Semnan University

Journal of Modeling in Engineering

Journal homepage: https://modelling.semnan.ac.ir/

ISSN: 2783-2538



Research Article

Out-of-Distribution Generalization in Graph Neural Networks via Contrastive Learning

Seyedeh Hamideh Erfani [a], Mohammad Javad Fadaeieslam [a,*], Reza Mortazavi [b], Mohammad Rahmanimanesh [a]

a Faculty of Electrical and Computer Engineering, Semnan University, Semnan, Iran
b School of Engineering, Damghan University, Damghan, Iran

UNCORRECTED PROOF

PAPER INFO

Paper history:

Received:
Revised:
Accepted:

Keywords:
Graph Neural Networks;
Out-of-Distribution Generalization;
Contrastive Learning;
Node Classification;
Siamese Architecture.

ABSTRACT

Graph Neural Networks (GNNs) excel at learning from graph-structured data but suffer significant performance degradation under distribution shifts between training and test environments. This paper proposes a Siamese-based contrastive learning framework for improving out-of-distribution (OOD) generalization in node classification tasks. Our approach generates positive samples through feature matrix perturbation without requiring negative samples, thereby reducing computational complexity. The model employs dual GCN encoders and MLP classifiers with shared weights, optimized using a three-component loss function that maximizes representation similarity, prediction consistency, and classification accuracy. Experimental evaluation on GOOD benchmark datasets across both covariate and concept shift scenarios demonstrates that our method outperforms baseline approaches. This work demonstrates that contrastive learning with Siamese architecture offers a computationally efficient and effective solution for enhancing GNN robustness under distribution shifts, with promising implications for real-world applications requiring reliable model performance in dynamic environments. The proposed method on average and in the GAP metric, has reduced the performance gap between IID and OOD scenarios by 19.75%, while also achieving an average OOD accuracy of 55.04%.

© 2024 Published by Semnan University Press.
This is an open access article under the CC-BY 4.0 license.(https://creativecommons.org/licenses/by/4.0/)

*** Corresponding author.**
E-mail address: fadaei@semnan.ac.ir

How to cite this article:

مقاله پژوهشی

تعمیم‌پذیری خارج‌ازتوزیع شبکه‌های عصبی گرافی با رویکرد یادگیری تقابلی

سیده حمیده عرفانی[۱]، محمدجواد فدائی‌اسلام[۱،*]، رضا مرتضوی[۲] و محمد رحمانی‌منش[۱]

اطلاعات مقاله

دریافت مقاله:
بازنگری مقاله:
پذیرش مقاله:

واژگان کلیدی:
شبکه‌های عصبی گراف،
تعمیم‌پذیری خارج‌ازتوزیع،
یادگیری تقابلی،
طبقه‌بندی گره،
تغییر توزیع.

چکیده

در این مقاله، چالش تعمیم‌پذیری خارج‌ازتوزیع (OOD) در شبکه‌های عصبی گرافی (GNN) برای مسئله طبقه‌بندی گره‌ها مورد بررسی قرار گرفته است. با توجه به کاربرد گستردهGNN ها در حوزه‌های مختلف علمی و صنعتی، ضعف این مدل‌ها در مواجهه با داده‌های خارج از توزیع محیط آموزش به‌عنوان یک محدودیت اساسی شناخته می‌شود. این مقاله با تمرکز بر یادگیری تقابلی، یک مدل مبتنی بر معماری سیامی پیشنهاد می‌دهد که با تقویت داده و تولید نمونه‌های مثبت از طریق به‌هم‌ریختگی ماتریس ویژگی گره‌ها و نیز با به‌کارگیری تابع اتلاف سه‌جزئی شامل بیشینه‌سازی شباهت بازنمایی، شباهت پیش‌بینی برچسب و دقت پیش‌بینی، بهبود قابل‌توجهی در تعمیم‌پذیری OOD حاصل می‌کند. مدل پیشنهادی بدون نیاز به تولید نمونه‌های منفی و با کاهش پیچیدگی محاسباتی، توانسته است عملکرد بهتری نسبت به روش‌های پایه از خود نشان دهد و به‌صورت میانگین در معیار GAP، شکاف عملکرد بین سناریوهای IID و OOD را تا ۱۹/۷۵٪ کاهش داده و ضمنا به میانگین دقت OOD برابر با ۵۵٪/۰۴ دست یافته است.

DOI:

© 2024 Published by Semnan University Press.
This is an open access article under the CC-BY 4.0 license.(https://creativecommons.org/licenses/by/4.0/)

۱-مقدمه[۱]

در سال‌های اخیر، داده‌های دارای ساختار گراف جایگاه ویژه‌ای در حوزه‌های مختلف علمی و صنعتی یافته‌اند و به ابزاری کلیدی برای مدل‌سازی روابط و وابستگی‌های پیچیده میان اجزای مختلف سیستم‌ها تبدیل شده‌اند[۱]. از برهم‌کنش ذرات در سیستم‌های فیزیکی و ساختار مولکولی پروتئین‌ها گرفته تا شبکه‌های ترافیکی و ارتباطی در مقیاس کلان، گراف‌ها نقش اساسی در نمایش و تحلیل داده‌ها ایفا می‌کنند. در این میان، شبکه‌های عصبی گرافی (GNNها) به عنوان یکی از مهم‌ترین ابزارهای یادگیری ماشین مبتنی بر گراف، پیشرفت‌های چشمگیری در وظایفی همچون طبقه‌بندی گره [۲] و پیش‌بینی یال [۳] داشته‌اند و کاربردهای متنوعی از جمله تحلیل شبکه‌های اجتماعی [۴]، سیستم‌های توصیه‌گر [۵]، بازنمایی دانش [۶] و پیش‌بینی ترافیک [۷]، و غیره را پوشش داده‌اند. با وجود موفقیت‌های قابل توجه GNNها، این مدل‌ها در مواجهه با تغییرات توزیع داده‌ها بین محیط‌های آموزش و آزمون با چالش جدی کاهش عملکرد روبرو هستند؛ مسئله‌ای که به ویژه در کاربردهای دنیای واقعی اهمیت می‌یابد. مطالعات نشان داده‌اند که روش‌های مبتنی بر GNN هنگام تقسیم‌بندی داده‌ها بر اساس تنظیمات خارج‌ازتوزیع (OOD)، افت کارایی قابل توجهی را تجربه می‌کنند [۸]. این آسیب‌پذیری عمدتاً ناشی از فرضیه

* پست الکترونیک نویسنده مسئول: fadaei@semnan.ac.ir
۱. دانشکده مهندسی برق و کامپیوتر، دانشگاه سمنان، سمنان، ایران
۲. دانشکده فنی و مهندسی، دانشگاه دامغان، سمنان، ایران

استناد به این مقاله: نحوه استناد فارسی در اینجا درج گردد.

