Archive of SID

مهندسی صنایع و مدیریت شریف، (زمستان ۱۳۹۴)
دوره‌ی ۱ - ۳۱، شماره‌ی ۲/۲، ص. ۹۳-۱۰۳

Original Article

ارائه‌ی یک رویکرد ترکیبی بر اساس شبکه‌ی نظرات و ویژگی موجودیت‌ها برای کشف نظرات جعلی مشتریان الکترونیکی

محمد فتحیان * **(استاد)**
محمدرضا غلامیان (استادیار)
اکرم رضایی (دانشجوی کارشناسی ارشد)
دانشکده‌ی مهندسی صنایع، دانشگاه علم‌وصنعت ایران

نظرات مشتریان در وب‌سایت‌های تجارت الکترونیکی به مشتریان بالقوه برای تصمیم‌گیری بهتر درباره‌ی خرید، و به فروشندگان برای ارتقاء محصولات و راه‌کارهای بازاریابی کمک می‌کند. ازاین‌رو متقلبان با جعل نظرات، محصولات خود را ترویج و توجه مشتریان را از محصولات رقبا منحرف می‌کنند. هدف این تحقیق طراحی یک سیستم ترکیبی برای کشف نظرات جعلی آنلاین (برخط) است، چنان که از ویژگی‌های اشیاء و روابط بین موجودیت‌های مختلف به‌طور هم‌زمان بهره گیرد. برای رسیدن به این هدف از داده‌کاوی رابطه‌یی با الگوریتم شبکه‌ی وابستگی ارتباطی استفاده شد. شبکه‌ی ارتباطی کاربران و نظرات براساس اعتماد، عدم‌اعتماد، بازخورد، کاربر و محصول مشترک ایجاد شد. این روش روی دو مجموعه داده‌ی نظرات اجرا شد. روش پیشنهادی علاوه‌بر آن که از نظر نتایج حاصله برتری نسبی به روش‌های دیگر دارد، روشی انعطاف‌پذیر برای انواع شبکه‌های نظرات براساس مجموعه داده‌ی مورد بررسی، بدون نیاز به فرضیات شهودی برای رابطه‌ی موجودیت‌ها، در اختیار می‌گذارد.

fathian@iust.ac.ir
gholamian@iust.ac.ir
rezaie.akram@gmail.com

واژگان کلیدی: نظر، مشتری الکترونیکی، نظرات جعلی، کشف نظرات جعلی.

۱. مقدمه

یکی از دلایلی که امروزه افراد احساس راحتی بیشتری برای خریدهای آنلاین[1] (برخط) نسبت به روش‌های سنتی می‌کنند، وفور اطلاعات درباره‌ی محصولات و فروشندگان است. یکی از کلیدی‌ترین انواع این اطلاعات، نظرات و امتیازات داده شده به محصولات و فروشندگان است. محققین به‌منظور بهره‌گیری از حجم زیاد اطلاعات تولیدی توسط کاربران، روش‌های مختلفی برای تحلیل نظرات توسعه داده‌اند که شامل پردازش زبان طبیعی و فنون داده‌کاوی است. تمامی این تحقیقات فرضیه‌ی یکسانی را مبنی بر موثق بودن تمامی نظرات مورد تحلیل می‌پذیرند، که در واقع با حقیقت همخوانی ندارد. با در نظرگرفتن دسترسی آسان نظرات و تأثیر قابل توجه آن‌ها بر فروش، انگیزه‌ی فزاینده‌یی برای دستکاری‌های سودآور نظرات توسط متقلبان وجود دارد. برای یک کاربر عادی تشخیص نظرات فریب‌آمیز تنها با خواندن آنها، اگر غیرممکن نباشد، بسیار دشوار است. اگر متقلبان نظرات مثبت ثبت کنند شانس فروش محصول را بالا می‌برند؛ اگر نظرات منفی بنویسند به احتمال زیاد کاربران محصولات دیگری را برای خرید انتخاب می‌کنند. مسئله وقتی بغرنج‌تر می‌شود که متقلبان نظرات منفی را برای محصولات با کیفیت، و نظرات مثبت را برای محصولات بی‌کیفیت ابراز کنند. تحقیقات اخیر نشان می‌دهد که ۸۳ درصد از مشتریان نظرات آنلاین را بررسی می‌کنند و ۸۰ درصد مشتریان تصمیم خرید خود را در پی مشاهده‌ی نظرات منفی تغییر داده‌اند.[1] وجود نظرات جعلی غیرقابل تشخیص باعث فریب مصرف‌کنندگان و انتخاب‌های غیربهینه‌ی آنان می‌شود؛ این امر موجب بی‌اعتمادی و بی‌توجهی به اطلاعات مفید ارسال شده توسط کاربران بی‌طرف می‌شود.

نظرات، محتوای تولیدشده توسط کاربران است که برای نشان دادن عقایدشان درباره‌ی اشیا ارائه می‌کنند.[2و3] در ادبیات، از لفظ نظرات جعلی[2] برای فریب خوانندگان یا سیستم‌های خودکار با ارائه‌ی عقاید مثبت یا منفی ساختگی برای بهبود یا تخریب شهرت محصول هدف استفاده می‌شود.[2]

کشف نظرات جعلی عبارت است از تشخیص نظرات تقلبی در سایت‌های برخط، با داشتن همه نظرات روی سایت، متن نظر، نویسنده، محصولی که نظر برای آن نوشته شده، زمان ارسال نظر و امتیاز اختصاص یافته.[4] در اکثر روش‌های موجود برای کشف نظرات جعلی، فرض مستقل بودن نظرات رعایت شده است، درحالی که این فرض به‌دلیل وجود ارتباطات ضمنی و صریح رد می‌شود. در این مطالعه این ارتباطات برای یادگیری رابطه‌یی، در جهت ارتقای کارایی طبقه‌بندی مورد استفاده قرار می‌گیرند. در این‌حالت از کلاس موجودیت‌های مرتبط نیز به‌عنوان

* نویسنده مسئول
تاریخ: دریافت ۱۳۹۲/۱۲/۲۵، اصلاحیه ۱۳۹۳/۴/۱۵، پذیرش ۱۳۹۳/۵/۱۴.

www.SID.ir

Archive of SID

متغیرهای پیش‌گویی‌کننده استفاده‌می‌شود. الگوریتم مورد استفاده، علاوه بر ارتباطات، از ویژگی‌های محلی موجودیت‌ها نیز استفاده‌ی داخلی می‌کند. در حالی که در تحقیقات پیشین نمودارهای رابطه فقط بین موجودیت‌های متفاوت شکل گرفته‌اند، در این تحقیق شبکه‌های جدیدی از نظرات و کاربران با به‌کارگیری چندین معیار ارتباط بین نظرات شامل اعتماد، عدم‌اعتماد، کاربر و محصول مشترک و بازخورد نظر توسعه داده شد. داده‌کاوی رابطه‌یی، در قالب الگوریتم شبکه‌ی وابستگی ارتباطی (RDN)[3] در این سیستم به‌کارگرفته شد. مزیت این روش انعطاف‌پذیری سیستم برای هر مجموعه داده با هر ترکیب از موجودیت‌هاست، درحالی‌که در روش‌های پیشین مبتنی بر شبکه، شبکه‌ها براساس یک مجموعه داده‌ی خاص شکل گرفته‌اند. از آنجا که داده‌های برچسب‌دار چالش اساسی برای سیستم‌های تشخیص نظرات جعلی است، در این تحقیق از یک روش ابتکاری برای برچسب‌زنی داده‌ها استفاده شد. این روش براساس ویژگی‌های رفتاری متقلبان و تحلیل متن عمل می‌کند. موفقیت این روش‌ها در مطالعات پیشین تضمین می‌کند که مجموعه‌ی موثقی از داده‌های برچسب‌دار در اختیار داشته باشیم.

در ادامه‌ی این مقاله، ادبیات موجود درباره‌ی موضوع به تفصیل مورد بحث قرار می‌گیرد (بخش دوم). سپس ضمن بیان روش تحقیق (بخش سوم)، روش پیشنهادی تشریح خواهد شد (بخش چهارم). با معرفی مجموعه داده‌های مورد استفاده، نتایج حاصل از اجرای روش روی داده‌ها ارائه می‌شود (بخش پنجم). پس از بحث درباره‌ی نتایج حاصل، روش پیشنهادی با سایر روش‌های مشابه مقایسه خواهد شد (بخش ششم).

۲. مرور ادبیات

در این بخش روش‌های کشف نظرات جعلی از دیدگاه‌های مختلف شامل نظارتی و غیرنظارتی بودن، روش‌های رفتاری، تحلیل متن نظرات و روش‌های مبتنی بر نمودار مورد بررسی قرار می‌گیرد و شیوه‌های خاص هرگروه بیان می‌شود. در انتهای بخش به چالش منحصر به فرد داده‌های برچسب‌دار در کشف نظرات جعلی خواهیم پرداخت.

