دانشگاه صنعتی خواجه نصیرالدین طوسی

دانشکده مهندسی نقشه برداری (ژئودزی و ژئوماتیک)

پیاده‌سازی GNN روی دیتاست METR-LA

استاد درس

خانم دکتر پرستو پیله‌فروشها

دانشجو

باربد عالمیان

تابستان ۱۴۰۵

فهرست

۱- معرفی دیتاست ۲
۲- معماری مدل‌ها ۲
۲-۱- مدل GCN ۲
۲-۲- مدل MLP (مدل غیرگرافی) ۳
۲-۳- مدل میانگین تاریخی (Historical Average) ۳
۳- آزمایش‌ها و نتایج ۳
۳-۱- آزمایش اول: GCN در مقابل مدل‌های غیرگرافی ۳
۳-۲- آزمایش دوم: اثر عمق GCN ۴
۳-۳- آزمایش سوم: گراف وزنی در مقابل گراف بدون وزن ۴
۳-۴- آزمایش چهارم: تحلیل خطای فضایی سنسورها ۵
۴- پاسخ به سوالات تمرین ۶
۴-۱- سوال ۱: آیا اطلاعات فضایی گراف بهبود پیش‌بینی ترافیک را به همراه دارد؟ ۶
۴-۲- آیا افزایش عمق GCN همواره بهبود عملکرد را به همراه دارد؟ ۶
۴-۳- آیا وزن‌های مبتنی بر فاصله مفید هستند؟ ۶
۴-۴- تحلیل فضایی خطا ۶
۵- جمع‌بندی نهایی ۷

در این تمرین، یک شبکه عصبی گرافی (GCN) برای پیش‌بینی سرعت ترافیک با استفاده از داده‌های سنسورهای جاده‌ای در لس‌آنجلس (دیتاست METR-LA) پیاده‌سازی شده است. هدف اصلی، بررسی تأثیر اطلاعات فضایی گراف بر دقت پیش‌بینی و تحلیل عوامل مؤثر بر عملکرد مدل است. چهار آزمایش مختلف شامل مقایسه GCN با مدل‌های غیرگرافی، بررسی اثر عمق شبکه، مقایسه گراف وزنی و بدون وزن، و تحلیل خطای فضایی سنسورها انجام شده است.

۱- معرفی دیتاست

دیتاست METR-LA شامل داده‌های سرعت ترافیک جمع‌آوری شده از ۲۰۷ سنسور جاده‌ای در لس‌آنجلس است. داده‌ها در بازه زمانی مارس تا ژوئن ۲۰۱۲ با گام‌های زمانی ۵ دقیقه‌ای جمع‌آوری شده‌اند.

ویژگی	مقدار
تعداد سنسورها	۲۰۷
تعداد گام‌های زمانی	۳۴,۲۷۲
تعداد یال‌های گراف	۱,۵۱۵
طول پنجره زمانی(History Length)	۱۲ گام (۱ ساعت)
تقسیم داده	٪۷۰ آموزش، ٪۱۰ اعتبارسنجی، ٪۲۰ تست

۲- معماری مدل‌ها

۲-۱- مدل GCN

شبکه عصبی گرافی پیاده‌سازی شده شامل لایه‌های کانولوشن گرافی (GCNConv) با تابع فعال‌سازی ReLU و Dropout است.

معماری پایه به صورت زیر است:

بخش	مشخصات
ورودی	ویژگی‌های هر گره (۱۲ گام زمانی)
لایه اول کانولوشن	GCNConv (۱۲, ٦٤)
لایه‌های میانی	GCNConv (٦٤, ٦٤)
لایه خروجی	GCNConv (٦٤, ۱)
Dropout	۰.۱

۲-۲- مدل MLP (مدل غیرگرافی)

یک پرسپترون چندلایه با ورودی تخت‌شده (۱۲ گام زمانی × ۲۰۷ سنسور = ۲۴۸۴ ویژگی) و دو لایه مخفی با ۶۴ نورون برای مقایسه با مدل GCN استفاده شده است.

