

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

گزارش کتبی سمینار درس شبکه های عصبی

شبکه های عصبی مبتنی بر رویکرد توجه گرافی

استاد درس: دکتر رضا صفا بخش

نگارش: رومینا ذاکریان

شماره دانشجویی: ۹۹۱۳۱۰۴۲

تیر ۱۴۰۰

چکیده

شبکه‌ها در دنیای واقعی بسیار رایج هستند و به طور گسترده مورد مطالعه قرار گرفته اند. متأسفانه ، ساختارهای نامنظم آن‌ها که در مقایسه با ساختارهای منظم شبکه مانند تصویر، ویدئو، صدا و غیره کاملاً متفاوت رفتار می‌کنند، مانع از پردازش مستقیم آنها توسط فناوری های مدرن یادگیری عمیق می شوند، که این فناوری‌ها به پیشرفت های چشمگیری در بسیاری از وظایف داده های ساختاری شبکه، مانند طبقه بندی تصویر، تشخیص شی و غیره رسیده است. اخیراً، شبکه های عصبی گرافی پردازش داده های شبکه را از طریق یادگیری عمیق به انجام می‌رساند. شبکه های عصبی گرافی به موفقیت فوق العاده ای در طبقه بندی گره ها به صورت نیمه نظارتی رسیده اند. در این گزارش، ابتدا دو روش اساسی پردازش داده ها با استفاده از شبکه های عصبی گرافی را تجزیه و تحلیل می‌کنیم تا فلسفه اساسی آنها آشکار کنیم. سپس در این گزارش چهار مدل مختلف از آن روش را بررسی می‌کنیم و با استناد به نتیجه اعمال آن در مسائل و مجموعه داده‌های متفاوت، عملکرد هر یک از این مدل ها را، هم از نظر کارایی نهایی و هم مقایسه با سایر روش های اعمال شده بر روی آن مسائل، بررسی می‌کنیم.

واژه های کلیدی:

شبکه های عصبی گرافی، شبکه های کانولوشنی گرافی، مکانیسم توجه، شبکه های توجه گرافی

فهرست مطالب

فصل اول	۹
مقدمه	۹
۱-۱ شبکه های گرافی	۱۰
۲-۱ سازماندهی گزارش	۱۰
فصل دوم	۱۲
تلاش های پیشین در شبکه های عصبی با رویکرد توجه	۱۲
فصل سوم	۱۴
شبکه های عصبی با رویکرد توجه	۱۴
۱-۳ شبکه های توجه گرافی	۱۵
۱-۱-۳ ۱-۱-۳ نشانه گذاری ها	۱۵
۲-۱-۳ شبکه کانولوشنی گرافی	۱۶
۳-۱-۳ شبکه ی توجه گرافی	۱۷
فصل چهارم	۱۹
شبکه های گرافی با رویکرد توجه مشارکتی	۱۹
۱-۴ شبکه های گرافی توجه مشارکتی	۲۰
۲-۴ نتایج	۲۲
۱-۲-۴ یادگیری نیمه نظارتی انتقالی	۲۲
۲-۲-۴ یادگیری نیمه نظارتی استقرایی	۲۳
فصل پنجم	۲۴
شبکه های گرافی با رویکرد توجه چند نمایه	۲۴

۱-۵ مکانیسم توجه	۲۵
۲-۵ تعبیه شبکه	۲۵
۳-۵ نتایج	۲۷
۱-۳-۵ پیش بینی پیوند	۲۷
۲-۳-۵ طبقه بندی گره ها	۲۸
فصل ششم	۳۰
۱-۶ گراف دانش	۳۱
۲-۶ آزمایش ها و نتایج	۳۴
۱-۲-۶ معیار های ارزیابی	۳۴
۲-۲-۶ نتایج پیش بینی پیوند	۳۴
فصل هفتم	۳۶
شبکه های توجه گرافی مکانی-زمانی: رویکرد یادگیری عمیق برای پیش بینی ترافیک	۳۶
۱-۷ مکانیسم توجه multi-head	۳۷
۲-۷ معماری شبکه	۳۷
۳-۷ آزمایش ها و نتایج	۳۸
مراجع	۴۰

فهرست اشکال

- شکل ۱: نمونه ای از GCN ۱۷
- شکل ۲: نمونه ای از GAT ۱۸
- شکل ۳: مقایسه استراتژی های مختلف توجه در گراف داده شده و گراف ویژگی ساخته شده. ۲۰
- شکل ۴: چارچوب شبکه های توجه گرافی چند نمایه ۲۶
- شکل ۵: یک زیر گراف از گراف دانش در نمایش سه تایی است که از روابط واقعی (خطوط آبی) بین موجودیت و روابط استنباط شده (نقطه چین قرمز) تشکیل شده است. ۳۱
- شکل ۶: یک زیرگراف از موجودیت "RAN" در مجموعه داده FB15k-237 ۳۲
- شکل ۷: یک تصویر از یک مثال ۳۲
- شکل ۸: معماری مدل LSA-GAT ۳۳
- شکل ۹: معماری شبکه های توجه به نمودار مکانی - زمانی (ST-GAT) ۳۸

فهرست جداول

- جدول ۱: مجموعه ی داده ۲۲
- جدول ۲: نتایج یادگیری انتقالی از نظر دقت..... ۲۲
- جدول ۳: نتایج یادگیری استقرایی از نظر **micro-F1 score** ۲۳
- جدول ۴: اطلاعات شش شبکه آزمایش شده..... ۲۷
- جدول ۵: نتایج روش های مختلف در پیش بینی پیوند ۲۸
- جدول ۶: نتایج روش های مختلف در طبقه بندی گره. ۲۹
- جدول ۷: اطلاعات مجموعه داده ها..... ۳۴
- جدول ۸: نتایج به دست آمده بر روی مجموعه آزمون ۳۵
- جدول ۹: مقایسه عملکرد **ST-GAT** و روش های پایه در مجموعه داده **PeMSD7**. ۳۹

فهرست نمادها

نماد	مفهوم
R^n	فضای اقلیدسی n بعدی
\parallel	الحاق
$\text{Diag}()$	ماتریس قطری
\star	کانولوشن
$6(.)$	تابع فعال ساز غیر خطی

فصل اول

مقدمه

۱-۱ شبکه های گرافی

شبکه های عصبی گرافی^۱ با استفاده از شبکه های عصبی عمیق در دامنه فوریه گراف ایجاد می شوند. متأسفانه، پیچیدگی بالای تجزیه ویژه، مانع از استفاده GNN ها در عمل، به ویژه در شبکه های با مقیاس بزرگ می شود. برای حل این مسئله، بسیاری از رویکردهای تقریب، مانند بسط کوتاه شده چند جمله ای چبیشف^۲، استفاده شده است. شبکه ی کانولوشنی گرافی^۳، که از تقریب مرتبه اول کانولوشن گرافی طیفی^۴ ایجاد می شود، فاصله ای را بین کانولوشن گرافی طیفی^۵ و انتشار ویژگی های مکانی ایجاد می کند و با مقیاس پذیری مناسب نتایج رضایت بخشی را به دست می آورد. با توجه به مکانیزم توجه، که با اختصاص وزن بیشتر به ورودی های مهم تأکید دارد، شبکه توجه گرافی^۶ با معماری مبتنی بر توجه پیشنهاد شده است که گر به تعیین وزن برای همسایه ها از همسایه های خود استفاده می کند. در این گزارش ابتدا با مقایسه دقیق قوانین انتشار GCN و GAT، که اساس و هسته سایر GNN های پیشنهادی اخیر هستند، می توان مشاهده کرد که فلسفه های اساسی آنها کاملاً متفاوت است، هرچند که آنها عملیات انتشار مشابهی دارند [۱]. ما شبکه های توجه گرافی را از منظر تقویت ویژگی-که تصویری از موفقیت آن را نشان می دهد- در مقایسه با GCN، تفسیر می کنیم. سپس به مسائل و مشکلات دیگر موجود در شبکه های مختلف گرافی می پردازیم و سعی می کنیم با استفاده از این شبکه های گرافی مبتنی بر رویکرد توجه، چالش های پیش آمده را به شیوه های مختلف حل کنیم و نسخه های مختلف منتشر شده با استفاده از این شبکه را معرفی کنیم.

۱-۲ سازماندهی گزارش

هدف از این گزارش بررسی پنج پژوهش در زمینه شبکه های عصبی گرافی با رویکرد توجه می باشد [۱،۲،۳،۴،۵]. در ادامه این گزارش، در فصل ۲، به بیان تلاش های انجام شده در زمینه شبکه های عصبی گرافی و شبکه هایی که از رویکرد توجه برای حل مسائل استفاده کرده اند، می پردازیم.

¹ Graph Neural Networks (GNNs)

² Chebyshev polynomials

³ Graph Convolutional Network

⁴ Spectra graph convolutions

⁵ spectral graph convolutions

⁶ Graph Attention Network (GAT)

فصل اول: مقدمه

در فصل ۳، به پایه و اساس مفهوم استفاده شده در این شبکه ها می پردازیم و در ادامه در فصل ۴، جدیدترین و مهم ترین فعالیت در این زمینه تحقیقاتی، یعنی شبکه های گرافی با رویکرد توجه مشارکتی را معرفی می کنیم و عملکرد آن را بر روی مجموعه داده های مختلف بررسی می کنیم و به ترتیب در فصل های ۵ و ۶ و ۷ به نسخه های دیگر منتشر شده در حوزه ی شبکه های توجه گرافی می پردازیم و به معرفی آزمایش هایی جهت مقایسه عملکرد این شبکه ها با هم می پردازیم، روش های ارائه شده را جمع بندی کرده و نتیجه گیری می کنیم.