۱.۲. دسته‌بندی روش‌های کشف نظرات جعلی

الف) روش‌های نظارتی و غیرنظارتی

رویکردهای موجود در کشف نظرات جعلی را می‌توان به دو دسته‌ی اصلی نظارتی و غیرنظارتی تقسیم کرد. به‌طور کلی، تشخیص نظرات جعلی را می‌توان به‌عنوان یک مسئله‌ی طبقه‌بندی با دو کلاس جعلی و غیرجعلی فرمول‌بندی کرد. روش‌های نظارتی از مجموعه داده‌های برچسب‌دار استفاده می‌کنند. بنابراین چالشی‌ترین بخش این دسته از روش‌ها به دست آوردن داده‌های زمینه‌یی برچسب‌دار است. الگوریتم‌های یادگیری نظارتی که در کشف نظرات جعلی به‌کار رفته‌اند، عمدتاً ماشین بردار پشتیبان، رگرسیون لجستیک و بیز ساده هستند.[۳-۵،۷] ویژگی مهم روش‌های غیرنظارتی عدم‌نیاز به داده‌های برچسب‌دار است. این ویژگی، به‌خصوص در کشف نظرات جعلی که به دست آوردن این داده‌ها در آن مشکل است، بسیار سودمند است.

در یکی از مطالعات انجام‌شده،[۲] از بین روش‌های غیرنظارتی مسئله‌ی شناسایی رفتارهای غیرمعمول کاربران مد نظر قرار گرفته است. این مسئله به‌صورت شناسایی گروهی از قواعد دور از انتظار فرموله می‌شود؛ مجموعه‌یی از انتظارات تعریف می‌شود و معیار غیرمنتظره بودن مرتبط نیز برای آن‌ها معرفی شده است. قواعد و گروه‌های دور از انتظار نشان‌گر رفتارهای غیرمعمول و غیرعادی نظردهندگان است. دور از انتظار بودن به‌صورت انحراف از انتظارات تعریف می‌شود. تکنیک داده‌کاوی کاوش اقلام مکرر[4] برای یافتن متقلبان گروهی در مطالعات انجام شده[۸] به‌کار رفته است با این استدلال که کاربران در گروه‌ها روی چندین محصول نظردهی می‌کنند.

در یکی از مطالعات انجام شده[۹] فرض اصلی این است که یک توزیع طبیعی در نظرات و امتیازات وجود دارد. یک موجودیت تجاری متقلب که افراد را برای نوشتن نظرات جعلی استخدام می‌کند، از این توزیع طبیعی فاصله می‌گیرد. در مطالعه‌یی دیگر،[۱۰] مسئله‌ی تشخیص تقلب در نظرات تکی بررسی شده است. نویسندگان این مسئله را به مسئله‌ی کشف الگوی زمانی[5] تبدیل کرده‌اند؛ آنها سه آماره‌ی تجمیعی شناسایی کرده‌اند که بر این نوع حمله دلالت می‌کنند. سپس یک سری زمانی چندبعدی[6] براساس برازش منحنی ایجاد می‌کنند. محققین نظردهندگان و همبستگی آنها در فوران‌ها را به‌صورت میدان تصادفی مارکوف مدل می‌کنند[۱۱] و از انتشار باور چرخه‌یی[7] برای استنتاج استفاده می‌کنند. در یک مدل غیرنظارتی[۱۲] با استفاده از یک رویکرد بیزین، مسئله به‌صورت یک مسئله‌ی خوشه‌بندی فرموله می‌شود. انگیزه‌ی اصلی در ابعاد متفاوت رفتار متقلبان نسبت به دیگران نهفته است؛ این امر باعث ایجاد واگرایی توزیعی بین دو خوشه‌ی متقلب و غیرمتقلب می‌شود. استنتاج مدل منجر به یادگیری توزیع جمعیتی دو خوشه می‌شود. با مشاهده‌ی تغییر در سبک نوشتار در طول زمان، دست‌کاری یا عدم‌دست‌کاری در نظرات به‌طور غیرنظارتی استنتاج می‌شود.[۱۳]

ب) مدل‌های رفتاری

با بررسی رویکردهای موجود در تشخیص نظرات جعلی، مشاهده می‌شود که بیشترین طرح‌ها بر مدل‌سازی رفتارهای مشکوک کاربران ــ ناشی از مشاهدات ــ تأکید دارند. در این مدل‌های رفتاری، روش‌ها به‌جای تشخیص نظرات جعلی بر تشخیص نظردهندگان متقلب متمرکز می‌شوند.[۱۴]

در مدل‌های رفتاری به هر کاربر درجه‌یی از متقلب بودن اختصاص می‌یابد و براساس آن کاربران رتبه‌بندی می‌شوند. کاربران با جایگاه بالا در این رتبه‌بندی متقلب، و نظرات آنها تقلب در نظر گرفته می‌شود. در واقع هرچه این درجه بالاتر باشد احتمال متقلب بودن آن‌ها بیشتر است. در برخی از مطالعات، از این مدل‌های رفتاری به‌عنوان ویژگی‌های ورودی یادگیرنده‌ها استفاده شده است.[۷] طبق مطالعات برخی از محققین، رفتار نظردهندگان ــ شامل امتیازات و نظرات چندتایی درمورد یک محصول یکتا یا گروهی از محصولات و نیز انحراف از امتیاز متوسط ــ را برای به دام‌انداختن رفتارهای متقلبانه یافتند.[۱۴] همچنین با هدف بهبود صحت تشخیص مطرح شده در مطالعات مذکور،[۱۴] دو فعالیت دیگر از کاربران شامل انتشار نظرات در یک دوره زمانی کوتاه و به‌مراتب کم‌تر بودن تعداد خریدها نسبت به تعداد نظرات را به‌عنوان ویژگی‌های رفتاری متقلبان اضافه کرده است.[۱۵]

ج) تحلیل محتوا

در ادبیات از شیوه‌های مختلفی برای ارزیابی سبک نوشتار در کشف نظرات جعلی استفاده شده است. الگوهای نحوی لغوی ساده نظیر دنباله‌های کلمات مجاور[8] و برچسب‌های بخش‌های گفتار[9] به‌عنوان شالوده‌هایی که به الگوریتم‌های یادگیری ماشین اجازه یادگیری از داده‌های متنی را می‌دهند مورد استفاده قرار گرفته‌اند.[۶،۷] فنگ و همکارانش[۱۶] با مطالعه سبک نحوی زبان برای تشخیص فریب، زاویه‌یی تا حدودی غیرمتعارف به مطالعات قبلی می‌افزاید.

مشابهت مندرجات نظرات در مدل‌های رفتاری مورد توجه قرارگرفته‌است. از

www.SID.ir

Archive of SID

آنجایی‌که متقلبان به‌اندازه‌ی کاربران قانونی برای نوشتن یک نظر کاملاً جدید زمان صرف نمی‌کنند، متن این نظرات ممکن است شبیه یا کاملاً یکسان باشد.[۸و۱۴] جنبه‌های مختلفی از متن شامل انسجام معنایی، خوانایی، قطبیت و تحلیل احساسی در این حوزه تحقیقاتی به‌کارگرفته شده‌اند.[۱۳،۱۶،۱۷و۱۸] مرجع[۳] مسئله را به‌صورت تشخیص فریب با استفاده از روانشناسی زبان می‌بینند. هزینه‌های بالای محاسباتی و تله‌های کارایی به دلیل تحلیل‌های ناصحیح متن از جمله دلایلی هستند که محققان از تحلیل متن نظرات اجتناب می‌کنند.

د) تحلیل‌های مبتنی بر نمودار

بیشترکارهای موجود در تشخیص نظرات جعلی روی تحلیل یک نظر یا یک نظردهنده در یک زمان متمرکز شده‌اند بدون این که ارتباطات بالقوه بین چند نظر یا نظردهنده را به‌حساب آورند. از سوی دیگر روش‌هایی که شبکه‌ی نظرات، نظردهندگان و محصولات را به‌کار می‌گیرند می‌توانند با ظرافت بیش‌تری علائم ساختاری را که فراتر از محتوای نظر و ابتکارات ساده‌اند را خلاصه‌سازی کنند.[۴] محققین روشی برای تشخیص گروه‌های متقلب روی محصولات هدف معرفی می‌کنند.[۸] در روش معرفی شده ابتدا کاوش اقلام مکرر برای یافتن مجموعه‌یی از گروه‌های کاندید به‌عنوان گروه‌های متقلب برچسب‌دار به‌کار گرفته می‌شود؛ و سپس از چندین مدل رفتاری برگرفته از روابط بین گروه‌ها، افراد و محصولات بهره می‌گیرد.

