تعیین جهت‌گیری نظرات در رسانه‌های اجتماعی فارسی زبان

حسین اکبریان[1]، مصطفی صالحی[2]، هادی ویسی[3]
[1]دانشگاه تهران، hosseinakbarian@ut.ac.ir
[2] دانشگاه تهران، mostafa_salehi@ut.ac.ir
[3] دانشگاه تهران، h.veisi@ut.ac.ir

چکیده - با افزایش کاربران اینترنت و ظهور وب ۲,۰، نظرات نوشته شده توسط کاربران به یکی از مهم‌ترین قابلیت‌های ارایه شده در رسانه‌های اجتماعی تبدیل شده است. برای مثال، این نظرات راهنمای مناسبی برای سایر کاربران در فروشگاه‌های اینترنتی هستند. با این وجود، حجم انبوه نظرات می‌تواند سبب گمراهی کاربران و خرید محصول نامناسب شود. از این رو در سال‌های اخیر حوزه‌ی نظرکاوی، به منظور تحلیل نظرات، مورد توجه بسیاری از پژوهشگران قرار گرفته است. تعیین جهت‌گیری نظرات، به عنوان یکی از زیرشاخه‌های این حوزه، به استخراج مثبت یا منفی-بودن نظرات کاربران نسبت به یک موضوع یا محصول گفته می‌شود. با وجود پژوهش‌های مختلفی که در زبان انگلیسی انجام شده است، روش‌های تعیین جهت‌گیری نظرات به دلیل ساختار متفاوت زبان فارسی با چالش‌های بسیاری روبرو هستند. از این‌رو در این مقاله چارچوب جدیدی برای استخراج ویژگی و دسته‌بندی نظرات با استفاده از موقعیت مکانی جملات اول و آخر، که تاثیر زیادی در جهت‌گیری نظرات دارد، پیشنهاد شده است. نتایج آزمایش‌های ما بر روی نظرات کاربران در یک فروشگاه اینترنتی نشان‌دهنده بهبود دقت روش‌های موجود است.

کلید واژه- نظرکاوی، تعیین جهت‌گیری نظرات، یادگیری ماشین، استخراج ویژگی، رسانه‌ی اجتماعی

۱- مقدمه

با توسعه‌ی رسانه‌های اجتماعی، امروزه محتوای زیادی توسط کاربران در این رسانه‌ها تولید می‌شود. نظرات کاربران در بسیاری از رسانه‌های اجتماعی، همچون رسانه‌های خبری و شبکه‌های اجتماعی، می‌توانند باعث به راه افتادن جریان‌های مختلف سیاسی و اجتماعی در مورد پدیده‌های مختلف شوند. در فروشگاه‌های اینترنتی، به عنوان یکی از انواع رسانه‌های اجتماعی، نظرات کاربران به مهم‌ترین راه‌حل برای افزایش اعتماد بین کاربران و فروشگاه‌ها تبدیل شده و علاه بر این راهنمای ارزشمندی برای بسیاری از کاربران می‌باشند.

با توجه به اهمیت تحلیل نظرات کاربران و بطورکلی محتوای ایجاد شده توسط آنها، در سال‌های اخیر حوزه‌ی نظرکاوی مورد توجه پژوهشگران قرار گرفته است. یکی از زیرشاخه‌های این حوزه، تعیین جهت‌گیری نظرات (موافق یا مخالف‌بودن، مثبت یا منفی بودن) برای تعیین علایق کاربران درباره‌ی یک موضوع می‌باشد. بطور مشخص، جهت‌گیری نظرات اهمیت بسزایی در خرید محصول توسط کاربران در یک فروشگاه اینترتی دارد. به عنوان مثال پژوهش [2] نشان می‌دهد، ۶۰ درصد کاربران بعد از خواندن نظرات سایر کاربران کالایی را خریداری می‌کنند، که با برندی که از ابتدا مدنظر آن‌ها بوده متفاوت است. همچنین پژوهش [17] نشان می‌دهد، نظرات منفی تاثیر بیشتری بر روی تصمیم نهایی کاربران در این فروشگاه‌ها دارد.

