Journal of Applied And Basic Machine Intelligence Research (ABMIR)

پژوهش‌های نظری و کاربردی هوش ماشینی

سال دوم، شماره اول، بهار و تابستان ۱۴۰۳، صفحات ۱۴۱ - ۱۳۱

رویکردی نوین در کاهش ابعاد گراف مبتنی بر یادگیری عمیق با استفاده از منطق فازی و گام‌های تصادفی

محمد مهدی کیخا*، سامان براهویی

دانشکده ریاضی، آماروعلوم کامپیوتر، دانشگاه سیستان و بلوچستان زاهدان، ایران

مقاله پژوهشی

تاریخ دریافت:
۱۴۰۳/۰۸/۳۰

تاریخ پذیرش:
۱۴۰۳/۱۰/۱۱

کلیدواژه‌ها:
کاهش ابعاد گراف، یادگیری عمیق، منطق فازی، گام‌های تصادفی تعبیه‌گذاری گراف، شبکه‌های پیچیده

نویسنده‌مسئول:
keikha@cs.usb.ac.ir

چکیده

ساختارهای گراف نقش مهمی در مدل‌سازی روابط در زمینه‌هایی مانند شبکه‌های اجتماعی و زیستی دارند، اما مقیاس‌پذیری و ثبت روابط غیرخطی همچنان چالش‌برانگیز است. روش‌های سنتی تعبیه‌گذاری گراف اغلب در مواجهه با داده‌های بزرگ یا ویژگی‌های ناکامل گره‌ها ناکارآمد هستند. این مقاله چارچوب جدیدی به نام FuzzyRandomNet را معرفی می‌کند که با ترکیب منطق فازی و گام‌های تصادفی، این چالش‌ها را برطرف می‌کند. این روش با بهینه‌سازی ویژگی‌های گره و افزودن لایه‌های غیرخطی، تعبیه‌گذاری‌های مقیاس‌پذیر و دقیقی ارائه می‌دهد. ارزیابی‌ها روی مجموعه‌داده‌های استاندارد نشان می‌دهد که FuzzyRandomNet با دقت بسیار خوبی در کاربردهایی مانند دسته‌بندی گره‌ها و پیش‌بینی پیوندها عمل کرده و اثربخشی بالای این چارچوب را در شبکه‌های بزرگ و پیچیده نشان می‌دهد.

: 10.22034/abmir.2025.22416.1075

E-ISSN: 2821-2037 /© 2023. Published by Yazd University This is an open access article under the CC BY 4.0 License (https://creativecommons.org/licenses/by/4.0/).

۱- مقدمه

ساختارهای گراف به‌طور گسترده‌ای در مدل‌سازی روابط پیچیده کاربرد دارند و در مسائل عملی از اهمیت بالایی برخوردار هستند. به‌عنوان مثال، گراف‌های دانش در شبکه‌های اجتماعی و شبکه‌های زیستی از جمله موارد استفاده از آن‌ها محسوب می‌شوند. در این گراف‌ها، گره‌ها نماینده موجودیت‌ها هستند و یال‌ها ارتباطات بین آن‌ها را نمایش می‌دهند. بنابراین، تحلیل مؤثر گراف‌ها نیازمند استخراج ویژگی‌های مهم است که معمولاً از طریق ایجاد نمایش های کاهش یافته صورت می‌گیرد؛ چراکه استفاده مستقیم از ماتریس های پیچیده مجاورت می‌تواند به تحلیل ناکارآمد و پردازش سنگین منجر شود.

برای مقابله با این چالش‌ها، تکنیک‌های تعبیه‌گذاری گراف [1]توسعه یافته‌اند که این فرآیند را بهبود می‌بخشند. این تکنیک‌ها با تبدیل گره‌ها یا حتی کل گراف به فضاهای با ابعاد پایین‌تر، تلاش می‌کنند تا ساختارها و روابط اساسی گراف را حفظ کنند. در نتیجه، این نمایش فشرده از گراف می‌تواند در کاربردهای مختلفی نظیر اندازه گیری شباهت گره‌ها، دسته‌بندی گره‌ها، پیش‌بینی پیوندها و بخش بندی گراف مورد استفاده قرار گیرد. استفاده از این رویکردهای کاهش ابعاد، تحلیل گراف‌ها را کارآمدتر کرده و امکان بهبود دقت در مسائل مختلف را فراهم می‌کند[۱،۲].

نمایش داده‌های ساختاریافته در فضای برداری پیوسته تا حد زیادی به تعبیه‌گذاری گراف وابسته است. این تعبیه‌گذاری‌ها امکان به کارگیری تکنیک‌های یادگیری ماشین را بر روی گراف‌ها فراهم می کنند و به‌طور موثری وظایفی مانند دسته‌بندی گره‌ها، پیش بینی پیوندها و خوشه‌بندی گراف‌ها را تسهیل می‌کنند. با تبدیل ساختارهای پیچیده گراف به بردارهایی با ابعاد پایین‌تر، تعبیه‌گذاری ها اطلاعات ارتباطی و ساختاری موجود در گراف‌ها را حفظ می کنند، که می‌تواند منجر به افزایش کارایی بسیاری از کاربردهای پایین‌دستی شود.

به‌عنوان نمونه، این رویکرد به‌طور چشمگیری در تحلیل‌هایی مانند سیستم‌های توصیه‌گر، شبکه‌های زیستی و شبکه‌های اجتماعی مؤثر است، چراکه با کاهش ابعاد داده‌ها، پردازش سریع‌تر و کارآمدتری را ممکن می‌سازد. در نتیجه، استفاده از تکنیک‌های تعبیه‌گذاری گراف منابع محاسباتی مورد نیاز را کاهش می‌دهد، که این امر در کاربردهای عملی بسیار ارزشمند است[۳،۵].

با وجود دستاوردهای این تکنیک‌ها، روش‌های سنتی تعبیه‌گذاری گراف با چالش‌های متعددی روبه‌رو هستند. درک روابط پیچیده و غیرخطی بین گره‌ها برای بسیاری از روش‌ها دشوار است، به‌ویژه در گراف‌هایی که در آن‌ها گره‌های متصل اغلب دارای برچسب‌ها یا ویژگی‌های متفاوتی هستند. به‌عنوان مثال، در یک شبکه اجتماعی، گره‌ها ممکن است نمایانگر کاربران مختلف باشند که هر کدام ویژگی‌هایی نظیر سن، جنسیت یا علاقه‌مندی‌های متفاوت دارند. این تفاوت‌ها می‌توانند تحلیل و درک روابط پیچیده بین گره‌ها را برای روش‌های سنتی مشکل‌ساز کنند. افزون بر این، روش های پیشین محاسبات سنگینی دارند و در مقیاس‌پذیری برای گراف های بزرگ با محدودیت مواجه هستند.