در مطالعات بعدی[۱۹] مفهوم جدید نمودار غیرمتجانس شامل سه نوع گره نظردهنده، نظر و فروشگاه پیشنهاد شده و یک مدل محاسباتی تکرارشونده برای شناسایی نظردهندگان مشکوک توسعه یافته است. در نمودار نظرات، ارتباطات با معرفی سه مفهوم اساسی قابلیت اعتماد نظردهندگان، درستی نظرات و اعتبار فروشگاه‌ها و تعاملات آنها مطرح می‌شوند.

در برخی از مطالعات انجام شده[۱۹] مسئله‌ی ارتباطات بین سه موجودیت لحاظ شده است، اما ارتباطات بین خود نظرات یا بین خود نظردهندگان در نظر گرفته نشده است. این نوع ارتباطات در مطالعات پیشین[۱۹] از طریق پیوند دادن نظردهندگان در یک فوران مد نظر قرار گرفته است. متقلبان برای آنکه بتوانند تأثیر عمده‌یی بر استنباط کاربران از کیفیت محصول مورد بررسی بگذارند، با یکدیگر به صورت گروهی کار می‌کنند. در این تحقیق یک نمودار، روش انتشار نمودار، و شاخص‌های رفتاری برای شناسایی متقلبان معرفی شده است؛ وقوع همکاری بین دو نظردهنده در یک جهش مشترک با یک بال اتصال‌دهنده بین گره‌های مخفی مربوط به آنها در میدان تصادفی مارکوف نشان داده می‌شود. در این مطالعه از الگوریتم انتشار پیام چرخه‌یی که برای حل مسائل استنتاج تقریبی روی نمودارهای غیرجهت‌دار کاربرد دارد استفاده شده است.

برخی از محققین مجموعه داده‌ی نظرات را به‌صورت یک شبکه‌ی دوبخشی بین کاربران و محصولات نشان داده‌اند.[۴] برچسب اشیاء در مجموعه داده به‌شدت همبسته‌اند چرا که هرکدام براساس دیگری تعریف شده است.

۲.۲. برچسب‌زنی داده

چالش منحصر به فرد در تشخیص نظرات جعلی در عدم‌وجود برچسبی برای تعیین جعلی بودن نظرات یا مشکوک بودن نظردهندگان است. بنابراین روش‌هایی برای تولید این داده پیشنهاد شده‌اند. رویکردهای مورد استفاده در ادبیات به سه گروه تقسیم می‌شود: ۱. رویکرد دستی؛ روشی که در بیشتر مطالعات موجود مورد استفاده قرار گرفته و در آن تفسیر نظرات توسط یک یا چند عامل انسانی خبره انجام می‌گیرد. رویکرد دستی به‌خصوص در مقیاس‌های بزرگ داده بسیار پرهزینه و دشوار است، چرا که نظرات توصیفاتی ذهنی‌اند که در آنها مرز بین شخصی بودن نظر، نظر با کیفیت پایین و نظرات جعلی نامعلوم است. برای ساده‌سازی این مشکلات شاخص‌های ابتکاری برای کمک به خبرگان در تصمیم‌گیری تهیه شده است. محققین زیرمجموعه‌یی از نظرات مشکوک و غیرمشکوک کشف شده توسط الگوریتم معرفی شده را به‌همراه سایر نظرات، برای بررسی در اختیار ارزیابان قرار می‌دهند.[۱۴] آنان نظرات را براساس امتیازات افراطی یعنی امتیاز ۱ یا ۵ ارزیابی می‌کنند،[۸و۲۰] و به‌صورت تصادفی جفت نظرات با مشابهت بالا را برای بازرسی دستی در اختیار متخصص قرار می‌دهند.[۲۱و۲۲] توانایی انسان برای تشخیص تقلب در نظرات بسیار ضعیف است، به‌طوری که به گواهی برخی از مطالعات[۳] کارایی برچسب‌زنی دستی از حالت انتخاب تصادفی بهتر نیست.

رویکردهای ابتکاری در برخی از دامنه‌ها تقریب‌های خوبی عرضه می‌کنند. پس از حذف انواع خاصی از نظرات غیرمرتبط، نظرات تکراری یا نیمه‌تکراری به‌عنوان نظرات جعلی برچسب می‌خورند.[۵] در مواردی یک معیار انحراف برای ارزیابی تأثیر روش‌ها در آشکارساختن نظرات جعلی هتل‌ها در «مشاوره سفر» معرفی شده است.[۲۳] قاعده‌ی کلی این است که نظرات متقلبانه موجب انحراف قابل توجه رتبه‌بندی محبوبیت هتل‌ها می‌شود. بنابراین سازوکاری که نظرات جعلی را حذف کند، این ترتیب محبوبیت را به‌طور چشم‌گیری در مقایسه با زمانی که مجموعه‌ی مشابهی از نظرات به‌طور تصادفی حذف شوند، تغییر می‌دهد.[۲۴] محققین نظرات فیلتر شده از yelp.com ــ یکی از بزرگ‌ترین سایت‌های میزبانی نظرات کاربران که نظرات مشکوک را فیلتر می‌کند ــ را جعلی و نظرات غیرفیلتر شده را غیرجعلی به حساب می‌آورند.[۲۵و۷] در مطالعات بعدی چارچوبی برای تشخیص نظرات تولیدشده به‌طور خودکار توسعه یافت[۱۶] و از داده‌های ساختگی برای داده‌های با برچسب جعلی استفاده شد. این داده‌ها با جایگزینی هر جمله در یک نظر با مشابه‌ترین جمله در مخزن نظرات تولید می‌شوند. برای ارزیابی نتایج مدل غیرنظارتی تشخیص نظرات جعلی بدون نیاز به هرگونه برچسب‌زنی دستی، یک راه جدید معرفی شد.[۱۲] برای این منظور از نتایج مدل غیرنظارتی برای داده‌های آموزش مدل نظارتی با مجموعه ویژگی‌های استفاده‌نشده در مدل اول استفاده شده است. برای ارزیابی تأثیر روش‌ها در آشکارسازی نظرات جعلی هتل‌ها نیز یک معیار انحراف معرفی شد.[۲۳]

رویکرد جمع‌سپاری عبارت است از به دست آوردن خدمات مورد نیاز، ایده‌ها، یا محتوا با درخواست همکاری از یک گروه بزرگ افراد به‌خصوص از یک اجتماع برخط به جای کارکنان یا تأمین‌کنندگان سنتی. در مورد نظرات برخط و با استفاده از پایگاه‌هایی همچون «تورک فنی[۱۰]» می‌توان محتوای متقلبانه‌ی استانداردی بسیار شبیه به آنچه در عمل مشاهده شده است تولید کرد. در برخی از تحقیقات انجام شده از تورک فنی برای تولید این داده‌ها استفاده شده است[۳] و در مطالعاتی دیگر مجموعه داده‌یی از نظرات مشابه با تمایلات منفی تولید شده است.[۲۶] گروهی از مطالعات نیز از این داده‌های تولید شده استفاده کرده‌اند.[۱۶و۱۷]

۳. روش تحقیق

هدف این تحقیق، توسعه‌ی سیستمی خودکار برای کشف نظرات جعلی است. تلاش‌های صورت‌گرفته در این حوزه با مقاله‌ی جیندال و همکارانش در سال ۲۰۰۸ آغاز شده است. از میان روش‌های موجود در ادبیات یک رویکرد بر ویژگی‌های محلی موجودیت‌ها ازجمله ویژگی‌های متن، ویژگی‌های نویسنده‌ی نظر و محصول، و نیز ویژگی‌های رفتاری نظردهندگان استوار است و مدل طبقه‌بندی مبتنی بر داده‌های

www.SID.ir

Archive of SID

آموزشی است که به‌صورت نیمه‌خودکار برچسب‌زنی شده‌اند.[۱و۲] در مقابل این روش‌ها، روش‌های مبتنی بر نمودار هستند. این مطالعات شبکه‌یی از متن نظر، نظردهنده، محصول و فروشگاه را تعریف کرده و براساس تأثیرات متقابل موجودیت‌ها در یک فرایند تکراری برچسب هریک از موجودیت‌ها را تعیین می‌کند.[۱۹و۲]

