزوین، دانشگاه آزاد اسلامی، قزوین، ایران

چکیده

مقاله پژوهشی

تاریخ دریافت مقاله: ۱۴۰۳/۴/۷
تاریخ پذیرش: ۱۴۰۳ /۸/۲۸

واژگان کلیدی:
پیش بینی لینک،
الگوریتم های فراابتکاری،
پیش پردازش داده ها،
مشکلات کلان داده،
شبکه های اجتماعی
کامپیوتری،

شبکه‌های اجتماعی عمدتاً در قالب نمودارهایی با تعداد زیادی راس و یال در قالب یک ماتریس مجاورت نمایش و تحلیل می‌شوند. لبه‌ها روابط بین افراد را نشان می‌دهند و به عنوان پیوند بین رئوس عمل می‌کنند. ویژگی‌های ساختاری هر شبکه با ویژگی‌های لبه‌ها و رئوس درون آن تعیین می‌شود. در این تحقیق که بر روی انواع داده‌های شبکه‌های اجتماعی از پایگاه داده دانشگاه استنفورد انجام شد، از روش پیش پردازش با استفاده از الگوریتم استعماری رقابتی برای عملیات انتخاب ویژگی‌هایی با بالاترین شایستگی (کمترین هزینه) استفاده شد. برای ارزیابی تأثیر انتخاب ویژگی بر خروجی نهایی، آزمایش‌هایی با و بدون عملیات انتخاب ویژگی با استفاده از الگوریتم‌های مختلف که معمولاً در این زمینه استفاده می‌شوند، انجام شد. شاخص‌های معتبر مانند دقت، تشخیص، حساسیت و عمده به طور مستقل بر روی نتایج خروجی با میانگین ۱۰ اجرای برنامه اندازه‌گیری شدند. مقایسه نتایج بین سناریوهای با و بدون انتخاب ویژگی تأثیر قابل توجهی بر همه شاخص‌های نتیجه نهایی نشان داد. بسیاری از ویژگی‌ها در مجموعه داده‌ها یا استفاده نشده بودند یا حاوی حداقل اطلاعات بودند. حذف نکردن این ویژگی‌ها نه تنها بار محاسباتی را افزایش داد، بلکه بر دقت نتایج خروجی به دلیل اجرای زمان‌بر تأثیر گذاشت.

مقدمه

پیش‌بینی روابط در شبکه‌های اجتماعی در مقیاس بزرگ به‌دلیل ماهیت پویای ساختار شبکه، جایی که لینک‌ها و گره‌ها به مرور زمان به شبکه اضافه می‌شوند، کاری چالش برانگیز است [۱]. روش‌های زیادی برای حل مشکل پیشنهاد شده است. در میان این روش‌ها، روش‌های مبتنی بر شباهت با درنظر گرفتن ساختار شبکه و استفاده به‌عنوان معیاری اساسی از تعداد همسایه‌های مشترک بین دو گره برای ایجاد شباهت ساختاری، کارایی خوبی از خود نشان می‌دهند. شباهت ساختاری بالا ممکن است نشان دهد که احتمالاً یک لینک بین دو گره ظاهر می‌شود [۲]. در مطالعات و پژوهش‌هایی که تاکنون در این حوزه صورت گرفته است عمدتاً چالش‌هایی در نتایج و ارزیابی‌ها بروز کرده است، ازجمله احتمال نرسیدن به پاسخ و نتیجهٔ نهایی که اغلب در به‌کارگیری الگوریتم‌های بدون ناظر و یا استفاده از ویژگی‌های غیرساختاری رخ می‌دهد [۳] و [۴] و یا پیچیدگی‌های محاسباتی که موجب زمانبر شدن حصول نتیجه نیز می‌شود و علی‌رغم دریافت نتایج دقیق‌تر و محتمل‌تر، استفاده از روش پیشنهادی به‌خصوص در شبکه‌های بزرگ غیرقابل استفاده است [۵] ، [۶] ، [۷] و [۸] .

در این پژوهش، یک روش جدید برای پیش‌بینی لینک در شبکه‌های اجتماعی بزرگ معرفی می‌شود. این روش دو چالش اساسی را در این مسئله برطرف می‌کند. اول با استفاده از یک الگوریتم فرااکتشافی، داده‌های بسیار بزرگ را پردازش می‌کند و موارد مفید را نگه می‌دارد (بهینه‌سازی می‌کند). دوم الگوریتم به کاربرده شده در مرحله اول به دلیل بیشترین شباهت کارکردی به ساختار شبکه‌های اجتماعی، در مقایسه با سایر الگوریتم‌های نظیر از نوع رقابت استعماری انتخاب شده و کارایی آن ارزیابی شده‌است.

پیشینه پژوهش

در مقالهٔ [۹]، از یادگیری عمیق برای پیش‌بینی لینک استفاده شد. در این تحقیق با استفاده از استخراج ویژگی‌ها که یکی از مراحل پیش‌پردازش داده‌هاست، بهبود دسته‌بندی‌ها انجام پذیرفت. در روش ارائه شده از الگوریتم‌های متاهیور ستیک مبتنی بر جمعیت ازقبیل شبکهٔ عصبی و الگوریتم ژنتیک برای جست‌وجوی زیرمجموعهٔ ویژگی‌ها استفاده و برای بهبود پیش‌بینی‌ها از داده‌های شبکهٔ اجتماعی Facebookاستفاده شد . در مقالهٔ [۱۰]، یک روش خوشه‌بندی برای انجام عملیات پیش‌بینی لینک ارائه شده است. در پیشینهٔ پژوهش دو شیوهٔ کارآمد پیش‌بینی لینک براساس خوشه‌بندی طیفی با استفاده از کی ـــ مدوئیدز (K-medoids) و لندمارک (Landmark) ارائه شد. شیوهٔ نخست از کی ـــ مدوئیدز برای خوشه‌بندی گره‌های گراف، براساس بردار ویژه به دست آمد که از ماتریس لاپلاسین نرمال، استفاده می‌کند. شیوهٔ دوم زیرمجموعه‌ای از نقطهٔ داده‌ها را به‌عنوان لندمارک انتخاب می‌کند و نقطهٔ داده‌های اولیه را به‌عنوان ترکیب‌های خطی این لندمارک‌ها به نمایش می‌گذارد. با توجه به اینکه هزینهٔ محاسباتی کی ـ مدوئیدز و لندمارک با دقت پیش‌بینی در ارتباط است، باید بهینه‌سازی‌هایی در این حوزه صورت پذیرد. بنابراین تحقیق در این حوزه ادامه دارد. در این پژوهش، روشی جهت پیش‌بینی لینک در پنج شبکه‌های اجتماعی ارائه می‌شود. روش پیشنهادی با استفاده از تکنیک خوشه‌بندی طیفی کی ـــ مدوئیدز و بهره‌گیری از الگوریتم بیشینه سازی انتظار در بهینه سازی نقاط لندمارک ارائه می شود. نتایج شبیه‌سازی در نرم‌افزار متلب نشان می‌دهد که روش پیشنهادی توانسته خطای پیش‌بینی لینک را در مقایسه با شیوهٔ جست‌وجوی محلی کی ـ مینز (خوشه‌بندی طیفی کی ـ مینز) به اندازهٔ ۳۱/۳۱ درصد، خوشه‌بندی طیفی کی ـــ مدوئیدز به اندازهٔ ۹۶/۱۸ درصد و روش آسما (Childhood Allergic Asthma) به اندازهٔ ۳۸/۳ درصد روی شبکهٔ دادهٔ تارو (Taro data network) کاهش دهد. در مقالهٔ [۱۱]، چارچوب جدیدی پیشنهاد می‌شود که پیچیدگی‌های گراف چندسطحی را در ساختار شبکهٔ محلی و ساختار هایپرگراف به شیوه‌ای یکپارچه درنظر می‌گیرد. روش پیشنهادی بر مشکل کمبود دادهٔ کار موجود غلبه می‌کند و لزوماً به اطلاعات جمعیتی کاربر متکی نیست. علاوه بر این، برای انطباق روش پیشنهادی برای توانایی مدیریت شبکه‌های اجتماعی در مقیاس بزرگ، ما یک مکانیسم تطبیق فضایی دوفازی را برای تراز کردن فضاهای تعبیه شده در هر دو آموزش موازی مبتنی بر پارتیشن‌بندی شبکه و تطبیق حساب در شبکه‌های اجتماعی مختلف پیشنهاد می‌کنیم. آزمایش‌های گسترده‌ای روی دو شبکهٔ اجتماعی واقعی در مقیاس بزرگ انجام شده است. نتایج تجربی نشان می‌دهد که روش پیشنهادی بهتر از مدل‌های پیشرفته با حاشیهٔ بزرگ عمل می‌کند.

