

بررسی مدل‌های نظرکاوی و تجزیه و تحلیل احساسات کاربران در محیط وب

نجمه شاه‌طالبی^۱، محمدجواد کارگر^۲، کمال میرزائی^۳

^۱ دانشجوی کارشناسی ارشد گروه مهندسی کامپیوتر، واحد میبد، دانشگاه آزاد اسلامی، میبد، ایران
n.shahtalebi@gmail.com

^۲ عضو هیأت علمی گروه مهندسی کامپیوتر، واحد میبد، دانشگاه آزاد اسلامی، میبد، ایران
showcaran@gmail.com

^۳ عضو هیأت علمی گروه مهندسی کامپیوتر، واحد میبد، دانشگاه آزاد اسلامی، میبد، ایران
k.mirzaie@maybodiau.ac.ir

چکیده

بدست آوردن نظرات مردم و مشتریان خود یک تجارت عظیم برای بازاریابی، روابط اجتماعی و حتی رقابت‌های انتخاباتی می‌باشد. در گذشته این اطلاعات از طریق برکه‌های نظرسنجی و یا آزمایش بر روی یک گروه خاص بدست می‌آمد ولی امروزه با گسترش روزافزون رسانه‌های اجتماعی (مثل بلاگ‌ها، توئیتر و فیسبوک) افراد و سازمان‌ها از محتوای این رسانه‌ها برای تصمیم‌گیری استفاده می‌کنند. این در حالی است که در دنیای واقعی شرکت‌ها و سازمان‌ها همواره به دنبال این هستند که نظرات مردم یا مشتریان را درباره‌ی محصولات و یا خدمات خود بدانند، از طرفی مشتریان نیز خواستار بهره‌مند شدن از نظرات سایر کاربران در رابطه با آن محصولات می‌باشند تا بتوانند بر اساس آن اقدام به خرید یا عدم خرید آن محصول کنند. در اینجاست که نیاز به یک سیستم جهت شناسایی و تجزیه و تحلیل احساسات در نظرات کاربران احساس می‌شود. در این مقاله مدل‌های مختلف تجزیه و تحلیل احساسات در سطح ویژگی مورد بررسی قرار گرفت و نقاط قوت و ضعف هر یک از مدل‌ها به طور کامل تشریح شد.

کلمات کلیدی

تجزیه و تحلیل احساسات، مدل‌های نظرکاوی، استخراج ویژگی‌ها، خلاصه‌سازی نظرات

وب‌پژوهی تلاش کرده‌اند تا روش‌های مختلفی در زمینه‌ی دسترسی و تجزیه و تحلیل احساسات ارائه دهند.

۱- مقدمه

منابع و صفحات زیادی از نظرات با حجم بالا در اینترنت موجود می‌باشد و رسیدن به نظرات دلخواه در این صفحات طولانی، یافتن جملات مورد نظر و خلاصه‌سازی و جمع‌بندی آنها کار هر خواننده‌ای نیست، همچنین یافتن اطلاعات احساسی در وب، دسته‌بندی کردن و نظم دادن به آنها کار آسانی نمی‌باشد [۱، ۲، ۳]. بنابراین ابزار خودکاری برای کشف و خلاصه کردن نظرات لازم است. با توجه به این نیاز، مفهوم کاهش نظرات یا تجزیه و تحلیل

پیشرفت فناوری ارتباطات و اطلاعات زمینه‌ای را فراهم کرده تا افراد در سراسر دنیا بتوانند از خدمات و محصولات شرکت‌ها و سازمان‌ها بهره‌مند شوند. در همین خصوص کاربران تلاش می‌کنند تا بتوانند از نظرات سایر مصرف‌کنندگان در رابطه با کالاها و خدمات بهره‌مند شده و در نهایت تصمیم درستی در رابطه با خرید یک کالا بگیرند. این در حالی است که متخصصان