در این میان، روش‌های مبتنی بر منطق فازی با بهبود در ثبت عدم قطعیت‌ها و مدیریت روابط پیچیده و ناهمگن، پتانسیل بالایی در حل این چالش‌ها و بهبود کیفیت تعبیه‌گذاری گراف‌ها دارند. این روش‌ها می‌توانند به‌طور مؤثر روابط مبهم و پیچیده بین گره‌ها را مدل‌سازی کرده و در نتیجه عملکرد بهتری را در تحلیل‌های مختلف ارائه دهند[۶،۸].

در این مقاله، روشی نوین با عنوان FRN [2]معرفی می‌شود که با استفاده از منطق فازی، گام‌های تصادفی، چالش‌های مذکور را برطرف می‌کند. این چارچوب به‌طور ویژه به مدیریت عدم‌قطعیت ها و روابط پیچیده در گراف‌ها پرداخته و برای اولین‌بار منطق فازی را با قدم‌زنی تصادفی تلفیق می‌نماید. این ترکیب باعث می‌شود تا اطلاعات ناهمگن و نامطمئن در شبکه به‌شیوه‌ای کارآمدتر و منعطف‌تر پردازش شود. افزون بر این،FRN با ارائه روشی جدید برای کاهش ابعاد و بهینه‌سازی نمایش گره‌ها، نه‌تنها به بیان‌پذیری و استحکام بالاتری دست یافته، بلکه محاسبات را نیز نزدیک‌تر به بهینه کرده است.

[1] Graph Embedding

[2] FuzzyRandomNet

۲- ادبیات و پیشینه پژوهش

در این بخش مهم‌ترین روش‌های تعبیه گذاری گراف مورد بررسی قرار گرفته‌اند که عبارتند از روش‌های سنتی تعبیه گذاری و روش‌های مبتنی بر شبکه‌های عصبی گرافی.

۲-۱ روش‌های سنتی تعبیه گذاری گراف

روش‌های کلاسیک تعبیه گذاری گراف برای نگاشت داده‌های گراف به فضایی با ابعاد کمتر طراحی شده‌اند که همزمان ساختار گراف را تا حد امکان حفظ می‌کنند. این روش‌ها برای تجزیه گراف های پیچیده ضروری هستند و تحلیل و استخراج الگوها را آسان‌تر می‌کنند. در ادامه برخی از مهمترین تحقیقات انجام شده در این گروه معرفی می‌شوند:

۲-۱-۱ روش‌های مبتنی بر مجاورت

روش‌های مبتنی بر مجاورت به تکنیک‌هایی گفته می‌شود که برای تحلیل و نمایش گراف‌ها از ماتریس‌های مجاورت استفاده می‌کنند. در این روش‌ها، ساختار گراف‌ها به‌صورت ماتریس‌هایی از روابط بین گره‌ها نمایش داده می‌شود و سپس با بهره‌گیری از این ماتریس ها، شباهت بین گره‌ها محاسبه می‌شود. این تکنیک‌ها در بسیاری از مسائل مانند دسته‌بندی گره‌ها، پیش بینی پیوندها و خوشه‌بندی گراف‌ها کاربرد دارند. در ادامه، مهم‌ترین روش‌های ارائه شده مبتنی بر مجاورت را بررسی می‌کنیم که به‌ویژه برای کاهش ابعاد و تحلیل ساختارهای پیچیده گراف‌ها مورد استفاده قرار می‌گیرند.

نقشه‌های لاپلاسی[1]: این تکنیک با استفاده از لاپلاسین گراف، شباهت بین گره‌ها را محاسبه می‌کند. در این روش، تجزیه مقادیر ویژه بر روی ماتریس لاپلاسین انجام می‌شود که گره‌ها را به فضای برداری پیوسته منتقل می‌کند؛ به طوری که در این فضا، گره‌های نزدیک به هم در گراف نیز به یکدیگر نزدیک هستند[۹].

نگاشت انتشار[2]: این روش از انتشار پویای گره‌ها برای ایجاد نقشه‌ای استفاده می‌کند که فاصله‌های بین گره‌ها را بر اساس تعداد مسیرهای احتمالی از یک گره به دیگری تعیین می‌کند. این روش به‌ویژه برای ثبت جریان‌ها و خوشه‌بندی گره‌ها مناسب است[۱۰].

۲-۱-۲ فاکتورگیری ماتریسی

روش‌های فاکتورگیری ماتریسی به تکنیک‌هایی اطلاق می‌شود که برای تجزیه و تحلیل گراف‌ها و داده‌ها، از ماتریس‌های مجاورت استفاده می‌کنند. هدف این روش‌ها، کاهش ابعاد داده‌ها به نحوی است که ویژگی‌های ساختاری و روابط مهم حفظ شوند. در این روش‌ها، ماتریس مجاورت به مجموعه‌ای از ماتریس‌های کوچک تر یا غیرمنفی تجزیه می‌شود تا امکان نمایش و پردازش داده‌ها با ابعاد کمتر فراهم گردد. در ادامه، به بررسی چند روش مهم در فاکتورگیری ماتریسی پرداخته می‌شود.

HOPE[3]: این روش با استفاده از فاکتورگیری مبتنی بر مجاورت مرتبه بالا، اطلاعات پیچیده‌تری از روابط بین گره‌ها استخراج می کند. این روش به‌طور ویژه برای گراف‌های بزرگ و پیچیده با روابط چندلایه، بسیار مؤثر است.HOPE قادر است ساختارهای پیچیده و الگوهای پنهان در گراف را به‌خوبی مدل‌سازی کند و نمایش‌های برداری دقیقی از گره‌ها ارائه دهد. این امر موجب بهبود عملکرد در وظایفی مانند دسته‌بندی گره‌ها، پیش‌بینی پیوندها، و تشخیص جوامع می‌شود و در نتیجه دقت بالاتری نسبت به روش های سنتی دارد[۱۱].

تجزیه مقادیر منفرد[4]: این روش با تجزیه ماتریس مجاورت به سه ماتریس کوچک‌تر، به کاهش ابعاد کمک می‌کند و همزمان ویژگی‌های ساختاری گراف را حفظ می‌نماید. به این ترتیب، می‌توان به نمایش با ابعاد کوچک تراز گره‌ها دست یافت.

