شبکه‌های گرافی توجه: گزارش فاز پیاده‌سازی

*محمد ابراهیمی، پارسا عباسی*

*دانشگاه علم و صنعت ایران، تهران* [*m\_ebrahimi74@comp.iust.ac.ir*](mailto:m_ebrahimi74@comp.iust.ac.ir)*,* [*parsa\_abbasi@comp.iust.ac.ir*](mailto:parsa_abbasi@comp.iust.ac.ir)

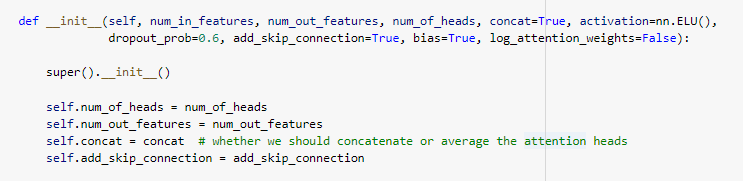
**1) لایۀ GAT**

ورودی‌های این لایه به شرح زیر هستند:

|  |  |
| --- | --- |
| **نام متغیر** | **مشخصات** |
| num\_in\_features | تعداد ویژگی‌های ورودی لایه |
| num\_out\_features | تعداد ویژگی‌های خروجی لایه |
| num\_of\_heads | تعداد هد |
| concat | مشخص کنندۀ اینکه خروجی هر هد باید با یکدیگر الحاق شود (True) یا از آن‌ها میانگین گرفته شود (False) |
| activation | تابع فعالسازی (پیش فرض برابر elu) |
| dropout\_prob | احتمال حذف تصادفی |
| add\_skip\_connection | استفاده از تکنیک skip connection |
| bias | اضافه کردن بایاس |
| log\_attention\_weights | تعیین کنندۀ لاگ گرفتن از وزن‌های توجه |

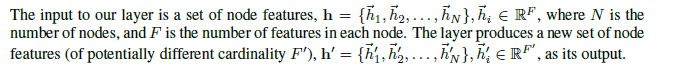
ماتریس خطی (معادل با در مقاله)، وزن‌های توجه (معادل با در مقاله) و بایاس (در مقاله بدان اشاره‌ای نشده اما در کد اصلی آن پیاده‌سازی شده است)، وزن‌های قابل آموزش این مدل هستند. از آنجا که چندین هد می‌تواند مورد استفاده قرار گیرد بجای اینکه برای هرکدام ماتریس وزن جداگانه تعریف کنیم، می‌توان به شکل زیر همه را در یک لایۀ مشترک جای داد.

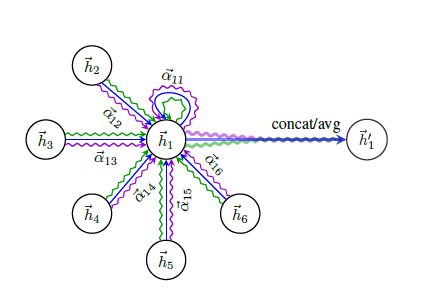
self.linear\_proj = nn.Linear(num\_in\_features, num\_of\_heads \* num\_out\_features, bias=False)



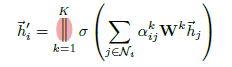
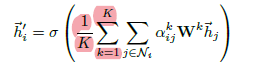
num\_in\_features: تعداد ویژگی‌های ورودی

num\_out\_features: تعداد ویژگی‌های خروجی



num\_of\_heads: تعداد هدها که برای stable یا پایدار کردن فرایند یادگیری از چند هد ممکن است استفاده کنیم. در شکل سه هد داریم که با سه رنگ بنفش، آبی و سبز نشان داده شده‌اند.

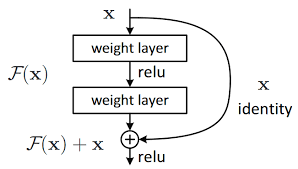
concat: این متغیر این را مشخص می‌کند که خروجی هدها را با هم concat کنیم یا از آن‌ها میانگین بگیریم



OR

activation: تابع فعالسازی که در این‌جا از تابع elu (exponential relu) استفاده شده است.

dropout\_prob: نرخ حذف تصادفی

add\_skip\_connection: استفاده از تکنیک connection skip :

bias: اضافه کردن bias که استفاده از آن در GAT ضروری نیست.

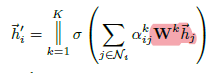
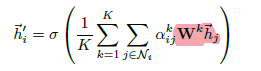
log\_attention\_weights: تعیین کننده اینکه از وزن‌های توجه لگاریتم بگیریم.