اسم‌هایی که تعداد تکرار زیاد در متن دارند را می‌توان به عنوان ویژگی در نظر گرفت. با این حال مشخص است که همه اسامی با تعداد تکرار زیاد لزوماً ویژگی نیستند. بنابراین خروجی روش‌های مبتنی بر تکرار برای رسیدن به لیستی دقیق‌تر به عملیات فیلترسازی خواهد داشت. مهم‌ترین مزیت روش‌های مبتنی بر تکرار این است که این روش‌ها ساده و در واقعیت کارا هستند. اما مشکل این روش‌ها این است که در فرایند استخراج ویژگی، تعداد زیادی ویژگی‌های زاید تولید می‌کنند، همچنین ویژگی‌هایی که تعداد تکرار آنها در متن کم باشند را نمی‌توانند استخراج کنند. یکی دیگر از مشکلات این روش‌ها این است که با اینکه این روش‌ها می‌توانند به صورت بدون نظارت یا با نظارت ضعیف تعریف شوند، زمانی که این روش‌ها از یک دامنه موضوع به دامنه موضوع دیگر برده می‌شوند نیاز به تنظیم دوباره دستی بسیاری از پارامترها (مانند مقادیر آستانه) دارند. تا به امروز روش‌های مختلفی به صورت مبتنی بر تکرار ارائه شده است.

۲-۱-۱- مدل کاوش و خلاصه‌سازی نظرات کاربران

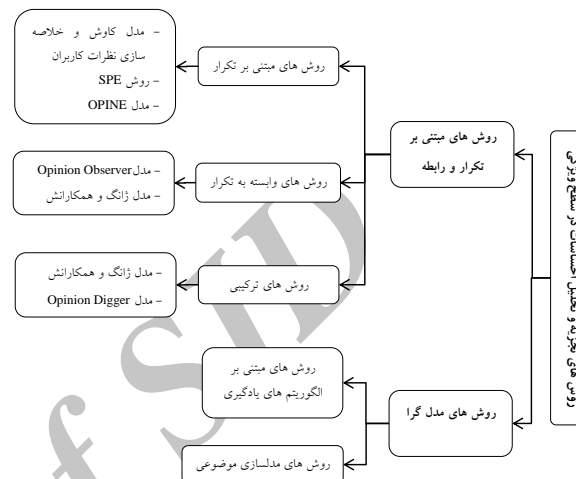
اولین مدل در زمینه استخراج ویژگی و احساس از متون، مدل کاوش و خلاصه‌سازی نظرات کاربران، ارائه شده توسط هو^۱ و لیو^۲ در سال ۲۰۰۴، می‌باشد که بر اساس روش‌های مبتنی بر تکرار تعریف شده است [۲۳، ۱۰]. این مدل ویژگی‌های بارز و کلمات احساس را استخراج می‌کند و برای تشخیص گرایش احساس کلمات احساس، از لیستی از کلمات احساس برچسب خورده شامل ۳۰ کلمه استفاده می‌کند. ایده این مدل این است که ویژگی‌ها، کلمات یا عباراتی هستند که دارای تعداد تکرار بالا هستند به طوری که در بسیاری از متون تکرار شده باشند و یا دارای تعداد تکرار خیلی کم هستند با این تفکر که یک کلمه با تعداد تکرار کم، به احتمال زیاد یک ویژگی خواهد بود. دو گام مهم و اصلی که در مدل کاوش و خلاصه‌سازی نظرات کاربران برای استخراج و احساس انجام می‌شود شامل:

یافتن اسم‌ها و عبارات اسمی با تکرار بالا به عنوان ویژگی: در این مدل، اسامی و عبارات اسمی با استفاده از یک برچسب زن (POS) به عنوان ویژگی‌های نامزد استخراج می‌شوند. سپس تعداد تکرار آنها محاسبه شده و ویژگی‌های نامزد با تکرار بالا بر اساس یک حد آستانه تعریف شده نگهداری می‌شوند. ایده نهفته در این روش این است که وقتی کاربران راجع به یک شی نظر می‌دهند معمولاً از واژگان مشابه به هم استفاده می‌کنند، بنابراین ویژگی‌هایی که کاربران زیادی در رابطه با آن‌ها نظر داده باشند ویژگی‌های مناسبی برای استخراج خواهند بود. ضمناً اسامی با تعداد تکرار کم با احتمال کمتری به عنوان ویژگی در نظر گرفته می‌شوند.

