ICWR2020
کنفرانس بین المللی وب پژوهی
ایران، تهران، ۲۴ خرداد ۱۳۹۹

Using Supervised Learning to Identify Opinion Spam in Persian Language

Sepideh Jamshidi_Nejad [1], Fatemeh Ahmadi Abkenari [2], Pyman Bayat [3]

[1] PhD candidate, Department of Computer Engineering and Information Technology, Islamic Azad University of Rasht, Rasht, Iran
jamshidi1270@phd.iaurasht.ac.ir

[2] Faculty, Department of Computer Engineering and Information Technology,Payame Noor University of Rasht, Rasht, Iran
Fateme.Abkenari@gilan.pnu.ac.ir

[3] Faculty, Department of Computer Engineering and Information Technology, Islamic Azad University of Rasht, Rasht, Iran
bayat@iaurasht.ac.ir

Abstract

Due to the increasing use of users' opinions in various domains on social networks and the importance of these opinions, their accuracy is very important, but unknown persons may use fake comments to promote or discredit products, services, Organizations or peoples. Since it is difficult and even impossible to identify only on through read, find data will be difficult to design and evaluate the algorithms for the identification of opinion spam too. Due to the challenge explained, the present paper, by innovating in the combination of opinion content, metadata and entity information, generates a set of data features and for the first time at the document and sentence level, recognizes opinion spam in Persian. In the following, the identification of opinion spam as a classification problem is introduced with two fake and non-fake categories and is modeled with six supervised learning methods. To evaluate the results, the Confusion matrix of each method is constructed and after calculating the precision, recall and accuracy and comparing the values, the best and most accurate classification will be introduced in identifying opinion spam.

Keywords: Identify Opinion Spam, Supervised Learning, Modeling Opinion Spam, Opinion Persian.

JCWR2020
کنفرانس بین المللی وب پژوهی
ایران، تهران، ۲۲ خرداد ۱۳۹۹

استفاده از یادگیری بانظارت برای شناسایی هرزنظر در زبان فارسی

سپیده جمشیدی نژاد*[۱]، فاطمه احمدی آبکناری[۲] و پیمان بیات[۳]

[۱] دانشجوی دکتری، گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشکده فنی و مهندسی، دانشگاه آزاد اسلامی واحد رشت
jamshidi1270@phd.iaurasht.ac.ir

[۲] هیات علمی، گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشکده فنی و مهندسی، دانشگاه پیام نور رشت
Fateme.Abkenari@gilan.pnu.ac.ir

[۳] هیات علمی، گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشکده فنی و مهندسی، دانشگاه آزاد اسلامی واحد رشت
bayat@iaurasht.ac.ir

چکیده

با توجه به استفاده روزافزون از نظرات درج‌شده کاربران در حوزه‌های مختلف در شبکه‌های اجتماعی و ارزشمند بودن این نظرات، صحت آنها بسیار مهم است اما افراد ناشناس بیان‌کننده نظر ممکن است با اهداف مخرب، نظرات جعلی و هرز را برای ترویج یا بی‌اعتبارکردن محصولات، خدمات، سازمان‌ها یا افراد، بیان نمایند. از آنجا که شناسایی هرزنظر تنها با خواندن، دشوار و حتی غیرممکن است یافتن داده‌هایی برای طراحی و ارزیابی الگوریتم‌های شناسایی هرزنظر نیز دشوار خواهد بود. با توجه به چالش مطرح‌شده، مقاله حاضر با نوآوری در ترکیب محتوای نظر، فراداده و اطلاعات موجودیت، مجموعه‌ای از ویژگی‌های داده‌ای را تولید می‌کند و برای اولین بار در سطح سند و جمله، هرزنظر را در زبان فارسی تشخیص می‌دهد. سپس شناسایی هرزنظر به عنوان یک مساله دسته‌بندی، با دو دسته جعلی و غیرجعلی معرفی و با شش روش یادگیری بانظارت، مدلسازی می‌شود. برای ارزیابی نتایج، ضمن محاسبه پارامترهای دقت، فراخوانی و صحت، ماتریس آشفتگی شش روش مدلسازی نیز ساخته شد و با مقایسه پارامترها، دسته‌بند جنگل تصادفی با ۹۸.۶۵٪، ۹۷.۲۷٪ و ۹۹.۰۹٪ به ترتیب برای دقت، فراخوانی و صحت، به عنوان بهترین و دقیق‌ترین دسته‌بند در شناسایی هرزنظر معرفی شد.

کلمات کلیدی

شناسایی هرزنظر، یادگیری بانظارت، مدلسازی هرزنظر، نظرات فارسی.

۱- مقدمه

افراد و سازمان‌ها بطور روز افزون در حال استفاده از نظرات درج شده در رسانه‌های اجتماعی، برای اتخاذ تصمیمات خرید، رای دادن در انتخابات، بازاریابی و طراحی محصول هستند. نظرات مثبت اغلب به معنای سود و شهرت برای کسب و کارها و افراد است و متاسفانه مشوقی برای ارسال نظرات یا مرورهای جعلی به منظور ترویج یا تخریب بعضی از محصولات، سرویس‌ها، سازمان‌ها، افراد و حتی ایده‌ها است بدون اینکه نیت واقعی و هویت فرد یا سازمان، افشا گردد. این افراد بیان‌کنندگان هرزنظر[۱] نامیده می‌شوند و فعالیت‌های آنها نیز بیان هرزنظر[۲] است [۱و۲]. به دلیل افزایش استفاده از نظرات در رسانه‌های اجتماعی، بیان هرزنظر، بیش از قبل شیوع یافته و پیچیده شده است و این خود چالشی عمده برای شناسایی آنها به شمار می‌رود. با این وجود باید آنها را تشخیص داد تا اطمینان حاصل شود که رسانه‌های اجتماعی، به جای اینکه مملو از هرزنظر[۳] شود یک منبع قابل اعتماد از نظرات عمومی باشد.

به طور کلی شناسایی هرزنظر در زمینه‌های متعددی مانند هرزوب[۴] و هرزنامه[۵] مورد مطالعه قرار گرفته است. با این وجود هرزنظر بسیار متفاوت است زیرا چالش اصلی در آن، این است که شناسایی، تنها با خواندن، بسیار دشوار و حتی ناممکن است و این امر یافتن داده‌های هرزنظر را برای ساخت الگوریتم‌های تشخیصی دشوار می‌سازد. در بعضی مواقع به عنوان مثال می‌توان یک مرور واقعی از یک رستوران خوب را نوشت و آن را به صورت یک مرور جعلی برای یک رستوران با کیفیت بد و با هدف ترویج آن، پست کرد. بدون توجه به اطلاعات فراداده از متن مرور، هیچ راهی برای شناسایی

این مرور جعلی وجود ندارد زیرا یک مرور نمی‌تواند همزمان هم واقعی و هم جعلی باشد [۳].