فصل دوم

تلاش های پیشین در شبکه های عصبی با رویکرد
توجه

اخیراً، سازوکار توجه، به طور گسترده ای برای طبقه بندی گره ها به صورت نیمه نظارت شده استفاده شده است. GNN مبتنی بر توجه [۶]، با توجه به شباهت کسینوس بین دو گره متصل به هم و بدون هیچ پارامتر قابل یادگیری، وزن توجه را به هر یال اختصاص می دهد. شبکه توجه گرافی [۷]، AGNN را با تخمین وزن توجه با تابع رگرسیون قابل یادگیری و استفاده از چندین هد توجه^۸ گسترش می دهد. شبکه توجه دروازه ای^۹ [۸]، از مکانیسم توجه به خود استفاده می کند، که برای هر سر وزن خاصی را محاسبه می کند. شبکه های توجه گرافی ناهمگن [۹]، توجه سلسله مراتبی، از جمله توجه در سطح گره و در سطح متافی^{۱۰} را به کار می گیرد. شبکه های کانولوشنی Motif، با اجازه دادن به هر گره برای حضور در مهم ترین محله ناشی از شکل، مکانیسم توجه را در خود جای داده است. شبکه های همگرایی توجه گرافی دوگانه [۱۰]، برای بهبود بیشتر عملکرد، از دو واحد توجه استفاده می کند. بسیاری از روش های اخیر بیشتر از تأثیر انواع مختلف اطلاعات توپولوژی مرتبه بالا، مانند مسیر، شکل و تعبیه، در تجمیع ویژگی ها بهره برداری می کنند. GeniePath [۱۱]، مکانیسم دروازه شبیه به حافظه های کوتاه مدت ماندگار^{۱۱} را برای جمع آوری اطلاعات در چندین لایه کانولوشن گرافی ارائه می دهد. SPAGAN [۱۲]، توجه مبتنی بر همسایگان را در GAT به توجه مبتنی بر مسیر گسترش می دهد که می تواند اطلاعات توپولوژی سراسری را به طور قوی و موثر کشف کند. شبکه های کانولوشنی motif [۱۳]، برای به دست آوردن همسایه های مرتبه بالاتر، از ماتریس های همسایگی موتیف multi-hop به صورت وزن دار استفاده کرده و مکانیسم توجه جدیدی را پیشنهاد می دهد تا هر گره مجزا بتواند مناسب ترین همسایه را برای انتشار انتخاب کند. اثر انگشت ساختاری تطبیقی^{۱۴} [۱۴]، با طراحی یک محله از همسایگی ها به صورت وزن دار و قابل یادگیری، که ساختارهای گرافیکی محلی متنوع را برای هر گره کد می کند، GAT را بهبود می بخشد. Geom-GCN [۱۵]، یک طرح تجمع هندسی تغییر ناپذیر جدید را پیشنهاد می کند که با در نظر گرفتن گراف تعبیه شده، همسایگی ها را تقویت می کند. اکثر رویکردهای موجود، که اطلاعات سراسری را در نظر می گیرند، فقط توپولوژی مرتبه بالا (مسیر یا دنباله گام ها تصادفی) را برای انتشار ویژگی ها کشف می کنند، در حالی که فعل و انفعالات مهم بین ویژگی های محلی / سراسری را نادیده می گیرند.

⁷ Attention-based GNN (AGNN)

⁸ attention heads

⁹ Gated Attention Network (GaAN)

¹ Self-attention

¹ metapath-level

¹ Motif Convolutional Networks (MCN)

¹ Dual Attention Graph Convolutional Networks (DAGCN)

¹ Long short term memory (LSTM)

¹ Adaptive Structural Fingerprint (ADSF)

فصل سوم

شبکه‌های عصبی با رویکرد توجه

۳-۱ شبکه‌های توجه گرافی

اخیراً، شبکه‌های عصبی گرافی روی پردازش داده‌های شبکه از طریق یادگیری عمیق درخشانند. GNN ها با استفاده از شبکه‌های عصبی عمیق بر روی دامنه فوریه گراف ایجاد می‌شود. متأسفانه، پیچیدگی بالای تجزیه ویژه^۶ مانع از استفاده GNN ها در عمل، به ویژه در شبکه‌های مقیاس بزرگ می‌شود. برای حل این مسئله، بسیاری از رویکردهای تقریب، مانند بسط کوتاه شده‌ی چند جمله‌ای چبیشف، استفاده شده است. شبکه‌ی کانولوشنی گرافی^۸ که از تقریب مرتبه اول کانولوشن گرافی طیفی^۹ ایجاد می‌شود، فاصله‌ای را بین کانولوشن گرافی طیفی و انتشار ویژگی‌های مکانی ایجاد می‌کند و با مقیاس پذیری مناسب نتایج رضایت بخشی را به دست می‌آورد. GCN به طور گسترده‌ای در طبقه بندی گره‌های نیمه نظارت شده، تعبیه شبکه^۲، خوشه بندی و پیش‌بینی پیوند و غیره استفاده شده است.

۳-۱-۱ نشانه گذاری‌ها^{۲۱}

یک شبکه که گره‌های آن از ویژگی‌های خاصی برخوردار است، می‌تواند به عنوان یک گراف $G=(V,E,X)$ مدل شود. $V=\{v_i | i = 1.2 \dots N\}$ مجموعه‌ای از N راس می‌باشد که هر راس آن به یک ویژگی $x_i \in R^f$ منتسب شده است. توپولوژی شبکه از مجموعه‌ای از یال‌های $E=\{e_i | i = 1. \dots M\}$ تشکیل شده است که هر کدام از آن‌ها دو راس را به یک دیگر متصل می‌کند. $X=[x_{ij}] \in R^{N \times N}$ مجموعه‌ی ویژگی‌های منتسب شده را شامل می‌شود. هر سطر از ماتریس X با ویژگی‌های یک گره x_i' مطابقت دارد. برای سهولت $x_i \in R^f$ و $x_j \in R^N$ به ترتیب برای نمایش سطر i ام (تمام صفات راس i) و j امین ستون (ویژگی j ام تمامی رئوس) استفاده شده است. همچنین ماتریس همجواری، $A=[a_{ij}] \in R^{N \times N}$ توپولوژی شبکه را نشان می‌دهد، که اگر یالی بین راس i و j وجود داشته باشد، $a_{ij}=1$ است و در غیر این صورت $a_{ij}=0$. ماتریس قطری $D=\text{diag}(d_1, d_2, \dots, d_n)$ ماتریس درجه‌های A می‌باشد که $d_i = \sum_j a_{ij}$ درجه راس i ام را بیان می‌کند. لاپلاسیان گراف و فرم نرمال شده آن به ترتیب به صورت $L=D-A$ و $L^{\wedge}=D^{-1/2}L D^{-1/2}$ تعریف می‌شود. برچسب‌های رئوس برچسب گذاری شده $v_l \subset V$ با استفاده از $Y=[y_{ik} \in R^{N \times k}]$ که k تعداد خوشه‌ها را

1 Graph neural network (GNN) 6
1 eigen-decomposition 7
1 Graph Convolutional Network 8
1 Spectra graph convolutions 9
2 Network embedding 0
2 Notations 1

نشان می دهد، نمایش داده می شود. اگر $y_{ik} = 1$ باشد راس i ام متعلق به v_l و k امین خوشه می باشد و در غیر این صورت مساوی صفر می باشد. برای راحتی فرض شده است، l گره اول دارای برچسب می باشند و یادگیری نیمه نظارت شده، بقیه گره های بدون برچسب را طبقه بندی می کند.

۳-۱-۲ شبکه کانولوشنی گرافی

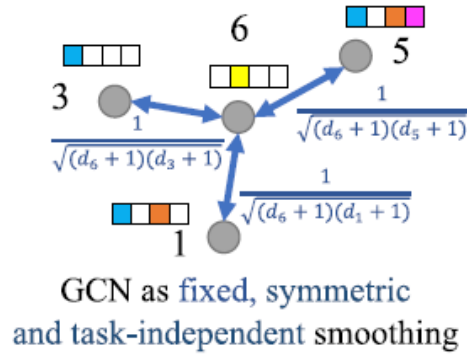
شبکه‌ی کانولوشنی گرافی، بر اساس تقریب مرتبه اول کانولوشن گرافی طیفی ایجاد می شود. کانولوشن طیفی بر روی گراف ها معادل با ضرب سیگنال $x \in R^N$ و یک فیلتر $g_\theta = \text{diag}(\theta)$ که پارامتر آن، θ می باشد در دامنه ی فوریه به صورت $g_\theta * x$ می باشد که با استفاده از تعمیم سیگنال $x \in R^N$ به یک سیگنال به صورت $x \in R^{N \times F}$ به تعداد کانال های ورودی F (ستون های ورودی) و گسترش یک فیلتر θ به k فیلتر $W \in R^{K \times F}$ ، کانولوشن های طیفی بر روی گراف ها با یک ترفند نرمال سازی مجدد توسط رابطه زیر تخمین زده می شود:

$$h'_i = \sigma \left(W \sum_{j \in N_i \cup \{i\}} \frac{1}{\sqrt{(d_i + 1)(d_j + 1)}} h_j \right) \quad (1-3)$$

که $\sigma(\cdot)$ یک تابع فعال ساز غیر خطی مثل سیگموئید یا $RELU$ می باشد و h_i ویژگی های گره v_i می باشد که با استفاده از x_i مقدار دهی اولیه شده است و با استفاده از انتشار ویژگی تکامل می یابد. پارامتر W نیز توسط حداقل کردن $cross\ entropy$ بین برچسب های پیش بینی شده و برچسب های واقعی از گره های دارای برچسب به صورت زیر یاد گرفته می شود.

$$\mathcal{L} = - \sum_{v \in V_l} \sum_{k=1}^K Y_{lk} \ln h_{l,(k)}. \quad (2-3)$$

GCN به عنوان یک هموارساز لاپلاسین معرفی شده است. با توجه به این تفاسیر، دو ویژگی اساسی انتشار در GCN وجود دارد، یعنی وزن انتشار ثابت و متقارن. اول اینکه، وزن تکثیر تحت تأثیر برچسب های داده شده قرار نمی گیرد و آنها به طور کامل توسط درجه دو گره متناظر تعیین می شوند و دوم اینکه، وزن انتشار دو جهت در یک یال یکسان است و برابر با مقدار $\frac{1}{\sqrt{(d_i+1)(d_j+1)}}$ است. همانطور که در شکل ۱ نشان داده شده است، به دلیل این دو ویژگی، ویژگی های پس از انتشار نرم تر از قبل از انتشار است.