یافتن ویژگی‌های با تعداد تکرار کم: از طریق استخراج رابطه بین ویژگی‌ها و کلمات دارای احساس: در گام اول ممکن است بعضی ویژگی‌های مهم که دارای تعداد تکرار کم هستند از دست برود. گام دوم در مدل کاوش و خلاصه‌سازی نظرات کاربران برای طرف کردن مشکل گام اول است. برای حل این مشکل، ایده‌ای که در اینجا از آن استفاده می‌شوند این است که بر اساس گام اول، پس از استخراج ویژگی با تعداد تکرار بالا، کلمه احساس متناظر آن و به کار رفت در متن نیز استخراج می‌شود، سپس با توجه به این نکته که هر کلمه احساسی که با یک ویژگی با تکرار بالا به کار رفته باشد با

احساسات در علم داده کاوی پدید آمده است. تجزیه و تحلیل احساسات مسأله‌ای در متن کاوی و پردازش زبان طبیعی است که به کاوش اطلاعات متنی دارای احساس، نظر یا عقیده می‌پردازد [۴، ۵].

بر همین اساس تا به حال روش‌های مختلفی در زمینه تجزیه و تحلیل احساسات در سطح ویژگی ارائه شده است [۲۲-۶]. در حالت کلی می‌توان این روش‌ها را به دو دسته‌ی روش‌های مبتنی بر تکرار و روش‌های مدل‌گرا به صورت زیر طبقه‌بندی کرد (شکل ۱):



شکل (۱): طبقه‌بندی تجزیه و تحلیل احساسات در سطح ویژگی

همان طور که گفته شد روش‌های تجزیه و تحلیل احساسات به دو دسته تقسیم‌بندی شده است که به علت گستردگی روش‌ها در این مقاله، تنها به ارایه و توضیح روش‌های مبتنی بر تکرار و رابطه پرداخته خواهد شد.

۲- روش‌های مبتنی بر تکرار و رابطه

اکثر روش‌هایی که در مسأله تجزیه و تحلیل احساسات در سطح ویژگی ارائه شده‌اند، مبتنی بر روش‌های تکرار می‌باشند [۸، ۱۲-۱۰، ۲۳]. این روش‌ها معمولاً محدودیت‌هایی را بر روی عبارات اسمی با تعداد تکرار بالا برای شناسایی ویژگی‌ها و احساسات قرار می‌دهند، در حالی که روش‌های وابسته به رابطه از روابط بین کلمات در متن استفاده می‌کنند تا ویژگی و احساسات را تشخیص دهند [۲۶-۲۴، ۱۳]. مطالعاتی در زمینه ترکیب این دو روش نیز انجام شده است که معمولاً بر اساس روابط بین کلمات حاصل از خروجی روش مبتنی بر تکرار، ویژگی‌ها و احساسات را فیلتر می‌کنند [۲۷، ۱۴]. در ادامه چند نمونه از روش‌های مبتنی بر تکرار شامل مدل کاوش و خلاصه‌سازی نظرات کاربران، روش اس پی ای و مدل OPINE تشریح خواهد شد.

۲-۱- روش‌های مبتنی بر تکرار

روش‌های مبتنی بر تکرار معمولاً از یک سری محدودیت‌ها بر روی اسم‌ها و عبارات اسمی با تعداد تکرار زیاد برای شناسایی ویژگی‌ها استفاده می‌کنند. اگرچه یک ویژگی ممکن است اسم، صفت، فعل یا قید در جمله باشد، اما معمولاً نزدیک به ۷۰ درصد ویژگی‌ها اسامی بارز در متن هستند [۱]. بنابراین



ابزار وردنت^{۱۳} و اطلاعات ریخت‌شناسی، لیست ویژگی‌های نهایی را برای کلاس محصول به‌دست می‌آورد. در ادامه مدل OPINE از این لیست ویژگی‌ها استفاده کرده تا لیست کلمات احساس را استخراج کند.

۳-۱-۲- روش‌های وابسته به رابطه

رویکردی که در روش‌های وابسته به رابطه وجود دارد این است که می‌توان از طریق ارتباط بین کلمات در متون، ویژگی‌ها و احساسات را استخراج کرد. فرضی که در این روش‌ها از آن تبعیت می‌شود این است که هر احساس به کار رفته در متن لزوماً درباره یک ویژگی خواهد بود. بنابراین از طریق تعریف ارتباط بین احساس و ویژگی می‌توان ویژگی‌های جدیدی از متون استخراج کرد. این روش‌ها در مقایسه با روش‌های مبتنی بر تکرار قابلیت استخراج ویژگی‌های با تعداد تکرار کم را از متون دارند. ضمناً این روش‌ها مشابه روش‌های مبتنی بر تکرار در فرایند استخراج ویژگی، تعداد زیادی ویژگی زاید تولید می‌کنند [۲۵، ۲۶].

