

This file has been cleaned of potential threats.

If you confirm that the file is coming from a trusted source, you can send the following SHA-256 hash value to your admin for the original file.

b5dbcc3fce3e73609259214ed47e6787a9f7cbe0aca93916c164f0476f610743

To view the reconstructed contents, please SCROLL DOWN to next page.



پایان‌نامه کارشناسی ارشد

گروه مهندسی کامپیوتر

نظر کاوی بر اساس استخراج و تطبیق الگوهای حسی جملات فارسی

نگارنده:

سید محمد اصغری نکاح

استاد راهنما:

دکتر محسن کاهانی

اردیبهشت ۹۴

سورة الاحقاف

اظهارنامه

اینجانب سید محمد اصغری نکاح دانشجوی دوره کارشناسی ارشد رشته مهندسی کامپیوتر دانشکده مهندسی دانشگاه فردوسی مشهد نویسنده پایان‌نامه نظر کاوی براساس استخراج و تطبیق الگوهای حسی جملات فارسی تحت راهنمایی جناب آقای دکتر محسن کاهانی متعهد می‌شوم:

- تحقیقات در این پایان‌نامه توسط اینجانب انجام شده و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهشهای محققان دیگر به مرجع مورد استفاده استناد شده است.
- مطالب مندرج در پایان‌نامه تاکنون توسط خود و یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است.
- کلیه حقوق معنوی این اثر متعلق به دانشگاه فردوسی مشهد می‌باشد و مقالات مستخرج با نام "دانشگاه فردوسی مشهد" و یا "*Ferdowsi University of Mashhad*" به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان‌نامه تاثیرگذار بوده‌اند در مقالات مستخرج از رساله رعایت شده است.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است، اصل رازداری، ضوابط و اصول اخلاق انسانی رعایت شده است.

تاریخ

امضای دانشجو

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه‌های رایانه‌ای، نرم‌افزارها و تجهیزات ساخته شده) متعلق به دانشگاه فردوسی مشهد می‌باشد. این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در پایان‌نامه بدون ذکر مرجع مجاز نمی‌باشد.

صور تجلسه دفاع از پایان نامه کارشناسی ارشد

جلسه دفاع از رساله آقای سید محمد اصغری نکاح دانشجوی دوره کارشناسی ارشد رشته مهندسی کامپیوتر گرایش نرم افزار در ساعت روز در محل دانشکده مهندسی با حضور امضا کنندگان ذیل تشکیل گردید. پس از بررسی هیأتی ای لازم، هیأت داوران پایان نامه نامبرده را با نمره به عدد به حروف و با درجه مورد تأیید قرار داد.

عنوان رساله

نظر کاوی بر اساس استخراج و تطبیق الگوهای حسی جملات فارسی

امضای هیئت داوران:

..... استاد راهنما: دکتر

..... داور: دکتر

..... داور: دکتر

..... نماینده تحصیلات تکمیلی: دکتر

..... مدیر گروه: دکتر

سپاسگزاران و ندی که در ذره ذره عالم پیداست و بدون لطف و مهربانی او قلم برکتی نداشت. از آنها که به مصداق حدیث: کسی که شکرگزار مخلوق نباشد، شکرگزار خالق نیز نیست؛ بر خود لازم می‌دانم که از زحمات تمام کسانی که من را در نگارش این پایان‌نامه یاری نموده‌اند تشکر نمایم.

تقست لازم است از زحمات **پدر و مادر عزیزم** در طول تمام این سال‌ها تشکر کنم. پدری که نشانه رحمت و مادری که نماد مهربانی خداوند در زمین هستند. از همه تلاش‌های این دو عزیز زندگی ام تشکر می‌کنم و این سعی بسیار کوچکم را به نشانه قدرشناسی از دریای محبت‌هایشان به ایشان تقدیم می‌کنم.

همچنین باید قدردانی خود را نسبت به **جناب پروفسور کاهانی** اعلام نمایم. نه تنها به علت راهنمایی‌های علمی ارزشمند ایشان در طول نگارش این پایان‌نامه؛ بلکه به خاطر فروتنی و افلاق ایشان. از ایشان تشکر می‌کنم که بدون تاثیر از مراتب علمی و دانشگاهی همچنان دلسوزانه و مخلصانه در جهت ارتقای علمی جامعه تلاش می‌کنند و بنده را نیز مورد راهنمایی‌های ارزشمند خود قرار دادند.

در نگارش این پایان‌نامه از فضای آزمایشگاه فناوری وب دانشگاه فردوسی و مشورت دوستانم در این آزمایشگاه بهره‌های فراوان برده ام. پس جای دارد از تمام این دوستان به خصوص از **آقای مهندس عسکریان** که در این مدت راهنمای بنده بوده‌اند، قدردانی کنم.

سپاس ویژه‌ای نیز به رایانه عزیزم در آزمایشگاه تقدیم می‌کنم که با وجود تمام کمبودهای سخت‌افزاری‌اش تا آخرین کام نگارش مرا همراهی کرد و از کار دست نکشید.

چکیده

با توجه به توسعه وسیع و رو به رشد فناوری وب، حجم نظرات در صفحات وب نیز به شدت در حال افزایش است. نظراتی که اکنون حجم عظیمی از فضای مجازی را به خود اختصاص داده‌اند، می‌توانند به عنوان یک منبع اطلاعاتی و آماری مهم، مورد توجه قرار گیرند. با این وجود پردازش تعداد **فراوانی زیادی از نظرات** برای کاربردی خاص، کاری بس دشوار و بعضاً ناممکن است. نظرکاوی به عنوان یک حیطة علمی نسبتاً جدید سعی در دسته‌بندی گرایشات و احساسات (نظرات) بیان شده توسط کاربران را دارد.

در روش پیشنهادی این تحقیق، به کشف دقیق حس نظرات از طریق کشف الگوهای نظری در واحد جمله پرداخته شده است. روش‌های مبتنی بر لغت‌نامه کنونی برای کشف عبارات حسی نظرات از چند الگوی ثابت چند کلمه‌ای استفاده می‌کنند. حال آنکه **حقیقتاً عملاً** نظرمی‌تواند به صورت پراکنده در طول ساختار جمله بیان شده باشد و به صورت چند کلمه متوالی نباشد. در روشی که در این تحقیق ذکر می‌شود با استفاده از نقش‌های نحوی کلمات و نیز تعریف برچسب‌های خاص حسی و ترکیب آن‌ها با یکدیگر، سعی می‌شود تا الگوهای کاملتری از بیان حس در جملات نظری به دست آید. بعلاوه **چون** این الگوها با توجه به نظرات کاربران در بخش آموزش پیکره و با ارائه الگوریتمی به صورت خودکار استخراج خواهد شد. **که** این روش قابلیت استفاده بر روی سایر پیکره‌ها را **نیز** دارد. در این کار تلاش شده تا علاوه بر ارائه روش جدید برای کشف حس، با چالش‌های خاص زبان فارسی نیز مقابله شود.

سپس الگوهای به دست آمده نهایی بر روی **بخش تست** پیکره **آزمون** مورد ارزیابی قرار گرفته‌اند که توانستند دقت ۹۶٪ در زبان فارسی را به نمایش بگذارند. به عنوان خط مبنا نیز از الگوریتم دسته‌بندی **SVM** استفاده شد. دقت به دست آمده ۵٪ بیش از دقت بهترین روش مشابه است.

همچنین پیکره مورد نظر یکبار با روش مبتنی بر الگوریتم بی‌زین نیز مقایسه شد که دقت به دست آمده ۸ درصد بهتر از این الگوریتم دسته‌بندی‌پروش بود. همچنین در مقایسه صورت گرفته با کارهای انجام شده برای زبان فارسی نشان داده شده که روش پیشنهادی در این پژوهش از کیفیت مناسبی نسبت به سایر سیستم‌های طراحی شده برای زبان فارسی برخوردار است.

کلید واژه ها: نظر کاوی، کشف الگوی نظری، برچسب‌گذاری اجزای کلام، نظرات مقایسه‌ای،

الگوی حسی جملات، پیکره زبان فارسی

فهرست مطالب

۱	مقدمه	۱
۲	مقدمه	۲
۳	تعریف مساله	۳
۴	چالش‌ها و موجود	۴
۸	راه‌حل پیشنهادی	۸
۹	جنبه جدید بودن نوآوری	۹
۱۱	ساختار پایان‌نامه	۱۱
۱۲	مرور بر کارهای گذشته	۱۲
۱۳	مقدمه	۱۳
۱۳	مفاهیم پایه	۱۳
۱۳	1-1-1- تعریف نظر	۱۳
۱۴	۱-۱-۲- سند، موضوع و گرایش	۱۴
۱۴	۱-۱-۳- انواع جمله	۱۴
۱۵	انواع نظر	۱۵
۱۶	۱-۱-۴- نظرات ساده و مقایسه‌ای	۱۶
۱۷	۱-۱-۵- نظرات صریح و ضمنی	۱۷
۱۷	اجزای نظر	۱۷
۱۹	معماری سیستم‌های نظر کاوی	۱۹
۲۰	سطوح نظر کاوی	۲۰
۲۱	۱-۱-۶- مبتنی بر سطح حسند	۲۱
۲۱	۱-۱-۷- مبتنی بر جمله	۲۱
۲۲	۱-۱-۸- مبتنی بر ویژگی	۲۲
۲۳	روش‌های نظر کاوی در سطح جمله	۲۳
۲۴	روش‌های مبتنی بر یادگیری با ناظر	۲۴
۲۶	مبتنی بر لغت‌نامه	۲۶
۳۰	کارهای انجام شده در فارسی	۳۰
۳۳	نظرات مقایسه‌ای	۳۳
۳۴	۱-۱-۹- تعریف مسئله	۳۴

۳۶	۱-۱-۱۰- شناسایی نظر اتمقایسه‌ای
۳۷	۱-۱-۱۱- دسته‌بندی نظر اتمقایسه‌ای
۳۸	خلاصه فصل
۳۹	۳- روش پیشنهادی
۴۰	مقدمه
۴۰	ساختار روش پیشنهادی
۴۳	پیش‌پردازش مجموعه داده
۴۳	۱-۱-۱۲- تبدیل محاوره به رسمی
۴۴	۱-۱-۱۳- یکسان‌سازی متن
۴۵	۱-۱-۱۴- جداسازی جملات
۴۶	۱-۱-۱۵- شناسایی کلمات
۴۶	برچسب‌گذاری دست‌نظرات
۴۷	شناسایی الگوهای جملات نظری
۴۷	۱-۱-۱۶- برچسب‌گذاری نقش‌نحو یا جزای سخن
۴۸	دسته‌بندی نظر اتو کشف الگو
۴۸	۱-۱-۱۷- برچسب‌های حسی
۵۱۵۴	۱-۱-۱۸- برچسب‌های نحوی
۵۳	۱-۱-۱۹- ترکیب برچسب‌ها و کشف الگوهای مناسب برای هر دسته
۵۴	۱-۱-۲۰- نظر اتمقایسه‌ای
۵۴	تطبیق الگوها بر نظر بر تبهدهی به سند نظری
۵۶	خلاصه فصل
۵۷۵۸	۴- پیاده‌سازی یوارز با سیستم
۵۸۵۹	مقدمه
۵۸۵۹	پیاده‌سازی
۵۸۵۹	۱-۱-۲۱- مجموعه داده
۶۱۶۴	۱-۱-۲۲- ابزارهای پیش‌پردازش و برچسب‌گذاری نحوی یا جزای کلام
۶۳۶۴	۱-۱-۲۳- برچسب‌های بیکره
۶۴۶۵	۱-۱-۲۴- دسته‌ها و الگوهای استخراج شده
۶۵۶۶	ارزیابی سیستم

۶۵۶۶	۱-۱-۲۵	مقایسه حجم پیکره
۶۷۶۸	۱-۱-۲۶	انتخاب دسته مناسب
۷۱۲۲	۱-۱-۲۷	دقت بر چسب نیم مثبت و منفی روش پیشنهادی
۷۳۲۴	۱-۱-۲۸	خط مینا و مقایسه خروجی سیستم
۷۵۲۶	۱-۱-۲۹	جایگاه سیستم پیشنهادی در زبان فارسی
۷۸۲۹	۱-۱-۳۰	نظرات مقایسه ای
۱۲۶۴		خلاصه فصل
۸۳۸۴	۵-	نتیجه گیری و ارائه پیشنهادها
۱۴۸۵		مقدمه
۱۴۸۵		نتیجه گیری
۱۶۸۷		کارهای آتی
۸۸۸۹	۶-	مراجع
۹۱۹۲	۷-	پیوست - الف
۱۲۴۴۵		الف-۱- مقدمه
۱۲۵۴۶	۸-	پیوست - ب
۱۲۷۴۸		ب-۱- مقدمه
۱۲۹۴۳		Title

فهرست شکل‌ها

- شکل ۱-۱: بخش‌های اصلی روش پیشنهادی ۹
- شکل ۱-۲: حوز هیژوهش‌های مبتنی بر تفکیک متن هدف ۱۵
- شکل ۲-۲: معمار بیکسیستم نظر کاوی - خروجی سیستم خلاصه‌ای از نظر اتارائه شده می باشد ۱۹
- شکل ۳-۲: گام‌های اساسی در نظر کاوی [۹] ۲۰
- شکل ۴-۲: نظر اتی به تفکیک برای تک‌تک ویژگی‌ها بیک محصول در دو دستم مثبت و منفی خلاصه شماره شمی شوند ۲۳۶۴
- شکل ۵-۲: ۵۲
- نظر اتی خلاصه هد در قالب نمودار می‌تواند به تصویر کشیده شوند و بر احتیما مقایسه بین ویژگی‌های مختلف محصولات از دید کار
بر انرا مشاهده نمود. ۲۳
- شکل ۱-۳: معمار یطر چپیشنهادی ۴۱
- شکل ۱-۴: برنامه چسبزن نحو یا جزا یکلام متعلقه‌ها ز مایشگاه فناور یویدانشگاه فر دوسی ۶۲۶۴
- شکل ۲-۴: نمودار مقایسه بین حجم پیکر هچندر و شنظر کاوی ۶۷۶۸
- شکل ۳-۴: مقایسه راهکار پیشنه‌ها دیا الگور یتم‌ها د یسته بند یبیز بنوماشینبر دار پشتیبان ۷۴۷۵
- شکل ۴-۴: مقایسه عملکر در و شپیشنهاد دیا الگور یتم *SVM* به عنوان خطمبنا یار ز یابی ۸۱۸۴

۱- مقدمه

Commented [m1]: این باید بشود: فصل اول :

مقدمه

مقدمه

Commented [m2]: اینجا همه باید بصورت سلسه مراتبی شماره گذاری شوند. مثلا ۱-۱

با توجه به گسترش روز افزون فناوری وب در سطح جهان حیطه‌های مختلف تحقیقاتی در این زمینه نیز در حال رشد و نمو هستند. در وب اولیه محتوای منتشر شده در سطح وب، به صورت ایستا و توسط دارندگان وب تولید می‌شد و در اکثر موارد بازدیدکنندگان آن تنها امکان دسترسی به منابع استفاده از آن را داشتند. اما با ظهور فناوری‌های برنامه‌نویسی جدید در وب ۲ امکان تعاملات و ایجاد بازخوردها در صفحات به شکل فزاینده‌ای رشد کرد. بخشی عمده‌ای از تعاملاتی که در سطح وب جریان دارد شامل نظرات کاربرانی می‌شود که منجر به بازخورد برای دارندگان صفحات وب می‌گردد. این بازخوردها موجب می‌شود که محصولات و خدمات مبتنی بر آن را به شیوه متناسب با گروه هدف خود نزدیک کرده و از ایده‌های کاربران خود در بهبود عملکردهای آتی خود بهره جویند. اما واضح است تحلیل این نظرات به خصوص در شبکه‌های اجتماعی با مخاطبان گسترده و جهانی به راحتی و به صورت دستی توسط کاربران انسانی امری بسیار دشوار، زمانبر و در برخی موارد ناممکن می‌باشد. برای غلبه بر مشکلات تحلیل و بررسی نظرات در سطح وب برخی از محققین سعی در بررسی نظرات منتشر شده به صورت خودکار داشتند که این امر موجب ظهور نظرکاوی^۱ شد.

تعریف مساله

حوزه نظرکاوی به بررسی خودکار نظرات وب و ارائه گزارش از حجم عظیمی از نظرات به صورت گزاره‌های مثبت و منفی و یا درصد گرایش مثبت و منفی می‌پردازد. در حقیقت نظرکاوی برای رفع چالش زیر ایجاد شده است:

✓ تجربه دیگران می‌تواند در فرآیند تصمیم‌گیری کاملاً مفید باشد اما با توجه به شناخت کم از افراد نظردهنده نظر یک یا تعداد محدودی از افراد نمی‌تواند معیار خوبی برای قضاوت باشد از طرفی بررسی حجم انبوه از نظرات منتشر شده در وب نیز کاری دشوار است. [۱]

نظرکاوی در زمینه‌های مختلف تجاری، سیاسی، فرهنگی، اجتماعی می‌تواند به بررسی نظرات پرداخته و خلاصه‌ای از احساسات بیان شده در این نظرات را به کاربر ارائه دهد. این خلاصه ارائه شده می‌تواند به جهت تصمیم‌گیری‌های مختلف (از جمله تصمیم‌گیری برای خرید محصول و یا سیاست‌گذاری‌های مختلف) و یا به جهت پیش‌بینی (نظیر بررسی نظرات کاربران کشوری خاص در شبکه اجتماعی برای تخمین زمان وقوع انقلاب احتمالی در آن کشور) مورد استفاده قرار گیرد. [۲] اما خود حیطه نظرکاوی نیز با مسائلی مواجه است که از جمله آن‌ها می‌توان به دقت نه چندان قابل قبول کارهای فعلی به علت ضعف در درک درست متن توسط ابزارهای پردازش زبان و نیز عدم کشف دقیق احساس در نظر بیان شده، اشاره کرد. [۳] در ادامه سیستمی تعریف می‌شود تا بتواند به نوعی این مسائل را برطرف کند و یا از تاثیر منفی آن‌ها بر فرآیند نظرکاوی بکاهد. بر اساس سیستم معرفی شده در این تحقیق، بانکی از نظرات پس از انجام عملیات پیش‌پردازشی در قالب جمله به همراه برچسب نحوی اجزای کلام آن‌ها ذخیره می‌شود.

همچنین این بانک با استفاده از برچسب‌های حسی که توسط کاربر برچسب‌گذاری شده است، برای کشف الگوهای حسی آجملات مورد استفاده قرار می‌گیرد. تلاش اصلی این پژوهش استخراج الگوهای نظری در واحد جمله و کشف دقیق جایگاه حس می‌باشد که بر اساس آن می‌توان نظرات را به صورت دقیقتر برچسب‌گذاری نمود. همچنین در این تحقیق سیستمی ارائه می‌شود که هر دو دسته نظر (عادی و مقایسه‌ای) پوشش داده شود.

چالش‌های موجود

علی‌رغم گرایش گسترده‌ای که در دهه اخیر در انتشار مقالات نظرکاوی وجود داشته، فرآیند نظرکاوی به صورت کلی با چندین مشکل عمده روبروست که در ذیل به تشریح آن‌ها پرداخته شده است [۵] و [۶]:

رابطهٔ وقتی یک خزنده^۳ به حوزه موضوعی خاصی محدود می‌شود و صفحات مربوطه به آن را به درستی شناسایی می‌کند، این بدان معنا نیست که تمامی نظرات مربوط به آن صفحه به همدیگر مرتبط هستند. این مشکل به خصوص در برخورد با شبکه‌های اجتماعی هنگامی که نظرات و بحث‌ها به سرعت از مسیر اصلی فاصله می‌گیرد، وجود دارد.

شناسایی ویژگی؛ یک سند ممکن است شامل نظرات مثبت و منفی از میان یک یا چند موضوع باشد، بنابراین، نظرکاوی در سطح سند احتیاج دارد تا به سطح عبارات و سطح ویژگی کاهش پیدا کند تا شناسایی دقیقی از ویژگی صورت گیرد.

Sentimental patterns
relevance
crawler
Aspect detecting

منفی کننده‌ها: طبقه‌بندهای لغات حسی به طور مشخص در غلبه بر منفی کننده‌ها دارای ضعف می‌باشند. کلماتی مثل «نه»، «نه چندان»، «عدم» و ... نمونه‌هایی از منفی کننده‌ها در زبان فارسی هستند.

اطلاعات وابسته به قراین: شمار عظیمی از نظرات بخصوص در شبکه‌های اجتماعی، حاوی اطلاعاتی است که به قراین خاصی مربوط می‌گردد. به عنوان مثال، اگر شما بیان کنید که کسی مانند «شرلوك هولمز» است، به این مطلب اشاره دارید که آن فرد دارای تیزهوشی و دقت نظر بسیار بالایی است. حال آنکه سیستم نظرکاوی نمی‌تواند براحتی این اطلاعات همراه قرینه را کشف نماید.

ناسزاها! ناسزاها از موارد دیگری است که شناخت و غلبه بر آن مشکل است. یک ناسزا گاه با ساختاری غیر متعارف در کلام یا متن ظاهر می‌شود.

مقایسه: گاهی اوقات مقایسه میان ویژگی از موجودیت داخل دامنه با موجودیتی خارج از دامنه وجود دارد که اطلاعات آن برای ما موجود نیست.

وابستگی به موجودیت: بعضی عبارات یکسان در کاربرد برای موجودیت‌ها و یا ویژگی‌های متفاوت دارای معانی متفاوتی‌اند. برای مثال در عبارت "نمایشگر کوچک"، کوچک دارای بار منفی و بیان کننده ضعف در ویژگی نمایشگر سیستم می‌باشد. درحالی‌که "صف کوچک" ویژگی مثبت کوچک را برای صف بیان می‌کند.

شدت گرایش: کلمات نمونه "عالی"، "خوب"، "مناسب" و "قابل قبول" برای بیان احساسات به کار می‌روند و در مجموعه‌ی کلمات حسی به عنوان کلماتی که حاوی نظراند، جای می‌گیرند. در مقالات ارائه شده سطح رضایت این چهار کلمه را یکسان در نظر می‌گیرند در حالی که در زبان طبیعی سطح رضایت کلمه "عالی" به مراتب بیشتر از سطح رضایت کلمه "خوب" می‌باشد و بنابراین این کلمات بیان‌کننده احساس نباید در یک دسته قرار گیرند و به کار بردن هر کدام از آن‌ها در تعیین رتبه‌ی کالا، تأثیر متفاوتی می‌گذارد. در مقالات مرتبط توجهی به این موضوع نشده و معمولاً کلمات بیان‌کننده‌ی احساس را در سه دسته‌ی مثبت و منفی و خنثی قرار داده‌اند، که دسته مثبت نشان‌دهنده رضایت کاربر و دسته منفی نشان‌دهنده عدم رضایت کاربر می‌باشد.

کشف ضمیر: به جمله «من یک دوربین از سایت آمازون خریدم. آن عکس‌های بسیاری خوبی می‌گیرد» دقت کنید. این که منظور از ضمیر «آن» در این جمله واقعاً کیست برای انسان ساده است، اما ماشین نمی‌تواند به آسانی درک کند که ضمیر دقیقاً به چه چیزی اشاره دارد. این مشکل بخصوص در نظرات شبکه‌های اجتماعی که موضوع اصلی از قبل مشخص نیست؛ قابل لمس است. کشف مرجع ضمیر از جمله مباحثی است که در حوزه پردازش زبان طبیعی نیز با وسعت زیادی مورد مطالعه است. اگرچه هنوز به عنوان یک چالش مهم مطرح است.

تطبیق مترادف‌ها: این امر خیلی رایج است که کاربران نظراتشان را با واژه‌های مختلفی که به یک معنی اشاره دارد و برایشان ملموس‌تر است بیان کنند، برای مثال (voice = sound quality). تطبیق این واژگان خود مستلزم زمان و هزینه بالایی است. عدم انطباق مناسب نیز، موجب افت دقت نهایی خواهد گردید.

دقت پایین: علی‌رغم تمام پژوهش‌ها و مطالعاتی که در زمینه نظرکاوی صورت گرفته است، هنوز سیستم‌های نظرکاوی از دقت بالایی برخوردار نیستند.

(۱) کیفیت صدا واقعا عالی است

(۲) من بلندگوها رو در آشپزخونه گذاشتم ولی به واضیحی میتونم صدای اونها رو در اتاق خواب که فاصله خیلی خوبی داره بشنوم

این که واقعا کدام یک از این دو جمله بیان مثبت‌تری از کیفیت صدا دارد، برای انسان نیز به راحتی قابل درک نیست. علاوه بر این در جملات بسیار ساده‌تری نیز که می‌بایست دقت بالاتری گزارش شود هنوز ناتوان هستیم.

عدم توجه کافی به نظرات مقایسه‌ای: از دیگر چالش‌هایی که نظرکاوی با آن روبرو است وجود نظرات مقایسه‌ای است. نظرات مقایسه‌ای بر خلاف نظرات مستقیم به بیان حس پیرامون یک موجودیت نمی‌پردازند و یک موجودیت را با یک و یا چند موجودیت دیگر مورد مقایسه قرار می‌دهند. از جمله چالش‌های خاصی که پیرامون بررسی نظرات مقایسه‌ای با آن روبرو هستیم می‌توان به موارد ذیل اشاره نمود:

- عدم اختصاص رتبه مثبت و منفی به محصول و صرفاً انجام مقایسه
- عدم وجود نظر در بسیاری از جملات مقایسه‌ای (لپ‌تاپاً از مودم داخلی و لپ‌تاپ ب از مودم خارجی پشتیبانی می‌کند)
- وجود قیاس‌های ضمنی (اپل سروره)

عدم کشف دقیق احساس: اکثر پژوهش‌های نظرکاوی در تلاش هستند که از مجموعه عبارات حسی به همراه چند الگوی نظری ثابت چند کلمه‌ای (نظیر آنچه در **جدول ۲-۲: الگوهای کشف حس [۲۱]** جدول ۴-۴ خواهد آمد) برای کشف نظر استفاده کنند. حال آنکه ممکن است در قالب یک جمله احساس خارج از این چهارچوب مشخص مطرح شود. در بعضی کارها نیز بدون داشتن الگوی

مشخص فاصله میان حس و ویژگی و یا تعداد وقوع عبارات حسی و ویژگیها مد نظر قرار گرفته است. بنابراین ممکن است درک دقیقی از حس جمله انجام نشده باشد.

✓ در این پژوهش در تلاش می‌شود تا به رفع دو چالش آخر یعنی "عدم توجه کافی به نظرات مقایسه‌ای" و "عدم کشف دقیق حس" پرداخته شود. همچنین با توجه به این که دامنه کاری این تحقیق نظرات فارسی است. با برخی چالشهای خاص زبان فارسی نیز مواجه شده است که از جمله آنها می‌توان موارد زیر را یاد کرد:

- نقص ابزارهای پردازش زبان فارسی
- دامنه گسترده وندهای تصریفی
- انواع فضاهای خالی
- حجم بالای کلمات محاوره‌ای و غیر رسمی

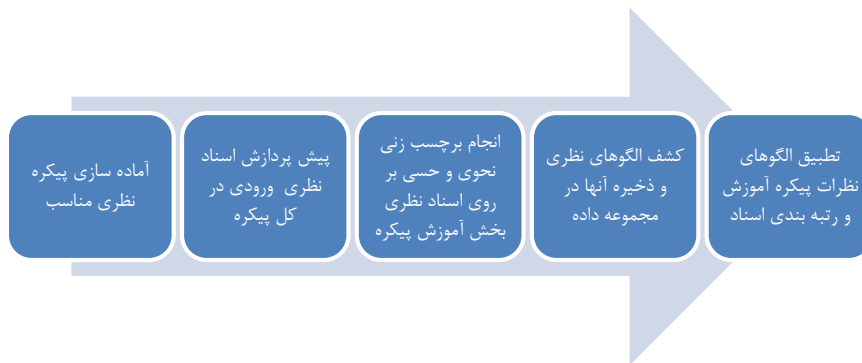
راه حل پیشنهادی

سیستم‌های فعلی برای رسیدن به دقت بالاتر در نظر کاوی، نیازمند درک عمیق‌تری از متن هستند. برای رسیدن به این درک عمیق‌تر در اینجا روشی بر مبنای استفاده از نقش نحوی کلمات معرفی خواهد شد. در این روش با استفاده از تکنیک‌های آماری و نیز نقش نحوی جملات نظری تلاش می‌شود تا الگوهایی که توسط کاربران به بیان نظر پرداخته می‌شود کشف شوند. در گام بعدی نیز از این الگوها برای ارزیابی دقیق عبارات تحت بررسی استفاده خواهد شد.

در حقیقت در روش‌هایی که تا کنون مورد استفاده قرار گرفته است، از چند الگوی نحوی چند کلمه‌ای به عنوان الگوهای نظری استفاده شده است. در روش پیشنهادی این پایان‌نامه قصد داریم تا به جای استفاده از این الگوهای ثابت چند کلمه‌ای الگوهای نحوی و حسی را در واحد جمله کشف و استخراج نماییم. مزیت اصلی این الگوها کشف آنها به صورت خودکار و بر اساس نرخ

تکرار الگو می‌باشد که از وابستگی به دامنه خاص جلوگیری شود، بعلاوه الگوها در واحد جمله کشف خواهند شد که ساختارهای مختلف بیان احساسات را پوشش خواهند داد.

برای رسیدن به مقصود مورد نظر از ابزار برچسب زن نحوی اجزای کلام بهره برده شده است. علاوه بر این، برچسب‌های حسی خاصی در این تحقیق مطرح گردیده که در کنار برخی برچسب‌های حسی سایر مقالات مورد استفاده قرار گرفته است. در نهایت الگوهای خاص نظری که از ترکیب این برچسب‌های نحوی و حسی ایجاد می‌شوند از پیکره آزمون استخراج گردیده است. همچنین مجموعه قواعدی برای دسته‌های مختلف الگوها تهیه شده است، تا با استفاده از آن پس از شناخت الگوی خاص بتوان آن را رتبه دهی نظری کرد. گام‌های اساسی که در روش پیشنهادی اجرا خواهد شد در شکل زیر بیان شده است.



شکل ۱-۱: بخشهای اصلی روش پیشنهادی

جنبه‌ی جدیدبودن و نوآوری

کارهایی که تاکنون در حوزه نظرکاوی منتشر گشته‌اند و از نقش نحوی لغات به عنوان یک ویژگی برای کشف حس استفاده کرده‌اند، از تعدادی الگوی نظری ثابت بهره می‌برند. نمونه‌ای از این الگوهای ثابت در جدول ۲-۲ بیان گردید. در طرح پیشنهادی تلاش بر این است که پس از برچسب‌گذاری نظرات مثبت و

منفی، برچسب‌های نحوی به تک‌تک کلمات در یک جمله اختصاص داده شود. سپس با توجه به جملات مختلف محل دقیق حس در قالب یک جمله شناسایی شود. در زیر از یک مثال برای تشریح نوآوری سیستم پیشنهادی استفاده می‌کنیم:

"قیمت این گوشی می‌تواند یک مثال خوب از ضعف آن باشد"

اکثر سیستم‌های ایجاد شده در نظرکاوی در طبقه‌بندی جمله فوق دچار مشکل خواهند شد، زیرا در این جمله دو کلمه حسی "ضعف" و "خوب" کنار یکدیگر در یک جمله پدیدار گشته‌اند. اما در روش پیشنهادی که بر اساس استخراج و تطبیق الگوی حسی در واحد جمله است در صورتی که جملات مشابه با این جمله در مجموعه داده وجود داشته باشد و یا الگوی مناسبی بر اساس جملات شبیه به این جمله در مجموعه داده استخراج شده باشد، می‌توان محل دقیق احساس را شناسایی کرد.

