شبکۀ توجه گرافي نامتجانس

چکیده:

شبکه‌‌ی عصبی گرافی، به عنوان یک روش بازنمایی گراف قدرتمند که مبتنی بر یادگیری عمیق است، عملکرد فوق‌العاده‌ای از خود نشان داده و علاقه تحقیقاتی قابل توجهی را به خود جلب کرده‌است. با این حال ، در شبکه‌‌ی عصبی گرافی ، گراف نامتجانس که شامل انواع مختلف رئوس و یال‌ها است ، به طور کامل در نظر گرفته‌ نشده ‌است. نامتجانس بودن و اطلاعات معنایی غنی چالش های بزرگی برای طراحی یک شبکه عصبی گرافی برای گراف نامتجانس به همراه دارد. اخیراً ، یکی از هیجان‌انگیزترین پیشرفتها در یادگیری عمیق ، مکانیزم توجه است که پتانسیل بزرگ آن به خوبی در زمینه های مختلف نشان داده شده است. در این مقاله ، ابتدا یک شبکه عصبی گرافی نامتجانس جدید مبتنی بر توجه سلسله مراتبی ، از جمله توجه در سطح رأس و توجه در سطح معنایی را ارائه می دهیم. به طور خاص ، هدف توجه در سطح رأس یادگیری اهمیت بین رأس و همسایگان مبتنی بر فرا-مسیر (meta-path) آن است ، در حالی که توجه سطح معنایی قادر به یادگیری اهمیت فرا-مسیرهای مختلف است. با توجه به اهمیت آموخته شده از توجه در سطح رأس و سطح معنایی ، می توان اهمیت رأس و فرا-مسیر را کاملاً در نظر گرفت. سپس مدل پیشنهادی می‌تواند با جمع آوری ویژگی‌ها از همسایگان مبتنی بر فرا-مسیر به صورت سلسله مراتبی ، تعبیه (embedding) رأس را ایجاد کند. نتایج تجربی گسترده در سه گراف نامتجانس در دنیای واقعی نه تنها عملکرد برتر مدل پیشنهادی ما را نسبت به لبه‌ی دانش نشان می دهد ، بلکه تفسیرپذیری بالقوه خوب آن برای تجزیه و تحلیل گراف را نیز نشان می دهد.

کلمات کلیدی :

شبکه‌ی اجتماعی، شبکه‌ی عصبی، تجزیه و تحلیل گراف

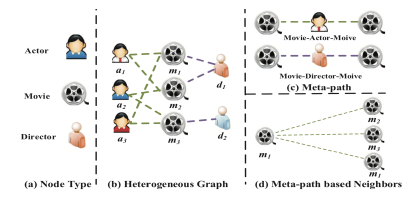
مقدمه:

داده های دنیای واقعی معمولا ساختار گرافی دارند، مانند شبکه های اجتماعی ، شبکه های استنادی و شبکه جهانی وب (وب جهان‌گستر) . شبکه عصبی گرافی (GNN) ، به عنوان یک روش یادگیری بازنمایی عمیق قدرتمند برای چنین داده‌های گرافی ، عملکرد برتری در تجزیه و تحلیل شبکه نشان داده و تحقیقات قابل توجهی را برانگیخته است. یک روند تحقیقی اخیر در یادگیری عمیق ، مکانیزم توجه است که با داده هایی با اندازه متغیر سروکار دارد و مدل را به تمرکز بر برجسته‌ترین قسمت‌های داده تشویق می کند. مکانیزم توجه اثر‌بخشی خود را در چارچوب شبکه عصبی عمیق نشان داده و به طور گسترده ای در کاربردهای مختلف مانند تجزیه و تحلیل متن ، گراف دانش و پردازش تصویر استفاده می شود. شبکه توجه گرافی (GAT) ، یک شبکه عصبی گرافی جدید مبتنی بر روش کانولوشن ، از مکانیزم توجه برای گراف متجانس استفاده می‌کند که فقط شامل یک نوع رأس یا یال است. با وجود موفقیت مکانیزم توجه در یادگیری عمیق ، در چارچوب شبکه عصبی گرافی برای گراف نامتجانس در نظر گرفته نشده است. گراف کاربردهای دنیای واقعی معمولاً دارای انواع مختلفی از رأس‌ها و یال‌ها است که به عنوان شبکه اطلاعات نامتجانس (HIN) نیز شناخته می شوند. برای راحتی ، ما در این مقاله آن را گراف نامتجانس می‌نامیم. از آنجا که گراف نامتجانس شامل اطلاعات جامع‌تر و معنایی غنی است ، به طور گسترده‌ای در بسیاری از کارهای داده‌کاوی استفاده شده است. فرا-مسیر یک رابطه ترکیبی است که دو شی را به هم متصل می کند ، ساختاری است که به طور گسترده ای برای ثبت معناها) (semantics استفاده می شود. برای مثال داده‌های فیلم IMDB که در شکل 1 (a) نشان داده شده است ، شامل سه نوع رأس فیلم ، بازیگر و کارگردان است. رابطه‌ی بین دو فیلم را می توان با فرا-مسیر