پژوهش‌های مرتبط با تعیین جهت‌گیری نظرات در زبان انگلیسی دارای پیشینه‌ی طولانی می‌باشند [3,4,5]. رویکردهای اولیه این پژوهش‌ها، بکارگیری روش‌های متن‌کاوی بوده و در سال‌های اخیر بر ترکیب استفاده از متن‌کاوی و روش‌های یادگیری ماشین متمرکز شده‌اند [18,19]. الگوریتم‌های یادگیری ماشین مورد استفاده در این پژوهش‌ها اغلب الگوریتم‌های نظارتی مانند ماشین بردار پشتیبان (SVM)[1]، تخصیص پنهان دیریکله[2]، بیز ساده[3] می‌باشد.

قابل ذکر است که با توجه به ساختار خاص متن نظرات در رسانه‌های اجتماعی، مانند استفاده از کلمات عامیانه، اصطلاحات دارای ایهام و عدم رعایت قوانین نگارشی، چالش‌های بسیاری برای تعیین جهت‌گیری نظرات در این رسانه‌ها وجود دارد. این در حالی است که در زبان فارسی این چالش‌ها شکل جدی‌تری به خود گرفته‌اند. ساختار متفاوت زبان فارسی در مقایسه با زبان انگلیسی سبب مشکلات متعددی شده است. وجود کلمات دو بخشی و کلمات دارای پسوند و پیشوند، که در آن‌ها نیم‌فاصله رعایت نمی‌شود، از جمله مشکلات این حوزه می‌باشد. با این حال، پژوهش‌های کمی در حوزه‌ی تعیین جهت‌گیری نظرات در زبان فارسی انجام گرفته است که دقت آنها نسبت به پژوهش‌های مشابه برای زبان انگلیسی کمتر است [11,12].

در این مقاله برای بهبود دقت پژوهش‌های پیشین ویژگی جدیدی برای استخراج از نظرات پیشنهاد نموده‌ایم. با توجه به طولانی بودن بسیاری از نظرات، استفاده از ویژگی‌های مبتنی بر کلمات باعث کاهش دقت پژوهش‌های این حوزه می‌باشد. استفاده از جملات اول و آخر نظرات می‌تواند سبب بهبود دقت تعیین جهت‌گیری نظرات شود. از این رو برای استفاده از این ویژگی جملات اول و آخر نظرات را استخراج نموده و با استفاده از الگوریتم یادگیری ماشین تحلیل تفکیک خطی (LDA) [۴] تعیین جهت‌گیری شده و سپس به عنوان ویژگی برای نظرات مورد استفاده قرار گرفته است. نتایج آزمایش‌های ما بر روی نظرات جمع‌آوری شده از فروشگاه اینترنتی دیجی‌کالا [24] نشان می‌دهد که استفاده از این ویژگی برای رسانه‌های اجتماعی‌ای که در آن‌ها نظرات دارای ساختار طولانی می‌باشند بسیار موثر بوده و باعث بهبود دقت در تعیین جهت‌گیری نظرات می‌شود.

۲- پژوهش‌های مرتبط

پژوهش‌های مرتبط با تعیین جهت‌گیری نظرات در رسانه‌های اجتماعی در حوزه‌های مختلفی انجام گرفته است. به عنوان مثال پژوهش [5] جملات طعنه‌آمیز و طنز را شناسایی می‌کند. میزان یکسان بودن معنای واقعی جملات با معانی‌ای که خواننده برداشت می‌کند را مشخص نموده و مدلی برای تشخیص جملات طعنه‌آمیز ارائه می‌دهد. پژوهش [6] عبارات مهم در نظرات را شناسایی کرده و از این رویکرد برای تعیین جهت‌گیری نظرات استفاده نمود است. در این پژوهش تمرکز برروی ساختار معنایی نظرات می‌باشد. پژوهش [7] بر روی تعیین جهت‌گیری نظرات بعد- محور[۵] متمرکز شده است. در این حوزه ابتدا موجودیت‌های داخل متن نظر و ابعاد آن مشخص می‌شود. سپس جهت‌گیری نظر نسبت به این ابعاد تعیین می‌شود. به عنوان مثال در مورد موجودیت "تلفن همراه"، ابعاد آن شامل "باطری" و "صفحه‌ی نمایش" می‌باشند.

