

تحقیق ۱:

روش های Kernel Initialization در لایه RNN:

روش پیش فرض: **glorot_uniform**

همچنین به عنوان مقداردهی اولیه Xavier شناخته می شود.

واریانس خروجی ها را در لایه ها متعادل می کند.

برای فعال سازی های tanh یا sigmoid در شبکه های پیش خور به خوبی کار می کند.

اما برای RNN ها، به خصوص با توالی های طولانی یا معماری های عمیق تر، ممکن است بهینه نباشد.

شروع کننده های جایگزین:

ویژگی ها و مزایا	مناسب برای چه نوع لایه ای؟	شروع کننده
افزایش سرعت همگرایی، مناسب برای شبکه های عمیق	لایه هایی با ReLU یا LeakyReLU	he_uniform
مناسب برای شبکه های کم عمق یا داده های نرمال شده	لایه هایی با sigmoid	lecun_uniform
حفظ بهتر گرادیان ها در طول زمان، پایدارتر	مخصوصاً برای RNN ها با tanh	orthogonal
نیاز به تنظیم دقیق دارد، ممکن است ناپایدار باشد	برای تست های تجربی یا مدل های تصادفی	random_normal
انعطاف پذیر، قابل تنظیم برای انواع مختلف لایه ها	مقداردهی تطبیقی بر اساس نوع لایه	variance_scaling

تحقیق ۲:

روش های موازی سازی محاسبات در لایه RNN:

محاسبات موازی در لایه های شبکه عصبی بازگشتی (RNN) به دلیل وابستگی ذاتی RNN ها به ترتیب، چالش قابل توجهی را ایجاد می کند. محاسبه هر گام زمانی به خروجی گام زمانی قبلی متکی است و پردازش همزمان چندین گام زمانی را به روش سنتی دشوار می کند. با این حال، چندین رویکرد و ملاحظات، درجه ای از موازی سازی را در داخل و اطراف لایه های RNN امکان پذیر می کند:

موازی سازی داده ها: این شامل آموزش چندین مدل RNN یا چندین مینی دسته از داده ها به صورت موازی است. هر فرآیند موازی، یک کپی مستقل از RNN را آموزش می دهد یا زیرمجموعه متفاوتی از داده ها را پردازش می کند و سپس نتایج آنها (مثلاً از طریق میانگین گیری گرادیان) برای به روزرسانی مدل تجمیع می شود.

¹ Data Parallelism

پردازش دسته‌ای^۲: در حالی که گام‌های زمانی منفرد در یک توالی را نمی‌توان به طور کامل موازی کرد، عملیات در توالی‌های مختلف در یک دسته را می‌توان به صورت موازی انجام داد. به عنوان مثال، ضرب‌های ماتریس-بردار برای گیت ورودی، گیت فراموشی و گیت خروجی در یک LSTM را می‌توان به صورت موازی برای همه توالی‌ها در یک دسته اجرا کرد.

موازی‌سازی گروهی^۳: آموزش گروهی از مدل‌های RNN به صورت موازی، که در آن هر مدل بر جنبه یا زیرمجموعه متفاوتی از داده‌ها تمرکز می‌کند، و سپس ترکیب پیش‌بینی‌های آنها، می‌تواند نوعی محاسبات موازی با RNN‌ها در نظر گرفته شود.

^۲ Batch Processing

^۳ Ensemble Parallelism