

بسم الله الرحمن الرحيم

پروژه تحلیل احساسات (Sentiment Analysis) درس NLP

موضوع پروژه : بررسی میزان شادی افراد در شب و در روز با استفاده از توییت های آن ها

این پروژه را بر روی توییت های افراد در توییتر انجام شده است. برای جمع آوری داده، با استفاده از کتابخانه `tweepy`، توییت هایی را با هشتگ های متفاوت به زبان فارسی از توییتر گرفتیم که تعداد آن ها تقریباً 10 هزار توییت بود (اگر هر توییت به طور میانگین 10 کلمه داشته باشد می شود 100 هزار کلمه) و بعد از نرمال کردن و تمیز کردن داده ها، آن ها را به دو دسته داده های مخصوص افراد فعال در شب (افرادی که شب از ساعت 1 تا 5 توییت می کنند) و افراد فعال در روز (افرادی که در روز از ساعت 7 صبح تا 11 توییت انجام می دهند) تقسیم کردیم. فایل `sentiment.py` مربوط به این بخش است.

داده ها را به صورت واحد واحد جدا کردیم که در اینجا هر واحد را یک توییت در نظر گرفتیم. سپس با استفاده از کتابخانه `hazm`، نقش کلمات را در آوردیم و با استفاده از این نقش ها، عبارات مهم (داری نقش اسم اسم و اسم صفت) را در هزار توییت اول فایل، در آورده و ذخیره نمودیم. و همچنین چند جفت کلمه را انتخاب کردیم و `polarity` هر عبارت را با استفاده از فرمول موجود در اسلایدها، به ازای هر کدام از آن جفت کلمات به دست آوردیم و در آخر از آن ها میانگین گرفتیم. فایل `polarity.py` مربوط به این بخش می باشد.

چون در فرمول `polarity`، `PMI` کلمه مثبت منهای `PMI` کلمه منفی می شود، بنابراین اگر این مقدار، بزرگتر از صفر شد، نتیجه میگیریم که افراد از شادی بیشتری در آن زمان برخوردارند.

نتایج `polarity` ها به ازای 4 جفت کلمه زیر به شکل زیر می باشد :

میانگین	خوب و بد	شاد و ناراحت	عشق و تنفر	خنده و گریه	
-1.03	-1.38	-1.56	-1.68	0.493	روز
0.32	0.119	0.392	0.392	0.392	شب

یعنی با توجه به نتایج به دست آمده مشخص میشود افرادی که در شب توییت می گذارند افراد شادتر و سرحال تر از افرادی هستند که روز توییت می گذارند.