در زمینه‌ی سایر زبان‌ها (به جز زبان انگلیسی)، در سال‌های اخیر پژوهش‌های بسیاری برای زبان‌های اسپانیایی، آلمانی، هندی، عربی انجام گرفته است. به عنوان مثال، در پژوهش [8] گزارش جامعی از تاثیر ویژگی‌های مختلف مورد استفاده در زبان اسپانیایی ارائه می‌شود. این ویژگی‌ها شامل ویژگی‌های مرتبط با فراوانی واژه- معکوس فراوانی سند (TFIDF)[۶]، الگوهای نحوی، ویژگی‌های معنایی، ویژگی‌های مرتبط با محل قرار گیری جملات در متن، می‌باشند. پژوهش [9] در زبان آلمانی و با استفاده از ویژگی‌های لغت‌نامه صورت گرفته است. این پژوهش به نقش کلمات (اسم، فعل، صفت، قید) در تعیین جهت‌گیری نظرات پرداخته است. همچنین با استفاده از یک لغت‌نامه جهت-گیری نظرات را مشخص نموده است. در پژوهش [10] برای کاهش مشکلات و تفاوت‌های زبانی هندی و انگلیسی، ابتدا نظرات را به زبان انگلیسی ترجمه نموده سپس جهت‌گیری آن را تعیین می‌نماید.

قابل ذکر است که پژوهش‌های اندکی در زمینه‌ی تعیین جهت‌گیری نظرات در زبان فارسی انجام گرفته است. مقاله [11] با استفاده از روش یادگیری ماشین تخصیص پنهان دیریکله و با تاکید بر ویژگی‌های لغت‌نامه و با استفاده از ترجمه‌ی خودکار عبارات انگلیسی، به تعیین جهت‌گیری آن‌ها پرداخته است. این پژوهش دارای دقت حدوداً ۷۷٪ می‌باشد. در پژوهش [12] با استفاده از ویژگی‌های n-گرام و با این فرضیه که کلماتی مانند صفت‌ها حداکثر دو بخشی هستند (مانند آنتن‌دهی)، با استفاده از دوتایی[۷] به تعیین جهت‌گیری نظرات پرداخته است. دقت گزارش شده در این پژوهش حدود ۸۴٫۷٪ می‌باشد.

در این مقاله ما علاوه بر ویژگی‌های مرتبط با لغت‌نامه و فراوانی کلمات، برای اولین بار از ویژگی جملات اول و آخر نظرات برای بهبود دقت در زبان فارسی بهره برده‌ایم.

۳- چارچوب پیشنهادی

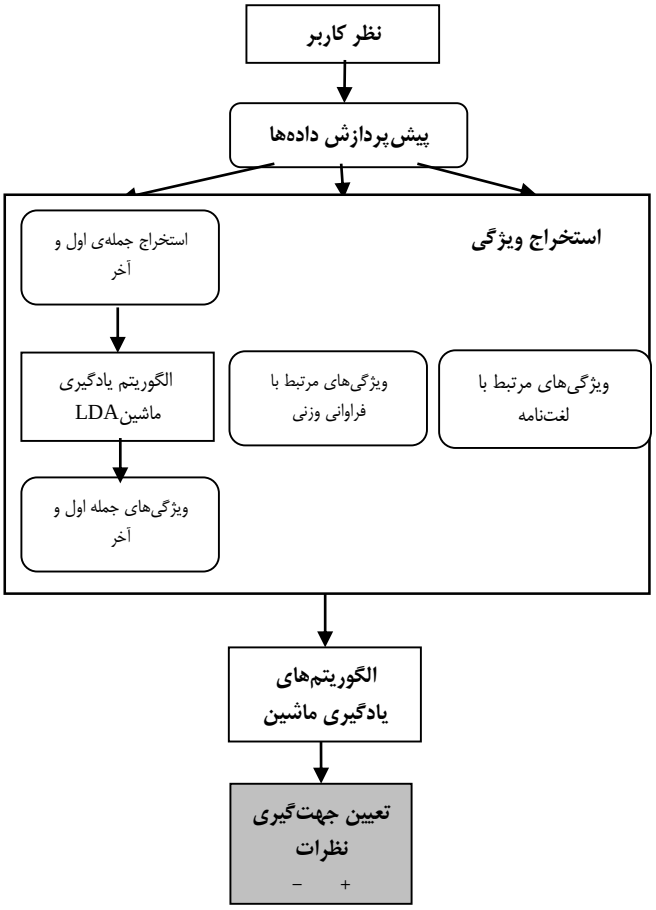
به طور کلی، تعیین جهت‌گیری نظرات شامل مراحل پیش‌پردازش، استخراج ویژگی و استفاده از الگوریتم‌های یادگیری ماشین می‌باشد. شکل (۱) چارچوب پیشنهادی ما را که بر این اساس می‌باشد، نشان می‌دهد. در ادامه به جزئیات هر یک از این مراحل پرداخته می‌شود.