یکی دیگر از تکنیک‌های رایج در فاکتورگیری ماتریسی، فاکتورگیری ماتریس غیرمنفی[5] است. این روش با تجزیه ماتریس مجاورت به دو ماتریس غیرمنفی کوچک‌تر، امکان بازسازی ماتریس اصلی را فراهم می‌آورد. به‌ویژه برای شناسایی جوامع و بهبود قابلیت تفسیر مدل بسیار کارآمد است[۱۲].

[1] Laplacian Maps
[2] diffusion map
[3] High-Order Proximity preserved Embedding
[4] SVD
[5] NMF

۲-۱-۳ روش‌های مبتنی بر قدم‌زنی تصادف

Deep walk : این روش به عنوان یکی از اولین مدل‌های مبتنی بر گام‌های تصادفی[1] شناخته می‌شود. در این مدل، مجموعه‌ای از گام‌های کوتاه تصادفی برای هر گره ایجاد می‌شود که دنباله‌هایی مشابه جملات در پردازش زبان طبیعی به وجود می‌آورد. این دنباله ها سپس برای به دست آوردن تعبیه گذاری‌هایی استفاده می‌شوند که ویژگی‌های محلی و سراسری گراف را حفظ می‌کنند[۱۳].

Node2vec: الگوریتم Node2vec یک روش یادگیری ویژگی های گره‌ها در گراف‌ها است که با استفاده از گام‌های تصادفی برای استخراج ویژگی‌های هر گره طراحی شده است. این الگوریتم با هدف یادگیری تعبیه گذاری گره‌ها به گونه‌ای که ارتباطات و شباهت‌های ساختاری بین گره‌ها را حفظ کند، عمل می‌کند. در فرآیند یادگیری، گراف را از طریق جستجوهای تصادفی پیمایش می‌کند و از این جستجوها برای آموزش مدل‌های زبانی مشابه با روش Word2Vec استفاده می‌کند. ویژگی منحصر به فرد Node2vec این است که از دو پارامتر پیکربندی به نام‌های p و q استفاده می‌کند که به مدل این امکان را می‌دهد تا بین جستجوی محلی و سراسری در گراف تعادل ایجاد کند، و بنابراین نتایج دقیق تری در یادگیری روابط بین گره‌ها به دست می‌آید[۱۴].

Wallets: یک روش یادگیری نمایه گره‌ها در گراف‌های بزرگ است که برای مقابله با مشکلات مقیاس‌پذیری جستجوهای تصادفی در گراف‌های پیچیده طراحی شده است. این روش به جای استفاده از جستجوهای تصادفی طولانی، از مسیرهای کوتاه‌تر استفاده کرده و ویژگی‌های آن‌ها را استخراج می‌کندWallets . با تقسیم جستجوها به بخش‌های کوچکتر و کمتر پیچیده، به کاهش زمان محاسباتی کمک می‌کند و در عین حال اطلاعات مهمی از روابط محلی و جهانی گره‌ها در گراف به دست می‌آورد. این الگوریتم برای گراف‌های بزرگ بهینه است و تعادلی بین دقت و مقیاس‌پذیری برقرار می‌کند[۱۵].

۲-۲ شبکه‌های عصبی گرافی

شبکه‌های عصبی گرافی به دلیل توانایی بالای آن‌ها در یادگیری نمایش‌های گرافی و کارایی در استخراج ساختارهای گراف، به عنوان ابزاری حیاتی برای یادگیری تعبیه گذاری در گراف‌ها شناخته می‌شوند[۱۶]. برخی از این شبکه‌ها شامل شبکه‌های پیچشی گراف هستند که در آن عملیات پیچش تعمیم داده شده و ویژگی‌ها از همسایگی محلی یک گره تجمیع می‌شوند. این شبکه‌ها در وظایف دسته‌بندی گره‌ها به‌صورت نیمه‌نظارت شده عملکرد بسیار خوبی از خود نشان داده‌اند. در شبکه‌های توجه گرافی، از مکانیسم توجه استفاده می‌شود تا وزن‌های هر لایه بر اساس گره‌های همسایه تنظیم شوند[۱۷].روش Graph SAGE با نمونه‌گیری و به‌روزرسانی ویژگی‌ها از طریق همسایگی نزدیک یک گره، به استخراج تعبیه گذاری گراف کمک می‌کند. این روش به دلیل رویکرد القایی خود که برای پیش بینی ویژگی‌های گره‌ها و گراف‌های نادیده مناسب است، به‌ویژه در گراف‌های پویا و بزرگ عملکرد مطلوبی ارائه می دهد[۱۷].

در ادامه این پیشرفت‌ها، منطق فازی به‌عنوان گسترشی ارزشمند برای منطق بولی سنتی معرفی‌شده است؛ به‌طوری که اطلاعات می توانند نه‌تنها به‌صورت کاملاً درست یا کاملاً نادرست، بلکه به درجات مختلف از هر دو حالت وجود داشته باشند. این مفهوم توسط لطفی زاده در سال 1965 مطرح شد[۱۸].و در مواجهه با ابهام و عدم قطعیت بسیار مؤثر است، به‌ویژه در سیستم‌های کنترلی، فرآیندهای تصمیم‌گیری و یادگیری ماشین. شبکه گراف پیچشی فازی عمیق با ترکیب منطق فازی در ساختار شبکه گراف پیچشی، بهبود قابل‌توجهی در دقت طبقه‌بندی ارائه می‌دهد. این روش عملکرد بالاتری را هم در نواحی ناهمگن و هم همگن تصاویر رادار قطبی نشان داده است[۱۹].

به‌طور کلی، روش‌های سنتی تعبیه‌گذاری گراف و شبکه‌های عصبی گرافی هر یک به‌نحوی به استخراج ویژگی‌ها و یادگیری ساختارهای پیچیده در گراف‌ها کمک می‌کنند. روش‌های مبتنی بر مجاورت و فاکتورگیری ماتریسی به حفظ ساختارهای محلی و جهانی گراف می‌پردازند، در حالی که روش‌های مبتنی بر قدم‌زنی

[1] Random Walks

تصادفی از ویژگی‌های احتمالاتی برای نمایش پیچیدگی‌های بیشتر استفاده می‌کنند. همچنین، شبکه‌های عصبی گرافی با تجمیع ویژگی ها از همسایگی گره‌ها و بهره‌گیری از مکانیسم توجه و نمونه‌گیری های القایی، عملکرد موفقی در دسته‌بندی گره‌ها و پیش بینی روابط از خود نشان داده‌اند.