UNCORRECTED PROOF

استاندارد توزیع یکسان و مستقل (IID) بین داده‌های آموزشی و آزمایشی است که در بسیاری از محیط‌های واقعی به دلیل ماهیت غیرقابل کنترل تولید داده‌ها نقض می‌شود و درنتیجه منجر به کاهش قابل‌توجه عملکرد هنگام اعمال مدل‌های یادگیری ماشین بر روی داده‌های خارج‌ازتوزیع می‌گردد [۹]. این چالش به‌ویژه در شبکه‌های عصبی گراف (GNN) حاد می‌شود، جایی که تعاملات پیچیده بین ساختارهای توپولوژیکی و ویژگی‌های گره، الگوهای وابستگی پیچیده‌ای را تولید می‌کند که ممکن است در توزیع‌های مختلف داده تعمیم نیابند. یکی از عوامل اصلی کاهش دقت مدل‌ها در شرایط تغییر توزیع، وجود همبستگی‌های کاذب میان ویژگی‌ها و برچسب‌ها است [۱۰] که به ویژه در روش‌های یادگیری نظارت‌شده و در شرایط کمبود داده‌های برچسب‌دار تشدید می‌شود. این چالش، توجه پژوهشگران را به سمت روش‌های یادگیری بدون نظارت و خودنظارتی جلب کرده است [۱۱]. اگرچه مسئله تعمیم‌پذیری خارج‌ازتوزیع در داده‌های اقلیدسی تا حدی مورد بررسی قرار گرفته، اما آثار محدودی به بررسی این موضوع در داده‌های گرافی پرداخته‌اند؛ در حالی که تغییرات توزیع در گراف‌های واقعی به طور گسترده رخ می‌دهد.با توجه به اهمیت و کاربرد گسترده GNNها در حوزه‌های مختلف هوش مصنوعی و کاهش عملکرد آن‌ها در مواجهه با داده‌های خارج‌ازتوزیع، توسعه مدل‌هایی با قابلیت تعمیم‌پذیری بالا و عملکرد پایدار تحت تغییرات توزیع، به ویژه در کاربردهای حساس مانند تشخیص پزشکی، تحلیل مالی و پیش‌بینی ساختار مولکولی، از اهمیت ویژه‌ای برخوردار است. هدف این مقاله، ارائه یک مدل شبکه عصبی گرافی مبتنی بر یادگیری تقابلی برای طبقه‌بندی گره‌ها است که بتواند در مواجهه با تغییرات توزیع ناشناخته، عملکرد رضایت‌بخشی ارائه دهد. در این راستا نوآوری‌های اصلی به‌صورت زیر خلاصه شده‌اند:

- ارائه روش یادگیری تقابلی بر پایه شبکه‌های سیامی[۲] به‌منظور به حداقل رساندن نقش برچسب‌ها در بازنمایی گره‌ها.
- به‌کارگیری تابع اتلاف سه‌جزئی شامل بیشینه‌سازی شباهت بازنمایی، شباهت پیش‌بینی برچسب و دقت پیش‌بینی برای دستیابی به بازنمایی قوی‌تر از گره‌ها.
- استفاده از تکنیک تقویت داده[۳] و تولید نمونه‌های مثبت از طریق به‌هم‌ریختگی ماتریس ویژگی گره‌ها به جهت دستیابی به مدل قوی‌تر و جلوگیری از وقوع بیش‌برازش[۴] در فرآیند یادگیری.

۲- کارهای مرتبط

در سال‌های اخیر شبکه‌های عصبی گرافی به دلیل موفقیت چشمگیر در بازنمایی داده‌های گرافی، توجه زیادی را به خود جلب کرده‌اند. این شبکه‌ها معمولا از یک الگوی ارسال پیام ساده استفاده می‌کنند که در آن ویژگی‌های گره‌ها و ساختار گراف به‌منظور بروزرسانی بازنمایی گره‌ها با یکدیگر ترکیب می‌شود [۱۲]. در ادامه به شرح مختصری از موضوعات مرتبط با روش پیشنهادی پرداخته شده است.

۲-۱- تقویت داده‌های گرافی

یک راه حل امیدوارکننده برای بهبود تعمیم‌پذیری خارج‌ازتوزیع و نیز مقابله با بیش‌برازش، تقویت داده است [۱۳]. تقویت داده‌ها، تبدیل‌هایی با حفظ برچسب را روی ورودی‌ها اعمال می‌کند، مانند تبدیل‌ها و بازتاب‌ها برای تصاویر. در نتیجه، تقویت داده به طور مؤثر مجموعه آموزشی را گسترش می‌دهد در حالی که سربار محاسباتی ناچیزی را متحمل می‌شود. با این حال، چگونگی تعمیم مؤثر مفهوم تقویت داده به GNNها همچنان یک مسئله باز است [۱۴]. اعمال تبدیلات روی تصاویر به شدت به ساختار تصاویر متکی هستند [۱۵] و طراحی تبدیلات کم‌هزینه که ساختارهای معنایی را برای کارهای غیر بصری مانند پردازش زبان طبیعی [۱۶] و یادگیری گراف حفظ کنند، چالش برانگیز است.