شکل ۱: نمونه ای از GCN

۳-۱-۳ شبکه‌ی توجه گرافی

با توجه به توسعه اخیر مکانیسم های توجه ، شبکه توجه گرافی برای محاسبه نمایش های نهان هر گره با حضور در همسایگان خود از طریق یک استراتژی توجه به خود، ارائه شده است. متفاوت از GCN، جایی که وزن یال ها توسط درجه هر دو گره متصل تعیین می شود، وزن یال در GAT را می توان از طریق ویژگی های هر دو گره متصل از طریق یک مدل رگرسیون نرمال آموخت. مدل رگرسیون به صورت زیر تعریف شده است:

$$e_{ij} = f(Wh_i, Wh_j) = \text{LeakyReLU}(b^T [Wh_i || Wh_j]), \quad (3-3)$$

که علامت $||$ به معنی عملیات الحاق است و $f(\cdot, \cdot)$ یک شبکه عصبی تک لایه پیشرو که با یک بردار وزن قابل یادگیری b پارامتر بندی شده است. همچنین نرمال سازی در همسایگی های محلی نیز با یک تابع softmax طبق فرمول زیر انجام شده است:

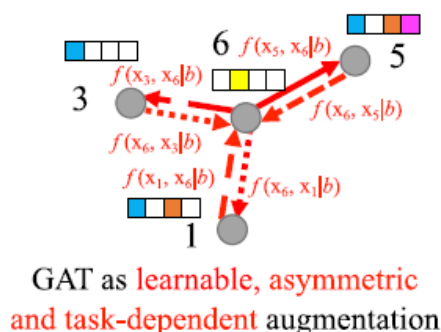
$$\alpha_{ij} = \text{softmax}_j(e_{ij}) = \frac{e_{ij}}{\sum_{k \in N_i} e_{ik}} \quad (4-3)$$

در نهایت انتشار ویژگی با استفاده از وزن یال های یادگرفته شده α_{ij} به صورت زیر انجام می شود:

$$h'_i = \sigma(W \sum_{j \in N_i} \alpha_{ij} h_j) \quad (5-3)$$

فصل سوم: شبکه‌های عصبی با رویکرد توجه

و پارامترهای W و b نیز با حداقل کردن معادله (۳-۲) همانند GCN به دست می‌آیند.



شکل ۲: نمونه ای از GAT

با مقایسه قوانین انتشار GAT در معادله (۳-۵) با GCN در معادله (۳-۱)، می‌توانیم مشاهده کنیم که فلسفه‌های اساسی آنها در دو جنبه زیر کاملاً متفاوت است، اگرچه آنها شکل‌های مشابه به هم دارند، همانطور که در شکل‌های ۱ و ۲ نشان داده شده است.

- ۱- وزن انتشار در GAT قابل یادگیری است. وزن یال یاد گرفته شده، همانطور که در معادله (۳-۴) نشان داده شده است، توسط هر دو ویژگی گره و برچسب‌های داده شده تحت تأثیر قرار می‌گیرند. بنابراین، GAT به جای هموار کردن مستقیم ویژگی‌های گره در همسایگی‌های محلی در GCN، با انتخاب ویژگی‌های همسایه که ممکن است در طبقه بندی گره مفید باشد، ویژگی‌های گره را تقویت می‌کند.
- ۲- وزن‌های انتشار در GAT نامتقارن است؛ زیرا $f(\cdot, \cdot)$ نامتقارن است. که این خاصیت به GAT درجه‌ی آزادی بیشتری برای انجام عملیات انتشار می‌دهد. به عنوان مثال برای دو گره متصل به هم v_i و v_j ، v_i ممکن است مقدار زیادی از ویژگی‌ها را از راس v_j بپذیرد ولی v_j مقدار ویژگی کمی را از v_i بپذیرد.

به طور کلی، وزن‌های انتشار و نامتقارن و قابل یادگیری، GAT را قادر می‌سازد تا از GCN در طبقه بندی گره نیمه نظارتی پیشی بگیرد.

مکانیسم GAT-Global هم با استفاده از اهرم سراسری تمام برچسب‌ها و گره‌های داده شده از همان خوشه، ویژگی‌های گره را تقویت می‌کند.

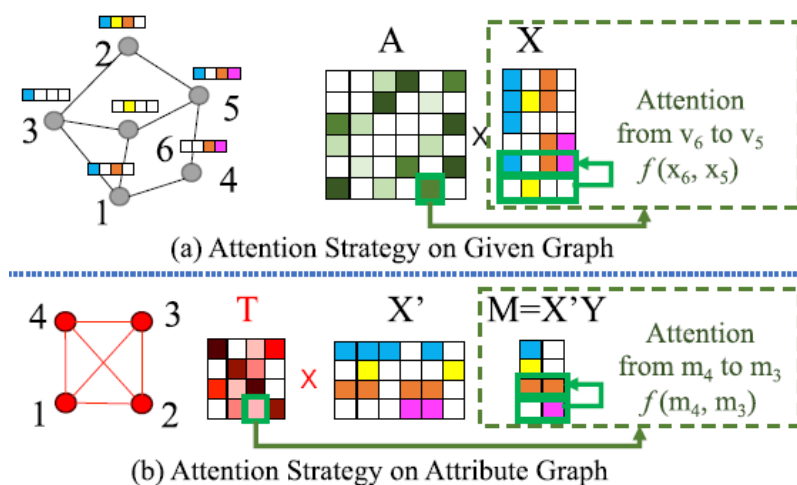
فصل چهارم

شبکه‌های گرافی با رویکرد توجه مشارکتی^{۲۲}

² Graph Co-Attention Network (Graph-CAT)

۴-۱ شبکه‌های گرافی توجه مشارکتی

شبکه‌های گرافی توجه مشارکتی^۲، از هر دو ویژگی محلی و سراسری برای تقویت ویژگی‌های گره بهره‌برداری می‌کند. در Graph-CAT، منظور از توجه مشارکتی به این معنی است که مکانیسم توجه به طور همزمان بر روی گراف داده شده و استخراج همبستگی ویژگی‌های سراسری اعمال می‌شود. تقویت ویژگی‌های محلی با اتخاذ مکانیسم توجه منظم در GAT انجام می‌شود. استخراج همبستگی ویژگی سراسری با استفاده از مکانیسم توجه به گراف ویژگی حاصل می‌شود. این گراف ویژگی همانطور که در شکل ۳ قسمت (b) نشان داده شده است، با ایجاد هر گره مربوط به یک ویژگی ساخته شده است. این گراف ویژگی ساخته شده، یک گراف کاملاً متصل با یک ماتریس همجواری قابل یادگیری $T = [t_{pq}] \in R^{P \times P}$ می‌باشد؛ که هر وزن یال قابل یادگیری t_{pq} بیانگر همبستگی بین دو ویژگی متناظر است. متفاوت از GAT، که هر ردیف از ماتریس ویژگی به عنوان سیگنال در هر گره از گراف داده شده در نظر گرفته می‌شود، همانطور که در شکل ۳ قسمت (a) نشان داده شده است، هر ستون از ماتریس ویژگی می‌تواند به عنوان سیگنال در هر گره از نمودار ویژگی ساخته شده، در نظر گرفته شود. از آنجا که بعد ویژگی‌ها ثابت است و معمولاً بسیار کمتر از تعداد گره‌ها در عمل است، محاسبه توجه در بین جفت ویژگی‌ها عملی است در حالی که در بین هر جفت گره اینگونه نیست. بنابراین، استخراج همبستگی ویژگی سراسری در مقایسه با GAT Global کارآمدتر و عملی‌تر است.



شکل ۳: مقایسه استراتژی‌های مختلف توجه در گراف داده شده و گراف ویژگی ساخته شده.

² Graph Co-Attention Network (Graph-CAT)

مکانیسم توجه، که دو ردیف متناظر از ماتریس ویژگی را به عنوان ورودی می‌گیرد، به دلیل دو مشکل زیر برای پردازش مستقیم گراف ویژگی مناسب نیست.

- ۱- بعد سیگنال‌ها در گراف ویژگی ساخته شده، یعنی بعد ستون ماتریس ویژگی X ، همان تعداد گره‌ها در شبکه داده شده است. متأسفانه، این بعد بسیار زیاد است و سیگنال‌ها کاملاً پراکنده هستند. این به بردار b که تعداد پارامترهای زیادی را پردازش کند، نیاز دارد، که ممکن است باعث ایجاد بیش برآزش شود.
- ۲- سیگنال‌های ورودی دارای نویز می‌باشند که ممکن است در تخمین همبستگی ویژگی‌ها، خطا ایجاد کند. همزمانی ویژگی‌ها در هر گره ممکن است از نظم آماری برخوردار نباشد. برعکس، توزیع دسته‌ای ویژگی‌ها، که همبستگی ویژگی‌های سراسری را منعکس می‌کند، معنادارتر است.