۳-۱- پیش‌پردازش نظرات

این مرحله شامل نرمال‌سازی[۸] (شامل حذف و اضافه کردن فاصله و نیم‌فاصله و علامت‌ها و نشانه‌های نامربوط)، ریشه‌یابی[۹] (تعیین ریشه‌ی افعال برای حذف پسوندهای افعال) و برچسب پاره-سخن[۱۰] (تعیین موقعیت کلمات در جمله) می‌شود. همچنین در این مرحله داده‌هایی که اطلاعات آن‌ها ناقص یا نامربوط هستند از مجموعه‌ی دادگان مورد مطالعه حذف می‌شود. یکی از مهم‌ترین عملیات پیش‌پردازش برروی داده‌ها حذف جملات نامربوط در اول و آخر نظرات می‌باشد. به عنوان مثال "سلام به

همه‌ی دوستان" و یا " با تشکر از سایت خوبتون" این جملات سبب ایجاد نویز در نتایج می‌شود و حذف آن‌ها تاثیر بسزایی در بهبود کارآیی الگوریتم دارد.

برای این مرحله از بسته‌ی پیش‌پردازش هضم [25] استفاده نموده‌ایم. این بسته شامل نرمال‌ساز، ریشه‌یاب، برچسب‌زن پاره‌سخن می‌باشد.



شکل ۱ - مراحل چارچوب پیشنهادی

۳-۲- ویژگی‌های مورد استفاده

ویژگی‌های لغت‌نامه: برای تعیین جهت‌گیری نظرات، پرکاربردترین راه استفاده از لغت‌نامه می‌باشد. با این وجود همه‌ی کلمات داخل یک جمله تاثیر یکسانی بر روی جهت‌گیری نظرات ندارند. بنابراین پس از تعیین برچسب پاره‌سخن و تعیین نقش کلمات در جمله با استفاده از لغت‌نامه، برای تعدادی از این کلمات امتیازی اختصاص می‌دهیم. پژوهش [9] نشان می‌دهد صفت‌ها و فعل‌ها بیشترین تاثیر را در تعیین جهت‌گیری نظرات دارند. با این وجود اسم‌ها نیز در بسیاری از اوقات دارای تاثیر بسزایی در جهت‌گیری نظرات دارند. از این‌رو ما سه گروه اسم، فعل و صفت را مورد بررسی قرار داده و با استفاده از لغت‌نامه به این گروه کلمات امتیاز مثبت یا منفی تخصیص داده‌ایم.

ویژگی‌های مربوط به فراوانی وزنی TF-IDF: برای این منظور پژوهش [16] یک رابطه برای تخصیص امتیاز به جملات ارائه نموده است. پیش از توضیح رابطه‌ی مورد نظر به تشریح متغیرهای آن می‌پردازیم.