داده‌های گرافی در یادگیری ماشین متشکل از اطلاعات ساختار گراف (ماتریس مجاورت) و ویژگی‌های گره/یال گراف (ماتریس ویژگی) هستند. تقویت داده به ازای یک گراف، می‌تواند در مواردی با اضافه کردن/حذف

[2] *Siamese Networks*

[3] *Data Augmentation*

[4] *Overfitting*

UNCORRECTED PROOF

یال‌ها/گره‌ها بر ساختار گراف تمرکز داشته باشد [۱۷، ۱۸]. در عین حال، تقویت داده تخاصمی، که اختلالات کوچکی را در فضای ویژگی‌های گراف اعمال می‌کند نیز می‌تواند به عنوان روش دیگر تقویت داده‌های گرافی محسوب شود [۱۹، ۲۰]. اخیراً توجه فزاینده‌ای به استفاده از اختلالات تخاصمی برای تقویت مجموعه داده‌ها و در نهایت کاهش بیش‌برازش معطوف شده است. به عنوان مثال، در [۲۱، ۲۲] نشان داده شده است که تقویت داده تخاصمی می‌تواند به تعمیم‌پذیری مدل‌های یادگیری ماشین در مواجهه با نمونه‌های خارج‌ازتوزیع کمک کند.

۲-۲- یادگیری تقابلی

اصول و پیشرفت‌های یادگیری تقابلی به عنوان یک پارادایم قدرتمند در یادگیری نمایش‌های خودنظارتی مطرح شده است. اصل اساسی این روش، حداکثر کردن شباهت میان جفت نمونه‌های مثبت (نمونه‌های مشابه) و کمینه کردن آن برای جفت نمونه‌های منفی (نمونه‌های غیرمشابه) است. این رویکرد در حوزه‌های بینایی ماشین (مانند SimCLR [۲۳] و MoCo [۲۴]) موفقیت‌های چشمگیری داشته و اخیراً به داده‌های گرافی نیز تعمیم یافته است [۲۵، ۲۶]. در حوزه گراف، یادگیری تقابلی معمولاً با ایجاد نماهای مختلف گراف از طریق نمونه‌گیری زیرگراف‌ها و نیز تکنیک‌های تقویت گراف جفت نمونه‌های مثبت و منفی ایجاد می‌گردند. جفت‌های مثبت معمولاً از طریق تقویت داده ایجاد می‌شوند [۱۷]. به‌عنوان مثال، یک گراف G را می‌توان با اعمال تبدیلات تصادفی به دو نمایش همبسته G_1 و G_2 افزایش داد (تقویت کرد). تبدیلات تصادفی مانند حذف گره، اختلال یال و یا پوشاندن ویژگی می‌توانند در تقویت داده‌های گرافی مورد استفاده قرار گیرند. نمایش‌های G_1 و G_2 (یا گره‌های مربوطه در داخل آن‌ها) یک جفت مثبت را تشکیل می‌دهند. نمایش‌های گراف‌های دیگر که به‌طور مستقل نمونه‌برداری شده‌اند، به‌عنوان جفت‌های منفی عمل می‌کنند.

۲-۳- شبکه‌های عصبی سیامی

یک الگوی معماری رایج برای پیاده‌سازی یادگیری تقابلی، شبکه سیامی [۲۷] است. شبکه‌های سیامی در کاربردهای دنیای واقعی توجه فزاینده‌ای را به خود جلب کرده‌اند. آن‌ها برای درک و اندازه‌گیری شباهت بین داده‌های ورودی آموزش دیده‌اند. این معماری شامل دو یا چند زیرشبکه، با معماری یکسان و وزن‌های مشترک است که نمونه‌های ورودی مختلف را پردازش می‌کنند. سپس خروجی‌های این زیرشبکه‌ها برای محاسبه میزان شباهت به‌منظور ارزیابی خطای تقابلی مقایسه می‌شوند. این شبکه‌ها با یادگیری یک تابع نگاشت که اطلاعات ورودی را به شکل جدیدی تبدیل می‌کند کار می‌کنند، به‌طوری که فاصله اندازه‌گیری شده بین ویژگی‌های نهفته یادگیری شده در فضای هدف، نشان‌دهنده شباهت معنایی در فضای ورودی است [۲۸]. این طراحی، آن‌ها را به ویژه برای کارهایی که شامل مقایسه موجودیت‌ها [۲۹]، از جمله یادگیری بازنمایی داده‌ها [۳۰-۳۲] هستند، مناسب می‌کند.

۳-روش پیشنهادی

مسئله تعمیم‌پذیری OOD در GNNها دو چالش اصلی را برای وظیفه طبقه‌بندی گره در بر می‌گیرد: یادگیری بازنمایی گره و طبقه‌بندی گره. در یادگیری بازنمایی گره، کیفیت و جامعیت بازنمایی‌های یادگرفته‌شده گره به‌طور مستقیم بر دقت طبقه‌بندی و ظرفیت مدل برای تعمیم به داده‌های آزمایشی OOD تأثیر می‌گذارد. از این جهت در این مقاله تمرکز روش پیشنهادی بر روی چالش یادگیری بازنمایی گره‌های گراف است. از طرفی، همبستگی‌های کاذب بین مؤلفه‌های نامرتبط بازنمایی‌های یادگرفته‌شده گره و برچسب‌های گره به‌عنوان یکی از عوامل اصلی عملکرد مدل هنگام وقوع جابجایی‌های توزیع بین داده‌های آموزشی و آزمون شناخته شده‌اند. در روش پیشنهادی با بکارگیری روش یادگیری تقابلی و معماری شبکه‌های سیامی سعی می‌شود تاثیر سوگیری‌های موجود در مجموعه آموزش را بر روی بازنمایی گره‌های گراف کاهش داده و بازنمایی معناداری از گره‌های گراف را بدست آوریم.