بر طبق نظر [۱] تاکنون روشی برای کشف دقیق حس وجود نداشته و نیاز به کارهای بیشتری برای کشف دقیق احساس وجود دارد. در حقیقت تفاوت اصلی این روش با روش‌های پیشین نگاهی دقیقتر به کشف حس، خارج از در نظر گرفتن الگوهای ثابت چند کلمه‌ای می‌باشد. در اینجا الگوها در واحد یک جمله ایجاد خواهند شد و نه بر اساس چند کلمه متوالی. همچنین روش‌های قبلی الگوهای محدود را به صورت دستی در نظر گرفته‌اند. در این تحقیق الگوها بر اساس تکرار آنها و همچنین ایجاد قواعدی به صورت نیمه خودکار (از روی مجموعه نظرات برچسب خورده) ایجاد خواهد شد تا به الگوهای جدید در قالب جمله دست یابیم. الگوهای واحد جمله توان حل مسائل پیچیده تری را خواهند داشت از جمله این که در صورت شناسایی چند صفت، کدام یک می‌تواند کاندید مناسبی برای کشف حس باشد. این الگوها همچنین می‌توانند تا حدودی در کشف جملات نظری که به ظاهر نظری نیستند و یا حذف جملات به ظاهر نظری نیز کاربرد داشته باشد.

در این پژوهش همچنین در تلاش هستیم الگوهای نظرات مقایسه‌ای را نیز در واحد جمله استخراج کنیم. برطبق جستجوهای که انجام شد، در منابع معدودی که در بین کارهای نظرکاوی به بررسی نظرات

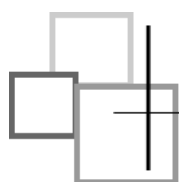
مقایسه‌ای پرداخته‌اند؛ توجهی به استخراج الگو برای نظرات مقایسه‌ای صورت نگرفته است. علاوه بر این مطالعه‌ای برای نظرکاوی نظرات مقایسه‌ای در زبان فارسی صورت نگرفته است.

با توجه به مطالب گفته شده در این بخش به صورت خلاصه می‌توان نوآوری‌هایی که این پژوهش دنبال می‌کند در تیتراهای زیر بیان نمود:

- توجه به کل واحد جمله و نه عبارت یا چند کلمه متوالی ثابت
- استخراج الگوهای جملات مقایسه‌ای
- استخراج الگوها به صورت نیمه خودکار

ساختار پایان‌نامه

در فصل دوم کارهای انجام شده در حوزه نظرکاوی در سه دسته الگوریتم‌های ناظر، الگوریتم‌های بدون ناظر و کارهای صورت گرفته بر روی زبان فارسی آورده شده است. در این فصل تلاش بر این بوده که از میان کارهای زیاد صورت پذیرفته مرتبط‌مرتبط‌ترین و معتبرترین منابع ذکر شود. پس از آن در فصل سوم به بسط روش پیشنهادی با توجه به نواقص گفته شده در فصل دوم پرداخته شده است. در فصل چهارم به صورت مراحل پیاده‌سازی روش پیشنهادی به صورت مرحله به مرحله شرح داده شده است. سپس در این فصل روش پیشنهادی مورد ارزیابی قرار گرفته است. در فصل پنجم نیز نتیجه گیری بیان شده است. همچنین در این فصل کارهای آتی قابل انجام در ادامه این تحقیق بیان شده‌اند.



۲- مروری بر کارهای گذشته

مقدمه

در این فصل ابتدا مفاهیم پایه‌ای که درک درست از نظرکاوی مستلزم درک صحیح آن‌ها است مرور خواهند شد. در ادامه سطوح نظرکاوی و انواع روش‌های آن بحث خواهد شد. پس از آن مفصل‌تر به روش‌های به کار گرفته شده مرتبط با روش پیشنهادی پرداخته شده است.

مفاهیم پایه

در این بخش به طور مختصر چند مفهوم پایه‌ای که مستلزم درک کامل این تحقیق هستند، بیان می‌شود.

۱-۱-۱- تعریف نظر

Commented [m3]: الان شماره این باید ۲-۲-۱ باشد

در [۶] نظر به جمله یا بخشی از آن اطلاق می‌گردد که به پاسخ این پرسش بپردازد " *How does X feel about Y?*" در فرهنگ لغات گوگل نیز نظر چنین تعریف می‌شود: دید یا قضاوت شکل گرفته در مورد چیزی که ضرورتاً بر اساس دانش و یا حقیقتی شکل نگرفته باشد.

با عنایت به تعاریف بالا و تعاریف دیگری که از نظرات در منابع مختلف موجود می‌باشد درمی‌یابیم که دو مفهوم کلیدی در تعریف اکثر نظرات به صورت ضمنی یا صریح بیان شده است که عبارتند از: "فردی و سلیقه‌ای بودن" و "مبتنی بر احساس بودن".

پس می‌توان تعریف زیر را ارائه نمود که در برگزیده مفاهیم بالا نیز باشد:

یک نظر عبارت است از دیدگاه، باور، قضاوت شکل گرفته فرد، پیرامون یک موضوع خاص که می‌تواند ریشه در تجربیات، تمایلات و احساسات فردی و سلیقه‌ای داشته باشد که ضرورتاً از دانشی بهره نچسته است.

۲-۱-۱- سند، موضوع و گرایش^۲

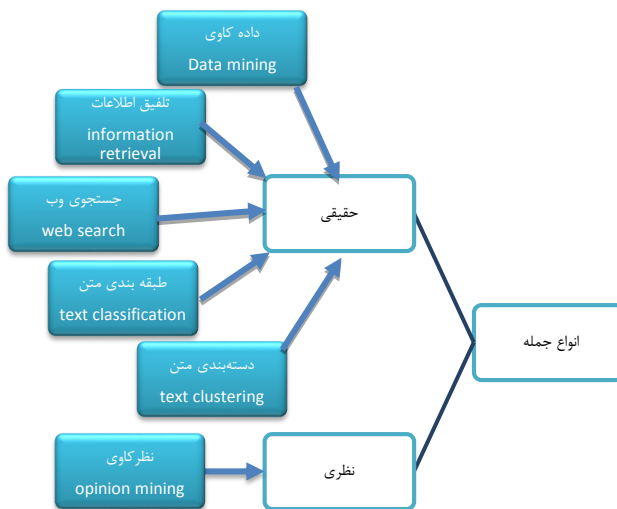
بر طبق تعریف ارائه شده توسط [۷]، سند D قطعه متنی در زبان طبیعی است. ما می‌پنداریم هر سند در مورد لااقل یک موضوع بحث می‌نماید، و تمام موضوعات بحث شده در یک سند الزاماً به هم مرتبط نمی‌باشند. موضوع T به واقع، موجودیت و یا مفهوم انتزاعی بیان شده در متن اشاره می‌کند. گرایش S هدف، نظر و تمایل نویسنده متن در مورد موضوع مربوطه را بیان می‌نماید.

۳-۱-۱- انواع جمله

جملات از دیدگاه مورد بحث ما در دو دسته جای می‌گیرند. دسته اول جملات عینی^۳ و دسته دوم جملات ذهنی^۴ است. جمله عینی به جمله‌ای گفته می‌شود که بیان‌کننده حقیقتی از جهان پیرامون باشد، مانند "من هفته پیش یک لپ‌تاپ خریدم". در مقابل جملات ذهنی جملاتی هستند که بیان‌کننده احساس، باور و یا دید شخصی باشند. مثال "من از این رستوران خوشم نمی‌آید". در فرآیند نظرکاوی غالباً به دنبال نظرات ذهنی هستیم، زیرا آن‌ها حاوی نظر می‌باشند. البته این را نیز باید مدنظر قرار داد که می‌توانیم جملات عینی داشته باشیم که بیان‌کننده نظر مثبت یا منفی نسبت به چیزی باشند و یا بالعکس. برای مثال جمله عینی "من دو روز پیش گوشی موبایلی خریدم که امروز از کار افتاد" حقیقتی را بیان می‌کند که در مورد یک گوشی موبایل کاملاً دارای گرایش منفی می‌باشد [2].

document
#topic
#sentiment
#objective
#subjective

در حقیقت اکثر کارهای صورت گرفته بر روی اطلاعات متنی، توجه خود را بر جملات حقیقی معطوف کرده‌اند و این وجه تمایز حوزه نظرکاوی با سایر زمینه‌های تحقیقاتی مرتبط است. تصویر زیر بیان‌کننده این حقیقت است.



شکل ۱-۲: حوزه پژوهش‌های متنی به تفکیک متن هدف

انواع نظر

نظرات را می‌توان از جنبه دلالت آن بر موجودیت به دو شکل ساده و مقایسه‌ای تقسیم نمود. اکثر کارهای انجام شده تا کنون بر روی نظرات ساده صورت پذیرفته است. همچنین به لحاظ

نحوه بیان نظر می‌توان آن‌ها را به نظرات صریح^۴ و ضمنی^۵ تقسیم بندی نمود. در ادامه به بررسی این دسته‌ها خواهیم پرداخت. [۱]

۴-۱-۱- نظرات ساده و مقایسه‌ای:

نظر ساده

یک نظر ساده اغلب نشان‌دهنده حس و تمایل به یک موجودیت است. خود نظر ساده دو نوع دارد که شامل نظر مستقیم و غیر مستقیم می‌گردد.

نظر مستقیم، نظری است که مستقیماً به یک موجودیت یا جنبه‌ای از آن اشاره می‌نماید. برای مثال: "سرعت پردازشگر این سیستم واقعا بالاست".

نظر ساده غیر مستقیم، به صورت غیر مستقیم به موجودیت و یا جنبه‌ای از آن اشاره دارد که بر اساس تاثیر بر موجودیت‌های دیگر حاصل گشته است. این دسته از نظرات در حوزه پزشکی شیوع فراوانی دارد. برای مثال، این جمله "بعد از تزریق دارو، مفاصلم احساس بدی داشتند" توصیف ناخوشایندی از تاثیر دارو بر روی مفاصل را بیان می‌دارد، که به صورت غیر مستقیم یک نظر منفی در مورد دارو می‌باشد. در این حالت، موجودیت مد نظر "دارو" و ویژگی آن "تاثیر بر مفاصل" است.

نظر مقایسه‌ای

این نظر بیان‌کننده شباهت و یا تفاوت میان دو موجودیت و یا جنبه‌های مشترکی از آن‌هاست. برای مثال، جملات "طعم پپسی بهتر از کوکاکولا است" و یا "کوکاکولا بهترین طعم را دارد" دو

^۴Explicit

^۵Implicit

۲- مروری بر کارهای گذشته

جمله مقایسه‌ای هستند. جملات مقایسه‌ای عموماً از صفات مقایسه‌ای یا تفضیلی استفاده می‌کنند.

۵-۱-۱- نظرات صریح و ضمنی

نظر صریح

نظر صریح یک بیان جمله ذهنی است که حاوی نظر مقایسه‌ای و یا ساده باشد. برای مثال، جملات "طعم پپسی بهتر از کوکاکولا است" و یا "کوکاکولا بهترین طعم را دارد" دو نظر صریح هستند.

نظر ضمنی

یک نظر ضمنی شامل بیان جمله عینی است که به صورت ضمنی به بیان نظری ساده یا مقایسه‌ای می‌پردازد. به عنوان نمونه، برای جملات ذهنی‌ای که حاوی حقایق خوشایند و یا ناخوشایند هستند می‌توان مثال مقابل را بیان نمود: "من هفته گذشته صندلی کامپیوتری خریداری کردم که تاثیر فرورفتگی ابرهای آن کاملاً ملموس است."

اجزا نظر

برای بررسی بهتر نظرات، جا دارد آن‌ها را به اجزا و بخشهای وابسته‌شان تقسیم بندی نماییم. بینگ لیو در کتابش اجزای نظر را بصورت چهارگانه و در تعریف کاملتری به صورت ۵ گانه زیر بیان می‌کند: [۱]

۲- مروری بر کارهای گذشته

(e, a, s, h, t)

هر کدام از این حروف، نماینده بخشی از نظر می‌باشند

e : مخفف entity، اشاره به موجودیتی دارد که نظر بیان‌کننده آن است.

a : مخفف aspect، جنبه یا ویژگی‌ای از موجودیت را بیان می‌کند.

s : مخفف sentiment، گرایش نظر را بیان می‌نماید که می‌تواند مثبت، منفی یا خنثی باشد.

همچنین ممکن است که این گرایش توسط اعداد (به عنوان مثال با اعداد ۱۰-الی ۱۰+) نمایش داده شود.

h : مخفف holder، شخصی که نظر را بیان می‌کند.

t : مخفف time، زمانی که نظر درج گردیده است.

نکته: هنگامی که از یک موجودیت به صورت کلی یاد می‌شود، جنبه یا ویژگی مورد نظر GENERAL برای نمایش به کار خواهد رفت. در اینجا مقدار e با a برابر خواهد بود.

در حقیقت کار اولیه‌ای که پس از دریافت یک سند انجام می‌گیرد، استخراج این اجزای پنجگانه نظر از داخل سند مربوطه است. به نوعی می‌توان این وظیفه را گذر از متن ساختارنیافته به داده‌های ساختاریافته دانست. از جمله فواید مهمی که این امر فراهم می‌سازد، می‌توان به موارد ذیل اشاره نمود:

- استفاده از ابزارهای قدیمی در نمایش داده‌ها و اعمال پردازش بهتر روی داده‌های تکه-تکه شده
- توانا ساختن تحلیل چهارگانه و پنجگانه

معماری سیستم‌های نظر کاوی

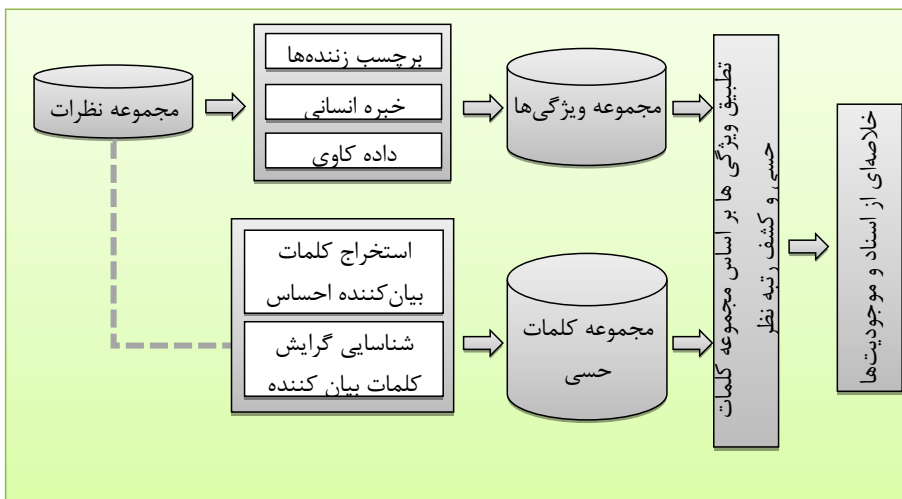
مدل‌های متفاوت نظر کاوی در مقالات ارائه شده است یکی از پرکاربردترین مدل‌های نظر کاوی مورد

استفاده، مدل ارائه شده در مقاله [۸] می‌باشد. مدل ارائه شده در این مقاله در **شکل ۲-۲**:

معماری یک سیستم نظر کاوی شکل ۲. معماری یک سیستم نظر کاوی خروجی سیستم خلاصه‌ای از نظرات ارائه شده می‌باشد. نمایش داده شده است.

شکل ۲-۲: معماری یک سیستم نظر کاوی خروجی سیستم خلاصه‌ای از نظرات ارائه شده می‌باشد نمونه

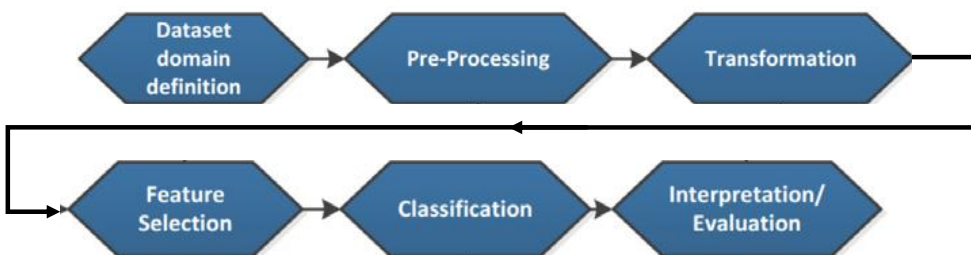
Commented [m4]: فقط ارجاع به شکل ۲ باید اینجا بیاید نه شرح شکل



ابتدا سیستم توسط یک خزنده تمام نظرات را از اینترنت دریافت کرده و سپس در داخل یک پایگاه داده ذخیره می‌کند. در مرحله بعد اجزاء تشکیل دهنده جملات تعیین می‌شود. این کار به منظور شناسایی ویژگی‌های یک کالا انجام می‌شود. در ادامه، بعد از برچسب زدن جملات و شناسایی ویژگی‌ها، کلمات بیان کننده نظرات استخراج و گرایش آن‌ها با استفاده از مجموعه‌ای که شامل کلمات بیان کننده نظرات و گرایش آن‌ها می‌باشد تعیین می‌شود. در مرحله آخر گرایش به ازاء هر جمله یا ویژگی تعیین می‌شود.

سطوح نظر کاوی

به صورت عمومی می‌توان بیان نمود که نظر کاوی در سه سطح اساسی مورد مطالعه قرار گرفته است: [۱]. شکل زیر ساختار کلی فرآیند نظر کاوی را به تصویر می‌کشد.



شکل ۲-۳: گامهای اساسی در نظر کاوی [۹]

خاص پیش‌پردازش شده و به صورت نرمال و ساختار یافته در می‌آیند. در مرحله سوم، کارهای خاص آماری بر روی سند و عبارات موجود در آن صورت می‌گیرد. مرحله چهارم که به عنوان یکی از گامهای اساسی در کاوش نظرات به حساب می‌آید انتخاب ویژگی‌های مناسب برای دسته‌بندی اسناد است. انتخاب مناسب این ویژگی‌ها می‌تواند موجب بهبود عملکرد الگوریتم‌های دسته‌بندی گردد. مرحله پنجم اجرای الگوریتم دسته‌بندی بر روی اسناد است. در نهایت در مرحله ششم به تحلیل نتایج به دست آمده از دسته‌بندی پرداخته می‌شود. این فرآیند می‌تواند در سه سطح زیر انجام گیرد.

مرحله اول تعیین دامنه کار کاوش نظرات است. در مرحله دوم داده‌های جمع‌آوری شده در حوزه خاص پیش‌پردازش شده و به صورت نرمال و ساختار یافته در می‌آیند. در مرحله سوم، کارهای خاص آماری بر روی سند و عبارات موجود در آن صورت می‌گیرد. مرحله چهارم که به عنوان یکی از گامهای اساسی در کاوش نظرات به حساب می‌آید انتخاب ویژگی‌های مناسب برای دسته‌بندی

اسناد است. انتخاب مناسب این ویژگی‌ها می‌تواند موجب بهبود عملکرد الگوریتم‌های دسته‌بندی گردد. مرحله پنجم اجرای الگوریتم دسته‌بندی بر روی اسناد است. در نهایت در مرحله ششم به تحلیل نتایج به دست آمده از دسته‌بندی پرداخته می‌شود. این فرآیندی می‌تواند در سه سطح زیر انجام گیرد.

۱. مبتنی بر واحد سند

۲. مبتنی بر واحد جمله

۳. سطح ویژگی یا سطح جنبه (از موجودیت)

۶-۱-۱- مبثنی بر سطح سند^{۱۱}

وظیفه این سطح از نظر کاوی این است که آنچه کل یک سند منتشر کرده را در قالب مثبت یا منفی دسته‌بندی نماید [۱۰]. برای مثال، یک نقد محصول داده می‌شود و سیستم باید تعیین کند که این نقد به صورت کلی دید مثبتی را نسبت به محصول نشان داده و یا منفی بوده است. این سطح از نظر کاوی همچنین به "طبقه‌بندی نظرات سطح سند"^{۱۲} نیز شناخته می‌شود. آنچه در این سطح از نظر کاوی مفروض است این است که سند، نمایش‌دهنده نظر پیرامون یک موجودیت واحد است. بنابراین، هنگامی که سند چندین موجودیت را ارزیابی یا مقایسه می‌کند، کاربرد ندارد.

۷-۱-۱- مبثنی بر جمله^{۱۳}

وظیفه این سطح این است که تعیین کند جمله یک نظر مثبت، منفی، یا خنثی را نشان می‌دهد. منظور از خنثی این است که جمله بیان‌کننده نظر نیست. این سطح به صورت نزدیکی با

^{۱۱} Document level

^{۱۲} Document-level sentiment classification

^{۱۳} Sentence level

"طبقه‌بندی ذهنیت" وابستگی دارد. در طی این فرآیند جملات ذهنی که بیان‌کننده دید یا نظر ذهنی هستند از جملات عینی که بیان‌کننده اطلاعات حقیقی هستند تمیز داده می‌شود.

۸-۱-۱- مبثنی بر ویژگی^{۱۵}

هر دو سطح قبلی یعنی سند و جمله به طور دقیق آنچه را که مردم دوست دارند و آنچه دوست ندارند را کشف نمی‌نماید. بجای نگرش به ساختار زبان (سند، پارگراف، جمله و ...)، سطح ویژگی مستقیماً خود نظر را دنبال می‌کند. در حقیقت بر این ایده استوار است که هر نظر شامل یک گرایش (مثبت یا منفی) و یک ویژگی از نظر قرار دارد. به این مثال توجه کنید: "اگر چه این رستوران سرویس چندان خوبی ارائه نمی‌کند ولی من هنوز به آن علاقه دارم" این نظر مشخصاً یک دید مثبت را بیان می‌کند ما نمی‌توانیم بگوییم که این نظر تماماً مثبت است. در حقیقت این نظر درباره رستوران به صورت کلی مثبت است و در مورد نحوه سرویس‌دهی آن دارای گرایش منفی می‌باشد. در مثال زده شده شکل ۲-۴ و شکل ۲-۵ نظرات جمع آوری شده در مورد یک گوشی موبایل نشان داده شده است [۱].

Feature Based Summary:

Feature1: Touch screen

Positive: 212

- The touch screen was really cool.
- The touch screen was so easy to use and can do amazing things.

...

Negative: 6

- The screen is easily scratched.
- I have a lot of difficulty in removing finger marks from the touch screen.

...

Feature2: battery life

...

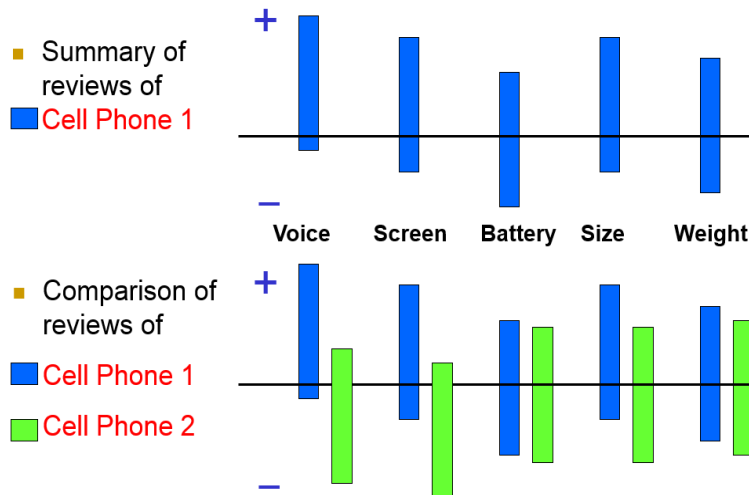
¹ Subjectivity classification

² Entity and aspect level

۲- مروری بر کارهای گذشته

شکل ۲-۴: نظرات به تفکیک برای تک تک ویژگیهای یک محصول و در دو دسته مثبت و منفی خلاصه شمارش میشوند

شکل ۲-۵: نظرات خلاصه شده در قالب نمودار می‌توانند به تصویر کشیده شوند و براحتی



مقایسه بین ویژگیهای مختلف محصولات از دید کاربران را مشاهده نمود.

Commented [m5]: این شرح ها خیلی بزرگ هستند . آنها را در متن بیاورید. اینجا میتوان نوشت: ارائه بصری خلاصه نظرات . این شکل ارجاع هم می خواهد

روش های نظر کاوی در سطح جمله

تا کنون مقالات فراوانی در حوزه نظر کاوی منتشر گشته است. با این وجود روش هایی که برای دسته بندی احساس نظرات مختلف بیان شده است، کاملاً متفاوت بوده و دقت های گزارش شده نیز (با توجه به زبان و دامنه سیستم) تفاوت های فراوانی دارد. همچنین همان طور که مطرح گردید، این منابع به سه سطح مختلف از نظر کاوی پرداخته اند، که با توجه به روش پیشنهادی که در واحد جمله به نظر کاوی می پردازد، منابع مهم در این بخش ذکر می شود. منابع انتشار یافته در

این زمینه در دو دسته الگوریتم‌های یادگیری باناظر^۶ و روش‌های مبتنی بر لغت‌نامه^۷ جای می‌گیرد. روش پیشنهادی نیز ساختاری بدون ناظر را برای نظرکاوی پیشنهاد می‌دهد، بنابراین در ادامه پس از بررسی مختصر مدل‌های مبتنی بر یادگیری باناظر، بر روی روش‌های بدون ناظر تمرکز خواهد شد. در ادامه نیز در بخشی جداگانه کارهای انجام شده بر روی زبان فارسی ذکر می‌شود.

روش‌های مبتنی بر یادگیری با ناظر

در مدل با ناظر طبقه‌بندی نظرات غالباً به صورت یک مسئله طبقه‌بندی دو-کلاس (مثبت و منفی) مد نظر گرفته می‌شود. در بعضی منابع نیز کلاس خنثی نیز در نظر گرفته شده است. پنگ و لی [۱۱] برای نخستین بار ایده استفاده از الگوریتم‌های با ناظر را برای دسته‌بندی نظرات مطرح کردند. در مقاله آن‌ها چندین ویژگی مختلف برای دسته‌بندی نظرات به الگوریتم‌های دسته‌بندی SVM ^۸ و NBA ^۹ داده شد و نشان داده شد که ویژگی کیف لغات می‌تواند در دسته‌بندی نظرات موثر واقع شود.

در حقیقت می‌توان گفت پس از انتشار ایده‌ی استفاده از الگوریتم‌های با ناظر برای دسته‌بندی نظرات، کارهای انجام شده نظرکاوی در مدل مبتنی بر نظارت که بیشترین حجم از مقالات را تشکیل می‌دهند؛ به دو دسته: کارهای انجام شده برای انتخاب ویژگی و بهبود روش‌های پردازش زبان برای این انتخاب‌ها و دسته دوم کارهای انجام شده برای افزایش دقت الگوریتم‌های

^۶machine learning

^۷dictionary-based

^۸support vector machine

^۹Naïve Bayes

^{۱۰}Bag-of-word

دسته‌بندی حوزه نظر‌کاوی تقسیم‌بندی کرد. در اینجا تمرکز ما بر روی تشریح کارهای دسته اول است. از جمله ویژگی‌هایی که معمولاً برای تحلیل احساسات از روی متن مورد استفاده قرار می‌گیرد می‌توان به موارد زیر اشاره کرد: عبارات و نرخ وقوع آن‌ها؛^{۱۰} برچسب اجزای کلام؛^{۱۱} عبارات و لغات حسی؛^{۱۲} نقش‌های نظری؛^{۱۳} تغییردهنده‌های معنایی و وابستگی نحوی^{۱۴}. برای استخراج این ویژگی‌ها از متن نیاز به ابزارهای پردازش زبان طبیعی است و از جمله مشکلات اساسی محققان در استخراج این ویژگی‌ها: موجود نبودن و یا عدم دقت کافی ابزارهای پردازش زبان طبیعی برای زبان‌های غیر انگلیسی است. حتی در بعضی از پژوهش‌ها تلاش شده که با استفاده از مترجم‌های ماشینی از ابزارهای پردازش زبان طبیعی ایجاد شده برای زبان انگلیسی استفاده شود. بر طبق [۱۲] این ترفند نیز به علت ضعف در ترجمه نتوانسته بهبود فراوانی در فرآیند استخراج اطلاعات از متن ایجاد کند.

حجم کارهای انجام شده در این حوزه بسیار زیاد است که برای نمونه به چند مورد مشهورتر اشاره می‌شود. در [۱۳] تمرکز بر روی شناخت ویژگی‌های مشترک نحوی و مستقل از زبان است (زبان مورد مطالعه انگلیسی و عربی). این ویژگی‌ها توسط الگوریتم ژنتیک^{۱۵} از انجمن‌های^{۱۶} ترکیبی و انگلیسی استخراج شده و از الگوریتم SVM برای دسته‌بندی نظرات استفاده شده است. در [۱۴] الگوریتم SVM به عنوان الگوریتم یادگیری انتخاب گشته و از پیکره‌های مختلف برای نظر‌کاوی

^{۱۰}Terms and their frequency

^{۱۱}Part of speech

^{۱۲}Sentiment words and phrases

^{۱۳}Rules of opinions

^{۱۴}Sentiment shifters

^{۱۵}Syntactic dependency

^{۱۶}Genetic algorithm

^{۱۷}Forum

استفاده شده است تا تاثیر این پیکره‌ها در یک روش ثابت مورد ارزیابی قرار گیرند. همینطور از ویژگی‌های مختلفی به عنوان ورودی الگوریتم دسته‌بندی استفاده شده و نیز یک ویژگی خاص برای شبکه‌های اجتماعی طرح گردیده و مورد استفاده قرار گرفته است. بر خلاف دو کار قبلی در [۹] ویژگی‌هایی که به عنوان ورودی به الگوریتم‌های دسته‌بندی داده می‌شود ثابت است و فقط مقایسه‌ای بر روی الگوریتم‌های مختلف انجام شده است. در این مطالعه سه الگوریتم *SVM*، *NBA* و شبکه عصبی برای دسته‌بندی نظرات مورد استفاده قرار گرفته‌اند. بر اساس این تحقیق کارایی الگوریتم شبکه عصبی بیشتر از ماشین بردار پایه و روش بیزین آلت است.

روش‌های مبتنی بر لغت‌نامه

اولین کاری که به شرح روشی مبتنی بر فرهنگ لغت برای نظرکاوی پرداخت توسط آقای تورنی [۱۰] مطرح شد. در گام اول این کار نظرات توسط برچسب‌زن نحوی لغات (POS^3) برچسب‌گذاری شده‌اند. همینطور چهار الگوی خاص به عنوان الگوی کاندید برای بیان نظر در نظر گرفته شده‌اند. به عنوان مثال الگوی کلمه اول صفت و کلمه دوم اسم در زبان انگلیسی می‌تواند الگویی باشد که به احتمال فراوان بیان‌کننده نظر است. در گام دوم با استفاده از معیار PMI^3 نسبت ارتباط کلماتی که در این الگو صدق کرده‌اند، مورد بررسی قرار می‌گیرد. این معیار به صورت زیر بیان می‌شود:

³SVM

³naïve Bayes

³part of speech tagger

³pointwise mutual information

$$PMI(term_1, term_2) = \log_2 \left(\frac{\Pr(term_1 \wedge term_2)}{\Pr(term_1) \Pr(term_2)} \right)$$

در این فرمول $\Pr(term_1 \wedge term_2)$ نسبت وقوع همزمان دو اصطلاح یک و دو را نشان می‌دهد و $\Pr(term_1) \Pr(term_2)$ نسبت وقوع این دو اصطلاح زمانی که به لحاظ آماری به یکدیگر وابسته نباشند را نشان می‌دهد. بنابراین گرایش یک کلمه بر اساس فرمول زیر استخراج می‌شود:

$$SO(phrase) = PMI(phrase, \text{"excellent"}) - PMI(phrase, \text{"poor"})$$

در گام سوم میانگین SO محاسبه شده برای تمام کلمات، گرایش مثبت و یا منفی سند را مشخص می‌کند.

بعد از تمرکزی که در این مقاله بر روی برجسبزن نحوی لغات به عنوان یک ابزار در تشخیص حس شد و با توجه به اینکه متون نظرات اصولاً متن‌های قاعده‌مندی نیستند. در [۱۵] روشی برای غلبه بر مشکل برجسبزنی این متون ارائه شد. بر طبق این مقاله اعمال پیش‌پردازشی خاصی برای نظرات در نظر گرفته شده است تا برجسبزن لغات بتواند عملکرد مناسبی بر روی متن حاوی نظر ارائه دهد و در عمل کارایی مناسبتری برای پردازش‌های نظر کاوی داشته باشد.