- $C_{t,d}$ تعداد دفعات تکرار کلمه‌ی t در نظر d می‌باشد.
- P_t تعداد نظراتی که واژه‌ی t در آن‌ها تکرار شده و دارای برچسب مثبت هستند.
- $|P|$ تعداد نظراتی که در داده‌ی آموزش برچسب مثبت دارند.
- N_t تعداد نظراتی که واژه‌ی t در آن‌ها تکرار شده و دارای برچسب منفی هستند.
- $V_{t,d}$ امتیاز ویژگی فراوانی وزنی است.

$$v_{t,d} = c_{t,d} \times \log_2\left(\frac{|N|}{N_t}\right) - c_{t,d} \times \log_2\left(\frac{P_t}{|P|}\right)$$
$$= c_{t,d} \times \log_2\left(\frac{|P_t|}{|N_t|}\right) \qquad (۱)$$

در رابطه (۱) با توجه به اینکه تعداد نظرات مثبت و منفی در مجموعه‌ی داده برابر می‌باشد، می‌توان دو متغیر |N| و |P| را از دو عبارت حذف کرد. این رابطه در پژوهش [16] برای تمام کلمات موجود در جمله مورد استفاده قرار می‌گیرد. این در حالی است که با این‌کار، تاثیر ویژگی‌های لغت‌نامه کاهش می‌یابد. از این‌رو ما این امتیاز را تنها برای صفت‌ها و فعل‌ها که تاثیر بیشتری در تعیین جهت‌گیری نظرات دارند، استفاده می‌نماییم. با توجه به این که هر جمله ممکن است دارای بیش از یک صفت یا فعل باشد رابطه‌ی (۱) را به شکل زیر تغییر داده‌ایم.

$$v_{t,d} = \frac{1}{N} \sum_{I=1}^{K} c_{t,d} * \log_2\left(\frac{P_t}{N_t}\right) \qquad (۲)$$

که در این رابطه، K برای صفت‌ها، نشان‌دهنده‌ی صفت‌های یک نظر و برای افعال، نشان‌دهنده‌ی افعال یک نظر می‌باشند. همچنین N تعداد صفت‌ها یا افعال یک نظر می‌باشد. در واقع میانگین فراوانی وزنی صفت‌ها و افعال یک نظر با استفاده از رابطه‌ی (۲) محاسبه می‌شود.

ویژگی‌های مربوط به محل قرار گرفتن جملات: پژوهش [14] نشان می‌دهد که محل قرار گرفتن جملات در یک نظر تاثیر بسزایی بر نقش آن در تعیین جهت‌گیری نظر دارد. به عنوان مثال می‌توان به این نظر اشاره کرد: "این تلفن همراه بسیار مناسب است، ولی مشکلاتی از جمله عمر کم باطری و جنس نامناسب بدنه دارد (" که در واقع به این معنا است: "با وجود اینکه این تلفن همراه مشکلاتی از جمله جنس نامناسب بدنه و عمر کم باطری دارد. گوشی مناسبی است.").

به منظور لحاظ کردن این ویژگی، در این مقاله ما روش جدیدی برای تعیین جهت‌گیری جمله‌ی اول و آخر پیشنهاد نمودیم. به این صورت که ابتدا جمله‌ی اول و آخر نظر را استخراج نموده و با استفاده از ویژگی‌های لغت‌نامه و الگوریتم یادگیری ماشین LDA، برای این دو جمله تعیین جهت‌گیری نموده و مقدار حاصل را به عنوان یک ویژگی در بردار ویژگی هر نظر استفاده نموده‌ایم. برای داده‌ی آموزش الگوریتم LDA برای جملات اول و آخر، ابتدا بخشی از نظرات را انتخاب نموده‌ایم. سپس ویژگی‌های لغت‌نامه‌ی آن را استخراج کرده و به عنوان داده‌ی آموزش برای تعیین جهت‌گیری جملات اول و آخر به کار گرفته‌ایم.

۴- ارزیابی کارایی چارچوب پیشنهادی

۴-۱- مجموعه‌ی داده

مجموعه داده مورد استفاد در این پژوهش شامل تعداد ۲۰۰۰ نظر درباره‌ی محصولات سایت دیجی‌کالا[11] می‌باشد. این نظرات شامل ۱۰۰۰ نظر مثبت و ۱۰۰۰ نظر منفی می‌باشد. که از این تعداد، ۹۰۰ نظر از مجموعه‌ی جمع‌آوری شده توسط نویسندگان مقاله‌ی [۱] و مابقی طی این پژوهش، از سایت مذکور جمع‌آوری و بعد از برچسپ‌زنی دستی به دادگان افزوده شده است.