این مقاله با استفاده از روش یادگیری بانظارت و با نوآوری در ساخت و استفاده از مجموعه ویژگی‌های داده‌ای، شناسایی هرزنظر را برای اولین بار در سطح سند و جمله، در زبان فارسی به عنوان یک مساله دسته‌بندی، با دو دسته جعلی و غیرجعلی مدلسازی می‌کند. ابتدا مجموعه داده متنی فارسی اولیه با مجموع ۱۱۰۰ نظر با طول متفاوت در حوزه هتلداری، جمع‌آوری و پیش‌پردازش شدند. مجموعه ویژگی‌ها با ده ویژگی، برای شناسایی هرزنظر ساخته شد و مجموعه داده اولیه به روش دستی و با استفاده از این مجموعه ویژگی‌ها برچسب‌گذاری شدند. سپس شش روش دسته‌بندی شامل درخت تصمیم[۶]، جنگل تصادفی[۷]، قوانین استقراء[۸]، ماشین بردار پشتیبان[۹]، شبکه عصبی[۱۰] با یک و دو لایه و K–نزدیک‌ترین همسایه[۱۱] با دو مقدار مختلف برای K، بر مجموعه داده متنی اعمال شدند. برای آموزش و آزمایش توزیع ۷۰/۳۰ منظور گردید و بعد از رسم ماتریس آشفتگی[۱۲]، نتایج خروجی دسته‌بندها با یکدیگر مقایسه شد تا بهترین دسته‌بند از لحاظ دقت[۱۳]، فراخوانی[۱۴] و صحت[۱۵] در شناسایی هرزنظر معرفی شود. ادامه مقاله به صورت زیر سازماندهی شده است: بخش ۲، مروری بر پژوهش‌های پیشین، بخش ۳، روش پیشنهادی، بخش ۴، نتیجه‌گیری و در انتها مراجع آمده است.

۲- مروری بر پژوهش‌های پیشین

جیندال[۱۶] و لیو[۱۷] (۲۰۰۸) در پژوهش خود از مرورهای تکراری استفاده کردند. آنها با مطالعه ۵/۸ میلیون مرور و ۲/۱۴ میلیون نویسنده مرور در سایتamazon.com ، تعداد زیادی مرورهای تکراری و تقریبا تکراری یافتند که نشان داد هرزنظر بسیار گسترده است. از مرورهای تکراری با چهار دسته مختلف شامل ۱- مرورهای تکراری از شناسه کاربری مشابه برای محصول مشابه، ۲- مرورهای تکراری از شناسه‌های کاربری مختلف برای محصول مشابه، ۳- مرورهای تکراری از شناسه کاربری مشابه برای محصولات مختلف، ۴- مرورهای تکراری از شناسه‌های کاربری مختلف برای محصولات مختلف، استفاده کردند. آنها بیان کردند که نوع اول می‌تواند ناشی از چند بار کلیک اشتباه مرورکنندگان بر روی دکمه ارسال مرور باشد با این وجود سه نوع دیگر احتمالا جعلی هستند. بنابراین به عنوان مرورهای جعلی بکار می‌روند و مابقی مرورها نیز به عنوان مرورهای غیرجعلی در داده‌های آموزشی برای یادگیری ماشینی استفاده شدند. علاه بر این، ویژگی‌های دیگری مانند نسبت تعداد مرورهای نوشته‌شده توسط نویسنده مرور، طول مرور، کلمات واقعی و _nگرم‌های مرور، تعداد دفعات ذکر اسامی برندها، درصد کلمات نظر نیز در این پژوهش استفاده شدند. همچنین آنها اعلام کردند که ویژگی تبلیغات و سایر متون بی‌ربط، هرزنظر نیستند زیرا نظرات کاربران را ارائه نمی‌کنند .[۲].

در پژوهش لی[۱۸] و همکاران (۲۰۱۱)، یک روش یادگیری بانظارت تلاش نمود تا مرورهای جعلی را شناسایی کند. در این مورد، مجموعه‌ای از بررسی‌های جعلی برچسب‌خورده به صورت دستی بر اساس مرورهای Epinions ایجاد شدند. در Epinions، بعد از ارسال یک مرور، کاربران می‌توانند مرور را با دادن امتیاز ارزیابی کنند. آنها همچنین می‌توانند نظراتی را درباره آن بنویسند. نویسندگان به صورت دستی یک سری مرورهای جعلی یا غیرجعلی را با خواندن مرورها و نظرات، برچسب زدند. برای یادگیری، چند نوع ویژگی پیشنهاد شد که مشابه با پژوهش [۳] بود اما با چند ویژگی دیگر مانند ویژگی‌های ذهنی و عینی، ویژگی‌های مثبت و منفی، مشخصات نویسنده مرور، امتیاز مرجع که با PageRank محاسبه شد. نویسندگان برای یادگیری، از دسته‌بند بیز ساده استفاده نمودند که نتایج امیدبخشی را به دنبال داشت. همچنین یک روش یادگیری نیمه نظارتی با این ایده که یک بیان‌کننده هرزنظر مرورهای جعلی متعددی دارد، در این مقاله آزمایش شد [۴].

در پژوهش اوت[۱۹] و همکاران (۲۰۱۱) از یادگیری بانظارت استفاده شد. در این مورد، نویسندگان از Amazon Mechanical Turk برای یافتن مرورهای جعلی بیست هتل در شیکاگو، استفاده نمودند. قوانینی تعیین شد تا از کیفیت مرورهای جعلی اطمینان حاصل شود. به عنوان مثال، آنها به هر کارمند اجازه دادند تا تنها یکبار مرور ارسال کند. کارمندان نیز باید در ایالات متحده باشند. همچنین به کارمندان این سناریو داده شد که در هتل‌ها کار می‌کنند و رئیس از آنها خواسته است تا برای ترویج هتل‌ها مرورهای جعلی بنویسند. مرورهای واقعی از وب سایت TripAdvisor بدست آمدند. نویسندگان، چند روش مانند شناسایی سبک، تشخیص نیرنگ روان‌شناختی–زبان‌شناختی و ترکیب n–گرم‌ها با روش‌های دسته‌بندی متن مانند بیز ساده و ماشین بردار پشتیبان خطی را امتحان کردند. تمام این کارها دارای بعضی ویژگی‌های پیشنهاد شده توسط پژوهشگران هستند. آزمایشات آنها نشان داد که دسته‌بندی متنی با تک_گرم و دو_گرم بر اساس توزیع ۵۰/۵۰ دسته جعلی و غیرجعلی، بهترین عملکرد را دارد. ویژگی‌های سنتی برای نیرنگ‌ها عملکرد خوبی نداشتند. با این وجود مانند مطالعات قبلی در اینجا نیز، داده‌هایی که برای ارزیابی بکار رفتند بی‌نقص نیستند. مرورهای جعلی در Amazon Mechanical Turk ممکن است واقعا مرورهای جعلی نباشند زیرا کارمندان، هتل‌ها را نمی‌شناسند اگرچه از آنها خواسته شد تا وانمود کنند که در این هتل‌ها مشغول به کار هستند. به علاوه، استفاده از توزیع ۵۰/۵۰ داده‌های جعلی/غیرجعلی برای آزمایش ممکن است توزیع واقعی در شرایط عملی را نشان ندهد. توزیع طبقات می‌تواند اثر چشمگیری بر دقت مرورهای جعلی تشخیص داده شده داشته باشد [۵].