در بعضی از روش‌های دیگر [مانند](#) [۱۶، ۱۷]، از برجسبگذار نحوی لغات به همراه تجزیه‌گر وابستگی^{۳۳} به عنوان ابزاری برای توسعه جنبه^{۳۴} و استخراج همزمان کلمات حسی و جنبه‌های مرتبط استفاده کرده است. این کارها از روش انتشار مضاعف^{۳۵} برای هدفشان بهره جسته‌اند. بنابر

^{۳۳}dependency parser

^{۳۴}spect

^{۳۵}double propagation

ارتباطی که بین نظر/احساس و هدف (یا جنبه) آن وجود دارد، هر لغت حسی^{۲۶} می‌تواند توسط هر جنبه که از متن استخراج می‌شود، مورد شناسایی قرار گیرد و هر جنبه نیز می‌تواند توسط لغت حسی تشخیص داده شود. بر این اساس روش انتشار مضاعف به دفعات قوانینی را برای کشف الگوهای خاص نحوی به کار می‌برد تا جنبه‌ها و لغات حسی بیشتری از متن استخراج شود. این روش تا آنجا ادامه پیدا خواهد کرد که دیگر هیچ الگوی جدیدی در متن قابل تشخیص نباشد. ۸ رابطه در این مقاله مورد استفاده قرار گرفته‌اند. برای مثال یک رابطه ذکر شده در این مقاله که در زیر آورده شده، بیان می‌کند که در صورت کشف لغت حسی با چه نوع وابستگی نحوی می‌توان جنبه خاص از نظر را که در مثال آورده شده، «صفحه نمایش» است کشف نمود.

جدول ۱-۲: روابط بین جنبه و احساس [۱۷]

	Observations	Output	Examples
RI_1 (OA-Rel)	$O \rightarrow O-Dep \rightarrow A$ s.t. $O \in \{O\}$, $O-Dep \in \{MR\}$, $POS(A) \in \{NN\}$	$a = A$	<i>The phone has a good "screen".</i> $good \rightarrow mod \rightarrow screen$

در [۱۸] تلاش بر این بوده است که توسط برچسب‌زنی لغات به کشف ویژگی‌های نحوی حاوی نظر - که از دامنه مستقل هستند - دست پیدا کنند. در این مقاله برچسب‌های نحوی (اسم، صفت فعل) مد نظر قرار گرفته‌اند. ایده اصلی این روش بر این مطلب استوار است که از میان ویژگی‌هایی که توسط برچسب‌زن نحوی لغات مورد استفاده قرار می‌گیرند تعدادی وابسته به دامنه و تعدادی مستقل از دامنه هستند. در این مقاله سعی شده تا با استفاده از برچسب‌زن نحوی لغات تجمیع شده^{۲۷} ویژگی‌های نحوی مستقل از دامنه شناسایی شود.

^{۲۶}sentiment words
^{۲۷}POS-based ensemble

در روش‌های مبتنی بر فرهنگ لغت تمرکز یا بر روی الگوی نحوی یا بر روی خود لغات است. در روش‌هایی که تاکنون بیان شد، تمرکز اصلی بر روی نقش نحوی بود ولی در [۱۹] مجموعه‌ای از لغات و عبارات به همراه گرایش آن‌ها مورد استفاده قرار گرفته است. همچنین در این روش این عبارات به همراه مجموعه‌ای از تشدیدکننده‌های معنا (نظیر *very*) و منفی‌کننده (نظیر *not*) برای تعیین گرایش نهایی کلمات مورد استفاده قرار گرفته است.

از میان کارهای دیگری که از برجسبزن نحوی به عنوان مبنای کار خود استفاده کرده‌اند، می‌توان به [۲۰]، [۲۱] و [۲۲] اشاره کرد. در تمام این کارها الگوهای ثابتی برای کشف نظر در نظر گرفته شده است جدول ۲-۲ و جدول ۳-۲، نشان‌دهنده الگوهای نحوی لغات است که نماینده ابراز نظر در متن می‌باشند. برای نمونه سطر اول جدول ۲-۲ بیان می‌کند که اگر کلمه اول صفت و کلمه بعد از آن اسم باشد، احتمال بیان یک نظر در این دو کلمه وجود دارد. جدول ۳-۲ نیز تعریف‌کننده علائم استفاده شده توسط برجسبزن نحوی می‌باشد.

جدول ۲-۲: الگوهای کشف حس [۲۱]

S.NO	First Word	Second Word	Third Word
1	JJ	NN or NNS	Anything
2	RB,RBR, or RBS	JJ	not NN n or NNS
3	JJ	JJ	not NN n or NNS
4	NN or NNS	JJ	not NN n or NNS
5	RB,RBR, or RBS	VB,VBD,VBN, or VBG	Anything

جدول ۲-۳: برچسب‌های مورد استفاده در برچسب‌زن نحوی لغات [۲۱]

S.NO	Tag	Description
1	NNP	Noun, proper, singular
2	NNPS	Noun, common, plural
3	RB	Adverb
4	JJ	Adjective or numeral, ordinal
5	JJR	Adjective, superlative
6	NN	Noun, common, singular
7	RBR	Adverb, comparative
8	VB	Verb, base, form
9	VBD	Verb, present participle, or gerund
10	WDT	WH-determiner
11	CC	Conjunction, coordinating
12	CD	Numeral, cardinal
13	DT	determiner

کارهای انجام شده زبان فارسی

در میان حجم انبوه مقالاتی که در دهه اخیر در حوزه نظرکاوی منتشر گشته‌اند، تعداد کمی از کارها به زبان‌های غیر انگلیسی پرداخته‌اند. در این میان کارهای محدودی نیز به زبان فارسی پرداخته‌اند که اولین کار ارائه شده پایان‌نامه‌ای در سال ۸۸ توسط خطیبی است [۲۳]. از روش مربع کای^۸ برای انتخاب ویژگی‌های مناسب برای الگوریتم دسته بند بهره برده است. مجموعه داده^۹ این تحقیق، نقدهای فیلم موجود در سایت ایران مانیا بوده است. اما متأسفانه ذکر دقیقی از آمار و خصوصیات این مجموعه نشده است. الگوریتم دسته‌بندی مورد استفاده در این کار نیز درخت تصمیم است. در نهایت در گزارش نهایی ارائه شده بیان شده است که استفاده از روش کای مربعی برای انتخاب مشخصه‌ها باعث بهبود نظرکاوی در زبان فارسی نمی‌شود.

پس از این تحقیق که از کیفیت چندان بالایی نیز برخوردار نمی‌باشد، تا چند سال در حوزه نظرکاوی زبان فارسی کار جدیدی ارائه نگردید. سپس در [۲۴] یک روش غیر ناظر مبتنی بر LDA به نام LDASA معرفی شده است. برای ایجاد منابع زبانی و بانک عبارات حسی از ترجمه بانک لغات انگلیسی استفاده شده است. در نهایت کار بر روی سه مجموعه داده‌ای (تلفن‌های همراه، هتلها و دوربینهای دیجیتال) که به صورت دستی از فروشگاه‌های اینترنتی گردآوری شده، اجرا گردیده است. دقت نهایی کار تا ۷۶ درصد نیز گزارش شده است. که در مقایسه‌ای که با روش با ناظر SVM-unigram به عنوان خط مبنا داشته است تا حدود ۹ درصد بهبود را به نمایش گذاشته است.

در یک ایده بسیار ساده [۲۵] مقایسه‌ای بین الگوریتم SVM و NBA بر روی یک پیکره محدود از نقدهای زبان فارسی صورت گرفته (فقط ۴۰۰ نظر!). ویژگی‌های ورودی به الگوریتم‌های دسته‌بندی نیز bigram, unigram و trigram می‌باشد.

در [۲۶] به پژوهش دقیق‌تری بر روی زبان فارسی پرداخته شده است. در این تحقیق بعضی مشکلات نظرکاوی که مخصوص زبان فارسی است مورد بررسی قرار گرفته است. از جمله این مشکلات می‌توان به پیشنوندهای صرفی مختلف، انواع فضا-خالیها، باین کلمه‌ای و درون کلمه‌ای و نیز حجم گسترده عبارات غیر رسمی و محاوره‌ای اشاره کرد. در این مقاله همچنین از متد اطلاعات متقابل^۱ برای استخراج بهترین ویژگی‌ها استفاده کرده است. الگوریتم یادگیری بی‌زین نیز در این تحقیق به کار رفته است. مقاله‌ی [۲۷] به نوعی دنباله این کار است. در این مقاله از چهار ویژگی به عنوان ورودی الگوریتم بی‌زین استفاده شده که شامل^۲ TFV ،^۳ DF ،^۴ MI و

^۱spaces

^۲mutual information

^۳Document Frequency

۲- مروری بر کارهای گذشته

$MMI^{۴۵}$ است که به صورت جداگانه استفاده شده است و در ارزیابی دقت نهایی؛ ویژگی جدید تعریف شده در این مقاله یعنی MMI توانسته اندکی بهبود را در مقایسه با سایر ویژگی‌ها ایجاد کند.

خلاصه تعدادی از روش‌های گفته شده در این بخش در جدول ۲-۴ آمده است.

جدول ۲-۴: خلاصه تعدادی از منابع ذکر شده

مقاله	پیکره	الگوهای نظری	خط مبنا	دقت	خلاصه روش	زبان
<i>Saraee</i> 2013 [۲۷]	۸۲۹ نظر-گوشی موبایل	تک کلمه حسی	SVM	$Precision < 78\%$ 90%	✓ بررسی برخی مشکلات خاص در نظر کاویزبان فارسی ✓ تعریف ویژگی جدید اطلاعات متقابل ✓ استفاده از الگوریتم بیزین	فارسی
<i>Jian</i> 2013 [۲۸]	هتل، فیلم و محصولات	عبارات ۳ کلمه ای	-	average accuracy = 69.1%	✓ استفاده از برجسبزن نحوی برای کشف نظر ✓ رتبه‌ای دهی نظر توسط SentiWordNet	انگلیسی
<i>Singh</i> 2013 [۲۰]	اخبار: استفاده از دو موضوع خبری	عبارات ۳ کلمه ای	SVM	$Accuracy < 67\%$ 74%	✓ استفاده از ابزار برجسبزن نحوی برای کشف نظر در موضوعات خبری	انگلیسی

Term Frequency Variance

Mutual Information

Modified Mutual Information

Shams 2012 [۲۴]	۴۰۰ نظر-هتل موبایل و دوربین تک کلمه حسی	SVM- uni	Accuracy < 78%	72%<	✓ روش غیر ناظر مبتنی بر LDA ✓ ترجمه مجموعه لغات حسی از انگلیسی به فارسی ✓ مقایسه با SVM و گزارش ۹ درصد افزایش در دقت نهایی	فارسی
Xia 2011 [۱۸]	حدود ۲۰۰۰ نظر DVD - تک کلمه حسی	SVM	Accuracy < 74%	72%<	✓ کشف الگوهای نحوی وابسته به دامنه و مستقل از دامنه	انگلیسی ی
Taboada 2011 [۱۹]	۱۹۰۰ نقد فیلم ۲۴۰۰ نظرسر (دوربین، چاپگر و کالسنکه کودک)	SVM	Accuracy < 72%	59%<	✓ بانک مجموعه لغات حسی به همراه گرایش حسی آن‌ها و قدرت آن ✓ بانک منفی‌کننده و تشدیدکننده‌ها	انگلیسی ی
Turney 2002 [۱۰]	اتومبیل- بانک - فیلم - سفر عبارات ۳ کلمه ای	-	Accuracy<84	65<	✓ برچسب‌گذاری نحوی اسناد ✓ شناسایی الگوهای خاص نظری ✓ تعیین گرایش مثبت و منفی از طریق مجاورت کلمات	انگلیسی ی

با توجه به جدول ۲-۴ روش‌های کنونی، حتی در زبان انگلیسی نیز از الگوهای نظری واحد جمله استفاده نکرده‌اند. اکثر روش‌ها یا بانک لغات حسی (تک کلمه حسی) و یا الگوهای ثابت (عبارات ۳ کلمه ای) مطرح شده توسط تورنی [۱۰] را مورد استفاده قرار داده‌اند.

نظرات مقایسه‌ای

جدای از نظراتی که مستقیماً بیان‌کننده احساس مثبت و منفی پیرامون یک موجودیت یا جنبه‌ای از آن هستند، نوع دیگری از نظرات وجود دارد که احساس خود را به صورت غیر مستقیم و با مقایسه دو موجودیت مشابه نشان می‌دهند، که نظر مقایسه‌ای نامیده می‌شود. نظرات مقایسه‌ای نه تنها به لحاظ معنایی، بلکه به

لحاظ ساختار نحوی نیز با نظرات عادی متفاوت‌اند. برای نمونه «کیفیت صدای این تلفن شگفت‌انگیز است» یک نظر عادی و «کیفیت صدای گوشی نوکیا بهتر از آیفون است» یک نظر مقایسه‌ای است. در حقیقت نظرات مقایسه هیچ اشاره مستقیمی به مثبت یا منفی بودن یک محصول ندارند و فقط به مقایسه بین دو یا چند موجودیت می‌پردازند. چه بسا در همین مثال کنونی، گوشی آیفون نیز دارای کیفیت صدای بسیار خوبی باشد، اما در این جمله در قیاس با گوشی نوکیا پایین‌تر دانسته شده است. [۱]

در این فصل بخش، ابتدا به چالش‌های این دسته خاص از نظرات اشاره می‌شود. سپس روش‌ها و کارهای صورت گرفته برای رفع آن بیان می‌شود. همچنین در این قسمت، به وجود دو دسته کلی نظرات مقایسه‌ای، یعنی نظرات قیاسی^{۴۶} و نظرات تفضیلی^{۴۷} توجه شده است. با توجه به اینکه به لحاظ ساختار این دو دسته نظر شبیه به هم هستند در اینجا از هر دو دسته تحت عنوان نظرات مقایسه‌ای یاد می‌شود.

۱-۱-۹ تعریف مقایسه

است. چندین نوع مختلف از نظرات مقایسه‌ای وجود دارد که آن‌ها را می‌توان تحت دو عنوان مقایسه‌ی قابل درجه بندی^{۴۸} و مقایسه‌ی غیر قابل درجه بندی^{۴۹} عنوان کرد. یک نظر مقایسه‌ای در واقع نشان‌دهنده یک رابطه از میزان شباهت یا تفاوت دو محصول یا بیشتر است. چندین نوع مختلف از نظرات مقایسه‌ای وجود دارد که آن‌ها را می‌توان تحت دو عنوان مقایسه‌ی قابل درجه بندی^{۴۸} و مقایسه‌ی غیر قابل درجه بندی^{۴۹} عنوان کرد.

مقایسه قابل درجه بندی:

^{۴۸}Comparative opinions
^{۴۹}Superlative opinions
^{۵۰}Gradable comparison
^{۵۱}Non-gradable comparison
^{۵۲}Gradable comparison
^{۵۳}Non-gradable comparison

۲- مروری بر کارهای گذشته

این نوع مقایسه امکان ترتیب بندی موجودیت‌های دخیل در مقایسه را فراهم می‌کند. سه زیر دسته از این نوع مقایسه قابل ذکر است.

✓ مقایسه غیر برابر: این مقایسه یک رابطه از نوع «بزرگ‌تر» و یا «کوچکتر» را بیان می‌کند. (عمر باتری الف بهتر از ب است)

✓ مقایسه برابر: بیان کننده رابطه از نوع «مساوی» است. (عمر باتری الف مانند ب عالی است)

✓ مقایسه برترین: بیان کننده رابطه از نوع «برتری یکی نسبت به همه» هست. (باتری الف بهترین طول عمر را بین گوشی‌های حال حاضر بازار دارد)

مقایسه غیر قابل درجه بندی:

Commented [m7]: در تیترو ارجاع نباید بیاید

مقایسه غیر قابل درجه بندی به مقایسه بین چند موجودیت گفته می‌شود به نوعی که موجودیت‌هایی که در مقایسه شرکت دارند قابل درجه بندی نیستند. این دسته خود شامل سه زیر دسته است:

۱. موجودیت یک نسبت به موجودیت دو متفاوت است. (گوشی الف با گوشی دو متفاوت است)

۲. موجودیت‌ها هر کدام ویژگی خاصی را دارند (لپ‌تاپ الف از مودم خارجی و لپ‌تاپ ب از مودم داخلی استفاده می‌کند)

۳. یک موجودیت دارای یک ویژگی است که موجودیت دیگر نیست. (تبلت الف با هدفون و تبلت ب بدون هدفون عرضه شده است)

در این تحقیق تمرکز بر روی مقایسه قابل درجه بندی معطوف است. مقایسه غیر قابل درجه بندی همچنان به عنوان یک چالش اساسی است که شناسایی و غلبه بر آن کاری بس دشوار است.

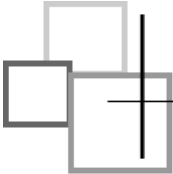
صورت نگرفته است. در طرح پیشنهادی این مقاله که در فصل آینده مطرح می‌شود. نظرات مقایسه‌ای و عادی به صورت مشترک مورد مطالعه قرار خواهند گرفت.

خلاصه فصل

در این فصل اهم کارهای صورت گرفته در حوزه نظرکاوی که با روش پیشنهادی ارتباط نزدیکی داشته‌اند، مرور گشت. ابتدا کارهای مطرحی که از الگوریتم‌های یادگیری ماشین بهره برده‌اند، یاد شد. سپس به الگوریتم‌های بدون ناظر پرداخته شد. در این بخش بیاد شد که بسیاری از روش‌هایی که تا کنون در این دسته جای گرفته‌اند، از چند الگوی ثابت چند کلمه‌ای استفاده نموده‌اند. این الگوهای به صورت دستی و توسط زبان شناس انتخاب شده‌اند و برای زبان انگلیسی و به عنوان نماینده بیان حس در جمله مورد استفاده قرار می‌گیرند.

در ادامه فصل کارهای انجام شده مربوط به زبان فارسی مورد بررسی قرار گرفت. همانطور که در این بخش ذکر شد تا کنون کارهای فراوانی برای شناسایی احساس نظرات فارسی صورت نگرفته است. بعلاوه محدود کارهای انجام شده نیز از پوشش و عمق مناسبی برخوردار نبوده‌اند و هنوز جای کار فراوانی در این حوزه برای محققان فارسی زبان موجود است.

در انتهای این فصل نیز برخی از تحقیقات صورت گرفته بر روی نظرات مقایسه‌ای ذکر گردید. همان طور که در این بخش بیان شد اکثر کارهای صورت گرفته بر روی نظرات، تمرکز خود را به نظرات عادی معطوف داشته‌اند و نظرات مقایسه‌ای کمتر مورد توجه قرار گرفته‌اند.



۳- روش پیشنهادی

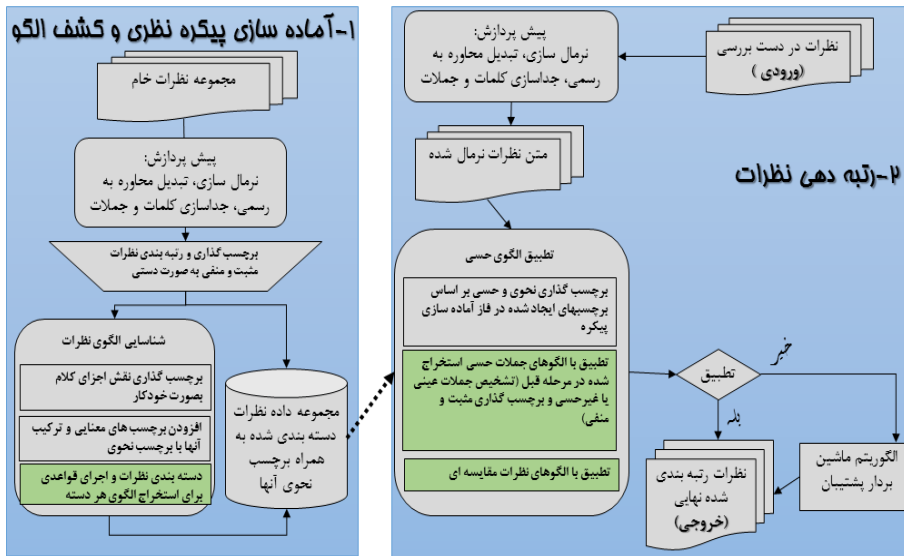
مقدمه

در فصل‌های پیشین چالش‌های جدی مطرح در نظر کاوی ذکر شد. در این فصل ابتدا در بخش ۳-۲ ساختار روش پیشنهادی به صورت کلی نمایش داده می‌شود و اجزای آن ذکر می‌شود. در این ساختار، تلاش شده تا اجزای مختلف سیستم و نحوه ارتباط بین آن‌ها به روشنی بیان شود. در ادامه این فصل، در سایر بخش‌ها اجزای مختلفی که در سیستم پیشنهادی مطرح شد، با جزئیات دقیق مورد بررسی قرار می‌گیرد.

ساختار روش پیشنهادی

همانطور که پیش‌تر مطرح گردید در اغلب روش‌های فعلی با انتخاب کلمات یا عبارات سعی در طبقه‌بندی نظرات به دسته‌های مثبت و منفی دارند. در روش پیشنهادی تلاش بر این است که با استفاده از برچسب‌زن نحوی اجزای کلام، بعد از تعیین نقش کلمات الگوهای بیان حس در واحد جمله شناسایی شوند. بدین منظور، ابتدا برچسب میزان حس مثبت و منفی نظر، توسط داوران برای هر جمله مشخص می‌شود. سپس با بررسی این نظرات و با استفاده از روش‌های آماری و همینطور برخی از ویژگی‌های شبکه‌ها و ژگان الگوهای خاص هر نظر مشخص می‌گردد. تلاش نهایی بر این است که از طریق الگوهای خاصی که در جملات برچسب خورده‌ی مجموعه داده شناسایی می‌شوند بتوان جایگاه دقیق حس در جمله را شناسایی نمود و سایر نظرات مورد بررسی را خارج از روش‌های بدون ساختار قبلی نظیر کیف لغات برچسب‌گذاری نمود.

در حقیقت در این تحقیق به نوعی سعی می‌شود که جایگاه احساس در واحد جمله به شکل دقیق شناسایی شود به جای آن که بر مبنای چند الگوی جزئی از پیش تعریف شده به دنبال کشف نظر باشیم. پس از کشف الگوهای مختلف و ذخیره آن‌ها در بانک داده، این الگوها با نظرات در دست بررسی مطابقت داده می‌شود و پس از کشف الگوی نظر، احساس مثبت و یا منفی آن تعیین می‌شود. در تصویر ذیل روند اجرای کار ترسیم شده است.



شکل ۱-۳: معماری طرح پیشنهادی

روند انجام کار مطابق شکل ۱-۳: معماری طرح پیشنهادی شکل ۱-۳ در زیر به صورت خلاصه تشریح می‌شود.

در ادامه این فصل هر یک از این گامها در زیر بخشهای جداگانه با جزئیات دقیق تر بحث می‌شود.

گام اول (جمع آوری مجموعه داده): گام اول جمع آوری مجموعه نظرات و یا انتخاب پیکره داده نظری متناسب با دامنه انتخابی است. در این تحقیق از مجموعه نظرات محصولات سایت فروشگاه دیجیکالا استفاده شده است. این مجموعه داده و جزئیات بیشتر آن در بخش (۱-۱) ۲۰-۱-۳-۴ مجموعه داده) شرح داده شده‌اند.

گام دوم (پیش پردازش): در این مرحله اعمال پیش پردازشی خاص بر روی جمله انجام می‌شود. از جمله این پردازشها می‌توان به تبدیل متن محاوره به رسمی، جدا کردن کلمات، جدا کردن جملات و اصلاح نامهای چندگانه محصولات اشاره کرد.

گام سوم (برچسب‌گذاری مثبت و منفی): در این مرحله جملات به کاربر داده می‌شود و جنبه‌های محصول به همراه احساس آن به صورت رتبه ۱ الی ۱۰ برچسب‌گذاری می‌شود. بعد از برچسب‌زنی احساسی توسط داوران؛ نظرات جمله‌بندی شده به همراه احساس و گرایش آن‌ها در مجموعه داده ذخیره خواهند شد.

گام چهارم (شناسایی الگوی نظرات): در این مرحله برچسب‌زنی نحوی لغات توسط ابزار «برچسب‌زن نحوی اجزای کلام» انجام خواهد گرفت. نظرات مقایسه‌ای نیز در قالبی جدا برچسب‌گذاری خواهند شد و همچنین ویژگی‌های خاص آماری از نظرات مثبت و منفی به صورت جداگانه استخراج خواهد شد. در این قسمت الگوها با توجه به جایگشت کلمات و برچسب‌های حسی و نحوی در جمله ایجاد خواهند شد. این قواعد به همراه اطلاعات آماری استخراج شده از ابزار برچسب‌زن نحوی، منجر به تولید الگوهای نهایی در واحد جمله خواهد شد. علاوه بر این، واحد مجزایی به ذخیره الگوهای جملات مقایسه‌ای خواهد پرداخت زیرا ماهیت جملات مقایسه‌ای متفاوت از جملات عادی است و الگوی متفاوتی نسبت به نظرات عادی خواهند داشت. در نهایت این الگوها نیز به همراه نظرات در مجموعه داده ذخیره خواهند شد.

گام پنجم (پیش‌پردازش نظرات تحت بررسی): در گام‌های قبلی به ساخت و اصلاح مجموعه داده خود پرداخته شد. در این گام مجموعه‌ی نظراتی را که برای تست در نظر گرفته شده است با توجه به پیش‌پردازش‌های صورت گرفته در گام دوم، به شکل یکسانی درآورده تا قابل مقایسه و پردازش باشند. بنابراین اعمال پیش‌پردازشی این در این گام شامل پردازش‌ها می‌توان به تبدیل متن محاوره به رسمی، جدا کردن کلمات، جدا کردن جملات و اصلاح نام‌های چندگانه محصولات می‌باشد.

گام ششم (کشف تطبیق الگوی حسی): در این مرحله با توجه به جملات نظری که در مجموعه داده به همراه الگوهای آن‌ها در اختیار است، تلاش می‌شود تا جملات نظری تحت بررسی به شبیه‌ترین الگو تخصیص داده شود و از این طریق حس دقیق جمله کشف شود و با توجه به آن برچسب مثبت و منفی به جمله اختصاص یابد. برای این کار در این مرحله از تطبیق جملات با جملات پایه و الگوهای ذخیره شده در مجموعه داده بهره برده شده است. همچنین از بخشی از روابط معنایی موجود در شبکه واژگان برای توسعه الگوی جمله در دست بررسی استفاده شده است. در این گام همچنین برچسب‌های نحوی هر نظر با برچسب‌های نظرات مقایسه‌ای نیز بررسی خواهد شد، تا در صورت تشخیص نظر مقایسه‌ای به آن به صورت جداگانه رسیدگی نماییم.

پیش‌پردازش مجموعه داده

پیش‌پردازش‌های اولیه که بر روی مجموعه نظرات جمع‌آوری شده اعمال شده است شامل جداسازی تبدیل محاوره به رسمی، جملات نظری و تشخیص کلمات است. ابزار مربوطه برای هر یک از این وظایف در ادامه بیان می‌شود.

۱-۱۱-۱-۱-۱-۱ تبدیل محاوره به رسمی

منظور از متن محاوره‌ای، نوشتن متن به شکلی است که معمولاً فارسی‌زبانان به آن شیوه، تکلم می‌کنند و با فارسی رسمی معیار متفاوت است. تبدیل متن محاوره‌ای به رسمی یکی از مسائل پیش‌رو برای پردازش زبان است که کاربردهای فراوانی در تحلیل و پردازش گفتار گوینده، بررسی وبلاگ‌ها، پیام‌های کوتاه و نظایر این‌ها می‌تواند داشته باشد.

اکثر نظرات نوشته شده در سایت‌های مختلف نیز با زبان محاوره‌ای بیانی می‌شود. برای نمونه تعدادی جمله محاوره‌ای در زیر آمده است.

- این گوشیش فوق لعادس
- بدتر ازین امکان نداشت واقعن تولید کنه

همان طور که در مثالها قابل ملاحظه است، جملات نظری محاوره‌ای می‌توانند ساختار نحوی زبان فارسی را تغییر دهند. بعلاوه ممکن است کلماتی با املاي غلط که در زبان محاوره است به کار رفته باشند. از آنجا که اکثر ابزارهای تولید شده برای پردازش متن، بر روی زبان رسمی تمرکز دارند، نیاز است که در این فاز جملات محاوره‌ای تا حد ممکن به جملات رسمی تبدیل شوند. برای تبدیل جملات محاوره به رسمی از ابزار تبدیل‌کننده جملات محاوره‌ای به رسمی تولید شده در آزمایشگاه فناوری وب دانشگاه فردوسی استفاده کردیم. خروجی مناسب برای جملات مثال پس از اعمال ابزار تبدیل‌کننده جملات محاوره‌ای باید به شکل زیر باشد.

- این گوشیش فوق لعادس ⇐ این گوشی آن (شرکت) فوق‌العاده است.
- بدتر ازین امکان نداشت واقعن تولید کنه ⇐ واقعاً بدتر از این، امکان نداشت تولید کند.

۱-۱-۱۲-۱-۱-۱۳-۱-۱-۱۲ یکسان سازی متن^۱

این گام تقریباً در تمام فرآیندهای پردازش زبان به چشم می‌خورد. در این گام به تغییر کاراکترهای متن به شکل یکنواخت آن پرداخته می‌شود. از آنجا که متون مختلف نظری توسط کاربران با سیستم‌های مختلفی نگاشته شده است، تعداد و نوع کاراکترهایی که به یک حرف خاص اشاره دارند کاملاً متفاوت است. برای نمونه چندین «ی» با استفاده از دکمه‌های صفحه کلید قابل

درج است و با کلمات «واقعا»، «واقعا»، «واقعا» هر سه به یک کلمه واحد اشاره دارند و نیاز است که با در نظر گرفتن یک شکل نگارش به وضعیت یکسانی نگاشت شوند.

۱۳-۱-۱-۱۴-۱ جداسازی جملات^۲

جداسازی جملات به عملی گفته می‌شود که بر اساس آن یک سند متنی به جملات صحیح شکسته می‌شود. از جمله علائمی که می‌تواند نشان‌دهنده مرز جملات باشد می‌توان به «»، «!»، «؟» و ... اشاره کرد. باید توجه داشت که هر کدام از این علائم با احراز شرایط خاص دیگری در جمله می‌توانند به عنوان جداکننده جملات در نظر گرفته شوند. برای نمونه علامت ویرگول باره‌امی‌تواند در میان جملات استفاده شود و یا علامت نقطه در عبارت «ه.ق.» یک جداکننده جمله محسوب نمی‌شود.

برای انجام این عمل از ابزاری به نام جداکننده جملات استفاده می‌شود. این ابزار باید با توجه به کاراکترهای جداکننده جمله در زبان فارسی، توانایی تشخیص جملات را در متن ورودی داشته باشد. برای ایجاد این ابزار باید ابتدا تمامی کاراکترها، نمادها و احیاناً قواعد دستوری که باعث شکسته شدن جملات می‌شوند، شناسایی گردند. با توجه به اینکه در بسیاری از پردازش‌های زبانی جمله مبنای پردازش است، خروجی دقیق این ابزار از درجه‌ی اهمیت بالایی برخوردار است.

نکته مهم دیگری در انجام این پیش‌پردازش برای متون نظری وجود دارد که حائز اهمیت است. از آنجا که متون نظری حاوی متن محاوره‌ای هستند در نگارش آن‌ها احتمال دارد توجه چندانی به علائم نگارشی نشود. از طرفی غلبه بر این امر در گام تبدیل محاوره به رسمی نیز کاری بس دشوار

است. بنابراین مناسب به نظر می‌رسد که بعد از انجام این مرحله، یک بار تمام جملات توسط نیروی انسانی مورد بازبینی قرار گیرد تا جملات صحیح از جملات ناصحیح جدا شوند.