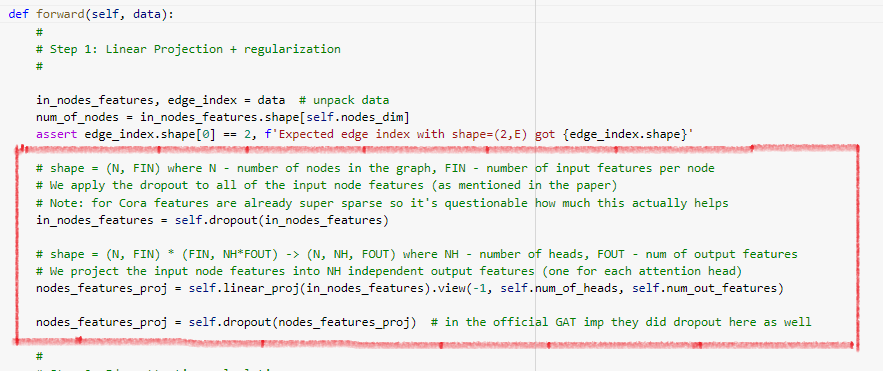
**گام اول: تبدیل خطی و regularization**



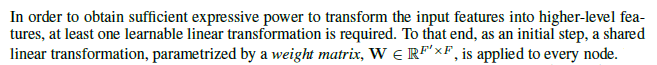
برای به دست آوردن ویژگی‌های high levelتر از روی ویژگی‌های ورودی روی آن‌ها حداقل یک تبدیل خطی اعمال می‌کنیم.

به تعداد هدها ماتریس وزن مستقل اعمال می‌شود یعنی به تعداد num\_of\_heads ماتریس وزن داریم.

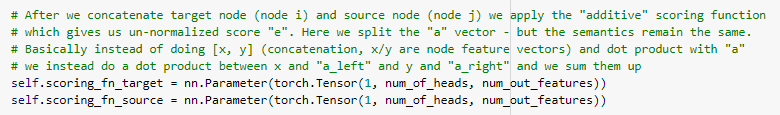
 

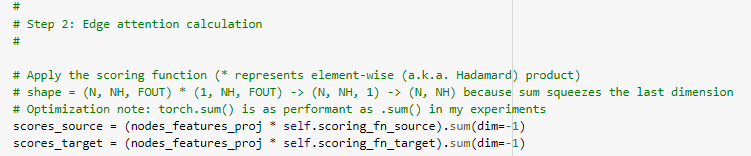


در قسمت فوق تبدیل خطی اولیه که در بالا به آن اشاره شد اعمال شده و قبل و بعد از آن روی گره‌ها dropout اعمال شده که در مقاله به dropout اشاره نشده است اما در کد اصلی مقاله هر دو dropout اشاره شده اعمال شده اند.

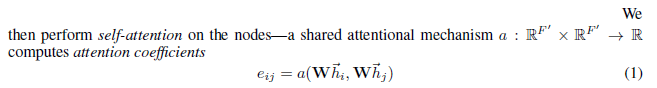


**گام دوم: محاسبه‌ی توجه**

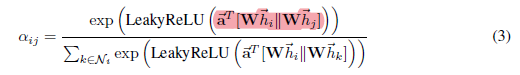




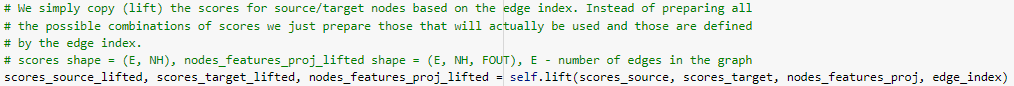
در این قست به محاسبه ضرایب توجه نرمال نشده یعنی eij در مقاله می‌پردازیم.



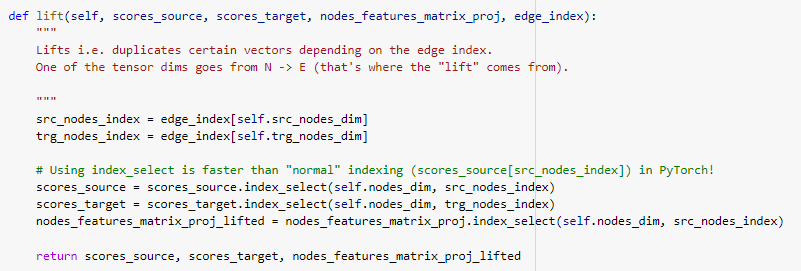
مقدار eij  برابر قسمت هایلایت شده در فرمول زیر است:



در کد پیاده‌سازی شده به جای اینکه بردار a را در الحاق دو بردار Whi و Whj ضرب داخلی کنیم، a را به دونیمه‌ی چپ و راست تقسیم می‌کنیم و سپس نیمه‌ی چپ را در Whi و نیمه راست را در Whj ضرب داخلی می‌کنیم و سپس نتایج حاصل را با هم جمع می‌کنیم.



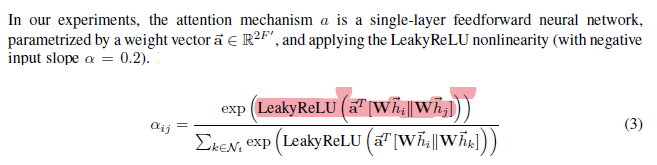
تابع lift که در کد فوق استفاده شده و پیاده‌سازی آن به شرح زیر است برای محاسبه‌ی ضرب‌های داخلی (a\_left . Whi) و (a\_right . Whj) با توجه به یالهای بین i و j است.