در مقالهٔ [۸]، یک الگوریتم جریمهٔ درجهٔ تکرارشونده (iterative degree penalty) برای پیش‌بینی پیوند بین‌لایه‌ای در شبکهٔ چندگانه توسعه داده شده است. اول، یک اصل جریمهٔ درجه پیشنهاد شد که وزن بیشتری را به یک همسایهٔ همسان مشترک با اتصالات کم‌تر اختصاص می‌دهد. دوم، ضرب ماتریس مجاورت گره برای به‌دست آوردن مؤثر درجهٔ تطابق همهٔ جفت‌های گره ناهمسان اعمال شده است. پس از آن، از روش حداکثر مقدار تأییدشده برای به‌دست آوردن نتایج پیش‌بینی پیوند بین‌لایه‌ای از

ماتریس درجهٔ تطبیق استفاده شده است. درنهایت، نتایج پیش‌بینی در مجموعهٔ جفت گره‌های بین‌لایه‌ای پیشینی درج شد و فرایندهای فوق به‌طور مکرر انجام شد تا زمانی که تمام گره‌های ناهمسان در یک لایه مطابقت پیدا کردند یا تمام درجات تطبیق جفت‌های گره ناهمسان برابر با ۰ شد. آزمایش‌ها نشان داد که IDP پیشرفته زمانی که میانگین درجهٔ شبکهٔ چندگانه و نرخ هم‌پوشانی گره کم باشد، الگوریتم به‌طور قابل‌توجهی از روش‌های مبتنی بر ساختار شبکهٔ فعلی بهتر عمل می‌کند.

در مقالهٔ [۴]، یک روش مبتنی بر شباهت همسایه برای مسئلهٔ پیش‌بینی لینک ارائه شد. در این پژوهش با بهره‌گیری از معیارهای محلی شباهت مبتنی بر همسایه و معیار عمومی شباهت مبتنی بر مسیر، که هر دو از ساختار گراف استفاده می‌کنند، معیارهای شباهت جدیدی معرفی شده است. نتایج کار بر روی مجموعه داده‌های مورد بررسی، برتری کار را نسبت‌به معیارهایی که تنها از مسیر و یا از همسایه بهره می‌جویند، نشان می‌دهد و مشاهده می شود که روش ارائه شده با ترکیب اطلاعات مسیر و همسایه‌ها می‌تواند براساس معیارهای precision و AUC (Area Under Curve) با دقت بیشتری نسبت‌به روش‌های قبل لینک‌ها را پیش‌بینی کند.

در مقالهٔ [۷]، از ساختارهای اجتماعی شبکه‌های اجتماعی برای بهبود دقت پیش‌بینی استفاده شده است و الگوریتم مبتنی بر فاصلهٔ انقباض کششی (the stretch shrink distance based algorithm) پیشنهاد شده است. حل مسئلهٔ پیش‌بینی پیوند را می‌توان تقریباً به دو دسته تقسیم کرد: روش‌های مبتنی بر شباهت و روش های مبتنی بر یادگیری. روش های مبتنی بر یادگیری دقت بالاتری دارند، اما پیچیدگی‌های زمانی آن‌ها برای شبکه‌های پیچیده بسیار زیاد است. با این حال، روش‌های مبتنی بر شباهت از مزیت زمان کم برخوردارند، بنابراین بهبود دقت آن‌ها به موضوعی کلیدی تبدیل می شود. در SSDBA، ابتدا جوامع یک شبکهٔ اجتماعی شناسایی و گره‌های فعال براساس آستانهٔ میانگین جامعه (community average threshold) و آستانهٔ میانگین گره (node average threshold) در هر جامعه شناسایی می شود. دوم، مدل فاصلهٔ انقباض کششی (the stretch shrink distance) برای محاسبهٔ تکراری تغییرات فواصل بین گره‌های فعال و همسایگان محلی آن‌ها پیشنهاد می شود. درنهایت، زمانی پیش‌بینی می‌شود که فاصلهٔ این پیوندها به هم نزدیک شوند. علاوه بر این، یادگیری پارامترهای گسترده در آزمایش‌ها انجام شده است. این مقاله SSDBA خود را با سایر رویکردهای رایج مقایسه کرده و نتایج تجربی اثربخشی و کارایی الگوریتم پیشنهادی را تأیید می‌کند.

در مقالهٔ [۶]، تکنیکی پیشنهاد شده است که مبتنی بر سنجش شباهت گره و مفهوم شبکهٔ عصبی مصنوعی است. پیش‌بینی ارتباط شبکه‌های اجتماعی تکنیکی کارآمد برای تحلیل تحول سازمان‌های اجتماعی و شکل‌گیری روابط شبکه‌های اجتماعی است. در این مقاله، ویژگی‌ها یا اطلاعات موجود در مجموعهٔ داده درنظر گرفته می‌شود، سپس امتیاز به همهٔ جفت گره‌ها اختصاص می‌یابد و مفهوم شبکهٔ عصبی مصنوعی به سیستم اعمال می‌شود. تکنیک پیشنهادی به‌دلیل استفاده از شبکه‌های عصبی مصنوعی چندلایه، دقت پیش‌بینی لینک را بهبود می‌بخشد.

در مقالهٔ [۱۲]، بر پیش‌بینی پیوند گمشده با رسمی‌سازی و توسعهٔ رویکردی تمرکز دارد که می‌تواند ویژگی هایی را از توپولوژی شبکه برای تجزیه و تحلیل امتیازهای مجاورت گره‌ها، پیوندها و ویژگی‌های آن‌ها در یک شبکه استخراج کند. گام اصلی در این مقاله، کشف یک الگوریتم جدید یال گمشده با معرفی تصادفی تعداد مساوی از پیوندهای گمشده یا «پیوندهای نادرست» در مجموعه داده است که در نمودار نیستند و طول کوتاه‌ترین مسیر آن‌ها بیشتر از ۲ است. مهندسی ویژگی فشرده برای استخراج بیش از ۳۸ ویژگی جدید که به‌عنوان مجموعه دادهٔ اصلی برای روش یادگیری معرفی شده‌اند، شامل ویژگی‌های خاص گراف، ویژگی های گره مانند و چند امتیاز دیگر انجام می‌شود. درنهایت، الگوریتم LightGBM برای آموزش و آزمایش مجموعه دادهٔ ساخته شده استفاده شده است. معیارهای عملکرد مبتنی بر آزمایش مانند AUC، F1، دقت و AP (Average Precision) نتایج چشمگیری در مجموعه داده های مختلف به‌دست آورده‌اند.