۳-۱-فرمول‌بندی مسئله و نمادگذاری

در مسئله طبقه‌بندی گره‌های گراف با استفاده از شبکه‌های عصبی گراف، یک گراف G با دو مولفه ماتریس

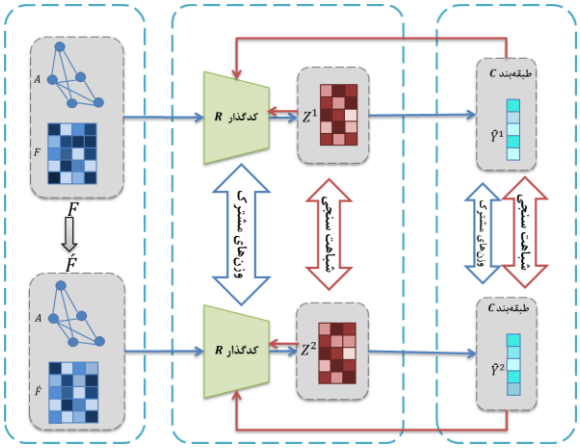
UNCORRECTED PROOF

استناد به این مقاله: نحوه استناد فارسی در اینجا درج گردد.

مجاورت A و ماتریس ویژگی گره F نمایش داده می‌شود. ورودی G= (A, F) توسط یک رمزگذار R پردازش می‌شود که به فضای بازنمایی D-بعدی Z نگاشت می‌شود، و سپس در ادامه یک طبقه‌بند C برچسب گره‌های گراف را پیش‌بینی می‌کند. با توجه به اینکه در مسئله تعمیم‌پذیری خارج‌ازتوزیع فرض بر این است که توزیع مجموعه آموزشی و توزیع مجموعه آزمون متفاوت است، هدف نهایی یادگیری پیش‌بینی‌کننده ترکیبی R○C است که شکاف عملکرد بین تنظیمات ارزیابی IID و OOD را به حداقل می‌رساند.

۳-۲-معماری سیامی مدل پیشنهادی

همان‌طور که در شکل ۱ طرح‌واره معماری روش پیشنهادی آورده شده است، مدل پیشنهادی از معماری سیامی شامل دو زیرشبکه یکسان استفاده می‌کند که در آن دو کدگذار GCN و دو طبقه‌بند MLP با وزن‌های مشترک قرار دارد. وظیفه کدگذار یافتن بازنمایی گره‌های گراف در فضای بازنمایی Z است. طبقه‌بند نیز جهت پیش‌بینی برچسب گره‌ها آموزش می‌بینند. از آنجایی که در معماری سیامی وزن‌ها بین دو زیر شبکه به اشتراک گذاشته می‌شود، سربار محاسباتی زیادی به مدل تحمیل نمی‌گردد.



شکل ۱- طرح‌واره معماری روش پیشنهادی

۳-۳-تقویت ماتریس ویژگی و تولید نمونه‌های مثبت

در ورودی معماری روش پیشنهادی، با اعمال به‌هم‌ریختگی بر روی ماتریس ویژگی F (سطرهای ماتریس) با نرخ ۴٪ ، اقدام به تولید نمونه‌های مثبت می‌شود. از آنجایی که ماتریس ویژگی در ورودی حاوی مقادیر صفر و یک است، در اعمال به‌هم‌ریختگی در هر سطر از ماتریس، ۴٪ از مولفه‌ها را به‌طور تصادفی صفر می‌شوند. لازم به توضیح است که در این معماری نیازی به تولید نمونه‌های منفی نبوده و به تولید نمونه‌های مثبت بسنده شده است. این امر پیچیدگی روش پیشنهادی را کاهش می‌دهد.

۳-۴-آموزش مدل پیشنهادی

در این بخش، یک تابع اتلاف سه مؤلفه‌ای پیشنهاد می‌شود که برای بهره‌برداری از شباهت بازنمایی گره، شباهت پیش‌بینی برچسب و دقت پیش‌بینی طراحی شده است. مؤلفه اول، شباهت کسینوسی بین بازنمایی‌های Z^1 و Z^2، که توسط کدگذارهای دوگانه $\mathcal{R}$ تولید شده‌اند، را به حداکثر می‌رساند. این امر کدگذارها را تشویق می‌کند تا ویژگی‌های مشابه و اساسی در بازنمایی‌ها را یاد بگیرند. مؤلفه دوم، شباهت کسینوسی برچسب‌های پیش‌بینی شده $\hat{Y}^1$ و $\hat{Y}^2$ را به حداکثر می‌رساند و پیش‌بینی‌های سازگار از دو شاخه را ارتقا می‌دهد. در نهایت، مؤلفه سوم مستقیماً دقت پیش‌بینی مدل را اندازه‌گیری می‌کند. ما شباهت کسینوسی را برای ارزیابی هم‌ترازی بردارها انتخاب می‌کنیم، زیرا به حداکثر رساندن این معیار، همپوشانی ویژگی‌ها را افزایش داده و منجر به بازنمایی‌های سازگارتر و قوی‌تر در چارچوب مدل پیشنهادی می‌شود. شباهت کسینوسی بین بردارهای x_1 و x_2 به‌صورت رابطه (۱) تعریف می‌شود.

$$D(x_1, x_2) = \frac{x_1 . x_2}{max(\|x_1\|_2 . \|x_2\|_2, \varepsilon)} \qquad (۱)$$

که در آن $\|.\|_2$ نشان دهنده نرم-l_2 و ε یک ثابت مثبت کوچک است. از آنجایی که هدف روش پیشنهادی دوم دستیابی به بازنمایی‌های مشابه از یک گره در هر دو شاخه شبکه است، هدف ما به حداکثر رساندن شباهت بین آن‌ها است. برای این منظور، تابع اتلاف شباهت بازنمایی گره، $\mathcal{L}_{rep}$، را همان‌طور که در معادله (۲) نشان داده شده است، تعریف می‌کنیم.