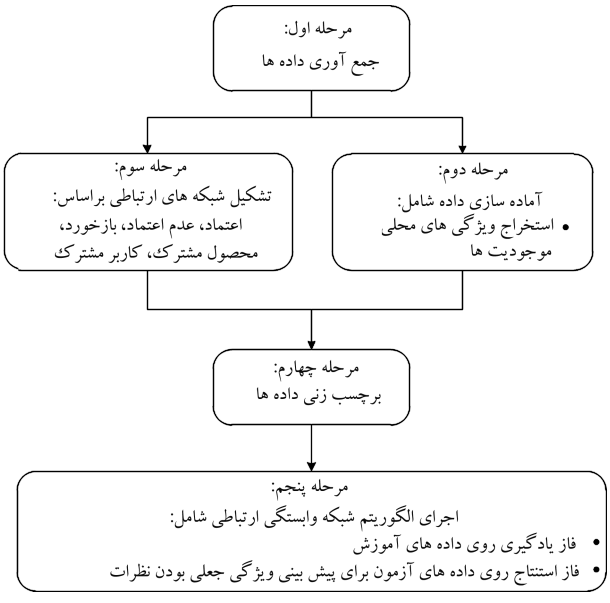
ویژگی‌های محلی موجودیت‌ها و تأثیرات شبکه‌یی بین آن‌ها هرکدام به‌تنهایی بخشی از نشانه‌های آشکارساز تقلب را در اختیار می‌گذارند، که بهره‌گیری از آن‌ها به‌طور همزمان برای توسعه‌ی سیستم تشخیص نظرات تقلبی در مطالعات پیشین صورت نگرفته است و هدف اصلی این پژوهش است. به‌همین منظور مطالعاتی درخصوص روش حل مسئله‌ی مطرح شده صورت گرفت و روشی برای توسعه‌ی سیستم مورد نظر معرفی شد که شبکه‌هایی از نوع نظر ـ نظر و کاربر ـ کاربر را با به‌کارگیری چندین معیار ارتباط بین نظرات ــ شامل اعتماد، عدم‌اعتماد، کاربر و محصول مشترک و بازخورد نظر ــ توسعه می‌دهد. سپس از داده‌کاوی رابطه‌یی در این شبکه‌ها، بر پایه‌ی الگوریتم شبکه‌ی وابستگی ارتباطی برای کشف نظرات جعلی استفاده شد. این الگوریتم مبتنی بر استفاده از وابستگی بین ارزش متغیرهای مشابه در موجودیت‌های مرتبط است که در مجموعه داده‌های نظرات برخط مشاهده می‌شود. ارزیابی و تحلیل نتایج این روش با استفاده از دو مجموعه‌ی داده در مرحله‌ی نهایی این تحقیق انجام پذیرفته است.

۴. معرفی روش پیشنهادی

در این بخش یک روش پنج مرحله‌یی برای کشف نظرات جعلی مشتریان، مطابق شکل ۱، پیشنهاد می‌شود. مراحل مختلف این روش در ادامه شرح داده می‌شود.

۱.۴. مرحله‌ی اول: جمع‌آوری داده‌ها

این داده‌ها ــ شامل ویژگی‌های نظرات و اشیاء مرتبط، نظیرکاربر و محصول ــ از وب‌سایت فروشگاه‌های اینترنتی یا وب‌سایت‌های نظرات مشتریان بازیابی می‌شود.



شکل ۱. روش پیشنهادی برای تشخیص نظرات جعلی.

۲.۴. مرحله‌ی دوم: آماده‌سازی داده‌ها

دراین مرحله ویژگی‌های محلی مورد نیاز موجودیت‌ها آماده‌سازی می‌شود. بدین‌منظور ویژگی‌های غیرمفید از داده‌ها حذف شده و ویژگی‌های مشتق از آن با استفاده از توابع تجمیعی و همچنین ویژگی‌های مربوط به متن به آن افزوده می‌شود. ویژگی‌های محلی موجودیت‌ها بخشی از نشانه‌های تقلب را آشکار می‌کنند.

۳.۴. مرحله‌ی سوم: تشکیل شبکه‌ی ارتباط

نفوذ تأثیرات اجتماعی در شکل‌دهی تصمیمات افراد در حوزه‌ی بازاریابی به‌خوبی شناخته شده است. در مطالعات انجام شده بر سیستم‌های توصیه‌گر مبتنی بر شبکه‌ی اجتماعی، از این روابط برای یادگیری ترجیحات دوستان استفاده شده است.[۴] این ارتباطات ضمنی بین نظردهندگان باید برای کشف نظرات جعلی در نظرگرفته شود. ما معتقدیم که از چنین ارتباطات اجتماعی می‌توان برای سنجش اعتبار یک نظر استفاده کرد.

ــ شبکه‌ی اعتماد: اعتماد یک احتمال ذهنی است که براساس آن شخص الف انتظار دارد فعالیت‌های شخص ب تأمین‌کننده‌ی رفاه وی باشد.[۲۴] به تعبیر دیگر اعتماد انتظار خوش‌بینانه‌ی یک فرد درباره‌ی نتیجه‌ی یک رویداد یا رفتار یک شخص است.[۲۷] براساس این تعاریف انتظار می‌رود کاربران عادی چنانچه نظرات مکتوب یک شخصی را قابل اعتماد و صادقانه ارزیابی کنند، به او اعتماد و در غیر این‌صورت او را حذف کنند. چنین روابطی بین کاربران متقلب به‌خصوص در فعالیت‌های گروهی مشاهده می‌شود. چراکه مشتریان بالقوه از نظرات افرادی که کاربران زیادی به آن اعتماد کرده‌اند و تعداد معدودی کاربر به آن بی‌اعتمادند، تأثیرپذیری بیشتری دارند. به‌طور شهودی می‌توان گفت که نظرات مرتبط برمبنای شبکه‌ی اعتماد و عدم‌اعتماد به یکدیگر، با احتمال بیشتری از لحاظ تقلبی و صحیح بودن در یک دسته قرارمی‌گیرند. از این جهت این نوع ارتباط را برای کشف نظرات جعلی مفید می‌دانیم.

ــ شبکه‌ی بازخورد نظر: علاوه بر شبکه‌ی اعتماد، شبکه‌ی دیگری نیز در سیستم نظرات برخط وجود دارد که امکان بهتری برای ارزیابی نظرات توسط افراد فراهم می‌آورد. اگرچه در شبکه‌ی اعتماد اشخاص را براساس مجموعه نظرات‌شان ارزیابی می‌کنیم، در شبکه‌ی بازخورد این امکان برای ارزیابی تک‌تک نظرات فراهم می‌شود. انتظار داریم ثبت بازخورد مثبت برای یک نظر متقلبانه یا بازخورد منفی برای یک نظر صحیح توسط یک کاربر، نشانه‌ی متقلب بودن بازخورد دهنده باشد و به استنتاج متقلبانه بودن نظرات او بینجامد. کاربران می‌توانند با قراردادن بازخورد روی نظرات، ارتباطاتی را به وجود آورند. نظرات معتبر بازخوردهای بیشتر و مثبت‌تری را دریافت می‌کنند.

ــ شبکه مبتنی برکاربرومحصول مشترک: متقلبان معمولاً به‌صورت گروهی فعالیت می‌کنند، حتی اگر یکدیگر را نشناسند ولی از طریق یک منبع مشترک هدایت می‌شوند. گروه متقلب نظرات خود را روی محصول هدف، به‌خصوص در یک بازه زمانی کوتاه، ارسال می‌کند. این گروه‌ها با منحرف ساختن امتیاز کلی محصول، تمایل به محصولات را تغییر می‌دهند. گروه‌های متقلب با نسخه‌برداری از نظرات سایر افراد، نظرات مرتبط با ویژگی‌های محلی مشابه ایجاد می‌کنند. از سوی دیگر نظراتی که با کاربر مشترک با هم مرتبط می‌شوند از نظر مشابهت متن نظر و مشابهت زمان ارسال (مثلاً در یک روز) و غیره، نسبت به حالت منفرد نشانه‌های

www.SID.ir

Archive of SID

بیشتری آشکار می‌کنند. براین اساس، علاوه بر ویژگی‌های هر نظر، کلاس هر نظر تحت تأثیر ویژگی‌های نظرات با کاربر و محصول مشترک و طبقه‌بندی آن‌هاست.

- نظر A با نظر B مرتبط است اگر توسط یک کاربر یا برای یک محصول مشترک ثبت شده باشد.
- نظر A با نظر B مرتبط است اگر کاربر ثبت‌کننده‌ی نظر B کاربر اعتمادکننده یا حذف‌کننده‌ی کاربر ثبت‌کننده‌ی نظر A باشد.
- نظر A با نظر B مرتبط است اگر کاربر ثبت‌کننده‌ی نظر B بازخوردی برای نظر A ثبت‌کرده باشد.

۴.۴. مرحله‌ی چهارم: برچسب‌زنی داده‌ها

یکی از مراحل اساسی و چالش‌برانگیز در توسعه و ارزیابی مدل‌های کشف نظرات جعلی، دست‌یابی به یک مجموعه داده برچسب‌دار است. رویکرد استفاده شده در این تحقیق را می‌توان در گروه رویکردهای ابتکاری قرار داد. در اینجا از روش‌های کشف نظرات جعلی که در مطالعات قبلی ارائه شده است بهره می‌بریم. این روش‌ها ازکارهای موفق در این زمینه انتخاب شده‌اند. مزیت این روش آن است که به‌دلیل انجام ارزیابی‌های موفق می‌توان مجموعه‌ی قابل اعتمادی از داده‌های برچسب‌دار در اختیار داشت.