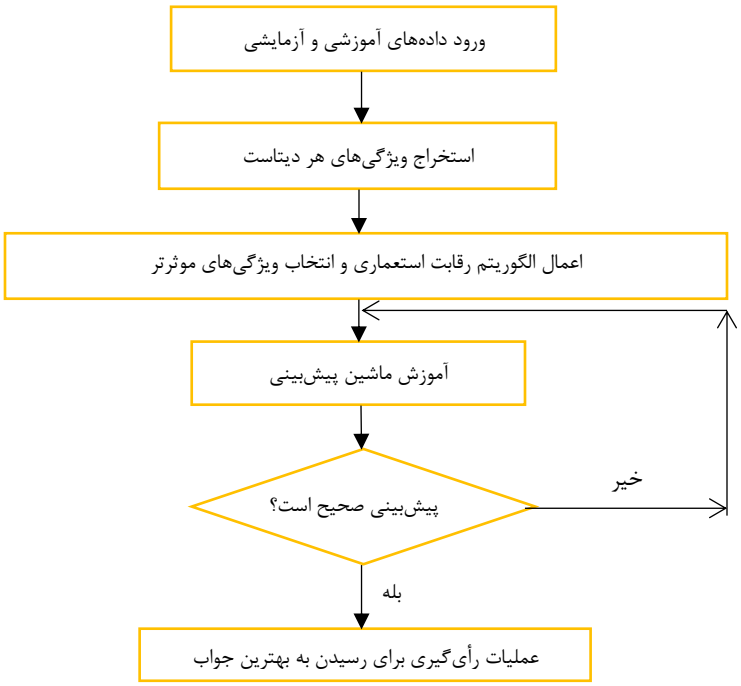
در جدول زیر تحقیقاتی که تاکنون در خصوص پیش بینی لینک انجام شده آورده شده است و نقاط قوت و ضعف آنها با هم مقایسه شده است تا ضرورت این تحقیق و ارائه روشی جدید ارائه شود.

جدول ۱: مقایسه کارهای انجام شده

منبع	روش پیشنهادی	مزایا	معایب
[۹]	یادگیری عمیق	سرعت بالا ـ قدرت پیش‌بینی بالا	در صـــورت کوچک بودن مجموعه داده، مناسب نیست
[۱۰]	خوشه‌بندی طیفی	کاهش خطای پیش‌بینی	نیاز به مقداردهی اولیه
[۱۱]	یک مکانیسم تطبیق فضایی دو فازی	مناسب برای شبکه‌های اجتماعی بزرگ	نیاز به دانش قبلی
[۸]	الگوریتم درجهٔ تکرارشونده	برای شبکه‌ها چندگانه بسیار مناسب است	زمانبر بودن
[۴]	روش مبتنی بر شباهت همسایگان	دقت مناسب ـ استفاده از ساختار کم‌هزینه	همیشه پاسخ درست نمی‌دهد
[۷]	الگوریتم مبتنی بر انقباض کششی	دقت بالا	پیچیدگی زمانی بالا
[۶]	شبکه‌های عصبی	بهبود دقت پیش‌بینی	مقداردهی اولیه پارامترها
[۱۲]	الگوریتم مبتنی بر یال گمشده LightGBM	بهبود دقت پیش‌بینی	نیاز به دانش قبلی _ برای شبکه‌های بزرگ مناسب نیست

پیاده سازی سیستم پیشنهادی

روش پیشنهادی برای بهبود نتایج پیش‌بینی لینک براساس علم یادگیری ماشین (machine learning) و بهینه‌سازی (optimization) است که مراحل آن در شکل ۱ نشان داده‌شده‌است.



شکل ۱: گام‌های روش پیشنهادی

از آنجا که وجه نوآوری این پژوهش اعمال عملیات پیش‌پردازش روی داده‌ها می‌باشد که تاکنون در تحقیقات پیشین به آن پرداخته‌نشده‌است، لذا مراحل مربوط به آن، طی چهار گام اصلی در ادامه عنوان می‌گردد.

گام اول: دیتاست

مجموعه داده های استفاده شده در این پژوهش، متشکل از ۷ شبکه اجتماعی بزرگ و معتبر که از سایت دانشگاه استنفورد گرفته شده‌است، می‌باشد. هر مجموعه داده به صورت یک فایل تکست که شامل ارتباط های نودها با یکدیگر است، می‌باشد. ابتدا در کد نوشته شده این تکست ها خوانده شده و سپس تبدیل به یک ماتریس مجاورت برای گراف می شوند تا راحت تر قابل استفاده باشند. با توجه به بزرگ بودن شبکه های اجتماعی و نودها، ذخیره سازی هر مجموعه داده پس از تبدیل به ماتریس مجاورت در یک فایل mat. در متلب ذخیره شده است.

جدول ۲: جزئیات دیتاست استفاده‌شده در پژوهش

شبکه	تعداد رأس	تعداد یال
CiteSeer	۳۸۴,۴۱۳	۱,۷۵۱,۴۶۳
Hep-ph	۳۴,۵۴۶	۴۲۱,۵۷۸
Wiki	۱۶۷,۵۲۵	۱,۱۶۴,۵۷۶
Emails	۱۶۷	۵,۷۸۴
Movielens	۲,۶۲۵	۱۰۰,۰۰۰
US airports	۱,۵۷۴	۲۸,۲۳۶
Twitter	۸۱,۳۰۶	۱,۷۶۸,۱۴۹

دیتاست‌های استفاده شده در این پژوهش در جدول شماره ۲ نشان داده‌شده‌است.

گام دوم: استخراج ویژگی (Feature Extraction)

این مرحله یکی از مراحل بسیار بااهمیت در پیش‌بینی لینک است، چراکه نتایج حاصل از این مرحله، مستقیماً بر روی کیفیت مرحلهٔ پیش‌بینی اثر می‌گذارد. استخراج ویژگی فرایندی است که در آن با انجام عملیاتی بر روی داده‌ها، ویژگی‌های بارز و تعیین‌کنندهٔ آن مشخص می‌شود. هدف استخراج ویژگی این است که داده‌های خام به‌شکل قابل استفاده‌تری برای پردازش‌های آماری بعدی درآیند. هرچه کیفیت ویژگی‌های استخراج‌شده بیشتر باشد، مدل پیش‌بینی ساخته‌شده قدرتمندتر خواهد بود؛ به عبارت دیگر، این گام تأثیر مستقیم در نتیجهٔ پایانی خواهد داشت. انتخاب روش مناسب برای استخراج ویژگی نیز امری بسیار مهم محسوب می‌شود [۱۸].