UNCORRECTED PROOF

$$\mathcal{L}_{rep} = 1 - \frac{1}{N}\sum_{i=1}^{N_{tr}} D(Z_{i*}^1, Z_{i*}^2) \quad (۲)$$

که در آن Z_{i*}^1 و Z_{i*}^2 مربوط به بازنمایی‌های گره i-ام هستند و N_{tr} تعداد گره‌ها در مجموعه آموزش را نشان می‌دهد. ما مدل پیشنهادی را تشویق می‌کنیم تا بازنمایی‌های مشابه Z^1 و Z^2 را برای هر گره از طریق تابع اتلاف $\mathcal{L}_{rep}$ تولید کند. متعاقباً، بازنمایی Z^1 و Z^2 به‌طور جداگانه به یک طبقه‌بند C_φ وارد می‌شود تا به ترتیب برچسب‌های $\hat{Y}^1$ و $\hat{Y}^2$ را پیش‌بینی کند. با توجه به اینکه هدف نهایی به دست آوردن برچسب‌های پیش‌بینی شده مشابه و دقیق است، دو عبارت جداگانه را در تابع اتلاف تعریف می‌کنیم: اتلاف شباهت پیش‌بینی برچسب $\mathcal{L}_{pred}$ و اتلاف دقت پیش‌بینی $\mathcal{L}_{entropy}$، همان‌طور که به ترتیب در معادله (۳) و معادله (۴) نشان داده شده است.

$$\mathcal{L}_{pred} = 1 - D(\hat{Y}^1, \hat{Y}^2) \quad (۳)$$

$$\mathcal{L}_{entropy} = -\sum_{i=1}^{N_{tr}} y_i \, log \frac{\hat{y}_i^1 + \hat{y}_i^2}{2} \quad (۴)$$

که در آن $\mathcal{L}_{pred}$ تابع اتلاف شباهت برچسب گره است که برای تشویق مدل برای به حداکثر رساندن شباهت کسینوسی بین دو برچسب پیش‌بینی‌شده $\hat{Y}^1$ و $\hat{Y}^2$ استفاده می‌شود. $\mathcal{L}_{entropy}$ تابع اتلاف آنتروپی متقاطع است که به‌ازای برچسب مورد نظر y_i و میانگین برچسب‌های پیش‌بینی‌شده $\hat{y}_i^1$ و $\hat{y}_i^2$ محاسبه شده است. تابع اتلاف نهایی مدل پیشنهادی در رابطه (۵) ارائه شده است.

$$\mathcal{L} = \mathcal{L}_{entropy} + \mathcal{L}_{pred} + \mathcal{L}_{rep} \quad (۵)$$

۴-نتایج تجربی

در این بخش، ما روش خود را به صورت تجربی ارزیابی کرده و آن را با روش‌های پایه پیاده‌سازی شده در چارچوب GOOD [۳۳] مقایسه می‌کنیم. در کلیه آزمایش‌ها، مدل‌ها را برای ۱۰۰ دوره با استفاده از بهینه‌ساز Adam با نرخ یادگیری 1e-3 آموزش می‌دهیم. همه آزمایش‌ها بر روی یک پردازنده گرافیکی NVIDIA GeForce GTX 1080 با ۱۶ گیگابایت حافظه انجام می‌شود. برای همه آزمایش‌ها، ما بهترین نقاط بررسی را برای ارزیابی روی داده‌های آزمون IID و OOD انتخاب می‌کنیم. جداول ۱، ۲ و ۳ نتایج را بر اساس دقت OOD و معیار GAP ارائه می‌دهند.

۴-۱-مجموعه‌داده‌ها

ارزیابی نتایج تجربی روش پیشنهادی بر روی چهار مجموعه داده طبقه‌بندی گره تطبیق‌یافته شامل، GOOD-Cora، GOOD-Arxiv و GOOD-WebKB و به‌ازای ۱۰ دامنه و تغییر توزیع متفاوت انجام شده است که از چارچوب GOOD [۳۳] استخراج گردیده‌اند تا از تعمیم‌پذیری و استحکام یافته‌ها در سناریوهای مختلف OOD برای وظیفه طبقه‌بندی گره اطمینان حاصل شود. در این مجموعه‌داده‌ها، هم تغییر توزیع متغیرهای مستقل و هم تغییر توزیع مفهومی با دقت خوبی در نظر گرفته شده است و به‌طور خاص برای نمایش تغییرات توزیع چالش‌برانگیز مربوط به داده‌های دارای ساختار گراف گردآوری شده‌اند.

۴-۲-معیارهای ارزیابی

در ارزیابی روش پیشنهادی، ما از معیار دقت OOD (Accuracy) و شکاف عملکرد بین تنظیمات ارزیابی IID و OOD را تحت عنوان معیار GAP [۳۴]، برای اندازه‌گیری تفاوت‌های عملکرد مطلق و نسبی استفاده می‌کنیم. نحوه محاسبه معیار GAP در رابطه (۶) آورده شده است که در آن IID_{te} و OOD_{te} به ترتیب عملکرد مدل روی داده‌های آزمون IID و OOD هستند. مقادیر پایین‌تر GAP نشان‌دهنده عملکرد تعمیم‌پذیری OOD برتر است.

$$GAP = \frac{IID_{te} - OOD_{te}}{IID_{te}} \times 100 \quad (۶)$$

۴-۳-روش‌های پایه

ما روش پیشنهادی خود را با چندین رویکرد پایه مقایسه می‌کنیم. به عنوان یک روش پایه اصلی، از کمینه‌سازی ریسک تجربی (ERM) [۳۵] استفاده می‌کنیم. همچنین

UNCORRECTED PROOF

استناد به این مقاله: نحوه استناد فارسی در اینجا درج گردد.

کمینه‌سازی ریسک ثابت (IRM) [۳۶] را در ارزیابی‌ها در نظر می‌گیریم. روش VREx [۳۷] دیگر روش پایه‌ای است که با به حداقل رساندن واریانس ریسک در محیط‌های آموزشی، تعمیم‌پذیری را بهبود می‌بخشد. ما همچنین دو روش DANN [۳۸] و Deep Coral [۳۹] را که بر به حداقل رساندن اختلافات ویژگی تمرکز دارند، مورد ارزیابی قرار می‌دهیم.