۱-۱۴-۱-۱۵-۱- شناسایی کلمات

از آن جا که برچسب‌گذاری نحوی در واحد کلمه صورت می‌گیرد، نیاز به انجام شناسایی کلمات وجود دارد. مشکل اصلی در این گام شناسایی دو دسته از کلمات است، کلمات مرکب نظیر «خوش حال» و نیز کلمات به همراه وند است که به شکل‌های مختلف «کتابها»، «کتاب‌ها» و «کتاب‌ها» قابل نگارش هستند. یک ابزار شناساگر کلمات با دقت بالایی می‌تواند تاثیر مناسبی را در سایر پردازش‌ها و دقت نهایی داشته باشد.

برچسب‌گذاری دستی نظرات

از آنجا که پیکره آماده برای نظرکاوی زبان فارسی در اختیار نیست، تمام پیش‌پردازش‌های گفته شده بر روی مجموعه نظرات جمع‌آوری شده توسط خزنده وب اعمال شده‌اند. بنابراین نیاز است که بعد از انجام پیش‌پردازش‌های اولیه نظراتی که در واحد جمله آماده شده‌اند توسط داور انسانی تعیین‌گرایش می‌شوند. در این مرحله چهار برچسب در نظر گرفته شده است. در زیر مثالی از هر دسته آورده شده است.

برچسب‌های نظری مثبت و منفی که برای جملات نظری عادی مورد استفاده قرار می‌گیرند:

۱. گوشه از امکانات خوبی برخوردار است. (نظری مثبت)

۲. من که زیاد از امکانات آن راضی نیستم. (نظری منفی)

برچسب مقایسه‌ای که برای جملات مقایسه‌ای مورد استفاده قرار می‌گیرد:

۳. این گوشه به نظر از گوشه الف مناسبتر به نظر می‌آید.

برچسب جملات غیر نظری برای آن دسته از جملاتی که بار احساسی یا نظری ندارند.

۴. من هفته گذشته این گوشی رو از سایت الف خریداری کردم و دیروز به دستم رسید.

شناسایی الگوی جملات نظری

در این فاز تلاشی شود که جملات نظری مختلف به نحو مناسبی دسته‌بندی شوند و با توجه به نقش نحوی هر دسته، الگوی نحوی مناسب را که نشان‌دهنده جملات آن دسته باشد، ارائه نمود. این الگوها در ادامه برای تشخیص جملات نظری دست بررسی و رتبه دهی به آن‌ها مورد استفاده قرار می‌گیرد.

۱-۱-۱۵- برچسب‌گذاری نقش نحوی اجزای سخن

برچسب‌گذاری اجزای واژگانی کلام عمل انتساب برچسب‌های واژگانی به کلمات و نشانه‌های تشکیل‌دهنده یک متن است، به صورتی که این برچسب‌ها نشان‌دهنده نقش کلمات و نشانه‌ها در جمله باشند. درصد بالایی از کلمات از نقطه‌نظر برچسب واژگانی دارای ابهام هستند، زیرا کلمات در جایگاه‌های مختلف برچسب‌های واژگانی متفاوت دارند. بنابراین برچسب‌گذاری واژگانی عمل ابهام‌زدایی از برچسب‌ها با توجه به زمینه‌مورد نظر است. [3]

در زیر مثالی از پردازش صورت گرفته توسط برچسب‌زن نحوی اجزای کلام زبان فارسی را ملاحظه می‌کنید. هر کدام از برچسب‌های اختصاری که در خروجی مشاهده می‌شود، حاوی یک نقش نحوی می‌باشد. به عنوان نمونه برچسب <N> نماینده اسم، برچسب <V> نماینده فعل، برچسب <PR> نماینده ضمیر می‌باشد.

✓ جمله ورودی:

من امروز یک گوشی فوق العاده خریداری کردم.

۳- سیستم پیشنهادی

از میان برجسب‌های این جدول برجسب‌های شماره ۹ تا ۱۱ برجسب‌های جدیدی هستند که برای اولین بار در این تحقیق مورد استفاده قرار گرفته‌اند. همچنین تعریف و یا نحوه استفاده برجسب‌های شماره ۵ و ۸ نیز متفاوت است که تفاوت آن در ادامه ذکر می‌شود.

جدول ۱-۳: برجسب‌های حسی تعریف شده

ردیف	نام برجسب	اختصار	توضیح	مثال
۱	حس	SENTI	حس اصلی جمله را بیان می‌کند.	«گوشی خوبی است» خوب
۲	فعل غیر حسی	V	افعال که نقش نظری ندارند.	«نوکیا این گوشی رو زیبا ساخته» ساخته
۳	شدت‌دهنده	INT	تاثیر حس را افزایش یا کاهش بسیار می‌دهد.	«گوشی بسیار خوبی است» بسیار
۴	منفی کننده	NEG	معنی حس را معکوس می‌کند.	«نه مزیتی دارد» نه
۵	موجودیت	ENT	موجودیتی که نظر حول آن بیان اکسپریا می‌شود.	«کیفیت صفحه نمایش اکسپریا..» اکسپریا
۶	جنبه	ASP	ویژگی مشخصی از موجودیت مانیتور است.	«کیفیت مانیتور اکسپریا..» مانیتور
۷	زیر جنبه	ASS	ویژگی مشخصی از جنبه است.	«کیفیت صفحه نمایش اکسپریا..» کیفیت
۸	فعل حسی	VS	فعلی که بیان کننده حس است.	«من واقعا این گوشی را پسندیدم» پسندیدم
۹	تشریح قابل حذف	EX	حذف آن در معنی جمله بی تاثیر به نظرم است.	«این گوشی به نظرم واقعا عالیه» حذف آن در معنی جمله بی تاثیر به نظرم است.
۱۰	تاکید	IM	شبه جمله‌ای که به عنوان تاکید شک نکند می‌آید.	«عالیه بخريدش، شك نكنيد» شک نکند
۱۱	دلیل آوری	RE	بین جمله نظری و جمله عینی به خاطر اینکه	

	می آید.			
--	---------	--	--	--

برچسب موجودیت: این برچسب در اکثر منابع نظر کاوی محصولات تجاری نشان دهنده نام محصول است، اعم از نام رسمی که کمپانی برای محصول تعیین می کند و نیز نام های مصطلح که مصرف کنندگان از آن برای اشاره به محصول مورد نظر استفاده می کنند. اما در این تحقیق با توجه به مشکلات زبان فارسی مانند نبود تشخیص دهنده ضمائر ضمیری که به موجودیت اشاره کنند نیز برچسب موجودیت خواهند خورد. به عنوان نمونه در جمله «این گوشی امکانات فوق العاده ای دارد» عبارت «این گوشی» که در نقش موجودیت است برچسب موجودیت خواهد خورد.

برچسب فعل حسی: در زبان انگلیسی اکثر افعالی که می توانند نقش حسی داشته باشند مانند «love» در نقش اسم و فعل به یک شکل و به صورت تک کلمه ای می آیند. حال آنکه در زبان فارسی افعال چند بخشی و مرکب زیادی نظیر «پشیمان شدن» وجود دارد که می توانند نقش حسی ایفا کنند. برای کشف دقیق تر الگوی نظرات در اینجا افعال حسی از سایر افعال جدا شده اند. از میان برچسب های جدول ۱-۳: برچسب های حسی تعریف شده جدول ۱-۳ برچسب های ۱، ۲، ۴، ۵، ۶، ۷ و ۸ در الگوی نهایی مورد استفاده قرار می گیرند و سایر برچسب ها برای حذف برخی از اجزای غیر نظری جمله مورد استفاده قرار می گیرند.

۱-۱-۱۷-۱-۱-۱۸-۱-۱-۱۹-۱-۱-۲۰-۱-۱-۲۱-۱-۲۲-۱-۲۳-۱-۲۴-۱-۲۵-۱-۲۶-۱-۲۷-۱-۲۸-۱-۲۹-۱-۳۰-۱-۳۱-۱-۳۲-۱-۳۳-۱-۳۴-۱-۳۵-۱-۳۶-۱-۳۷-۱-۳۸-۱-۳۹-۱-۴۰-۱-۴۱-۱-۴۲-۱-۴۳-۱-۴۴-۱-۴۵-۱-۴۶-۱-۴۷-۱-۴۸-۱-۴۹-۱-۵۰-۱-۵۱-۱-۵۲-۱-۵۳-۱-۵۴-۱-۵۵-۱-۵۶-۱-۵۷-۱-۵۸-۱-۵۹-۱-۶۰-۱-۶۱-۱-۶۲-۱-۶۳-۱-۶۴-۱-۶۵-۱-۶۶-۱-۶۷-۱-۶۸-۱-۶۹-۱-۷۰-۱-۷۱-۱-۷۲-۱-۷۳-۱-۷۴-۱-۷۵-۱-۷۶-۱-۷۷-۱-۷۸-۱-۷۹-۱-۸۰-۱-۸۱-۱-۸۲-۱-۸۳-۱-۸۴-۱-۸۵-۱-۸۶-۱-۸۷-۱-۸۸-۱-۸۹-۱-۹۰-۱-۹۱-۱-۹۲-۱-۹۳-۱-۹۴-۱-۹۵-۱-۹۶-۱-۹۷-۱-۹۸-۱-۹۹-۱-۱۰۰ برچسب های نحوی

برچسب های نحوی که برای ساخت الگوهای نظری مورد استفاده قرار خواهند گرفت در جدول

۲-۳: برچسب های نحوی استفاده شده به همراه فراوانی آنها [۳۳] جدول ۲-۳ آماده اند.

جدول ۲-۳: برچسب‌های نحوی استفاده شده به همراه فراوانی آن‌ها [۲۳]

برچسب اجزای سخن	تعداد حضور	درصد حضور
PREP	۵۷۹۶۰	۱۱/۶۴
N	۱۹۵۴۳۱	۳۹/۲۴
PUNC	۴۵۵۹۷	۹/۱۵
ADJ	۳۶۸۳۳	۷/۳۹
PREM	۱۰۷۹۵	۲/۱۷
CONJ	۲۲۸۴۵	۴/۵۹
V	۶۲۸۷۶	۱۲/۶۲
SUBR	۱۴۱۸۲	۲/۸۵
PRENUM	۶۱۹۵	۱/۲۴
POSTP	۱۵۱۸۹	۳/۰۵
PR	۱۸۹۷۳	۳/۸۱
ADR	۱۸۲	۰/۰۴
ADV	۸۵۵۵	۱/۷۲
IDEN	۸۵۱	۰/۱۷
PSUS	۵۲۶	۰/۱۱
POSNUM	۵۷۱	۰/۱۱
PART	۵۲۰	۰/۱۰

این برچسب‌ها مطابق با برچسب‌های به دست آمده نقش نحوی لغات از ابزار برچسب‌زن اجزای کلام هستند. همچنین این برچسب‌ها، در پیکره «وابستگی نحوی زبان فارسی دادگان» برای نقش‌های مختلف کلمات در یک جمله انتخاب شده است.

۱۱۸-۱-۱-۱۱۸ ترکیب برچسب‌ها و کشف الگوی مناسب برای هر دسته

در این بخش ترکیب دو دسته برچسب قبلی و کشف الگو در قالب یک مثال به تصویر کشیده شده است. جمله‌ای در **جدول ۳-۳: ترکیب برچسب‌های حسی و برچسب‌های نحوی جدول ۳-۳** به همراه برچسب‌های برچسب‌زن نحوی، برچسب‌های حسی تعریف شده در این مقاله و برچسب‌های ترکیب شده که منجر به تولید الگومی شود، به نمایش گذاشته شده است.

جدول ۳-۲: ترکیب برچسب‌های حسی و برچسب‌های نحوی

گوشیش عالی هست.	این گوشی اکسپریا زد مناسب است.	این گوشی از نظر من بسیار عالی است.	جمله
NADJV	PREMNNADJV	PREMNPREP NPRADVADJV	برچسب نحوی
(گوشیش) (مناسب) (است) موجودیت + حس + فعل	(این گوشی اکسپریا زد) (مناسب) (است) موجودیت + حس + فعل	(این گوشی) (بسیار) (عالی) (است) موجودیت + شدت‌دهنده + حس + فعل	برچسب حسی
موجودیت + حس + فعل	موجودیت + حس + فعل	موجودیت + N + PREP + N + شدت‌دهنده + حس + فعل	برچسب‌های ترکیب شده
موجودیت + حس + فعل			دسته
الگوی اول: موجودیت + N + PREP + N + حس + فعل الگوی دوم: موجودیت + حس + فعل			الگوهای محتمل
الگوی اول: V + SENTI + N + PREP + N + ENT الگوی دوم: V + SENTI + ENT			الگوها با اختصار

Commented [m11]: دقت شود که رنگ در چاپ دیده نخواهد شد

همانگونه که در **جدول ۳-۳: ترکیب برچسب‌های حسی و برچسب‌های نحوی جدول ۳-۳** مشخص است در بخش برچسب‌گذاری نحوی، سه جمله آورده شده کاملاً متفاوت به نظر می‌رسند. حال آنکه پس از برچسب‌گذاری حسی سه جمله بسیار شبیه‌تر به هم شدند. از آنجا که در بخش برچسب‌گذاری حسی به این مطلب اشاره شد که شدت‌دهنده جزو برچسب‌های تعیین‌کننده

دسته الگو نمی باشد، با حذف این برچسب هر سه جمله در یک دسته یکسان به لحاظ محل قرارگیری برچسب‌های حسی واقع خواهند شد. این دسته شامل سه برچسب است که برچسب اول موجودیت، برچسب دوم حس و برچسب سوم فعل باشد. پس از تشخیص دسته با استفاده از برچسب ترکیب شده تلاشی شود تا الگو یا الگوهای متعلق به هر دسته شناسایی شود. البته امر شناسایی الگو نیاز به حجم داده بیشتری دارد تا احتمالات دقیقتری قابل استخراج باشد اما در فضای کوچک مسئله طرح شده با توجه به برچسب‌های ترکیبی دو الگو محتمل است. این دو الگو در سطر انتهایی جدول آمده است.

۱-۱۹-۱-۱-۱- نظرات مقایسه‌ای

در بخش در نظر گرفته شده برای نظرات مقایسه‌ای تمام مراحل برچسب‌زنی حسی، دسته‌بندی جملات نظری و کشف الگوی جملات به مانند جملات عادی صورت می‌گیرد. تنها تفاوتی که در این بخش وجود دارد این است که نظرات مقایسه‌ایمانند نظرات عادی به دو گروه مثبت و منفی تقسیم نمی‌شوند. نظرات مقایسه‌ایمی‌توانند شامل گروه‌های زیر باشند که کشف الگو در هر دسته به صورت جداگانه انجامی‌شود.

- ✓ برابری (این دو گوشه واقعا هیچی از هم کم ندارند)
- ✓ برتری (از هر لحاظی از گوشه ب سرتراست)
- ✓ بهترین (این گوشه واقعا بهترین گوشه حال حاضر در بازار است)

تطبيق الگوهای نظری و رتبه دهی به سند نظری

پس از کشف الگوهای نظری در مرحله قبل، در این گام نیاز است که اسناد نظری توسط الگوهای کشف شده رتبه دهی شوند. برای انجام این عمل، ابتدا اسناد تحت بررسی برچسب‌گذاری خواهند شد (حسی و نحوی). سپس برچسب ترکیبی ایجاد شده با مجموعه الگوهای موجود در مجموعه

داده مقایسه خواهد شد. در این مرحله دو حالت ممکن است رخ دهد: اول تطبیق برچسب با الگوهای موجود، دوم عدم تطبیق برچسب ترکیبی جمله مد نظر با هیچ یک از الگوها. در حالت اول که الگوی نظر به دست آمده باشد با توجه قاعده مربوطه در **جدول ۳-۴: قواعد مورد استفاده پس از کشف الگو برای رتبه دهی به سند جدول ۳-۴** گرایش جمله نظری شناسایی خواهد شد. مجموعه نظرات تحت بررسی که الگویی برای آن‌ها کشف نشود به الگوریتم یادگیری SVM داده خواهند شد تا توسط این الگوریتم گرایش حسی آن‌ها مشخص شود.

جدول ۳-۴: قواعد مورد استفاده پس از کشف الگو برای رتبه دهی به سند

نوع الگو	تعریف	قواعد مورد استفاده برای تعیین گرایش
الگوهای الف	الگوهایی که در آن‌ها برچسب SENTI وجود ندارد	<ul style="list-style-type: none"> ✓ این الگوها بر طبق مثبت و منفی بودن فعل جمله تعیین گرایش خواهند شد. ✓ در صورت عدم وجود فعل منفی‌کننده‌ها بار حسی جمله را منفی خواهند کرد و در غیر این صورت مثبت ✓ در صورت وجود فعل حسی، بار معنایی جمله توسط آن تعیین می‌شود.
الگوهای ب	الگوهایی که در آن‌ها یک SENTI وجود دارد	<ul style="list-style-type: none"> ✓ این الگوها بر اساس گرایش SENTI تعیین گرایش می‌شوند. ✓ فعل منفی گرایش جمله را تغییر خواهد داد. ✓ منفی‌کننده گرایش SENTI را تغییر خواهد داد.
الگوهای ج	الگوهایی که در آن‌ها بیش از یک SENTI وجود دارد که هم گرایش هستند	<ul style="list-style-type: none"> ✓ این الگوها بر اساس گرایش یکی از SENTIها تعیین گرایش می‌شوند. ✓ فعل منفی گرایش جمله را تغییر خواهد داد. ✓ منفی‌کننده گرایش SENTI را تغییر خواهد داد.
الگوهای د	الگوهایی که در آن‌ها بیش از	<ul style="list-style-type: none"> ✓ بر مبنای مثبت و منفی بودن SENTIها در جملات متعلق به الگو تعیین گرایش می‌شود.

<p>✓ هر تغییر <i>SENTI</i> بار گرایشی جمله را تغییر خواهد داد</p> <p>✓ فعل منفی گرایش جمله را تغییر خواهد داد.</p> <p>✓ منفی کننده گرایش <i>SENTI</i> را تغییر خواهد داد.</p>	<p>یک <i>SENTI</i> با گرایش های</p> <p>مختلف وجود دارد.</p>
---	---

مدنظر است. پس از اعمال برچسب زنی، برچسب این جمله (از راست به چپ) به شکل زیر خواهد بود:

فرض کنید جمله نظری « این گوشی سونی قیمت بالایی عرضه شده است» برای برچسب زنی مدنظر است. پس از اعمال برچسب زنی، برچسب این جمله (از راست به چپ) به شکل زیر خواهد بود:

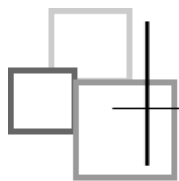
«موجودیت+ جنبه+ حس+ فعل» یا به اختصار «*V+SENTI+ASP+ENT*»

همچنین فرض نمایید که این برچسب در مجموعه الگوها وجود دارد. با توجه به اینکه در این الگو یک برچسب *SENTI* موجود است، این جمله در الگوهای نوع «ب» **جدول ۳-۴: قواعد مورد استفاده پس از کشف الگو برای رتبه دهی به سند جدول ۳-۴** قرار می گیرد. پس به سراغ قواعد این نوع الگو می رویم. با توجه به اینکه فعل مثبت است و منفی کننده ای نیز در این الگو وجود ندارد، تنها قاعده اول این ردیف مورد استفاده است. با توجه به اینکه کلمه «بالا» برای «قیمت» یک صفت منفی است بار معنایی این جمله منفی خواهد شد.

خلاصه فصل

در این فصل ساختار کلی روش پیشنهادی مورد بحث قرار گرفت. در ابتدا در مورد پیکره نظر کاوی و چگونگی ایجاد آن بحث شد. روش پیشنهادی در دو مرحله اصلی آماده سازی پیکره نظری و کشف حس نظرات مورد بررسی تشریح گردید. در مرحله اول به آماده سازی پیکره و برچسب گذاری نحوی و حسی

پرداخته شد. در این مرحله پیش پردازش‌های اولیه بر روی نظرات جمع‌آوری شده اعمال می‌شود. سپس برچسب‌گذاری بر روی اسناد صورت می‌گیرد. این نظرات برچسب خورده به همراه الگوهای کشف شده از آن‌ها در مجموعه داده ذخیره می‌شوند. در مرحله دوم نظرات ورودی با الگوهای ایجاد شده در مرحله قبل تطبیق داده شده و نظر ورودی تعیین حس می‌گردد. قواعد مناسب برای تطبیق الگوها با جملات نظری نیز در انتهای این فصل آمده است.



۴- پیاده سازی و ارزیابی سیستم

نماید. در زیر پایگاه‌هایی را که معمولاً برای این حوزه مورد بررسی قرار می‌گیرد شرح داده شده است [۳۴]:

وبلاگ‌ها! وبلاگ‌ها سایت‌های کوچکی هستند که اشتراک‌گذاری اطلاعات و ایجاد مطلب در آن براحتی امکان‌پذیر است. از این روی تعداد زیادی از افراد بسته به نوع علاقه و زمینه فعالیت خود نسبت به ایجاد و مدیریت وبلاگشان اقدام می‌کنند. در این منابع توده انبوهی از اطلاعات نهفته است که بسیاری از مطالعات آن را بعنوان منبع داده خود، انتخاب نموده‌اند.

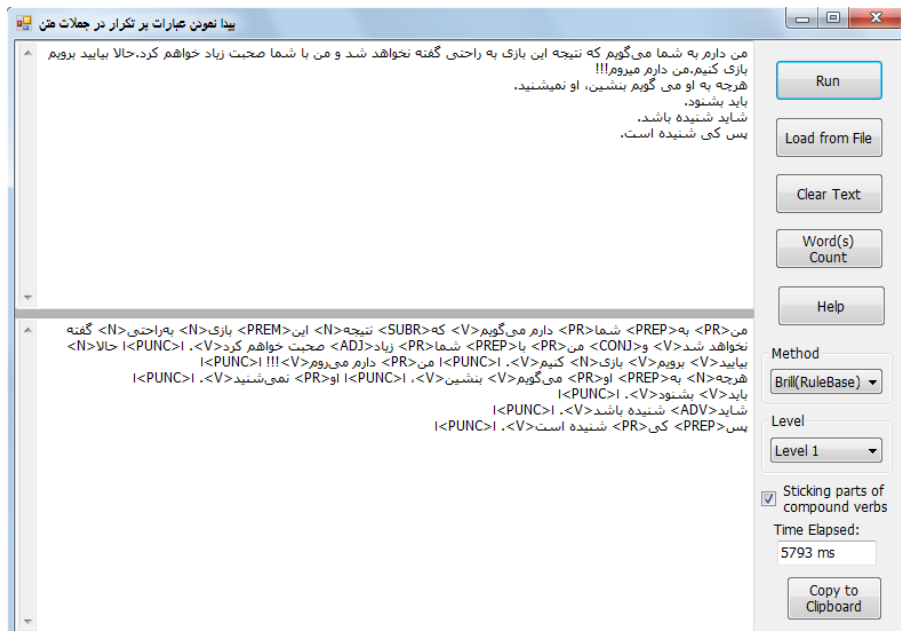
میکرووبلاگ‌ها: شبکه‌های اجتماعی نظیر Facebook, Twitter, Tumbler و ... به عنوان یکی از پرکاربردترین ابزارها در تبادل روزانه داده در اینترنت محسوب می‌شوند. نظرات فراوانی که در این میان تولید می‌گردد می‌تواند منبع مناسبی برای مجموعه داده حوزه نظرکاوی باشد.

نقدها: حجم وسیعی از نقدها در اینترنت موجود می‌باشد. یکی از کاربردهای مهم آن‌ها، تسهیل تصمیم‌گیری خرید برای کاربران است. سایتهای www.gsmarena.com (نقد موبایل)، www.amazon.com (نقد محصول)، CNETdownload.com (نقد محصول) و ... میزبان میلیون‌ها نقد از محصولات مختلف توسط کاربران هستند.

در حال حاضر مجموعه‌ی داده نظراتی از سایت دیجیکالا^۱ در آزمایشگاه فناوری وب دانشگاه فردوسی تهیه گشته است، که بخشی از آن برای این پروژه مورد استفاده قرار گرفت. علت انتخاب نظرات سایت دیجیکالا این است که به نسبت سایر سایت‌های فروش محصولات آنلاین از شهرت بیشتری برخوردار است و حجم نظرات مناسبی در آن قرار دارد. مجموعه نظرات جمع‌آوری شده ۷۹۰۰ نوع محصول در ۱۰ دسته (دوربین عکاسی، تلفن همراه و ...) را در بر می‌گیرد. ۳۱۷۳۰ نظر

^۱blogs

^۲<http://www.digikala.com>



شکل ۴-۱: برنامه برچسب زن نحوی اجزای کلام متعلق به آزمایشگاه فناوری وب دانشگاه فردوسی

در شکل ۴-۱، برچسب‌های نحوی اجزای کلام متعلق به آزمایشگاه فناوری وب دانشگاه فردوسی (مبتنی بر سیستم برچسب‌زنی نحوی اجزای کلام متعلق به آزمایشگاه فناوری وب دانشگاه فردوسی)

Commented [m13]: در تمامی این ارجاعات باید فقط مثلا نوشته شود: در شکل ۴-۱ نمونه ای ...

نمونه‌ای از ابزار استفاده شده نشان داده شده است. در این شکل در کادر بالا جمله ورودی نشان داده شده و پس از انجام تنظیمات با فشردن دکمه *run* خروجی در کادر پایین قابل مشاهده است. همچنین کتابخانه آماده این مجموعه ابزار نیز در دسترس است که از آن برای انجام پروژه بهره گرفته شد. لازم به ذکر است که در طول اجرای پروژه تلاش شد با ایجاد گزارشهای مناسب، از خطاهای مشاهده شده در پیاده‌سازی دقت ابزارهای پیش‌پردازش بهبود داده شود تا این ابزارها برای سایر پژوهش‌های پردازش زبانی بر روی زبان فارسی از دقت و کیفیت بالاتری برخوردار باشند.

لازم به ذکر است که در حال حاضر سه داور بر روی پیکره‌ای که در گام اول ایجاد شده، برچسب‌گذاری نموده‌اند. قریب به ۲۰ هزار جنبه و احساس نیز توسط این سه داور رتبه‌بندی شده‌اند که از آنها نیز به عنوان برچسب‌های حسی استفاده شده است.

دسته‌ها و الگوهای استخراج شده ۱-۲۳-۱-۱-۲۴-۱-۱

در بخش روش پیشنهادی اشاره شد که برای شناسایی الگوی جملات نظری ابتدا دسته‌های جملات تعیین می‌شود. این دسته‌ها از جایگشت‌های برچسب‌های حسی که در متن مشاهده شده‌اند به دست می‌آید. در **جدول ۴-۴: نمونه دسته‌های پرتکرار جدول ۴-۴** تعدادی از دسته‌های نظری که بیشترین نرخ وقوع در بخش آموزش پیکره را داشتند، نشان داده است.

جدول ۴-۴: نمونه دسته‌های پرتکرار

تعداد وقوع در بخش آموزش	برچسب چهارم	برچسب سوم	برچسب دوم	برچسب اول
236			V	SENTI
147		V	SENTI	ENT
120		V	SENTI	ASP
76		V	SENTI	SENTI
45	V	SENTI	SENTI	ENT
36			V	ENT

همچنین بیشترین برچسب‌های نحوی که بین هر دو برچسب حسی می‌توانند واقع شوند و در شکل‌گیری الگو موثر باشند، در **جدول ۴-۵: برخی برچسب‌های حسی و برچسب نحوی مجاز بین آن‌ها جدول ۴-۵** آمده‌اند. در حقیقت این جدول نشان دهنده نقش‌های نحوی مجازی است که می‌توانند بین دو برچسب حسی خاص واقع شوند. در این جدول *empty* به معنای امکان نبودن هیچ نقش نحوی بین دو برچسب حسی است. برچسب‌های *start* و *end* نیز برچسب‌های خاتمه جمله هستند.

جدول ۴-۵: برخی برچسب‌های حسی و برچسب نحوی مجاز بین آن‌ها

برچسب حسی دوم	برچسب نحوی مجاز بین دو برچسب حسی	برچسب حسی اول
v	Empty	senti

v	empty	end
v	PUNC	end
start	empty	asp
v	empty	senti
start	empty	ex
senti	N + PUNC	v
start	empty	in
senti	CONJ + PUNC	senti

در پیوست (جداول جدول ۷-۲: دسته‌های ایجاد شده برای استخراج الگو بر اساس جایگشت برچسب‌های حسی ۷-۳: نقش‌های نحوی مجاز بین دو برچسب حسی که با ترکیب با برچسب حسی منجر به ساخت الگو میشوند. ۷-۴) به ترتیب لیست کامل دسته‌ها، نقش‌های نحوی مجاز بین هر دسته و الگوهای نهایی به دست آمده را نشان می‌دهد.

ارزیابی سیستم

در این بخش تلاش می‌شود تا ارزیابی جامعی از کیفیت و کمیت سیستم پیشنهادی صورت گیرد. بعلاوه سعی شده تا با مقایسه سیستم پیشنهادی با سایر سیستم‌های زبان فارسی، جایگاه این سیستم در نظرکاوی زبان فارسی روشن شود.

۲۴-۱-۱-۲۵-۱-۱-۱ مقایسه حجم پیکره

در زبان انگلیسی پیکره‌های مختلفی برای نظرکاوی وجود دارد. بیشترین کارهای صورت گرفته بر روی پیکره نقد فیلم [۱۱] است. حجم این پیکره‌ها تفاوت فراوانی دارد که در کارهای مختلف با توجه به نیاز به پیش‌پردازش‌های خاصی از پیکره انتخاب می‌شود و برای نظرکاوی مورد استفاده قرار می‌گیرد.

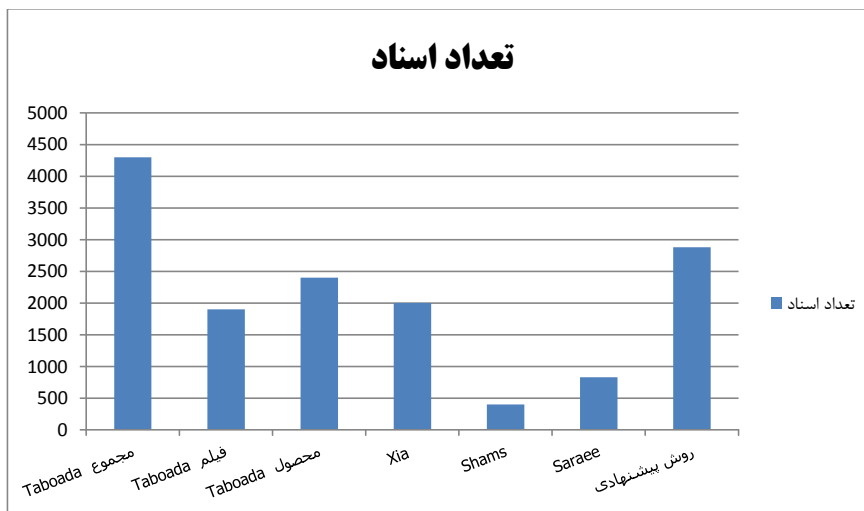
جدول ۴-۶: حجم پیکره در چند روش نظر کاوی جدول ۴-۶ حجم پیکره مورد استفاده در چند سیستم نظر کاوی را نشان می‌دهد. دو روش مربوط به زبان انگلیسی و دو روش مربوط به زبان فارسی است.

جدول ۴-۶: حجم پیکره در چند روش نظر کاوی

روش	تعداد اسناد نظری	دامنه نظرات	زبان کار شده
روش پیشنهادی	۲۸۸۳	گوشی موبایل	فارسی
<i>Saraee2013</i> [۲۷]	۸۲۹	گوشی موبایل	فارسی
<i>Shams2012</i> [۲۴]	۴۰۰	هتل موبایل و دوربین	فارسی
<i>Xia2011</i> [۱۸]	۲۰۰۰	DVD	انگلیسی
<i>Taboada2011</i> [۱۹]	۱۹۰۰ + ۲۴۰۰	۱۹۰۰ نقد فیلم و ۲۴۰۰ نظر (دوربین، چاپگر و کالسنکه کودک)	انگلیسی

جدول ۴-۶ نشان می‌دهند که نسبت نظرات استفاده شده در روش پیشنهادی به نسبت حجم کل نظراتی که در روش‌های متداول نظر کاوی استفاده شده است، از کمیت قابل قبولی برخوردار است. به علاوه در قیاس روش پیشنهادی با سایر کارهای زبان فارسی می‌توان مشاهده کرد که حجم پیکره استفاده شده از دیگر

روش‌های نظرکاوی زبان فارسی به وضوح بالاتر است. در نمودار فلان این مقایسه به صورت دیگری نمایش داده شده است.