وو[۲۰] و همکاران (۲۰۱۰) یک روش بدون نظارت را برای شناسایی مرورهای جعلی بر اساس معیار تحریف (و نه رفتارهای مرورکنندگان) پیشنهاد دادند. ایده آنها این بود که مرورهای جعلی باعث تحریف در رتبه محبوبیت کلی یک مجموعه از موجودیت‌ها می‌شود. پاک کردن یک مجموعه مرور انتخاب شده به صورت تصادفی نباید لیست رتبه‌بندی شده موجودیت‌ها را مختل سازد در حالیکه حذف مرورهای جعلی می‌تواند رتبه‌بندی موجودیت‌ها را به شدت تغییر داده یا تحریف کند تا رتبه‌بندی واقعی نشان داده شود. این

تحریف را می‌توان با مقایسه رتبه‌های محبوبیت، قبل و بعد از حذف، با استفاده از همبستگی رتبه، سنجید [۶].

وانگ[۲۱] و همکاران (۲۰۱۴) تجزیه و تحلیل ناهنجاری‌های موجود و نیز روش‌های تجزیه و تحلیل نظرات را بررسی کردند. همچنین محدودیت‌ها و چالش‌های آنها را نیز تحلیل کردند. برای مقابله با چالش‌های موجود نویسندگان، یک روش پیشرفته طبقه‌بندی نظرات، ارائه دادند که از طریق تجزیه و تحلیل نظرات در مورد داده‌های رسانه اجتماعی، امکان استفاده از این روش پیشنهادی را برای شناسایی ناهنجاری‌ها نیز مطالعه کردند. نویسندگان از طریق تجزیه و تحلیل توئیت‌ها، کاربردپذیری و قابلیت اطمینان روش پیشنهادی خود را بررسی نمودند [۷].

سان[۲۲] و همکاران (۲۰۱۸) با استفاده از توزیع گاوسی چند متغیره احساسات کاربران درباره میکروبلاگ‌ها را مدلسازی و تجزیه و تحلیل نمودند و حالت احساسات غیرعادی را شناسایی کردند. با اندازه‌گیری مقدار چگالی احتمال مشترک و اعتبارسنجی ساختار، دقت شناسایی ناهنجاری یک کاربر برابر با ۸۳.۴۹٪ و با استفاده از این روش برای ماه‌های مختلف برابر با ۸۷.۸۴٪ خواهد بود. نتایج به دست آمده به کمک آزمون توزیع، نشان داد که احساس خنثی، شادی و غم در هر کاربر، به توزیع نرمال بستگی دارد اما احساس تعجب و خشم اینطور نیست. علاوه بر این، نویسندگان بیان کردند که احساسات میکروبلاگ‌هایی که از طریق گروه‌ها منتشر می‌شوند از توزیع قانون توانی تبعیت می‌کند اما احساسات فردی اینطور نیستند. این مقاله، برای شناسایی احساسات غیرعادی در رسانه‌های اجتماعی، یک روش کمیت ارائه داد که همبستگی بین ویژگی‌های مختلف احساسات را بطور خودکار نشان می‌دهد و با محاسبه چگالی احتمال مشترک برای مجموعه داده‌ها، مقدار مشخصی از زمان را صرفه جویی می‌کند [۸].

پاتیل[۲۳] و همکاران (۲۰۱۴) رویکرد شناسایی ناهنجاری در شبکه‌های اجتماعی و بررسی در مورد شناسایی ناهنجاری را مطرح کردند. نویسندگان از ترکیب دو روش شناسایی لینک ناهنجاری و شناسایی ناهنجاری متن، استفاده نمودند. در مدل شناسایی ناهنجاری این مقاله از داده‌های جمع آوری شده از پروفایل، متن، کلمات، URL های کاربر که توسط خود او به اشتراک گذاشته شد استفاده شده است. نویسندگان از مدل احتمالی بیز برای دسته‌بندی حوادث ناشناخته و حوادث غیرعادی یا عادی استفاده کردند [۹].

پنگ[۲۴] (۲۰۱۴) ازمتن زبان طبیعی و با استفاده از یک تجزیه کننده وابستگی سطحی رتبه احساسات و نظرات را محاسبه کرد. در این مقاله بیشتر به رابطه بین امتیاز احساسات و تبلیغات_اسپم پرداخته شد. مجموعه‌ای از قوانین تبعیض‌آمیز با مشاهدات بصری ایجاد شد و در نهایت، این مقاله با ترکیب سری‌های زمانی با قوانین تبعیض‌آمیز، تبلیغات_اسپم را به نحو موثری شناسایی کرد [۱۰].

بصیری و همکاران (۲۰۱۹) مرورهای فارسی گرفته شده از سایت دیجی‌کالا را در مورد تلفن‌های همراه برای یافتن مرورهای جعلی و مرورهای مربوط به مارک‌ها مطالعه کردند. در چارچوب پیشنهادی، یک مجموعه داده دارای برچسب، ابتدا با استفاده از رأی اکثریت در پاسخ به پرسشنامه‌ای با ۱۱ سوال که قبلا برای شناسایی اسپم توسط حاشیه‌نویسان انسانی طراحی شده‌است ایجاد شد. سپس چندین مرحله پیش‌پردازش برای زبان فارسی برای اصلاح اطلاعات انجام گرفت. در آخر ویژگی‌های مبتنی بر مرور و فراداده‌ها (ویژگی‌های مبتنی بر مرور شامل کلمات احساساتی و نمرات آنها، کلمات مکرر و فرکانس آنها، برچسب بخشی از گفتار [۲۵] از کلمات و فرکانس های آنها و ویژگی‌های فراداده شامل تعداد بازخوردهای مثبت، تعداد بازخورد منفی، عنوان مرور، طول مرور، امتیاز کاربر اختصاص داده شده است، نرخ کل محصول، تفاوت بین نرخ کلی و امتیاز کاربر) استخراج شدند و عملیات دسته‌بندی با سه روش ماشین بردار پشتیبان، درخت تصمیم و بیز ساده انجام گرفت. نتایج به دست آمده از ۳۰۰۰ مرور نشان داد که بیشترین دقت در درخت تصمیم با F-measure= 0.78 است. علاوه بر این، نتایج نشان داد که SVM برای داده‌های نامتوازن و درخت تصمیم برای داده‌های متوازن، هنگامی که فراداده و ویژگی های مبتنی بر مرور آموزش می‌گیرند عملکرد بهتری دارند [۱۱].