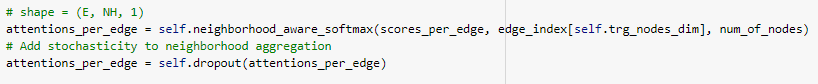




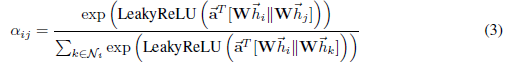


سپس روی حاصل جمع دو عبارت (a\_left . Whi) و (a\_right . Whj) تابع leakyReLU اعمال شده است.

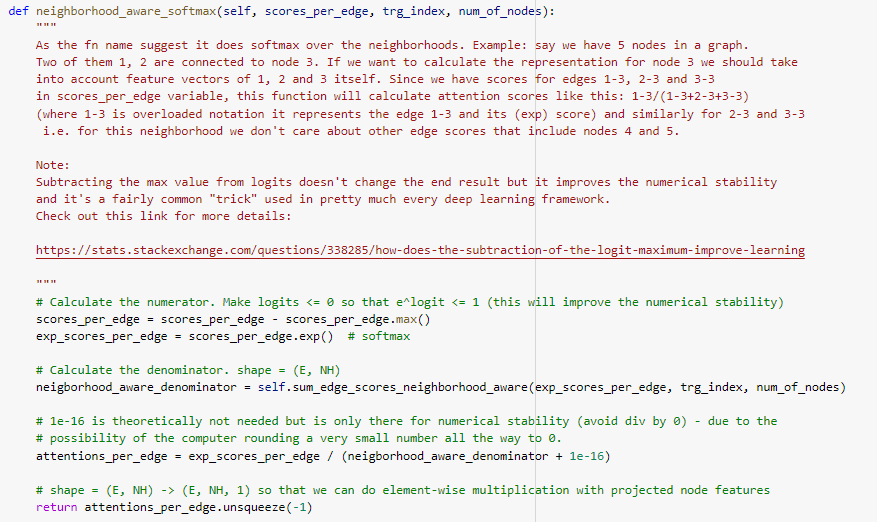




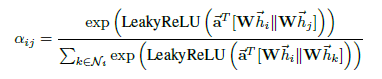
در نهایت با اعمال تابع softmax روی هر گره با توجه به همسایگانش ( که خود گره را نیز شامل می‌شود) ضرایب توجه نرمال شده را به دست می‌آوریم:



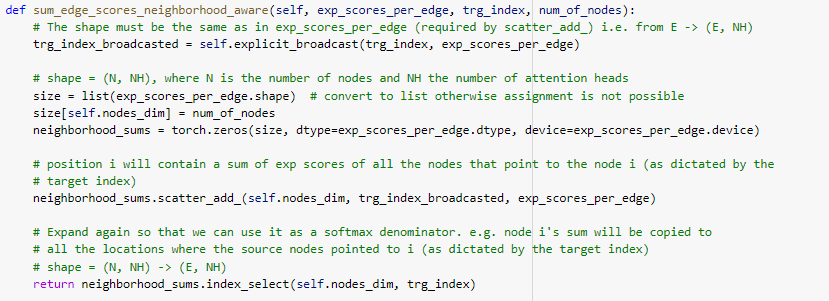
پس از به دست آوردن softmax روی خروجی آن که در گام بعدی استفاده می‌شود dropout اعمال شده است.



تابع فوق مربوط به اعمال softmax روی ضرائب توجه و به دست آوردن ضرائب توجه نرمال شده است.

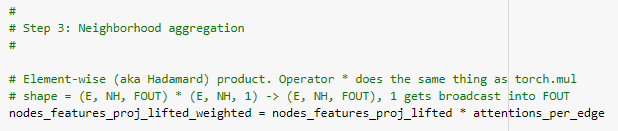


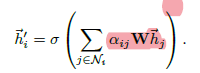
مخرج رابطه‌ی softmax فوق با استفاده از تابع زیر به دست آمده‌است.

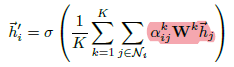


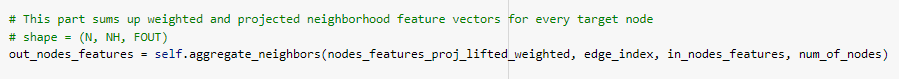
**گام سوم: انباشتن همسایه‌ها**

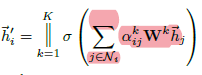
بعد از به دست آوردن ضرایب توجه نرمال شده جهت محاسبه‌ی بردار ویژگی جدید هر گره یا همان بردار ویژگی high levelتر این ضرایب را در بردار ویژگی تبدیل‌یافته ضرب می‌کنیم.