با درنظر گرفتن موارد فوق، در این بخش ویژگی‌هایی که در این کار از گراف شبکهٔ اجتماعی استخراج شده و مورد استفاده قرارگرفته است، ارائه می‌شود. اگر $G = (V.E)$ گرافی باشد که ساختار توپولوژیک شبکهٔ اجتماعی را نشان می‌دهد، هر لبهٔ گراف با $e = (u.v) \in E$ نمایش داده می‌شود که $u.v \in V$ هستند. هدف ما ایجاد یک دسته‌بندی مناسب با استفاده از تکنیک‌های یادگیری ماشین است، به گونه‌ای که برای هر دو رأس $u.v$ بتواند پیش‌بینی کند که آیا اتصال بین این دو رأس دارای احتمال بالایی است یا خیر. لذا برای هر لبهٔ کاندید برای دسته‌بندی، یک مجموعه از ویژگی‌های برگرفته از ساختار توپولوژیک شبکه استخراج شده است. این ویژگی‌ها در جدول ۳ نمایش داده شد.

جدول ۳: ویژگی‌های ساختاری مربوط به جفت رأس‌های گراف شبکهٔ اجتماعی

نام ویژگی	نحوه محاسبه
ویژگی‌های درجه رأس	(1)$d(v) = \|\rho(v)\|$ (2)$\rho(v) = \{u\|(u.v)\epsilon\ E\ or\ (v.u)\epsilon E\}$
ویژگی‌های زیر گراف رأس	(3)$nh - subgraph(v) = \{(x.y)\epsilon E\|x.y\ \epsilon\ \rho(v)\}$ (4)$subgraph - Edge - Number(v) = \|nh - subgraph(v)\|$ (5)$Densiry - nh - subgraph(v) = {}^{d(v)}/_{\|nh - subgraph(v)\|}$
ویژگی‌های دوستان مشترک (common friends)	(6)$common - friends(u.v) = \|\rho(v) \cap \rho(u)\|$
شاخص آدامیک/آدار (Adamic/Adar Index)	(7)$adamic - adar - index(u.v) = \sum_{z\epsilon\rho(v)\cap\rho(u)} \frac{1}{\log\|\rho(z)\|}$
معیار همه دوستان	(8)$total - friends(u.v) = \|\rho(v) \cup \rho(u)\|$
ضریب جاکارد	(9)$jaccard - coefficient(u.v) = \frac{\|\rho(v)\cap\rho(u)\|}{\|\rho(v)\cup\rho(u)\|}$
اندیس ترفیع هاب (HPI) (Hub Promoted index)	(10)$HPI(u.v) = \frac{\|\rho(v)\cap\rho(u)\|}{\min\{\rho(v).\rho(u)\}}$
اندیس فشرده هاب (HDI) (Hub Depressed index)	(11)$HDI(u.v) = \frac{\|\rho(v)\cap\rho(u)\|}{\max\{\rho(v).\rho(u)\}}$
شاخص سالتون (Salton index)	(12)$Salton - index(u.v) = \frac{\|\rho(v)\cap\rho(u)\|}{\sqrt{\|\rho(v)\|\times\|\rho(u)\|}}$
شاخص سورنسون (Sorensen index)	(13)$sorenson - index(u.v) = \frac{\|\rho(v)\cap\rho(u)\|}{\|\rho(v)\|+\|\rho(u)\|}$
شاخصLHN (Leichet-Holme-Newman index)	(14)$LHN(u.v) = \frac{\|\rho(v)\cap\rho(u)\|}{\|\rho(v)\|\times\|\rho(u)\|}$
شاخص RA (Resource Allocation)	(15)$RA - index(u.v) = \sum_{z\epsilon\rho(v)\cap\rho(u)} \frac{1}{\rho(z)}$
امتیاز الحاق ترجیحی (Preferential attachment Score)	(16)$preferential - attachment(u.v) = \|\rho(v)\| \times \|\rho(u)\|$
معیار دوستان	(17)$freinds - measure(u.v) = \sum_{x\epsilon\rho(u)}\sum_{y\epsilon\rho(v)}\delta(x.y)$ $\delta(x.y) = \begin{cases}1 & if\ x = y\\ 0 & otherwise\end{cases}$, $w\ (u) = \frac{1}{\sqrt{1+\|\rho(u)\|}}$
معیار وزن لبه (Edge weight)	(18)$w(v) = \frac{1}{\sqrt{1+\|\rho(v)\|}}$
جمع وزن ها	(19)$w(u.v) = w(u) + w(v)$
ضرب وزن ها	(20)$w(u.v) = w(u) \times w(v)$
ضریب خوشه بندی (Clustering co-efficient)	(21)$clustering - coefficient(u.v) = \frac{\sum_{z\neq u,v} a_{uz}a_{zv}}{\min\{\rho(v).\rho(u)\}}$
همبستگی درجه (Degree correlation)	(22)$degree - correlation(u.v) = \frac{4.\|\rho(v)\|.\|\rho(u)\|-\|\rho(v)\|-\|\rho(u)\|}{2.\|\rho(v)\|^2+2.\|\rho(u)\|^2-\|\rho(v)\|-\|\rho(u)\|}$
روش FriendLink	(23)$friend - link(u.v) = \sum_{i=2}^{l} \frac{1}{i-1}.\frac{\|paths_{u,v}^{i}\|}{\prod_{j=2}^{l} n-j}$

گام سوم: انتخاب ویژگی (Feature Selection)

مسئلهٔ انتخاب ویژگی، یکی از مسائلی است که در مبحث یادگیری ماشین و همچنین شناسایی آماری الگو، مطرح است. این مسئله در بسیاری از کاربردها (مانند طبقه‌بندی) اهمیت بسزایی دارد، زیرا در این کاربردها تعداد زیادی ویژگی وجود دارد که بسیاری از آن‌ها یا بلااستفاده هستند و یا اینکه بار اطلاعاتی چندانی ندارند. حذف نکردن این ویژگی‌ها مشکلی از لحاظ اطلاعاتی ایجاد نمی‌کند، ولی بار محاسباتی را برای کاربرد مورد نظر بالا می‌برد؛ و علاوه بر این باعث می‌شود که اطلاعات غیرمفید زیادی را به همراه داده‌های مفید ذخیره کنیم [۱۹]. برای مسئلهٔ انتخاب ویژگی، راه‌حل‌ها و الگوریتم‌های فراوانی ارائه شده است که بعضی از آن‌ها قدمت سی یا چهل ساله دارند. مشکل بعضی از الگوریتم‌ها در زمانی که ارائه شده بودند، بار محاسباتی زیاد آن‌ها بود، اگرچه امروزه با ظهور کامپیوترهای سریع و منابع ذخیره‌سازی بزرگ، این مشکل به چشم نمی‌آید، ولی از طرف دیگر، مجموعه‌های داده‌ای بسیار بزرگ برای مسائل جدید باعث شده است که همچنان پیدا کردن الگوریتمی سریع و کم‌هزینه برای این کار مهم باشد. بسیاری از روش‌های مختلف انتخاب ویژگی کلاسیک، تلاش می‌کنند تا از میان 2^N زیرمجموعهٔ کاندید، بهترین زیرمجموعه را پیدا کنند. در تمام این روش‌ها براساس کاربرد و نوع تعریف، زیرمجموعه‌ای به‌عنوان جواب انتخاب می‌شود که بتواند مقدار یک تابع ارزیابی را بهینه کند. با وجود اینکه هر روشی سعی می‌کند که بتواند، بهترین ویژگی‌ها را انتخاب کند، اما با توجه به وسعت جواب‌های ممکن و اینکه این مجموعه‌های جواب به‌صورت توانی با N افزایش پیدا می‌کنند، پیدا کردن جواب بهینه، مشکل و در N های متوسط و بزرگ بسیار پرهزینه است. به‌طور کلی روش‌های مختلف انتخاب ویژگی را براساس نوع جست‌وجو به دسته‌های مختلفی تقسیم‌بندی می‌کنند. در بعضی روش‌ها تمام فضای ممکن جست‌وجو می‌شود. در سایر روش‌ها که می‌تواند مکاشفه‌ای و یا جست‌وجوی تصادفی باشد، در ازای از دست دادن مقداری از کارایی، فضای جست‌وجو کوچک‌تر می‌شود.

