



تحلیل شبکه‌های پیچیده

پروژه

{Graph Neural Network}



برای ارسال پروژه به نکات زیر توجه کنید.

۱. ملاک اصلی انجام پروژه گزارش آن است و ارسال کد بدون گزارش فاقد ارزش است. برای پروژه یک فایل گزارش در قالب pdf تهیه کنید و در آن برای هر سوال، توضیحات مربوط به آن را ذکر کنید. سعی کنید توضیحات کامل و جامعی تهیه کنید.
۲. زبان برنامه‌نویسی برای انجام پروژه، پایتون (Python) در نظر گرفته شده است.
۳. برای انجام قسمت‌های مختلف از کتابخانه pytorch-geometric استفاده کنید.
۴. در نظر داشته باشید کدهای شما باید قابلیت اجرا در هنگام ارائه را داشته باشند. همچنین بر روی کدهای خود مسلط باشید
۵. کدهای ارسالی خود را برای افزایش خوانایی و درک بهتر به صورت مناسب کامنت‌گذاری کنید.
۶. پاسخ سوالات تشریحی، فایل‌های کد و گزارش خود را در یک فایل فشرده قرار داده و با نام با فرمت Project_StudentID.zip ارسال نمایید.
۷. در صورت وجود ابهام یا اشکال می‌توانید از طریق [کانال](#) با ما در ارتباط باشید (داخل کانال لینک گروه نیز موجود می‌باشد)
۸. برای اطلاع از ددلاین پروژه، به تاریخ اعلام شده در سایت توجه کنید. (با توجه به زمان‌بندی ترم امکان تمدید فراهم نخواهد بود.)
۹. مطابق با قوانین دانشگاه هر نوع کپی‌برداری و تقلب در انجام تکالیف و پروژه‌های این درس مجاز نبوده و در صورت مشاهده نمره طرفین صفر منظور می‌گردد.
۱۰. از بین سوال دو و سه یک سوال را به دلخواه انتخاب کرده و انجام دهید. هر دو سوال از نظر دشواری و زمان لازم برای پیاده‌سازی یکسان هستند.
۱۱. زمان و نحوه تحویل پروژه متعاقباً اعلام خواهد شد.
۱۲. برای پیاده‌سازی مقالات می‌توانید از پیاده‌سازی‌های موجود در اینترنت نیز الهام بگیرید. توجه داشته باشید که تسلط کامل بر روی کدهای تحویل داده شده الزامی است. کپی صرف از منابع موجود موجب کاهش شدید نمره خواهد شد.
۱۳. حداکثر نمره قسمت امتیازی سوال دو و سه برابر ۱۰ درصد کل بارم پروژه است.

سوال اول: پیاده‌سازی شبکه‌های عصبی گرافی پایه (۴۵ نمره)

هدف از این بخش آشنایی با شبکه‌های عصبی گرافی پایه، GCN، GAT، و GATv2 و پیاده‌سازی آن‌ها بر روی مجموعه داده‌های متداول حوزه یادگیری گرافی می‌باشد.

الف) مجموعه داده‌های CoraFull و CiteSeer را با کمک کتابخانه pytorch-geometric بارگیری کنید. سپس تعداد کلاس‌ها، گره‌ها، یال‌ها و تعداد ویژگی‌های هر گره را گزارش کنید. تقسیم‌بندی مجموعه داده برای سه قسمت آموزش^۱، اعتبار سنجی^۲ و آزمون^۳ را به ترتیب برابر با ۷۰٪، ۱۰٪ و ۲۰٪ در نظر بگیرید. (۳ نمره)

در تمامی قسمت‌ها آموزش مدل با قسمت آموزش و پیدا کردن بهترین معماری و ابر پارامترها با قسمت اعتبار سنجی خواهد بود. در آخر دقت نهایی بهترین معماری هر مدل را بر روی قسمت آزمون به دست بیاورید.