۲-۳- مدل میانگین تاریخی (Historical Average)

یک مدل پایه ساده که سرعت آینده را برابر با میانگین سرعت گذشته در نظر می‌گیرد.

۳- آزمایش‌ها و نتایج

۳-۱- آزمایش اول: GCN در مقابل مدل‌های غیرگرافی

هدف: بررسی اینکه آیا اطلاعات فضایی گراف باعث بهبود پیش‌بینی ترافیک می‌شود یا خیر.

نتایج:

مدل	MAE	RMSE
میانگین تاریخی	۰.۵۳۹۲	۰.۸۴۹۵
MLP	**۰.۲۹۷۳**	**۰.۵۳۰۳**
GCN	۰.۵۲۶۴	۰.۸۴۲۵

تحلیل نتایج:

نتایج نشان می‌دهد که مدل MLP با MAE برابر ۰.۲۹۷۳ عملکرد بهتری نسبت به GCN دارد. این نتیجه برخلاف انتظار اولیه است، زیرا انتظار می‌رفت که مدل GCN با استفاده از اطلاعات فضایی، عملکرد بهتری داشته باشد. اما نکات زیر قابل توجه است:

- **زمان آموزش**: مدل GCN حدود ۶ برابر بیشتر از MLP زمان آموزش نیاز دارد (۱۳۸۶ ثانیه در مقابل ۲۳۵ ثانیه).
- **دلیل عملکرد ضعیف‌ترGCN** : با توجه به اینکه داده‌های ورودی به GCN به صورت (تعداد سنسورها × طول پنجره زمانی) است، این مدل تنها از اطلاعات یک گام زمانی استفاده می‌کند، در حالی که MLP از تمام ۱۲ گام زمانی به‌عنوان ویژگی‌های ورودی استفاده می‌کند. این تفاوت در ساختار ورودی، مزیت قابل توجهی به MLP داده است.
- **نتیجه‌گیری نهایی**: در این پیاده‌سازی خاص، اطلاعات فضایی گراف به تنهایی نتوانسته از اطلاعات زمانی در MLP پیشی بگیرد. اما GCN همچنان از مدل میانگین تاریخی (MAE=۰.۵۳۹۲) بهتر عمل کرده است.

۳-۲- آزمایش دوم: اثر عمق GCN

هدف: بررسی اینکه آیا افزایش عمق شبکه GCN همواره باعث بهبود عملکرد می‌شود یا خیر (بررسی پدیده Over-smoothing).

تعداد لایه‌ها	MAE	RMSE	Cosine Similarity
۱	۰.۵۲۷۲	۰.۸۴۲۵	۰.۱۸۸۹
۲	**۰.۵۲۷۰**	**۰.۸۴۲۲**	۰.۲۸۶۸
۳	۰.۵۳۱۹	۰.۸۵۳۷	۰.۳۴۲۵
۴	۰.۵۳۳۴	۰.۸۵۹۰	۰.۲۹۴۹

تحلیل نتایج:

- **بهترین عمق**: عمق ۲ لایه بهترین عملکرد را با MAE=۰.۵۲۷۰ دارد.
- پدیده Over-smoothing: با افزایش عمق از ۲ به ۳ و ۴، خطا افزایش می‌یابد. این نشان می‌دهد که عمق بیشتر باعث می‌شود تا ویژگی‌های گره‌ها بیش از حد با یکدیگر مخلوط شوند (Over-smoothing) و تمایز بین گره‌ها از بین برود.
- Cosine Similarity: با افزایش عمق، شباهت بین بردارهای گره‌ها ابتدا افزایش و سپس کاهش می‌یابد. در عمق ۳ بیشترین شباهت (۰.۳۴۲۵) مشاهده می‌شود که نشان‌دهنده شروع Over-smoothing است.
- نتیجه‌گیری: عمق ۲ لایه بهترین تعادل بین قدرت یادگیری و جلوگیری از Over-smoothing را ارائه می‌دهد.