۴-۲- نتایج ارزیابی‌ها

پس از پیش‌پردازش و استخراج ویژگی، ویژگی‌های استخراج شده از متن نظر را به الگوریتم‌های یادگیری ماشین نظارتی می‌دهیم. این الگوریتم‌ها نظرات را برچسب زده و خروجی مثبت یا منفی بودن نظر را نشان می‌دهد. برای این منظور، الگوریتم‌های ماشین بردار پشتیبان (SVM) ، بیز ساده، تحلیل تفکیک خطی (LDA)، بیز ساده و شبکه‌ی عصبی، مورد استفاده قرار گرفته‌اند. در الگوریتم‌های یادگیری مذکور از ۸۰ درصد داده‌ها برای آموزش و از ۲۰ درصد آنها به عنوان داده‌ی آزمون استفاده شده است. همچنین ۵۰ بار به صورت تصادفی داده‌ی آموزش و آزمون از مجموعه داده استخراج شده و میانگین نتایج استخراج شده است. در روش شبکه‌ی عصبی تعداد لایه‌های مخفی ۱ و تعداد نورون‌های این لایه ۱۰ در نظر گرفته شده است. در روش ماشین بردار پشتیبان، مقدار هزینه 2^{12} و گاما[13] ۰,۰۰۰۷ در نظر گرفته شده است. همچنین برای پیاده سازی ماشین بردار پشتیبان از کتابخانه‌ی LibSVM [20] استفاده شده است.

جدول (۱)، نتایج قبل و بعد از بکارگیری ویژگی‌های پیشنهادی (موقعیت مکانی جملات) را، بر اساس معیارهای دقت[14]، صحت[15]، فراخوانی[16]، و امتیاز F[17]، نشان می‌دهد.

همان‌طور که در جدول (۱) دیده می‌شود، کارایی روش بیز ساده از سایر روش‌ها پایین‌تر است و سه روش ماشین بردار پشتیبان، و شبکه عصبی و LDA دارای کارایی تقریباً مشابهی هستند. این موضوع هم قبل از به‌کارگیری ویژگی‌های پیشنهادی و پس از آن وجود دارد، هرچند که رتبه روش‌های برتر قبل و بعد از اعمال ویژگی‌های پیشنهادی تغییر کرده است.

آنچه به صورت سازگار قابل مشاهده است، این است که پس از بکارگیری ویژگی پیشنهادی (جملات اول و آخر)، نتایج در همه الگوریتم‌ها و با همه معیارها بهبود یافته است که بیانگر موثر بودن این ویژگی‌هاست. بعد از اعمال ویژگی‌های پیشنهادی، روش ماشین بردار پشتیبان نتایج بهتری را نسبت به سایر روش‌های یادگیری ماشین می‌دهد. پس از این روش، شبکه‌ی عصبی و سپس LDA دارای کارایی بالاتر می‌باشد. همچنین پیشرفت محسوسی در میانگین معیار F دو کلاس مثبت و منفی نسبت به سایر پژوهش‌های فارسی مشاهده می‌شود [11,12]. این نتایج فرض اولیه ما در مورد ویژگی‌های پیشنهادی را تایید می‌کند. در این فرض بیان می‌شود، که بیشتر کاربران در ارائه نظرات خود، در جمله اول بخش عمده‌ای از نظر و ایده خود در مورد محصول را اظهار می‌کنند. همچنین در جمله پایانی، جمع‌بندی نظر خود را اعلام می‌کنند.

پیش از اضافه نمودن ویژگی پیشنهادی بهترین نتایج به دلیل یک جنس بودن ویژگی‌ها (همه‌ی ویژگی‌ها از جنس لغت است) با استفاده از الگوریتم شبکه‌ی عصبی حاصل شده است. اما پس از اضافه کردن ویژگی پیشنهادی به دلیل افزایش فضای ویژگی نتایج الگوریتم SVM به دلیل استفاده از مرز تصمیم در فضای چندبعدی بهتر از سایر روش‌ها بوده است.