فیلم-بازیگر-فیلم (MAM) نشان داد که رابطه بازیگر مشترک بودن در دو فیلم را توصیف می کند ، در حالی که فیلم-کارگردان-فیلم (MDM) به معنای کارگردانی آنها توسط همان کارگردان است. همانطور که مشاهده میشود ، بسته به فرا-مسیرها ، رابطه بین رئوس در گراف نامتجانس می‌تواند معناشناسی متفاوتی داشته باشد.

به دلیل پیچیدگی گراف نامتجانس ، شبکه عصبی گرافی معمولی را نمی توان مستقیماً روی گراف نامتجانس اعمال کرد. بر اساس تجزیه و تحلیل فوق ، هنگام طراحی معماری شبکه عصبی گرافی با مکانیزم توجه برای گراف نامتجانس ، ما الزامات جدید زیر را برطرف می کنیم.

**نامتجانس بودن گراف.** نامتجانس بودن یک ویژگی ذاتی گراف نامتجانس است ، یعنی ، انواع مختلف رئوس و یال‌ها. به عنوان مثال ، انواع مختلف رئوس صفات مختلفی دارند و ممکن است ویژگی‌های آنها در فضای مختلف ویژگی‌ها قرار داشته باشد.



شکل 1: مثالی گویا از گراف نامتجانس (IMDB). (a) سه نوع رأس (یعنی ، بازیگر ، فیلم ، کارگردان).

(b) گراف نامتجانس IMDB شامل سه نوع رأس و دو نوع اتصال است. (c)دو فرا-مسیر درگیر

در IMDB (به عنوان مثال فیلم-بازیگر-فیلم و فیلم-کارگردان-فیلم). (d) فیلم

𝑚1 و همسایگان متا مسیر آن (به عنوان مثال ، 𝑚1 ، 𝑚2 و 𝑚3).

هنوز IMDB را به عنوان مثال در نظر می گیریم ، ویژگی بازیگر ممکن است شامل جنس ، سن و ملیت باشد. از طرف دیگر ، ویژگی فیلم ممکن است شامل طرح (plot) و بازیگران باشد. چگونگی کنترل چنین اطلاعات پیچیده ساختاری و حفظ اطلاعات متنوع ویژگی‌ها به صورت همزمان، یک مشکل ضروری است که باید حل شود.

**توجه در سطح معنایی.** اطلاعات معنایی معنی‌دار و پیچیده مختلفی در گراف نامتجانس دخیل هستند که معمولاً توسط فرا-مسیرها منعکس می شوند. فرا-مسیرهای مختلف در گراف نامتجانس ممکن است اطلاعات معنایی متنوعی را استخراج کند. نحوه انتخاب معنادارترین فرا-مسیرها و ترکیب اطلاعات معنایی برای کار خاص ، مسئله‌ای باز است. هدف توجه در سطح معنایی آن است که اهمیت هر فرا-مسیر را بیاموزد و وزن‌های مناسبی را به آن‌ها اختصاص دهد. هنوز IMDB را به عنوان مثال در نظر می گیریم ، The Terminator یا می تواند از طریق فیلم-بازیگر-فیلم (MAM) به The Terminator 2 متصل شود (در هر دو شوارتزنگر بازیگر است) یا از طریق فیلم-سال ساخت-فیلم (MYM) به Birdy متصل شود (هر دو در سال 1984 فیلمبرداری شده اند). با این حال ، هنگام شناسایی ژانر فیلم The Terminator ، MAM معمولاً نقش مهم‌تری دارد تا MYM. بنابراین ، برخورد یکسان با فرا-مسیرهای مختلف سودمند نیست و اطلاعات معنایی ارائه شده توسط برخی از متا مسیرهای مفید را ضعیف می کند.