ب) یک مدل MLP با حداکثر ۴ لایه برای دسته‌بندی ویژگی گره‌های هر مجموعه داده پیاده‌سازی کنید. توجه کنید که در این حالت از اطلاعات ساختاری گراف استفاده نخواهد شد و تنها ویژگی‌های اولیه هر گره به عنوان ورودی به شبکه داده خواهد شد. دقت^۴ بهترین معماری را گزارش کنید. (۶ نمره)

ج) با استفاده از کتابخانه pytorch-geometric، مدل GCN دو لایه را پیاده‌سازی کنید. بررسی کنید که بهترین تعداد ابعاد لایه پنهان چند می‌باشد. حداقل ۴ مقدار (مانند ۱۶، ۳۲، ...) را به عنوان ابعاد پنهان بررسی کنید. سپس مدل GCN یک لایه و سه لایه را نیز پیاده‌سازی و ارزیابی کنید. برای مدل GCN سه لایه نیز حداقل ۴ مقدار متفاوت برای ابعاد لایه‌های پنهان امتحان کنید. دقت بهترین مدل GCN خود را بر روی قسمت آزمون گزارش کنید. (۱۰ نمره)

د) قسمت ج را به طور مشابه برای مدل GAT به جای مدل GCN تکرار کنید. پس از یافتن بهترین مدل، حال بررسی کنید که بهترین تعداد سرهای توجه^۵ چه تعداد است. حداقل ۳ مقدار متفاوت را بررسی کنید. (۱۰ نمره)

ه) قسمت د را برای مدل GATv2 تکرار کنید. (۷ نمره)

و) نتایج قسمت‌های بالا را از جنبه دقت و زمان اجرا برای هر دو مجموعه داده مقایسه و تحلیل کنید. (۴ نمره)

ز) پیچیدگی زمانی و حافظه‌ای برای یک شبکه گرافی یک لایه با n گره، m یال، d بعد به عنوان ویژگی اولیه و f بعد به عنوان ابعاد خروجی را به دست آورید. مدل GCN، GAT و GATv2 را از این لحاظ با یکدیگر مقایسه کنید. (۵ نمره)

^۱ train

^۲ validation

^۳ test

^۴ accuracy

^۵ attention-head

سوال دوم: پیاده‌سازی مقاله اول (۵۵ نمره)

* توجه: از بین این سوال و سوال سوم تنها یک سوال را به دلخواه انجام دهید.

این سوال در مورد پیاده‌سازی مقاله زیر است:

[لینک](#) PREDICT THEN PROPAGATE: GRAPH NEURAL NETWORKS MEET PERSONALIZED PAGERANK

الف) ایده کلی مقاله، نوآوری و تفاوت آن با شبکه GCN چیست؟ همچنین ارتباط مقاله را با الگوریتم PageRank بیان کنید. (۷ نمره)

توضیحات بخش الف بایستی در حدود نصف صفحه باشد.

ب) over-smoothing چیست، چه زمان رخ می‌دهد و روش عملی برای تشخیص این پدیده چیست؟ در روش پیشنهادی این مقاله چه تدبیری برای جلوگیری از over-smoothing اندیشیده شده است؟ (۴ نمره)

ج) مدل پیشنهادی این مقاله را پیاده‌سازی کنید.

د) مدل پیاده‌سازی شده را بر روی مجموعه داده CoraFull و CiteSeer مطابق با همان تقسیم‌بندی سوال اول آموزش داده و ارزیابی کنید. (بخش ج و د ۲۴ نمره)

ه) نتایج حاصل از قسمت ج را با نتایج مدل‌های پایه که در سوال اول به دست آمده است مقایسه کنید. آیا مدل پیاده‌سازی شده توانسته است دقت مدل‌های پایه را بهبود بخشد؟ چرا؟ (۴ نمره)