در راهبرد اول از ویژگی‌های رفتاری استفاده می‌شود. بخش زیادی از سیستم‌های موجود در ادبیات، به معرفی ویژگی‌های رفتاری متعدد برای متقلبان و مطالعه‌ی نتایج آن اختصاص یافته است.[۱۲] این مطالعات نشان‌گر امیدبخش بودن برخی از ویژگی‌های معرفی شده برای کشف متقلبان است. با فرموله‌بندی این الگوهای رفتاری درجه‌یی از قابلیت اعتماد به هرکاربر اختصاص می‌یابد. نظرات جعلی اغلب توسط افرادی که برای این کار پاداش دریافت می‌کنند نوشته می‌شود. می‌توان نتیجه گرفت که نظرات منتسب به کاربران با قابلیت اعتماد کم، نظرات جعلی‌اند. قبل از تشریح ویژگی‌های رفتاری نمادهای مورد استفاده را معرفی می‌کنیم:

u: یک نمونه ازکاربران؛

O: مجموعه‌ی همه‌ی نظرات؛

o: یک نمونه از نظرات؛

o^u: مجموعه‌ی نظرات کاربر u؛

$o^{u,p}$: نظرکاربر u برای محصول p؛

P: مجموعه‌ی همه محصولات؛

p: یک نمونه از محصولات؛

P^u: مجموعه‌ی محصولات نظردهی شده توسط u؛

R^p: مجموعه‌ی امتیازات ۱ تا ۵ داده شده به محصول p؛

$r^{u,p}$: امتیازکاربر u به محصول p؛

$R^{u,p}$: مجموعه امتیازات کاربر u به محصول p؛

S_i^u: شاخص رفتاری iام.

۱.۴.۴. شاخص‌های رفتاری تقلب

چهار شاخصی که رفتارهای مشکوک درکاربران را شناسایی می‌کند عبارت‌اند از:

۱. امتیاز چندتایی. متقلبان برای منحرف ساختن نتایج سیستم‌های کاوش نظرات که خروجی آن‌ها رتبه‌بندی محصول یا ارزیابی کلی کاربران از محصول است و برای تغییر احساسات خوانندگان نظرات اقدام به تزریق مجموعه‌ی مشابهی از امتیازات و نظرات می‌کنند. شاخص امتیاز چندگانه برای نظرات به‌صورت رابطه‌ی ۱ تعریف می‌شود. در این رابطه اختلاف بیشترین و کم‌ترین امتیاز با تقسیم بر حداکثر اختلاف ممکن، نرمال شده و با کم شدن از ۱، میزان مشابهت محاسبه می‌شود. براساس معیار فوق هرچه میزان مشابهت و تعداد امتیازهای یک کاربر برای یک محصول بیشتر باشد، میزان اعتماد به آن نظرات کم‌تر است. بر اساس این شاخص نظرات چندگانه تقلبی در نظرگرفته می‌شود.[۱۲]

$$S_1^{O^{u,p}} = n(R^{u,p}) \times (1 - \frac{1}{4}(\max(R^{u,p}) - \min(R^{u,p}))) \quad (1)$$

۲. انحراف ازامتیاز متوسط. انتظار می‌رود امتیاز یک کاربر عادی برای یک محصول نزدیک به امتیاز سایرکاربران به آن محصول باشد. از آنجا که متقلب سعی در ترفیع یا تنزیل یک محصول دارد، امتیازدادنش به یک محصول کاملاً متفاوت از سایرین است. رابطه‌ی ۲ برای سنجش این تفاوت تعریف می‌شود. این معیار میانگین انحراف امتیاز کاربر نسبت به میانگین امتیازات سایرین برای همه‌ی محصولاتی است که کاربر به آن‌ها امتیاز داده است.[۱۲]

$$S_2^u = \underset{p \in P^u}{\text{Avg}} |r^{u,p} - \text{Avg}(R^p)| \quad (2)$$

۳. بیشترین تعداد نظرات در یک روز. بیشتر متقلبان حجم زیادی از نظرات را در یک روز ارسال می‌کنند. بیشترین تعداد نظرات ارائه شده توسط کاربر در یک روز، طبق رابطه‌ی ۳ شاخص خوبی برای تشخیص متقلبین است:[۱۲]

$$S_3^u = \max_{t:\text{every one day}} n(O_t^u) \quad (3)$$

۴. پنجره‌ی فعالیت. متقلبان معمولاً کاربران بلندمدتی نیستند؛ آن‌ها معمولاً نظرات خود را در یک پنجره‌ی زمانی کوتاه و به تعداد زیاد ثبت می‌کنند. درحالی که کاربران عادی نظرات خود را هراز چند گاهی و در بازه‌های زمانی بلندمدت ارائه می‌کنند. شاخص نشان‌گر این رفتار با محاسبه‌ی اختلاف زمانی اولین و آخرین نظر ثبت شده برای هرکاربر در رابطه‌ی ۴ محاسبه می‌شود:[۱۲]

$$S_4^u = \text{Last}(R^u) - \text{First}(R^u) \quad (4)$$

کاربرانی که هریک از شاخص‌های فوق برای آن‌ها از یک حد آستانه بالاتر باشد به‌عنوان متقلب برچسب می‌گیرند. نظرات مربوط به متقلبان به‌عنوان نظرات جعلی شناخته می‌شود.

در راهبرد دوم با استفاده از متن نظرات و مجموعه ویژگی‌های دنباله‌ی کلمات مجاور دوکلمه‌یی با استفاده ازالگوریتم طبقه‌بندی SVM، نظرات طبقه‌بندی می‌شوند. این روش برگرفته از تکنیک تشخیص نظرات جعلی است.[۳]

۵.۴. مرحله‌ی پنجم: یادگیری و استنتاج با استفاده از الگوریتم شبکه‌ی وابستگی ارتباطی

هرچند بسیاری از مجموعه داده‌های جمع‌آوری شده توسط سازمان‌ها ماهیت‌شان رابطه‌یی است، هنوز بسیاری از مطالعات یادگیری ماشین روی داده‌های گزاره‌یی مسطح تمرکز کرده‌اند. در داده‌های گزاره‌یی، نمونه‌ها ویژگی‌های اشیاء متجانس و مستقل به‌لحاظ آماری را حفظ می‌کنند، در حالی که در داده‌های رابطه‌یی نمونه‌ها حافظ ویژگی‌های اشیای نامتجانس و روابط بین آن‌ها هستند.[۲۸]

وجود خودهمبستگی موجب انگیزه‌ی لازم برای استفاده از تکنیک‌های رابطه‌یی در یادگیری و استنتاج است. خودهمبستگی، وابستگی بین ارزش متغیرهای مشابه

www.SID.ir

Archive of SID

در موجودیت‌های مرتبط است و این تقریباً ویژگی همه‌ی مجموعه داده‌های رابطه‌یی است. زمانی که داده‌های رابطه‌یی خودهمبستگی نشان می‌دهند، فرصتی است برای بهره‌مندی از ارتقاء کارایی مدل، چرا که استنتاج درباره‌ی یک شیء می‌تواند به استنتاج درباره‌ی اشیاء مرتبط کمک کند.[۲۸] مجموعه داده نظرات مشتریان برخط نمونه‌یی از یک مجموعه داده‌ی رابطه‌یی با اشیاء نامتجانس است. به‌دلیل ماهیت رابطه‌یی مجموعه داده نظرات، از داده‌کاوی رابطه‌یی به‌کمک الگوریتم شبکه‌ی وابستگی ارتباطی در این مطالعه استفاده می‌شود. به‌منظور درک چگونگی کاربرد این الگوریتم در ترکیب ویژگی‌های محلی و روابط موجودیت‌ها، در این بخش معرفی کوتاهی از این الگوریتم ارائه می‌شود.