۳-۱-۱- مدل Opinion Observer

در سال ۲۰۰۵، لیو و همکارانش مدل Opinion Observer را که از معروف‌ترین روش‌های وابسته به رابطه است، ارائه دادند [۲۸]. این مدل بر روی متون نظری کوتاه ارائه شده است و به عنوان خروجی، محصولات مختلف را بر روی ویژگی‌هایشان مقایسه می‌کند. مدل Opinion Observer از قوانین کاوش نظارتی برای استخراج ویژگی‌ها استفاده کرده است به طوری که در این مدل ابتدا یک مجموعه داده به طور دستی برچسب‌گذاری می‌شود. برای این برچسب‌گذاری ابتدا برچسب‌زنی POS انجام می‌شود و اعداد از مجموعه متون حذف می‌شوند. سپس ویژگی‌های واقعی در نظرات با عبارت [feature] جایگزین می‌شوند، جملات به بخش‌های مختلفی که شامل n-gram باشند تقسیم شده و اگر در یک جمله برچسب‌های یکسان و تکراری وجود داشته باشد این برچسب‌ها شماره‌گذاری می‌شوند. در انتها کلمات در متون ریشه‌یابی^{۱۴} می‌شوند. با انجام این عملیات مجموعه داده مورد نظر برای آموزش آماده شده است. سرانجام لیو و همکارانش از الگوریتم کاوش قوانین انجمنی برای استخراج ویژگی‌ها استفاده می‌کنند. با توجه به این که مدل Opinion Observer بر روی متون نظری کوتاه ارائه شده است برای استخراج کلمات احساس فرض را بر این می‌گذارد که کاربران نظرات مثبت و منفی خود را در متونی جداگانه ارائه می‌دهند و نیاز به استخراج آن‌ها توسط مدل وجود ندارد.

۳-۱-۲- مدل ژانگ^{۱۵} و همکارانش

مدل ژانگ و همکارانش از ایده ارتباط بین کلمات احساس و ویژگی‌ها بر اساس روابط وابستگی نحوی برای استخراج ویژگی‌ها و احساسات استفاده کرده است [۲۶]. این مدل، روابط وابستگی را برای شناسایی ویژگی‌ها و کلمات احساس از متون نظری مربوط به فیلم‌ها به کار برده است. شکل (۲) ساختار مدل معرفی شده توسط ژانگ و همکارانش را معرفی می‌کند.

یک ویژگی با تکرار کم نیز ممکن است به کار رفته باشد، ویژگی‌های با تعداد تکرار کم استخراج می‌شوند.

برای مثال اگر کلمه "picture" به عنوان یک ویژگی با تعداد تکرار بالا از جمله زیر استخراج شده باشد:

"The pictures are amazing"

می‌توان به این نتیجه رسید که کلمه "amazing" دارای احساس متناسب با ویژگی استخراج شده است، بنابراین از طریق این کلمه احساس می‌توان ویژگی با تعداد تکرار کمی مانند "software" را از جمله زیر استخراج کرد.

"The software is amazing"

مدل کاوش و خلاصه‌سازی نظرات کاربران برای یافتن ویژگی‌های با تعداد تکرار بالا از الگوریتم کاوش قوانین انجمنی^{۱۶} با حداقل میزان پشتیبان^{۱۷} یک درصد استفاده می‌کند. همچنین این مدل برای استخراج کلمات دارای احساس، از لیست از پیش تعریف شده‌ای از کلمات احساس بهره گرفته و عبارات با برچسب POS از نوع صفت را از متن استخراج می‌کند. این مدل پس از تهیه لیست ویژگی‌ها از دو روش هرس کردن^{۱۸} استفاده می‌کند تا به لیست خروجی بهتری برسد.

