

تمرین: مقایسه معماری‌های دارای گیت - LSTM، GRU و BiLSTM

در این تمرین، چهار معماری Simple RNN، LSTM، GRU و BiLSTM را برای پیش‌بینی سری زمانی پیاده‌سازی و مقایسه خواهید کرد. دیتاست مورد استفاده Household Electric Power Consumption است که از آدرس <https://www.kaggle.com/datasets/uciml/electric-power-consumption-data-set> در دسترس می‌باشد. هدف اصلی درک عملکرد مکانیزم‌های گیت در معماری‌های مختلف و شناسایی تفاوت‌های عملی آنها است. تمام مدل‌ها باید با معماری یکسان پیاده‌سازی شوند: دو لایه بازگشتی با ۶۴ واحد مخفی در هر لایه و یک لایه کاملاً متصل برای خروجی. از رویکرد پنجره لغزان استفاده کنید که ۶۰ گام زمانی متوالی برای پیش‌بینی مقدار در گام ۶۱ به کار می‌رود. داده‌ها را به ۷۰٪ آموزش، ۱۵٪ اعتبارسنجی و ۱۵٪ تست تقسیم کرده و ترتیب زمانی را حفظ کنید. ویژگی‌ها را با استانداردسازی (میانگین صفر و واریانس یک) نرمال کنید که فقط روی مجموعه آموزش محاسبه شود. تمام مدل‌ها با هاپیرپارامترهای یکسان شامل نرخ یادگیری، اندازه batch، بهینه‌ساز Adam و تعداد epoch آموزش داده شوند.

الف) جدول مقایسه جامع عملکرد

پس از آموزش چهار مدل، آنها را روی مجموعه تست ارزیابی کرده و جدول مقایسه‌ای ایجاد کنید که شامل معیارهای زیر باشد: میانگین مربعات خطا (MSE) و میانگین قدرمطلق خطا (MAE) روی تست، میانگین زمان آموزش هر epoch به ثانیه، تعداد کل پارامترهای قابل آموزش و سرعت همگرایی که به عنوان تعداد epoch لازم برای رسیدن به ۹۰٪ بهترین عملکرد اعتبارسنجی تعریف می‌شود. بر اساس جدول مقایسه خود به سؤالات زیر پاسخ دهید: کدام معماری بهترین توازن بین دقت، هزینه محاسباتی و سرعت همگرایی را ارائه می‌دهد و چرا؟ BiLSTM از نظر این معیارها چگونه با LSTM مقایسه می‌شود و آیا بهبود عملکرد احتمالی ارزش دو برابر شدن پارامترها را دارد؟ در چه سناریوهایی GRU را به LSTM یا BiLSTM ترجیح می‌دهید با در نظر گرفتن محدودیت‌های زمانی آموزش و نیازهای استقرار مدل؟

ب) بصری‌سازی مکانیزم گیت‌ها

کدی پیاده‌سازی کنید که مقادیر فعال‌سازی گیت‌های داخلی LSTM و GRU را در حین استنتاج استخراج کند. برای LSTM سه گیت forget، input و output را در هر timestep استخراج کنید و برای GRU دو گیت update و reset را بدست آورید. سه توالی نمونه با طول ۶۰ timestep از مجموعه تست انتخاب کرده و برای هر توالی heatmap یا نمودار خطی رسم کنید که محور افقی موقعیت timestep، محور عمودی نوع گیت و رنگ یا ارتفاع خط مقدار فعال‌سازی گیت بین ۰ تا ۱ را نشان دهد. به سؤالات زیر پاسخ دهید: forget gate در LSTM چه زمانی به ۱ نزدیک می‌شود (یادآوری) و چه زمانی به ۰ نزدیک می‌شود (فراموشی) و چه اتفاقی در توالی ورودی در آن timestep رخ می‌دهد که این رفتار را ایجاد می‌کند؟ گیت‌های update و reset

در GRU را با گیت‌های forget و input در LSTM مقایسه کنید و توضیح دهید که الگوهای فعال‌سازی آنها چگونه متفاوت است و آیا رفتار آنها همبسته است یا مستقل؟ timestep‌های خاصی را شناسایی کنید که در آن مدل تصمیم به یادآوری اطلاعات مهم می‌گیرد و این الگوها را با مقادیر واقعی سری زمانی در آن timestep‌ها مرتبط کنید.