صدیقی و همکاران (۲۰۱۷) یک روش مبتنی بر درخت تصمیم برای شناسایی هرزنظر به نام RLOSD پیشنهاد کردند. مجموعه داده‌های اولیه از سایت Yelp گرفته شد. این روش ترکیبی از یادگیری بدون نظارت به همراه روش‌های انتخاب ویژگی سنتی است که بعد از استخراج ویژگی‌ها، با یک درخت تصمیم ارزیابی انجام می‌گیرد. این مدل برای انتخاب ویژگی‌های مناسب با توجه به همبستگی داده‌ها، سه ویژگی تکرار کلمات، برچسب زنی بخشی از گفتار و استفاده از n-گرم‌ها را انتخاب کرد. در انتها مقایسه روش پیشنهادی مقاله با سه روش ماشین بردار پشتیبان، بیز ساده و رگرسیون نشان‌دهنده کارایی روش RLOSD بود [۱۲].

حیدری و همکاران (۲۰۱۹) چالش‌های تحلیل احساسات در زبان فارسی را بررسی نمودند [۱۳]. الله‌کرمی و همکاران (۲۰۱۹) تحلیل روش‌های مبتنی بر پیوند و محتوا را برای نظرکاوی در زبان فارسی انجام دادند [۱۴]. کریمی و همکاران (۲۰۱۹) با نمایش زبان قبل از آموزش و یادگیری عمیق، به تحلیل احساس بر متون فارسی پرداختند [۱۵]. بصیری و همکاران (۲۰۱۹) سیستم ترکیبی برای نظرکاوی در زبان فارسی ارائه دادند [۱۶]. ذبیدی و همکاران (۲۰۱۹) نظرکاوی به زبان فارسی را با روش استخراج ویژگی ترکیبی مبتنی بر شبکه عصبی کانولوشن، انجام دادند [۱۷]. عطایی و همکاران (۲۰۱۹) مجموعه داده‌های تحلیل احساس مبتنی بر جنبه را در فارسی ایجاد کردند [۱۸]. بصیری و همکاران (۲۰۱۸) بهبود تحلیل احساس در زبان فارسی را با پالایش واژگان، انجام دادند [۱۹]. عسکریان و همکاران (۲۰۱۸) تأثیر ویژگی‌های احساس بر دسته‌بندی قطبیت احساس را در مرورهای فارسی بررسی کردند [۲۰]. دهکردی و همکاران (۲۰۱۸) با استفاده از عبارات قطب مخالف، تحلیل احساس در زبان فارسی را بهبود بخشیدند [۲۱]. شمس و همکاران (۲۰۱۷) ترکیبی از تخصیص دیریکله پنهان با تحلیل همزمان وقایع در استخراج جنبه را ارائه دادند [۲۲].

۳- متدولوژی پژوهش

این مقاله مطابق متدولوژی شکل(۱)، در سه مرحله، بهترین روش مدلسازی در شناسایی هرزنظر را با ساخت مجموعه ترکیبی از داده‌ها و با توجه به سطح سند و جمله در زبان فارسی معرفی می‌کند. مرحله اول جمع‌آوری مجموعه داده اولیه و پیش‌پردازش آن، مرحله دوم ساخت مجموعه ویژگی‌ها، آماده‌سازی مجموعه داده و برچسب‌زنی داده‌ها و مرحله سوم مدلسازی با ۶ روش پیشنهادی است. در ادامه هر یک از این مراحل توضیح داده خواهند شد.

۳-۱- جمع‌آوری مجموعه داده اولیه و پیش‌پردازش آن

داده‌های متنی، فرمت اصلی داده‌های رسانه‌های اجتماعی است که بازخورد، نگرش، نظرات و افکار کاربران را نسبت به محصولات، خدمات، سیاست‌ها و سایر موضوعات نشان می‌دهند [۳]. داده‌های متنی این مقاله، نظرات فارسی کاربران در حوزه هتلداری در بازه زمانی پنج سال است و از دو سایت iranhotelonline.com و egardesh.com با خزنده وب جمع‌آوری و در یک پایگاه داده ذخیره شد. برای شناسایی هرزنظر، ۱۱۰۰ نظر از کاربران مختلف و با طول متفاوت، بصورت تصادفی از پایگاه داده استخراج و در یک فایل اکسل ذخیره شد.

عملیات پیش‌پردازش شامل نرمالسازی متن و شناسایی کلمات عامیانه است که بر مجموعه داده متنی اولیه اعمال شد.

۳-۲- ساخت مجموعه ویژگی‌ها، آماده‌سازی مجموعه داده و برچسب‌زنی داده‌ها

تا کنون از سه نوع داده اصلی برای شناسایی هرزنظر در مرورها و نظرات استفاده شده است که عبارت است از [۳]:

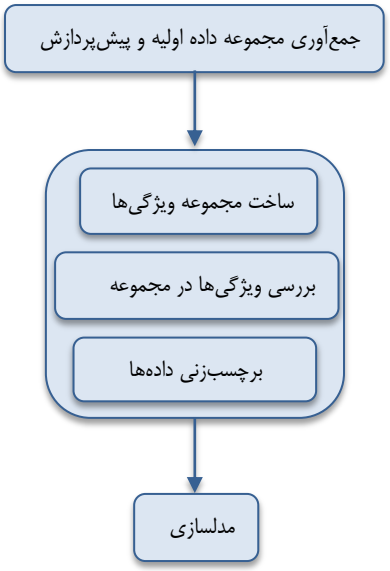
۱- محتوای مرور و نظر: این دسته شامل ویژگی‌های زبان‌شناختی و یا معناشناختی برای فریب و نیرنگ کاربران است. مانند متون تکراری.

۲- فراداده در مرور و نظر: از این داده‌ها می‌توان انواع متعددی از الگوهای رفتاری غیرعادی کاربران را کشف کرد. مانند شناسه کاربری نویسنده.

۳- اطلاعات درباره محصول: این داده‌ها اطلاعاتی درباره محصول یا موجودیت هستند و آن را شرح می‌دهند. مانند رتبه‌فروش و قیمت محصول یا موجودیت.