در حالت کلی این مسئله دارای راه‌حل قطعی نیست و تاکنون روشی دقیق برای حل آن پیشنهاد نشده است. رویکردهای مختلفی به صورت کلاسیک برای این مسئله‌ها پیشنهاد شده‌اند، که معمولاً کیفیت پاسخ آن‌ها به صورت عمومی چندان مناسب و مطلوب نیست، اما در مقابل، روش‌های بهینه‌سازی هوشمند می‌توانند در حل این مسائل، راه‌حل‌های به مراتب بهتری را ارائه دهند. از این رو، یکی از روش‌های مؤثر و سازنده در مسیر حل مسائل انتخاب ویژگی و مسائل مرتبط با آن، استفاده از روش‌های بهینه‌سازی فراابتکاری و الگوریتم‌های تکاملی است.

الگوریتم رقابت استعماری (Imperialist Competitive Algorithm) روشی در حوزهٔ محاسبات تکاملی است که به یافتن پاسخ بهینهٔ مسائل مختلف بهینه‌سازی می‌پردازد. این الگوریتم با مدل‌سازی ریاضی فرایند تکامل اجتماعی ـــ سیاسی، الگوریتمی برای حل مسائل ریاضی بهینه‌سازی ارائه می‌دهد. پایه‌های اصلی این الگوریتم را سیاست همسان‌سازی (assimilation)، رقابت استعماری (imperialistic competition) و انقلاب (revolution) تشکیل می‌دهند [۲۰]. این الگوریتم با تقلید از روند تکامل اجتماعی، اقتصادی و سیاسی کشورها و با مدل سازی ریاضی بخش‌هایی از این فرایند، عملگرهایی را در قالب منظم به‌صورت الگوریتم ارائه می‌دهد که می‌توانند به حل مسائل پیچیدهٔ بهینه‌سازی کمک کنند. درواقع این الگوریتم جواب‌های مسئلهٔ بهینه‌سازی را در قالب کشورها رصد و سعی می‌کند طی فرایندی تکرارشونده این جواب‌ها را رفته رفته بهبود دهد و درنهایت به جواب بهینهٔ مسئله بر ساند. همانند سایر روش‌های بهینه سازی تکاملی، کار این الگوریتم نیز با تعدادی جمعیت اولیه شروع می شود. به‌طور خلاصه در این الگوریتم، هر عنصر جمعیت، یک کشور نامیده می‌شود. کشورها به دو دسته مستعمره و استعمارگر تقسیم می شوند. هر استعمارگر، بسته به قدرت خود، تعدادی از کشورهای مستعمره را به سلطهٔ خود درمی‌آورد و آن‌ها را کنترل می‌کند. سیاست جذب و رقابت استعماری، هستهٔ اصلی این الگوریتم را تشکیل می‌دهند [۲۱].

این روش در مقابل سایر الگوریتم‌های کلاسیک که مجموعه حالات توانی 2^N داشته و زمان محاسبات را بسیار افزایش داده و محاسبات را پیچیده می‌کنند ـ دارای مجموعه‌حالات چندجمله‌ای N^2 است. برای انجام روش پیشنهادی، ابتدا ماتریس مجاورت گراف شبکهٔ اجتماعی به‌عنوان ورودی به برنامه اعمال می‌شود. سپس با استفاده از روش‌های استخراج ویژگی، ویژگی‌های تمام یال‌های موجود و غیرموجود محاسبه و ذخیره می‌شود. با توجه به حجم بالای گراف‌ها و ازدیاد تعداد یال‌ها، از ۵۰۰،۰۰۰ هزار یال

در هر بار اجرا استفاده می‌شود. این تعداد یال‌ها به دو دسته یال های موجود و غیرموجود تقسیم‌بندی می‌شوند. پس از جداسازی داده‌های آزمایش و آموزشی، وارد مرحلهٔ انتخاب ویژگی می‌شوند تا بهترین ویژگی‌ها استخراج شوند.

مراحل الگوریتم رقابت استعماری

۱. مقداردهی اولیه به پارامترها: در این قسمت پارامترهای الگوریتم ازجمله کشورهای اولیه، مقداردهی می‌شوند. بدین منظور هر نود ایجاد شده در شبکه، به یکی از کشورها اختصاص داده شده و به‌عنوان مکان اولیهٔ کشورها محسوب می شود [۲۲]. در این پژوهش پارامترهای اولیهٔ الگوریتم رقابت استعماری به شرح زیر مقداردهی شده‌اند. گفتنی است نحوهٔ مقداردهی این پارامترها طبق روال الگوریتم‌های فراابتکاری عموماً به دو صورت می‌تواند صورت گیرد: پیروی از سایر مقالات مشابه که با توجه به تفاوت در داده‌ها و اثرگذاری ساختار آن‌ها در نتایج، چندان توصیه نمی‌شود، یا روش آزمون و خطا که با درنظر گرفتن ماهیت غیرقطعی این‌گونه الگوریتم‌ها، متداول‌ترین روش است. در این پژوهش به روش آزمون و خطا انجام شده است.

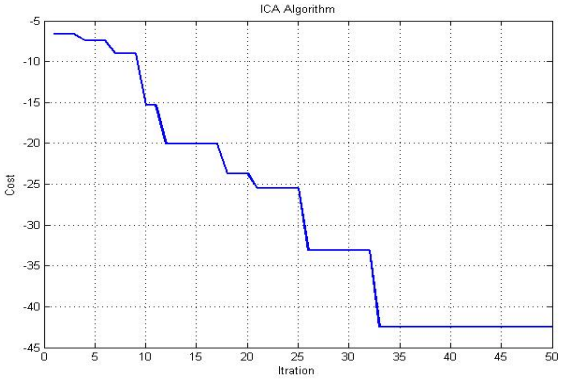
جدول ۴: مقداردهی پارامترهای اولیهٔ الگوریتم رقابت استعماری

پارامتر	مقدار
تعداد جمعیت اولیه	۲۰۰
تعداد تکرارهای برنامه	۱۰۰۰
تعداد امپراتوری‌ها	۳۰
تعداد مستعمره‌ها	۱۷۰
پارامتر بتا (نرخ حرکت مستعمره‌ها به سمت امپراتوری‌ها)	۰/۲
پارامتر زتا (نرخ محاسبهٔ میانگین قدرت مستعمره‌ها)	۰/۱
نرخ انقلاب	۰/۱