شکل ۴-۲: نمودار مقایسه بین حجم پیکره چند روش نظرکاوی

انتخاب دسته مناسب ۱-۲۵-۱-۱-۱

برای انجام ارزیابی پیکره نظرات به دو بخش تست و آموزش دسته‌بندی شد. در بخش آموزش ۱۷۲۴ نظر عادی مثبت و ۴۳۳ نظر عادی منفی و در مجموع ۲۱۵۷ نظر عادی وجود دارد. در فصل گذشته بیان شد که برای ایجاد الگو ابتدا دسته‌های نظرات تشکیل می‌شود. این دسته‌ها همان جایگشت‌های مختلف برچسب‌های حسی هستند که در بخش آموزش مشاهده شده‌اند. جدول ۲-۷: دسته‌های ایجاد شده برای استخراج الگو بر اساس جایگشت برچسب‌های حسی جدول ۴-۷ در پیوست تعداد کل دسته‌ها را نشان می‌دهد. از این دسته‌ها تعدادی با تکرار زیاد و تعدادی با تکرار محدود در بخش آموزش مشاهده شده‌اند. جدول ۴-۷: تعداد تکرار دسته‌های مختلف جدول ۴-۷ خلاصه‌ای از دسته‌های ایجاد شده به همراه تعداد وقوع آن‌ها را بیان می‌کند.

جدول ۴-۷: تعداد تکرار دسته های مختلف

تعداد دفعات مشاهده دسته	تعداد دسته‌هایی که با این تعداد تکرار موجود هستند
236	1
147	1
120	1
76	1
45	1
36	1
35	1
34	1
32	1
28	1
24	1
23	1
22	1
21	1
20	2
19	2
18	1
17	1
14	1
13	2
12	4
10	2
9	6
8	3
7	10
6	7
5	11
4	16
3	29
2	94
1	482

Commented [m14]: اعداد فارسی شوند

به عنوان نمونه سطر اول جدول ۴-۷ بیان می‌کند که یک دسته وجود دارد که ۲۳۶ بار در اسناد دیده شده است. بنابراین این دسته معتبر است و می‌توان از آن با اطمینان به عنوان دسته ایجاد

کننده یک الگو استفاده کرد. سطر آخر جدول نیز بیان می‌کند که ۴۸۲ دسته وجود دارد که هر کدام فقط یک بار در پیکره اتفاق افتاده‌اند. همانطور که مشخص است استفاده از این دسته‌ها برای ساخت الگو با ریسک بسیار بالایی همراه هست زیرا احتمال زیادی وجود دارد که این دسته‌هایی که فقط یک بار تکرار شده‌اند به علت خطا در مرحله برچسب زنی حسی ایجاد شده باشند. بنابراین نیاز به تعریف یک آستانه وجود دارد که دسته‌ها با تعداد تکرار مشخص به عنوان دسته‌های صحیح شناخته شوند و سایر دسته‌هایی که تعداد تکرار آنها کمتر از میزان آستانه است به عنوان دسته‌های ناصحیح در نظر گرفته می‌شوند و برای ساخت الگو مورد استفاده قرار نمی‌گیرند.

تعیین آستانه انتخاب دسته‌ها: برای تعیین آستانه انتخاب دسته‌ها دقت نهایی کار ملاک قرار داده شد. البته دقت نهایی در بخش (۲۶-۱-۳-۴-) به صورت مفصل بحث خواهد شد. در اینجا فقط اشاره‌ای شده که با در نظر گرفتن چه آستانه‌ای از تعداد تکرار دسته‌ها دقت مناسب‌تری به دست آمده است. در جدول ۴-۸ بیان شده که به ازای تغییر آستانه انتخاب دقت نهایی نظر کاوی چه تغییری می‌کند. سطر اول این جدول بیان می‌کند که در صورتی که آستانه یک باشد، به این معنا که دسته‌هایی با تعداد تکرار یک به عنوان ایجاد کننده الگو در نظر گرفته شوند، ۲۶۹ الگو برای ۳۵۹ نظر در بخش ارزیابی شناسایی خواهد شد که تنها ۲۱۵ نظر به صورت صحیح تعیین حس می‌شوند. بعلاوه ۵۴ الگو شناسایی شده‌اند که نتوانستند حس سند را به درستی شناسایی کنند. لازم به ذکر است همانطور که در فصل قبل بیان شد. نظراتی که هیچ الگویی برای آنها شناسایی نشود، به الگوریتم SVM سپرده خواهند شد تا دسته‌بندی گردند.

جدول ۴-۸: تاثیر آستانه تکرار دسته‌ها بر دقت نهایی در شناسایی الگو در بخش تست (۳۵۹ نظر) و دسته‌بندی حسی نظرات

Commented [m15]: این جدول همه اش در یک صفحه باشد

مقدار آستانه انتخاب شده برای دسته‌ها در بخش آموزش	تعداد الگوهای شناسایی شده در بخش تست	تعداد الگوهای که منجر به تعیین گرایش صحیح شده‌اند	دقت نهایی (به درصد)
آستانه یک	۲۶۹	۲۱۵	۸۱,۲۳
آستانه دو	۲۲۰	۲۰۵	۹۲,۹۵
آستانه سه	۲۰۲	۲۰۱	۹۶,۶۶
آستانه چهار	۱۸۸	۱۸۷	۹۵,۵۱
آستانه پنج	۱۷۸	۱۷۸	۹۴,۹۹

Commented [m16]: این جدول باید مرتب شود فاصله های خیلی زیاد است

همانطور که در جدول ۴-۸ مشخص است بهترین میزان آستانه به دست آمده (آستانه=۳) است. در آستانه‌های کوچک‌تر به علت افزایش الگوهای ناکارآمد دقت کاهش پیدا نموده است. در آستانه‌های بزرگ‌تر از سه نیز الگوها دقیق‌تر می‌شوند و با دقت بالاتری عملی می‌کنند. اما از طرفی کاهش میزان فراخوانی الگوها موجب کاهش دقت نهایی است. بنابراین آستانه‌ای که در اینجا مورد استفاده قرار گرفته آستانه سه هست و دسته‌های کم تکرارتر از ۳ برای استخراج الگو استفاده

نخواهند شد. با توجه به این مطلب و **جدول ۷-۴: تعداد تکرار دسته‌های مختلف جدول ۷-۴** از مجموع دسته‌های موجود ۱۱۱ دسته به عنوان دسته‌های ایجاد کننده الگو مورد استفاده قرار گرفته‌اند.

دقت برچسب زنی مثبت و منفی روش پیشنهادی ۱-۱-۲۶-۱-۱-۲۶

در بخش **۱-۱-۲۵-۱-۱-۲۵-۴** بیان شد که ۱۱۱ دسته نظر از بخش آموزش به عنوان دسته‌های ایجاد کننده الگو معرفی شدند. با توجه به نقش‌های نحوی مجاز بین هر دو برچسب الگوهای نهایی نظری برای کشف حس به دست آمد. این الگوها از مجموع ۲۱۵۷ نظر موجود در بخش آموزش به دست آمد. در حالت کلی کارایی طبقه‌بندی نظرات توسط چهار شاخص ارزیابی می‌گردد. این شاخص‌ها عبارتند از Accuracy, Precision, Recall and F1-score هستند. یک راه ساده برای محاسبه این چهار شاخص بر اساس ماتریس پیچیدگی است که در جدول ۹-۴ آمده است [۷].

Commented [m17]: این چرا انگلیسی است؟ ترجمه شود

#	Predicted positives	Predicted negatives
Actual positive instances	Number of True Positive instances (TP)	Number of False Negative instances (FN)
Actual negative instances	Number of False Positive instances (FP)	Number of True Negative instances (TN)

جدول ۹-۴: ماتریس پیچیدگی [۷]

این چهار شاخص می‌توانند به شکل زیر تعریف گردند

- $Accuracy = \frac{TN + TP}{TN + TP + FP + FN}$
- $Precision = \frac{TP}{TP + FP}$
- $Recall = \frac{TP}{TP + FN}$
- $F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$

† confusion matrix

برای ارزیابی سیستم این الگوها برای تشخیص بار حسی نظرات بر روی پیکره تست مورد آزمایش قرار گرفت. **جدول ۱۰-۴: دقت روش پیشنهادی جدول ۳-۴** جزئیات آزمایش مورد نظر را نشان می‌دهند. در این جدول گزارشی از تعداد نظراتی که توسط الگو شناسایی و برچسب گذاری حسی شده‌اند، ارائه شده است. قسمتی از نظرات که در هیچ الگویی صدق نکنند به الگوریتم دسته بند ماشین پشتیبان بردار داده خواهد شد و مجموع نظرات که با استفاده از الگوها و ماشین بردار پشتیبان دسته‌بندی حسی شوند؛ عملکرد نهایی سیستم را تشکیل خواهند داد. در این جدول خروجی دسته‌بندی هر کدام از این مراحل به تفکیک بیان شده است.

جدول ۱۰-۴: دقت روش پیشنهادی

دقت نهایی				باقیمانده نظرات داده شده به SVM				شناسایی شده با الگو				تعداد	
precision	recall	نام صحیح	مجموع	precision	recall	نام صحیح	مجموع	precision	recall	نام صحیح	مجموع		
۹۴,۵۴	۹۸,۲۳	۵	۲۷۷	۶۸,۳۷	۹۵,۸۰	۵	۱۱۴	۹۹,۳۹	۱۰۰	۰	۱۶۳	۲۸۲	نظرات مثبت
۹۲,۴۲	۷۹,۲۲	۱۶	۶۱	۸۲,۱۴	۶۰,۵۳	۱۵	۲۳	۱۰۰	۹۷,۴۴	۱	۳۸	۷۷	نظرات منفی
۹۶,۶۶	۹۶,۳۶	۲۱	۳۳۸	۸۹,۳۲	۸۹,۹۲	۲۰	۱۳۷	۹۹,۵۱	۹۹,۵۰	۱	۲۰۱	۳۵۹	مجموع نظرات

با توجه به گزارشی که در **جدول ۴-۱۰: دقت روش پیشنهادی جدول ۴-۱** ارائه شده است. دقت نهایی روش پیشنهادی توسط دو معیار دیگر $Accuracy=94.15$ و $F-measure=96.51$ بدست می‌آید.

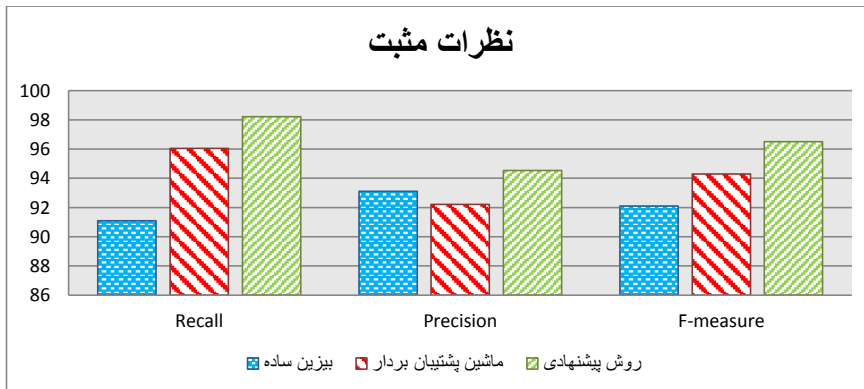
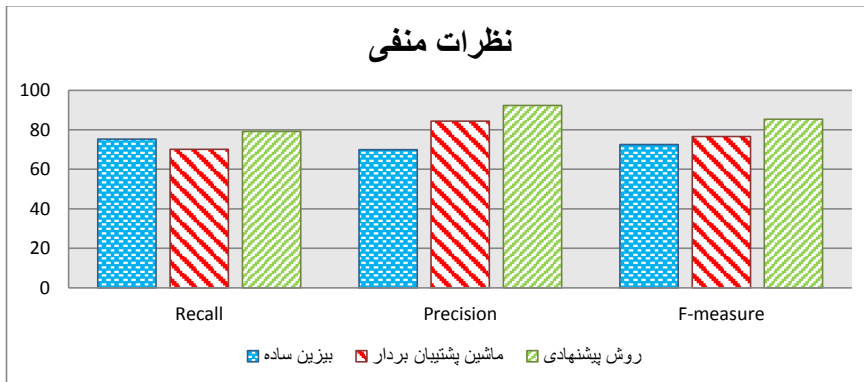
خط مبنا و مقایسه خروجی سیستم ۱-۲۷-۱-۲۸-۱-۱-۱

به طور معمول سیستم‌های نظرهاوی برای ارزیابی میزان بهبود روش پیشنهادی خود از یک الگوریتم یادگیری ماشین به عنوان خط مبنا استفاده می‌کنند و درصد بهبود نسبی که بر روی خط مبنا انجام داده‌اند را ملاک عملکرد سیستم خود می‌دانند. در سال‌های اخیر بسیاری از روش‌ها از الگوریتم ماشین بردار پشتیبان به عنوان خط مبنا استفاده کرده‌اند. در این جا نیز از دو الگوریتم NBA و SVM به عنوان خط مبنا استفاده شده است. **جدول ۴-۱۱: گزارش عملکرد الگوریتم‌های یادگیری ماشین بیزین و ماشین بردار پشتیبان بر روی پیکره سیستم پیشنهادی به عنوان خط مبنا جدول ۴-۱۱** نتیجه عملکرد این الگوریتم‌ها بر روی پیکره روش پیشنهادی را نشان می‌دهد.

جدول ۴-۱۱: گزارش عملکرد الگوریتم‌های یادگیری ماشین بیزین و ماشین بردار پشتیبان بر روی پیکره سیستم پیشنهادی به عنوان خط مبنا

الگوریتم NBA				الگوریتم SVM				
precision	recall	تأمین	میان	precision	recall	تأمین	میان	
۹۳,۱	۹۱,۱	۲۵	۲۵۷	۹۲,۲	۹۶,۰۵	۱۰	۲۷۲	نظرات مثبت
۶۹,۹	۷۵,۳	۱۹	۵۸	۸۴,۴	۷۰,۱	۲۳	۵۴	نظرات منفی
۸۸,۱	۸۷,۷	۴۴	۳۱۵	۹۰,۵	۹۰,۸	۳۳	۳۲۶	مجموع نظرات

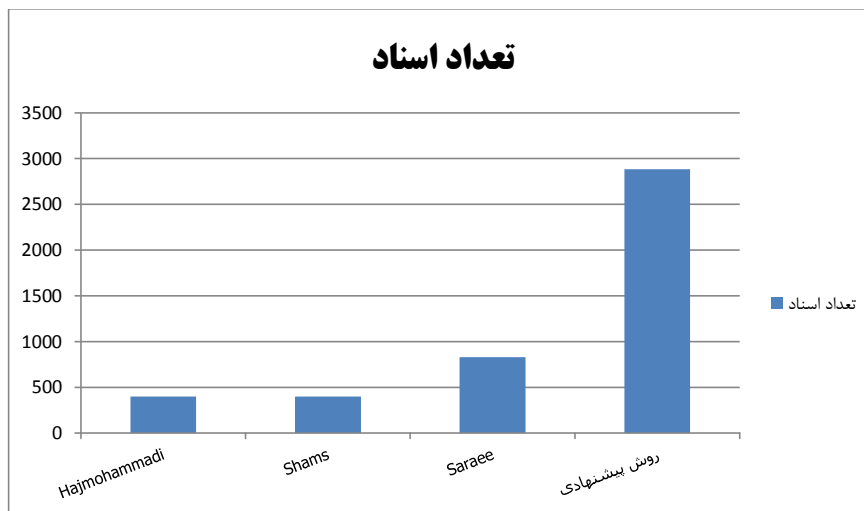
با توجه به جدول ۱۰-۴ و جدول ۱۱-۴ می توان مقایسه‌ای داشت میان روش پیشنهادی و الگوریتم‌هایی که در این جا به عنوان خط مبنا استفاده شده‌اند، داشت.



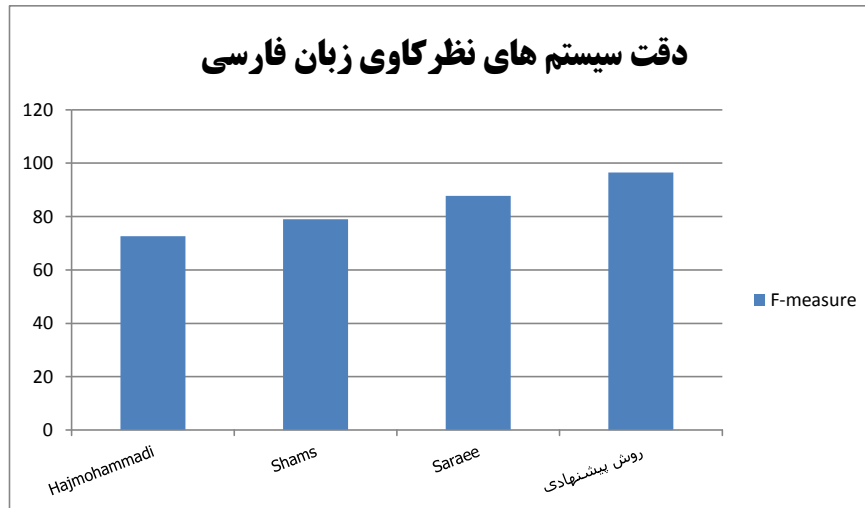
در روشی که سرایی و باقری در پیش گرفتند، ابتدا در [۲۶] یک ویژگی جدید به عنوان ورودی الگوریتم دسته‌بندی بیزین تعریف کردند. این ویژگی جدید تعریف شده، به همراه دو ویژگی دیگر مورد مقایسه قرار گرفت. نتیجه به دست آمده این بود که ویژگی تعریف شده دقتی حدود ۸۸ درصد را در نظر کاوی زبان فارسی به دست می‌دهد که از دو ویژگی دیگر بالاتر بود. در ادامه همین کار، این دو نویسنده در مقاله‌ی [۲۷] ویژگی معرفی شده خود را مورد ارزیابی دقیقتری قرار دادند.

(نمودار ۱: مقایسه حجم پیکره نظرات مورد استفاده با سایر کارهای انجام شده در زبان فارسی. نمودار ۲: درصد دقت گزارش شده روش‌ها استفاده از معیار *F-measure* نمودار ۳: درصد دقت گزارش شده روش‌ها استفاده از معیار *F-measure* به ترتیب حجم دقت روش استفاده شده در این پژوهش را با دقت گزارش شده در روش‌های دیگر زبان فارسی مقایسه می‌کند.)

Commented [m18]: همه باید شکل شوند و کلمه نمودار نداریم



نمودار ۱: مقایسه حجم پیکره نظرات مورد استفاده با سایر کارهای انجام شده در زبان فارسی

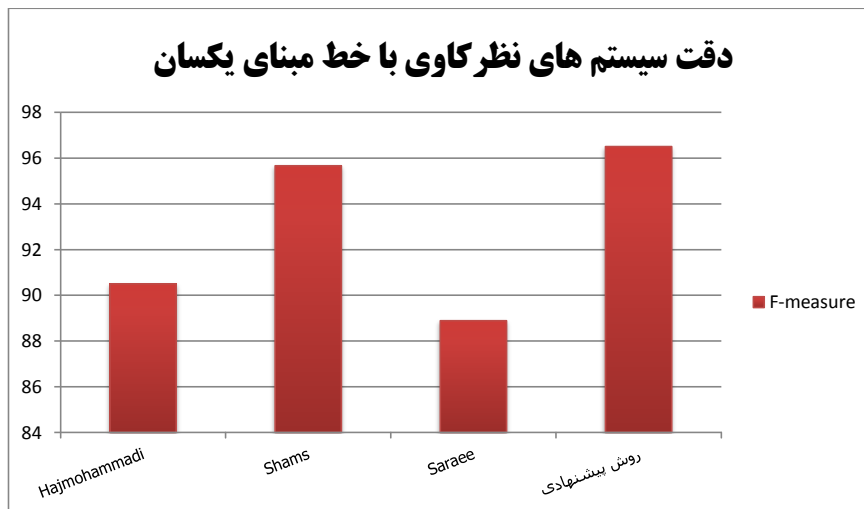


نمودار ۲: درصد دقت گزارش شده روش‌ها با استفاده از معیار *F-measure*

همانطور که در نمودار قابل ملاحظه است دقت روش پیشنهادی از سایر کارهای انجام شده بالاتر است. البته این دقت، دقتی است که توسط منبع منتشر شده به عنوان بالاترین دقت به دست آمده بیان شده است، اما باید توجه داشت که پیکره‌های مورد استفاده این روش‌ها نیز با یکدیگر متفاوت است. در حقیقت یک روش بر روی چند پیکره مختلف ممکن است جواب‌های کاملاً متفاوتی داشته باشد. بنابراین بهتر آن است که روش‌های ارائه شده این مقالات بر روی پیکره روش پیشنهادی مورد مقایسه قرار گیرد و با روش پیشنهادی بر روی پیکره مورد استفاده در روش‌های دیگر اجرا گردد. متأسفانه هیچ کدام از دو طرح فوق به علت عدم انتشار پیکره و یا کدی با دسترسی باز ممکن نیست. در ادامه برای اینکه تا حدودی بر این مشکل غلبه شود، ارزیابی جدیدی پیشنهاد می‌شود.

در این قسمت علاوه بر توجه صرف به دقت الگوریتم ارائه شده، تلاش می‌شود تا روش‌های مختلف با یک خط مبنا یکسان مورد ارزیابی قرار گیرند. در روش [۲۶] از بی‌زین استفاده شده است. در روش‌های [۴] و

[۵] نیز SVM با ویژگی *unigram* به عنوان خط مبنا مورد استفاده قرار گرفته است. در روش پیشنهادی نیز هم SVM و هم *NBA* به عنوان خط مبنا استفاده شده‌اند. بنابراین در اینجا تلاش می‌شود تا با یکسان کردن خط مبنا میان روش‌های فوق مقایسه دقیق‌تری انجام گیرد.



نمودار ۳: مقایسه معیار *F-measure* پس از یکسان سازی خط مبنا

همان‌طور که در (نمودار ۳) مقایسه معیار *F-measure* پس از یکسان سازی خط مبنا نمودار ۳ مقایسه معیار *F-measure* پس از یکسان سازی خط مبنا (ملاحظه می‌شود) پس از این که تناسبی میان خط مبناهای روش‌های مورد بررسی با روش پیشنهادی ایجاد شد، دقت‌های بیان شده کاملاً متفاوت گشته‌اند. بر طبق این نمودار روش که در [۲۴] ارائه شده بود بسیار نزدیکتر به روش پیشنهادی بوده است. دو روش دیگر همچنان اختلاف قابل توجهی با روش پیشنهادی دارند.

۲۹-۱-۱-۳-۱-۱ نظرات مقایسه‌ای

اکثر کارهای نظر کاوی توجه خود را بر روی نظرات عادی متمرکز ساخته‌اند و تعداد انگشت شمار از آنها به نظرات مقایسه‌ای توجه کرده‌اند. در فصل قبل مطرح شد که روش پیشنهادی قابلیت بکارگیری برای هر دو

مدل نظر یعنی نظرات عادی و نظرات مقایسه‌ای را دارا می‌باشد. در این بخش گزارشی از چگونگی عملکرد روش پیشنهادی بر روی نظرات مقایسه‌ای ارائه می‌گردد.

برای نظرات ساده یازده برچسب حسی در بخش قبل معرفی شد. با توجه به ماهیت متفاوت نظرات مقایسه‌ای، برچسب‌های زیر مختص نظرات مقایسه‌ای تعریف شده‌اند.

جدول ۴-۱۲: برچسب‌های حسی نظرات مقایسه‌ای

تعداد	اختصار	نام برچسب	
این برچسب نشان دهنده صفات مقایسه‌ای نظیر زیباتر، ضعیفتر و ... است.	COM	مقایسه‌ای	۱
این برچسب نشان دهنده صفات عالی نظیر بزرگترین، گرانترین و ... است.	SUP	عالی	۲
موجودیتی که نظر بر آن واقع است	ASPM	موجودیت اصلی	۳
موجودیت‌هایی که با موجودیت اصلی قیاس شده است.	ASPO	موجودیت فرعی	۴

نظرات مقایسه‌ای پیرامون موجودیت اصلی مد نظر می‌توانند به پنج کلاس بهترین، بدترین، بهتر، بدتر و تساوی تقسیم شوند. در روش پیشنهادی کلاس «بدترین» به علت تعداد کم نظرات انتساب داده شده به این کلاس حذف گردید. نظرات موجود در کلاس «بدتر» با تعویض محل موجودیت اصلی و موجودیت فرعی با کلاس بهتر ادغام گردید. بنابراین در این پژوهش کلاس‌های باقیمانده برای نظرات مقایسه‌ای شامل کلاس‌های زیر می‌شوند:

کلاس «بهتر»: موجودیت اصلی از موجودیت فرعی بهتر دانسته شده است.

کلاس «بهترین»: موجودیت اصلی از تمام موجودیت‌های مشابه بهتر دانسته شده است.

کلاس «تساوی»: موجودیت‌ها برابر دانسته شده‌اند.

۴- ارزیابی سیستم

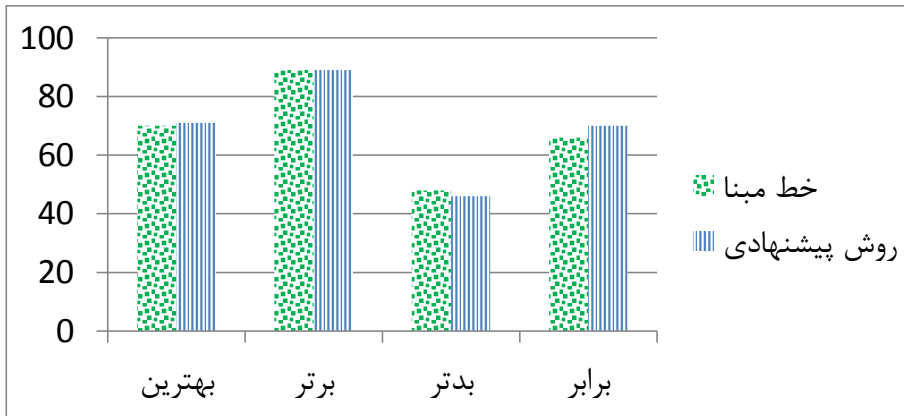
مراحل تشخیص الگو مطابق آنچه برای نظرات عادی اعمال شد، در این بخش نیز اعمال گردید. در این بخش با توجه به آستانه انتخاب سه دسته‌های زیر برای نظرات مقایسه‌ای کشف گردید و سایر دسته‌ها به علت فرکانس پایین رخداد از میان دسته‌های منتخب حذف گشته‌اند. دسته‌های منتخب نهایی که منجر به تولید الگو خواهند شد در **جدول ۴-۱۳: دسته‌های نماینده الگوهای نظرات مقایسه‌ای جدول ۴-۱۳. دسته‌های نماینده الگوهای نظرات مقایسه‌ای** آمده‌اند.

جدول ۴-۱۳: دسته‌های نماینده الگوهای نظرات مقایسه‌ای

برچسب اول	برچسب دوم	برچسب سوم	برچسب چهارم	برچسب پنجم	برچسب ششم	برچسب هفتم
ent	comp	verb				
ent	asp	comp	verb			
key	ent	senti	verb			
key	ent	comp	verb			
ent	asp	super	verb			
ent	super	verb				
ent	key	ent	comp	verb		
asp	ent	comp	verb			
asp	ent	key	asp	ent	comp	verb
asp	ent	key	ent	comp	verb	
super	ent	verb				
super	verb					

Commented [m19]: این جدول باید راست چین شود. یعنی این ستون به سمت راست برود و بقیه هم همینطور

دقت نهایی سیستم، پس از اعمال الگوها برای شناسایی کلاس‌های نظرات مقایسه‌ای در شکل ۴-۴ گزارش شده است.



شکل ۴-۴: مقایسه عملکرد روش پیشنهادی با الگوریتم SVM به عنوان خط مبنا ارزیابی

همان‌طور که مشاهده می‌شود میزان موفقیت در نظرات مقایسه‌ای نسبت به نظرات عادی کمتر می‌باشد زیرا دسته‌های منتخب یافته شده برای نظرات مقایسه‌ای کمتر از نظرات عادی است. برخی از علت‌های ایجاد دسته‌های کمتر برای نظرات مقایسه‌ای نسبت به نظرات عادی در زیر لیست گشته‌اند.

- شمار کمتر این نظرات نسبت به نظرات عادی: نظرات مقایسه‌ای حجم کمتری از نظرات را به خود اختصاص می‌دهند و اکثر نظرات بیان شده پیرامون یک موجودیت نظرات مستقیم هستند.
- استفاده بیشتر از کنایه‌ها: در بسیاری از موارد نظرات مقایسه‌ای توسط کاربرانی بیان می‌شود که نسبت به یک موجودیت احساس خوشایند و نسبت به دیگری احساس متضاد دارند. در بسیاری از موارد این تناقض به صورت بیان کنایه به موجودیت منفی بیان می‌شود.
- موجودیت اضافه: در بعضی موارد، مقایسه صورت گرفته بین بیش از دو موجودیت مشابه انجام می‌گیرد و در چنین مواردی کشف موجودیت برتر بسیار مشکل خواهد شد.
- مشکل در شناسایی ضمیر: در بخش نظرات عادی یک موجودیت در اختیار بود و در صورت شناسایی ضمیر به راحتی امکان ایجاد ارتباط آن با موجودیت مد نظر وجود داشت. حال آن که برای

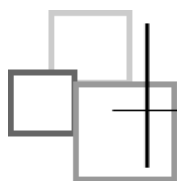
نظرات مقایسه‌ای بیش از یک موجودیت وجود دارد و ضمیر می‌تواند به هر کدام از موجودیتها اشاره داشته باشد. تشخیص صحیح ضمیر به خصوص در زبان فارسی یکی از چالش‌های باز در پردازش زبان طبیعی است.

خلاصه فصل

در این فصل به جزئیات پیاده سازی روش پیشنهادی و پس از آن به ارزیابی روش پرداخته شد. در بخش پیاده سازی، ابتدا از چگونگی جمع‌آوری مجموعه داده برای این تحقیق صحبت شد. در ادامه ابزارهای مورد استفاده در این بخش بیان گردید. پس از برچسب‌های حسی مورد استفاده در این تحقیق بیان شد. همچنین تعدادی از الگوهای به دست آمده آورده شد.

در بخش ارزیابی، ابتدا ویژگی‌های کمی از پیکره مورد استفاده بیان گردید. پس از آن پیکره مورد استفاده به لحاظ کمی با چند پیکره دیگر مورد مقایسه قرار گرفت. در ادامه دقت به دست آمده از روش پیشنهادی ذکر گردید. به عنوان خط مبنای ارزیابی الگوریتمهای *SVM* و بیزین ساده بر روی پیکره پیاده سازی گشت و دقت آن با روش پیشنهادی مقایسه شد. همچنین بر اساس این خط مبنای روش پیشنهادی با سایر کارهای صورت گرفته در زبان فارسی مقایسه گردید. همان‌طور که در این قسمت نشان داده شد، روش پیشنهادی دقت مناسبی را نسبت به سایر روش‌های موجود از خود نشان داده است.

۵- نتیجه‌گیری و ارائه پیشنهادها.....



۵- نتیجه‌گیری و ارائه پیشنهادها

مقدمه

در این فصل تلاش می‌شود خلاصه‌ای از انجام کار و نتایج به دست آمده بیان شود. در انتها نیز زمینه‌های تحقیقاتی پیش رو بیان خواهد شد.