برای حل مشکلات فوق، توزیع‌های دسته‌ای ویژگی‌ها برای محاسبه وزن توجه استفاده می‌شود، همانطور که در شکل ۳ قسمت (b) نشان داده شده است. بنابراین، یک ماتریس $M \in R^{F \times K}$ ، که هر ردیف از آن توزیع دسته‌ای یک ویژگی را نشان می‌دهد، به صورت $M = X^T Y$ محاسبه می‌شود که Y ماتریس برچسب‌ها می‌باشد. سپس توجه بین ویژگی‌ها بر اساس ردیف‌های M به جای ستون‌های X ، به صورت زیر تخمین زده می‌شود:

$$t_{pq} = \frac{\text{LeakyReLU}(c^T [m_p || m_q])}{\sum_r \text{LeakyReLU}(c^T [m_p || m_r])}, \quad (1-4)$$

که $c \in R^{2K}$ با k که برابر با تعداد خوشه‌ها می‌باشد و t_{pq} که یک درایه از ماتریس T است و m_p برابر با p امین ردیف ماتریس M می‌باشد. در معادله (۱-۴)، همبستگی سراسری بین دو ویژگی را می‌توان توسط یک تابع محاسبه کرد، که با c پارامتر می‌شود و توزیع‌های دسته‌ای دو ویژگی را به عنوان ورودی می‌گیرد و بعد در ادامه ویژگی‌های هر گره به صورت زیر محاسبه می‌شود:

$$h'_{i.(p)} = \sigma \left(\sum_q t_{pq} h_{i.(q)} \right) \quad (2-4)$$

که $h'_{i.(p)}$ ، P امین درایه از گره v_i است و ماتریس آن به صورت $h'_i = \sigma(T h_i)$ محاسبه می‌شود. با ترکیب فرمول (۵-۳) و ماتریس h'_i که به ترتیب تقویت محلی و سراسری را فرموله می‌کند، ویژگی خروجی نهایی Graph-CAT را می‌توان به صورت زیر بدست آورد:

$$h'_i = \sigma(WT \sum_{j \in N_i} \alpha_{ij} h_j) \quad (3-4)$$

Graph-CAT هم از تابع هزینه cross-entropy همانند GAT و GCN استفاده می‌کند. ابتدا T به عنوان ماتریس هویت مقدار دهی اولیه می‌شود و پارامترهای b و W بهینه می‌شوند. سپس، $c, \{b, W\}$ از طریق بهینه‌ساز آدام^۴ به روز می‌شوند.

۴-۲ نتایج

برای اعتبار بخشیدن به اثربخشی Graph-CAT پیشنهادی، به طور تجربی عملکرد آن را در هر دو طبقه بندی گره نیمه نظارتی انتقالی^۵ و استقرایی^۴ ارزیابی می‌کنیم.

۴-۲-۱ یادگیری نیمه نظارتی انتقالی

برای ارزیابی عملکرد برای یادگیری نیمه نظارتی انتقالی، سه شبکه‌ای که معمولاً مورد استفاده قرار می‌گیرند (Cora, CiteSeer و PubMed) و یک شبکه بزرگ دو بخشی (NELL) اتخاذ می‌شوند، همانطور که در جدول ۱ نشان داده شده است.

جدول ۱: مجموعه ی داده

Dataset	#Nodes	#Edges	#Classes	#Features
CiteSeer	3,327	4,732	6	3,703
Cora	2,708	5,429	7	1,433
PubMed	19,717	44,338	3	500
NELL	65,755	266,144	210	5,414

جدول ۲: نتایج یادگیری انتقالی از نظر دقت.

Methods	Cora	Citeseer	Pubmed	NELL
MLP	55.1%	46.5%	71.4%	22.9%
ManiReg	59.5%	60.1%	70.7%	21.8%
SemiEmb	59.0%	59.6%	71.7%	26.7%
LP	68.0%	45.3%	63.0%	26.5%
DeepWalk	67.2%	43.2%	65.3%	58.1%
ICA	75.1%	69.1%	73.9%	23.2%
Planetoid	75.7%	64.7%	77.2%	61.9%
Chebyshev	81.2%	69.8%	74.4%	-
MoNet	81.7%	69.9%	78.8%	64.2%
GCN	81.5%	70.3%	79.0%	66.0%
GAT	83.0%	72.5%	79.0%	64.5%
LGCN	83.3%	73.0%	79.5%	-
GWNN	82.8%	71.7%	79.1%	-
DGI	82.3%	71.8%	76.8%	-
SPAGAN	83.6%	73.0%	79.6%	-
Graph-CAT	84.8%	73.6%	80.5%	68.7%

² Adam optimizer

² Transductive semi-supervised node classification

² Inductive semi-supervised node classification

همانطور که از نتایج جدول ۲ مشخص است، عملکرد Graph-CAT پیشنهادی، نسبت به روشهای پیشرفته از جمله GCN و GAT در همه شبکه‌ها بهتر است. از آنجا که برای جبران تقویت ویژگی محلی، یک تقویت ویژگی سراسری پیشنهاد شده است، Graph-CAT می‌تواند با موفقیت تجمع خطا را کاهش دهد. عملکرد برجسته این روش، نه تنها اهمیت اطلاعات سراسری را نشان می‌دهد، بلکه نشان‌دهنده اثربخشی آن، در تلفیق اطلاعات سراسری است [۱].

۲-۲-۴ یادگیری نیمه نظارتی استقرایی

مجموعه داده‌های (PPI) و مجموعه داده‌های Reddit برای یادگیری استقرایی استفاده می‌شود و نتایج برای مجموعه داده‌ی PPI به شرح زیر است:

جدول ۳: نتایج یادگیری استقرایی از نظر micro-F1 score

Methods	PPI
Random	0.396
Logistic Regression	0.422
GraphSAGE-mean	0.598
GraphSAGE-LSTM	0.612
GraphSAGE-pool	0.600
Inductive GCN	0.500
LGCN	0.772
GAT-Const	0.934
DGI	0.638
GAT	0.973
Graph-CAT	0.981

افزایش عملکرد GAT در مقایسه با GAT-Const که فقط از یک وزن ثابت توجه استفاده می‌کند، برتری تقویت نسبت به هموارسازی را نشان می‌دهد. همبستگی ویژگی سراسری، که از گراف‌های آموزشی یاد گرفته می‌شود، با توجه به افزایش عملکرد رضایت بخش Graph-CAT در مقایسه با GAT می‌تواند با موفقیت به گراف‌های آزمون در کار یادگیری استقرایی منتقل شود. می‌توان مشاهده کرد که پیشرفت‌های Graph-CAT پیشنهادی در کار یادگیری استقرایی قابل توجه‌تر از یادگیری انتقالی است که به دلایل زیر است. در مرحله اول، اطلاعات برجسته در یادگیری استقرایی به میزان کافی در مقایسه با یادگیری انتقالی ارائه می‌شود. ثانیاً، نسبت تعداد ویژگی‌ها به تعداد گره‌ها در یادگیری استقرایی نسبت به یادگیری انتقالی کمتر است. هر دوی آنها تمایل دارند که تخمین توزیع‌های دسته‌ای ویژگی‌ها را بهبود بخشند. بنابراین، تخمین همبستگی ویژگی در مقایسه با یادگیری انتقالی قوی‌تر خواهد بود [۱].

فصل پنجم

شبکه‌های گرافی با رویکرد توجه چند نمایه^{۲۷}

² Multi-view Graph Attention Networks

۵-۱ مکانیسم توجه

مکانیسم توجه با موفقیت در وظایف مبتنی بر دنباله و توالی‌ها اعمال شده است. این مکانیسم اجازه می‌دهد تا وابستگی‌ها بدون در نظر گرفتن فاصله بین دنباله‌ی ورودی و خروجی مدل سازی شود. مزیت مکانیسم توجه این است که مهم نیست که اندازه ورودی چیست؛ سازوکار توجه بطور مداوم بر روی مهم‌ترین قسمت متمرکز می‌شود. وقتی مکانیسم توجه برای یادگیری بازنمایی یک دنباله منفرد معرفی می‌شود، به طور کلی به عنوان درون توجه^{۲۸} یا توجه به خود^{۲۹} توصیف می‌شود. با داشتن شبکه‌های حافظه کوتاه مدت (LSTM) و شبکه‌های عصبی بازگشتی (RNN)، توجه به خود برای انجام کارهای یادگیری ماشین مانند ترجمه ماشین و خواندن ماشین کارآمد است.

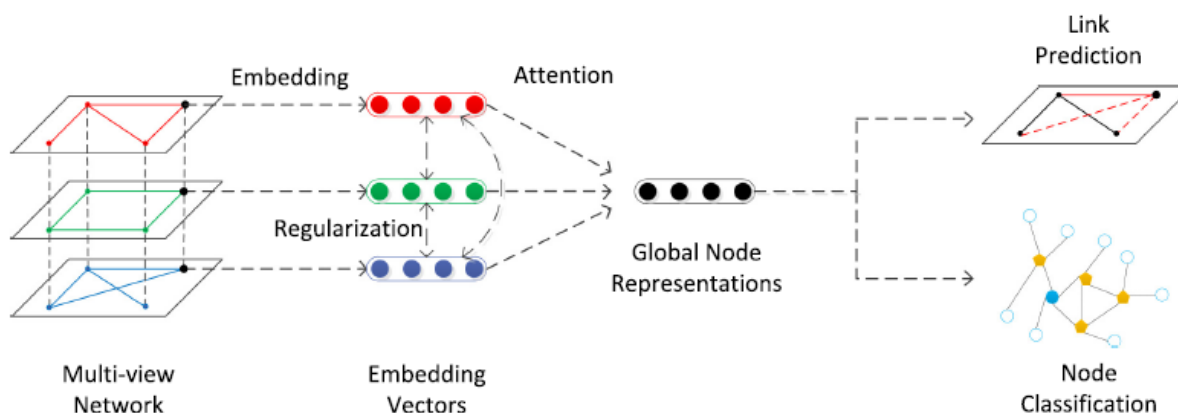
۵-۲ تعبیه شبکه^{۳۰}

تعبیه شبکه تلاش می‌کند تا یک شبکه پیچیده را در فضایی با ابعاد پایین تر برای تجزیه و تحلیل بیشتر شبکه‌ها رمزگذاری کند. رویکردهای اولیه تعبیه شبکه شامل اجزای محاسباتی پیچیده‌ای برای تجزیه و تحلیل خصوصیات طیف گراف‌ها است. با این حال، هنگامی که شبکه گرافی بزرگ باشد، چنین روش‌هایی ناکارآمد خواهند بود، زیرا محاسبه تجزیه ویژه لاپلاسیان گراف در حوزه فوریه هزینه‌بر است. اخیراً، پیرو افکار رویکردهای مبتنی بر random walk، DeepWalk، node2vec و skip-gram را در تعبیه گراف معرفی می‌کنند. این روش‌ها از مسیرهای نمونه به دست آمده با random walk برای یادگیری بردارهای تعبیه‌شده استفاده می‌کنند. شبکه‌های عصبی گرافی (GNN) به دلیل برتری آموزش انتها به انتها^{۳۱} بدون تکیه بر ویژگی‌های دست‌ساز، به طور فزاینده‌ای رایج شده‌است. به عنوان نمونه ای از GNN ها می‌توان به شبکه‌های کانولوشن گراف (GCN) و شبکه‌های کانولوشن گراف رابطه‌ای^{۳۲} و GAT نام برد که در فصل‌های گذشته به آن اشاره شد. عملکرد GAT کارآمد بوده است. با اختصاص وزن به همسایه‌ها، می‌توان GAT را روی گره‌های با درجه‌های مختلف اعمال کرد. اما به دنبال بیشتر روش‌های دیگر تعبیه گراف، GAT فقط مربوط به یادگیری