۴-۴-تحلیل نتایج

عملکرد روش پیشنهادی در مقایسه با چندین روش پایه و تحت هر دو سناریوی تغییر توزیع متغیرهای مستقل و تغییر توزیع مفهومی ارزیابی گردید. گزارشی از عملکرد روش پیشنهادی در جدول ۱ تا ۳ به‌ازای مجموعه‌داده‌های GOOD-Arxiv، GOOD-Cora و GOOD-WebKB، آورده شده است. سنجش عملکرد تعمیم‌پذیری خارج‌ازتوزیع الگوریتم‌ها بر مبنای معیار دقت و نیز معیار عملکرد نسبی GAP ارائه شده است. در کلیه جداول نتایج، نماد ↑ نشان می‌دهد که مقادیر بالاتر بیانگر عملکرد بهتر هستند و ↓ نشان می‌دهد که مقادیر پایین‌تر بیانگر عملکرد بهتر هستند. ضمنا اعداد پررنگ نشان دهنده بهترین نتایج و مقادیر زیرخط‌دار بیان کننده دومین بهترین نتایج در هر ستون هستند. با توجه به جداول ۱ تا ۳، روش پیشنهادی (Proposed Method) مبتنی بر تقویت داده در معماری سیامی شبکه عصبی گرافی، به طور مداوم نسبت به روش‌های پایه عملکرد بهتری ارائه داده و توانسته است به بهترین نتیجه و یا دومین بهترین نتیجه دست یابد. به عنوان مثال، روش پیشنهادی در مجموعه‌داده GOOD-Arxiv با دقت OOD برابر ۶۹.۸۷٪ و GAP برابر ۱.۸۸٪، کمترین شکاف عملکرد را در بین تمامی روش‌ها به دست آورده است. در مجموعه‌داده GOOD-Cora، روش پیشنهادی با دقت OODبرابر ۶۵.۰۶٪، بهترین عملکرد را کسب کرده است. عدم موفقیت روش پیشنهادی در برابر روش‌های پایه به ازای تغییر توزیع مفهومی در دامنه Degree و به ازای مجموعه‌داده GOOD-Cora و در دامنه University از مجموعه‌داده GOOD-WebKB می‌تواند به علت ساختار خاص تغییر توزیع در آن دامنه خاص باشد.

جدول ۱. مقایسه عملکرد روش پیشنهادی در مقایسه با رویکردهای پایه در مجموعه داده GOOD-Arxiv.

Algorithm	GOOD-Arxiv							
	covariate shift				concept shift			
	Time		degree		time		Degree	
	OOD↑	GAP↓	OOD↑	GAP↓	OOD↑	GAP↓	OOD↑	GAP↓
ERM	<u>71.08</u>	2.09	57.04	25.24	<u>65.34</u>	<u>9.31</u>	62.12	14.29
IRM	71.04	2.29	57.00	25.26	65.12	9.69	62.15	14.75
VREx	**71.12**	2.11	<u>57.39</u>	**24.24**	65.31	9.54	<u>62.54</u>	**12.53**
DANN	69.89	<u>1.91</u>	57.09	24.52	65.10	10.36	61.99	14.62
Deep Coral	69.66	2.41	57.16	24.58	64.23	11.68	62.17	14.63
Proposed Method	69.87	**1.88**	**57.76**	<u>24.31</u>	**66.10**	**9.27**	**62.83**	<u>14.21</u>

جدول ۲. مقایسه عملکرد روش پیشنهادی در مقایسه با رویکردهای پایه در مجموعه داده GOOD-Cora.

Algorithm	GOOD-Cora							
	covariate shift				concept shift			
	word		degree		word		Degree	
	OOD↑	GAP↓	OOD↑	GAP↓	OOD↑	GAP↓	OOD↑	GAP↓
ERM	<u>64.96</u>	8.41	56.40	22.76	64.13	<u>2.42</u>	60.63	11.50
IRM	64.04	<u>8.23</u>	**56.57**	23.63	64.27	3.35	**61.37**	**10.30**
VREx	63.85	8.96	55.30	23.47	64.48	3.09	<u>61.26</u>	<u>11.46</u>
DANN	64.93	8.41	56.37	22.69	64.07	**2.29**	60.13	11.73
Deep Coral	64.75	8.50	<u>56.51</u>	**22.45**	<u>64.31</u>	2.90	59.83	12.24
Proposed Method	**65.06**	**7.70**	56.46	**22.45**	**64.32**	3.23	59.96	13.26

UNCORRECTED PROOF

جدول ۳. مقایسه عملکرد روش پیشنهادی در مقایسه با رویکردهای پایه در مجموعه داده GOOD-WebKB.

Algorithm	GOOD-WebKB			
	covariate shift		concept shift	
	University			
	OOD↑	GAP↓	OOD↑	GAP↓
ERM	14.92	67.06	29.36	**51.00**
IRM	17.62	61.85	28.62	52.48
VREx	**27.78**	**34.82**	26.61	60.08
DANN	19.84	54.69	29.36	52.89
Deep Coral	16.66	60.71	**29.91**	52.19
Proposed Method	21.43	45.52	26.61	55.65

تحلیل جامع نتایج تجربی نشان می‌دهد که روش پیشنهادی توانسته است در اکثر سناریوهای ارزیابی، عملکرد قابل توجهی در بهبود تعمیم‌پذیری OOD داشته باشد. با این حال، همچنان فضایی برای بهبود در مجموعه داده‌های خاص و سناریوهای چالش‌برانگیز وجود دارد که می‌تواند موضوع تحقیقات آتی باشد.