نتیجه‌گیری

در اکثر پژوهش‌های صورت گرفته در حوزه نظرکاوی تلاش بر این است که دقت نهایی دسته‌بندی نظرات (به کلاس‌های مثبت و منفی یا طبقات بیشتر) افزایش یابد. برای رسیدن به این مقصود از الگوریتم‌های مختلفی استفاده می‌شود. این الگوریتم‌ها را می‌توان به صورت عمده به دو دسته الگوریتم‌های باناظر و بدون ناظر تقسیم بندی کرد. در روش‌های بدون ناظر ویژگی‌های خاصی از متن به همراه لغت‌نامه برای کشف احساس سند نظری مورد استفاده قرار می‌گیرند. یکی از ویژگی‌های مهمی که این گونه کارها از آن استفاده می‌کنند نقش نحوی لغات است.

در این مقاله نیز تاکید خود را بر ویژگی نقش نحوی لغات در ایجاد الگوی حسی قرار دادیم. در اینجا تلاش شد تا با در نظر گرفتن تعدادی خاص از نقش‌های نحوی و افزودن نقش‌های جدید حسی و ترکیب آن با نقش‌های نحوی به الگوهای با کیفیت بالاتری دست یابیم که از اعوجاج کمتری برخوردار باشد و پایایی آن بالاتر باشد. الگوهایی که در این تحقیق ساخته شده‌اند الگوهایی در واحد جمله هستند برخلاف الگوهایی که پیش از این از نقش چند کلمه متوالی (حداکثر سه کلمه) به عنوان الگو استفاده می‌شد. نکته‌ی دیگری که در این کار حائز اهمیت است این است که کارهای صورت گرفته در زبان فارسی انگشت‌شمار است، بنابراین این تحقیق برای اولین بار به کشف الگوهای حسی برای زبان فارسی پرداخته است.

الگوهایی که در این مقاله ایجاد شده‌اند، می‌توانند به عنوان یک ویژگی برای کشف نظر مورد استفاده سایر پژوهش‌ها قرار گیرند. این الگوها با روشی خودکار ایجاد شده‌اند بنابراین برای سایر دامنه‌ها نیز می‌توان این روش را به کار برد و الگوهای متناسب با دامنه مورد بررسی ایجاد نمود. علاوه بر این از این الگوها می‌توان برای بسط نیمه خودکار مجموعه داده نیز بهره برد و با آن بخشی از پیکره نظر کاوی را به صورت نیمه خودکار ایجاد کرد.

در روشی که در اینجا به کشف احساس اسناد پرداخته شد به هر دو دسته نظر (نظرات عادی و نظرات مقایسه‌ای) توجه شد. در اکثر روش‌ها فقط به نظرات عادی پرداخته شده و از نظرات مقایسه‌ای صرف نظر شده است. در شمار اندکی از کارهای دیگر نیز روشی خاص برای نظرات مقایسه‌ای ارائه شده است. در روشی که در اینجا پیاده‌سازی و ارزیابی گردید یک ایده با شمول عام مطرح گردید که با تغییرات بسیار جزئی برای هر دو دسته نظر قابل اعمال است.

به صورت خلاصه می‌توان خصوصیات پژوهش انجام شده را به صورت زیر بیان کرد.

- تعریف سیستم نظر کاوی بدون ناظر
- ایجاد الگوهای نظری در واحد جمله نه بر مبنای چند کلمه متوالی
- ایجاد الگوی نظری برای اولین بار در زبان فارسی
- توجه به دو دسته نظر عادی و مقایسه‌ای به صورت همزمان
- آماده سازی پیکره نظرات فارسی محصولات تجاری در واحد جمله

نتایج به دست آمده که در فصل قبل نیز به صورت مفصل بیان شد، حاکی از آن است که این سیستم از دقت مناسبی نسبت به سایر سیستم‌های موجود برای زبان فارسی برخوردار است. همچنین به لحاظ حجم نظرات مورد استفاده و وجود دو دسته نظر مقایسه‌ای و عادی، پیکره مورد استفاده به وضوح بر سایر کارهای زبان فارسی برتری دارد. در اکثر کارهای صورت گرفته در این

۵- نتیجه‌گیری و ارائه پیشنهادها

حوزه نیز ارزیابی نهایی خود را مقایسه دقت نهایی سیستم با یک الگوریتم دسته‌بندی رایج را ملاک قرار داده‌اند. الگوریتم‌های دسته‌بندی ماشین پشتیبان بردار و بی‌زین ساده بیشتر مورد استفاده قرار گرفته شده است. در این تحقیق نیز هر دو الگوریتم بر روی پیکره مورد استفاده مورد ارزیابی قرار گرفته‌اند. نتیجه به دست آمده نشان می‌دهد روش ارائه شده ۵٪ نسبت به الگوریتم ماشین پشتیبان بردار و ۸٪ نسبت به الگوریتم بی‌زین ساده بهبود داشته است.

Commented [m20]: تحلیل ثان را در مورد نظرات مقایسه ای هم

بگویند

کارهای آتی

در این پژوهش تلاش شد تا گام کوچکی در جهت بهبود نتایج حاصل در تحقیقات نظرکاوی برداشته شود. همچنین تلاش شد نوآوری‌های طرح پیشنهادی بر روی زبان فارسی اعمال شود تا به غنای بیشتر پردازش‌های نظرکاوی در این زبان کمک نماید. چالش اساسی که در انجام این امر وجود دارد، عدم وجود پیکره مناسب در زبان فارسی است. در حالی که در زبان انگلیسی چندین پیکره مختلف (نقد فیلم، نقد محصولات، نظرات ایجاد شده در توییتر و ...) نظرکاوی در حوزه‌های متفاوت وجود دارد. برخی از مشکلاتی که نبود یک پیکره استاندارد شده و واحد ایجاد می‌کند در زیر لیست شده است.

- زمان بر بودن مرحله جمع آوری داده اولیه
- اعمال سلیقه در داده‌ها منطبق بر روش پیشنهادی
- غیرقابل استفاده بودن داده‌های جمع آوری شده
- کاهش میزان اطمینان به خروجی‌های گزارش شده نهایی
- عدم امکان مقایسه روشن بین روش‌های مختلف پیاده‌سازی شده

۵- نتیجه‌گیری و ارائه پیشنهادها.....

با توجه به چالش‌های فوق به نظر می‌رسد مهمترین مسئله حل نشده‌ای که در نظر کاوی وجود دارد و نیاز به پرداختن به آن احساس می‌شود، حل مسئله پیکره استاندارد نظر کاوی برای زبان فارسی باشد.

در طرح پیشنهادی این پژوهش الگوهای جملات نظری برای زبان فارسی در حوزه محصولات تجاری استخراج گردید. از آنجا که در دامنه‌های مختلف طرز بیان احساسات (استفاده از کنایه‌ها و جملات وابسته به قراین) متفاوت است، این نیاز احساس می‌شود تا در دامنه‌های دیگر این روش به کار گرفته شود. استفاده از روش پیشنهادی در دامنه‌های دیگر ممکن است به کشف الگوهای جدیدی بیانجامد. همچنین ممکن است تعدادی الگوی مشترک با الگوهای کشف شده در این جا به دست آید. به این ترتیب الگوهای مختلف وابسته به دامنه و خارج از دامنه قابل دسته‌بندی خواهند بود و می‌توان از آن‌ها به عنوان ویژگی‌های کشف نظر در سایر روش‌های نظر کاوی و در حوزه‌های مختلف استفاده کرد.

۶-مراجِع

- [۱] B. Liu, "Sentiment analysis and opinion mining," Synthesis Lectures on Human Language Technologies, vol. 5, pp. 1-167, 2012.
- [۲] B. Liu, "Sentiment analysis and subjectivity," Handbook of natural language processing, vol. 2, p. 568, 2010.
- [۳] R. Feldman, "Techniques and applications for sentiment analysis," Communications of the ACM, vol. 56, pp. 82-89, 2013.
- [۴] V. Y. Karkare and S. R. Gupta, "A Survey on Product Evaluation using Opinion Mining," International Journal Of Computer Science And Applications, vol. 6, 2013.
- [۵] P. Sharma and K. Srivastava, "OPINIONS ON OPINION MINING: A SURVEY," International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE), vol. 2, pp. pp: 684-689, 2013.
- [۶] S. Bethard, H. Yu, A. Thornton, V. Hatzivassiloglou, and D. Jurafsky, "Automatic extraction of opinion propositions and their holders," 2004 AAAI Spring Symposium on Exploring Attitude and Affect in Text, p. 2224, 2004.
- [۷] S. Padmaja and S. S. Fatima, "Opinion Mining and Sentiment Analysis-An Assessment of Peoples' Belief: A Survey," International Journal, 2013.
- [۸] X. Ding, B. Liu, and P. S. Yu, "A holistic lexicon-based approach to opinion mining," in Proceedings of the 2008 International Conference on Web Search and Data Mining, 2008, pp. 231-240.
- [۹] R. Moraes, J. F. Valiati, and W. P. Gavião Neto, "Document-level sentiment classification: An empirical comparison between SVM and ANN," Expert Systems with Applications, vol. 40, pp. 621-633, 2013.
- [۱۰] P. D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," in Proceedings of the 40th annual meeting on association for computational linguistics, 2002, pp. 417-424.
- [۱۱] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up :?sentiment classification using machine learning techniques," in Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10, 2002, pp. 79-86.
- [۱۲] K. Duh, A. Fujino, and M. Nagata, "Is machine translation ripe for cross-lingual sentiment classification?," in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2, 2011, pp. 429-433.

Commented [m21]: در مراجِع دقت شود همه به شکل استاندارد باشد. صفحه شروع و خاتمه در مورد مقالات (ارجاع ۲) و حروف بزرگ و کوچک در متن (ارجاع ۵) نوشتن اول اسم فامیل مولف و بعد اول اسم کوچک - کامل بودن (ارجاع ۷) دقت در محتوا (مرجع ۲۳) و کلا برای لاتین ها عدد شماره ارجاع لاتین و برای فارسی ها، فارسی باشد. header, footer, broder هم طبق استاندارد گروه باشد برای کل تز

- [۱۳] A. Abbasi, H. Chen, and A. Salem, "Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums," *ACM Transactions on Information Systems (TOIS)*, vol. 26, p. 12, 2008.
- [۱۴] E. Kouloumpis, T. Wilson, and J. Moore, "Twitter sentiment analysis: The Good the Bad and the OMG!," in *ICWSM*, 2011.
- [۱۵] L. Dey and S. M. Haque, "Opinion mining from noisy text data," *International Journal on Document Analysis and Recognition (IJ DAR)*, vol. 12, pp. 205-226, 2009.
- [۱۶] G. Qiu, B. Liu, J. Bu, and C. Chen, "Expanding Domain Sentiment Lexicon through Double Propagation," in *IJCAI*, 2009, pp. 1199-1204.
- [۱۷] G. Qiu, B. Liu, J. Bu, and C. Chen, "Opinion word expansion and target extraction through double propagation," *Computational linguistics*, vol. 37, pp. 9-27, 2011.
- [۱۸] R. Xia and C. Zong, "A POS-based Ensemble Model for Cross-domain Sentiment Classification," in *IJCNLP*, 2011, pp. 614-622.
- [۱۹] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," *Computational linguistics*, vol. 37, pp. 267-307, 2011.
- [۲۰] V. K. Singh, M. Mukherjee, and G. K. Mehta, "Sentiment and mood analysis of weblogs using POS tagging based approach," in *Contemporary Computing*, ed: Springer, 2011, pp. 313-324.
- [۲۱] A. K. Jain and Y. Pandey, "Analysis and Implementation of Sentiment Classification Using Lexical POS Markers," *International Journal*, vol. 2, 2013.
- [۲۲] N. D. Patel and C. Chand, "Selecting Best Features Using Combined Approach in POS Tagging for Sentiment Analysis," 2014.
- [۲۳] ت. خطیبی، م. سپهری، and ب. حمیدپور، "بررسی تاثیر روش مربع کای برای انتخاب مشخصه‌ها در نظر کاوی متن فارسی،" *presented at the* "دومین همایش ملی مهندسی برق، کامپیوتر و فناوری اطلاعات، همدان، مرکز آموزش و فرهنگی سما همدان، ۱۳۸۷.
- [۲۴] M. Shams, A. Shakery, and H. Faili, "A non-parametric LDA-based induction method for sentiment analysis," in *Artificial Intelligence and Signal Processing (AISP)*, 2012 16th CSI International Symposium on, 2012, pp. 216-221.
- [۲۵] M. S. Hajmohammadi and R. Ibrahim, "A SVM-based method for sentiment analysis in Persian language," in *2012 International Conference on Graphic and Image Processing*, 2013, pp. 876838-876838-5.
- [۲۶] A. Bagheri, M. Saraee, and F. de Jong, "Sentiment classification in Persian: Introducing a mutual information-based method for feature selection," in *Electrical Engineering (ICEE)*, 2013 21st Iranian Conference on, 2013, pp. 1-6.

- [۲۷] M. Saraee and A. Bagheri, "Feature selection methods in Persian sentiment analysis," in *Natural Language Processing and Information Systems*, ed Springer, 2013, pp. 303-308.
- [۲۸] N. Jindal and B. Liu, "Mining comparative sentences and relations," in *AAAI*, 2006, pp. 1331-1336.
- [۲۹] N. Jindal and B. Liu, "Identifying comparative sentences in text documents," in *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, 2006, pp. 244-251.
- [۳۰] S. Yang and Y. Ko, "Extracting comparative entities and predicates from texts using comparative type classification," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, 2011, pp. 1636-1644.
- [۳۱] S. Li, C.-Y. Lin, Y.-I. Song, and Z. Li, "Comparable entity mining from comparative questions," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 25, pp. 1498-1509, 2013.
- [۳۲] m. mohseni, "Persain pos tagger," *graduate, Elm Va Sanat university, Tehran*, 1391.
- [۳۳] گ. پ. دادگان, "پروژه دادگان وابستگی زبان فارسی," ۱۳۹۱ تابستان ۹۱.
- [۳۴] A. Buche, D. Chandak, and A. Zadgaonkar, "Opinion Mining and Analysis: A survey," *arXiv preprint arXiv:1307.3336*, 2013.
- [۳۵] R. Feldman, "Techniques and applications for sentiment analysis," *Communications of the ACM*, vol. 56, pp. 82-89, 2013.

۷- پیوست - الف

Commented [m22]: پیوست ها فقط عنوانش اینجا بیاید و متن در
cd بیاید

جدول ۷-۱: برجسبهای حسی استفاده شده برای کشف نظر

اختصار	برچسب	اختصار	برچسب	اختصار	برچسب
asp	شبکه بی‌سیم Wi - Fi	ent	گوشی موبایل سونی اکسپریا ال	vs	گمراه خواهد کرد
asp	ضد آب و گردوغبار بودن	ent	گوشی موبایل سونی اکسپریا پی	vs	مجنوب خودش کرده
asp	سنسورها و سیستم‌های تعبیه‌شده	ent	گوشی موبایل سونی اکسپریا وی	vs	خط و خش می‌افتد
asp	ضد آب و گرد و غبار بودن	ent	Samsung Galaxy Note II	vs	راضی نگه داشته
asp	ضد آب و گرد و غبار بودن	ent	گوشی موبایل ال جی جی ۲	vs	پیشنهاد می‌کنم
asp	کار با برنامه‌های مختلف	ent	Samsung Galaxy Grand	vs	خیره کننده است
asp	اجرای بازی‌های سنگین	ent	سامسونگ گالاکسی نوت ۲	vs	احساس نمی‌شود
asp	چراغ Notification	ent	Samsung Galaxy S III	vs	پایین می‌آورد
asp	تعداد برنامه‌هایی	ent	سونی اکسپریا تی ایکس	vs	تحسین می‌کنم
asp	LED هشدار دهنده	ent	سامسونگ گالاکسی اس	vs	تعجب کرده بود
asp	برنامه‌های سنگین	ent	سامسونگ گلکسی اس ۳	vs	پشیمان بشوید
asp	سرعت کار کردن با	ent	اکسپریا ال	vs	کثیف می‌شود
asp	باز کردن درپوش	ent	اکسپریا پی	vs	اذیت می‌شوی
asp	درپوش‌های محافظ	ent	اکسپریا زد	vs	درست می‌شود
asp	از بقیه‌ی جهات	ent	اکسپریا وی	vs	ضعیف می‌شود
asp	پورت‌های گوشی	ent	ال جی جی ۲	vs	معرکه می‌شد
asp	حافظه داخلی‌اش	ent	گالاکسی اس	vs	لذت می‌برد
asp	گواهی‌نامه ۵۷/P	ent	گلکسی اس ۳	vs	هنگ می‌کند
asp	صفحه‌ی نمایش	ent	گلکسی اس ۴	vs	آسوده گشت
asp	خروجی اسپیکر	ent	گلکسی گرند	vs	بهتر کرده
asp	سطح این شیشه	ent	هوآوی اسند	vs	حرف ندارد
asp	فیلم‌برداری	ent	Galaxy S4	vs	دوست دارم
asp	رابط کاربری	ent	Samsung S	vs	مرده بودن
asp	زاویه‌ی دید	ent	Xperia SP	vs	به صرفست
asp	صفحات نمایش	ent	Xperia TX	vs	جلب کرد
asp	صفحه نمایش	ent	Xperia ZR	senti	مقرون به صرفه‌ترین
asp	صفحه نمایش	ent	این محصول	senti	درخشندگی بی‌نظیری
asp	صفحه نمایشی	ent	شرکت سونی	senti	یک لایه سفید رنگ
asp	قاب شیشه‌ای	ent	موتو ایکس	senti	ماندن اثر انگشت
asp	کیفیت نمایش	ent	Galaxy S	senti	بالارفتن حرارت
asp	از همه نظر	ent	iPhone 5	senti	غیر قابل تعویض
asp	بگیرد دستش	ent	Xperia L	senti	جذب اثر انگشت
asp	زاویه دیدش	ent	Xperia P	senti	سنگین‌تر بودن
asp	زد آب بودن	ent	Xperia V	senti	ضد بازتاب نور
asp	زوایای دید	ent	xperia z	senti	غیرقابل تعویض
asp	سخت افزار	ent	اچ تی سی	senti	فوق العاده‌ای
asp	سخت‌افزاری	ent	این گوشی	senti	فوق العاده ای
asp	شارژ باتری	ent	سامسونگ	senti	۱۳ مگاپیکسلی
asp	صفحه کلیدش	ent	فیلثا	senti	استفاده مفید
asp	صفحه نمایش	ent	گوشیتان	senti	شاهکار واقعی
asp	صفحه‌نمایش	ent	گوشی‌های	senti	منحصر به فرد

انتخاب عالی	<i>senti</i>	موتورولا	<i>ent</i>	فناوری‌های	<i>asp</i>
فوق‌العاده	<i>senti</i>	Apple 5	<i>ent</i>	فیلمبرداری	<i>asp</i>
کیفیت نهایی	<i>senti</i>	HTC One	<i>ent</i>	کیفیت ساخت	<i>asp</i>
۳۲۰ Adreno	<i>senti</i>	Note II	<i>ent</i>	مقابل ضربه	<i>asp</i>
احساس خوبی	<i>senti</i>	Optimus	<i>ent</i>	نرم‌افزاری	<i>asp</i>
آزار دهنده	<i>senti</i>	Sony TX	<i>ent</i>	۴۴۰ پیکسل	<i>asp</i>
بهینه‌سازی	<i>senti</i>	Sony ZR	<i>ent</i>	امکاناتش	<i>asp</i>
شگفت‌انگیز	<i>senti</i>	xperia z	<i>ent</i>	این گوشی	<i>asp</i>
فوق‌العاده	<i>senti</i>	اپتیموس	<i>ent</i>	بدنه‌های	<i>asp</i>
فوق‌العاده	<i>senti</i>	اسند جی	<i>ent</i>	حافظه RAM	<i>asp</i>
قابل تعویض	<i>senti</i>	این یکی	<i>ent</i>	زاویه دید	<i>asp</i>
گریبان گیر	<i>senti</i>	آیفون ۵	<i>ent</i>	زمان شارژ	<i>asp</i>
ویژگی مثبت	<i>senti</i>	دستگاه	<i>ent</i>	سخت افزار	<i>asp</i>
یک حس خاصی	<i>senti</i>	موبایل	<i>ent</i>	کار کردن	<i>asp</i>
۴ هسته‌ای	<i>senti</i>	یک گوشی	<i>ent</i>	نرم افزار	<i>asp</i>
اثر انگشت	<i>senti</i>	Ascend	<i>ent</i>	نمایشگرش	<i>asp</i>
اشکالاتی	<i>senti</i>	Huawei	<i>ent</i>	نور آفتاب	<i>asp</i>
انتظارات	<i>senti</i>	Moto X	<i>ent</i>	استفاده	<i>asp</i>
بالایشان	<i>senti</i>	note 2	<i>ent</i>	امکانات	<i>asp</i>
بدون مشکل	<i>senti</i>	Sony L	<i>ent</i>	این گوشی	<i>asp</i>
پرقدرتی	<i>senti</i>	Sony V	<i>ent</i>	پایداری	<i>asp</i>
پلاستیکی	<i>senti</i>	اس نری	<i>ent</i>	پردازنده	<i>asp</i>
سایه سفید	<i>senti</i>	اس فور	<i>ent</i>	دوربینش	<i>asp</i>
قدرتمندی	<i>senti</i>	گالکسی	<i>ent</i>	سرعت عمل	<i>asp</i>
کم رنگ‌تر	<i>senti</i>	LG G2	<i>ent</i>	سی پی یا	<i>asp</i>
معرکه‌ای	<i>senti</i>	not 2	<i>ent</i>	صفحه‌اش	<i>asp</i>
نقاط ضعفی	<i>senti</i>	note2	<i>ent</i>	کار کردن	<i>asp</i>
واقعی‌تر	<i>senti</i>	S III	<i>ent</i>	محصولات	<i>asp</i>
با کیفیت	<i>senti</i>	اس پی	<i>ent</i>	نمایشگر	<i>asp</i>
بهره‌رسانی	<i>senti</i>	اس سه	<i>ent</i>	همه چیزش	<i>asp</i>
تمام شدن	<i>senti</i>	اس ۴	<i>ent</i>	اسپیکر	<i>asp</i>
جذابیتش	<i>senti</i>	خریدن	<i>ent</i>	اسپیکرش	<i>asp</i>
خوش ساخت	<i>senti</i>	خودش	<i>ent</i>	انتخاب	<i>asp</i>
شیشه‌ای	<i>senti</i>	سونی	<i>ent</i>	اندازه	<i>asp</i>
ضعف‌های	<i>senti</i>	کالا	<i>ent</i>	باتریش	<i>asp</i>
ضعیف‌تر	<i>senti</i>	گالکسی	<i>ent</i>	باطریش	<i>asp</i>
عادلانه	<i>senti</i>	گوشی	<i>ent</i>	بلندگویی	<i>asp</i>
عالی‌ای	<i>senti</i>	نوت ۲	<i>ent</i>	تصاویر	<i>asp</i>
عمق رنگ	<i>senti</i>	not2	<i>ent</i>	تصویرش	<i>asp</i>
عیب‌های	<i>senti</i>	اس ۴	<i>ent</i>	تو دستت	<i>asp</i>
قدرتمند	<i>senti</i>	اس ۳	<i>ent</i>	توانایی	<i>asp</i>
کاربردی	<i>senti</i>	اس ۴	<i>ent</i>	خیس شدن	<i>asp</i>

کم آوردن	senti	آن	ent	دوربین	asp
کم کیفیت	senti	پی	ent	دوربینش	asp
مورد نظر	senti	نوت	ent	دوربینی	asp
نقاط ضعف	senti	G2	ent	رنگ‌ها	asp
نقاط قوت	senti	LG	ent	صفحه‌ی	asp
Full HD	senti	s3	ent	ضد ضربه	asp
ایرادی	senti	آن	ent	ضد غبار	asp
بالایی	senti	سامسونگ گلکسی اس ۴	ent	طراحی	asp
بهترین	senti	سامسونگ گلکسی نند	ent	عکس‌ها	asp
بی‌عیب	senti	سونی اکسپریا اس پی	ent	عکس‌های	asp
تیک زدن	senti	سونی اکسپریا زد آر	ent	عملکرد	asp
حس خوبی	senti	موتورولا موتو ایکس	ent	قابلیت	asp
خوش دست	senti	Samsung Galaxy S4	ent	کارایی	asp
زیبایی	senti	Samsung Galaxy S	ent	کارکردش	asp
شاهکار	senti	اچ تی سی ال‌تی‌میت	ent	کنتراس	asp
طولانی	senti	Motorola Moto X	ent	کیفیتش	asp
عیب‌های	senti	Samsung Note II	ent	گرافیک	asp
کارآمد	senti	اکسپریا تی ایکس	ent	مقاومت	asp
کاربردی	senti	این مدل سونی را	ent	باتری	asp
کمبودی	senti	سونی اکسپریا ال	ent	باتریش	asp
متفاوت	senti	سونی اکسپریا پی	ent	باتری	asp
متناقض	senti	سونی اکسپریا زد	ent	باتری	asp
مشکلات	senti	سونی اکسپریا وی	ent	برنامه	asp
مناسبی	senti	مدل‌های xperia	ent	تصاویر	asp
هوشمند	senti	Apple iPhone 5	ent	تصویر	asp
آشنایی	senti	Samsung Grand	ent	خریدم	asp
بالای	senti	Sony Xperia SP	ent	دوربین	asp
بزرگی	senti	Sony Xperia TX	ent	زیبایی	asp
بیشتر	senti	Sony Xperia Z	ent	سیستم	asp
پایین	senti	Sony Xperia ZR	ent	صدایش	asp
جذابی	senti	در این گوشی‌ها	ent	طراحی	asp
جولوتر	senti	Galaxy Grand	ent	طراحی	asp
دلچسب	senti	Huawei Ascend	ent	ظاهرش	asp
عاشقش	senti	Samsung S III	ent	ظاهری	asp
قشنگی	senti	Sony Xperia L	ent	قیمتش	asp
کاملی	senti	Sony Xperia P	ent	کیفیت	asp
کم نور	senti	Sony Xperia V	ent	اجرای	asp
کوته	senti	اکسپریا اس پی	ent	امنیت	asp
متناسب	senti	اکسپریا زد آر	ent	بازی	asp
محدود	senti	ال جی اینیموس	ent	بدنه	asp
مزایای	senti	گالاکسی نوت ۲	ent	تاچش	asp
مستحکم	senti	HTC Ultimate	ent	خرید	asp

مشکلات	<i>senti</i>	اچ تی سی وان	<i>ent</i>	در کل	<i>asp</i>
مشکلی	<i>senti</i>	این مدل سونی	<i>ent</i>	ساخت	<i>asp</i>
معایب	<i>senti</i>	Sony Xperia	<i>ent</i>	سایز	<i>asp</i>
مقاوم	<i>senti</i>	اپل آیفون ۵	<i>ent</i>	سرعت	<i>asp</i>
مناسب	<i>senti</i>	LG Optimus	<i>ent</i>	صدای	<i>asp</i>
وسیع	<i>senti</i>	Samsung S4	<i>ent</i>	صفحه	<i>asp</i>
اچ دی	<i>senti</i>	سنسور تنظیم کننده‌ی نور	<i>sub</i>	ضد آب	<i>asp</i>
اعلا	<i>senti</i>	امکانات سخت‌افزاری	<i>sub</i>	ضدآب	<i>asp</i>
بالا	<i>senti</i>	تکنولوژی Bravia	<i>sub</i>	ظواهرش	<i>asp</i>
بالای	<i>senti</i>	سنسور تنظیم نور	<i>sub</i>	عکاسی	<i>asp</i>
بزرگ	<i>senti</i>	جذب اثر انگشت	<i>sub</i>	قیمت	<i>asp</i>
بهرتر	<i>senti</i>	زاویه‌های دید	<i>sub</i>	گوشی	<i>asp</i>
جالب	<i>senti</i>	نمایش تصاویر	<i>sub</i>	ویژگی	<i>asp</i>
حساس	<i>senti</i>	توان مانور	<i>sub</i>	RAM	<i>asp</i>
خوبش	<i>senti</i>	کیفیت تصویر	<i>sub</i>	بدنه	<i>asp</i>
خوبی	<i>senti</i>	زوایای دید	<i>sub</i>	خرید	<i>asp</i>
دوست	<i>senti</i>	نرم‌افزاری	<i>sub</i>	دقت	<i>asp</i>
راحت	<i>senti</i>	اثر انگشت	<i>sub</i>	سرعت	<i>asp</i>
راضی	<i>senti</i>	در پوش‌های	<i>sub</i>	ظاهر	<i>asp</i>
رقیب	<i>senti</i>	زاویه دید	<i>sub</i>	قیمت	<i>asp</i>
زیاد	<i>senti</i>	Full HD	<i>sub</i>	عکس	<i>asp</i>
زیادی	<i>senti</i>	کنتر است	<i>sub</i>	وزن	<i>asp</i>
زیبا	<i>senti</i>	جزئیات	<i>sub</i>	و این یعنی رفع یکی از مهمترین مشکلات این محصول	<i>emph</i>
سختی	<i>senti</i>	رزولوشن	<i>sub</i>	اصلا مهم تلقی نمی‌شود.	<i>emph</i>
سریع	<i>senti</i>	شفافیت	<i>sub</i>	که چیز زیاد مهمی نیست	<i>emph</i>
شفاف	<i>senti</i>	عملکرد	<i>sub</i>	و خیلی هم فراتر رفته.	<i>emph</i>
ضعفش	<i>senti</i>	کنتر است	<i>sub</i>	اطمینان داشته باشید	<i>emph</i>
ضعفی	<i>senti</i>	کیفیتش	<i>sub</i>	تنها عیبی هم که دارد	<i>emph</i>
ضعیف	<i>senti</i>	مایعات	<i>sub</i>	شکی وجود ندارد	<i>emph</i>
عاشق	<i>senti</i>	رنگ‌ها	<i>sub</i>	مهم این هست که	<i>emph</i>
عالی	<i>senti</i>	زیر اب	<i>sub</i>	از همه جهات	<i>emph</i>
عیش	<i>senti</i>	طراحی	<i>sub</i>	از همه مهمتر	<i>emph</i>
کارا	<i>senti</i>	عملکرد	<i>sub</i>	از دستش ندید	<i>emph</i>
کاهش	<i>senti</i>	کیفیت	<i>sub</i>	که مهم نیست	<i>emph</i>
کوتاه	<i>senti</i>	مقدار	<i>sub</i>	حرف ندارد	<i>emph</i>
محشر	<i>senti</i>	بدنه	<i>sub</i>	زیرا	<i>emph</i>
محکم	<i>senti</i>	زاویه	<i>sub</i>	اما علی رغم	<i>neg</i>
مرده	<i>senti</i>	سایز	<i>sub</i>	می‌توانست	<i>neg</i>
مزیت	<i>senti</i>	ضربه	<i>sub</i>	امیدوارم	<i>neg</i>
مشکل	<i>senti</i>	غبار	<i>sub</i>	مشخص نیست	<i>neg</i>