² intra-attention	8
² self-attention	9
³ Network embedding	0
³ End-to-end	1
³ Relational graph convolutional networks (R-GCNs)	

فصل پنجم: شبکه‌های گرافی با رویکرد توجه چند نمایه

نمایش شبکه تک نمایه^۳ است. با استفاده از این مورد، می‌توان این روش کارآمد تعبیه گراف مبتنی بر توجه را از شبکه‌های تک نمایه به شبکه‌های چند نمایه تبدیل کرد. اثبات شده است که تمام رویکردهای ذکر شده در شبکه‌های گرافی تک نمایه مفید هستند، اما برای شبکه‌های چند نمایه چالش برانگیز، عملکرد خوبی ندارند. با این وجود، معمولاً انواع مختلفی از روابط بین اشیاء در دنیای واقعی وجود دارد و تعداد زیادی از سیستم‌های دنیای واقعی به صورت شبکه‌های ترکیبی^۴ سازمان یافته‌اند. برای شبکه‌های اطلاعاتی چند نمایه، یک شبکه با تعداد نمای K را به صورت $G = \{V, E_1, E_2, \dots, E_K\}$ نمایش می‌دهند. تعبیه شبکه چند نمایه، به دنبال یادگیری نمایش‌های تعبیه شده گره سراسری $\{h_i\}_{v_i \in V} \subseteq R^d$ با $d \ll |V|$ می‌باشد. از آنجا که برای نمایش گره‌ها برای هر نما به صورت جداگانه، تعاملات بین لایه‌های مختلف در نظر گرفته نمی‌شود، منطقی است که از طریق همکاری بهتر بین نماهای متنوع، نماهای گراف سراسری را یاد بگیریم. برای تعبیه شبکه‌های چند نمایه، یک مدل شبکه توجه گرافی چند نمایه^۵ طراحی می‌شود و چارچوب آن در شکل ۴ نشان داده شده است.



شکل ۴: چارچوب شبکه‌های توجه گرافی چند نمایه

در ابتدا تک تک نماها را تعبیه کرده و مجموعه‌ای از نمایش گره‌ها با نمای خاص هر گره را یاد می‌شود. از آنجا که تعاملات اطلاعاتی متقابل بین نماهای مختلف وجود دارد، یک عبارت منظم سازی را برای ایجاد ترغیب در شباهت پارامترهای شبکه با نماهای مختلف ابداع می‌شود. سپس یک مدل مبتنی بر توجه مورد بررسی قرار می‌گیرد تا وزن نماها را بیاموزد و گره‌ها را قادر سازد تا بر نمایی حاوی بیشترین اطلاعات تمرکز کنند. سرانجام،

³ single-view 3
³ multiplex 4
³ Multi-view Graph Attention Network (MGAT)

فصل پنجم: شبکه‌های گرافی با رویکرد توجه چند نمایه

نمایش گره‌ها برای هر نمای جداگانه، توسط وزن‌های آموخته شده در یک نمایش گره سراسری ادغام می‌شوند. پس از آن، از نمایش سراسری برای کارهای بعدی تجزیه و تحلیل گراف استفاده می‌شود. این مدل MGAT، شامل دو جز اصلی است: یک مولفه تعبیه گراف برای نمای تکی و یک مولفه تجمع مبتنی بر توجه برای قطعات چند نمایه. به طور خاص، مولفه تعبیه شده در گراف، مجاورهای هر گره را استخراج می‌کند، که طی آن یک مدل محدود GAT برای رمزگذاری هر نمای خاص استفاده می‌شود. مولفه تجمع مبتنی بر توجه نماهای مختلفی را برای نمایش گره‌های جامع‌تر ادغام می‌کند.

۳-۵ نتایج

این روش پیشنهادی و روش‌های پایه در شش شبکه چند نمای جهان واقعی ارزیابی می‌شود که عبارتند از: CKM، LAZEGA، CiteSeer، CoRA، DBLP و Twitter. آمار دقیق گره‌ها، لبه‌ها، برچسب‌ها و نماها در جدول ۴ خلاصه شده است.

جدول ۴: اطلاعات شش شبکه آزمایش شده

Network	#Nodes	#Edges	#Labels	#Views
CKM	246	1551	4	3
LAZEGA	71	2571	3	3
CiteSeer	3312	24,233	6	2
CoRA	2708	21,190	7	2
DBLP	69,110	1,884,236	8	3
Twitter	456,626	15,183,974	/	2

۱-۳-۵ پیش‌بینی پیوند

پیش‌بینی پیوند گاربرد اساسی تعبیه شبکه است. با حذف برخی از پیوندهای یک شبکه داده شده، هدف از پیش‌بینی پیوند، پیش‌بینی موارد گم‌شده است. برای شبیه‌سازی، بخشی از پیوندهای موجود در گراف به طور تصادفی به عنوان پیوندهای گم‌شده انتخاب می‌شوند. مدل بر روی پیوندهای باقیمانده آموزش داده شده تا پیش‌بینی شاخص‌های از دست رفته را ارائه دهد و عملکردش با مقایسه آن با ماتریس مجاورت اصلی شبکه، ارزیابی شود [۲]. از نتایج جدول ۵، مشاهدات اصلی زیر را داریم:

³ Link prediction

فصل پنجم: شبکه‌های گرافی با رویکرد توجه چند نمایه

جدول ۵: نتایج روش‌های مختلف در پیش بینی پیوند

Methods	Algorithms	CKM	LAZEGA	CoRA	CiteSeer	Twitter
Embedding methods	DeepWalk	0.761	0.785	0.688	0.685	0.504
	LINE	0.630	0.654	0.631	0.620	0.517
	node2vec	0.769	0.781	0.712	0.691	0.506
	NetMF	0.793	0.812	0.650	0.703	0.523
	MELL	0.846	0.848	0.768	0.725	0.598
	MNE	0.857	0.842	0.815	0.795	0.625
Structure-based methods	CN	0.862	0.809	0.781	0.752	0.590
	JC	0.861	0.827	0.772	0.755	0.521
	AA	0.864	0.813	0.778	0.764	0.591
Our methods	MGAT	0.915	0.853	0.826	0.835	0.638

- 1- با مقایسه تمام روش‌های پایه، می‌توان دریافت که برای کارهای پیش بینی پیوند، روش‌های تعبیه طراحی شده برای شبکه‌های چند نمایه نتایج رقابتی دارند و روش‌های مبتنی بر ساختار، عملکرد بهتری نسبت به روش‌های طراحی شده برای شبکه‌های تک نمایه دارند. این منطقی است زیرا اطلاعات ارتباط از شبکه‌های گرافی تک نمایه بسیار ساده است و ممکن است در نماهای مختلف مکمل یکدیگر شوند.
- 2- تقریباً برای همه مجموعه‌های داده شبکه چند نمایه، در نظر گرفتن انواع مختلف رابطه به طور مشترک مفید است [۲].
- 3- به طور خلاصه، مدل پیشنهاد شده، تقریباً از همه روش‌های اساسی کار پیش بینی پیوند بهتر است. دلیل اصلی این است که با محدود کردن پارامترهای شبکه عصبی و یادگیری وزن‌های مختلف نماهای مختلف به صورت مشترک، می‌توانیم توانایی تعمیم تعبیه شده در شبکه را افزایش دهیم، که از اهمیت زیادی برای پیش بینی پیوند برخوردار است [۲].

۵-۳-۲ طبقه‌بندی گره‌ها^{۳۷}

در شبکه‌های دنیای واقعی، در بیشتر موارد، فقط چند گره برچسب‌گذاری می‌شوند و بقیه آنها برای تعیین برچسب مربوط به خودشان، باقی مانده‌اند. با نظارت بر عملکرد وظایف طبقه‌بندی گره، می‌توان به طور مستقیم اثر الگوریتم‌های تعبیه شده در شبکه را آشکار کرد. نتایج طبقه‌بندی گره در جدول ۶ نشان داده شده است و ما مشاهدات اصلی زیر را داریم.

³ Node classification

فصل پنجم: شبکه‌های گرافی با رویکرد توجه چند نمایه

جدول ۶: نتایج روش‌های مختلف در طبقه بندی گره.