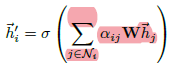
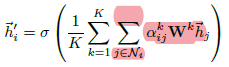


این محاسبات در فرمول‌های زیر مشخص شده‌اند:

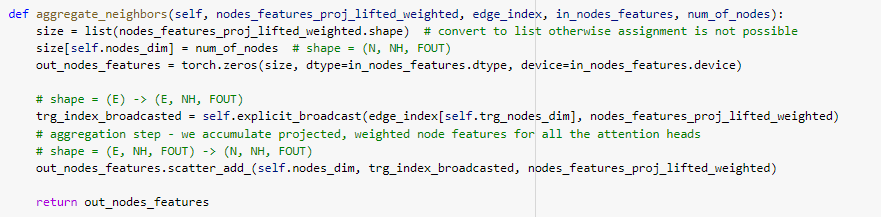




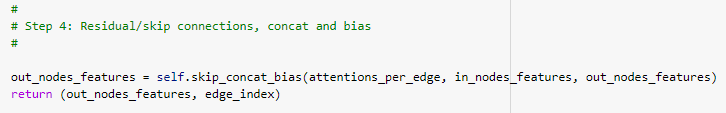
در آخرین قسمت از این مرحله برای به دست آوردن بردار ویژگی جدید هر گره، بردار ویژگی تبدیل‌ یافته‌ی وزن‌دار همسایگان آن را با هم جمع می‌کنیم که توسط تابع aggregate\_neighbors انجام شده است.

پیاده‌سازی تابع aggregate\_neighbors به صورت زیر است:



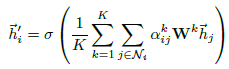
**گام چهارم: skip\_connection, concat, bias**

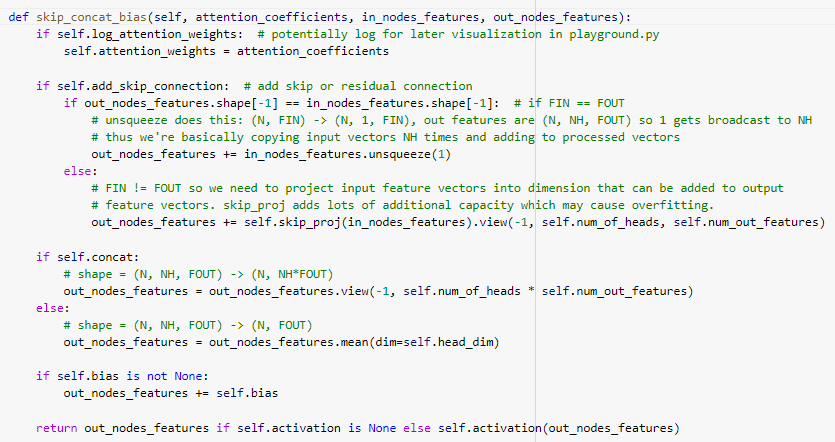


در این مرحله با توجه به متغیرها درباره سه مورد زیر تصمیم می‌گیریم:

* آیا از تکینک skip connection استفاده کنیم؟
* خروجی هدها را با هم الحاق کنیم یا از آن‌ها میانگین بگیریم؟
* استفاده از bias در لایه‌ی Gat

خروجی نهایی خروجی این مرحله است.



استفاده یا عدم استفاده از log\_attention\_weights مربوط به بصری سازی است که در این پیاده‌سازی توجه چندانی به آن نشده است.

استفاده یا عدم استفاده از skip connection، bias و همچنین استفاده از concat برای خروجی هدها یا میانگیری از آنها همگی در ابتدا توضیح داده شده‌اند فقط پیاده‌سازی آن‌ها در تابع فوق انجام شده است.

**2) مدل GAT**

**3) مجموعه داده**

در مقالۀ اصلی GAT چهار مجموعه داده مورد آزمایش قرار گرفته است که دو مورد از آن به عنوان نمونه‌هایی از رویکرد یادگیری Transductive و Inductive جهت بررسی صحت عملکرد مدل در این پیاده‌سازی نیز مورد ارزیابی قرار گرفته‌اند. مشخصات این دو مجموعه داده به شرح زیر است:

**3-1) CORA**

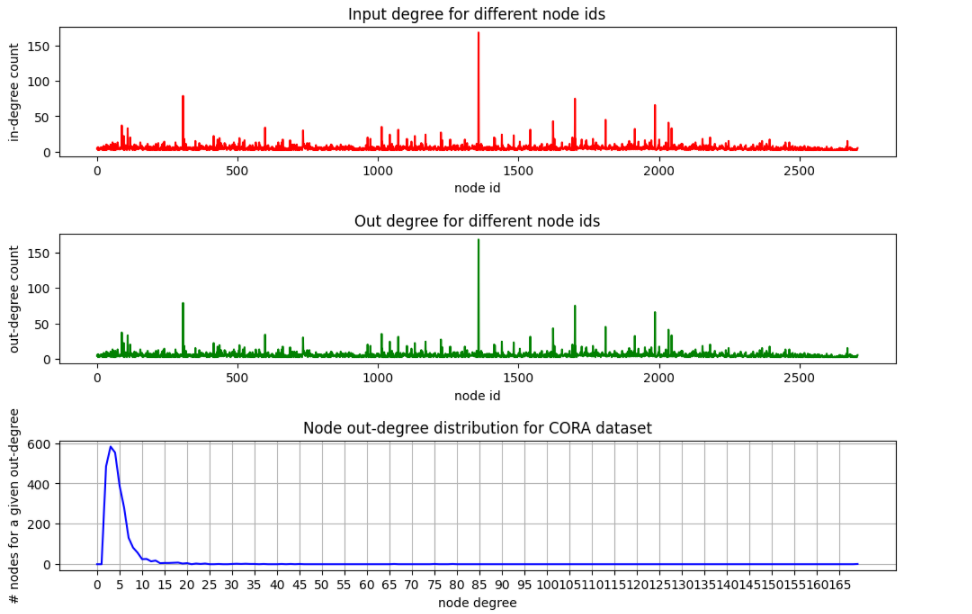
مجموعه داده CORA شامل مقالات مربوط به یادگیری ماشین است. این مقالات به هفت کلاس طبقه بندی می‌شوند. در این مجموعه داده ، گره‌ها به اسناد یا مقالات و یال‌ها به استناد‌[[1]](#footnote-1) (بدون جهت) مربوط هستند.ویژگی‌های گره مطابق با اجزای نمایش BOW[[2]](#footnote-2) یک مقاله هستند. پس از تبدیل کلمات به ریشه‌ی آن‌ها[[3]](#footnote-3) و حذف کلمات توقف با یک مجموعه‌ی واژگان شامل 1433 کلمه منحصر بفرد مواجه هستیم. مقالات به گونه‌ای انتخاب شده‌اند که در پیکره نهایی هر مقاله حداقل یک مقاله دیگر به آن استناد کند یا از آن استناد شود. در کل پیکره 2708 مقاله وجود دارد.