سپس براساس میزان شایستگی هر نود (در این مسئله میزان شایستگی هر نود، میزان شباهت به سایر نودهاست) استعمارگر و مستعمره مشخص می‌شود. میزان شایستگی هر نود از طریق تابع شایستگی محاسبه می‌شود. تابع شایستگی می‌تواند به روش‌های مختلفی تعیین شود. یکی از این روش‌ها، روش تصادفی است. در این روش هر بار به‌طور تصادفی تعدادی از ویژگی ها انتخاب می‌شوند و الگوریتم روی آن‌ها اجرا و خروجی گرفته می‌شود، درنهایت ویژگی هایی که با انتخاب آن ها خروجی مطلوب با کم‌ترین هزینه به‌دست آید، انتخاب می شود. روش دیگر استفاده از روابط است. در این روش وابستگی درون کلاسی ویژگی‌ها محاسبه می شود و از بین آن‌ها ویژگی‌هایی که کم‌ترین وابستگی را به سایر ویژگی‌ها دارند، حذف می‌شوند و ویژگی‌های دارای بیشترین وابستگی به هم، باقی می‌مانند. این مبحث در حوزهٔ هوش مصنوعی با عنوان تشخیص الگو (pattern recognition) شناخته می‌شود. در این پژوهش به‌منظور حصول بیشترین کارآیی، جهت تعیین میزان شایستگی ویژگی‌ها، از روش دوم استفاده شده است. تابع شایستگی مورد استفاده در الگوریتم رقابت استعماری در این پژوهش، ارزش هر ویژگی در بردار ویژگی را مشخص کرده و در انتها، پرارزش‌ترین ویژگی‌ها به‌منظور انجام عملیات دسته‌بندی نهایی استفاده می شوند. در رابطهٔ ۱ این تابع نشان داده شده است.

$$f(x) = Mu^{+} + Mu^{-}/Var^{+} + Var^{-} \qquad (۱)$$

بنابر رابطهٔ ۱، Mu^{+} برابر با میانگین ویژگی های مثبت، Mu^{-} بیانگر میانگین ویژگی‌های منفی، Var^{+} بیانگر واریانس ویژگی‌های مثبت و Var^{-} بیانگر واریانس ویژگی‌های منفی است. بعد از جای گذاری ویژگی ها در رابطهٔ فوق و محاسبهٔ تابع شایستگی آن‌ها، ادامهٔ مراحل اجرای الگوریتم انجام می شود. در نمودار ۱ اجرای این الگوریتم با ۵۰ دور تکرار نشان داده شده است.



نمودار ۱: نمودار حرکتی الگوریتم رقابت استعماری

در نمودار ۱، محور Xها بیانگر تعداد تکرارهای برنامه و محور Yها بیانگر میزان هزینهٔ محاسبه شده توسط تابع شایستگی بهدست آمده از تعداد ویژگی‌های انتخاب شده در تکرار مربوطه است. این نمودار بهترین مقدار بهدست آمده در کل تکرارها را نشان می‌دهد. به وضوح مشخص است با حرکت الگوریتم به سمت جلو، نتایج بهتری حاصل می‌شود.

۲. مستعمرات را به سمت کشور استعمارگر حرکت بده (سیاست همسان سازی): براساس قدرت هر استعمارگر (میزان شایستگی) مستعمره‌ها به استعمارگرها اختصاص داده می شوند. در ابتدای برنامه، این اختصاص به‌صورت برابر بین تمام استعمارگران خواهد بود. درنتیجه نودهای موجود در شبکه، براساس شایستگی، به دو دسته تقسیم می‌شوند و چندین امپراتوری تشکیل می‌دهند.

۳. اگر مستعمره‌ای در یک امپراتوری وجود داشته باشد که هزینه‌ای کم‌تر از امپریالیست داشته باشد؛ جای مستعمره و امپریالیست را باهم عوض کن. پس از اجرای الگوریتم و تغییرات موردنظر در هر کشور، میزان شایستگی هر کشور تغییر خواهد کرد. در صورتی که میزان شایستگی هر مستعمره بیشتر از استعمارگر شود، جای این دو باهم عوض می‌شود. هر نود با توجه به شایستگی در میزان شباهت در صورت انتخاب، جایش در امپراتوری عوض می‌شود.

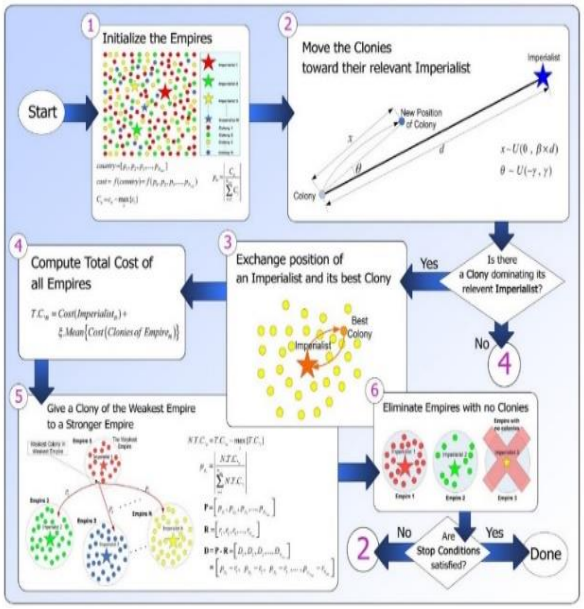
۴. هزینهٔ کل یک امپراتوری را حساب کن (با درنظر گرفتن هزینهٔ استعمارگر و مستعمراتشان): برای فهمیدن قدرت هر امپراتوری، باید هزینهٔ کل را محاسبه کرد. هر امپراتوری که شامل چندین نود در شبکه است، میزان هزینه‌اش محاسبه می‌شود.

۵. یک مستعمره از ضعیف‌ترین امپراتوری انتخاب کن و آن را به امپراتوری‌ای که بیشترین احتمال تصاحب را دارد، بده: پس از محاسبهٔ قدرت هر امپراتوری، استعمارگرهای قوی‌تر، مستعمره‌های استعمارگر ضعیف‌تر را تصاحب می‌کنند. بین نودهای شبکه این جابه‌جایی‌ها انجام می‌شود.

۶. امپراتوری‌های ضعیف را حذف کن: ضعیف‌ترین امپراتوری حذف می‌شود و استعمارگر این امپراتوری، به‌عنوان یک مستعمره درنظر گرفته می‌شود.

۷. تا پایان تعداد تکرارهای برنامه، این مراحل را تکرار کن: پس از اتمام تعداد تکرارهای برنامه، قوی‌ترین امپراتوری باقی می‌ماند که نود مختص به آن، همان نود رأس انتخاب‌شده برای خوشه است.

در شکل ۲ روال این الگوریتم نشان داده شده است [۲۲].



شکل ۲: شمایی از ترتیب انجام الگوریتم رقابت استعماری

گام چهارم: آموزش سیستم با داده‌ها و ویژگی‌های انتخاب‌شده

در این گام عملیات پیش‌بینی لینک بر روی دیتاست‌ها براساس ویژگی های بهدست‌آمده، اجرا می‌شود. در این پژوهش این عملیات یک بار با اعمال مرحلهٔ «انتخاب ویژگی» و بار دیگر بدون اعمال آن اجرا شد. علاوه بر آن این روش با استفاده از چندین الگوریتم متفاوت که از الگوریتم‌های شناخته‌شده و پرکاربرد در این حوزه هستند ازجمله الگوریتم‌های: KNN، Random Forest، SVM، Decision Tree، Bayes تکرار شد. شبه‌کد الگوریتم نیز به‌صورت زیر است:

۱. Separate training and testing data

۲. Operation ICA Algorithm

- For i=1 to max iteration
  - Initialize the parameters used
  - Create a Primary Population
  - Assimilation: Colony Move toward imperialist

- Revolution: Random Chang occur in the characteristics of some country
- Exchange Between City and Country
- Imperialist Competitions: all imperialist compete to take position of colonies of each others
- End for
- Select best output and use it