مشکلتش	<i>senti</i>	کیفیت	<i>sub</i>	بعید هست	<i>neg</i>
مشکلی	<i>senti</i>	جنس	<i>sub</i>	نه چندان	<i>neg</i>
مطلوب	<i>senti</i>	سرعت	<i>sub</i>	ولی اگر	<i>neg</i>
معایب	<i>senti</i>	عمر	<i>sub</i>	عیبی	<i>neg</i>
مناسب	<i>senti</i>	مصرف	<i>sub</i>	اما	<i>neg</i>
موفق	<i>senti</i>	نویز	<i>sub</i>	عدم	<i>neg</i>
نقصی	<i>senti</i>	رنگ	<i>sub</i>	جز	<i>neg</i>
واضح	<i>senti</i>	در شرایط مختلف نورانی	<i>ex</i>	نه	<i>neg</i>
<i>TFT</i>	<i>senti</i>	را هم که من تست کردم	<i>ex</i>	که راحت می‌شود ...	<i>rea</i>
آسان	<i>senti</i>	قبل از خرید می‌دانستم	<i>ex</i>	چون معمولاً ...	<i>rea</i>
بالا	<i>senti</i>	که اولین محسوب می‌شود	<i>ex</i>	که کاربر را مجبور به ...	<i>rea</i>
بروز	<i>senti</i>	و بعد اینکه خشکش کردی	<i>ex</i>	به دلیل داشتن ...	<i>rea</i>
بزرگ	<i>senti</i>	از جمله بازی آسفالت	<i>ex</i>	و خودم شخصا ...	<i>rea</i>
بودن	<i>senti</i>	از نظر دقت و شارپ نس	<i>ex</i>	به علت ...	<i>rea</i>
جالب	<i>senti</i>	به همراه تعدادی <i>Fix</i>	<i>ex</i>	به دلیل ...	<i>rea</i>
جذب	<i>senti</i>	نسبت به قابلیت هایش	<i>ex</i>	که باعث ...	<i>rea</i>
خوب	<i>senti</i>	من با آن کار کردم	<i>ex</i>	چون ...	<i>rea</i>
خوبی	<i>senti</i>	نسبت به سخت افزارش	<i>ex</i>	به خاطر ...	<i>rea</i>
دقیق	<i>senti</i>	با استفاده متوسط	<i>ex</i>	اینقدر زیاد	<i>in</i>
دوام	<i>senti</i>	به نظر می‌آید که	<i>ex</i>	بسیار پایین	<i>in</i>
راضی	<i>senti</i>	که می‌بایست بگویم	<i>ex</i>	تعداد زیادی	<i>in</i>
زیبا	<i>senti</i>	مهمترین عوارض آن	<i>ex</i>	فوق‌العاده	<i>in</i>
سبک	<i>senti</i>	می‌توانم بگویم که	<i>ex</i>	تا این حد	<i>in</i>
شکیل	<i>senti</i>	نسبت به امکاناتش	<i>ex</i>	کمی بیشتر	<i>in</i>
شیک	<i>senti</i>	البته به نظر من	<i>ex</i>	کوچکترین	<i>in</i>
ضعف	<i>senti</i>	توسط افراد عادی	<i>ex</i>	یک جورایی	<i>in</i>
ضعیف	<i>senti</i>	وقتی آن را دیدم	<i>ex</i>	بهراحتی	<i>in</i>
عالی	<i>senti</i>	خلاصه مطلب آنکه	<i>ex</i>	بینهایت	<i>in</i>
عیب	<i>senti</i>	در هنگام مکالمه	<i>ex</i>	خیلی زود	<i>in</i>
قوی	<i>senti</i>	نسبت به امکانات	<i>ex</i>	فقط ۲ تا	<i>in</i>
کمی	<i>senti</i>	یا اصطلاحاً قبلیت	<i>ex</i>	مهمترین	<i>in</i>
لذت	<i>senti</i>	نسبت به رقبايش	<i>ex</i>	نسبتاً کم	<i>in</i>
محکم	<i>senti</i>	از همه مهمتر	<i>ex</i>	یک خورده	<i>in</i>
مشکل	<i>senti</i>	راستش من اصلاً	<i>ex</i>	این همه	<i>in</i>
مکت	<i>senti</i>	نسبت به ابعاد	<i>ex</i>	بمقدری	<i>in</i>
نقص	<i>senti</i>	نسبت به قیمتش	<i>ex</i>	خیلی کم	<i>in</i>
هنگ	<i>senti</i>	هر چه بخواهید	<i>ex</i>	مقداری	<i>in</i>
<i>HDR</i>	<i>senti</i>	ولی به نظر من	<i>ex</i>	همه چیز	<i>in</i>
بد	<i>senti</i>	از حد واقعی	<i>ex</i>	هیچوقت	<i>in</i>
تک	<i>senti</i>	از نظر من که	<i>ex</i>	واقعاً	<i>in</i>
خوب	<i>senti</i>	برای بعضی‌ها	<i>ex</i>	یک قدری	<i>in</i>
داغ	<i>senti</i>	چند عدد نکته	<i>ex</i>	اندکی	<i>in</i>

زود	<i>senti</i>	نسبت به رقبا	<i>ex</i>	بسیار	<i>in</i>
قوی	<i>senti</i>	از این بابت	<i>ex</i>	تأحدی	<i>in</i>
کم	<i>senti</i>	از همه نظر	<i>ex</i>	چندان	<i>in</i>
کم	<i>senti</i>	استفاده از	<i>ex</i>	خیلیم	<i>in</i>
		با همه آنها	<i>ex</i>	زیادی	<i>in</i>
		در یک لحظه	<i>ex</i>	کاملاً	<i>in</i>
		روزهای اول	<i>ex</i>	نسبتاً	<i>in</i>
		مورد دوم هم	<i>ex</i>	واقعا	<i>in</i>
		<i>NFC</i> و <i>DLNA</i>	<i>ex</i>	اصلاً	<i>in</i>
		از نظر دقت	<i>ex</i>	بسیار	<i>in</i>
		از همه جهت	<i>ex</i>	تنها	<i>in</i>
		از همه نظر	<i>ex</i>	خیلی	<i>in</i>
		توسط کاربر	<i>ex</i>	زیاد	<i>in</i>
		ولی حیف که	<i>ex</i>	کامل	<i>in</i>
		از هر نظر	<i>ex</i>	کاملاً	<i>in</i>
		به نظر من	<i>ex</i>	نسبتاً	<i>in</i>
		دوم اینکه	<i>ex</i>	خیلی	<i>in</i>
		ولی در کل	<i>ex</i>	زود	<i>in</i>
		با آپدیت	<i>ex</i>	کمی	<i>in</i>
		بنظر من	<i>ex</i>	هیچ	<i>in</i>
		در اینکه	<i>ex</i>	زود	<i>in</i>
		در مجموع	<i>ex</i>	هیچ	<i>in</i>
		ولی باز	<i>ex</i>		
		در خودش	<i>ex</i>		
		من خودم	<i>ex</i>		
		نسبت به	<i>ex</i>		
		در کل	<i>ex</i>		
		نقطاتی	<i>ex</i>		

جدول ۷-۲: دسته‌های ایجاد شده برای استخراج الگو بر اساس جایگشت برجسبهای حسی

دسته‌ها (از چپ به راست)	نرخ وقوع
<i>senti + v</i>	236
<i>ent + senti + v</i>	147
<i>asp + senti + v</i>	120
<i>senti + senti + v</i>	76
<i>ent + senti + senti + v</i>	45
<i>ent + v</i>	36
<i>sub + senti + v</i>	35
<i>asp + senti + senti + v</i>	34
<i>senti + ent + v</i>	32
<i>senti + senti + senti + v</i>	28
<i>v + ent + v</i>	24
<i>v + senti + v</i>	23
<i>vs</i>	22
<i>asp + senti</i>	21
<i>asp + v</i>	20
<i>ent + senti</i>	20
<i>v</i>	19
<i>senti + ent + senti + v</i>	19
<i>senti + v + v</i>	18
<i>sub + asp + senti + v</i>	17
<i>ent + asp + senti + v</i>	14
<i>senti + asp + senti + v</i>	13
<i>ent + v + senti + v</i>	13
<i>senti + asp + v</i>	12
<i>v + v</i>	12
<i>asp + ent + senti + v</i>	12
<i>sub + v</i>	12
<i>sub + asp + senti + senti + v</i>	10
<i>ent + ent + senti + v</i>	10
<i>senti + ent</i>	9
<i>senti</i>	9
<i>senti + senti + senti</i>	9
<i>sub + ent + senti + v</i>	9
<i>ent + senti + senti + senti + v</i>	9
<i>sub + senti + senti + v</i>	9
<i>senti + sub + v</i>	8
<i>asp + v + senti + v</i>	8
<i>ent + v + v</i>	8
<i>sub + asp + v</i>	7

<i>senti + senti</i>	7
<i>sub + senti + asp</i>	7
<i>sub + senti</i>	7
<i>senti + senti + ent + v</i>	7
<i>ent</i>	7
<i>ent + senti + senti</i>	7
<i>v + v + senti + v</i>	7
<i>senti + v + senti + v</i>	7
<i>ent + v + senti + senti + v</i>	7
<i>asp + vs</i>	6
<i>asp</i>	6
<i>asp + senti + senti</i>	6
<i>sub + asp + senti</i>	6
<i>asp + sub + senti + v</i>	6
<i>asp + ent + v</i>	6
<i>neg + senti + v</i>	6
<i>sub + ent + senti + senti + v</i>	5
<i>asp + ent + senti + senti + v</i>	5
<i>rea + senti + senti + v</i>	5
<i>ent + senti + v + senti + v</i>	5
<i>sub + senti + senti + senti + v</i>	5
<i>v + v + v</i>	5
<i>v + senti + senti + v</i>	5
<i>asp + v + v</i>	5
<i>asp + senti + ent + v</i>	5
<i>senti + asp + senti + senti + v</i>	5
<i>v + asp + senti + senti + v</i>	5
<i>sub + ent + v</i>	4
<i>senti + vs</i>	4
<i>rea + senti + v</i>	4
<i>ent + ent + v</i>	4
<i>v + v + ent + senti + v</i>	4
<i>v + ent</i>	4
<i>senti + v + ent + v</i>	4
<i>asp + senti + senti + senti + v</i>	4
<i>senti + sub + ent + v</i>	4
<i>ent + senti + ent + v</i>	4
<i>senti + senti + ent + v + v</i>	4
<i>ent + senti + v + v</i>	4
<i>ent + senti + asp + senti + senti + v</i>	4
<i>v + asp + senti + v</i>	4
<i>asp + senti + ent</i>	4

<i>senti + senti + ent + asp + v</i>	4
<i>senti + asp + senti</i>	3
<i>asp + senti + v + v</i>	3
<i>senti + senti + v + v</i>	3
<i>asp + asp + v</i>	3
<i>ent + ent + senti + ent + v</i>	3
<i>ent + asp + asp + v</i>	3
<i>ent + asp + v</i>	3
<i>sub + senti + asp + v</i>	3
<i>senti + v + ent</i>	3
<i>senti + asp</i>	3
<i>senti + v + asp</i>	3
<i>asp + v + senti + senti + v</i>	3
<i>senti + senti + v + senti + v</i>	3
<i>asp + ent + senti</i>	3
<i>senti + senti + v + senti</i>	3
<i>senti + v + v + v</i>	3
<i>rea + ent + senti + v</i>	3
<i>senti + v + senti</i>	3
<i>ent + v + ent</i>	3
<i>senti + senti + senti + ent + v</i>	3
<i>ent + v + v + senti + senti + v</i>	3
<i>ent + senti + senti + v + ent + v</i>	3
<i>asp + asp + senti + v</i>	3
<i>asp + senti + senti + senti</i>	3
<i>senti + senti + asp + v</i>	3
<i>neg + senti + senti + v</i>	3
<i>senti + ent + asp + v</i>	3
<i>senti + ent + v + senti + v</i>	3
<i>ent + senti + senti + senti</i>	3
<i>vs + sub</i>	2
<i>ent + asp</i>	2
<i>ent + senti + senti + v + senti + v</i>	2
<i>sub + sub + senti + senti + senti</i>	2
<i>sub + asp + senti + senti</i>	2
<i>v + senti + asp</i>	2
<i>senti + ent + v + v</i>	2
<i>neg + sub + asp</i>	2
<i>neg + senti + ent + senti + v</i>	2
<i>senti + senti + v + senti + senti + v</i>	2
<i>ent + senti + asp + v</i>	2
<i>asp + senti + ent + v + v</i>	2

<i>vs + v</i>	2
<i>ent + vs</i>	2
<i>sub + ent + senti + senti + senti + v</i>	2
<i>senti + asp + vs</i>	2
<i>asp + ent + sub + senti + v</i>	2
<i>senti + senti + asp + senti + v</i>	2
<i>senti + v + asp + senti + v</i>	2
<i>v + asp + ent + senti + v</i>	2
<i>v + ent + v + senti + v</i>	2
<i>ent + sub + asp + senti + v</i>	2
<i>senti + senti + senti + asp + v</i>	2
<i>vs + ent + v</i>	2
<i>ent + senti + v + senti + senti + v</i>	2
<i>senti + senti + v + asp + v</i>	2
<i>asp + senti + vs</i>	2
<i>senti + sub</i>	2
<i>senti + ent + v + ent</i>	2
<i>v + asp + senti</i>	2
<i>v + senti + asp + senti + v</i>	2
<i>senti + v + senti + senti + senti + senti + v + v</i>	2
<i>senti + senti + v + senti + ent</i>	2
<i>sub + asp + vs</i>	2
<i>senti + ent + senti + v + v + ent + v</i>	2
<i>rea + senti + v + v</i>	2
<i>sub + senti + senti + v + in</i>	2
<i>v + senti + senti + senti + v</i>	2
<i>v + asp + v + v + senti + v</i>	2
<i>v + v + senti</i>	2
<i>ent + senti + ent + v + senti + v</i>	2
<i>senti + senti + senti + senti + v + senti</i>	2
<i>senti + v + v + ent + v</i>	2
<i>v + senti + ent + v</i>	2
<i>asp + senti + senti + v + v</i>	2
<i>ent + senti + senti + ent + v</i>	2
<i>asp + asp + senti + senti + senti + v</i>	2
<i>senti + senti + ent + v + ent</i>	2
<i>senti + ent + senti + senti + ent + v</i>	2
<i>senti + ent + v + v + senti + v</i>	2
<i>v + senti</i>	2
<i>senti + v + v + senti + v</i>	2
<i>asp + v + senti + ent + v</i>	2
<i>senti + senti + ent + senti + v</i>	2

<i>rea + asp + v</i>	2
<i>ent + senti + v + senti + ent + v</i>	2
<i>asp + sub</i>	2
<i>v + ent + senti + v</i>	2
<i>sub + v + senti + v</i>	2
<i>neg + asp + senti + senti + v</i>	2
<i>senti + ent + senti + senti + v</i>	2
<i>ent + v + in</i>	2
<i>senti + senti + v + ent + v</i>	2
<i>ent + senti + senti + senti + senti + senti + v</i>	2
<i>ent + senti + v + sub + asp + senti + v</i>	2
<i>ent + senti + v + senti</i>	2
<i>ent + senti + v + asp + v</i>	2
<i>sub + sub + senti + senti + v</i>	2
<i>ent + senti + v + ent + v</i>	2
<i>rea + v</i>	2
<i>asp + asp + senti + senti + v</i>	2
<i>sub</i>	2
<i>asp + senti + asp + v</i>	2
<i>ent + ent + senti + senti + senti + v</i>	2
<i>sub + senti + senti</i>	2
<i>v + senti + v + v</i>	2
<i>sub + senti + senti + ent + v</i>	2
<i>sub + asp + ent + v</i>	2
<i>senti + ent + ent + v</i>	2
<i>senti + sub + asp + senti + v</i>	2
<i>senti + ent + senti + v + senti + asp + v</i>	2
<i>ent + ent + v + senti + v</i>	2
<i>sub + asp + v + senti + v</i>	2
<i>ent + v + asp + senti + senti + v</i>	2
<i>v + asp + v</i>	2
<i>ent + senti + ent + senti + v</i>	2
<i>neg + ent + senti + v</i>	2
<i>senti + asp + senti + v + v</i>	2
<i>asp + senti + ent + senti + senti + senti + v</i>	2
<i>asp + v + senti</i>	2
<i>sub + asp + senti + senti + senti + v</i>	2
<i>neg + senti + senti + senti + v</i>	2
<i>ent + senti + asp + ent + v</i>	2
<i>senti + senti + senti + senti + v</i>	2
<i>asp + senti + senti + sub + sub + senti</i>	1
<i>asp + senti + senti + asp + senti</i>	1

<i>asp + asp + ent + v</i>	1
<i>senti + v + asp + ent + senti + senti + senti</i>	1
<i>asp + ent + rea + senti + v</i>	1
<i>asp + senti + sub + ent + v</i>	1
<i>asp + asp + ent + vs</i>	1
<i>sub + sub + asp + senti + v</i>	1
<i>senti + sub + asp + v</i>	1
<i>v + asp + v + neg</i>	1
<i>rea + senti + ent + senti + ent + v</i>	1
<i>emph + asp + ent + sub + senti + senti + v</i>	1
<i>rea + senti + asp + v + v</i>	1
<i>sub + senti + senti + asp + ent + vs</i>	1
<i>asp + senti + senti + v + senti + ent + senti + v + asp + v</i>	1
<i>rea + ent + senti + sub + sub + senti + v</i>	1
<i>ent + senti + v + emph</i>	1
<i>asp + ent + senti + v + v + v + sub + senti + v</i>	1
<i>sub + senti + asp + senti + senti</i>	1
<i>sub + senti + asp + senti + sub</i>	1
<i>neg + senti</i>	1
<i>ent + ent + senti + senti + v + emph</i>	1
<i>senti + senti + rea + v + asp</i>	1
<i>sub + asp + ent</i>	1
<i>v + senti + asp + ent</i>	1
<i>v + asp + ent + senti + senti + senti + v</i>	1
<i>rea</i>	1
<i>senti + senti + senti + ent + asp + senti + v</i>	1
<i>senti + senti + v + senti + v + asp + senti + v + asp</i>	1
<i>asp + senti + asp + senti + senti + senti + senti + v + v</i>	1
<i>asp + ent + senti + sub + senti + v</i>	1
<i>asp + senti + asp</i>	1
<i>senti + v + in</i>	1
<i>vs + asp</i>	1
<i>neg + vs + neg + senti + v</i>	1
<i>senti + senti + sub + asp + senti + senti + v + v</i>	1
<i>ent + sub + v + senti + ent + senti + v</i>	1
<i>senti + v + senti + v + v</i>	1
<i>asp + v + vs</i>	1
<i>neg + v + ent + vs</i>	1
<i>asp + ent + vs</i>	1
<i>ent + senti + v + vs</i>	1
<i>asp + senti + v + sub + senti + v</i>	1
<i>neg + v + neg + asp</i>	1

<i>asp + sub + ent + senti + sub + senti + senti + v</i>	1
<i>neg + v + senti + v</i>	1
<i>senti + senti + v + emph</i>	1
<i>v + sub + asp + vs</i>	1
<i>ent + v + asp + senti + v + senti + senti + v</i>	1
<i>ent + senti + v + senti + senti + asp + asp + v</i>	1
<i>sub + asp + senti + vs</i>	1
<i>emph + rea + v + senti + v + v + senti + ent + v</i>	1
<i>ent + asp + senti + senti + senti + vs</i>	1
<i>neg + ent + asp + v</i>	1
<i>neg + v + asp + ent + senti + senti + v</i>	1
<i>ent + senti + v + asp + senti + senti + v</i>	1
<i>emph + ent + asp + v</i>	1
<i>asp + senti + v + sub + senti</i>	1
<i>ent + senti + vs</i>	1
<i>asp + v + sub + senti + v</i>	1
<i>asp + senti + senti + senti + asp + senti + asp + senti</i>	1
<i>asp + sub + senti + sub + sub + senti + v</i>	1
<i>senti + v + ent + v + ent + v + senti + v</i>	1
<i>senti + ent + v + asp + senti + senti + v + sub + asp + senti + v</i>	1
<i>+</i>	
<i>ent + senti + senti + sub + senti + senti + asp + senti + senti + v</i>	1
<i>neg + senti + v + senti + senti + senti + ent + v</i>	1
<i>senti + ent + sub + asp + senti + v</i>	1
<i>senti + ent + ent + v + senti + asp + v</i>	1
<i>asp + ent + senti + v + emph + asp + asp + asp + asp</i>	1
<i>sub + asp + senti + v + senti + v + sub + senti + senti + v</i>	1
<i>sub + asp + senti + v + emph</i>	1
<i>asp + senti + sub + senti + asp + asp + senti + senti</i>	1
<i>senti + senti + asp + senti</i>	1
<i>asp + asp + v + ent + senti + v</i>	1
<i>senti + asp + v + senti + v</i>	1
<i>senti + v + vs + v + ent + v</i>	1
<i>senti + sub + v + vs</i>	1
<i>v + neg + senti + ent + v</i>	1
<i>neg + ent + v</i>	1
<i>neg + ent + v + senti + senti + v</i>	1
<i>ent + senti + v + neg + senti + v</i>	1
<i>senti + senti + ent + asp + senti + v</i>	1
<i>ent + senti + sub + senti + v + senti + v</i>	1
<i>neg + ent + asp + ent + v</i>	1

<i>asp + asp + sub + asp + senti</i>	1
<i>ent + ent + ent + senti + v</i>	1
<i>ent + senti + senti + vs + v</i>	1
<i>neg + senti + senti + ent + v + v + v</i>	1
<i>sub + senti + asp + asp</i>	1
<i>asp + senti + asp + senti + senti</i>	1
<i>asp + v + sub + senti + ent + v</i>	1
<i>sub + ent + ent + v</i>	1
<i>asp + senti + ent + sub + senti + senti + v</i>	1
<i>asp + v + v + senti + senti + v</i>	1
<i>rea + v + senti + ent + senti + senti + asp + asp + v</i>	1
<i>v + v + v + v + senti + v</i>	1
<i>ent + senti + senti + senti + senti + v</i>	1
<i>neg + v + senti + senti + ent + v</i>	1
<i>neg + asp + ent + senti + senti + v</i>	1
<i>rea + ent + v + sub + senti + v</i>	1
<i>rea + v + senti + v</i>	1
<i>asp + asp + v + v + senti + v</i>	1
<i>sub + senti + ent + v</i>	1
<i>sub + ent</i>	1
<i>neg + asp + senti + v + senti + v</i>	1
<i>rea + senti</i>	1
<i>sub + v + v</i>	1
<i>senti + sub + senti + asp + v</i>	1
<i>neg + senti + ent + asp + ent + v</i>	1
<i>asp + v + in</i>	1
<i>asp + ent + senti + senti + senti + senti + v</i>	1
<i>asp + senti + v + sub + asp + v</i>	1
<i>sub + sub + ent + v</i>	1
<i>rea + asp + senti + v</i>	1
<i>v + senti + ent + vs</i>	1
<i>ent + senti + senti + v + v</i>	1
<i>senti + asp + ent + senti + v</i>	1
<i>senti + ent + senti + v + v</i>	1
<i>senti + asp + v + senti</i>	1
<i>senti + asp + ent + v + v</i>	1
<i>ent + v + senti + senti + ent + v</i>	1
<i>ent + v + ent + rea + senti + ent + v</i>	1
<i>ent + senti + v + sub + senti + senti + senti + v</i>	1
<i>v + sub + senti + v</i>	1
<i>v + senti + senti + sub + v</i>	1
<i>ent + v + asp + senti + v</i>	1

<i>asp + sub + asp + senti + v</i>	1
<i>ent + senti + v + senti + asp</i>	1
<i>senti + v + neg + senti + senti</i>	1
<i>sub + neg + senti</i>	1
<i>senti + sub + sub + senti + v</i>	1
<i>neg + v + v + v + senti</i>	1
<i>v + in</i>	1
<i>ent + ent + vs</i>	1
<i>senti + v + senti + senti + v</i>	1
<i>asp + v + asp + senti + v</i>	1
<i>senti + senti + v + ent + senti + asp</i>	1
<i>ent + senti + ent</i>	1
<i>sub + asp + ent + senti + v</i>	1
<i>sub + ent + ent + senti + senti + v</i>	1
<i>senti + v + v + sub + senti + v</i>	1
<i>sub + ent + sub + v + neg + v</i>	1
<i>rea + senti + v + asp + senti + v</i>	1
<i>senti + v + v + senti + sub + sub + v</i>	1
<i>senti + v + ent + senti + senti + v</i>	1
<i>neg + ent + senti + vs</i>	1
<i>sub + ent + senti + v + sub + senti + v</i>	1
<i>senti + neg + senti</i>	1
<i>senti + ent + v + ent + v</i>	1
<i>rea + asp + v + v</i>	1
<i>ent + senti + senti + v + asp + senti + v</i>	1
<i>vs + senti + senti + v</i>	1
<i>senti + senti + sub + senti + v</i>	1
<i>ent + senti + v + ent + asp</i>	1
<i>ent + v + senti + sub + asp + senti + v</i>	1
<i>v + v + v + senti + v</i>	1
<i>senti + senti + v + sub + ent + v</i>	1
<i>sub + senti + senti + asp + asp + senti</i>	1
<i>v + sub + senti + v + ent</i>	1
<i>v + senti + senti + senti</i>	1
<i>senti + senti + ent + sub + senti + ent + v</i>	1
<i>senti + senti + ent + v + senti</i>	1
<i>senti + v + senti + ent + v</i>	1
<i>senti + ent + sub + v</i>	1
<i>neg + sub + v</i>	1
<i>asp + asp</i>	1
<i>ent + v + asp + v</i>	1
<i>neg + senti + v + senti + senti + senti + v</i>	1

<i>rea + v + v</i>	1
<i>v + ent + v + ent + senti + senti + v</i>	1
<i>v + v + senti + ent + v + senti + v</i>	1
<i>senti + ent + v + asp + v</i>	1
<i>senti + senti + senti + v + senti + ent + v</i>	1
<i>senti + v + senti + ent + asp + v</i>	1
<i>ent + sub + v</i>	1
<i>asp + senti + ent + v + senti + v</i>	1
<i>asp + v + senti + senti + asp + v</i>	1
<i>senti + asp + v + senti + senti + v</i>	1
<i>asp + senti + v + asp</i>	1
<i>asp + senti + senti + v + neg + v + v + neg + v + v</i>	1
<i>asp + senti + senti + senti + senti + v + senti</i>	1
<i>senti + v + sub + asp + asp + v + senti + senti + v</i>	1
<i>senti + senti + asp + asp + ent + v</i>	1
<i>v + senti + senti + senti + senti + v</i>	1
<i>asp + ent + senti + v + sub</i>	1
<i>v + asp + v + v</i>	1
<i>neg + asp + asp + vs</i>	1
<i>senti + ent + sub + senti + sub + v</i>	1
<i>asp + senti + asp + senti</i>	1
<i>senti + neg + senti + v</i>	1
<i>neg + v + sub + asp + senti + v</i>	1
<i>senti + asp + sub + v</i>	1
<i>v + v + ent</i>	1
<i>v + v + senti + ent + asp + senti + v</i>	1
<i>ent + v + ent + v + ent + senti + v</i>	1
<i>v + sub + sub + senti + v</i>	1
<i>v + senti + senti + v + senti + ent + senti + v</i>	1
<i>senti + ent + sub + ent + v</i>	1
<i>sub + senti + asp + asp + senti</i>	1
<i>ent + v + senti + v + v</i>	1
<i>ent + senti + senti + v + v + senti</i>	1
<i>ent + senti + senti + v + asp + v + v + senti + v</i>	1
<i>asp + senti + senti + senti + v + v</i>	1
<i>v + ent + v + senti + ent + v + v</i>	1
<i>senti + v + asp + senti + senti + ent + v</i>	1
<i>senti + asp + senti + senti + v + v</i>	1
<i>senti + v + vs</i>	1
<i>v + v + senti + v + v</i>	1
<i>senti + asp + v + v</i>	1
<i>senti + ent + senti + asp + senti + v</i>	1

<i>ent + senti + ent + v + senti + ent</i>	1
<i>senti + asp + ent + v</i>	1
<i>neg + ent + senti + v + v</i>	1
<i>ent + v + ent + v</i>	1
<i>senti + v + senti + senti + senti + senti + v</i>	1
<i>neg + ent + v + v + v</i>	1
<i>senti + v + senti + ent + senti + v</i>	1
<i>senti + senti + senti + v + v</i>	1
<i>asp + asp + vs</i>	1
<i>senti + asp + senti + ent + v + senti + v + v</i>	1
<i>senti + v + ent + emph</i>	1
<i>v + v + senti + senti + asp</i>	1
<i>ent + senti + v + neg + v + neg + senti + v</i>	1
<i>neg + senti + v + ent + v + v</i>	1
<i>ent + ent + senti + v + v</i>	1
<i>v + ent + v + v</i>	1
<i>senti + ent + asp + senti + v + senti</i>	1
<i>senti + senti + senti + v + senti + senti + senti</i>	1
<i>senti + sub + senti + senti</i>	1
<i>senti + senti + asp + ent + asp + ent + v</i>	1
<i>ent + sub + senti + senti + v</i>	1
<i>sub + sub + asp + senti + sub + senti + v</i>	1
<i>asp + sub + senti + senti + v</i>	1
<i>asp + asp + asp</i>	1
<i>senti + ent + v + senti + ent + v</i>	1
<i>senti + v + asp + asp + senti + v</i>	1
<i>senti + v + senti + v + ent + v</i>	1
<i>sub + senti + senti + asp + v</i>	1
<i>senti + senti + ent + asp + senti + senti + v</i>	1
<i>asp + asp + v + neg + v</i>	1
<i>v + asp + senti + v + asp + v + v</i>	1
<i>senti + ent + v + v + senti + v + v</i>	1
<i>senti + v + senti + senti</i>	1
<i>neg + ent + senti + v + asp + senti + v</i>	1
<i>senti + asp + ent + senti + senti + v</i>	1
<i>ent + v + v + senti + ent</i>	1
<i>ent + ent + v + senti + senti + senti + v</i>	1
<i>ent + v + ent + senti + v</i>	1
<i>ent + senti + senti + asp + senti + v + senti + v</i>	1
<i>ent + senti + v + v + senti + ent + v</i>	1
<i>ent + senti + v + asp + asp + ent + v + asp</i>	1
<i>senti + asp + senti + v + ent + v</i>	1

<i>asp + senti + v + senti + senti + ent + v + v</i>	1
<i>senti + ent + senti + asp + senti + senti + v</i>	1
<i>senti + ent + senti + senti + senti + v</i>	1
<i>asp + ent + v + ent + senti + v</i>	1
<i>senti + v + asp + senti</i>	1
<i>ent + ent + ent + v</i>	1
<i>v + senti + senti + asp + v</i>	1
<i>senti + senti + senti + asp + senti + v</i>	1
<i>ent + senti + senti + ent + senti + v</i>	1
<i>asp + senti + senti + v + senti + v + v</i>	1
<i>ent + v + senti</i>	1
<i>asp + asp + senti + v + ent</i>	1
<i>ent + senti + senti + ent + v + senti</i>	1
<i>senti + ent + ent + v + senti + sub + senti + v</i>	1
<i>asp + ent + v + senti + v</i>	1
<i>senti + asp + ent + v + v + ent + senti + v</i>	1
<i>senti + ent + asp + ent + v</i>	1
<i>v + senti + ent + ent + v</i>	1
<i>sub + ent + senti + ent + senti + v</i>	1
<i>ent + neg + senti + ent + v + senti + ent</i>	1
<i>v + v + senti + senti + ent + v + senti + v</i>	1
<i>neg + asp + senti + sub + v</i>	1
<i>ent + senti + sub + ent + senti + v</i>	1
<i>senti + senti + vs</i>	1
<i>senti + asp + v + ent</i>	1
<i>ent + neg + senti + senti + v</i>	1
<i>neg + senti + v + v</i>	1
<i>senti + senti + sub + senti</i>	1
<i>ent + senti + v + senti + ent + senti + asp</i>	1
<i>senti + sub + senti + v + ent</i>	1
<i>v + v + senti + ent + v</i>	1
<i>asp + ent</i>	1
<i>ent + v + senti + v + senti + ent + v</i>	1
<i>sub + asp + sub + ent + senti + v + ent + senti + senti</i>	1
<i>senti + ent + asp</i>	1
<i>senti + senti + asp + senti + ent + v</i>	1
<i>v + sub + senti</i>	1
<i>ent + sub + v + v</i>	1
<i>v + senti + ent + v + v</i>	1
<i>asp + senti + senti + senti + senti + ent + v</i>	1
<i>ent + senti + senti + v + v + v</i>	1
<i>asp + senti + v + senti + ent + senti + v</i>	1