مجموعه داده برای مدلسازی، در یازده ستون آماده شد که ستون اول نظرات کاربران و ستون ۲ تا یازده داده‌های اصلی یا ویژگی‌ها هستند. در این مقاله مجموعه ویژگی‌ها برای اولین بار در زبان فارسی با ترکیب سه داده اصلی یعنی محتوای مرور، فراداده و اطلاعات محصول یا موجودیت [۳] ساخته شد. این مجموعه شامل ویژگی‌هایی است که در زبان فارسی می‌توانند در شناسایی هرزنظر موثر باشند. در جدول (۱) مجموعه ویژگی‌ها به تفکیک آمده است.

از مجموعه ویژگی‌های ساخته شده در این مقاله (جدول ۱)، چهار ویژگی "شناسه کاربری تکراری، مرور تکراری، زمان ارسال نظر و تعداد مرورهای نویسنده"، برای شناسایی هرزنظر در زبان انگلیسی درمورد محصولات و در سطح کلمه در [۲] استفاده شده است. در حالیکه این مقاله از آنها در جملات فارسی و برای ارزیابی در سطح سند و جمله استفاده می‌کند. همچنین چهار ویژگی "نظرات کلی، قطبیت نظر، مرور با جملات خاص مشابه و تطبیق قطبیت نظر با رتبه کلی هتل"، برای اولین بار در این مقاله معرفی شدند. از طرفی برخلاف نظر جیندال و لیو [۲] ما معتقدیم می‌توان از دو ویژگی "تبلیغات و سایر متون بی‌ربط"، برای شناسایی هرزنظر استفاده کرد به شرط آنکه جملات تبلیغی و بی‌ربط در کنار سایر جملات ذهنی در یک نظر، استفاده شده باشند. بنابراین در این مقاله علاوه بر نوآوری در شناسایی هرزنظر در سطح سند و جمله در زبان فارسی شش ویژگی جدید "نظرات کلی، قطبیت نظر، مرور با جملات خاص مشابه، تطبیق قطبیت نظر با رتبه کلی هتل، تبلیغات و سایر متون بی‌ربط " نیز در شناسایی استفاده شد.



شکل (۱) : متدولوژی پیشنهادی شناسایی هرزنظر

جدول (۱) : ویژگی‌های استفاده شده در شناسایی هرزنظر

محتوای مرور	فراداده	اطلاعات موجودیت
تبلیغات	شناسه کاربری تکراری	قطبیت نظر
متون بی‌ربط	زمان ارسال نظر از زمان افتتاح هتل	نظرات کلی
متون تکراری	تطبیق قطبیت نظر با رتبه کلی هتل	مرور با جملات خاص مشابه
		تعداد مرورهای نویسنده نسبت به کل مرورها

بعد از انتخاب ویژگی‌ها، ۱۱۰۰ نظر به صورت دستی مطالعه شد. سپس ده ویژگی جدول ۱ در نظرات مورد بررسی قرار گرفت. به عنوان مثال اگر کاربر یک نظر کلی ارائه کرده باشد ویژگی نظرات کلی برابر Yesو در غیراینصورت برابر No خواهد بود و یا اگر نظر کلی کاربر مثبت باشد ویژگی قطبیت نظر Yes در غیراینصورت برابر No خواهد بود. لازم به ذکر است که این بررسی در سطح کلی سند و سپس در سطح جمله انجام گرفته است. به ویژگی‌هایی مانند کلی بودن نظرات یا قطبیت آنها توجه نمایید. این ویژگی‌ها، از سطح کلمه خارج شده است و با بررسی کلی نظرات بیان‌شده کاربران، آنها را به طور کامل و در سطح سند، پوشش می‌دهد و در نهایت، دقت و صحت خروجی را فزایش می‌دهد.

قبل از ورود به مرحله بعد و انجام مدلسازی، لازم است برای داده‌ها، برچسبی به نام هرزنظر اضافه شود تا با ویژگی‌های جدول (۱)، جعلی بودن یا نبودن نظرات در این برچسب تعیین گردد. در ستون هرز نظر مقدار Yes به معنای جعلی بودن و مقدار No به معنای جعلی نبودن است.

۳-۳- مدلسازی

شناسایی هرزنظر به دلیل تحلیل رفتار انسانی فقط یک مساله پردازش زبان طبیعی[۲۶] نیست بلکه یک مساله داده‌کاوی[۲۷] نیز هست. بنابراین می‌توان با استفاده از یادگیری بانظارت، آن را به عنوان یک مساله دسته‌بندی، با دو دسته جعلی و غیرجعلی تعریف کرد. با این وجود همانطور که ذکر شد یکی از مشکلات اصلی، شناسایی سخت و حتی غیرممکن مرورهای جعلی با خواندن آنها است زیرا یک بیان‌کننده هرزنظر می‌تواند به دقت مروری جعلی بنویسد که شبیه یک مرور واقعی است. به دلیل این مشکل، هیچ داده مطمئنی درباره مرورهای جعلی و غیرجعلی وجود ندارد تا با آن الگوریتم یادگیری ماشینی، شناسایی مرورهای جعلی را یاد بگیرد [۳]. یکی از راهکارها در رفع این مشکل، استفاده از ویژگی‌های متعدد برای مدلسازی است. بنابراین این مقاله در مرحله قبل مجموعه ویژگی‌های ترکیبی را با ده ویژگی از محتوای مرور، فراداده و اطلاعات محصول ساخت (جدول (۱)). بعد از تکمیل مجموعه داده اولیه با مجموعه ویژگی‌ها، شش روش مدلسازی یعنی درخت تصمیم، جنگل تصادفی، قوانین استقراء، ماشین بردار پشتیبان، شبکه عصبی با یک و دو لایه پنهان و K_نزدیک‌ترین همسایه با دو مقدار مختلف برای K، بر مجموعه داده اعمال شدند تا بهترین و دقیق‌ترین دسته‌بند در شناسایی هرزنظر، ارائه شود. این شش روش در ادامه مورد بررسی قرار می‌گیرند.