با وجود پیشرفت‌های چشمگیر در روش‌های سنتی تعبیه‌گذاری گراف و شبکه‌های عصبی گرافی، این روش‌ها همچنان در ثبت روابط غیرخطی و مقیاس‌پذیری برای گراف‌های عظیم با چالش هایی مواجه هستند و اغلب نمی‌توانند به‌خوبی با عدم قطعیت و ابهامات موجود در داده‌ها مقابله کنند. در این زمینه، منطق فازی به عنوان رویکردی قدرتمند برای مدیریت عدم قطعیت‌ها معرفی‌شده است و با ترکیب آن در شبکه‌هایی نظیر شبکه گراف پیچشی فازی عمیق بهبود قابل‌توجهی در عملکرد طبقه‌بندی به‌دست‌آمده است. با این حال، تاکنون استفاده از منطق فازی برای کاهش ابعاد گراف کمتر مورد توجه قرارگرفته و هیچ رویکرد جامعی در این زمینه ارائه نشده است؛ این مسئله به‌عنوان فرصتی برای توسعه روش‌های نوآورانه مطرح می‌شود. در این راستا، روش FRN با ترکیب منطق فازی و گام‌های تصادفی، رویکردی جامع برای یادگیری تعبیه گذاری در گراف‌ها ارائه داده است که بهبود دقت و پایداری نتایج را در دسته‌بندی گره‌ها ممکن ساخته است.

۳- روش پیشنهادی

در این بخش به تشریح چارچوب پیشنهادی برای تولید پیوندها با استفاده از پیاده‌سازی گام تصادفی همراه با پردازش فازی و ارزیابی از طریق رگرسیون لجستیک پرداخته شده است. در این رویکرد، ابتدا با نمونه‌برداری از پیاده‌سازی‌های تصادفی، ویژگی‌های ساختاری گراف استخراج می‌شوند. سپس، پردازش فازی برای بهبود کیفیت ویژگی‌های تولیدشده اعمال می‌گردد. درنهایت، رگرسیون لجستیک به‌منظور ارزیابی اثربخشی پیوندهای یادگرفته‌شده به کار گرفته می‌شود.

۳-۱ پیاده‌سازی گام تصادفی

در این مرحله، مراحل ایجاد گام‌های تصادفی برای استخراج ویژگی های گراف شرح داده‌شده است تا از آن‌ها به‌منظور بهره‌برداری از ویژگی‌های ساختاری گراف استفاده شود. سپس، مراحل یادگیری تولید و روابط شبکه با استفاده از یادگیری عمیق آغاز شده است.

۳-۱-۱ تولید قدم زنی تصادفی

اولین مرحله شامل تولید قدم زنی‌های تصادفی در گراف برای به‌دست آوردن روابط بین گره‌ها است. برای هر گره، ما چندین قدم زنی تصادفی انجام می‌دهیم که هر کدام شامل تعداد ثابتی از گام‌های Tاست. فرآیند تولید قدم زنی تصادفی به‌صورت زیر تعریف می‌شود:

مقداردهی اولیه به گره فعلی:

$$\text{current_node} \leftarrow v_i$$

برای هر گام $t \in [1, T]$:

گره بعدی را از همسایه‌های گره فعلی بر اساس وزن‌های یال $\mathcal{W}(v_i, v_j)$نمونه‌برداری کنید:

$$P(\text{next}_{\text{node}} = v_i \mid \text{current}_{\text{node}} = v_j) = \frac{\mathcal{W}(v_i, v_j)}{\sum_{v_k \in N(v_i)} \mathcal{W}(v_i, v_j)} \quad (۱)$$

در اینجا $\mathcal{W}(v_i, v_j)$ وزن یال بین گره‌های v_iو v_jاست. $N(v_i)$ مجموعه همسایگان گره v_jاست. احتمال انتخاب گره v_iبه‌عنوان گره بعدی از همسایگان گره v_jبر اساس وزن یال‌ها محاسبه می‌شود.

۳-۱-۲ آموزش مدل Skip-gram

با استفاده از توالی‌های تولیدشده از قدم زنی‌های تصادفی، مدل Skip-gramبرای یادگیری پیوندهای گره به کار گرفته می‌شود. هدف مدل Skip-gram حداکثر کردن احتمال مشاهده گره‌های زمینه با توجه به یک گره هدف است. به‌طور رسمی، برای یک گره هدف و گره‌های زمینه، تابع هدف به‌صورت زیر تعریف می‌شود:

$$\mathcal{L} = -\sum_{(v_i, v_j) \in D} log\, P(v_i | v_j) \quad (۲)$$

احتمال با استفاده از تابع softmaxمحاسبه می‌شود:

$$P(v_i | v_j) = \frac{e^{h_{v_j}^T h_{v_i}}}{\sum_{v_k \in V} e^{h_{v_k}^T h_{v_i}}} \quad (۳)$$

که در آن و به ترتیب پیوندهای گره‌های و هستند و T نمایانگر گام در فرآیند قدم زنی تصادفی است این پیوندها با استفاده از

تکنیک‌های کاهش گرادیان برای حداقل کردن تابع هزینه بهینه می شوند.

۳-۲ پیش‌پردازش فازی

در این بخش، مراحل پردازش فازی به‌منظور بهبود استخراج ویژگی‌ها از گره‌ها ارائه شده است. نخستین گام در این مرحله، تعریف تابع عضویت فازی است که در ادامه تشریح می‌شود.

۳-۲-۱ تعریف توابع عضویت فازی

مقادیر پارامترهای مثلثی به‌صورت a=0، b=0.5، c=1 انتخاب شده‌اند، زیرا این مقادیر بازه طبیعی ویژگی‌ها را پوشش می‌دهند و امکان نرمال‌سازی بهتر را فراهم می‌کنند.

برای بهبود نمایش ویژگی‌های گره‌ها، توابع عضویت فازی بر اساس ویژگی‌های گره‌ها تعریف می‌شوند. تابع عضویت مثلثی با پارامترهای a، b و c به‌صورت زیر تعریف می‌شود:

$$\mu(x;a,b,c)=\begin{cases}0 & \text{if } c\le x\le a\\ \dfrac{x-a}{b-a} & \text{if } a<x\le b\\ \dfrac{c-x}{c-b} & \text{if } b<x<c\end{cases} \quad (٤)$$

در اینجا، x به‌عنوان ورودی تابع عضویت فازی در نظر گرفته می‌شود، که ویژگی‌های عددی گره‌ها را نمایندگی می‌کند و از این تابع جهت محاسبه ویژگی‌های فازی هر گره استفاده می‌شود که در گام بعدی تشریح شده است.