<i>v + v + ent + senti + ent + v</i>	1
<i>neg + sub + asp + in</i>	1
<i>ent + senti + senti + v + asp + v</i>	1
<i>asp + sub + senti</i>	1
<i>senti + senti + v + v + senti</i>	1
<i>emph + senti + v</i>	1
<i>senti + ent + ent + ent + ent + v</i>	1
<i>senti + v + asp + v + senti + ent + asp + v</i>	1
<i>senti + asp + senti + asp</i>	1
<i>ent + senti + asp + senti + v</i>	1
<i>asp + senti + senti + asp + v</i>	1
<i>asp + asp + ent + senti + v</i>	1
<i>asp + asp + asp + senti</i>	1
<i>asp + asp + v + senti</i>	1
<i>ent + ent + senti + v + v + sub</i>	1
<i>ent + v + senti + v + vs</i>	1
<i>v + v + senti + senti + ent</i>	1
<i>sub + asp + senti + v + asp + senti + v</i>	1
<i>sub + asp + sub + v</i>	1
<i>senti + asp + sub + sub + asp + senti + asp + senti</i>	1
<i>senti + v + senti + ent + asp + asp + senti + v</i>	1
<i>ent + senti + sub + sub + senti + v</i>	1
<i>senti + asp + senti + senti</i>	1
<i>senti + sub + senti</i>	1
<i>asp + senti + senti + ent + v</i>	1
<i>sub + senti + asp + senti + v</i>	1
<i>ent + senti + v + senti + v + asp + senti + senti + v +</i>	1
<i>ent + senti + senti + v + asp</i>	1
<i>sub + sub + ent + senti + v</i>	1
<i>senti + senti + ent + ent + senti + senti + v</i>	1
<i>senti + ent + senti + senti + senti + senti + v</i>	1
<i>neg + ent + senti + v + senti + v</i>	1
<i>asp + senti + senti + v + ent + ent + v</i>	1
<i>v + asp + senti + v + asp + ent</i>	1
<i>v + ent + senti + asp + senti + v</i>	1
<i>asp + v + v + v</i>	1
<i>v + sub + senti + senti + v</i>	1
<i>v + ent + senti + v + senti + v</i>	1
<i>senti + asp + senti + v + asp + asp + senti</i>	1
<i>asp + senti + v + ent + v</i>	1
<i>sub + senti + v + sub + v</i>	1
<i>senti + sub + asp + senti + senti + v</i>	1

<i>sub + asp + asp + v</i>	1
<i>v + ent + in</i>	1
<i>senti + senti + asp + asp + ent + senti + v</i>	1
<i>sub + asp + asp + ent + v</i>	1
<i>v + v + v + asp + v + v + ent</i>	1
<i>senti + ent + ent + ent + sub + senti + v</i>	1
<i>ent + senti + asp + senti</i>	1
<i>ent + sub + ent + senti + v</i>	1
<i>asp + senti + ent + asp + v</i>	1
<i>senti + v + ent + senti</i>	1
<i>senti + v + asp + asp</i>	1
<i>asp + v + ent + v + ent + senti + v</i>	1
<i>ent + senti + ent + v + v</i>	1
<i>senti + senti + v + senti + senti</i>	1
<i>senti + v + ent + senti + v</i>	1
<i>senti + ent + asp + vs</i>	1
<i>asp + asp + senti + asp + senti + v</i>	1
<i>ent + ent + senti + senti + v</i>	1
<i>senti + sub + senti + v</i>	1
<i>ent + sub + sub + sub + senti + v</i>	1
<i>asp + ent + asp + senti + senti + v</i>	1
<i>ex</i>	1
<i>neg + senti + senti + ent + sub + v</i>	1
<i>ent + senti + v + asp + senti + v</i>	1
<i>ent + senti + asp</i>	1
<i>ent + v + ent + senti + v + v + v</i>	1
<i>ent + v + vs</i>	1
<i>sub + senti + v + v</i>	1
<i>rea + asp + ent + senti + v</i>	1
<i>ent + senti + asp + ent + sub + v</i>	1
<i>sub + asp + asp + senti + senti + v</i>	1
<i>senti + v + senti + senti + senti + v</i>	1
<i>senti + asp + senti + senti + senti</i>	1
<i>senti + asp + ent</i>	1
<i>senti + senti + asp + senti + senti + asp</i>	1
<i>senti + senti + senti + asp + senti + senti + senti</i>	1
<i>ent + v + v + senti + v</i>	1
<i>v + senti + ent + senti + v</i>	1
<i>senti + senti + senti + sub + v</i>	1
<i>asp + ent + asp + senti + v</i>	1
<i>sub + sub + senti + senti</i>	1
<i>v + senti + ent + asp + v</i>	1

<i>sub + v + ent + senti + v</i>	1
<i>senti + v + asp + ent + v</i>	1
<i>asp + senti + senti + sub + senti + ent + asp</i>	1
<i>ent + v + ent + ent + senti + v</i>	1
<i>asp + senti + senti + v + ent</i>	1
<i>senti + asp + v + ent + v</i>	1
<i>senti + senti + sub</i>	1
<i>asp + v + senti + v + v + ent + v</i>	1
<i>ent + senti + v + ent</i>	1
<i>senti + ent + ent + v + senti + v + senti</i>	1
<i>sub + sub + senti + v</i>	1
<i>senti + ent + asp + senti + asp + v</i>	1
<i>asp + senti + sub + senti + v + ent</i>	1
<i>sub + senti + asp + senti + sub + senti + v</i>	1
<i>senti + senti + v + asp + senti + senti + v</i>	1
<i>asp + asp + v + v</i>	1
<i>ent + senti + ent + v + v + senti</i>	1
<i>v + senti + v + ent + v</i>	1
<i>sub + ent + senti</i>	1
<i>asp + v + senti + v + senti + v</i>	1
<i>asp + v + v + senti + v</i>	1
<i>asp + senti + senti + sub</i>	1
<i>ent + senti + v + asp + senti</i>	1
<i>senti + senti + ent + v + asp + v</i>	1
<i>senti + ent + asp + senti + senti + v</i>	1
<i>ent + vs + ent</i>	1
<i>ent + sub + asp + senti + senti + v + senti + v</i>	1
<i>senti + senti + senti + senti + ent + v</i>	1
<i>ent + v + v + senti + asp + senti + v</i>	1
<i>ent + senti + senti + senti + senti</i>	1
<i>senti + neg + v + v + v</i>	1
<i>senti + senti + senti + senti + senti</i>	1
<i>ent + ent + v + sub + senti + senti + v</i>	1
<i>ent + senti + senti + v + v + senti + v</i>	1
<i>v + ent + senti + v + ent + v + senti + senti + v + v</i>	1
<i>v + v + ent + v</i>	1
<i>sub + senti + senti + sub + sub + senti</i>	1
<i>v + senti + senti + v + asp + v + senti + senti + v</i>	1
<i>in</i>	1
<i>asp + senti + asp + senti + senti + v</i>	1
<i>asp + asp + sub</i>	1
<i>asp + senti + senti + v + senti + senti + v</i>	1

<i>v + asp</i>	1
<i>v + ent + v + senti + senti</i>	1
<i>sub + ent + senti + senti + asp + senti + v</i>	1
<i>v + asp + sub + v + senti + v</i>	1
<i>asp + senti + v + senti + asp + senti</i>	1
<i>ent + v + senti + sub + senti + v</i>	1
<i>senti + asp + asp + senti + v</i>	1
<i>asp + v + asp + v</i>	1
<i>neg + v + senti + asp + v</i>	1
<i>asp + ent + senti + v + senti + v</i>	1
<i>v + asp + ent + vs</i>	1
<i>vs + v + ent + v</i>	1
<i>senti + asp + sub + sub + senti + v</i>	1
<i>asp + asp + sub + senti + senti + ent + senti + v</i>	1
<i>v + v + asp + senti + v</i>	1
<i>senti + v + senti + ent + senti + senti + asp + senti</i>	1
<i>neg + ent + v + ent + v + ent + sub + asp + senti + v</i>	1
<i>v + senti + senti + ent</i>	1
<i>neg + senti + asp + asp + ent + v</i>	1
<i>v + senti + asp + v</i>	1
<i>ent + ent + senti + senti + v + v</i>	1
<i>senti + senti + senti + senti + senti + v</i>	1
<i>sub + senti + ent + senti + senti + senti + v</i>	1
<i>senti + asp + senti + ent + senti + senti + v</i>	1
<i>neg + ent + v + v + v + v + v</i>	1
<i>sub + senti + senti + ent + senti + v</i>	1
<i>senti + senti + senti + v + senti + senti + v</i>	1
<i>senti + senti + asp + asp + v</i>	1
<i>senti + senti + v + ent + senti + v</i>	1
<i>senti + v + asp + ent + senti + senti + v</i>	1
<i>senti + ent + rea + v + ent</i>	1
<i>sub + sub + senti</i>	1
<i>asp + senti + ent + senti + v</i>	1
<i>neg + ent + v + ent + senti + asp + in</i>	1
<i>ent + v + senti + v + senti + v</i>	1
<i>senti + senti + asp + senti + senti + v</i>	1
<i>ent + senti + v + v + ent + v + ent + v</i>	1
<i>v + v + sub + senti + sub</i>	1
<i>ent + senti + sub + v</i>	1
<i>v + senti + ent + senti + senti + v</i>	1
<i>v + v + asp + sub + sub + senti + v</i>	1
<i>sub + asp + senti + v + asp + senti + v + senti</i>	1

<i>senti + v + ent + in</i>	1
<i>asp + senti + v + senti</i>	1
<i>neg + ent + v + senti + v</i>	1
<i>ent + sub</i>	1
<i>asp + senti + senti + senti + sub + senti + senti + asp + senti</i>	1
<i>senti + asp + v + vs</i>	1
<i>ent + senti + senti + senti + v + asp</i>	1
<i>ent + v + v + asp + senti + v</i>	1
<i>ent + asp + senti + senti + v</i>	1
<i>rea + asp + senti + sub + ent + senti + v</i>	1
<i>asp + senti + ent + senti + senti + v</i>	1
<i>ent + senti + senti + senti + sub + sub + senti + asp + senti + v</i>	1
<i>senti + senti + ent + ent + v + v</i>	1
<i>neg + senti + senti + v + senti</i>	1
<i>rea + asp + senti + v + asp + asp</i>	1
<i>ent + ent + v + senti</i>	1
<i>asp + sub + ent + sub + ent + senti + ent + senti + v</i>	1
<i>v + v + ent + asp + senti + v</i>	1
<i>ent + senti + v + ent + ent + v</i>	1
<i>sub + senti + asp + senti</i>	1

جدول ۳-۷: نقشهای نحوی مجاز بین دو برچسب حسی که با ترکیب با برچسب حسی منجر به ساخت الگو میشوند.

برچسب حسی اول	نقش نحوی مجاز بین دو برچسب حسی	برچسب حسی دوم
asp	empty	asp
asp	CONJ + PUNC	asp
asp	PUNC	asp
asp	empty	end
asp	PUNC	end
asp	N + PUNC	end
asp	empty	ent
asp	N + PUNC	ent
asp	empty	ex
asp	empty	in
asp	N + PUNC	in
asp	empty	senti
asp	ADV + PUNC	senti
asp	N + PUNC	senti
asp	SUBR + PUNC	senti
asp	PREP + N + PUNC	senti
asp	CONJ + N + PUNC	senti
asp	PUNC	senti
asp	POSTP + PUNC	senti
asp	CONJ + PUNC	senti
asp	PRENUM + N + PUNC	senti
asp	ADJ + PUNC	senti
asp	empty	sub
asp	CONJ + PUNC	sub
asp	PUNC	sub
asp	empty	v
asp	N + PUNC	v
asp	ADJ + PUNC	v
asp	N + N + PUNC	v
asp	POSTP + PUNC	v
asp	SUBR + PUNC	v
asp	PRENUM + N + PUNC	v
asp	empty	vs
emph	empty	end
ent	empty	asp
ent	POSTP + PUNC	asp
ent	PREP + PUNC	asp

ent	PREP + N + PUNC	asp
ent	N + PUNC	end
ent	PUNC	end
ent	empty	end
ent	POSTP + V + PUNC	end
ent	ADJ + N + PUNC	end
ent	empty	ent
ent	N + PUNC	ent
ent	empty	ex
ent	empty	in
ent	empty	senti
ent	N + PUNC	senti
ent	ADV + PUNC	senti
ent	POSTP + PUNC	senti
ent	number	senti
ent	N + N + PUNC	senti
ent	PREP + PUNC	senti
ent	PREP + N + PUNC	senti
ent	CONJ + PUNC	senti
ent	SUBR + PUNC	senti
ent	empty	sub
ent	N + PUNC	sub
ent	empty	v
ent	N + PUNC	v
ent	POSTP + PUNC	v
ent	ADJ + PUNC	v
ent	POSTP + N + PUNC	v
ent	PR + PUNC	v
ent	N + N + PUNC	v
ent	number	v
ent	N + ADJ + PUNC	v
ent	PREP + N + N + N + PUNC	v
ent	PREP + N + PUNC	v
ent	N + N + N + PUNC	v
ent	empty	vs
ex	empty	asp
ex	empty	end
ex	PUNC	end
ex	empty	ent
ex	PREP + PUNC	ent
ex	empty	ex

<i>ex</i>	<i>empty</i>	<i>in</i>
<i>ex</i>	<i>empty</i>	<i>senti</i>
<i>ex</i>	<i>N + N + PUNC</i>	<i>senti</i>
<i>ex</i>	<i>empty</i>	<i>sub</i>
<i>ex</i>	<i>empty</i>	<i>v</i>
<i>ex</i>	<i>N + PUNC</i>	<i>v</i>
<i>ex</i>	<i>N + N + PUNC</i>	<i>v</i>
<i>in</i>	<i>empty</i>	<i>asp</i>
<i>in</i>	<i>N + PUNC</i>	<i>end</i>
<i>in</i>	<i>empty</i>	<i>ent</i>
<i>in</i>	<i>PREP + PUNC</i>	<i>ent</i>
<i>in</i>	<i>empty</i>	<i>ex</i>
<i>in</i>	<i>empty</i>	<i>senti</i>
<i>in</i>	<i>N + PUNC</i>	<i>senti</i>
<i>in</i>	<i>N + N + PUNC</i>	<i>senti</i>
<i>in</i>	<i>empty</i>	<i>sub</i>
<i>in</i>	<i>empty</i>	<i>v</i>
<i>in</i>	<i>N + PUNC</i>	<i>v</i>
<i>in</i>	<i>ADJ + PUNC</i>	<i>v</i>
<i>in</i>	<i>PREP + PUNC</i>	<i>v</i>
<i>in</i>	<i>empty</i>	<i>vs</i>
<i>neg</i>	<i>empty</i>	<i>asp</i>
<i>neg</i>	<i>empty</i>	<i>ent</i>
<i>neg</i>	<i>empty</i>	<i>ex</i>
<i>neg</i>	<i>empty</i>	<i>senti</i>
<i>neg</i>	<i>N + PUNC</i>	<i>senti</i>
<i>neg</i>	<i>empty</i>	<i>v</i>
<i>rea</i>	<i>empty</i>	<i>asp</i>
<i>rea</i>	<i>empty</i>	<i>ent</i>
<i>rea</i>	<i>empty</i>	<i>in</i>
<i>rea</i>	<i>empty</i>	<i>senti</i>
<i>rea</i>	<i>empty</i>	<i>v</i>
<i>senti</i>	<i>empty</i>	<i>asp</i>
<i>senti</i>	<i>PR + PUNC</i>	<i>asp</i>
<i>senti</i>	<i>N + PUNC</i>	<i>asp</i>
<i>senti</i>	<i>PUNC</i>	<i>asp</i>
<i>senti</i>	<i>CONJ + PUNC</i>	<i>asp</i>
<i>senti</i>	<i>N + PREP + PUNC</i>	<i>asp</i>
<i>senti</i>	<i>PREP + N + PUNC</i>	<i>asp</i>
<i>senti</i>	<i>PREP + PUNC</i>	<i>asp</i>
<i>senti</i>	<i>V + PUNC</i>	<i>asp</i>
<i>senti</i>	<i>PUNC</i>	<i>end</i>

<i>senti</i>	<i>empty</i>	<i>end</i>
<i>senti</i>	<i>N + PUNC</i>	<i>end</i>
<i>senti</i>	<i>V + PUNC</i>	<i>end</i>
<i>senti</i>	<i>CONJ + ADJ + PUNC</i>	<i>end</i>
<i>senti</i>	<i>empty</i>	<i>ent</i>
<i>senti</i>	<i>PREP + PUNC</i>	<i>ent</i>
<i>senti</i>	<i>N + PUNC</i>	<i>ent</i>
<i>senti</i>	<i>ADJ + PUNC</i>	<i>ent</i>
<i>senti</i>	<i>SUBR + PUNC</i>	<i>ent</i>
<i>senti</i>	<i>V + PUNC</i>	<i>ent</i>
<i>senti</i>	<i>CONJ + ADJ + PUNC</i>	<i>ent</i>
<i>senti</i>	<i>empty</i>	<i>ex</i>
<i>senti</i>	<i>N + PUNC</i>	<i>ex</i>
<i>senti</i>	<i>empty</i>	<i>in</i>
<i>senti</i>	<i>empty</i>	<i>senti</i>
<i>senti</i>	<i>CONJ + PUNC</i>	<i>senti</i>
<i>senti</i>	<i>N + PUNC</i>	<i>senti</i>
<i>senti</i>	<i>PUNC</i>	<i>senti</i>
<i>senti</i>	<i>ADV + PUNC</i>	<i>senti</i>
<i>senti</i>	<i>PRENUM + N + PUNC</i>	<i>senti</i>
<i>senti</i>	<i>PR + PUNC</i>	<i>senti</i>
<i>senti</i>	<i>POSTP + PUNC</i>	<i>senti</i>
<i>senti</i>	<i>ADJ + PUNC</i>	<i>senti</i>
<i>senti</i>	<i>N + N + PUNC</i>	<i>senti</i>
<i>senti</i>	<i>PREP + N + PUNC</i>	<i>senti</i>
<i>senti</i>	<i>SUBR + PUNC</i>	<i>senti</i>
<i>senti</i>	<i>CONJ + N + PUNC</i>	<i>senti</i>
<i>senti</i>	<i>empty</i>	<i>sub</i>
<i>senti</i>	<i>PREP + PUNC</i>	<i>sub</i>
<i>senti</i>	<i>PUNC</i>	<i>sub</i>
<i>senti</i>	<i>CONJ + PUNC</i>	<i>sub</i>
<i>senti</i>	<i>PREP + N + PUNC</i>	<i>sub</i>
<i>senti</i>	<i>empty</i>	<i>v</i>
<i>senti</i>	<i>N + PUNC</i>	<i>v</i>
<i>senti</i>	<i>ADJ + PUNC</i>	<i>v</i>
<i>senti</i>	<i>N + N + PUNC</i>	<i>v</i>
<i>senti</i>	<i>ADV + PUNC</i>	<i>v</i>
<i>senti</i>	<i>CONJ + N + PUNC</i>	<i>v</i>
<i>senti</i>	<i>PREP + N + PUNC</i>	<i>v</i>
<i>senti</i>	<i>POSTP + PUNC</i>	<i>v</i>
<i>senti</i>	<i>N + POSTP + PUNC</i>	<i>v</i>
<i>senti</i>	<i>SUBR + PUNC</i>	<i>v</i>

<i>senti</i>	<i>PREP + PR + N + PUNC</i>	<i>v</i>
<i>senti</i>	<i>POSTP + N + PUNC</i>	<i>v</i>
<i>senti</i>	<i>PRENUM + N + PUNC</i>	<i>v</i>
<i>senti</i>	<i>N + N + N + PUNC</i>	<i>v</i>
<i>senti</i>	<i>PREP + N + N + PUNC</i>	<i>v</i>
<i>senti</i>	<i>V + PUNC</i>	<i>v</i>
<i>senti</i>	<i>PR + N + PUNC</i>	<i>v</i>
<i>senti</i>	<i>ADJ + N + PUNC</i>	<i>v</i>
<i>senti</i>	<i>CONJ + N + N + PUNC</i>	<i>v</i>
<i>senti</i>	<i>CONJ + ADJ + PUNC</i>	<i>v</i>
<i>senti</i>	<i>empty</i>	<i>vs</i>
<i>start</i>	<i>empty</i>	<i>asp</i>
<i>start</i>	<i>number</i>	<i>asp</i>
<i>start</i>	<i>CONJ + PUNC</i>	<i>asp</i>
<i>start</i>	<i>PREP + N + PUNC</i>	<i>asp</i>
<i>start</i>	<i>N + PUNC</i>	<i>asp</i>
<i>start</i>	<i>SUBR + PUNC</i>	<i>asp</i>
<i>start</i>	<i>ADV + PUNC</i>	<i>asp</i>
<i>start</i>	<i>PUNC</i>	<i>asp</i>
<i>start</i>	<i>PR + PREM + N + PUNC</i>	<i>asp</i>
<i>start</i>	<i>CONJ + N + PUNC</i>	<i>asp</i>
<i>start</i>	<i>PREP + PUNC</i>	<i>asp</i>
<i>start</i>	<i>PR + PUNC</i>	<i>asp</i>
<i>start</i>	<i>SUBR + PREP + PUNC</i>	<i>asp</i>
<i>start</i>	<i>N + N + PUNC</i>	<i>asp</i>
<i>start</i>	<i>CONJ + PREP + N + PUNC</i>	<i>asp</i>
<i>start</i>	<i>empty</i>	<i>emph</i>
<i>start</i>	<i>empty</i>	<i>ent</i>
<i>start</i>	<i>N + PUNC</i>	<i>ent</i>
<i>start</i>	<i>PR + PUNC</i>	<i>ent</i>
<i>start</i>	<i>SUBR + PUNC</i>	<i>ent</i>
<i>start</i>	<i>number</i>	<i>ent</i>
<i>start</i>	<i>CONJ + N + PUNC</i>	<i>ent</i>
<i>start</i>	<i>CONJ + PUNC</i>	<i>ent</i>
<i>start</i>	<i>SUBR + PREP + PUNC</i>	<i>ent</i>
<i>start</i>	<i>PR + PREP + PUNC</i>	<i>ent</i>
<i>start</i>	<i>PREP + N + PUNC</i>	<i>ent</i>
<i>start</i>	<i>CONJ + PREP + N + PUNC</i>	<i>ent</i>
<i>start</i>	<i>N + N + PUNC</i>	<i>ent</i>
<i>start</i>	<i>SUBR + N + PUNC</i>	<i>ent</i>

<i>start</i>	<i>ADJ + PUNC</i>	<i>ent</i>
<i>start</i>	<i>PREP + PUNC</i>	<i>ent</i>
<i>start</i>	<i>N + N + N + PUNC</i>	<i>ent</i>
<i>start</i>	<i>PR + SUBR + PREP + PUNC</i>	<i>ent</i>
<i>start</i>	<i>CONJ + PREP + PUNC</i>	<i>ent</i>
<i>start</i>	<i>N + PUNC + N + PUNC</i>	<i>ent</i>
<i>start</i>	<i>empty</i>	<i>ex</i>
<i>start</i>	<i>CONJ + PUNC</i>	<i>ex</i>
<i>start</i>	<i>SUBR + PUNC</i>	<i>ex</i>
<i>start</i>	<i>N + PUNC</i>	<i>ex</i>
<i>start</i>	<i>empty</i>	<i>in</i>
<i>start</i>	<i>CONJ + PUNC</i>	<i>in</i>
<i>start</i>	<i>SUBR + PUNC</i>	<i>in</i>
<i>start</i>	<i>N + PUNC</i>	<i>in</i>
<i>start</i>	<i>ADV + PUNC</i>	<i>in</i>
<i>start</i>	<i>PR + PUNC</i>	<i>in</i>
<i>start</i>	<i>empty</i>	<i>neg</i>
<i>start</i>	<i>empty</i>	<i>rea</i>
<i>start</i>	<i>empty</i>	<i>senti</i>
<i>start</i>	<i>CONJ + PUNC</i>	<i>senti</i>
<i>start</i>	<i>SUBR + PUNC</i>	<i>senti</i>
<i>start</i>	<i>N + PUNC</i>	<i>senti</i>
<i>start</i>	<i>CONJ + N + PUNC</i>	<i>senti</i>
<i>start</i>	<i>N + N + PUNC</i>	<i>senti</i>
<i>start</i>	<i>ADV + PUNC</i>	<i>senti</i>
<i>start</i>	<i>PREP + N + PUNC</i>	<i>senti</i>
<i>start</i>	<i>number</i>	<i>senti</i>
<i>start</i>	<i>CONJ + PREP + N + PUNC</i>	<i>senti</i>
<i>start</i>	<i>PR + PUNC</i>	<i>senti</i>
<i>start</i>	<i>N + PREP + PUNC</i>	<i>senti</i>
<i>start</i>	<i>ADV + N + PUNC</i>	<i>senti</i>
<i>start</i>	<i>SUBR + N + PUNC</i>	<i>senti</i>
<i>start</i>	<i>PUNC</i>	<i>senti</i>
<i>start</i>	<i>PREP + PUNC</i>	<i>senti</i>
<i>start</i>	<i>N + PREP + N + PUNC</i>	<i>senti</i>
<i>start</i>	<i>ADJ + PUNC</i>	<i>senti</i>
<i>start</i>	<i>SUBR + PR + PUNC</i>	<i>senti</i>
<i>start</i>	<i>PREP + PRENUM + N + PUNC</i>	<i>senti</i>
<i>start</i>	<i>CONJ + ADV + PUNC</i>	<i>senti</i>


<i>start</i>	<i>N + ADV + PUNC</i>	<i>senti</i>
<i>start</i>	<i>PR + PR + PUNC</i>	<i>senti</i>
<i>start</i>	<i>PR + SUBR + PUNC</i>	<i>senti</i>
<i>start</i>	<i>CONJ + N + N + PUNC</i>	<i>senti</i>
<i>start</i>	<i>PREM + N + PUNC</i>	<i>senti</i>
<i>start</i>	<i>PREP + N + N + PUNC</i>	<i>senti</i>
<i>start</i>	<i>empty</i>	<i>sub</i>
<i>start</i>	<i>CONJ + PUNC</i>	<i>sub</i>
<i>start</i>	<i>SUBR + PUNC</i>	<i>sub</i>
<i>start</i>	<i>number</i>	<i>sub</i>
<i>start</i>	<i>PREP + N + PUNC</i>	<i>sub</i>
<i>start</i>	<i>CONJ + PREP + N + PUNC</i>	<i>sub</i>
<i>start</i>	<i>PUNC</i>	<i>sub</i>
<i>start</i>	<i>ADV + PUNC</i>	<i>sub</i>
<i>start</i>	<i>SUBR + PREP + PUNC</i>	<i>sub</i>
<i>start</i>	<i>ADJ + PUNC</i>	<i>sub</i>
<i>start</i>	<i>empty</i>	<i>v</i>
<i>start</i>	<i>N + PUNC</i>	<i>v</i>
<i>start</i>	<i>PREP + PUNC</i>	<i>v</i>
<i>start</i>	<i>SUBR + PUNC</i>	<i>v</i>
<i>start</i>	<i>SUBR + N + PUNC</i>	<i>v</i>
<i>start</i>	<i>CONJ + PUNC</i>	<i>v</i>
<i>start</i>	<i>ADV + PUNC</i>	<i>v</i>
<i>start</i>	<i>PREP + N + PUNC</i>	<i>v</i>
<i>start</i>	<i>CONJ + N + PUNC</i>	<i>v</i>
<i>start</i>	<i>CONJ + PREP + PUNC</i>	<i>v</i>
<i>start</i>	<i>CONJ + PREP + N + PUNC</i>	<i>v</i>
<i>start</i>	<i>empty</i>	<i>vs</i>
<i>start</i>	<i>SUBR + PUNC</i>	<i>vs</i>
<i>sub</i>	<i>empty</i>	<i>asp</i>
<i>sub</i>	<i>CONJ + PUNC</i>	<i>asp</i>
<i>sub</i>	<i>N + PUNC</i>	<i>asp</i>
<i>sub</i>	<i>N + PUNC</i>	<i>end</i>
<i>sub</i>	<i>PUNC</i>	<i>end</i>
<i>sub</i>	<i>empty</i>	<i>ent</i>
<i>sub</i>	<i>N + PUNC</i>	<i>ent</i>
<i>sub</i>	<i>empty</i>	<i>ex</i>
<i>sub</i>	<i>empty</i>	<i>in</i>
<i>sub</i>	<i>empty</i>	<i>senti</i>
<i>sub</i>	<i>N + PUNC</i>	<i>senti</i>

<i>sub</i>	<i>POSTP + PUNC</i>	<i>senti</i>
<i>sub</i>	<i>N + ADV + PUNC</i>	<i>senti</i>
<i>sub</i>	<i>empty</i>	<i>sub</i>
<i>sub</i>	<i>CONJ + PUNC</i>	<i>sub</i>
<i>sub</i>	<i>empty</i>	<i>v</i>
<i>sub</i>	<i>N + PUNC</i>	<i>v</i>
<i>sub</i>	<i>POSTP + PUNC</i>	<i>v</i>
<i>sub</i>	<i>N + N + PUNC</i>	<i>v</i>
<i>v</i>	<i>empty</i>	<i>asp</i>
<i>v</i>	<i>N + PUNC</i>	<i>asp</i>
<i>v</i>	<i>PREP + PUNC</i>	<i>asp</i>
<i>v</i>	<i>ADV + PUNC</i>	<i>asp</i>
<i>v</i>	<i>PUNC</i>	<i>asp</i>
<i>v</i>	<i>PR + PREM + N + PUNC</i>	<i>asp</i>
<i>v</i>	<i>empty</i>	<i>emph</i>
<i>v</i>	<i>empty</i>	<i>end</i>
<i>v</i>	<i>PUNC</i>	<i>end</i>
<i>v</i>	<i>N + PUNC</i>	<i>end</i>
<i>v</i>	<i>PR + PUNC</i>	<i>end</i>
<i>v</i>	<i>number</i>	<i>end</i>
<i>v</i>	<i>ADJ + N + PUNC</i>	<i>end</i>
<i>v</i>	<i>empty</i>	<i>ent</i>
<i>v</i>	<i>PREP + PUNC</i>	<i>ent</i>
<i>v</i>	<i>N + PUNC</i>	<i>ent</i>
<i>v</i>	<i>ADJ + PUNC</i>	<i>ent</i>
<i>v</i>	<i>empty</i>	<i>ex</i>
<i>v</i>	<i>empty</i>	<i>in</i>
<i>v</i>	<i>CONJ + PUNC</i>	<i>in</i>
<i>v</i>	<i>empty</i>	<i>neg</i>
<i>v</i>	<i>empty</i>	<i>senti</i>
<i>v</i>	<i>N + PUNC</i>	<i>senti</i>
<i>v</i>	<i>N + N + PUNC</i>	<i>senti</i>
<i>v</i>	<i>CONJ + PUNC</i>	<i>senti</i>
<i>v</i>	<i>PR + PUNC</i>	<i>senti</i>
<i>v</i>	<i>ADV + PUNC</i>	<i>senti</i>
<i>v</i>	<i>PUNC</i>	<i>senti</i>
<i>v</i>	<i>empty</i>	<i>sub</i>
<i>v</i>	<i>CONJ + PUNC</i>	<i>sub</i>
<i>v</i>	<i>empty</i>	<i>v</i>
<i>v</i>	<i>N + PUNC</i>	<i>v</i>
<i>v</i>	<i>ADJ + PUNC</i>	<i>v</i>
<i>v</i>	<i>PREP + PUNC</i>	<i>v</i>

v	<i>POSTP + PUNC</i>	v
v	<i>PR + PUNC</i>	v
v	<i>ADV + PUNC</i>	v
v	<i>ADJ + N + PUNC</i>	v
v	<i>N + N + PUNC</i>	v
v	<i>empty</i>	<i>vs</i>
<i>vs</i>	<i>PUNC</i>	<i>end</i>
<i>vs</i>	<i>empty</i>	<i>end</i>

پیوست - الف

الف-۱- مقدمه



۸- پیوست - ب

بیوست - ب

ب-۱- مقدمه

Title of Dissertation

Abstract

Write it here.



MSc/PhD Dissertation

Title

By:

Supervised by:

June 2013

Title of Dissertation

Title of Dissertation