**توجه در سطح رأس.** در یک گراف نامتجانس ، رئوس را می‌توانند از طریق انواع مختلف رابطه ، به عنوان مثال ، فرا-مسیر اتصال داشته‌باشند. با توجه به یک فرا-مسیر ، هر رأس همسایگان مبتنی بر فرا-مسیر زیادی دارد. نحوه تشخیص تفاوت ظریف همسایگان آنها و انتخاب برخی همسایگان حاوی اطلاعات مفید الزامی است. برای هر رأس هدف توجه در سطح رأس ، یادگیری اهمیت همسایگان مبتنی بر فرا-مسیر و اختصاص دادن مقادیر مختلف توجه به آنها است. باز IMDB را به عنوان مثال در نظر می گیریم ، هنگام استفاده از فرا-مسیر فیلم-کارگردان-فیلم (Movie-Director-Movie) (فیلم ها با یک کارگردان هستند) ، The Terminator از طریق کارگردان James Cameron به Titanic و The Terminator 2 متصل می شود. برای شناسایی بهتر ژانر The Terminator به عنوان فیلم علمی-تخیلی ، مدل باید بیش از Titanic به The Terminator 2 توجه کند.

بنابراین ، نحوه طراحی مدلی که بتواند تفاوتهای ظریف همسایه‌ها را کشف کند و وزن آنها را به درستی یاد بگیرد ، مطلوب خواهد بود. در این مقاله ، ما یک شبکۀ توجه گرافي نامتجانس جدید

به نام HAN را پیشنهاد می کنیم که توجهات در سطح رأس و سطح معنایی را در نظر می گیرد. به طور خاص ، با توجه به ویژگی های رأس به عنوان ورودی ، ما برای پیش‌بینی انواع مختلف ویژگی‌های رأس در یک فضا از ماتریس تحول نوع خاص (type-specific transformation matrix) استفاده می کنیم. سپس توجه در سطح گره قادر به یادگیری مقادیر توجه بین رئوس و همسایگان مبتنی بر فرا-مسیر آنها است ، در حالی که توجه در سطح معنایی برای یادگیری مقادیر توجه فرا-مسیرهای مختلف برای کار مشخص در گراف نامتجانس است. بر اساس مقادیر توجه آموخته شده از نظر دو سطح ، مدل ما می‌تواند ترکیبی بهینه از همسایگان و چندین فرا-مسیر را به صورت سلسله مراتبی بدست آورد ، که رئوس تعبیه (embedding) یاد گرفته شده را قادر می‌سازد تا ساختار پیچیده و اطلاعات معنایی غنی را در گراف نامتجانس بهتر به دست آورند. پس از آن ، می‌توان مدل کلی را از طریق backpropagation به روش انتها به انتها (end-to-end) بهینه کرد.

سهم‌های کار ما به شرح زیر است:

* تا جایی که ما می‌دانیم، این اولین تلاش برای مطالعه شبکه عصبی گراف نامتجانس بر اساس مکانیزم توجه است. کار ما این امکان را فراهم می‌کند که شبکه عصبی گرافی به طور مستقیم بر روی گراف نامتجانس اعمال شود ، و کاربرد‌های مبتنی بر گراف نامتجانس را بیشتر تسهیل می کند.
* ما یک شبکۀ توجه گرافي نامتجانس جدید (HAN) را ارائه می‌کنیم که هر دوی توجه در سطح رأس و توجه در سطح معنایی شامل می‌شود. با بهره گیری از چنین توجهات سلسله مراتبی ، HAN پیشنهادی می‌تواند اهمیت رئوس و فرا-مسیرها را به طور همزمان در نظر بگیرد. علاوه بر این ، مدل ما با پیچیدگی خطی با توجه به تعداد جفت رأس‌های مبتنی بر فرا-مسیر ، دارای کارایی بالا است ، که می‌تواند برای گراف نامتجانس در مقیاس بزرگ اعمال شود.
* ما آزمایش‌های گسترده‌ای را برای ارزیابی عملکرد مدل پیشنهادی انجام می دهیم. نتایج برتری مدل پیشنهادی را در مقایسه با مدل‌های لبه‌ی دانش ، نشان می‌دهد. از همه مهمتر ، با تجزیه و تحلیل مکانیزم توجه سلسله مراتبی ، HAN پیشنهادی تفسیر‌پذیری بالقوه خوب خود را برای تجزیه و تحلیل گراف نامتجانس نشان می دهد.

**مقدمات**

گراف نامتجانس نوعی خاص از شبکه اطلاعاتی است که شامل انواع مختلفی از اشیا یا انواع مختلفی از یال‌ها است.

تعریف گراف نامتجانس: یک گراف نامتجانس ، که با G = (V, E) مشخص می‌شود ، از یک مجموعه شی V و یک مجموعه یال E تشکیل شده است. یک گراف نامتجانس همچنین با یک تابع نگاشت نوع رأس: 𝜙 : V → A و یک تابع نگاشت نوع یال 𝜓: E → R مرتبط است. A و R مجموعه‌ای از انواع شی و انواع یال از پیش تعریف شده را نشان می دهد ، بطوری‌ که |A| + |R| > 2.

**مثال**. همانطور که در شکل (a) 1 نشان داده شده است ، ما یک گراف نامتجانس برای مدل کردن IMDB می‌سازیم که شامل انواع مختلفی از اشیا ((بازیگر (A) ، فیلم (M) ، کارگردان (D)) و روابط (رابطه ساخت فیلم بین فیلم ها و کارگردانان ، رابطه نقش آفرینی بین بازیگران و فیلم ها) است.

در گراف نامتجانس ، دو شی از طریق مسیرهای معنایی مختلف می‌توانند متصل باشند که به آنها فرا-مسیر می‌گویند.

تعریف فرا-مسیر : فرا-مسیر Φ به عنوان یک مسیر به شکل A1 R1 A2 R2 … Rl Al+1 (به اختصار 𝐴1𝐴2 · · · 𝐴𝑙 + 1) ، که یک رابطه ترکیبی 𝑅= 𝑅1 ◦ 𝑅2 ◦ · · · ◦ 𝑅𝑙 بین اشیا 𝐴1 و 𝐴𝑙 + 1 را توصیف می کند ، جایی که ◦ عملگر ترکیب را بر روابط نشان می دهد.

**مثال**. همانطور که در شکل (a) 1 نشان داده شده است ، می توان دو فیلم را از طریق چندین فرا-مسیر به هم متصل کرد ، به عنوان مثال فیلم-بازیگر-فیلم (MAM) و فیلم-کارگردان-فیلم (MDM). فرا-مسیرهای مختلف همیشه معانی مختلفی را نشان می‌دهند. به عنوان مثال ، MAM به معنای بازیگر مشترک بودن در دو فیلم است ، در حالی که فیلم-کارگردان-فیلم (MDM) بدان معنی است که توسط همان کارگردان کارگردانی می شوند.

با توجه به فرا-مسیر Φ ، مجموعه‌ای از همسایگان مبتنی بر فرا-مسیر هر رأس وجود دارد که می‌تواند اطلاعات مختلف ساختار و معانی غنی را در یک گراف نامتجانس نشان دهد.

**تعریف همسایگان مبتنی بر فرا-مسیر**. با توجه به رأس 𝑖 و فرا-مسیر Φ در گراف نامتجانس ، همسایگان مبتنی بر فرا-مسیر رأس) i ( N𝑖Φ به عنوان مجموعه‌ای از رئوس تعریف می شوند که از طریق فرا-مسیر Φ با رأس i متصل می شوند. توجه داشته باشید که همسایگان رأس خود رأس را نیز شامل می شود.