۳-۱-۲- روش SPE^{۱۹}

در سال ۲۰۱۰، Wei و همکارانش روشی را برای استخراج ویژگی‌ها و تجزیه و تحلیل آنها ارائه دادند [۱۱]. این روش بر اساس مدل مبتنی بر تکرار کاوش و خلاصه‌سازی نظرات کاربران [۱۰] ارائه شده و سعی در بهبود نتایج آن دارد. روش پیشنهادی Wei و همکارانش، SPE نامیده شده است به طوری که مدل کاوش و خلاصه‌سازی نظرات کاربران را بدین صورت توسعه می‌دهد که لیستی از کلمات احساس مثبت و منفی را بر اساس یک فرهنگ لغت به عنوان واژگان احساس استخراج کرده و بر اساس این لیست ویژگی‌هایی که زاید و یا نامربوط به کلمات احساس باشند را هرس می‌کند. در نهایت این روش چنین نتیجه می‌گیرد که استفاده از لیستی برچسب‌گذاری شده از کلمات احساس در کنار فرایند استخراج ویژگی و احساس باعث می‌شود ویژگی‌هایی مرتبط با دامنه موضوع به‌دست آیند.

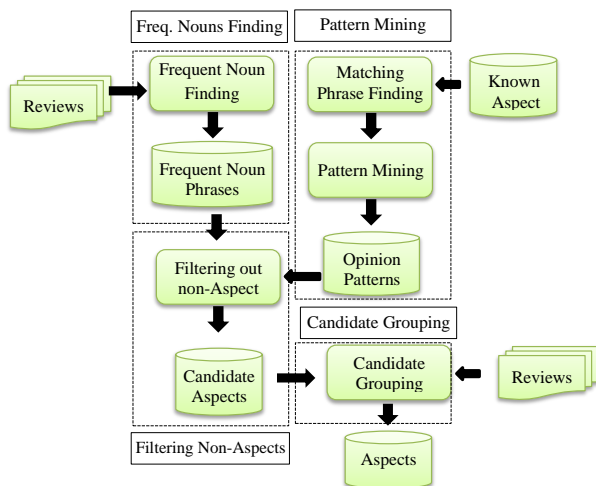
۳-۱-۲- مدل OPINE

یکی از معروف‌ترین مدل‌هایی که در زمینه استخراج ویژگی و کلمه احساس معرفی شده است مدل OPINE می‌باشد که در سال ۲۰۰۵ توسط پاپ اسکویو^{۱۰} و اتزیونی^{۱۱} ارائه شده است [۱۲]. این مدل برای استخراج ویژگی‌های بارز طراحی شده است. برای این کار ابتدا کلاس‌هایی از محصولات تعریف شده سپس برای هر مجموعه داده نظری بخش‌ها و خصوصیات کلاس محصول داده شده شناسایی می‌گردند تا جایی که دیگر نامزدی برای ویژگی‌ها وجود نداشته باشد. برای این کار، OPINE ابتدا عبارات اسمی با تعداد تکرار بالا را استخراج کرده سپس بر مبنای یک معیار ارزیابی تعریف شده مبتنی بر معیار اطلاعات متقابل PMI^{۱۱} رابطه بین ویژگی استخراج شده و خصوصیات تعیین شده، کلاس محصول را استخراج می‌کند. مدل OPINE با استفاده از

عمل می کند. در این الگوریتم از روش رتبه بندی تعریف شده برای صفحات وب HITS^{۱۶} استفاده می کند.

۴-۱-۲- مدل Opinion Digger

در سال ۲۰۱۰، مدل Opinion Digger توسط مقدم و استر بر مبنای روش ترکیبی از مبتنی بر تکرار و رابطه معرفی شده است [۲۷]. ایده این مدل این است که از مزایای هر دو روش مبتنی بر تکرار و رابطه استفاده کند. شکل (۳) ساختار این مدل را برای استخراج ویژگی نشان می دهد.

