

به نام خدا



دانشگاه صنعتی اصفهان

دانشکده برق و کامپیوتر

پیشنهاد پروژه ی درس یادگیری ماشین

رشته ی مهندسی کامپیوتر / گرایش هوش مصنوعی و رباتیک

مدل نظارت شده ی ضعیف برای تشخیص مشترک احساس – موضوع از متن

مسعود فاطمی

جناب آقای دکتر پالهنهنگ

پاییز ۱۳۹۴

۱: هدف

هدف در این پروژه پیاده سازی یک مدل احتمالاتی برای تشخیص همزمان احساس و موضوع از داده های متنی می باشد. مدل نهایی و پیاده سازی شده یک مدل نظارت شده ی ضعیف می باشد که در دسته ی مدل های موضوعی قرار می گیرد.

۲: ورودی

پایگاه داده ای که در اینجا به عنوان ورودی برای مدل مورد استفاده قرار می گیرد نسخه ی ۲ پایگاه داده ی "بازبینی فیلم" (Movie Review) می باشد که شامل ۱۰۰۰ بازبینی مثبت و ۱۰۰۰ بازبینی منفی در مورد فیلم ها می باشد که از بایگانی سایت IMDB جمع آوری شده است با متوسط ۳۰ جمله در هر بازبینی.

۳: خروجی

خروجی که از این مدل بدست می آید توزیع احتمالاتی هر سند به همراه توزیع موضوعات و احساسات همراه با آن ها می باشد که از آن می توان برای کاربردهای مختلفی از جمله طبقه بندی یک مجموعه متن استفاده کرد.

۴: ابزار

برای پیاده سازی مدل مورد نظر از نرم افزار متلب استفاده می شود و کد های مورد نظر در محیط این نرم افزار نوشته می شوند.

۵: زمان بندی

- مطالعه ی منابع و مقالات مرتبط و مقاله ی اصلی و تسلط بر موضوع: هفته ی سوم آبان.

- پیاده سازی و انجام آزمایش هفته ی سوم آذر ماه.

- گزارش نویسی و ارائه ی پروژه هفته ی چهارم آذر ماه.

۶: روش یادگیری

در این مدل از یادگیری نظارت شده ی ضعیف استفاده می شود. برای بدست آوردن پارامترهای مدل از روش نمونه برداری گیبس (Gibbs Sampling) استفاده می شود که به تخمین پارامترهای مورد نیاز در فرآیند آموزش می پردازد.

۷: آزمایش ها

از خروجی بدست آمده از مدل نهایی برای طبقه بندی یک مجموعه متن بر اساس احساس آن ها استفاده می شود، که با برچسب زنی به هر سند آن را در دسته ی مثبت، منفی یا بی طرف قرار می دهد.

۸: روش پیاده‌سازی

برای پیاده‌سازی این روش همان‌طور که ذکر شد از نرم‌افزار متلب استفاده می‌شود. ابتدا مدل معروف تخصیص دیریکله ی پنهان (Latent Dirichlet Allocation) پیاده‌سازی شده و به مدل مربوطه گسترش داده می‌شود. فرآیند آموزش این مدل هم یک فرآیند نظارت شده ی ضعیف می‌باشد.