هنگام مدل‌سازی داده‌های رابطه‌یی با یک مدل گرافیکی، سه نمودار داده، مدل و استنتاج مرتبط با مدل وجود دارد. شمای مجموعه داده‌ی رابطه‌یی توسط یک نمودار $G_D = (V_D, E_D)$ ــ که در آن گره‌ها نماینده‌ی اشیاء، و یال‌ها روابط بین اشیاء است ــ نشان داده می‌شود. هر گره و هر یال یک نوع دارد؛ هر نوع از اقلام شامل گره و یال تعدادی ویژگی مرتبط دارد. بنابراین هر نمونه از یال‌ها یا گره‌ها با یک مجموعه از ارزش ویژگی که با نوع آن تعیین می‌شود، مرتبط می‌شود. وابستگی‌های بین ویژگی‌ها در نمودار مدل $G_M = (V_M, E_M)$ نشان داده می‌شوند. ویژگی‌های یک شیء می‌تواند به‌صورت احتمالی به ویژگی‌های دیگر همان شیء همچنین به ویژگی‌های سایر اشیا و پیوندهای مرتبط در G_D وابسته باشد. ارتباطات در G_D برای محدودکردن جست‌وجو برای وابستگی‌های آماری ممکن و درنتیجه مجموعه یال‌هایی که در G_M ظاهر می‌شوند استفاده می‌شود. یک رابطه بین دو شیء در G_D لزوماً موجد وابستگی احتمالاتی بین ویژگی‌های آن‌ها در G_M نیست. به‌جای ایجاد ساختار وابستگی روی ویژگی‌های اشیای خاص، شبکه‌ی وابستگی ارتباطی یک ساختار وابستگی کلی در سطح نوع شیء تعریف می‌کند. هر گره به یک توزیع احتمال مشروط به سایر متغیرها مرتبط است. والدین X_k^t، یعنی kامین ویژگی از نوع t، سایر ویژگی‌های اقلام نوع t_k یا ویژگی‌های آیتم‌هایی از نوع t_j هستند که در آن آیتم t_j با آیتم t_k در نمودار G_D مرتبط است.[۲۸]

در داده‌های گزاره‌یی، یادگیرنده‌ها مجموعه‌ی ثابتی از ویژگی‌های درونی مرتبط به هرشیء را مدل می‌کنند. در مقابل در داده‌های رابطه‌یی، یادگیرنده‌ها باید تصمیم بگیرند که چه تعدادی را مدل کنند. یعنی چه میزان از همسایگی رابطه‌یی اطراف یک آیتم می‌تواند بر توزیع احتمال ویژگی‌های یک آیتم تأثیر بگذارد.[۲۸]

در طول استنتاج، شبکه‌ی وابستگی ارتباطی از یک نمودار مدل و یک نمودار داده برای نمونه‌سازی یک نمودار استنتاج $G_I = (V_I, E_I)$ در فرایندی که «گسترده شدن» نامیده می‌شود، استفاده می‌کند. G_I وابستگی‌های احتمالاتی بین همه‌ی متغیرها در یک مجموعه آزمون را نشان می‌دهد. ساختار G_I توسط هر دو نمودار G_D و G_M تعیین می‌شود. در اینجا، G_D متفاوت از نمودار داده‌ی مورد استفاده برای آموزش است. شبکه‌ی وابستگی ارتباطی از یک نمودار مدل جهت‌دار G_M با مجموعه‌یی از توزیع احتمال‌های شرطی P استفاده می‌کند. هر گره v_i عضو V_M برابر است با یک X_k^t که در آن t عضو T (مجموعه‌ی نوع‌ها) است و با یک توزیع شرطی $P(x_k^t|PA_{x_k^t})$ مرتبط است.[۲۸]

ورودی الگوریتم شامل یک نمودار داده، یک یادگیرنده‌ی رابطه‌یی شرطی R و یک مجموعه از پرس‌وجوها به‌نام Q_t که همسایگی در نظرگرفته شده در R برای هر نوع t را مشخص می‌کند. برای هر نوع t و هر ویژگی از هر نوع (X_k^t)، الگوریتم از R برای یادگیری یک توزیع احتمال شرطی (CPDs) برای X_k^t با داشتن مجموعه ویژگی‌های همسایگی آن استفاده می‌کند؛ این همسایگی توسط Q مشخص می‌شود. از توزیع‌های احتمال شرطی حاصل، برای شکل دادن به G_M استفاده می‌شود.

یادگیرنده‌ی رابطه‌یی شرطی R برای برآورد پارامترها و یادگیری ساختار در شبکه‌های وابستگی ارتباطی کاربرد دارد. متغیرهای انتخاب شده توسط R در یال‌های G_M بازتاب پیدا می‌کند. اگر R همه‌ی ویژگی‌های موجود را انتخاب کند، شبکه‌ی وابستگی ارتباطی متصل کامل خواهد بود.[۲۸]

نمودار استنتاج شبکه‌ی وابستگی ارتباطی بسیار بزرگ‌تر از نمودار داده‌ی اصلی است. برای ساخت G_I مجموعه توزیع‌های احتمال شرطی الگو روی نمودار داده مجموعه آزمون گسترده می‌شود. از نمونه‌گیری گیبز برای استنتاج در شبکه‌ی وابستگی ارتباطی استفاده می‌شود. نمونه‌گیری گیبز به رویه‌یی اشاره می‌کند که در آن یک ترتیب تصادفی از متغیرها انتخاب می‌شود و هر متغیر به یک مقدار دلخواه مقداردهی می‌شود. سپس هر متغیر مکرراً به‌ترتیب ملاقات می‌شود و مقدار آن براساس توزیع شرطی آن مقداردهی مجدد می‌شود. برآورد یک توزیع توأم باگسترش مدل G_M روی مجموعه داده‌ی هدف G_D و شکل‌دهی نمودار G_I شروع می‌شود. ارزش‌های همه‌ی متغیرهای مشاهده نشده به ارزش‌های برگرفته از توزیع‌های پیشین که به‌طور تجربی از داده‌های آموزش یادگیری شده‌اند، مقداردهی اولیه می‌شوند. سپس نمونه‌گیری گیبز به‌طور تکرارشونده هر متغیر مشاهده نشده را براساس توزیع شرطی محلی آن، با داشتن وضعیت فعلی باقی‌مانده نمودار، مجدداً برچسب‌زنی می‌کند. پس از تعداد تکرار مناسب، ارزش‌ها از یک توزیع پایدار استخراج می‌شود و می‌توان نمونه‌ها را برای برآورد احتمالات استفاده کرد.[۲۹]

۵. اجرا و ارزیابی روش معرفی شده

روش معرفی شده در این مطالعه با استفاده از نرم‌افزار متن‌باز «پراکسیمیتی[۱۱]» با استفاده از یک پردازنده‌ی Core(TM) i5, ۱/۷۳ GHz و RAM ۴/۰۰ GB روی مجموعه داده‌های آمازون[۱۲][۶] و epinions [۱۳] اجرا شد.

۱.۵. مجموعه داده‌ی آمازون

در جدول ۱ جزئیات مربوط به مجموعه داده‌های آمازون، یکی از بزرگ‌ترین و موفق‌ترین وب‌سایت‌های تجارت الکترونیکی، آمده است. از این داده‌ها ۲۰۰۰ محصول به‌طور تصادفی انتخاب، و نظرات و کاربران مربوط به این محصولات استخراج شد.

مطابق مرحله‌ی اول روش، داده‌ها آماده‌سازی شد و مجموعه ویژگی‌های اصلی و مشتق شده برای مجموعه داده آمازون استخراج شد: ویژگی‌های کاربر (شامل رتبه‌ی کاربر، میانگین امتیازات داده شده توسط کاربر، و میانگین طول نظرات ارائه شده توسط کاربر)، ویژگی‌های محصول (شامل کیفیت محصول بر اساس میانگین امتیاز داده شده به محصول، محبوبیت محصول بر اساس تعداد نظرات هر محصول، قیمت)، ویژگی‌های نظرات (شامل موقعیت در ترتیب نظرات، تعداد بازخوردها، تعداد بازخوردهای مثبت، درصد بازخوردهای مثبت، امتیاز، طول عنوان نظر، طول متن نظر، اولین نظر بودن، تنها نظر بودن)، یک ویژگی دودویی که نشان می‌دهد آیا امتیاز داده شده ۱ یا ۵ هست یا خیر و پیچیدگی متن نظر که بر اساس شاخص خوانایی خودکار طبق فرمول ۵ محاسبه می‌شود و میزان قابل درک بودن متن را مشخص می‌کند. C تعداد حروف، W تعداد کلمات و S تعداد جملات متن است:[۱۳]

$$ARI = ۴/۷۱(C/W) + ۰/۵(W/S) = ۲۱/۴۳ \qquad (۵)$$

راهبرد اول شرح داده شده در بخش قبل، برای برچسب‌زنی داده‌های آمازون کاربرد دارد. شاخص‌های رفتاری روی مجموعه داده‌ها محاسبه می‌شود. مقادیر آستانه برای

www.SID.ir

Archive of SID

جدول ۱. مشخصات مجموعه داده‌ی آمازون.

	تعداد	ویژگی‌ها
نظرات	۵۸۳۸۹۲۳	شناسه کاربر، شناسه محصول، تاریخ، تعداد بازخوردها، تعداد بازخوردهای مثبت، امتیاز، عنوان نظر، متن نظر
کاربران	۲۲۳۷۷۶۵	شناسه، نام کاربری، نام، رتبه، تاریخ تولد، موقعیت مکانی
کالاها	۶۷۲۶۶۴۹	شناسه، نام، نوع، برند، قیمت، شرح کالا

جدول ۲. مقادیر آستانه برای برچسب زنی.