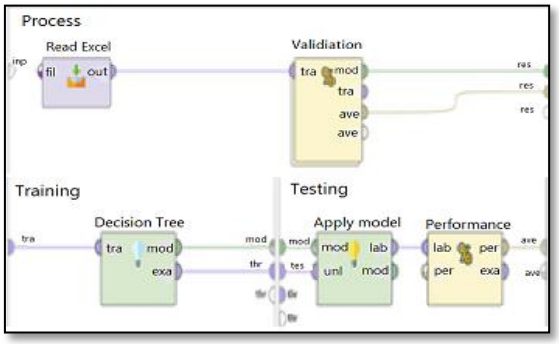
در این مقاله برای مدلسازی از نرم‌افزار RapidMiner استفاده شد. در تمام روش‌های مدلسازی، توزیع ۷۰ به ۳۰ برای داده‌های آموزشی و آزمایشی در دسته‌بندها لحاظ گردید. تصویر کلی عملگرهای مورد استفاده در شکل (۲) آمده است. البته باید توجه کرد که دسته‌بند مورد استفاده در بخش Training عملگر Validation مطابق با روش مدلسازی تغییر خواهد کرد. عملگر Read Excel مجموعه داده‌ها را از ورودی دریافت می‌کند سپس برای ارزیابی هر دسته‌بند، در بخش آموزش عملگر Validation، یکی از شش روش نام برده شده برای مدلسازی استفاده خواهد شد و در بخش آزمایش آن، دو عملگر Apply Model و Performance به ترتیب برای اعمال مدل بر مجموعه داده‌ها و ارزیابی کارآیی دسته‌بند استفاده می‌شوند.

مدلسازی با درخت تصمیم

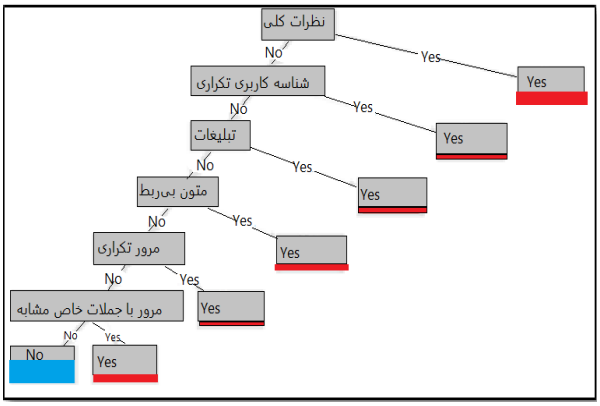
درخت تصمیم با تحلیل داده‌های بزرگ در زمان کوتاه و توانایی کار با داده‌های گسسته، یک مدل جعبه سفید بوده و توصیف شرایط آن به آسانی و با منطق بولی امکان‌پذیر است. بنابراین به عنوان یکی از دسته‌بندها برای مدلسازی استفاده شد. عملگر Decition Tree برای مدلسازی در نرم‌افزار Rapid Miner با تنظیمات Criterion= accuracy، maximal depth=10 و Confidence=0.1 استفاده شد. در شکل (۳) درخت تصمیم خروجی نمایش داده شده است. صحت این روش ۹۸.۴۸٪ است.

مدلسازی با جنگل تصادفی

طراحی یک درخت تصمیم بهینه، دشوار بوده و کارایی هر درخت به چگونگی طراحی آن بستگی دارد بنابراین برای طراحی درخت بهینه در این مرحله، مدلسازی با جنگل تصادفی انتخاب شد. عملگر Random Forest برای مدلسازی در نرم‌افزار Rapid Miner با تنظیمات number of trees=80، Criterion= accuracy و maximal depth=10 استفاده شد. صحت مدلسازی در این روش ۹۹.۰۹٪ است.



شکل (۲): نمونه‌ای از عملگرها در مدلسازی (درخت تصمیم) در نرم‌افزار داده‌کاوی Rapid Miner



شکل(۳): درخت تصمیم خروجی

مدلسازی با قوانین استقراء

به عنوان یکی از دسته‌بندهای مبتنی بر قانون، مدلسازی با قوانین استقراء در این مرحله انجام شد. عملگر Rule Induction با Criterion= accuracy و sample ratio=1.0 برای مدلسازی در نرم‌افزار Rapid Miner استفاده شد. بخشی از قوانین استخراج شده این مدل در شکل (۴) آمده است. صحت مدلسازی با این روش، ۹۶.۶۷٪ است.

مدلسازی با ماشین بردار پشتیبان

در این بخش عملگر SVM با گزینه svm type=nu_svc برای وظایف دسته‌بندی تنظیم شد. همچنین به دلیل اینکه نتایج دسته‌بند غیرخطی SVM، عملکرد بهتری نسبت به خطی دارد لذا تنظیمات عملگر در گزینه kernel type روی rbf تنظیم شد تا نمونه‌سازی غیرخطی انجام شود. همچنین در مدلسازی با SVM به دلیل اینکه عملگر با فیلدهای عددی اجرا می‌شود در فایل ورودی مجموعه داده به جای yes عدد ۱ و به جای No عدد صفر جایگزین شد. شکل(۵) مدل هسته استخراج شده از این روش را نشان می‌دهد. صحت مدلسازی با این روش، ۹۶.۰۶٪ است.

مدلسازی با شبکه عصبی

توسط الگوریتم شبکه‌های عصبی، می‌توان مدل‌های مختلف و پیچیده‌ای را شناخت و دسته‌بندی را با دقت خوب انجام داد. در این روش مدلسازی از عملگر Neural Net استفاده شد. خروجی شبکه عصبی با یک لایه پنهان، در شکل (۶) نشان داده شده است. صحت مدلسازی برای یک لایه پنهان ۹۸.۶۸٪ و با دو لایه پنهان ۹۶.۶۷٪ است.

مدلسازی با K-نزدیک‌ترین همسایه

دسته‌بند K-نزدیک‌ترین همسایه را به عنوان یک دسته‌بند تنبل[۲۸] برای مدلسازی استفاده کردیم. در تنظیم مقادیر عملگر K_NN از اعداد زوج استفاده نشد زیرا با وجود برچسب زوج با دو مقدار جعلی و غیرجعلی، توازن بین دسته‌ها با ورود حجم داده‌ای بیشتر در یک دسته از بین می‌رفت. بنابراین k با ۳ و ۵ مقداردهی شد. صحت خروجی با K=3 و K=5 در این روش به ترتیب ۹۷.۲۷٪ و ۹۷.۸۸٪ است.

٤- ارزیابی و مقایسه نتایج

بعد از انجام سه مرحله برای شناسایی هرزنظر، یک مرحله ارزیابی و مقایسه نتایج نیز برای معرفی دسته‌بند بهتر از لحاظ دقت، فراخوانی و صحت انجام شد. جدول (۲) در نمایش تعداد تشخیص‌های درست و نادرست هرزنظر، دارای سطرها و ستون‌هایی برای نمایش ماتریس آشفتگی است. همچنین دقت و فراخوانی دسته‌بندها نیز به ترتیب اولویت در این جدول آمده است.

بر اساس جدول (۲) نمودار مقایسه دقت، فراخوانی و صحت روش‌های مدلسازی در شکل ۷ رسم شد. لازم به ذکر است که هر یک از دو روش مدلسازی شبکه عصبی و K_نزدیکترین همسایه، در دو حالت بررسی شدند که در مجموع هشت خروجی از مدلسازی به دست آمد و این خروجی‌ها با هم مقایسه و ارزیابی نهایی انجام شد.