۳-۳- آزمایش سوم: گراف وزنی در مقابل گراف بدون وزن

هدف: بررسی اینکه آیا استفاده از وزن‌های مبتنی بر فاصله (سنسورهای نزدیک‌تر وزن بیشتر) بهبود عملکرد را به همراه دارد یا خیر.

نوع گراف	MAE	RMSE
بدون وزن	۰.۵۲۶۴	۰.۸۴۲۵
وزنی (فاصله‌محور)	**۰.۴۸۲۲**	**۰.۷۸۱۲**

تحلیل نتایج:

- بهبود عملکرد: گراف وزنی با MAE=۰.۴۸۲۲ عملکرد بهتری نسبت به گراف بدون وزن (MAE =۰.۵۲۶۴) دارد. این بهبود حدود ۸.۳۹٪ است.
- دلیل بهبود: در گراف وزنی، سنسورهای نزدیک‌تر به یکدیگر وزن بیشتری دارند. این موضوع باعث می‌شود که اطلاعات سنسورهای نزدیک‌تر در پیش‌بینی سرعت یک سنسور خاص تأثیر بیشتری داشته باشند. این با واقعیت فیزیکی جاده‌ها همخوانی دارد که سنسورهای نزدیک به هم رفتار ترافیکی مشابهی دارند.
- نتیجه‌گیری: استفاده از وزن‌های مبتنی بر فاصله در گراف، عملکرد مدل GCN را به طور قابل توجهی بهبود می‌بخشد.

۳-۴- آزمایش چهارم: تحلیل خطای فضایی سنسورها

هدف: تحلیل توزیع خطای پیش‌بینی در بین سنسورها و شناسایی عواملی که بر خطای سنسورها تأثیر می‌گذارند.

معیار	مقدار
میانگین خطا	۰.۵۲۶۴
انحراف معیار	۰.۲۵۹۷
حداکثر خطا	۲.۰۹۶۶ (سنسور ۵۶)
حداقل خطا	۰.۱۷۷۳ (سنسور ۲۰۱)
تعداد سنسورهای پرخطا (۱۰٪ بالای خطا)	۲۱
میانگین خطای سنسورهای مرکزی	۰.۶۱۷۴
میانگین خطای سنسورهای محیطی	۰.۴۹۹۷

تحلیل نتایج:

- سنسورهای پرخطا: ۲۱ سنسور (۱۰٪ بالای داده‌ها) دارای خطای بالایی هستند. سنسور ۵۶ با خطای ۲.۰۹۶۶ بیشترین خطا را دارد.
- سنسورهای مرکزی در مقابل محیطی: سنسورهای مرکزی (با درجه بالاتر در گراف) خطای بیشتری دارند (۰.۶۱۷۴) نسبت به سنسورهای محیطی (۰.۴۹۹۷).
- دلیل این تفاوت: سنسورهای مرکزی اطلاعات بیشتری از همسایگان خود دریافت می‌کنند و این اطلاعات ممکن است نویز بیشتری داشته باشد. همچنین ممکن است سنسورهای مرکزی در نقاط شلوغ‌تری قرار داشته باشند که پیش‌بینی آن‌ها دشوارتر است.
- سنسورهای ایزوله: سنسورهای با درجه پایین‌تر (محیطی) معمولاً خطای کمتری دارند، زیرا رفتار آن‌ها مستقل‌تر و قابل پیش‌بینی‌تر است.
- نتیجه‌گیری GCN : در مناطق با تراکم سنسور بالا (مرکزی) عملکرد ضعیف‌تری دارد، زیرا اطلاعات زیاد و نویز بالا باعث سردرگمی مدل می‌شود. در مناطق با تراکم کم (محیطی)، مدل عملکرد بهتری دارد.