۳-۲-۲ محاسبه ویژگی‌های فازی

برای هر ویژگی در مجموعه ویژگی‌ها، نمایش فازی آن با استفاده از توابع عضویت تعریف‌شده محاسبه می‌شود. این فرآیند به تولید یک ماتریس ویژگی فازی منجر می‌شود که هر عنصر مقدار عضویت فازی ویژگی برای گره است:

$$M_{ij}=\mu_i(f_j) \quad (٥)$$

با ترکیب این ویژگی‌ها یک بردار فازی برای هر گره به دست می آید. در محاسبه ویژگی‌های فازی برای هر گره، i و j به ترتیب از ۱ تا n واز ۱ تا m تغییر می‌کنند، که n تعداد گره‌ها و m تعداد ویژگی‌ها را نشان می‌دهد

بردار ویژگی فازی برای گره به‌صورت زیر نمایش داده می‌شود:

$$M_i=[M_{i1},M_{i2},\dots,M_{im}] \quad (٦)$$

۳-۲-۳ نرمال‌سازی ویژگی‌های فازی

برای اطمینان از سازگاری، نرمال‌سازی مانند مقیاس بندی مین-ماکس به ماتریس ویژگی فازی اعمال می‌شود:

$$M_{ij}'=\frac{M_{ij}-min(M_i)}{max(M_i)-min(M_i)} \quad (٧)$$

که در آن M_i نمای i -ام از ماتریس‌است M_i. بردار ویژگی فازی نرمال‌شده به‌صورت زیر است

$$M_i'=[M_{i1}',M_{i2}',\dots,M_{im}'] \quad (٨)$$

۳-۲-۴ تجمیع فازی

به‌عنوان آخرین مرحله در پردازش فازی گراف، نمای کلی فازی هر گره با استفاده از مجموع وزنی ویژگی‌های فازی همسایگان آن محاسبه می‌شود. به‌طور رسمی، اگر مجموعه همسایگان گره $N(i)$ باشد، نمای کلی فازی گره $Magg_i$ به‌صورت زیر محاسبه می‌شود:

$$Magg_i=\sum_{j\in N(i)}\mathcal{W}(v_i,v_j)\cdot M_j' \quad (٩)$$

در فرآیند پردازش فازی گراف، نمای کلی فازی هر گره با استفاده از مجموع وزنی ویژگی‌های فازی همسایگان آن محاسبه می‌شود. وزن $\mathcal{W}(v_i,v_j)$ با روشی نظیر شباهت ویژگی‌ها تعیین می‌شود. نمای کلی فازی گره $Magg_i$ نشان‌دهنده تجمیع ویژگی‌های فازی همسایگان آن گره است و می‌تواند به‌عنوان ترکیب یا نمای نهایی از اطلاعات فازی مربوط به گره i در نظر گرفته شود.

۳-۳ ارزیابی رگرسیون لجستیک

۳-۳-۱ آماده‌سازی مجموعه‌های آموزش و آزمایش

ویژگی‌های نرمال‌شده فازی برای آماده‌سازی مجموعه‌های آموزش و آزمایش استفاده می‌شوند. در این مرحله یک مسئله طبقه‌بندی دودویی تعریف می‌شود که در آن متغیر هدف نشان‌دهنده برچسب های کلاس برای هر گره است. به‌طور رسمی، بردار هدف دوتایی در نظر گرفته می‌شود، به‌طوری که نشان‌دهنده وجود یک ویژگی خاص و نشان‌دهنده عدم وجود آن است.

۳-۳-۲ آموزش مدل رگرسیون لجستیک

از مدل رگرسیون لجستیک برای ارزیابی اثربخشی پیوندهای یادگرفته‌شده استفاده می‌شود. مدل رگرسیون لجستیک به‌صورت زیر بیان می‌شود:

$$P(y = 1 \mid X) = 1\frac{1}{1 + e^{-\beta^{T}x}} \quad (۱۰)$$

در فرآیند مدل‌سازی، x نمایانگر ویژگی‌ها یا پیوندها، y متغیر هدف و βضرایب مدل هستند که ارتباط میان ویژگی‌ها و متغیر هدف را مشخص می‌کنند.

ضرایب β با حداکثر کردن تابع احتمال لگاریتمی تخمین زده می‌شوند:

$$\mathcal{L}(\beta) = \sum_{i=1}^{n}(Y_i log(P(y = 1 \mid X_i;\beta)) + (1 - Y_i)log(1 - P(y = 1 \mid X_i;\beta))) \quad (۱۱)$$

تابع $\mathcal{L}(\beta)$ یک تابع هدف در رگرسیون لجستیک است که برای به‌دست آوردن مقادیر مدل استفاده می‌شود. این تابع احتمال لگاریتمی را برای هر نمونه i محاسبه کرده و مجموع آن‌ها را برای تمامی نمونه‌ها n می‌گیرد. در این تابع، Y_iبرچسب واقعی کلاس برای نمونه i است و $P(y = 1 \mid X_i;\beta))$ احتمال پیش‌بینی‌شده برای کلاس ۱ برای ویژگی‌های X_iاست. هدف این است که مقادیر مدل به‌گونه‌ای تنظیم شوند که این احتمال‌ها بیشترین تطابق را با برچسب‌های واقعی Y_i داشته باشند.

۳-۳-۳ ارزیابی مدل

درنهایت، عملکرد مدل رگرسیون لجستیک با استفاده از معیار F1 ارزیابی می‌شود. ارزیابی به‌گونه‌ای صورت می‌گیرد که از مقادیر پیش‌بینی‌شده و واقعی برای محاسبه مقادیر مربوط به‌دقت، فراخوانی و F1 استفاده می‌شود. این ارزیابی به‌منظور سنجش کیفیت مدل در وظایف دسته‌بندی و پیش‌بینی پیوندها انجام می‌شود.