هر گره دارای یک برچسب کلاس است. ما اجازه می‌دهیم فقط 20 گره در هر کلاس برای آموزش استفاده شود - با این وجود ، طبق به تنظیمات Transductive ، الگوریتم آموزش به بردار ویژگی همه گره‌ها دسترسی دارد.

آمار کلی این مجموعه داده مطابق با جدول زیر است:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **تعداد راس‌ها** | **تعداد یال‌ها** | **اندازۀ ویژگی هر راس** | **نمونه‌های آموزشی** | **نمونه‌های اعتبارسنجی** | **نمونه‌های آزمون** |
| 2708 | 5429 | 1433 | 140 | 500 | 1000 |

نمودارهای میزان درجه ورودی هر راس، درجه خروجی هر راس و توزیع درجات خروجی در زیر نمایش داده شده است. همانطور که مشاهده می‌شود، دو نمودار اول مشابه با همدیگر بوده و این به دلیل بدون جهت بودن گراف است (با اینکه به طور طبیعی باید به عنوان یک گراف جهتدار مدل شود).بعضی از گره‌ها دارای تعداد زیادی یال هستند (قله در وسط) ، اما بیشتر گره ها دارای یال‌های بسیار کمتری هستند.در نمودار سوم نیز به خوبی قابل مشاهده است که رئوسی با درجات بسیار بالا، نادر هستند و پراکندگی در تعداد یال‌های کمتر، بیشتر است.



**3-2) Protein-Protein Interaction (PPI)**

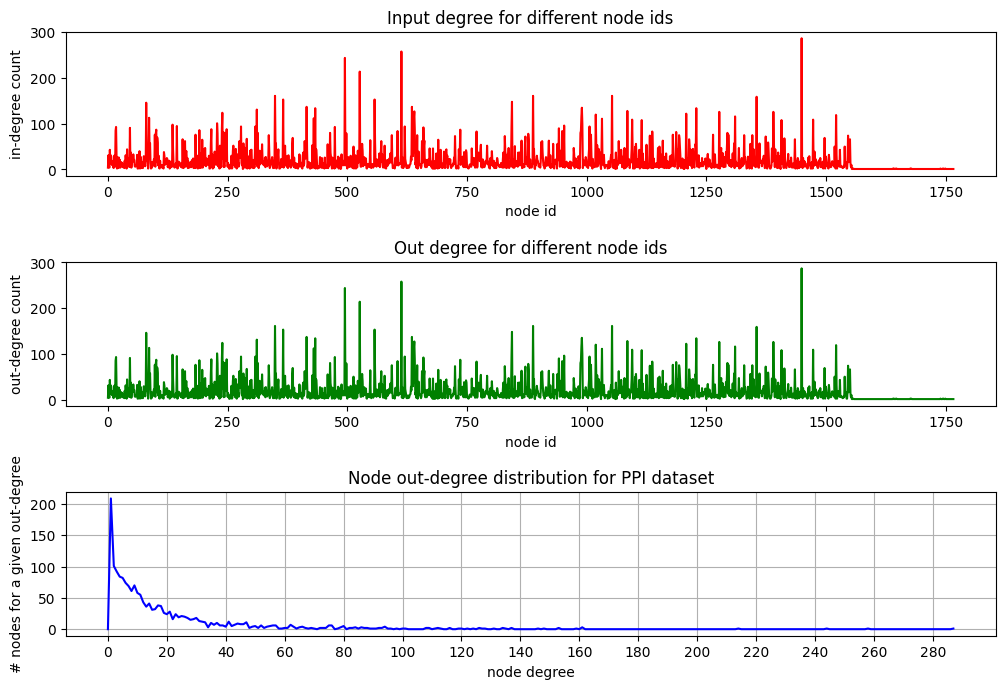
در مسئله‌ای که برای این مجموعه داده در نظر گرفته می‌شود بجای اینکه یادگیری به منظور کسب دانشی درباره ساختار جامعه انجام گیرد بر نقش راس‌ها در آن تاکید دارد، و هدف اصلی دستیابی به عمومیت[[4]](#footnote-4) بر روی گراف‌ها است. در اینجا قصد بر این است تا نقش پروتئین‌ها در گراف‌های مختلف تعاملات پروتئینی-پروتئینی (PPI)، که هر گراف نمایانگر یک بافت متفاوت انسانی است، رده‌بندی شود [1].