معیار	امتیازدهی چندتایی	انحراف از امتیاز متوسط	تعداد نظرات در یک روز	پنجره فعالیت
مقدار آستانه	بیشتر از ۲٫۰۱	بیشتر از ۲٫۵ امتیاز	بیشتر از ۵ نظر	کمتر از ۲۸ روز

ویژگی‌های رفتاری، روی یک مجموعه برچسب‌دار موجود یادگیری شده است.[۱۲] مقدار آستانه برای شاخص‌های رفتاری مطابق جدول ۲ در نظر گرفته می‌شود.[۱۲]

در مرحله‌ی دوم، شبکه‌یی از نظرات برای مجموعه آمازون معرفی می‌شود. این کار در الگوریتم RDN با تعریف یک پرس‌وجو انجام می‌گیرد. پرس‌وجو معیار تطابق را برای یک شیء مرکزی و همسایگی ارتباطی آن مشخص می‌کند. بر اساس این پرس‌وجو یک زیرنمودار از کل نمودار داده ایجاد می‌شود (شکل ۲). در این شبکه «نظر اصلی» شیء مرکزی است که با دو موجودیت کاربر و محصول صریحاً در ارتباط است. علاوه بر ارتباطات فوق، ارتباط بین نظرات براساس معیار ضمنی کاربر و محصول مشترک نیز به‌کار گرفته می‌شود.

برای بررسی کارایی شبکه‌ی معرفی شده، شبکه‌یی بدون ارتباط نظرات تشکیل شد. این حالت فقط محصول و کاربر را برای همسایگی هر نظر مورد استفاده قرار می‌دهد. این شبکه که به‌عنوان مبنایی برای ارزیابی شبکه ارتباط نظرات استفاده می‌شود، در شکل ۳ نشان داده شده است.

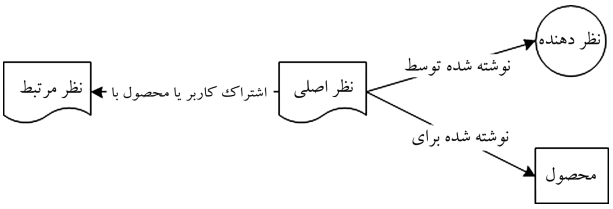
الگوریتم RDN برای دو شبکه اجرا می‌شود. در شکل ۴ RDN حاصل برای شبکه‌ی اصلی، و در شکل ۵ RDN حاصل برای شبکه‌ی مبنا نشان داده شده است.

الگوریتم RDN وابستگی‌ها را برای شمول در مدل انتخاب می‌کند. کمان‌های داخل قاب وابستگی‌های بین ویژگی‌های یک شیء واحد را بیان می‌کند. به‌طور مثال امتیاز یک نظر بر ویژگی تقلبی بودن نظر تأثیرگذار است. کمان‌هایی که از مرزهای قاب عبور می‌کنند، وابستگی‌های بین ویژگی‌های اشیاء مرتبط را نشان می‌دهند.

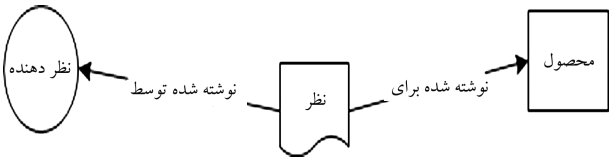
۲.۵. مجموعه داده‌ی Epinions

این مجموعه از داده‌ها اطلاعات غنی‌تری از ارتباطات را در اختیار می‌گذارد. جزئیات مربوط به مجموعه داده در جدول ۳ آمده‌است. از این مجموعه تعداد ۲۰۰۰ نظر به همراه سایر موجودیت‌های مرتبط به طور تصادفی انتخاب شد.

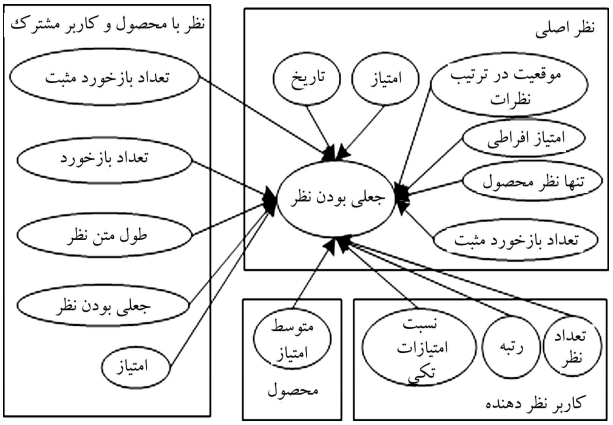
برچسب زنی با استفاده از متن نظرات انجام شده است. در این روش از یک



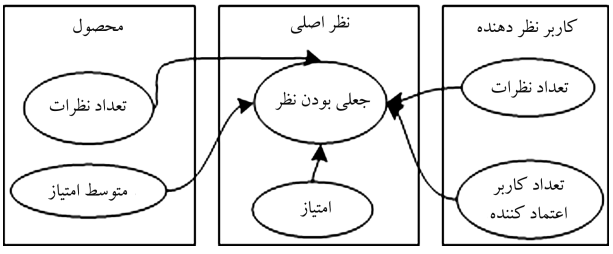
شکل ۲. شبکه‌ی اصلی روی داده آمازون.



شکل ۳. شبکه مبنا روی داده آمازون.



شکل ۴. RDN حاصل برای شبکه اصلی آمازون.



شکل ۵. RDN حاصل برای شبکه مبنا.

جدول ۳. مشخصات مجموعه داده Epinions.

موجودیت	تعداد	ویژگی‌ها
نظرات	۱۰۲۳۷۸	شناسه نظر، شناسه کاربر، شناسه محصول، متن نظر
بازخوردها	۱۳۶۶۸۳۲۰	شناسه نظر، شناسه کاربر، بازخورد: (۱. not helpful، ۲. somwhat helpful، ۳. helpful، ۴. very helpful، ۵. most helpfull)، وضعیت (تمایل کاربر به نمایش یا عدم‌نمایش بازخورد ارائه شده)، تاریخ ایجاد، تاریخ ویرایش
اعتماد	۸۴۱۳۷۲	شناسه کاربر اعتماد (مسدود) کننده، شناسه کاربر اعتماد (مسدود) شونده، ارزش اعتماد، تاریخ اعتماد

www.SID.ir

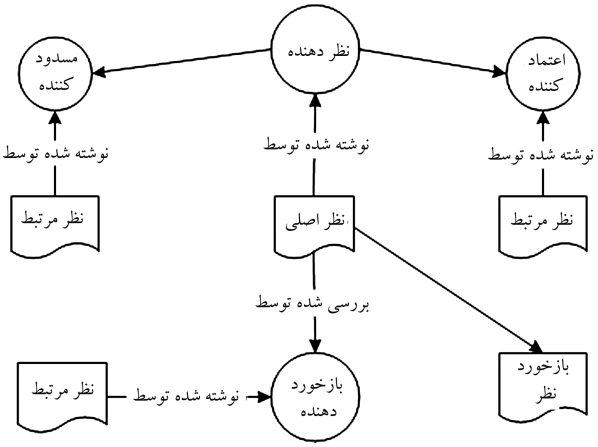
Archive of SID

مجموعه برچسب‌دار استفاده می‌شود. این مجموعه با استفاده از رویکرد جمع‌سپاری توسط مایل‌اَت و همکارانش تهیه شده است و تنها مجموعه‌ی برچسب‌دار موجود در زمینه‌ی نظرات برخط است.[۳و۲۶] این داده‌ها حاوی ۱۶۰۰ متن نظر است که ۸۰۰ نظر واقعی و ۸۰۰ نظر جعلی است. از این مجموعه به‌عنوان داده‌های آموزش استفاده شده و مدلی را با الگوریتم SVM و مجموعه ویژگی دنباله کلمات مجاور دوکلمه‌یی[۳] آموزش داده شد. این مدل روی مجموعه داده‌ی epinions برای کسب برچسب‌ها اعمال شد.

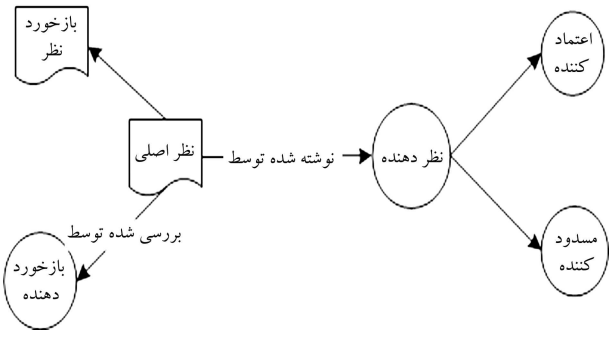
دومین شبکه معرفی شده در این مطالعه در شکل ۶ نمایش داده شده است. سه نوع از ارتباطات نظر ـ نظر و کاربر ـ کاربر با معیارهای ضمنی اعتماد، عدم‌اعتماد و بازخورد در این شبکه وجود دارد. افزون بر این از ارتباطات صریح نظر ـ کاربر و نظر ـ بازخورد نیز استفاده می‌شود. بنابراین تمامی ارتباطات فوق به‌صورت مجتمع کاربرد دارد.