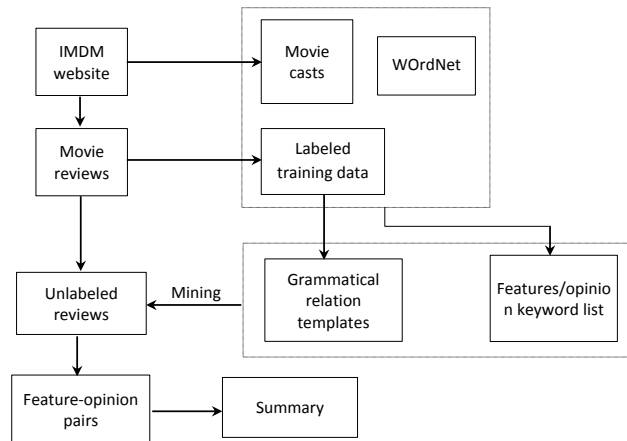


شکل (۳): ساختار مدل Opinion Digger بر اساس ترکیب روش های مبتنی بر تکرار و رابطه [۲۷]

همان طور که در شکل (۳) نشان داده شده، مدل Opinion Digger پس از انجام پیش پردازش های اولیه، ویژگی های با تکرار بالا را از متون استخراج می کند. سپس با استفاده از الگوهای نحوی از قبل تعریف شده ای این لیست ویژگی های نامزد را فیلتر می کند. در ادامه ویژگی های نامزد به دست آمده را گروه بندی کرده و به هر گروه، یک برجسب نسبت می دهد. روش های مختلفی مبتنی بر فرهنگ لغات، برای گروه بندی ویژگی ها ارائه شده است. با این حال مدل Opinion Digger در گروه بندی ویژگی ها کلمات مترادف را در یک گروه قرار می دهد. برای این کار، این مدل از روش MMR^{۱۷} برای این عمل گروه بندی استفاده می کند. روش MMR برای رتبه بندی اسناد متنی در بازایی اطلاعات تعریف شده است به طوری که تلاش می کند ارتباط بین دو متن را بیشینه کند در حالی که شباهت کمینه شود [۲۷، ۲۹].

۵- بحث و نتیجه گیری

در زمینه پردازش زبان طبیعی تلاش می شود از طریق فرآیندهای محاسباتی خودکار، به درک بهتری از زبان طبیعی دست یافت. در این موارد هدف این است که زبان طبیعی را با یک قالب ساختارمند نمایش دهند، به طوری که برای تحلیل توسط کامپیوتر ساده تر باشد. در همین راستا نظر کاوی و تحلیل احساسات استخراج نظرات کاربران و تشخیص قطبیت آنها، درون متون ذهن-گرایانه می پردازد. به دلیل افزایش نظرات آنلاین مشتریان در مورد کالاها در



شکل (۲): ساختار مدل ژانگ و همکارانش [۲۶]

همان طور که شکل (۲) نشان می دهد در این مدل ابتدا کلمات در یک جمله با استفاده از تجزیه وابستگی نحوی پیش پردازش شده سپس روابط بین آن ها در قالب گراف وابستگی مشخص می شود.

۴- روش های ترکیبی

روش های ترکیبی، خصوصیات دو روش مبتنی بر تکرار و وابسته به رابطه را با هم ترکیب می کنند. این روش ها بر اساس روابط بین کلمات حاصل از خروجی روش مبتنی بر تکرار، ویژگی ها و احساسات را فیلتر می کنند. این روش ها بر اساس روابط بین کلمات حاصل از خروجی روش مبتنی بر تکرار، ویژگی ها و احساسات را فیلتر می کنند. ایده ای که در این روش ها وجود دارد این است که ویژگی ها معمولاً اسمی و عبارات اسمی با تعداد تکرار بالا هستند که همراه با کلمات احساس در جملات به کار می روند [۲۷، ۱۴]. در مقایسه با روش های مبتنی بر تکرار و وابسته به رابطه، روش های ترکیبی تعداد کمتری ویژگی زاید تولید می کنند. با این حال ممکن است ویژگی های با تعداد تکرار کم را از دست دهند. یکی دیگر از مشکلات این روش ها این است که زمانی که این روش ها از یک دامنه موضوع به دامنه موضوع دیگر برده می شوند نیاز به تنظیم دوباره دستی بسیاری از پارامترها (مانند مقادیر آستانه) دارند [۱].