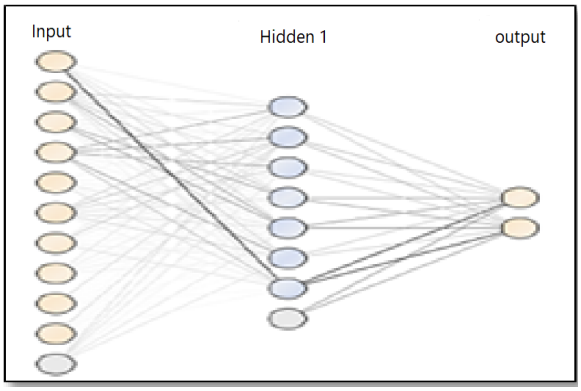
همان‌طور که گفته شد معیار مقایسه و سنجش برتری دسته‌بندها، سه پارامتر دقت، فراخوانی و صحت است که این مقادیر در جدول ۲ به ترتیب اولویت نشان داده شدند و در نمودار شکل (۷) با هم مورد مقایسه قرار گرفتند. با مقایسه مقادیر این پارامترها، در شناسایی هرزنظر، دسته‌بند جنگل تصادفی با دقت ۹۸.۶۵٪ فراخوانی ۹۷.۲۷٪ و صحت ۹۹.۰۹٪ بهترین عملکرد را در شش دسته‌بند دیگر داشت بنابراین به عنوان دقیق‌ترین دسته‌بند با بهترین عملکرد معرفی شد.

if تطبیق قطبیت نظر با رتبه کلی ≤ 0.500 and نظرات کلی ≤ 0.500 and ربط متون بی ≤ 0.500 and شناسه کاربری تکراری ≤ 0.500 then No (199 / 0)
if نظرات کلی ≤ 0.500 and شناسه کاربری تکراری ≤ 0.500 and مرور تکراری ≤ 0.500 and ربط متون بی ≤ 0.500 and تبلیغات ≤ 0.500 and مرورهایی با جملات خاص مشابه ≤ 0.500 then No (533 / 5)
else Yes (2 / 361)

شکل (٤): بخشی از قوانین استخراج‌شده از مدلسازی با قوانین استقراء

Total number of Support Vectors: 555
Bias (offset): 0.102
w[نظرات کلی] = 344.255
w[تبلیغات] = 46.243
w[ربط‌متون بی] = 51.381
w[شناسه کاربری تکراری] = 102.763
w[مرور تکراری] = 71.934
w[مرورهایی با جملات خاص مشابه] = 51.381
w[زمان ارسال نظر نسبت به افتتاح هتل] = 849.337
w[تعداد مرورهای نوشته شده توسط نویسنده نسبت به کل مرورها در مورد هتل] = 159.282
w[قطبیت نظر] = 943.620
w[تطبیق قطبیت نظر با رتبه کلی هتل] = 948.758
number of classes: 2
number of support vectors for class No: 278
number of support vectors for class Yes: 277

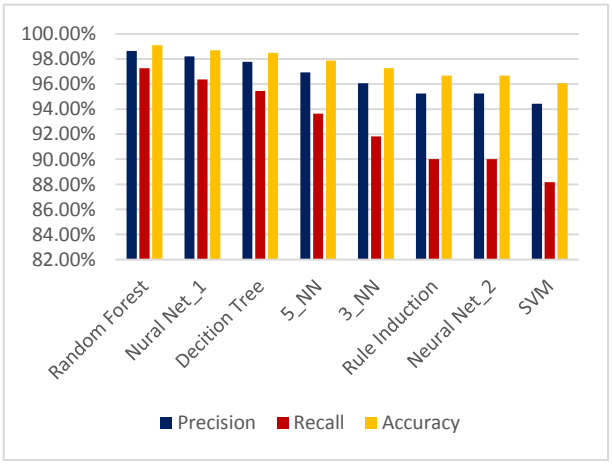
شکل (٥): مدل هسته استخراج‌شده از مدلسازی با ماشین بردار پشتیبان



شکل (٦): شبکه عصبی مدلسازی شده با یک لایه پنهان

جدول (۲): ماتریس آشفتگی، دقت و فراخوانی شش دسته‌بند در شناسایی هرزنظر

		true No	true Yes	class precision
Random Forest	pred. No	220	3	98.65%
	pred. Yes	0	107	100.00%
	class recall	100.00%	97.27%	
Neural Net_1	pred. No	220	4	98.21%
	pred. Yes	0	106	100.00%
	class recall	100.00%	96.36%	
Decition Tree	pred. No	220	5	97.78%
	pred. Yes	0	105	100.00%
	class recall	100.00%	95.45%	
K_NN (K=5)	pred. No	220	7	96.92%
	pred. Yes	0	103	100.00%
	class recall	100.00%	93.64%	
K_NN (K=۳)	pred. No	220	9	96.07%
	pred. Yes	0	101	100.00%
	class recall	100.00%	91.82%	
Rule Induction	pred. No	220	11	95.24%
	pred. Yes	0	99	100.00%
	class recall	100.00%	90.00%	
Neural Net_2	pred. No	220	11	95.24%
	pred. Yes	0	99	100.00%
	class recall	100.00%	90.00%	
SVM	pred. No	220	13	94.42%
	pred. Yes	0	97	100.00%
	class recall	100.00%	88.18%	



شکل (۷): نمودار مقایسه دقت، فراخوانی و صحت شش روش مدلسازی با هشت خروجی

۵- نتیجه

همگام با توسعه شبکه‌های اجتماعی، استفاده روزافزون کاربران از نظرات درج‌شده برای تصمیم‌گیری در حوزه‌هایی چون خرید، رای‌دادن در انتخابات و ... به پدیده‌ای با کاربرد چشمگیر مبدل شده است. نظرات مثبت اغلب به معنای سود و شهرت برای کسب و کارها و افراد و متاسفانه مشوقی برای ارسال نظرات جعلی به منظور ترویج یا تخریب محصولات، سرویس‌ها، خدمات، افراد و سازمان‌ها خواهد بود. یکی از چالش‌های هرزنظر، این واقعیت است که یک فرد بدون ترس از افشای هویت واقعی از هر نقطه جهان می‌تواند به راحتی به بیان نظرات خود با اهداف مخرب در رسانه‌های اجتماعی اقدام کند و با توجه به عدم شناخت آسان نظرات جعلی، چالش بعدی کمبود داده‌هایی برای شناسایی هرزنظر است که این مقاله به رفع چالش‌ها می‌پردازد.