۴- پاسخ به سوالات تمرین

۴-۱- سوال ۱: آیا اطلاعات فضایی گراف بهبود پیش‌بینی ترافیک را به همراه دارد؟

در این پیاده‌سازی، GCN عملکرد بهتری نسبت به مدل میانگین تاریخی (MAE: ۰.۵۲۶۴ در مقابل ۰.۵۳۹۲) داشت، اما از (MAE: ۰.۲۹۷۳) MLP ضعیف‌تر عمل کرد. این نتیجه نشان می‌دهد که اطلاعات فضایی گراف به‌تنهایی نمی‌تواند جایگزین اطلاعات زمانی شود، اما در ترکیب با اطلاعات زمانی (با اصلاح معماری) می‌تواند مفید باشد. برای بهبود GCN، می‌توان از ترکیب اطلاعات زمانی و فضایی با استفاده از معماری‌هایی مانند ST-GCN یا T-GCN استفاده کرد.

۴-۲- آیا افزایش عمق GCN همواره بهبود عملکرد را به همراه دارد؟

خیر. با افزایش عمق از ۲ به ۳ و ۴، خطا افزایش یافته است. عمق ۲ لایه بهترین عملکرد را داشت. عمق بیشتر باعث پدیده Over-smoothing می‌شود که در آن ویژگی‌های گره‌ها بیش از حد شبیه به یکدیگر می‌شوند و تمایز بین گره‌ها از بین می‌رود. Cosine Similarity نیز با افزایش عمق افزایش یافته و نشان‌دهنده این پدیده است.

۴-۳- آیا وزن‌های مبتنی بر فاصله مفید هستند؟

بله. گراف وزنی با MAE =۰.۴۸۲۲ عملکرد بهتری نسبت به گراف بدون وزن با MAE = ۰.۵۲۶۴ داشت. این بهبود حدود ۸.۳۹٪ است. وزن‌دهی مبتنی بر فاصله باعث می‌شود سنسورهای نزدیک‌تر به یکدیگر تأثیر بیشتری در پیش‌بینی داشته باشند که با واقعیت فیزیکی جاده‌ها همخوانی دارد.

۴-۴- تحلیل فضایی خطا

سنسورهای مرکزی (با درجه بالاتر) خطای بیشتری دارند (۰.۶۱۷۴) نسبت به سنسورهای محیطی (۰.۴۹۹۷). این نشان می‌دهد که مناطق با تراکم بالای سنسور دارای اطلاعات بیشتر اما نویز بالاتری هستند که پیش‌بینی را دشوارتر می‌کند. سنسورهای ایزوله (محیطی) رفتار قابل پیش‌بینی‌تری دارند و خطای کمتری در آن‌ها مشاهده می‌شود.

۵- جمع‌بندی نهایی

آزمایش	نتیجه اصلی
GCN در مقابل مدل‌های غیرگرافی	MLPعملکرد بهتری داشت (MAE: ۰٫۲۹۷۳ در مقابل ۰.۵۲۶۴) دلیل اصلی استفاده از تمام ۱۲ گام زمانی توسط MLP است.
اثر عمقGCN	عمق ۲ لایه بهترین عملکرد را داشت. عمق بیشتر باعث Over-smoothing می‌شود.
گراف وزنی در مقابل بدون وزن	گراف وزنی با بهبود ٪۸.۳۹ عملکرد بهتری داشت.
تحلیل خطای فضایی	سنسورهای مرکزی خطای بیشتری دارند. مناطق پرتراکم و پرنویز باعث کاهش دقت می‌شوند.

پیشنهادات برای بهبود:

- استفاده از معماری‌های پیشرفته‌تر مانند ST-GCN یا T-GCN که ترکیبی از اطلاعات زمانی و فضایی را به کار می‌برند.
- تنظیم دقیق‌ترHyperparameter ها مانند نرخ یادگیری، تعداد لایه‌ها و.Dropout
- استفاده از Batch Normalization برای بهبود سرعت همگرایی.
- آزمایش با طول پنجره‌های زمانی مختلف.