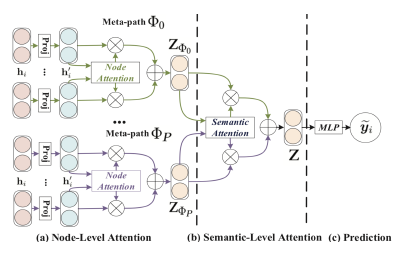
**مثال.** با استفاده از شکل (d) 1 به عنوان مثال ، با توجه به فرا-مسیر فیلم-بازیگر-فیلم، همسایگان مبتنی بر فرا-مسیر 𝑚1 شامل (خود) 𝑚1 ، 𝑚2 و 𝑚3 هست. به طور مشابه، همسایگان 𝑚1 بر اساس فرا-مسیر فیلم-کارگردان-فیلم شامل 𝑚1 و 𝑚2 است. بدیهی است که همسایگان مبتنی بر فرا-مسیر می‌توانند از جنبه های مختلف اطلاعات ساختاری در گراف نامتجانس بهره ببرند. ما می‌توانیم که همسایگان مبتنی بر فرا-مسیر را با ضرب دنباله‌ای از ماتریس‌های مجاورت بدست آوریم. شبکه عصبی گرافی برای کار کردن با داده‌هایی با ساختار گرافی دلخواه ارائه شده است. با این حال ، همه‌ی آنها برای شبکه متجانس طراحی شده اند. از آنجا که فرا- مسیر و همسایگان مبتنی بر فرا-مسیر دو ساختار اساسی در یک گراف نامتجانس هستند ، در مرحله بعدی ، ما یک شبکه عصبی گرافی جدید برای گراف داده‌ی نامتجانس ارائه خواهیم داد ، که قادر است از تفاوت ظریف رئوس و فرا-مسیرها بهره ببرد. علائمی که در مقاله استفاده خواهیم کرد در جدول 1 خلاصه شده است.

جدول 1

|  |  |  |
| --- | --- | --- |
| Notation/ علامت | Explanation | توضیحات |
| Φ | Meta-path | فرا-مسیر |
| h | Initial node feature | ویژگی اولیه رأس |
| MΦ | Type-specific transformation matrix | ماتریس تبدیل وابسته به نوع |
| h′ | Projected node feature | ویژگی پیش بینی شده رأس |
| 𝑒Φ𝑖𝑗 | Importance of meta-path based node pair (𝑖,𝑗 ) | اهمیت زوج رأس مبتنی بر فرا-مسیر (𝑖,𝑗) |
| aΦ | Node-level attention vector for meta-path Φ | بردار توجه در سطح رأس برای فرا-مسیر Φ |
| αΦ𝑖𝑗 | Weight of meta-path based node pair (𝑖,𝑗 ) | وزن زوج رأس مبتنی بر فرا-مسیر (𝑖,𝑗) |
| NΦ | Meta-path based neighbors | همسایگان مبتنی بر فرا-مسیر |
| ZΦ | Semantic-specific node embedding | تعبیه رأس وابسته به معنا |
| q | Semantic-level attention vector | بردار توجه در سطح معنایی |
| wΦ | Importance of meta-path Φ | اهمیت فرا-مسیر (𝑖,𝑗) |
| 𝛽Φ | Weight of meta-path Φ | وزن فرا-مسیر (𝑖,𝑗) |
| Z | The final embedding | تعبیه نهایی |

**مدل پیشنهادی**

در این بخش ، ما یک شبکه عصبی گرافی نیمه نظارت شده (semi-supervised) جدید را برای گراف نامتجانس پیشنهاد می کنیم. مدل ما از یک ساختار توجه سلسله مراتبی پیروی می‌کند: توجه در سطح رأس ← توجه در سطح معنایی. شکل 2 چارچوب کلی HAN را نشان می دهد.