Methods	Algorithms	CKM	LAZEGA	CoRA	CiteSeer	DBLP
Embedding methods	DeepWalk	0.925	0.927	0.691	0.585	0.711
	LINE	0.804	0.828	0.517	0.491	0.682
	node2vec	0.891	0.920	0.683	0.582	0.716
	NetMF	0.935	0.931	0.733	0.573	0.693
	R-GCNs	0.966	0.953	0.813	0.681	0.741
	MELL	0.960	0.944	0.836	0.684	0.732
	MNE	0.973	0.965	0.824	0.706	0.737
Our methods	MGAT-conte	0.951	0.924	0.827	0.724	0.727
	MGAT-noattn	0.962	0.947	0.831	0.715	0.734
	MGAT	0.977	0.963	0.856	0.729	0.750

مدل MGAT می تواند در اکثر مجموعه‌های داده، نتایج قانع کننده ای کسب کند. برای CoRA و CiteSeer با هزاران گره، این مدل بهترین عملکرد را در بین تمام روش‌های مبتنی بر تعبیه دارد و برای CKM و LAZEGA شامل صدها گره، روش پیشنهادی کمی بهتر از روش‌های پیشرفته است [۲].

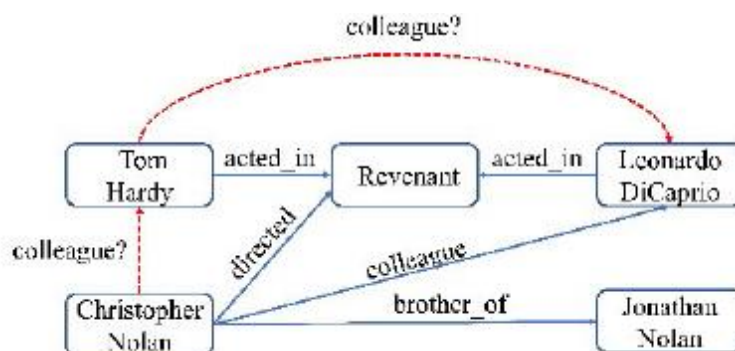
فصل ششم

شبکه های توجه گرافی با آگاهی از ساختار محلی برای تکمیل گراف دانش^{۳۸}

³ Graph Attention Networks With⁸ Local Structure Awareness for Knowledge Graph Completion

۶-۱ گراف دانش^{۳۹}

گراف های دانش (KGS) حامل های امروزه برای دانش های مختلف و رایج هستند. آنها به عنوان هسته اصلی بسیاری از پیشرفته ترین راه حل های پردازش زبان طبیعی برای بسیاری از کاربردهای عملی از جمله پاسخ به سوال، درک مطلب و غیره عمل می کنند. با این حال، این واقعیت که گراف دانش معمولاً ناقص است و فاقد واقعیت ها، موجودیت ها یا روابط است، می تواند به طور چشمگیری مانع عملکرد خوب شود. بنابراین، تکمیل گراف دانش ضروری و حیاتی است. شکل ۵ یک زیر گراف از گراف دانش را با حقایق واقعی متشکل از روابط واقعی بین دو موجودیت (خطوط آبی رنگ) نشان می دهد. روابط به صورت (t؛ r؛ h) سه گانه سازمان یافته است (به عنوان مثال "Tom Hardy"؛ h = "acted_in"؛ r = "Revenant"؛ t = "Christopher Nolan"). هدف از تکمیل گراف دانش، استدلال در مورد سه گانه موجود در گراف برای یافتن پیوندهای گمشده (به عنوان مثال خط نقطه چین قرمز در شکل ۵) است [۳].

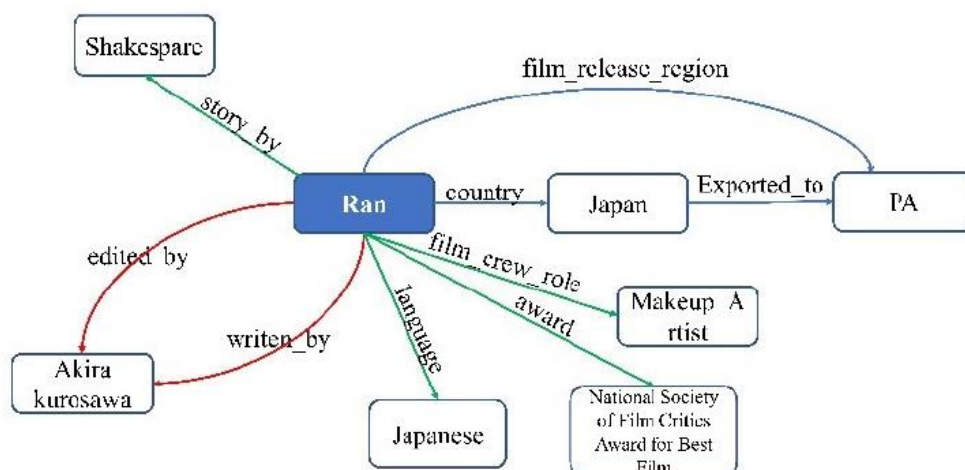


شکل ۵: یک زیر گراف از گراف دانش در نمایش سه تایی است که از روابط واقعی (خطوط آبی) بین موجودیت و روابط استنباط شده (نقطه چین قرمز) تشکیل شده است.

در واقع، درک موجودیت در گراف دانش می تواند از فضای محلی آن شروع شود. در گراف دانش، برخی از ساختارهای محلی حاوی اطلاعات معنایی زیادی هستند، که می توانند به پیش بینی پیوند کمک کنند. شکل ۶، یک زیرنویس از موجودیت "Ran" را در مجموعه داده FB15k-237 نشان می دهد، می توانیم ببینیم که ساختار محلی اطراف موجودیت عمدتاً به سه نوع تقسیم شده است: (۱) از طریق روابط مختلف به یک موجودیت واحد متصل شوید. (پیوندهای قرمز) (۲) از طریق یک مسیر و یک رابطه دیگر همزمان به موجودیت وصل شوید. (پیوندهای آبی) (۳) از طریق روابط مختلف به نهادهای مختلف متصل شوید [۳]. (پیوندهای سبز)

³ Knowledge Graphs (KGs)

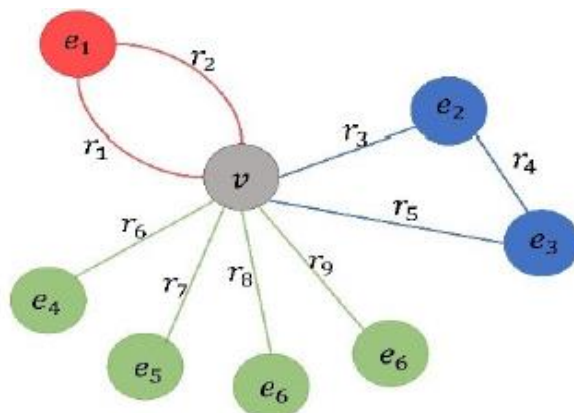
فصل ششم: شبکه های توجه گرافی با آگاهی از ساختار محلی برای تکمیل گراف دانش



شکل ۶: یک زیرگراف از موجودیت "RAN" در مجموعه داده FB15k-237

به همین ترتیب، می توان کل گراف را خلاصه کرد و ساختارهای محلی زیر را تعریف کرد:

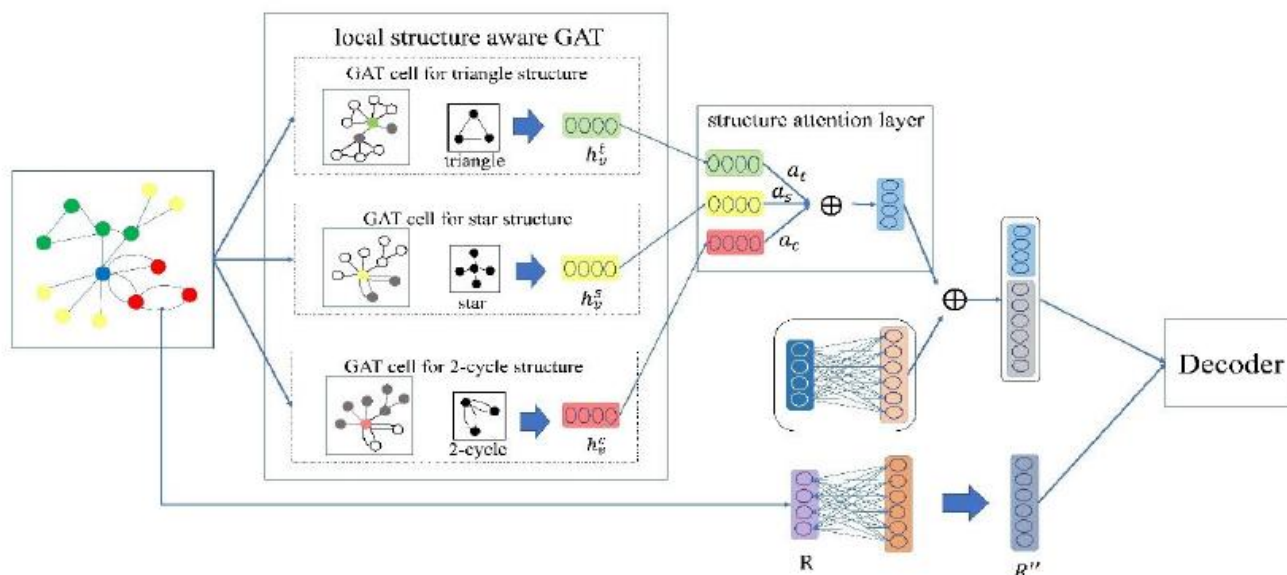
- ۱- 2-Cycle (به رنگ قرمز): موجودیت v و موجودیت دیگر t از طریق دو رابطه متفاوت به هم متصل می شوند. به طور رسمی، دو ترکیب سه گانه $(e1; r1; v)$ و $(e1; r2; v)$ وجود دارد که فقط رابطه در آنها متفاوت است.
- ۲- مثلث (به رنگ آبی): موجودیت v و موجودیت t به ترتیب از طریق یک رابطه و یک مسیر به هم متصل می شوند. به طور رسمی، $(e3; r5; v)$ رابطه بین موجودیت v و موجودیت $e3$ را نشان می دهد، $(e2; r3; v) \wedge (e2; r4; e3)$ نشان دهنده مسیر است.
- ۳- Star (به رنگ سبز): موجودیت v از طریق روابط مختلف با موجودیت های مختلفی پیوند می خورد. به طور رسمی، برای یک موجودیت خاص v ، دو ترکیب سه تایی وجود دارد $(e4; r6; v)$ و $(e5; r7; v)$ و $(e6; r8; v)$ و $(e7; r9; v)$ که فقط موجودیت اصلی یکی است. که در شکل ۷ نیز پیداست.