جدول (۱) : نتایج ارزیابی تعیین جهت‌گیری نظرات کاربران پیش و پس از بکارگیری ویژگی‌های پیشنهادی

الگوریتم یادگیری ماشین	جهت‌گیری نظر	نتایج بدست آمده بدون ویژگی پیشنهادی				نتایج بدست آمده با استفاده از ویژگی پیشنهادی (موقعیت مکانی جملات)			
		دقت (درصد)	صحت (درصد)	فراخوانی (درصد)	معیار F (درصد)	دقت (درصد)	صحت (درصد)	فراخوانی (درصد)	معیار F (درصد)
SVM	منفی	٪۸۴,۱۹	۸۳,۰۷	۸۲,۷۸	۸۲,۹۲	٪۸۶,۰۴	۸۵,۶۸	۸۶,۴۹	۸۶,۰۸
	مثبت		۸۵,۲۰	۸۴,۶۲	۸۴,۹۱		۸۶,۴۱	۸۵,۹۶	۸۶,۱۹
شبکه‌ی عصبی	منفی	٪۸۴,۹۹	۸۴,۶۷	۸۴,۵۲	۸۴,۶۰	٪۸۵,۴۸	۸۴,۹۰	۸۶,۰۵	۸۵,۴۷
	مثبت		۸۵,۳۹	۸۴,۲۸	۸۴,۳۳		۸۶,۰۳	۸۴,۹۶	۸۵,۴۹
LDA	منفی	٪۸۴,۱۵	۸۵,۹۱	۸۵,۹۱	۸۴,۳۴	٪۸۴,۷۱	۸۶,۰۱	۸۳,۸۰	۸۴,۸۹
	مثبت		۸۲,۴۰	۸۵,۵۷	۸۳,۹۶		۸۳,۴۱	۸۵,۷۰	۸۴,۵۴
بیز ساده	منفی	٪۶۷,۱۶	۵۷,۵۵	۷۱,۴۷	۶۳,۷۶	٪۷۰,۹۸	۶۵,۱۵	۷۳,۱۷	۶۸,۹۳
	مثبت		۷۶,۸۲	۶۴,۳۰	۷۰,۰۱		۷۶,۶۷	۶۹,۲۵	۷۲,۷۷

۵- نتیجه‌گیری و پژوهش‌های آتی

در این پژوهش تلاش کردیم به منظور حل مساله تعیین جهت-گیری نظرات فارسی، علاوه‌بر بهبود کارایی مرحله‌ی پیش‌پردازش داده‌ها، در مراحل استخراج ویژگی نیز ایده‌های جدیدی اعمال کنیم. از این‌رو ویژگی جدیدی که مربوط به جایگاه قرارگیری جملات در نظر کاربران می‌باشد را با استفاده از الگوریتم یادگیری LDA پیشنهاد نمودیم. همچنین با استفاده از الگوریتم‌های یادگیری ماشین به دسته‌بندی نظرات پرداخته‌ایم. نتایج حاصل، بهبود در دقت تعیین جهت‌گیری را نشان می‌دهد. برای تحقیقات آتی در زبان فارسی، استفاده از سایر ویژگی‌های مربوط به نظرات مانند ویژگی‌های الگوی نحوی [21]، ویژگی‌های معنایی [22] و تشخیص ابعاد موجودیت‌های نظرات و تعیین جهت‌گیری نظرات نسبت به این ابعاد [23] پیشنهاد می‌شود.

مراجع

[۱] پدرام حسینی و دیگران. (۱۳۹۳). پیکره‌ی فارسی تحلیل احساس سِنتی پِرس. ارائه‌شده در سومین همایش ملی زبان‌شناسی رایانشی، تهران: دانشگاه صنعتی شریف

[2] Williams, Bardford. "Buy it, Try it, Rate It: Study of Consumer Electronics Purchase Decisions in the Engagement Era." KRC Research, available at www.krcresearch.Com, 2012.

[3] Kouloumpis, Efthymios, Wilson, Theresa, and Moore, Johanna. Twitter sentiment analysis: The good the bad and the omg! Proceedings of the ICWSM, 11, 538–541, 2011.

[4] Minqing Hu and Bing Liu. Mining and summarizing customer reviews. Proceedings of the ACM SIGKDD international conference on Knowledge discovery and data mining – KDD, 2004.

[5] Kennedy, Alistair, and Diana Inkpen."Sentiment classification of movie reviews using contextual valence shifters." Computational intelligence, 22(2), 110-125, 2006.

