

## ۱- نظر کاوی (opinion mining) در پایتون

نظر کاوی یکی از جدیدترین حوزه‌های پژوهشی در پردازش زبان طبیعی<sup>۱</sup>، بازیابی اطلاعات<sup>۲</sup> و متن کاوی<sup>۳</sup>، است. یکی از پیچیدگی‌های نظر کاوی ناشی از آن است که نظر افراد، یک حقیقت مسلّم (Facts) نیست بلکه امری ذهنی و حسی (Sentiment) است. به بیانی دیگر، نظرات بر خلاف حقایق (که اطلاعات صریح و مشخصی را در رابطه با یک موجودیت یا رویداد و ویژگی‌های آنها بیان می‌کنند)، معمولاً غیرشفاف و بصورت ضمنی، حسی، سلیقه‌ای، گرایش یا تجربه و دیدگاه یک شخص را درباره یک شیء یا رویداد، مشخص می‌نمایند.

## ۲- هدف پروژه:

هدف کلی: دریافت یک فایل متنی حاوی نظرات و مشخص کردن گرایش هر نظر به یکی از دو دسته منفی و مثبت.

هدف دیگر: علاوه بر مشخص کردن گرایش، یک امتیاز نیز محاسبه کند و درجه مثبت و منفی بودن یک نظر را مشخص کند.

هدف دیگر: نظر کاوی در سطح ویژگی. یعنی بگوییم نظر در رابطه با کدام ویژگی‌های محصول مثبت و برای کدام ویژگی‌ها منفی است. مثلاً دوربین یک موبایل مورد رضایت است اما باتری آن راضی کننده نبوده است.

## ۳- بررسی یک نمونه و ورودی و خروجی مطلوب:

[t]excellent phone , excellent service .

##i am a business user who heavily depend on mobile service .

phone[+3], work[+2]##there is much which has been said in other reviews about the features of this phone , it is a great phone , mine worked without any problems right out of the box .

##just double check with customer service to ensure the number provided by amazon is for the city / exchange you wanted .

at&t customer service[-2]##after several years of torture in the hands of at&t customer service i am delighted to drop them , and look forward to august 2004 when i will convert our other 3 family-phones from at&t to t-mobile !

[t].....

## شرح علائم:

[t] : نشان دهنده عنوان (title) یک نظر است. همچنین شروع یک نظر را مشخص می‌کند.

xxxx[+|-n] : نام یک ویژگی از محصول مورد نظر است و عدد درون براکت نشان می‌دهد که نظر فرد درباره این ویژگی مثبت و یا منفی است و همچنین امتیازی بین ۱ الی ۳ برای قدرت مثبت یا منفی بودن مشخص می‌کند.

## : شروع یک جمله را نمایش می‌دهد.

---

<sup>1</sup> Natural Language Processing

<sup>2</sup> Information Retrieval

<sup>3</sup> Text Mining

تذکر: از یک سیستم نظر کاوی واقعی انتظار داریم که خودکار ویژگی ها را تشخیص دهد و به آن ها امتیاز دهد. همچنین ابتدای جمله را خود تشخیص دهد. بنابراین در یک سیستم واقعی، علائم ## و XXXX[+|-n] وجود ندارند. در واقع نمونه فوق یک نمونه از جواب مورد قبول است که می توان از آن برای ارزیابی میزان صحت عملکرد سیستم طراحی شده استفاده کرد.

#### ۴- فاز اول پروژه:

تحقیق کنید که برای طراحی یک سیستم نظر کاوی چه مراحل باید برای پردازش متن یک نظر انجام شود. (حداقل یک و حداکثر ۲ صفحه) مهلت تحویل فاز اول: ۵ آبان ساعت ۲۴

#### ۵- فاز دوم پروژه:

الف) فرض کنید یک فایل ورودی متنی مانند نمونه ذکر شده (به همراه تمام علائم راهنما) داشته باشیم. یعنی فرض می کنیم که الگوریتمی دیگر وجود داشته که ویژگی های یک محصول را تشخیص و امتیازدهی کرده است. برنامه ای بنویسید که با استفاده از امتیاز ویژگی ها، گرایش کلی هر نظر را تعیین کند. همچنین پیشنهادی برای خریدن یا نخریدن محصول ارائه دهد. خروجی باید فایلی به فرمت زیر باشد:

1:

sentiment= (negative/positive)

score:

2:

sentiment = ...

score:

3: ...

Overall:

Recommendation: (buy/ do not buy) (با در نظر گرفتن تمام نظرات ارائه شده، آیا پیشنهاد خرید می دهید؟)

Recommendation score: (درصد) (پیشنهاد خریدن یا نخریدن می دهید؟)

Best feature: نام ویژگی ای که بیشترین محبوبیت را در میان کل نظرات داشته است. اگر بیش از یکی بودند، همه ( آنها را لیست کند)

Worst feature: (نام ویژگی که کمترین امتیاز (محبوبیت) را در مجموع نظرات داشته است.)

در واقع هر شماره نمایش دهنده یک نظر است. برای هر نظر باید گرایش و درجه (امتیاز) آن گرایش محاسبه شود. همچنین در بخش overall باید با توجه به مجموع نظرات درباره یک محصول به یک پیشنهاد خرید/نخریدن برسیم و برای آن پیشنهاد نیز یک درجه امتیاز محاسبه شود.

#### ۶- فاز سوم پروژه: نظر کاوی یک متن خام بدون علائم راهنما (بدون داشتن ویژگی ها و امتیازات آنها)

در این فاز دو فایل کمکی شامل لیست کلمات با گرایش مثبت positive و لیست کلمات با گرایش منفی negative در اختیار شما قرار میگیرد. در این فاز شما مجاز به استفاده از ویژگی‌ها و امتیازات مشخص شده در فایل نظرات نیستید. بلکه باید تعداد کلمات مثبت و منفی (دو فایل کمکی) را در نظرات جستجو کنید و به ازای هر کلمه مثبت و منفی به ترتیب امتیاز +1 و -1 در نظر بگیرید. و برای هر نظر یک امتیاز و گرایش محاسبه کنید. و سپس گرایش کلی محصول را ارزیابی کنید.

فرمت فایل خروجی متنی:

1:

sentiment= (negative/positive)

score:

2:

sentiment = ...

score:

3: ...

Overall:

Recommendation: (buy/ do not buy) (با در نظر گرفتن تمام نظرات ارائه شده، آیا پیشنهاد خرید می‌دهید؟)

Recommendation score: (با چه امتیازی (درصد) پیشنهاد خریدن یا نخریدن می‌دهید؟)

- برای سنجش دقت روش خود می‌توانید امتیازات و گرایش‌ها را با امتیازات و گرایش‌هایی که به عنوان خروجی فاز دوم محاسبه کردید، مقایسه کنید.
- ممکن است دقت نتایج در این مرحله، به خوبی فاز ۲ نباشد. تعجب نکنید!
- ممکن است هم خانواده یک کلمه از فایل‌های کمکی، در متن موجود باشد و نه خود آن کلمه. راهکاری برای حل این مشکل پیدا کنید.