شکل ۷: یک تصویر از یک مثال

فصل ششم: شبکه های توجه گرافی با آگاهی از ساختار محلی برای تکمیل گراف دانش

این سه ساختار محلی مختلف می توانند اطلاعات معنایی مختلفی را نشان دهند. به عنوان مثال، ساختار 2-cycle نشان می دهد که روابط مختلف متصل به یک موجودیت واحد، معنایی مشابه دارند. برای یادگیری اطلاعات ساختاری، یک مدل GAT آگاه از ساختار محلی به عنوان LSA-GAT پیشنهاد شده است. شکل ۸ نمای کلی معماری این مدل را نشان می دهد.



شکل ۸: معماری مدل LSA-GAT

مدل پیشنهادی، به طور عمده شامل چهار واحد است: (۱) واحد توجه گراف؛ (۲) واحد نمایندگی ساختار محلی؛ (۳) واحد ترکیب ویژگی (۴) رمزگشا.

واحد توجه گراف، اهمیت هر سه همسایه را در اطراف موجودیت مورد نظر اندازه گیری می کند. واحد نمایندگی ساختار محلی فرایندهای مختلفی را بر روی ساختارهای مختلف محلی برای بدست آوردن تعبیه های خاص ساختار انجام می دهد. واحد ترکیب ویژگی، تعبیه های ساختار خاص را برای بدست آوردن تعبیه نهایی برای موجودیت ها ادغام می کند. واحد ۱ و ۲، قسمت رمزگذار مدل را تشکیل می دهد؛ واحد رمزگشایی از تعبیه شده های یادگرفته شده برای محاسبه نمره صحت ترکیب سه تایی، بر اساس یک مدل پیشرفته CNN استفاده می کند [۳].

⁴ local structure-aware GAT 0

۶-۲ آزمایش ها و نتایج

این مدل بر روی دو مجموعه داده استاندارد برای تکمیل گراف دانش ارزیابی شده است، FB15k-237 و WN18RR. جدول ۷، اطلاعات مجموعه داده ها را نشان می دهد:

جدول ۷: اطلاعات مجموعه داده ها

Dataset	#Entities	#Relations	#training triples	#validation triples	#test triples
WN18RR	40943	11	86835	3034	3134
FB15k-237	14541	237	272115	17535	20466

۶-۲-۱ معیار های ارزیابی

هر ترکیب سه تایی آزمون $(t; r; h)$ ، به دو قالب پیش بینی پیوند تبدیل می شود: پیش بینی پیوند سر $(h; r; ?)$ و پیش بینی پیوند انتها $(t; r; ?)$. برای هر قالب، موجودیت صحیح در میان تمام موجودیت های KG به استثنای مجموعه موجودیت واقعی دیگر برای سه گانه مشاهده شده در مجموعه های آموزش و اعتبار سنجی رتبه بندی می شود. در این کار، میانگین رتبه^۴، میانگین رتبه متقابل^۴ موجودیت صحیح، و $Hits @ N$ که دقت در N پیش بینی برتر است، گزارش می شود. در اینجا، N را به عنوان ۱، ۳ و ۱۰ انتخاب می شود [۳].

۶-۲-۲ نتایج پیش بینی پیوند

جدول ۸ نتایج پیش بینی پیوند را در دو مجموعه داده نشان می دهد. در مقایسه با نتایج قبلی، مدل پیشنهادی LSA-GAT، در بیشتر شاخص ها عملکرد بهتری به دست می آورد که نشان می دهد مدل ما توانایی یادگیری نمایش قوی تری را در گراف های دانش دارد. به طور خاص، در مجموعه داده های FB15K-237، LSA-GAT در مقایسه با دومین بهترین نتایج، بهبود ۱۵ درصدی در $Hits @ 1$ ، ۱۱٪ در $Hits @ 3$ ، ۷٪ در $Hits @ 10$ و ۱۲٪ را در MRR کسب می کند. در مجموعه داده های WN18RR، LSA-GAT در $Hits @ 10$ بهبود ۴٪ و ۶۰٪ در MR را به دست می آورد. این نتایج به وضوح ثابت می کند که مدل LSA-GAT، در چهار معیار در مجموعه های FB15K-237 و دو معیار در مجموعه های WN18RR نسبت به نتایج پیشرفته، به طور قابل توجهی پیشی می گیرد. می توانیم ببینیم که در مجموعه داده های FB15k-237 پیشرفت قابل توجهی وجود دارد،

⁴ Mean Rank (MR) 1

⁴ Mean Reciprocal Rank (MRR) 2

فصل ششم: شبکه های توجه گرافی با آگاهی از ساختار محلی برای تکمیل گراف دانش

عمدتاً به این دلیل که مجموعه داده های FB15k-237 دارای ساختار گراف پیچیده تری هستند (انواع بیشتری از رابطه و انواع کمتری از موجودات که به معنی وابستگی های پیچیده تر هستند)، و سپس در آنجا ساختارهای محلی بیشتری در اطراف یک موجودیت هستند. پس این مدل، در تجمیع اطلاعات درون این ساختارهای محلی بهتر است. بهبود مجموعه داده WN18RR به اندازه مجموعه داده FB15k-237 نیست؛ که دلیل این امر این است که ساختار گراف مجموعه داده WN18RR نسبتاً مشابه است، بنابراین ساختار محلی که تعریف شده است، بیشتر بر روی مجموعه داده ثابت می شود (رابطه یا مسیری که در ساختار محلی رخ داده، ثابت است). اگرچه، این مدل هنوز هم در مقایسه با مدل های اصلی جاری موجود در مجموعه داده WN18RR، به ویژه در میانگین رتبه (MR)، به عملکرد رقابتی دست پیدا می کند، این مدل با فاصله زیادی بهترین نتیجه را کسب می کند [۳].

جدول ۸: نتایج به دست آمده بر روی مجموعه آزمون

Dataset	FB15k-237					WN18RR				
	Hits@N			MRR	MR	Hits@N			MRR	MR
	1	3	10			1	3	10		
Distmult(Yang et al., 2014)	0.16	0.26	0.42	0.24	254	0.39	0.44	0.49	0.43	5110
ComplEx (Trouillon et al., 2016)	0.16	0.28	0.43	0.25	339	0.41	0.46	0.51	0.44	5261
ConvE (Dettmers et al. 2017)	0.23	0.35	0.50	0.32	<u>244</u>	0.40	0.44	0.52	0.43	4187
ConvKB (Nguyen et al., 2018)	0.20	0.32	0.52	<u>0.40</u>	257	0.06	0.45	0.53	0.25	<u>2554</u>
R-GCN (Schlichtkrull et al., 2018)	0.15	0.26	0.42	0.25	--	--	--	--	--	--
SACN (Shang et al., 2019)	<u>0.27</u>	<u>0.40</u>	<u>0.55</u>	0.36	--	<u>0.43</u>	0.48	0.54	0.47	--
CompGCN(Shikhar Vashishth et al., 2020)	0.26	0.39	0.53	0.35	197	0.44	0.49	0.54	0.47	3533
LSA-GAT(our proposed method)	0.41	0.50	0.60	0.47	273	0.35	0.49	0.58	<u>0.44</u>	1947

فصل هفتم

شبکه های توجه گرافی مکانی-زمانی: رویکرد یادگیری
عمیق برای پیش بینی ترافیک^{۴۳}

⁴ Spatial-Temporal Graph Attention Networks: A Deep Learning Approach for Traffic Forecasting (ST-GAT)

۱-۷ مکانیسم توجه multi-head

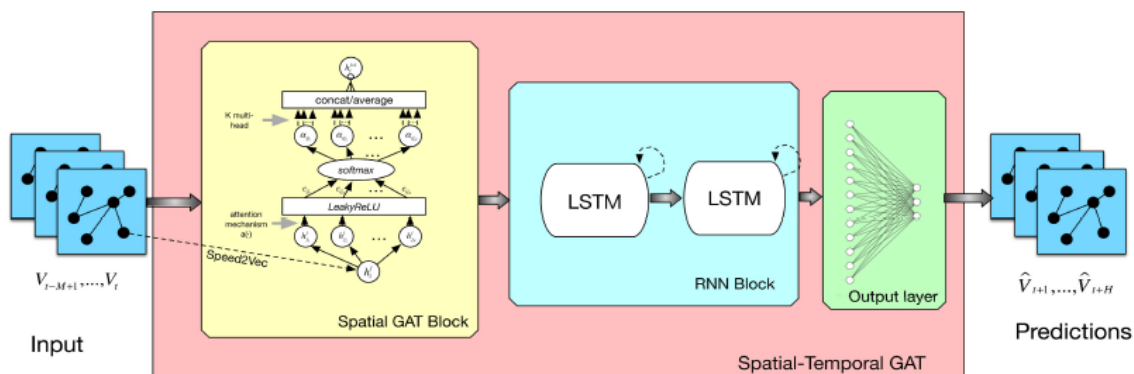
مکانیسم توجه multi-head مدل را قادر می سازد تا ضریب توجه را از طریق چندین زیرفضای نمایشی متعدد بیاموزد. به منظور مقاوم سازی روند یادگیری توجه به خود،^{۴۴} استراتژی های مکانیسم توجه multi-head معمولاً اتخاذ می شوند. خروجی نهایی با میانگین گیری از همسان سازی^{۴۵} نمایش ویژگی به دست می آید [۵]. این فرایند به طور رسمی به صورت زیر تعریف می شود:

$$\begin{cases} h_i^{l+1} = \parallel_{K=1}^K \sigma \left(\sum_{j \in N(i)} \alpha_{ij}^K \phi^K h_j^l \right), & \text{Concatenation} \\ h_i^{l+1} = \sigma \left(\frac{1}{K} \sum_{K=1}^K \sum_{j \in N(i)} \alpha_{ij}^K \phi^K h_j^l \right), & \text{Averaging,} \end{cases} \quad (1-7)$$