[6] Molina-González, M. Dolores, et al. "Semantic orientation for polarity classification in Spanish reviews." Expert Systems with Applications, 40(18), 7250-7257, 2013.

[7] Singh, V. K., et al. "Sentiment analysis of movie reviews: A new feature-based heuristic for aspect-level sentiment classification" Proceedings of the International Multi-Conference on. IEEE, 2013.

[8] Chenlo, Jose M., and David E. Losada. "An empirical study of sentence features for subjectivity and polarity classification." Information Sciences, 280, 275-288, 2014.

[9] Remus, Robert, Uwe Quasthoff, and Gerhard Heyer. "SentiWS-A Publicly Available German-language Resource for Sentiment Analysis." Proceedings of the LREC, 2010.

[10] Joshi, Aditya, A. R. Balamurali, and Pushpak Bhattacharyya. "A fall-back strategy for sentiment analysis in hindi: a case study." Proceedings of the 8th ICON, 2010.

[11] Shams, Mohammadreza, Azadeh Shakery, and Heshaam Faili. "A non-parametric LDA-based induction method for sentiment analysis." Proceedings of the conference Artificial Intelligence and Signal Processing (AISP), 2012 16th CSI International Symposium on. IEEE, 2012.

[12] Bagheri, Ayoub, and Mohamad Saraee. "Persian Sentiment Analyzer: A Framework based on a

[1] SVM (Support Vector Machine)

[2] Latent Dirichlet Allocation

[3] Naive Bayes

[4] Linear discriminant analysis (LDA)

[5] Aspect-Based

[6] Term Frequency – Inverse Document Frequency (TF-IDF)

[7] Bigram

[8] Normalization

[9] Stemming

[10] POS Tagging

[11] DigiKala.com

[12] Cost

[13] Gamma

[14] Accuracy

[15] Precision

[16] Recall

[17] F-Measure

Novel Feature Selection Method." Procedia Computer Science, 50, 16-23, 2015.

[13] Turney, Peter D. "Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews." Proceedings of the 40th annual meeting on association for computational linguistics. Association for Computational Linguistics, 2002.

[14] Beineke, Philip, et al. "Exploring sentiment summarization." Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications, 2004.

[15] http://www.sobhe.ir/ 2015.

[16] Martineau, Justin, and Tim Finin. "Delta TFIDF: An Improved Feature Space for Sentiment Analysis." Proceedings of the ICWSM, 2009.

[17] Berger, Jonah, Alan T. Sorensen, and Scott J. Rasmussen. "Positive effects of negative publicity: When negative reviews increase sales." Marketing Science, 29(5), 815-827, 2010.

[18] Bogle, S. A., and W. D. Potter. "SentAMaL-A Sentiment Analysis Machine Learning Stock Predictive Model." Proceedings on the International Conference on Artificial Intelligence (ICAI), 2015.

[19] Montejo-Ráez, A., Martínez-Cámara, E., Martín-Valdivia, M. T., & Ureña-López, L. A. "Ranked wordnet graph for sentiment polarity classification in twitter". Computer Speech & Language, 28(1), 93-107, 2014.

[20] Chang, Chih-Chung, and Chih-Jen Lin. "LIBSVM: A library for support vector machines." ACM Transactions on Intelligent Systems and Technology (TIST), 2(3), no 27, 2011.

[21] Liu, Bing. "Sentiment Analysis and Opinion Mining (Synthesis Lectures on Human Language Technologies)." Morgan & Claypool Publishers, 2012.

[22] Singh, V. K., Piryani, R., Uddin, A., & Waila, P. "Sentiment analysis of movie reviews: A new feature-based heuristic for aspect-level sentiment classification". Proceedings of the International Multi-Conference on IEEE, 2013.

[23] Garcia-Moya, Lisette, Henry Anaya-Sanchez, and Rafael Berlanga-Llavori. "Retrieving product features and opinions from customer reviews." IEEE Intelligent Systems, 3, 19-27 , 2013.

[24] http://www.digikala.com/, 2015

[25] Nourian, Alireza, et al. "On the Importance of Ezafe Construction in Persian Parsing." Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics , 877–882 , 2015