بطور کلی در این مقاله ابتدا مجموعه داده متنی با مجموع ۱۱۰۰ نظر با طول متفاوت در حوزه هتلداری، جمع‌آوری و پیش‌پردازش شد. سپس این مقاله با سه نوآوری زیر اقدام به شناسایی هرزنظر نمود. اول ساخت مجموعه ویژگی‌ها با ترکیب محتوای نظر، فراداده و اطلاعات محصول، دوم شناسایی هرزنظر در سطح سند و جمله و در نهایت شناسایی هرزنظر در زبان فارسی.

مجموعه داده اولیه به روش دستی و با ده ویژگی از مجموعه ویژگی‌ها برچسب‌زنی شد. سپس شش روش دسته‌بندی شامل درخت تصمیم، جنگل تصادفی، قوانین استقراء، ماشین بردار پشتیبان، شبکه عصبی با یک و دو لایه و K–نزدیک‌ترین همسایه با دو مقدار مختلف برای K، بر مجموعه داده متنی اعمال گردید. برای آموزش و آزمایش توزیع ۷۰/۳۰ منظور شد. بعد از رسم ماتریس آشفتگی و محاسبه سه پارامتر دقت، فراخوانی و صحت، نتایج خروجی دسته‌بندها با یکدیگر مقایسه گردید. با مقایسه نتایج، دسته‌بند جنگل تصادفی به عنوان بهترین و دقیق‌ترین دسته‌بند در شناسایی هرزنظر با دقت، فراخوانی و صحت به ترتیب ۹۸.۶۵٪، ۹۷.۲۷٪ و ۹۹.۰۹٪ معرفی شد.

این پژوهش در گام‌های بعدی قصد استخراج جنبه از نظرات درست کاربران، تعیین قطبیت نظرات در قالب عددی و سپس رفع چالش‌های موجود در زبان فارسی را دارد.

مراجع

[1] Jindal, N., Liu, B., "*Review spam detection*", in Proceedings of WWW (Poster paper), 2007.
[2] Jindal, N., Liu, B., *"Opinion spam and analysis"*, in Proceedings of the Conference on Web Search and Web Data Mining (WSDM-2008), 2008.
[3] Liu, B., *"Sentiment Analysis and Opinion Mining"*, Morgan & Claypool Publishers, May 2012.
[4] Li, F., Huang, M., Yang, Y., Zhu, X., *"Learning to Identify Review Spam"*, in Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-2011), 2011.
[5] Ott, M., Choi, Y., Cardie, C., Hancock, J., *"Finding deceptive opinion spam by any stretch of the imagination"*, in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL-2011), 2011.
[6] Wu, Guangyu, Greene, D., Smyth, B., Cunningham, P., "Distortion as a validation criterion in the identification of

suspicious reviews", in Proceedings of Social Media Analytics, 2010.

[7] WANG, Z., TONG, V., XIN, X., CHIN, H., *"Anomaly Detection through Enhanced Sentiment Analysis on Social Media Data"* , IEEE 6th International Conference on Cloud Computing Technology and Science, 2014.

[8] Sun, X., Zhang, C., Li, G., Sun, D., Ren, F., Zomaya, A., Ranjan, R., *"Detecting users' anomalous emotion using social media for businessintelligence"*, Journal of Computational Science 25 (2018) 193–200, 2018.

[9] Patil, B., Bedi, P., *"Survey on Anomaly Detection Techniques in Social Networking"*, International Journal of Engineering Research & Technology (IJERT), Vol. 3 Issue 11, November-2014.

[10] Peng, Q., *"Detecting Spam Review through Sentiment Analysis"*, DOI:10.4304/jsw.9.8.2065-2072, August 2014.

[11] Basiri, M., Safarian, N., Khosravi Farsani, H., *"A Supervised Framework for Review Spam Detection in the Persian Language"*, 5th International conference on Web Research, Tehran, Iran, April 2019.

[12] Sedighi, Z., Ebrahimpour-Komleh, h., Bagheri, A., *"RLOSD: Representation Learning based Opinion Spam Detection"*, 3rd Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS), 2017.

[13] Heydari, M., *"Sentiment Analysis Challenges in Persian Language"*, July 2019.

[14] Allahkaram, N., Yari, A., *"Comparative Analysis of Link based and Content-based Methods for Opinion Mining in Persian language"*, International Journal of Web Research, 2018.

[15] Karimi, s., Sadat Shahrabadi, F., *"Sentiment analysis using BERT (pre-training language representations) and Deep Learning on Persian texts"*, Iran University of Science and Technology Deep Learning, Spring 2019.

[16] Basiri, M., Kabiri, A., *"Words Are Important: Improving Sentiment Analysis in the Persian Language by Lexicon Refining"*, Journal ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP) TALLIP, Volume 17 Issue 4, August 2018.

[17] Zobeidi, S., Naderan, M., Alavi, S., *"Opinion mining in Persian language using a hybrid feature extraction approach based on convolutional neural network"*, Multimedia Tools and Applications, DOI: 10.1007/s11042-019-07993-4, November 2019.

[18] Shangipour Ataei, T., Darvishi, K., Minaei-Bidgoli, B., Eetemadi, S., *"Pars-ABSA: an Aspect-based Sentiment Analysis dataset for Persian"*, Cornel university, Submitted on 26 Jul 2019.

[19] Basiri, M., Kabiri, A., *"Uninorm operators for sentence-level score aggregation in sentiment analysis"*, 4th International Conference on Web Research (ICWR), 2018.

[20] Asgarian, E., Kahani, M., Sharifi, S., *"The Impact of Sentiment Features on the Sentiment Polarity Classification in Persian Reviews*", Volume 10, Issue 1, pp 117–135, February 2018.

[21] Botshekanan Dehkordi, B., Basiri, M., *"Improving Persian Sentiment Analysis Using Opposing Polarity Phrases"*, 4th International Conference on Web Research, 2018.

[22] Shams, M., Baraani-Dastjerdi, A., *"Enriched LDA (ELDA): combination of latent Dirichlet allocation with word co-occurrence analysis for aspect extraction"*, Published in Expert Syst, DOI:10.1016/ j. eswa. 2017. 02. 038, Appl 2017.

زیرنویس‌ها

[1] Opinion spammers
[2] Opinion spamming
[3] Opinion spam
[4] Web spam
[5] Email spam
[6] Decition Tree
[7] Random Forest
[8] Rule Induction
[9] Support vector machine (SVM)
[10] Neural Net
[11] K Nearest Neighbors (KNN)
[12] Confusion Matrix
[13] Precision
[14] Recall
[15] Accuracy
[16] Jindal
[17] Liu
[18] Li
[19] Ott
[20] Wu
[21] Wang
[22] Sun
[23] Patil
[24] Peng
[25] Part Of Speech (POS)
[26] Natural Language Processing (NLP)
[27] Data Mining
[28] Lazy