۵-نتیجه گیری

این تحقیق نشان داده است که ترکیب معماری سیامی با یادگیری تقابلی و تقویت داده مبتنی بر به‌هم‌ریختگی ویژگی گره‌ها می‌تواند به طور مؤثری مشکل کاهش عملکرد شبکه‌های عصبی گرافی در مواجهه با داده‌های خارج‌ازتوزیع را برطرف کند. مدل پیشنهادی با تکیه بر یادگیری بازنمایی‌های مقاوم به سوگیری‌های مجموعه آموزش و استفاده از تابع اتلاف سه‌جزئی مناسب، توانسته است تعمیم‌پذیری بالاتری در محیط‌های متنوع و ناشناخته ارائه دهد. سادگی معماری پیشنهادی، کارایی محاسباتی مناسب و عملکرد برتر در اکثر سناریوهای ارزیابی، آن را به یک راه‌حل عملی و قابل اجرا برای کاربردهای واقعی تبدیل می‌کند. با توجه به رشد روزافزون کاربردهای GNNها در مسائل حساس و داده‌های واقعی که غالباً با تغییرات توزیع مواجه هستند، استفاده از رویکردهای مبتنی بر تقویت داده و یادگیری تقابلی می‌تواند مسیر مناسبی برای توسعه مدل‌های پایدارتر و قابل اطمینان‌تر باشد. نتایج این پژوهش نویدبخش کاربرد وسیع‌تر و مطمئن‌تر GNNها در حوزه‌هایی چون پزشکی، مالی، زیست‌مولکولی و شبکه‌های اجتماعی است و افق‌های جدیدی را برای پژوهش‌های آتی در زمینه تعمیم‌پذیری مدل‌های گرافی می‌گشاید. امید است که این کار پایه‌ای برای تحقیقات بیشتر در جهت توسعه GNN های قوی‌تر و قابل اعتمادتر باشد که بتوانند به طور مؤثر در محیط‌های دائماً در حال تغییر دنیای واقعی عمل کنند.

از آنجایی که نتایج ما امیدوارکننده است، چندین مسیر برای تحقیقات آینده از این کار پدیدار می‌شود. اول، گسترش این روش‌ها برای مدیریت انواع تغییر توزیع و بررسی کاربرد آن‌ها در سایر سطوح وظایف فراتر از طبقه‌بندی گره، مانند طبقه بندی سطح گراف. دوم، بررسی مقیاس‌پذیری به‌ازای مجموعه‌داده‌های دارای ساختار گراف بزرگتر و پیچیده‌تر برای پرداختن به کاربردهای نوظهور در زمینه‌هایی مانند شبکه‌های بیولوژیکی و تجزیه و تحلیل رسانه‌های اجتماعی در مقیاس بزرگ ضروری خواهد بود.

مراجع

[1] H. Li, X. Wang, Z. Zhang and W. Zhu. "Out-of-Distribution Generalization on Graphs: A Survey." IEEE Transactions on Pattern Analysis and Machine Intelligence 47, no. 11 (2024): 10490-10512.

[2] Jiaxin Zhang, Yiqi Wang, Xihong Yang, and En Zhu. “A Fully Test-time Training Framework for Semi-supervised Node Classification on Out-of-Distribution Graphs.” ACM Transactions on Knowledge Discovery from Data 18, no. 7 (2024): 1-19.

UNCORRECTED PROOF

[3] Tong Zhao and Gang Liu and Daheng Wang and Wenhao Yu and Meng Jiang. “Counterfactual graph learning for link prediction.” The Tenth International Conference on Learning Representations (ICLR), 2022.
[4] Jiezhong Qiu, Jian Tang, Hao Ma, Yuxiao Dong, Kuansan Wang, and Jie Tang. “Deepinf: Social influence prediction with deep learning.” Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018.
[5] Shiwen Wu, Fei Sun, Wentao Zhang, Xu Xie, and Bin Cui. “Graph neural networks in recommender systems: a survey.” ACM Computing Surveys 55, no. 97 (2022): 1 – 37.
[6] Q. Wang, Z. Mao, B. Wang and L. Guo. “Knowledge graph embedding: A survey of approaches and applications.” IEEE Transactions on Knowledge and Data Engineering 29, no. 12 (2017): 2724-2743.
[7] Bing Yu, Haoteng Yin, and Zhanxing Zhu. “Spatiotemporal graph convolutional networks: a deep learning framework for traffic forecasting.” Proceedings of the 27th International Joint Conference on Artificial Intelligence, 2018.
[8] Shurui Gui, Meng Liu, Xiner Li, Youzhi Luo, and Shuiwang Ji. “Joint learning of label and environment causal independence for graph out-of-distribution generalization.” In Thirty-seventh Conference on Neural Information Processing Systems, 2023.
[9] Yoshua Bengio, Tristan Deleu, Nasim Rahaman, Rosemary Ke, Sébastien Lachapelle, Olexa Bilaniuk, Anirudh Goyal, and Christopher Pal. “A meta-transfer objective for learning to disentangle causal mechanisms.” International Conference on Learning Representations, 2020.
[10] Xingxuan Zhang, Peng Cui, Renzhe Xu, Linjun Zhou, Yue He, and Zheyan Shen. “Deep stable learning for out-of-distribution generalization.” IEEE/CVF Conference on Computer Vision and Pattern Recognition In CVPR, pp. 5372–5382. IEEE, 2021.
[11] Haoyang Li, Xin Wang, Zeyang Zhang, Haibo Chen, Ziwei Zhang, and Wenwu Zhu. “Disentangled graph selfsupervised learning for out-of-distribution generalization.” In Forty-first International Conference on Machine Learning, pp. 28890-28904, 2024.
[12] Haoyang Li, Xin Wang, Ziwei Zhang, and Wenwu Zhu. “Ood-gnn: Outof-distribution generalized graph neural network.” IEEE Transactions on Knowledge and Data Engineering 35, no. 7 (2023):7328-7340.
[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton. “Imagenet classification with deep convolutional neural networks.” In Advances in neural information processing systems, pp. 1097-1105, 2012.
[14] K Kong, G Li, M Ding, Z Wu, C Zhu, B Ghanem, G Taylor, T Goldstein. “Robust optimization as data augmentation for large-scale graphs.” Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 60-69. IEEE, 2022.
[15] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton. “A simple framework for contrastive learning of visual representations.” In International conference on machine learning, pp. 1597-1607, 2020.
[16] J. Wei and K. Zou. “Eda: Easy data augmentation techniques for boosting performance on text classification tasks.” arXiv preprint arXiv:1901.11196, (2019).
[17] Y. You, T. Chen, Y. Sui, T. Chen, Z. Wang, and Y. Shen. “Graph contrastive learning with augmentations.” Advances in Neural Information Processing Systems 33, (2020).
[18] J. Godwin, M. Schaarschmidt, A. L. Gaunt, A. Sanchez-Gonzalez, Y. Rubanova, P. Veli˘ckovi´c, J. Kirkpatrick, and P. Battaglia. “Simple GNN regularisation for 3d molecular property prediction and beyond.” In International Conference on Learning Representations, 2022.
[19] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu. “Towards deep learning models resistant to adversarial attacks.” International Conference on Learning Representation, 2018.
[20] Y. Balaji, T. Goldstein, and J. Hoffman. “Instance adaptive adversarial training: Improved accuracy tradeoffs in neural nets.” arXiv preprint arXiv:1910.08051, (2019).
[21] Yun Zhu, Haizhou Shi, Zhenshuo Zhang, and Siliang Tang. “MARIO: Model Agnostic Recipe for Improving OOD Generalization of Graph Contrastive Learning.” In Proceedings of the ACM Web Conference, 2024.
[22] Lu, Bin and Zhao, Ze and Gan, Xiaoying and Liang, Shiyu and Fu, Luoyi and Wang, Xinbing and Zhou, Chenghu. “Graph Out-of-Distribution Generalization With Controllable Data Augmentation.” IEEE Transactions on Knowledge and Data Engineering 36, no. 11 (2024): 6317–6329.
[23] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. “Momentum contrast for unsupervised visual representation learning.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9729–9738, 2020.
[24] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. “A simple framework for contrastive learning of visual representations.” arXiv preprint arXiv:2002.05709, (2020).
[25] Lu, Bin and Zhao, Ze and Gan, Xiaoying and Liang, Shiyu and Fu, Luoyi and Wang, Xinbing and Zhou, Chenghu. “Graph Out-of-Distribution Generalization With Controllable Data Augmentation.” IEEE Transactions on Knowledge and Data Engineering 36, no. 11 (2024): 6317–6329.