۲-۷ معماری شبکه

طی تحقیقات و پژوهش ها، با استفاده از شبکه عصبی، یک پیش بینی کننده سرعت ترافیک ترکیبی به نام ST-GAT پیشنهاد شده است. همانطور که در شکل ۹ نشان داده شده است، ST-GAT شامل یک بلوک GAT مکانی برای استخراج همبستگی مکانی، یک بلوک RNN برای یادگیری ویژگی زمانی و همچنین پیش بینی سری زمانی و یک لایه خروجی برای تولید خروجی توالی است. به طور خاص، در بلوک مکانی GAT، از مکانیسم توجه multi-head استفاده می شود که مدل را قادر می سازد تا به طور مشترک وابستگی های مکانی را از طریق چندین بلوک توجه مستقل یاد بگیرد تا از روند یادگیری بهره مند شود. در بلوک RNN، یک شبکه LSTM ۲ لایه برای استخراج ویژگی سری زمانی به کار گرفته می شود. پیش بینی های نهایی توسط یک شبکه عصبی کاملاً متصل^{۴۶} در لایه خروجی نهایی ایجاد می شود [۵].

⁴ Self-attention	4
⁴ concatenation	5
⁴ Fully connected	6



شکل ۹: معماری شبکه های توجه به نمودار مکانی - زمانی (ST-GAT)

۷-۳ آزمایش ها و نتایج

در این آزمایش از مجموعه داده ی PeMSD7 استفاده شده است. PeMSD7 مجموعه داده ای است که توسط بیش از ۳۹۰۰۰ ایستگاه حسگر در منطقه ۷ کالیفرنیا از سیستم اندازه گیری عملکرد Caltrans، جمع آوری شده است. نمونه های داده از هر فاصله ۳۰ ثانیه ای در بازه های ۵ دقیقه ای جمع می شوند. در نهایت مجموعه داده ای انتخاب شده است که در یک مقیاس متوسط شامل ۲۲۸ ایستگاه PeMSD7 نمونه برداری می شود. مدت زمان مجموعه داده از اول ماه مه تا ۳۰ ژوئن ۲۰۱۲ است که فقط شامل روزهای هفته برای جلوگیری از ترافیک غیرمعمول است. با تنظیم فاصله زمانی جمع آوری اطلاعات به ۵ دقیقه، هر سنسور در شبکه جاده، ۲۸۸ مشاهده در روز تولید می کند. اگر که داده از دست رفته باشد، برای بازیابی داده های از دست رفته از درون یابی خطی استفاده شده است. علاوه بر این، داده ها با روش Z-Score نرمال می شوند. مجموعه های آموزش، اعتبارسنجی و تست به همین ترتیب توسعه یافته اند که هر یک شامل ۶۰٪، ۲۰٪ و ۲۰٪ کل داده ها هستند. ST-GAT بر اساس بهینه ساز آدام برای ۱۵۰ دوره آموزش داده می شود. نرخ یادگیری اولیه $2e^{-4}$ با کاهش وزن $5e^{-4}$ است. اندازه دسته ۵۰ تعیین شده است. برای جلوگیری از بیش برآزش، از حذف تصادفی^۴ و توقف زود هنگام^۸ استفاده می شود. علاوه بر این، از نرمال سازی دسته ای و مقداردهی اولیه پارامتر Xavier [۱۶]. برای ثبات روند یادگیری استفاده شده است و همچنین برای ارزیابی همه مدل های یادگیری، از خطای میانگین

⁴ dropout 7
⁴ earlystopping 8

فصل هفتم: شبکه های توجه گرافی مکانی-زمانی

مربع ریشه (RMSE)، میانگین خطای مطلق (MAE) و درصد میانگین مطلق (MAPE) استفاده شده است [۵].
نتایج در جدول ۹ ذکر شده است:

جدول ۹: مقایسه عملکرد ST-GAT و روش های پایه در مجموعه داده PeMSD7.

Model	15 min			30 min			45 min		
	RMSE	MAE	MAPE (%)	RMSE	MAE	MAPE (%)	RMSE	MAE	MAPE (%)
HA	7.20	4.01	10.61	7.20	4.01	10.61	7.20	4.01	10.61
ARIMA	9.00	5.55	12.92	9.13	5.86	13.94	9.38	6.27	15.20
LSVR	4.55	2.50	5.81	6.67	3.63	8.88	8.28	4.54	11.50
FNN	4.75	2.74	6.38	6.98	4.02	9.72	8.58	5.04	12.38
FC-LSTM	6.20	3.57	8.60	7.03	3.94	9.55	7.51	4.16	10.10
GCGRU	4.21	2.37	5.54	5.96	3.31	8.06	7.13	4.01	9.99
STGCN	4.04	2.25	5.26	5.70	3.03	7.33	6.77	3.57	8.69
ST-GAT	3.45	2.01	4.76	4.68	2.76	6.57	5.30	3.20	7.86

از تحلیل نتایج و مقایسه به مشاهدات زیر می‌رسیم:

۱- ST-GAT به بهترین دقت پیش بینی در مورد تمام معیارها برای سه پنجره پیش بینی، حتی در مقایسه با پیشرفته ترین مدل مبتنی بر کانولوشن گرافی مانند STGCN و GCGRU با سطح قابل توجه ۹۹٪ دست می‌یابد. به طور خاص، معیار MAPE در ST-GAT از STGCN به اندازه ۵٪ (۱۵ دقیقه)، ۷۶٪ (۳۰ دقیقه) و ۸۳٪ (۴۵ دقیقه) بهتر عمل می‌کند. این اثربخشی سازوکار توجه را در یادگیری ویژگی های مکانی مبتنی بر گراف نشان می‌دهد [۵].

۲- از روش های سنتی آماری و یادگیری ماشینی، خصوصاً برای پیش بینی بلند مدت، بسیار بهتر عمل کرده اند. به عنوان مثال، با مقایسه نتایج LSVR و FC-LSTM عملکرد بهتری را در ۱۵ دقیقه پیش بینی به دست می‌آورد. با این حال، در ۴۵ دقیقه پیش بینی، عملکرد آن نسبت به FC LSTM بدتر است. این تا حدی به دلیل ناتوانی آنها در یادآوری توالی طولانی و یادگیری مکانی-مکانی بر روی داده های پیچیده است [۵].

مراجع

- [1] Yang, Liang, et al. "Graph-CAT: Graph Co-Attention Networks via local and global attribute augmentations." *Future Generation Computer Systems* 118 (2021): 170-179.
- [2] Xie, Yu, et al. "Mgat: Multi-view graph attention networks." *Neural Networks* 132 (2020): 180-189.
- [3] Ji, Kexi, Bei Hui, and Guangchun Luo. "Graph Attention Networks With Local Structure Awareness for Knowledge Graph Completion." *IEEE Access* 8 (2020): 224860-224870.
- [4] Liu, Xiyang, et al. "RAGAT: Relation Aware Graph Attention Network for Knowledge Graph Completion." *IEEE Access* 9 (2021): 20840-20849.
- [5] Zhang, Chenhan, J. Q. James, and Yi Liu. "Spatial-temporal graph attention networks: A deep learning approach for traffic forecasting." *IEEE Access* 7 (2019): 166246-166256.
- [6] K.K. Thekumparampil, C. Wang, S. Oh, L. Li, Attention-based graph neural network for semi-supervised learning, CoRR abs/1803.03735 (2018) arXiv: 1803.03735.
- [7] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Liò, Y. Bengio, Graph attention networks, in: 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings, 2018.
- [8] J. Zhang, X. Shi, J. Xie, H. Ma, I. King, D. Yeung, GaAN: Gated attention networks for learning on large and spatiotemporal graphs, in: proceedings of the Thirty-Fourth Conference on Uncertainty in Artificial Intelligence, UAI 2018, Monterey, California, USA, August 6-10, 2018, 2018, pp. 339–349.
- [9] X. Wang, H. Ji, C. Shi, B. Wang, Y. Ye, P. Cui, P.S. Yu, Heterogeneous graph attention network, in: The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019, 2019, pp. 2022–2032, <http://dx.doi.org/10.1145/3308558.3313562>.

- [10] F. Chen, S. Pan, J. Jiang, H. Huo, G. Long, DAGCN: Dual attention graph convolutional networks, in: International Joint Conference on Neural Networks, IJCNN 2019, Budapest, Hungary, July 14-19, 2019, 2019, pp. 1–8, <http://dx.doi.org/10.1109/IJCNN.2019.8851698>.
- [11] Z. Liu, C. Chen, L. Li, J. Zhou, X. Li, L. Song, Y. Qi, GeniePath: Graph neural networks with adaptive receptive paths, in: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019, 2019, pp. 4424–4431.
- [12] Y. Yang, X. Wang, M. Song, J. Yuan, D. Tao, SPAGAN: Shortest path graph attention network, in: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019, 2019, pp. 4099–4105, <http://dx.doi.org/10.24963/ijcai.2019/569>.
- [13] J.B. Lee, R.A. Rossi, X. Kong, S. Kim, E. Koh, A. Rao, Graph convolutional networks with motif-based attention, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM 2019, Beijing, China, November 3-7, 2019, 2019, pp. 499–508.
- [14] K. Zhang, Y. Zhu, J. Wang, J. Zhang, Adaptive structural fingerprints for graph attention networks, in: 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020, 2020.
- [15] H. Pei, B. Wei, K.C. Chang, Y. Lei, B. Yang, Geom-GCN: Geometric graph convolutional networks, in: 8th International Conference on Learning
- [16] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.