مجموعه ژن‌های موضعی[[5]](#footnote-5)، ژن‌های موتیف[[6]](#footnote-6) و امضاهای ایمونولوژیک[[7]](#footnote-7) به عنوان ویژگی و مجموعه آنتولوژی ژن‌ها[[8]](#footnote-8) به عنوان برچسب (مجموعاً 121تا) در نظر گرفته شده است، که همگی از دیتابیس Molecular Signatures Database [2]استخراج شده‌اند.

در مجموع 24 گراف در این مجموعه داده موجود است که 20 گراف آن به منظور آموزش مدل استفاده شده و میانگین امتیازات F1 نیز بر روی 2 گراف آزمون گزارش شده است (2 گراف دیگر نیز به منظور اعتبارسنجی مورداستفاده قرار گرفته است). میانگین درجه هر راس برابر با 28.8 است و به هر راس می‌توان از میان 121 کلاس ممکن، چندین کلاس نیز نسبت داد (رده‌بندی چند-برچسبی[[9]](#footnote-9)).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **تعداد راس‌ها** | **تعداد یال‌ها** | **اندازۀ ویژگی هر راس** | **نمونه‌های آموزشی** | **نمونه‌های اعتبارسنجی** | **نمونه‌های آزمون** |
| 56944 | 1644208 | 50 | 44906 | 6514 | 5524 |

لازم به ذکر است که به دلیل حجم بالای داده در این مجموعه داده، گراف‌ها به شکل دسته‌های دوتایی به مدل داده شده و آموزش پیش خواهد رفت. در نمودارهای زیر، میزان درجات ورودی، درجات خروجی و توزیع آن‌ها برای یکی از گراف‌ها نمایش داده شده است. همانطور که از نمودار اول و دوم مشاهده می‌شود بسیاری از راس‌ها در مقابله با مجموعه داده CORA دارای تعداد یال بیشتری هستند. همچنین طبق نمودار سوم می‌توان گفت که اکثر راس‌ها دارای 1 تا 20 یال هستند.



علاوه بر این، مدل پیاده‌سازی شده بر روی دو مجموعه داده جدید نیز مورد آزمایش قرار گرفته است که جزئیات هرکدام از آن‌ها به شرح زیر است:

**3-3) Reddit**

شبکه اجتماعی Reddit یکی از بزرگترین انجمن‌های اینترنتی است که کاربران به انتشار محتوا و نظرات خود پیرامون بسیاری از موضوعات در آن می‌پردازند. در مسئله‌ای که برای این شبکه مطرح می‌شود سعی بر این است تا پیش‌پینی کنیم که هر پست Reddit به کدام جامعه (موضوع) تعلق دارد.

در سال 2014 یک مجموعه داده گرافی از پست‌های این شبکه ساخته شد. برچسب‌های هر راس در این گراف، جامعه یا subreddit ای است که پست بدان تعلق دارد. گراف ساخته شده به شکل پست-به-پست بوده و دو پست در صورتی با یک یال به یکدیگر متصل هستند که کاربر مشابهی بر روی هر دوی آن‌ها دیدگاه خود را ثبت کرده باشد [3].

در مجموع این مجموعه داده شامل 232965 پست با میانگین درجه 492 است که در مقایسه با سایر مجموعه داده‌های گرافی میزان درجه قابل توجهی به شمار می‌آید. داده‌های مربوط به 20 روز نخست به عنوان مجموعۀ آموزشی و سایر روزها به عنوان آزمایش (30% آن به عنوان اعتبارسنجی) در نظر گرفته شده است.

بردار ویژگی هر راس مطابق با بردارهای کلمات 300 بعدی GloVe CommonCrawl ساخته شده است. به ازای هر پست، ویژگی‌های زیر با یکدیگر الحاق شده‌اند و بردار نهایی را تشکیل داده‌اند:

* میانگین تعبیۀ عنوان پست
* میانگین تعبیۀ تمام دیدگاه‌های پست
* امتیاز پست
* تعداد دیدگاه‌های پست

آمار کلی این مجموعه داده مطابق با جدول زیر است:

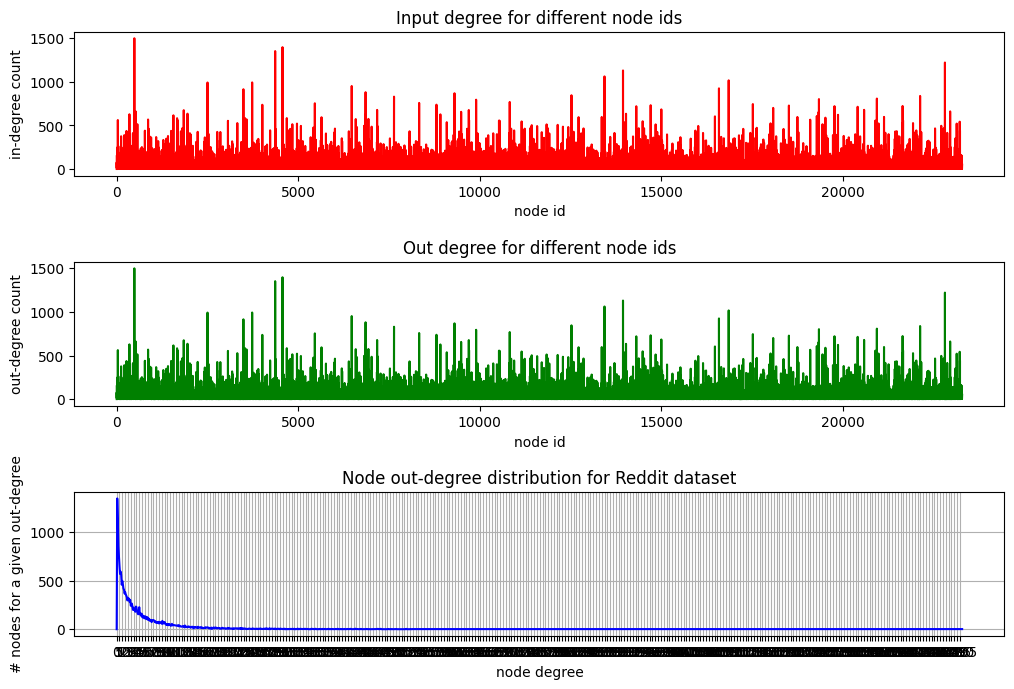
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **تعداد راس‌ها** | **تعداد یال‌ها** | **اندازۀ ویژگی هر راس** | **نمونه‌های آموزشی** | **نمونه‌های اعتبارسنجی** | **نمونه‌های آزمون** |
| 232965 | 114615892 | 602 | 153431 | 23831 | 55703 |

با این حال، به دلیل اندازۀ بسیار بزرگ مجموعۀ داده و عدم دسترسی به سخت‌افزار کافی برای آموزش مدل بر روی این حجم از داده، در آزمایش‌های صورت گرفته تنها به ده درصد آن اکتفا شده و آمار آن به شرح موجود در جدول زیر است:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **تعداد راس‌ها** | **تعداد یال‌ها** | **اندازۀ ویژگی هر راس** | **نمونه‌های آموزشی** | **نمونه‌های اعتبارسنجی** | **نمونه‌های آزمون** |
| 23296 | 1052560 | 602 | 15343 | 2383 | 5570 |

باید توجه داشت که در برخی مقالات پیشین که بر روی این مجموعه داده نیز آزمایش خود را انجام داده‌اند از رویکرد یادگیری Inductive استفاده شده است. این بدین معنی است که در طول فرآیند آموزش مدل، رئوس مربوط به نمونه‌های آزمایشی به کلی نادیده گرفته می‌شوند. با این حال از آنجا که در این آزمایش مجموعه داده کوچک شده، آموزش مدل به شکل Transductive و مشابه با آنچه که بر روی مجموعه CORA اتفاق افتاد، انجام خواهد شد.

نمودارهای میزان درجه ورودی هر راس، درجه خروجی هر راس و توزیع درجات خروجی در زیر نمایش داده شده است. همانطور که مشاهده می‌شود، دو نمودار اول مشابه با همدیگر بوده و این به دلیل بدون جهت بدون گراف است. در نمودار سوم نیز به خوبی قابل مشاهده است که رئوسی با درجات بسیار بالا، نادر هستند و پراکندگی در تعداد یال‌های کمتر، بیشتر است.



**3-4) Wiki\_CS**

مجموعه داده‌ Wiki-CS که در پژوهش [4] معرفی شده است، یک مجموعه داده‌ مبتنی بر ویکی‌پدیا برای آزمودن[[10]](#footnote-10) شبکه های عصبی گرافی است. این مجموعه داده از دسته‌بندی‌های ویکی‌پدیا ، به طور خاص 10 کلاس مربوط به شاخه های علوم کامپیوتر[[11]](#footnote-11) ، با اتصال بسیار بالا ساخته شده است. ویژگی‌های گره از متون مقاله‌های مربوطه گرفته شده است. این ویژگی‌ها به عنوان میانگین تعبیه کلمه از پیش آموزش دیده GloVe محاسبه شدند. در نتیجه ویژگی‌های گره 300 بعدی حاصل می‌شود که می‌تواند برای آموزش مدل‌های بزرگ روی GPU یک مزیت باشد. مقالاتی که می‌توانستند به چند کلاس تعلق داشته باشند از این مجموعه داده حذف شده‌اند.