امتیاز Macro-F1با استفاده از فرمول زیر محاسبه می‌شود:

$$Macro - F1 = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + \frac{1}{2}(FP_i + FN_i)} \quad (۱۲)$$

در تحلیل عملکرد یک مدل، T نشان‌دهنده تعداد پیش‌بینی‌های مثبت صحیح،F تعداد مثبت‌های کاذب و N تعداد کل دسته‌ها را نشان می‌دهد. همچنین، تعداد ارزیابی‌های منفی کاذب به معنای عدم توانایی مدل در شناسایی موارد مثبت است، که به درک دقیق‌تر عملکرد مدل کمک می‌کند.

٤- ارزیابی روش پیشنهادی

٤-١ مجموعه داده آزمایش

در این بخش، چندین آزمایش به‌منظور ارزیابی اثربخشی چارچوب FRNبا استفاده از چهار مجموعه داده پرکاربرد در حوزه تعبیه گذاری گراف [۲۰] ، انجام شده است: دو مجموعه داده شبکه استنادی شامل Cora وCiteseer ، و دو مجموعه داده شبکه صفحات وب شامل Chameleon و Squirrel این مجموعه داده ها فرصت‌های گسترده‌ای را برای آزمایش مدل در شبکه‌هایی با معماری‌ها و ویژگی‌های مختلف گره‌ها فراهم می‌کنند.

جدول (۱): مجموعه دادهای آزمایش

Datasets	Nodes	Classes	Edges	Homophil e Rate
Cora	2708	7	5429	0.809
Citeseer	3327	6	4739	0.721
Chameleon	2277	5	36101	0.233
Squirrel	5201	5	217073	0.203

در جدول ۱، اطلاعات مربوط به چهار مجموعه داده پرکاربرد در حوزه تعبیه‌گذاری گراف ارائه شده است. این داده‌ها شامل تعداد گره‌ها، کلاس‌ها، پیوندها و نرخ هموفیلیتی برای هر مجموعه داده است.

شبکه‌های استنادی ساختارهای گرافی مهمی را تشکیل می‌دهند که در آن گره‌ها نمایانگر مقالات علمی هستند و یال‌ها ارتباطات استنادی میان این مقالات را نشان می‌دهند. گره‌ها اغلب ویژگی هایی مانند مجموعه کلمات برای محتوای مقاله دارند و برچسب‌ها گره‌ها را به حوزه‌های تحقیقاتی مختلف طبقه‌بندی می‌کنند.

مجموعه داده Cora شامل ۲۷۰۸ مقاله علمی است که به‌عنوان گره‌ها در نظر گرفته می‌شوند. این مقالات به هفت گروه مختلف تقسیم می‌شوند که عبارت‌اند از: استدلال مبتنی بر مورد، الگوریتم

های ژنتیکی، شبکه‌های عصبی، روش‌های احتمالاتی، یادگیری تقویتی، یادگیری قواعد و نظریه Cora شامل ۵۴۲۹ یال است که ارتباطات استنادی میان مقالات را نشان می‌دهد. هر مقاله در Cora دارای یک بردار ویژگی ۱۴۳۳ کلمه‌ای است که محتوای مقاله را برای مطالعه عمیق و دسته‌بندی درون مجموعه داده‌ها ثبت می کند[۲۱].

کتابخانه دیجیتال Citeseer به‌عنوان منبع برای مجموعه داده Citeseerعمل می‌کند. این مجموعه شامل ۳۷۰۳ کلمه مختلف به عنوان ویژگی‌های هر مقاله است و ۳۳۲۷ گره دارد که هر گره نمایانگر یک مقاله است. مجموعه داده Citeseer شامل ۱۰۵۴۳ یال است که پیوندهای استنادی بین مقالات را نشان می‌دهد. مقالات در این مجموعه به شش حوزه طبقه‌بندی می‌شوند :عامل‌ها، هوش مصنوعی، پایگاه داده‌ها، بازیابی اطلاعات، یادگیری ماشین و تعامل انسان و کامپیوتر. این مجموعه نیز همانند Cora از کیسه کلمات برای نمایش ویژگی‌های گره‌ها استفاده می‌کند[۲۱].

شبکه‌های Wikipedia برای Chameleon و Squirrel از ویکی پدیای انگلیسی در دسامبر ۲۰۱۸ به‌دست‌آمده‌اند. در این شبکه‌ها، مقالات به عنوان گره‌ها و پیوندهای مشترک بین مقالات به عنوان یال‌ها در نظر گرفته شده‌اند. ویژگی‌های گره از حضور اسامی در مقالات و میانگین ترافیک ماهانه از اکتبر ۲۰۱۷ تا نوامبر ۲۰۱۸ به‌دست‌آمده‌اند. هدف این مجموعه داده‌ها پیش‌بینی لگاریتم میانگین ترافیک ماهانه برای دسامبر ۲۰۱۸ است[۲۲].

شبکه Squirrelدارای ۵۲۰۱ گره در ۵ گروه است و شامل ۲۱۷۰۷۳ یال است که ارتباطات بین مقالات را نشان می‌دهند. شبکه Chameleon شامل ۲۲۷۷ گره در ۵ دسته است و دارای ۲۳۲۵ ویژگی و ۳۶۱۰۱ یال است که پیوندهای بین مقالات را نشان می‌دهد [۲۲].

جدول ۲ عملکرد مدل FRN را در زمینه طبقه‌بندی گره‌ها در گراف های همگن و ناهمگنشان می‌دهد. به‌ویژه، امتیاز F1 مدل FRN در دیتاست‌های Cora و Citeseer به ترتیب معادل ۷۶،۴ و ۷۰،۲ به‌دست‌آمده است. این نتایج نشان‌دهنده این است که FRN توانسته است روابط بین ویژگی‌ها را به‌خوبی شناسایی و استفاده کند و عملکرد طبقه‌بندی را در سناریوهای همگن بهبود بخشد. با این حال، عملکرد این مدل در گراف‌های ناهمگن نشان‌دهنده برخی محدودیت‌ها است، زیرا FRN در دیتاست‌های Squirrel و Chameleon به ترتیب امتیاز ۳۴،۰ و ۵۳،۱ را به‌دست آورده است. این کاهش چشم‌گیر تأکید می‌کند که FRN در مواجهه با ساختارهای پیچیده ناهمگن با چالش‌هایی روبرو است و نیاز به بهبودهای بیشتری برای افزایش اثربخشی آن در این زمینه‌ها احساس می‌شود.