UNCORRECTED PROOF

[26] Zhu, Yun and Shi, Haizhou and Zhang, Zhenshuo and Tang, Siliang. “MARIO: Model Agnostic Recipe for Improving OOD Generalization of Graph Contrastive Learning.” Proceedings of the ACM Web Conference 2024. Singapore.
[27] Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Sckinger, and Roopak Shah. “Signature verication using a "siamese" time delay neural network.” In Proceedings of the 6th International Conference on Neural Information Processing Systems, NIPS'93, pp. 737-744, 1993.
[28] Nikita Mehrotra, Navdha Agarwal, Piyush Gupta, Saket Anand, David Lo, and Rahul Purandare. “Modeling functional similarity in source code with graph-based siamese networks.” IEEE Transactions on Software Engineering 48, no. 10 (2022): 3771-3789.
[29] Xinlei Chen and Kaiming He. “Exploring simple siamese representation learning.” In IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 15745-15753, IEEE. 2021.
[30] Junjie Lu, Shengyang Li, Weilong Guo, Manqi Zhao, Jian Yang, Yunfei Liu, and Zhuang Zhou. “Siamese graph attention networks for robust visual object tracking.” Computer Vision and Image Understanding 229, (2023).
[31] Chenrui Wu, Long Chen, Zaixing He, and Junjie Jiang. “Pseudo-siamese graph matching network for textureless objects 6-d pose estimation.” IEEE Transactions on Industrial Electronics 69, no.3 (2022):2718-2727.
[32] Necva Blc, Burcu Can, and Harun Artuner. “A siamese neural network for learning semantically-informed sentence embeddings.” Expert Systems with Applications 214, (2023).
[33] Shurui Gui , Xiner Li, Limei Wang and Shuiwang Ji. “GOOD: A Graph Out-of-Distribution Benchmark.” NIPS'22: Proceedings of the 36th International Conference on Neural Information Processing Syste:, pp. 2059 – 2073, 2022.
[34] Seyedeh Hamideh Erfani, Mohammad Javad Fadaeieslam, Reza Mortazavi, Mohammad Rahmanimanesh. “SSGNN: Simple Siamese Graph Neural Networks for Out-of-Distribution Generalization”. Knowledge-Based Systems 329, https://doi.org/10.1016/j.knosys.2025.114290, 2025.
[35] David Krueger, Ethan Caballero, Joern-Henrik Jacobsen, Amy Zhang, Jonathan Binas, Dinghuai Zhang, Remi Le Priol, and Aaron Courville. “Out-of-distribution generalization via risk extrapolation (REx).” In Proceedings of Machine Learning Research, (2021): 5815–5826.
[36] Martin Arjovsky, Léon Bottou, Ishaan Gulrajani, David Lopez-Paz. “Invariant Risk Minimization.” arXiv:1907.02893, (2019).
[37] Qi Zhu, Natalia Ponomareva, Jiawei Han, and Bryan Perozzi. “Shift-robust gnns: Overcoming the limitations of localized graph training data.” Advances in Neural Information Processing Systems 34, (2021): 27965-27977.
[38] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, Franois Laviolette, Mario Marchand, and Victor Lempitsky. “Domain-adversarial training of neural networks.” The Journal of Machine Learning Research 17, no. 1 (2016): 2096 - 2030.
[39] Baochen Sun and Kate Saenko. “Deep coral: Correlation alignment for deep domain adaptation.” arXiv:1607.01719, (2016).

UNCORRECTED PROOF