پ) آزمایش وابستگی بلندمدت

آزمایشی طراحی کنید که بررسی کند هر معماری چقدر توالی‌های با طول مختلف را به خوبی مدیریت می‌کند. سه نسخه از دیتاست با طول‌های توالی ۳۰، ۶۰ و ۱۲۰ گام زمانی برای پیش‌بینی مقدار بعدی ایجاد کنید. مدل‌های جداگانه‌ای برای هر طول توالی آموزش دهید در حالی که مشخصات معماری و هابیر پارامترها ثابت بماند. تمام مدل‌ها را روی مجموعه تست ارزیابی کرده و دقت پیش‌بینی (MSE یا MAE) برای Simple RNN، GRU، LSTM و BiLSTM در هر طول توالی را ثبت کنید. نمودار رسم کنید که محور افقی طول توالی و محور عمودی خطای تست باشد با چهار خط برای چهار معماری. شواهد کمی ارائه دهید که معماری‌های دارای گیت (GRU، LSTM و BiLSTM) عملکرد خود را در توالی‌های بلندتر بهتر از Simple RNN حفظ می‌کنند. افت عملکرد را برای هر معماری با افزایش طول توالی از ۳۰ به ۱۲۰ محاسبه کنید. کدام معماری پایدارترین عملکرد را در طول‌های مختلف نشان می‌دهد؟ آیا BiLSTM برتری خود را در توالی‌های بلندتر حفظ می‌کند یا مزایا کاهش می‌یابد؟ آیا GRU با وجود پارامترهای کمتر، توانایی LSTM در مدیریت توالی‌های بلند را دارد؟

ت) تحلیل نقادانه دوطرفه بودن

LSTM تک‌جهته را با BiLSTM در وظیفه پیش‌بینی سری زمانی مقایسه کنید. به دقت تحلیل کنید که آیا دسترسی BiLSTM به timestep‌های آینده در حین آموزش مزیت واقعی ارائه می‌دهد یا سناریوی غیرواقعی ایجاد می‌کند که به پیش‌بینی دنیای واقعی منتقل نمی‌شود. گزارش دهید که آیا BiLSTM عملکرد پیش‌بینی سری زمانی را نسبت به LSTM تک‌جهته بهبود می‌بخشد و اعداد خطای تست هر دو معماری را ارائه کنید. درباره علیت در پیش‌بینی سری زمانی تفکر نقادانه کنید: هنگام پیش‌بینی مقدار در زمان t ، آیا مدل باید به اطلاعات از زمان‌های $t+1$ ، $t+2$ و بعدتر دسترسی داشته باشد؟ در سناریوهای پیش‌بینی دنیای واقعی، اطلاعات آینده هنوز در دسترس نیست. توضیح دهید که چرا ممکن است عملکرد بهتر BiLSTM در مجموعه تست به استقرار واقعی منتقل نشود. این را با سناریوی طبقه‌بندی متن مقابله کنید و توضیح دهید که چرا زمینه دوطرفه برای وظایف طبقه‌بندی که کل توالی در دسترس است مشروع و مفید است اما برای پیش‌بینی سری زمانی که مقادیر آینده هنوز وجود ندارد مشکل‌ساز است. توصیف کنید که چه نوع وظایف توالی BiLSTM را مناسب‌ترین می‌کند.

ث) بررسی حذف گیت

سه نسخه تغییر یافته از LSTM پیاده‌سازی کنید که در آنها مکانیزم گیت برای درک اینکه کدام گیت‌ها حیاتی‌تر هستند تغییر داده شده است: LSTM با فقط forget gate فعال (input gate و

output gate همیشه ۱)، LSTM با فقط input gate فعال (forget gate و output gate همیشه ۱) و LSTM کامل با هر سه گیت. هر سه نسخه را آموزش داده و عملکرد تست آنها را مقایسه کنید. کدام گیت برای عملکرد در این وظیفه پیش‌بینی سری زمانی حیاتی‌تر است؟ سه نسخه را بر اساس خطای تست رتبه‌بندی کنید. اگر حذف یک گیت خاص باعث افت شدید عملکرد شود، توضیح دهید که آن گیت چه نقشی دارد و چرا برای این وظیفه خاص مهم است. آیا اهمیت نسبی گیت‌ها با انتظار شما بر اساس نقش‌های نظری آنها مطابقت دارد؟