شکل 2: چارچوب کلی HAN پیشنهادی. (a) انواع رئوس به یک فضای مشخصه واحد منتقل

می شوند و وزن جفت رأس مبتنی بر فرا-مسیر را می‌توان از طریق توجه در سطح رأس یاد

گرفت. (b) یادگیری مشترک وزن هر فرا-مسیر و ترکیب کردن تعبیه رأس خاص معنایی از طریق

توجه در سطح معنایی. (c) محاسبه ضرر و بهینه سازی انتها به انتها برای HAN پیشنهادی

در ابتدا ، ما توجه در سطح رأس را برای یادگیری وزن همسایگان مبتنی بر فرا-مسیر و جمع‌آوری آنها برای بدست آوردن تعبیه رأس خاص معنایی را پیشنهاد می‌کنیم. پس از آن ، HAN می‌تواند از طریق توجه در سطح معنایی ، تفاوت فرا-مسیرها را تشخیص دهد و از ترکیب بهینه وزن‌دار تعبیه رأس خاص معنایی برای کار مشخص ، استفاده کند.

**توجه در سطح رأس**

قبل از جمع آوری اطلاعات از همسایگان فرا-مسیر برای هر رأس ، باید توجه کنیم که همسایگان مبتنی بر فرا-مسیر هر رأس نقش متفاوتی دارند و اهمیت مختلفی را در یادگیری تعبیه رأس برای کار خاص نشان می‌دهند. توجه در سطح رأس می‌تواند اهمیت همسایگان مبتنی بر فرا-مسیر را برای هر رأس در یک گراف نامتجانس بیاموزد و بازنمایی این همسایگان معنی‌دار را برای تشکیل تعبیه رأس جمع کند.

به دلیل نا متجانس بودن رئوس، رئوس متفاوت دارای فضای ویژگی متفاوتی هستند. بنابراین برای هر نوع رأس ( مثلا رأس با نوع 𝜙𝑖) ماتریس تبدیل وابسته به نوع M𝜙𝑖 را طراحی می‌کنیم تا ویژگی رأس‌های متفاوت را به یک فضای ویژگی یکسان تصویر کند. فرایند تصویر کردن می‌تواند به صورت زیر باشد:



که h𝑖 و h′𝑖 به ترتیب ویژگی اصلی و تصویر شده رأس i هستند. با عمل تصویرکردن وابسته به نوع ، توجه در سطح رأس می تواند رئوس با نوع دلخواه را کنترل کند. پس از آن ، ما از توجه به خود slef attention استفاده می‌کنیم تا وزن بین انواع رئوس را یاد بگیریم. اگر جفت رأس (𝑖, 𝑗 ) که از طریق فرا-مسیر Φ به هم وصل هستند را داشته‌باشیم ، توجه در سطح رأس 𝑒Φ𝑖𝑗 می تواند اهمیت 𝑒Φ𝑖𝑗 را بیاموزد ، این به این معنی است که رأس j برای رأس i چقدر مهم خواهد بود. اهمیت جفت رأس (𝑖, 𝑗 ) مبتنی بر فرا-مسیر را می‌توان به صورت زیر فرموله کرد:



در اینجا 𝑎𝑡𝑡𝑛𝑜𝑑𝑒 نشان دهنده شبکه‌ی عصبی عمیقی است که توجه در سطح رأس را انجام می دهد. با توجه به فرا-مسیر Φ ، 𝑎𝑡𝑡𝑛𝑜𝑑𝑒 برای همه‌ی جفت رأس‌های مبتنی بر فرا-مسیر به اشتراک گذاشته می‌شود.