۳. Train Support Vector Machine and create model

۴. Use test data and Validation model

۵. Voting Answers for Access Final Solution

نتایج و تحلیل

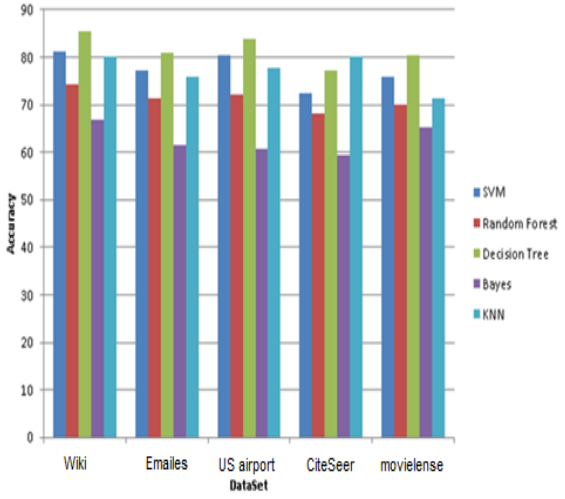
به‌منظور ارزیابی تأثیر انتخاب ویژگی در خروجی نهایی، در هر اجرا درصد صحت به‌عنوان شاخص اصلی و شاخص‌های Sensitivity، Specificity، F-Measure به‌منظور افزایش ضریب اتکا به نتایج، در شرایط با انتخاب ویژگی و بدون آن، محاسبه شد و نتایج در جدول ۵ نشان داده شده است:

جدول ۵: مقایسهٔ نتایج در حالت انتخاب ویژگی و بدون انتخاب ویژگی

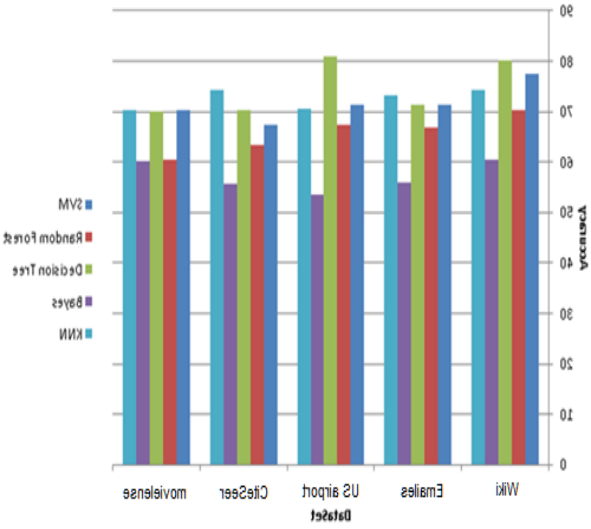
دیتاست	معیار	Sensitivity	Specificity	F-Measure	Accuracy
CiteSeer	با انتخاب ویژگی	۸۷/۲۲	۹۰	۸۸/۵۹	۸۷/۶۲
	بدون انتخاب ویژگی	۸۳/۰۶	۸۸/۵۳	۸۵/۷۱	۸۲/۸۶
Hep-ph	با انتخاب ویژگی	۸۰/۴۸	۹۰	۸۵	۸۰/۴۸
	بدون انتخاب ویژگی	۸۸/۸۹	۹۰	۸۴/۱۲	۸۸/۸۹
Wiki	با انتخاب ویژگی	۸۷/۰۶	۸۵	۸۸/۵۱	۸۷/۰۶
	بدون انتخاب ویژگی	۸۵/۵۹	۸۶	۸۷/۷۴	۸۴/۵۱
Emails	با انتخاب ویژگی	۸۱/۱۱	۸۸/۸۵	۸۴/۵۴	۸۱/۴۸
	بدون انتخاب ویژگی	۷۴/۴۴	۸۳/۰۸	۷۸/۳۵	۷۴/۰۷
Movielens	با انتخاب ویژگی	۸۲/۱۶	۸۷/۹۲	۸۴/۹۵	۹۱/۸
	بدون انتخاب ویژگی	۸۰/۲	۸۲	۸۱/۰۹	۸۵/۲۵
US airport	با انتخاب ویژگی	۸۵/۵۴	۸۱/۱۱	۸۲/۴۳	۸۴/۲۴
	بدون انتخاب ویژگی	۸۳/۷۱	۷۴/۰۷	۷۸/۰۳	۷۹/۸۹

با مشاهدهٔ جدول ۵ و مقایسهٔ نتایج بین دو حالت انتخاب ویژگی و بدون انتخاب ویژگی درمی‌یابیم که استفاده از این عملیات می‌توانند در نتایج نهایی تأثیر بسزایی بگذارد، زیرا در این دیتاست‌ها تعداد زیادی ویژگی وجود دارد، که بسیاری از آن‌ها یا بلااستفاده هستند و یا اینکه بار اطلاعاتی چندانی ندارند. حذف نکردن این ویژگی‌ها نه‌تنها بار محاسباتی را برای مسئلهٔ مورد نظر بالا می‌برد، بلکه صحت نتایج خروجی را نیز با توجه به زمان‌بر کردن اجرا، تحت تأثیر قرار می‌دهد و علاوه بر این باعث می‌شود که اطلاعات غیرمفید زیادی به همراه داده‌های مفید ذخیره شود. پس درنتیجه استفاده از این عملیات بسیار مفید خواهد بود. برای رسیدن به درکی بهتر از تأثیر عملیات انتخاب ویژگی، دو نمودار برای ۵ روش دسته‌بندی شناخته‌شده و پرکاربرد طراحی شده است که تأثیر آن را فارغ از نوع الگوریتم دسته‌بندی نمایش می‌دهد. با این آزمایش اهمیت انتخاب ویژگی به‌عنوان عاملی

مستقل در بهبود روند پیش‌بینی پیوند به خوبی نمایش داده می‌شود. نمودار ۲ بیانگر میزان درصد صحت هر روش دسته‌بندی در صورت استفاده از عملیات انتخاب ویژگی و نمودار ۳ بیانگر میزان درصد صحت هر روش دسته‌بندی در صورت استفاده نکردن از عملیات انتخاب ویژگی با دیتاست‌های مختلف است.



نمودار ۲: درصد صحت روش‌های مختلف دسته‌بندی در صورت استفاده از عملیات انتخاب ویژگی



نمودار ۳: درصد صحت روش‌های مختلف دسته‌بندی در صورت استفاده نکردن از عملیات انتخاب ویژگی

همانطور که در نمودارهای ۲ و ۳ نشان می‌دهد، استفاده از عملیات انتخاب ویژگی بسیار کارا و مفید است و نتایج را به‌صورت قابل توجهی بهبود می‌بخشد. در نمودار ۲ بیشتر نتایج، از نمودار ۳ عملکرد بهتری را از خود نشان داده و این امر نشان‌دهندهٔ برتری عملیات انتخاب ویژگی است.