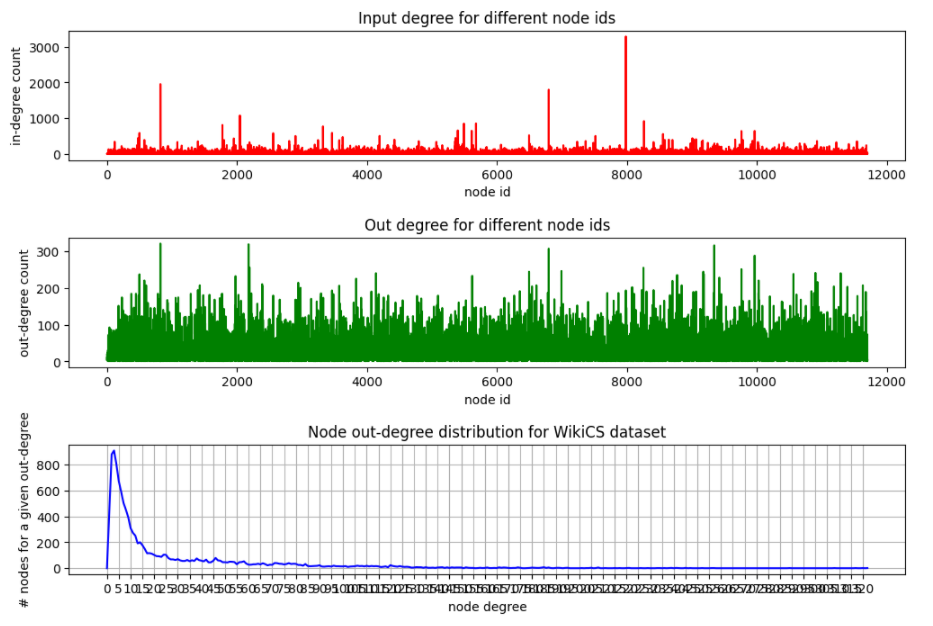
گره‌ها را در هر کلاس به دو مجموعه تقسیم می‌شوند، 50٪ برای مجموعه آزمون و 50٪ به طور بالقوه قابل مشاهده[[12]](#footnote-12). از مجموعه قابل مشاهده ، 20 تقسیم مختلف برای آموزش ، اعتبارسنجی و توقف زودرس[[13]](#footnote-13) ایجاد شده‌است: در هر تقسیم 5٪ از گره‌های هر کلاس برای آموزش ، 22.5٪ برای ارزیابی معیار توقف زودرس مورد استفاده قرار گرفت و 22.5٪ به عنوان مجموعه اعتبارسنجی برای تنظیم هایپر پارامتر استفاده شده‌است.

ما در آزمایشات خود فقط از یکی از 20 تقسیم مختلف برای برای آموزش و اعتبارسنجی استفاده کرده و از داده‌های مربوط به توقف زودرس استفاده نمی‌کنیم.

آمار کلی این مجموعه داده مطابق با جدول زیر است:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **تعداد راس‌ها** | **تعداد یال‌ها** | **اندازۀ ویژگی هر راس** | **نمونه‌های آموزشی** | **نمونه‌های اعتبارسنجی** | **نمونه‌های آزمون** |
| 11701 | 216123 | 300 | 580 | 1769 | 5847 |

نمودارهای میزان درجه ورودی هر راس، درجه خروجی هر راس و توزیع درجات خروجی در زیر نمایش داده شده است. همانطور که مشاهده می‌شود، نمودار اول و دوم متفاوت هستند و این به دلیل جهتدار بودن گراف مربوطه است. در نمودار سوم نیز به خوبی قابل مشاهده است که رئوسی با درجات بسیار بالا، نادر هستند و پراکندگی در تعداد یال‌های کمتر، بیشتر است.



**4) نتایج**

در مقالۀ

# References

|  |  |
| --- | --- |
| [1] | M. Zitnik and L. Jure, "Predicting multicellular function through multi-layer tissue networks," in *Proceedings of the 25th International Conference on Intelligent Systems for Molecular Biology (ISMB)*, 2017. |
| [2] | A. T. Subramanian, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander and J. P. Mesirov, "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles," in *Proceedings of the National Academy of Sciences*, 2005. |
| [3] | W. L. Hamilton, R. Ying and J. Leskovec, "Inductive representation learning on large graphs," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017. |
| [4] | P Mernyei, C Cangea, "Wiki-CS: A Wikipedia-Based Benchmark for Graph Neural Networks," *arXiv preprint arXiv:2007.02901,* 2020. |

1. Citation [↑](#footnote-ref-1)
2. bag-of-words [↑](#footnote-ref-2)
3. stemming [↑](#footnote-ref-3)
4. Generalization [↑](#footnote-ref-4)
5. Positional gene sets [↑](#footnote-ref-5)
6. Motif gene sets [↑](#footnote-ref-6)
7. Immunological signatures [↑](#footnote-ref-7)
8. Gene ontology sets [↑](#footnote-ref-8)
9. Multi-label classification [↑](#footnote-ref-9)
10. Benchmarking [↑](#footnote-ref-10)
11. Computer Science [↑](#footnote-ref-11)
12. Visible [↑](#footnote-ref-12)
13. Early Stopping [↑](#footnote-ref-13)