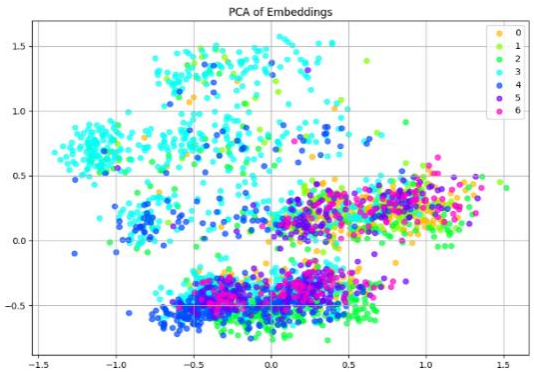
جدول (۲): طبقه‌بندی گره‌ها با استفاده از معیار ماکرو F1 در گراف‌های همگن و ناهمگن است

Chameleon	Squirrel	Citeseer	Cora	Datasets
NA	NA	41.6	67.2	Deep Walk
NA	NA	51.1	72.3	node2vec
73.2	24.1	73.2	85.8	GCN
74.4	30.0	74.4	86.5	GAT
71.5	36.2	71.5	88.1	Graph SAGE
53.1	34.0	70.2	76.4	FRN

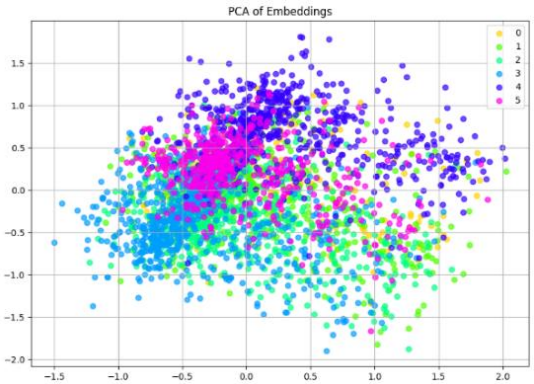
موفقیت متعادل مدل FRN به‌عنوان یک گزینه مناسب در مقایسه با روش‌های سنتی مانند Deep Walk و node2vec که در این زمینه‌ها عملکرد ضعیفی دارند، تأکید می‌کند. توانایی FRN در حفظ سطح عملکرد قوی در دیتاست‌های Cora و Citeseer نشان دهنده کارایی رویکردهای مبتنی بر ویژگی در بهره‌برداری از توپولوژی گراف است. با این حال، کاهش محسوس عملکرد در دیتاست‌های ناهمگن نشان‌دهنده این است که ممکن است FRN نیاز به بهبودهای اضافی یا استراتژی‌های ترکیبی داشته باشد که رویکرد متمرکز بر ویژگی آن را با سایر تکنیک‌های یادگیری گراف ترکیب کند. این امر می‌تواند به افزایش سازگاری و قدرت مدل کمک کند، به‌ویژه در سناریوهایی که در آن‌ها گره‌ها روابط متنوع و پیچیده‌ای دارند. بنابراین، تحقیقات بیشتری در مورد معماری FRNو کاربرد آن در زمینه‌های هتروفیلیک ضروری به‌نظر می‌رسد تا به‌طور کامل از قابلیت‌های آن بهره‌برداری شود.

۴-۲ تجسم تعبیه گذاری FRN

نمودارها توزیع‌های تعبیه‌های تولیدشده توسط FRN برای مجموعه داده‌های Cora (شکل ۱) و Citeseer (شکل ۲) نشان داده می‌شوند. در مجموعه دادهCora ، تعبیه‌ها به وضوح تشکیل خوشه



شکل (۱): تجسم مجموعه داده Cora (A)



شکل (۲): تجسم مجموعه داده Citeseer (B)

می‌دهند که نشان‌دهنده توانایی FRN در جداسازی مرزهای کلاس‌ها و گروه‌های مختلف در گراف است. در مجموعه داده Citeseer، با اینکه هم‌پوشانی بیشتری بین نقاط مشاهده می‌شود، اما همچنان تمایل به خوشه‌بندی قابل مشاهده است که نشان‌دهنده توانایی مدل در شناسایی ساختارهای جامعه‌ای در داده‌های پیچیده‌تر و دارای نویز بیشتر است. این نتایج، قدرت FRN را در تفکیک روابط پیچیده و ابعاد بالا در شبکه‌های ناهمگن با استفاده از منطق فازی برای یادگیری تعبیه‌های بهبودیافته نشان می‌دهند.

۵- نتیجه‌گیری

FRN به‌عنوان یک روش پیشرفته و نوآورانه برای تعبیه‌های گراف معرفی می‌شود که با ترکیب منطق فازی و قدم‌زنی تصادفی، توانایی مناسبی در مدیریت عدم قطعیت و ابهام موجود در داده‌های شبکه‌ای دارد. این مدل قادر است نمایه‌های باکیفیت و قابل تفسیری تولید کند که ساختارهای گروهی و طبقه‌بندی‌های پنهان در داده‌های پیچیده را به‌طور مؤثری نمایان می‌سازد. نتایج حاصل از تحلیل PCA بر روی مجموعه داده‌های Cora و Citeseer نشان داد که FRN توانایی بالایی در شناسایی خوشه‌ها و الگوهای پنهان با دقت و کارایی مطلوب، حتی در گراف‌های پیچیده با هم‌پوشانی زیاد داشته است. سرعت و دقت بالای این روش در طبقه‌بندی، آن را به گزینه‌ای مناسب برای کار با گراف‌های بزرگ و پیچیده تبدیل کرده بود. درنهایت، نتایج نشان دادند که FRN به دلیل استفاده از منطق فازی و انعطاف‌پذیری بالا، توانسته است با چالش‌های رایج در نمایه‌سازی گرافی مانند نویز و ساختارهای پیچیده مقابله کند. علاوه بر این، این مدل توانایی بهتری در یادگیری از داده‌های ناهمگن نشان داده و برای کاربردهایی همچون شبکه‌های اجتماعی، سیستم‌های توصیه‌گر و تحلیل داده‌های علمی مناسب بوده است. FRNپتانسیل توسعه بیشتری دارد و می‌توان آن را با مدل‌های پیچیده‌تری مانند GNN و GCN ترکیب کرد تا عملکرد بهتری حاصل شود. ترکیب FRN با مدل‌هایGNN و GCN می‌تواند دقت، مقیاس‌پذیری و مقاومت در برابر نویز را بهبود بخشد. این ادغام از انعطاف‌پذیری FRN در مدیریت ابهام و توانایی یادگیری عمیق GNN و GCN بهره می‌برد و تحلیل ساختارهای پیچیده و ناهمگن را تسهیل می‌کند. این ترکیب می‌تواند کاربردهای گسترده‌تری در پیش‌بینی لینک، طبقه‌بندی گره‌ها، و تحلیل شبکه‌های بزرگ ارائه دهد واین امر افق‌های جدیدی را برای استفاده ازFRN در مدل‌های پیشرفته و تحلیل گراف‌های ناهمگن فراهم می‌آورد.