و) آیا نتایج حاصل از پیاده‌سازی شما با نتایج گزارش شده در مقاله همخوانی دارد یا نه؟ توضیح دهید. (۲ نمره)

ز) در معادله ۳ مقاله، یک ماتریس قابل یادگیری به صورت زیر به روش پیشنهادی اضافه کرده و پیاده‌سازی کنید. نتایج را با نتایج قسمت ج مقایسه کنید. (۱۴ نمره)

$$Z_{PPNP} = w * \text{softmax}(\alpha(I_n - (1 - \alpha)\hat{A})^{-1}H)$$

ح) اگر ایده‌ای برای بهبود عملکرد روش پیشنهادی دارید، آن را پیاده‌سازی کنید. نتایج حاصل از پیاده‌سازی روش اصلی در قسمت ج با روش بهبود یافته را مقایسه کنید. دلیل بهبود عملکرد روش پیشنهادی خود را بیان کنید. این قسمت اختیاری و دارای نمره امتیازی است.

سوال سوم: پیاده‌سازی مقاله دوم (۵۵ نمره)

* توجه: از بین این سوال و سوال دوم تنها یک سوال را به دلخواه انجام دهید.

این سوال در مورد پیاده‌سازی مقاله زیر است:

DropEdge: Towards Deep Graph Convolutional Networks on Node Classification ([لینک](#))

الف) ایده کلی مقاله، نوآوری آن چیست؟ تفاوت‌ها و شباهت‌های ایده روش پیشنهادی با ایده Dropout در یادگیری عمیق چیست؟ توضیح دهید. (۶ نمره)

توضیحات بخش الف بایستی در حدود نصف صفحه باشد.

ب) روش پیشنهادی این مقاله را پیاده‌سازی کنید. مدل پایه برای اعمال روش پیشنهادی را GCN دو لایه در نظر بگیرید.

ج) مدل پیاده‌سازی شده در قسمت ب را بر روی مجموعه داده CoraFull و CiteSeer مطابق با همان تقسیم‌بندی سوال اول آموزش داده و ارزیابی کنید. (بخش ب و ج ۲۱ نمره)

د) نتایج حاصل از قسمت ج را با نتایج مدل‌های پایه که در سوال اول به دست آمده است مقایسه کنید. آیا روش DropEdge توانسته است دقت بهتری نسبت به GCN بدون DropEdge کسب کند؟ چرا؟ (۳ نمره)

ه) آیا نتایج حاصل از پیاده‌سازی شما با نتایج گزارش شده در مقاله همخوانی دارد یا نه؟ توضیح دهید. (۲ نمره)

و) over-smoothing چیست، چه زمان رخ می‌دهد و روش عملی برای تشخیص این پدیده چیست؟ (۴ نمره)

ز) حال تعداد لایه‌های GCN به همراه DropEdge را برابر با هشت لایه قرار دهید. مدل را بر روی مجموعه داده CiteSeer آموزش دهید و آن را ارزیابی کنید. آیا DropEdge همچنان در جلوگیری از over-smoothing موثر است؟ (۷ نمره)

ح) یکی از راه‌های جلوگیری از over-smoothing استفاده از روش skip-connection است؛ ایده skip-connection را بر روی یک GCN هشت لایه اعمال کنید. آیا روش skip-connection توانسته است در جلوگیری از over-smoothing موثر باشد؟ از بین DropEdge و skip-connection کدام یک در جلوگیری از over-smoothing بهتر عمل می‌کند؟ توضیح دهید. (۱۲ نمره)

ط) اگر ایده‌ای برای بهبود عملکرد روش پیشنهادی دارید، آن را پیاده‌سازی کنید. نتایج حاصل از پیاده‌سازی روش اصلی در قسمت ج با روش بهبود یافته را مقایسه کنید. دلیل بهبود عملکرد روش پیشنهادی خود را بیان کنید. این قسمت اختیاری و دارای نمره امتیازی است.