۴-۱-۱- مدل ژانگ و همکاران

مدلی که توسط ژانگ و همکارانش ارائه شده، بر مبنای روشی ترکیبی برای استخراج ویژگی ها تعریف شده است [۱۴]. اساس این مدل بر مبنای مدل Double Propagation تعریف شده است [۱۳]. مدل Double Propagation به دنبال استخراج مجموعه ای از کلمات احساس و ویژگی ها می باشد که برای این کار از مجموعه ساده از قبل تعریف شده ای برای کلمات احساس استفاده می کند. این مدل روابط ساختاری بین کلمات احساس و ویژگی ها را بر اساس گرامر وابستگی شناسایی می کند. مدل ژانگ و همکارانش از دو الگوی "part-whole" و "no" برای بهبود مدل Double Propagation در استخراج ویژگی ها استفاده می کند. در این مدل از یک الگوریتم رتبه بندی برای مرتب کردن ویژگی های استخراج شده استفاده می شود. این الگوریتم رتبه بندی بر اساس تعداد تکرار ویژگی ها و ارتباط آن ها

- Systems and E-Business Management, Vol. 8, No. 2, pp. 149-167, 2010.
- [12] Popesc A. M. and Etzioni O., "Extracting Product Features and Opinions from Reviews," in *Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)*, Vancouver, pp. 339-346, 2005.
- [13] Qiu G., Liu B., Chen C., "Opinion word expansion and target extraction through double propagation", *Computational linguistics*, vol. 37, No.1, pp. 9-27, 2011.
- [14] Zhang L., Liu B., Lim S. H., and Eamonn O'Brien-Strain., "Extracting and ranking product features in opinion documents", In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters, COLING' 10*, Stroudsburg, PA, USA, pp. 1462-1470, 2010.
- [15] Somprasertsri G., Lalitrojwong P., "Automatic product feature extraction from online product reviews using maximum entropy with lexical and syntactic features", In *Proceedings of IEEE International Conference on Information Reuse and Integration*, pp. 250-255, 2008.
- [16] Lin C., He Y., "Join sentiment/topic model for sentiment analysis", In *Proceedings of ACM International Conference on Information and Knowledge Managemnt (CIKM-2009)*, pp. 375-384, 2009.
- [17] Lin C., He Y., Everson R., Ruger S., "Weakly supervised joint sentiment-topic detection fro, text", *IEEE Transactions on Knowledge and Data Engineering.*, vol. 24, No. 6, pp. 1134-1145, 2012.
- [18] Jo Y., A. H., "Aspect and sentiment unification model for online review analysis", In *Proceedings of the fourth ACM nternational Conference on Web search and data mining*, pp. 815-824, 2011.
- [19] Zhai Z., Liu B., Xu H., Jia P., "Constrained LDA for grouping product features in opinion mining", In *Proceedings of 15th Pacific-Asia Conference, Advances in Knowledge Discovery and Data Mining*, pp. 448-459, 2011.
- [20] Zhao W., Jiang J., Yan H., Li X., "Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid", In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2010)*, Association for Computational Linguistics, pp. 56-65, 2010.
- [21] Titov I., McDonald R., "Modeling online reviews with multigrain topic models", In *Proceeding of ACM International Conference on World Wide Web (WWW-2008)*, pp. 111-120, 2008.
- [22] Mei Q., Ling X., Wondra M., Su H., Zhai C., "Topic sentiment mixture: modeling facts and opinion in weblogs", In *Proceeding of ACM International Conference on World Wide Web (WWW-2007)*, pp. 171-180, 2007.
- [23] Hu, M. ; Liu, B., "Mining and Summarizing Customer Reviews.", *The ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004)*, 2004.
- [24] Guo H., Zhu H., Guo Z., Zhang X., Su Z., "Product feature categorization with multilevel latent semantic association", In *Proceeding of ACM International Conference on Information and Knowledge Management (CIKM-2009)*, pp. 1087-1096, 2009.
- [25] Liu B., Hu M. and Cheng J., "Opinion Observer: Analyzing and Comparing Opinions on the Web," in *WWW*, Chiba, Japan, 2005.
- محیط وب، نیاز به سازمان‌دهی این نظرات، متن‌کاوی و کشف احساسی متون، توجه فزاینده‌ای را در سال‌های اخیر به خود معطوف کرده است. در این میان تکنیک‌های خلاصه‌سازی خودکار متن نیز با هدف ایجاد یک قالب فشرده و خلاصه از اسناد که شامل جنبه‌های کلیدی سند اصلی است، ایجاد شده‌اند. این روش‌ها معمولاً بر اساس انتخاب و استخراج جملاتی که دارای اطلاعات مهم تر هستند تشکیل می‌شوند و سپس یک خلاصه از مجموع این جملات استخراج می‌شود. برای استخراج جملات و تجزیه و تحلیل احساسات در سطح ویژگی، از روش‌های مختلفی می‌توان بهره‌مند شد و در این میان متخصصان وب‌پژوهی تلاش نموده‌اند تا روش‌های مختلفی در این زمینه ارایه دهند. از جمله مهم‌ترین این روش‌ها، روش مبتنی بر تکرار و رابطه و روش مدل‌گرا می‌باشد. با توجه به گستردگی و اهمیت روش مبتنی بر تکرار و رابطه تلاش شد تا زیرمجموعه‌های موجود در این روش به طور کامل تشریح شده و مراحل انجام، نقاط قوت و ضعف و ویژگی‌های هر یک مورد بررسی قرار گیرد.