References

[1] J. Smith and R. Johnson, "Advances in graph embedding techniques for efficient analysis," Journal of Machine Learning Research, vol. 25, no. 3, pp. 45–60, 2024
[2] L. Chen and S. Kumar, "Dimensionality reduction in graphs: Methods and applications," IEEE Transactions on Knowledge and Data Engineering, vol. 36, no. 5, pp. 1012–1025, 2024.
[3] Wang, Y., & Li, J. (2023). Graph embedding techniques for complex network analysis: A comprehensive review. IEEE Transactions on

Neural Networks and Learning Systems, 34(1), 15-32.
[4] Zhang, H., & Chen, L. (2023). Advancements in node classification using graph embedding methods. Journal of Machine Learning Research, 24(2), 123-145.
[5] Liu, X., & Zhao, Y. (2023). Link prediction in social networks through graph embedding: A survey. ACM Computing Surveys, 55(3), Article 45.
[6] Wang, F., & Zhang, T. (2023). Fuzzy logic-based graph embeddings for uncertain networks. IEEE Transactions on Fuzzy Systems, 31(2), 245-258
[7] Chen, L., & Yang, H. (2023). Modeling complex relationships in heterogeneous graphs using fuzzy embeddings. Knowledge-Based Systems, 256, Article 110234.
[8] Li, J., & Xu, K. (2023). Enhancing graph representation with fuzzy logic: A survey. ACM Transactions on Intelligent Systems and Technology, 14(3), Article 56.
[9] M. Belkin and P. Niyogi/ "Laplacian eigenmaps for dimensionality reduction and data representation/" Neural computation/ vol. 15/ no. 6/ pp. 1373-1396/ 2003.
[10] R. R. Coifman and S. Lafon/ "Diffusion maps/" Applied and computational harmonic analysis/ vol. 21/ no. 1/ pp. 5-30/ 2006.
[11] M. Coskun/ "A high order proximity measure for linear network embedding/" Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi/ vol. 11/ no. 3/ pp. 477-483/ 2022.
[12] D. D. Lee and H. S. Seung/ "Learning the parts of objects by non-negative matrix factorization/" nature/ vol. 401/ no. 6755/ pp. 788-791/ 1999.
[13] B. Perozzi/ R. Al-Rfou/ and S. Skiena/ "Deepwalk: Online learning of social representations/" in Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining/ 2014/ pp. 701-710.
[14] A. Grover and J. Leskovec/ "node2vec: Scalable feature learning for networks/" in Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining/ 2016/ pp. 855-864.
[15] B. Perozzi/ V. Kulkarni/ and S. Skiena/ "Walklets: Multiscale graph embeddings for interpretable network classification/" arXiv preprint arXiv:1605.02115/ pp. 043238-23/ 2016.
[16] P. Veličković/ G. Cucurull/ A. Casanova/ A. Romero/ P. Lio/ and Y. Bengio/ "Graph attention networks/" arXiv preprint arXiv:1710.10903/ 2017.
[17] W. Hamilton/ Z. Ying/ and J. Leskovec/ "Inductive representation learning on large graphs/" Advances in neural information processing systems/ vol. 30/ 2017.
[18] L. A. Zadeh/ "Fuzzy sets/" Information and Control/ 1965.
[19] H. Liu/ T. Zhu/ F. Shang/ Y. Liu/ D. Lv/ and S. Yang/ "Deep fuzzy graph convolutional networks for PolSAR imagery pixelwise classification/" IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing/ vol. 14/ pp. 504-514/ 2020.
[20] Z. Yang, W. Cohen, and R. Salakhudinov, "Revisiting semi-supervised learning with graph embeddings," in International Conference on Machine Learning, PMLR, Jun. 2016, pp. 40–48.
[21] A. K. McCallum/ K. Nigam/ J. Rennie/ and K. Seymore/ "Automating the construction of internet portals with machine learning/" Information Retrieval/ vol. 3/ pp. 127-163/ 2000.
[22] D. Bo/ X. Wang/ C. Shi/ M. Zhu/ E. Lu/ and P. Cui/ "Structural deep clustering network/" in Proceedings of the web conference 2020/ 2020/ pp. 1400-1410.
[23] B. Rozemberczki/ C. Allen/ and R. Sarkar/ "Multi-scale attributed node embedding/" Journal of Complex Networks/ vol. 9/ no. 2/ p. cnab014/ 2021.

Journal of Applied And Basic Machine Intelligence Research (ABMIR)

Applied and Basic Machine Intelligence Research

Vol. 2, No. 1, Spring & Summer 2024, pp. 131 - 141

A Novel Deep Learning-Based Approach for Graph Dimensionality Reduction Using Fuzzy Logic and Random Walks

Mohammad Mehdi Keikha , Saman Barahouei*

Faculty of Mathematics, Statistics, and Computer Science, University of Sistan and Baluchestan, Zahedan, Iran

Article Information

Original Research Paper

Received:
2024 November 20

Accepted:
2024 December 31

Keywords:
Graph Dimensionality Reduction, Deep Learning, Fuzzy Logic, Random Walks, Graph Embedding, Complex Networks

Corresponding Author*:
keikha@cs.usb.ac.ir

Abstract

Graph structures play a critical role in modeling relationships across domains such as social networks and complex systems. However, scalability and capturing nonlinear relationships remain persistent challenges. Traditional graph embedding methods often struggle with large-scale data or incomplete node features, limiting their effectiveness. This paper introduces FuzzyRandomNet, a novel framework that integrates fuzzy logic with random walks to overcome these challenges. By optimizing node features and incorporating nonlinear layers, this method delivers scalable and precise embeddings. Evaluations on benchmark datasets demonstrate that FuzzyRandomNet achieves superior performance in tasks like node classification and link prediction, highlighting its effectiveness for large-scale and complex networks.

: 10.22034/abmir.2025.22416.1075

E-ISSN: 2821-2037 /© 2023. Published by Yazd University This is an open access article under the CC BY 4.0 License (https://creativecommons.org/licenses/by/4.0/).