

به نام خدا



دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده برق و کامپیوتر



تمرین کامپیوتری شماره ۴ مهلت تحویل : ۲۴ فروردین

در این تمرین قصد داریم با روش‌های مختلف تعیین ادات سخن (part-of-speech) به صورت دنباله‌ای آشنا شویم.

دادگان

داده مورد نیاز این تمرین در سامانه بارگذاری شده است. در این مجموعه داده جملات به همراه ادات سخن هر کلمه به صورت مشخص وجود دارد. این مجموعه شامل ۴۲ ادات سخن مختلف می‌باشد و در `pos_list` قابل مشاهده می‌باشد.

برای بخش `Rnn` از فایل `Rnn_train` برای آموزش و از `Rnn_test` برای تست استفاده کنید.

برای بخش `Hmm` از فایل `viterbi_train` برای آموزش و از `viterbi_test` برای تست استفاده کنید.

مدل

پس از پیاده‌سازی هریک از روش‌های زیر نحوه پیاده‌سازی و عملکرد مدل را در گزارش نهایی خود بیان کنید.

روش `RNN`: در این روش با استفاده از کتابخانه‌های موجود برای یادگیری عمیق می‌توانید عمل یادگیری را انجام دهید.

روش HMM : در این روش باید الگوریتم ویتربی پیاده‌سازی گردد. برای ساخت ماتریس گذار از بایگرم کلمات استفاده گردد. در این بخش سعی شود از روش‌های موجود برای برطرف کردن حل مشکل احتمال صفر استفاده شود و در گزارش نهایی توضیح داده شود.

ارزیابی

برای دو روش پیاده‌سازی شده موارد زیر را به صورت جداگانه گزارش کنید.

- برای ارزیابی هر یک از مدل‌های پیاده‌سازی شده گزارش صحت (accuracy) بر روی داده تست الزامی است.
- همچنین دو معیار دقت و فراخوانی (recall, precision) را برای هر یک از انواع ادات سخن موجود ارایه کنید.
- پس تعیین ادات سخن کلمات موجود در داده تست این تگ‌ها را در ستون مشخص شده در فایل تست ذخیره کنید و همراه گزارش خود ارسال کنید.

نکات تکمیلی

- گزارش نهایی، کدهای موجود و فایل خروجی را در سامانه بارگذاری کنید.
- گزارش نهایی شما از اهمیت بالایی برخوردار می‌باشد سعی شود تمامی موارد انجام شده به صورت کامل و در قالب گزارش علمی ارایه گردد.
- استفاده از تمامی کتابخانه‌های موجود برای پیاده‌سازی امکان پذیر می‌باشد اما نمونه برداری کامل از کدهای موجود در اینترنت و دانشجویان دیگر مشمول جریمه می‌گردد.
- برای سهولت در پیاده‌سازی بخش Rnn پیشنهاد می‌شود از [keras](#) استفاده کنید.

در صورت هرگونه سوال و یا ابهام از طریق ali96ebrahimi@ut.ac.ir با من در تماس باشید.