مراجع

- [1] Liu B., Zhang L., *A survey of opinion mining and sentiment analysis*, Mining Text Data, pp. 415-463, 2012.
- [2] Turney, P., "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews", In *Proceeding of Annual Meeting of the Association for Computational Linguistics (ACL-2002), The 40th Annual*, Philadelphia, 2002, pp. 417-424, 2002.
- [3] Pang, B; Lee, L; Vaithyanathan, S., "Thumbs up?: Sentiment Classification Using Machine Learning Techniques", In *Proceeding of the ACL-02 Conference on Empirical Methods In Natural Language Processing* , pp. 79-86, 2002.
- [4] Pang B., Lee L., *Opinion mining and sentiment analysis. J Found Trends Inf Retrieval* 2(1-2):1 135, 2008.
- [5] Reyes A., Rosso P., "Making objective decisions from subjective data: Detecting irony in customer reviews", *Decision Support System*, vol. 53, pp. 754 - 760, 2012.
- [6] Thet T. T., Na J. C., Khoo C.S., "Aspect-based sentiment analysis of movie reviews on discussion boards", *Journal pf Information Science*, vol. 36, No.6, pp. 823 - 848, 2010.
- [7] Titov I., McDonald R., "A joint model of text and aspect rating for sentiment summarization", In *Proceeding of the Annual Meeting on Association for Computational Linguistics and the Human Language Technology Conference (ACL-HLT)*, pp. 308-316, 2008.
- [8] Zhu J., Wang H., Zhu M., Tsou B.K., Ma M., "Aspect-based opinion polling from customer reviews", *IEEE Transactions on Affective Computing*, Vol. 2, No. 1, pp. 37-49, 2011.
- [9] Brody S., Elhadad N., "An unsupervised aspect-sentiment model for online reviews", In *Proceeding of Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 804-812, 2010.
- [10] Hu M., Liu B., "Mining opinion features in customer reviews", In *Proceedings of 19th National Conference on Artificial Intelligence, AAAI Press*, pp. 755-760, 2004.
- [11] Wei, C. P., Chen Y. M., Yang C.S., "Understanding what concerns consumers: a semantic approach to product feature extraction from consumer reviews", *Information*

- [26] Zhuang L., Jing F., Zhu X., “Movie review mining and summarization”, *In Proceeding of the ACM International Conference on Information and Knowledge Management (CIKM-2006)*, pp. 43-50, 2006.
- [27] Moghaddam S., Ester M., “Opinion digger: an unsupervised opinion miner from unstructured product reviews”, *In Proceedings of ACM International conference on Information and Knowledge Management (CIKM-2010)*, pp.1035-1045, 2010.
- [28] Wang X., McCallum A., “A note on topical n-grams”, Technical Report UM-CS-071, Department of Computer Science University of Massachusetts Amherst, 2005.
- [29] Zhang L., Liu B., “Aspect and Entity Extraction for Opinion Mining”, *Data Mining and Knowledge Discovery for Big Data*, pp. 1-40. Springer, 2014.

زیر نویس ها

- 1 Hu
- 2 Liu
- 3 Frequent Features
- 4 Part-Of-Speech Tagging
- 5 Infrequent features
- 6 Association rule mining
- 7 Support
- 8 Aspect Pruning
- 9 Semantic-based Product feature Extraction
- 10 Popescu
- 11 Etzioni
- 12 Pointwise Mutual Information
- 13 WordNet
- 14 Stemming
- 15 Zhuang
- 16 Hyperlink-Induced Topic Search
- 17 Maximal Marginal Relevance