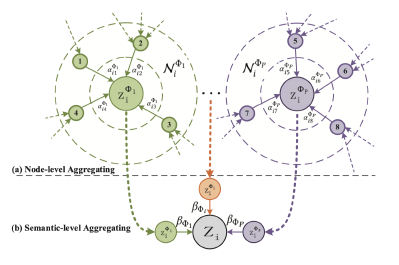
این به این دلیل است که برخی الگوهای اتصال مشابه در زیر یک فرا-مسیر وجود دارد. معادله فوق نشان می‌دهد که با توجه به فرا-مسیر Φ ، وزن جفت رأس‌ (𝑖, 𝑗 ) مبتنی بر فرا-مسیر به ویژگی‌های آنها بستگی دارد. لطفا توجه داشته باشید که ، 𝑒Φ𝑖𝑗 نامتقارن است ، یعنی ، اهمیت رأس‌ 𝑖 برای رأس‌ 𝑗 و اهمیت رأس‌ 𝑗 برای رأس‌ 𝑖 می تواند کاملاً متفاوت باشد. این نشان میدهد توجه در سطح رأس‌ می‌تواند نامتقارن بودن که از ویژگی‌های مهم گراف نامتجانس است را حفظ کند. سپس اطلاعات ساختاری را از طریق توجه پوشیده masked attention به مدل تزریق می کنیم ، بدین معنی که ما فقط eΦ𝑖j را برای رئوس 𝑗 ∈ NΦ𝑖 محاسبه می‌کنیم ، که NΦ𝑖 نشان‌دهنده همسایگان مبتنی بر فرا-مسیر رأس 𝑖 است (شامل خود 𝑖). پس از به دست آوردن اهمیت بین جفت رأس‌های مبتنی بر فرا-مسیر ، آنها را نرمال می کنیم تا ضریب وزن 𝛼Φ𝑖𝑗 را از طریق تابع softmax بدست آوریم:



که در آن 𝜎 نشانگر تابع فعال سازی است ، ∥ نشان دهنده عملیات اتصال می باشد و aΦ بردار توجه در سطح گره برای فرا-مسیر Φ است. همانطور که از معادله می بینیم. (3) ، ضریب وزن (𝑖, 𝑗 ) به ویژگی های آنها بستگی دارد. همچنین توجه داشته باشید که ضریب وزن 𝛼Φ𝑖𝑗 نامتقارن است به این معنی که آنها کمک متفاوتی به یکدیگر می‌کنند. نه تنها به دلیل ترتیب اتصال در صورت ، بلکه به این دلیل که آنها همسایگان مختلفی دارند بنابراین ترم نرمال (مخرج) کاملاً متفاوت خواهد بود. سپس ، تعبیه مبتنی بر فرا-مسیر رأس 𝑖 می‌تواند توسط ویژگی‌های تصویر شده همسایه با ضرایب متناظر به شرح زیر جمع شود:



که در آن zΦ𝑖 تعبیه یاد گرفته شده‌ی رأس i برای فرا-مسیر Φ است.



شکل 3: توضیح فرآیند جمع کردن (انباشتن) در سطح گره و معنایی.

هر تعبیه رأس با همسایگانش جمع می‌شود. از آنجایی که وزن توجه 𝛼Φ𝑖𝑗 برای یک فرا-مسیر واحد تولید می‌شود ، خاص معنایی است و می تواند نوعی از اطلاعات معنایی را ضبط کند. از آنجا که گراف نامتجانس خاصیت بدون مقیاس بودن را دارد ، واریانس (پراکندگی) داده‌های گراف کاملاً زیاد است. برای مقابله با مشکل فوق ، ما توجه در سطح رأس را به توجه چند سره (multihead) گسترش می‌دهیم تا روند آموزش با ثبات‌تر باشد. به طور خاص ، ما توجه در سطح رأس را برای 𝐾 بار تکرار می‌کنیم و تعبیه‌های آموخته شده را به عنوان تعبیه ویژه معنایی به هم متصل می‌کنیم:



با توجه به مجموعه فرا-مسیر {Φ1, . . . , Φ𝑃 } ، پس از دادن ویژگی‌های رأس به توجه در سطح رأس ، می‌توانیم 𝑃 گروه از تعبیه رئوس خاص معنایی را بدست آوریم ، به عنوان {ZΦ1, . . . , ZΦ𝑃 } نشان داده شده است.