نتیجه‌گیری

از آنجا که پژوهش‌های متعددی درخصوص پیش‌بینی لینک در حوزهٔ شبکه‌های اجتماعی انجام شده است و هر کدام با الگوریتم‌های متفاوتی به آن پرداخته‌اند که مزایا و معایبی را داشته است، در این پژوهش از بُعد دیگری به این چالش نگریسته شد و حجم بسیار زیاد داده‌های مورد پردازش که موجب می‌شود این گونه مسائل علاوه‌بر حوزهٔ یادگیری ماشین، در حوزهٔ مسائل داده‌های بزرگ (big data) نیز قرار گیرند، مورد نظر قرار گرفت و برای کاهش حداکثری عملیات پردازشی روی آن، روش انتخاب ویژگی پیشنهاد شد که ورای الگوریتم‌ها و روش‌های متعددی که می‌توان با آن‌ها به پیش‌بینی احتمال برقراری لینک پرداخت، موجب کاهش مؤثر و بسزایی در میزان اطلاعات مورد پردازش می‌شود. انتخاب ویژگی از میان تمام ویژگی‌های استخراج‌شده است که برای انجام آن الگوریم فرااکتشافی رقابت استعماری پیشنهاد شد و چنانچه در نتایج مشاهده گردید مطابق جدول ۵ اعمال این پیشنهاد، بهبود قابل ملاحظه ای را موجب شده که با معیارهای ارزیابی Sensitivity، Specificity، F-Measure و Accuracy قابل مشاهده است. همچنین به‌منظور بررسی دقیق میزان تأثیرگذاری این پیشنهاد، این بار مقایسهٔ دیگری با چند الگوریتم متفاوت و شناخته شده در این حوزه ازجمله الگوریتم های Random Forest، Decision Tree، Bayes و KNN صورت گرفت که نتایج آن در نمودارهای ۲ و ۳ نشان داده شد و به‌طور شگفت‌آوری نشان می‌دهد که تنها با اعمال این مورد، سایر روش های کنونی نیز بهبود قابل ملاحظه‌ای خواهند داشت. مقایسهٔ نمودارهای ۲و۳ که فاکتور صحت را با شرایط انتخاب ویژگی و بدون آن با دسته‌بندهای متفاوت نمایش می‌دهد، نتیجه می‌دهد که فارغ از انتخاب نوع دسته‌بند پیش‌بینی لینک، انتخاب ویژگی، در صحت خروجی تأثیرگذار است. برای نمونه در دیتاست CiteSeer در حالت با انتخاب ویژگی، صحت خروجی با مدل پیشنهادی که در جدول ۵ نشان داده شده است، ۸۷٫۶۲ درصد است و با همین شرط در ۵ دسته‌بند غیرترکیبی دیگر نیز که در نمودار ۲ نشان داده شده است، بالاترین درصد صحت را داراست. در سایر دیتاست‌ها نیز به همین صورت بوده و با استفاده از انتخاب ویژگی به نتایج بهبودیافته رسیدیم.

مراجع

[1] Kumar, A., et al. Link prediction techniques, applications, and performance: A survey. Physica A: Statistical Mechanics and its Applications, 553(6) (2020). https://doi.org/10.1016/j.physa.2020.124289.

[2] Shao, H., et al. Link prediction algorithms for social networks based on machine learning and HARP. IEEE Access, 7. pp. 122722-122729 (2019). https://doi.org/10.1109/ACCESS.2019.2938202.

[3] Sherkat, E., Rahgozar, M., Asadpour, M., Ant Colony Approach to Link Prediction in Social Networks, The CSI Journal on Computer Science and Engineering 12 (1) (2014) . 1-9. https://doi.org/10.1016/j.swevo.2018.03.001.

[4] Saadinezhad, H., Parvinniya, A., New link prediction criteria based on node composition and network structure. Soft Computing and Information Technology, 2, pp. 41-52 (1401). https://jscit.nit.ac.ir/article_151735_c67dc613817a84d127c6e844a821a0a3.pdf.

[5] Jalili, M., Orouskhani, Y., Asgari, M., Alipourfard, N., Perc, M. Link prediction in multiplex online social networks. Royal Society Open Science, 2: pp.160863 (2017). https://doi.org/10.1098/rsos.160863.

[6] Sharma, S., Singh, A. An efficient method for link prediction in weighted multiplex networks. Computational Social Networks, 3, no. 1: pp.7 (2016). https://doi.org/10.1186/s40649-016-0034-y.

[7] Yan, R., et al., SSDBA: the stretch shrink distance based algorithm for link prediction in social networks, Frontiers of Computer Science, 15(1) (2021), 1-8.

[۸] Tang, R., et al., Interlayer link prediction in multiplex social networks: an iterative degree penalty algorithm, Knowledge-Based Systems. 194 (2020). https://doi.org/10.1016/j.knosys.2020.105598.

[9] Sayyarifard, S., Presenting a new link prediction method in social networks using deep learning methods. Third Conference on Electrical, Computer and Mechanical Engineering, pp. 39-50(1399).

[10] Zare, H., Shokrzade, H., Link prediction in social networks using the clustering method using the expectation maximization algorithm. The fourth national conference of new technologies (1400).

[11] Chen, H., et al., Multi-level graph convolutional networks for cross-platform anchor link prediction, Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining (2020), 1503–1511. https://doi.org/10.1145/3394486.3403201.

[12] Alzubaidi, M. N. 2023. LightGBM for link prediction based on graph structure attributes. ICIC Express Letters, Vol 14, 3: pp. 303-3011 (2023).

[۱۳] Zare, H., & Shakorzadeh, H., Link prediction in social networks using clustering method with maximum expectation maximization algorithm, Fourth National Conference on Advanced Technologies in Electrical, Computer, and Mechanical Engineering in Iran. (2021). https://civilica.com/doc/1292870

[14] Sharma, A., et al., Link Prediction in Social Network using Artificial Neural Network. Int. J. Comput. Appl. 174 (2021), 26-30.

[15] Parvazeh, F., A. Harounabadi, and M. A. Naizari, A Recommender System for Making Friendship in Social Networks Using Graph Theory and users profile, Journal of Current Research in Science. 1(2016), 535.

[۱۶] Piltan, Y., & Mojarrad, M., Introducing a new method for link prediction in social networks based on metaheuristic algorithms. Kahraba Quarterly. 6(25) (2019). https://civilica.com/doc/1442443

[17] Ahuja, R., et al., Using hierarchies in online social networks to determine link prediction, Soft Computing and Signal Processing, Springer (2019), 67-76.

[18] Siyari Fard, S., Presenting a new method for link prediction in social networks using deep learning techniques, Third Conference on Electrical, Computer, and Mechanical Engineering, (2020), https://civilica.com/doc/118916

[19] Shaeghi, H., & A. Ghasemi, Modeling a multi-input multi-output system for simultaneous prediction of price and load in intelligent networks with load management, Computational Intelligence in Electrical Engineering, 6 (4) (2015), 1-87.

[20] Luo, J., J. Zhou, and X. Jiang, A Modification of the Imperialist Competitive Algorithm With Hybrid Methods for Constrained Optimization Problems, Journal of PeerJ Computer Science. 8 (2021), e1075. https://doi.org/10.7717/peerj-cs.1075.

[21] Li, X., J. Chen, L. Sun, and J. Li, A new imperialist competitive algorithm with spiral rising mechanism for solving path optimization problems, Journal of PeerJ Computer Science. 8(2022), e1075. https://doi.org/10.7717/peerj-cs.1075.

[22] Tao, Xin-rui., J.-Q. Li, T.-H. Huang and P. Duan, Discrete imperialist competitive algorithm for the resource-constrained hybrid flowshop problem with energy consumption, Journal of Complex & Intelligent Systems. 7(2021), 311–326. https://doi.org/10.1007/s40747-020-00193-w.

COPYRIGHTS
©2024 by the authors. Published by the Islamic Azad University, Khodabandeh Branch, Zanjan. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0) https://creativecommons.org/licenses/by/4.0