**توجه در سطح معنایی**

به طور کلی ، هر رأس در یک گراف نامتجانس شامل انواع مختلفی از اطلاعات معنایی است و تعبیه رأس خاص معنایی فقط می تواند رأس را از یک جنبه منعکس کند. برای یادگیری یک تعبیه رأس جامع‌تر ، باید چندین معنا را که می تواند توسط فرا-مسیرها مشخص شود ، ادغام کنیم. برای پرداختن به چالش انتخاب فرا-مسیر و ادغام معنایی در یک گراف نامتجانس ، ما توجه جدیدی در سطح معنایی را پیشنهاد می‌کنیم تا به طور خودکار اهمیت فرا-مسیرهای مختلف را یاد بگیریم و آنها را برای کار خاص ادغام کنیم. با در نظر گرفتن 𝑃 گروه تعبیه رأس‌های خاص معنایی از توجه در سطح رأس به عنوان ورودی ، وزن های یاد گرفته شده برای هر فرا-مسیر (𝛽Φ1 ،.. ، 𝛽Φ𝑃) را می توان به شرح زیر نشان داد:



در اینجا 𝑎𝑡𝑡𝑠𝑒𝑚 نشان دهنده شبکه عصبی عمیق است که توجه درسطح معنایی را انجام می دهد. این نشان می دهد که توجه در سطح معنایی می تواند انواع مختلفی از اطلاعات معنایی را در پشت یک گراف نامتجانس ضبط کند. برای یادگیری اهمیت هر فرا-مسیر ، ما ابتدا تعبیه خاص معنایی را از طریق یک تبدیل غیرخطی (به عنوان مثال ، MLP یک لایه) تبدیل می‌کنیم. سپس ما اهمیت تعبیه خاص معنایی را به عنوان شباهت تعبیه تبدیل شده با یک بردار توجه در سطح معنایی q اندازه گیری می‌کنیم. علاوه بر این ، ما از اهمیت کلیه تعبیه رئوس معنایی خاص را که می‌تواند به عنوان اهمیت هر یک از فرا-مسیرها توضیح داده شود ، میانگین می‌گیریم.

اهمیت هر فرا-مسیر، که به عنوان 𝑤Φ𝑖 مشخص می شود ، به شرح زیر نشان داده شده است:



که W ماتریس وزن است ، b بردار بایاس است ، q بردار توجه در سطح معنایی است. توجه داشته باشید که برای مقایسه معنی‌دار ، تمام پارامترهای فوق برای تمام فرا-مسیرها و تعبیه خاص معنا به اشتراک گذاشته می‌شوند. پس از به دست آوردن اهمیت هر فرا-مسیر ، آنها را از با تابع softmax نرمال می‌کنیم. وزن فرا-مسیر Φ𝑖 ، که به عنوان 𝛽Φ𝑖 مشخص می شود ، می تواند با نرمال سازی اهمیت تمام فرا-مسیرهای بالا با استفاده از تابع softmax بدست آید ،



که می تواند به عنوان سهم فرا-مسیر Φ𝑝 برای یک کار خاص تفسیر شود. بدیهی است که هرچه 𝛽Φ𝑝 بیشتر باشد ، فرا-مسیر Φ𝑝 مهمتر است. توجه داشته باشید که برای کارهای مختلف ، فرا-مسیر Φ𝑝 ممکن است وزن های مختلفی داشته باشد. با وزن های آموخته شده به عنوان ضرایب ، می توانیم این تعبیه های خاص معنایی را برای بدست آوردن تعبیه نهایی Z به صورت زیر ادغام کنیم:



برای درک بهتر روند تجمیع در سطح معنایی ، ما همچنین توضیح مختصری در شکل 3 (ب) ارائه می دهیم. تعبیه نهایی توسط جمع تمام تعبیه‌های خاص معنایی می شود. سپس می‌توانیم تعبیه نهایی را برای کارهای خاص اعمال کنیم و توابع ضرر مختلفی را طراحی کنیم. برای طبقه بندی رأس نیمه نظارت شده ، ما می توانیم Cross-Entropy را روی تمام رئوس برچسب خورده بین مقدار واقعی و پیش بینی به حداقل برسانیم:



که در آن C پارامتر طبقه بند است ، Y𝐿 مجموعه اندیس‌های گره‌ای است که دارای برچسب هستند ، Y𝑙 و Z𝑙 برچسب ها و تعبیه‌های گره‌های دارای برچسب هستند. با راهنمای داده های دارای برچسب ، ما می توانیم مدل پیشنهادی را از طریق backpropagation بهینه کرده و تعبیه گره‌ها را یاد بگیریم. روند کلی HAN در الگوریتم 1 نشان داده شده است.