برای مقایسه‌ی تأثیر ارتباطات بر کارایی روش، حالتی از شبکه که تنها ارتباط کاربران را براساس اعتماد، عدم‌اعتماد و بازخورد به‌حساب آورده است، به‌عنوان مبنا ایجاد شد. این شبکه که هیچ ارتباطی از نوع نظر ـ نظر در آن مورد استفاده نیست، در شکل ۷ شبکه نشان داده شده است.

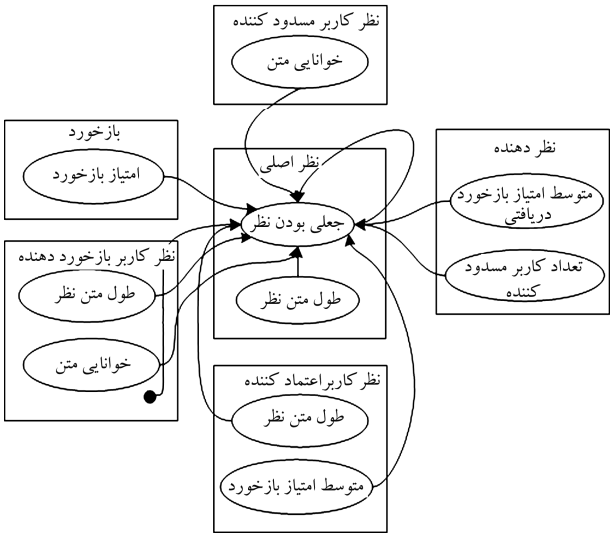
شکل‌های ۸ و ۹، RDNهای یادگیری شده را برای حالت‌های مختلف شبکه به تصویر می‌کشند. هرکدام از اشیاء در یک قاب جداگانه نشان داده شده است. هر کمان نشان‌دهنده‌ی یک وابستگی است. در شکل ۸ کمانی که از دایره‌ی توپر در موجودیت «نظر کاربر بازخورددهنده» به ویژگی جعلی بودن نظر رفته است، به‌معنی وابستگی برچسب جعلی‌بودن نظر به تعداد نظرات کاربر بازخورددهنده‌ی نظر است. با این که تعداد نظرات کاربر بازخورددهنده از ویژگی‌های



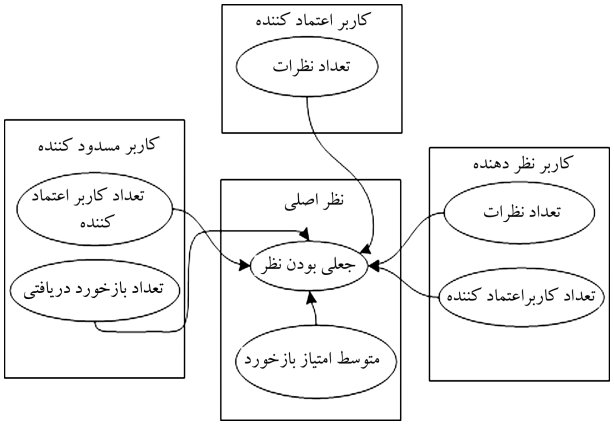
شکل ۶. شبکه‌ی اصلی برای داده‌ی epinions.



شکل ۷. شبکه‌ی مبنا برای داده‌ی epinions.



شکل ۸. RDN حاصل برای شبکه‌ی اصلی.



شکل ۹. RDN حاصل برای شبکه‌ی مبنا.

نظرات نبوده، الگوریتم آن را به‌عنوان یک ویژگی ساختاری از شبکه‌ی یادگیری کرده است.

۶. نتیجه‌گیری

نتایج حاصل برای معیارهای ACC و AUC در دو مجموعه داده به‌صورت نمودار در شکل‌های ۱۰ و ۱۱ آمده است. منحنی ROC یک نمودارگرافیکی است که از طریق رسم نسبت TPR در برابر FPR در آستانه‌های مختلف به دست می‌آید. معیار AUC با محاسبه‌ی مساحت زیر منحنی ROC، این منحنی را به یک عدد خلاصه می‌کند. معیار ACC با رابطه‌ی ۶ به دست می‌آید.

$$\text{Accuracy (Acc)} = \frac{TP + TN}{TP + FP + TN + FN} \qquad (۶)$$

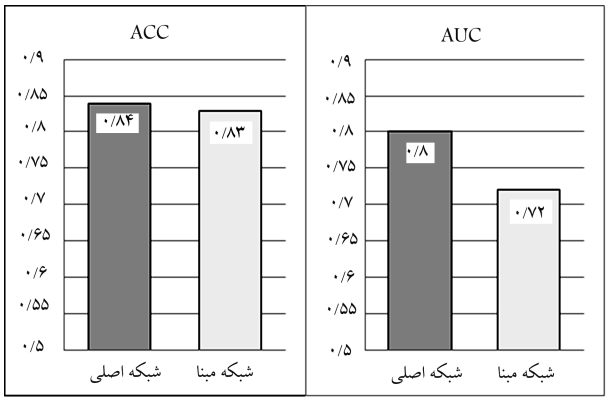
چنان که در شکل‌های ۱۰ و ۱۱مشاهده می‌شود، درداده‌های آمازون با بهره‌گیری از تأثیرات شبکه‌یی نظرات مرتبط، کارایی مدل در شبکه‌ی اصلی نسبت به حالت مبنا برتری قابل ملاحظه‌یی دارد. این نتیجه نشان‌گر وجود رابطه بین نظرات یک کاربر مشترک و نیز نظرات کاربران مختلف روی نظرات یک محصول است. چرا که کاربر متقلبی که برای ارائه‌ی نظرات جعلی پاداش دریافت می‌کند، اغلب نظرات جعلی ثبت می‌کند، هرچند ممکن است برخی نظرات آن‌ها صحیح باشد. از طرفی

www.SID.ir

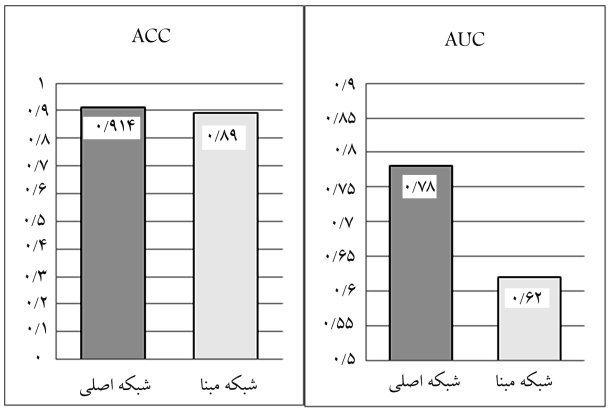
Archive of SID

جدول ۴. مشاهدات در داده‌ها.

مجموعه داده	معیار	مشاهده در نظرات جعلی تشخیص داده شده توسط روش پیشنهادی	مشاهده در نظرات غیرجعلی تشخیص داده شده توسط روش پیشنهادی
آمازون	میانگین درصد تعداد بازخوردهای مثبت	۴۱	۶۷
	درصد اولین نظرات	۳۶	۱۷
	درصد نظرات تکی	۱۳	۹
epinions	میانگین درصد تعداد بازخوردهای مثبت	۶۰	۹۱
	میانگین درصد تعداد بازخوردهای منفی	۱۲	۲
	میانگین امتیاز بازخورد نظرات	۲/۴۵	۴/۱۲۴



شکل ۱۰. نتایج داده آمازون.



شکل ۱۱. نتایج داده **epinions**.

این نتایج مؤید فعالیت گروهی متقلبان است. متقلبان برای تأثیرگذاری هرچه بیشتر بر امتیاز کلی محصول، معمولاً به‌صورت گروهی و در حملات کوتاه‌مدت روی یک محصول هدف اقدام می‌کنند.

نتایج هر دو شبکه‌ی epinions در معیار ACC بالای ۰/۸۹ قرار می‌گیرد که نشان‌دهنده‌ی کارایی کلی روش است. در معیار AUC که معیار بهتری برای مقایسه‌ی مدل‌هاست،[۳۰] شبکه‌ی اصلی که از همه‌ی ارتباطات بهره می‌گیرد، نتیجه‌ی بهتری به دست آورده است. تفاوت چشم‌گیر در نتایج حاصل بین شبکه‌ی معرفی شده‌ی اصلی و شبکه‌های مبنا در مجموعه داده‌ی epinions نیز مفید بودن بهره‌گیری از نظرات مرتبط را سودمند ارزیابی می‌کند.

برای نشان‌دادن کارایی پیش‌بینی برخی مشاهدات در نتایج را بررسی می‌کنیم.

جدول ۵. مقایسه‌ی نتایج حاصل از روش پیشنهادی با مطالعات پیشین.

شماره مرجع	مجموعه داده	برچسب‌زنی	بهترین حالت (٪)	
			ACC	AUC
[۵]	آمازون	تکرار نظرات	-	۷۸
[۸]	آمازون	دستی	-	۹۵
[۹]	TripAdvisor	جمع‌سپاری	۷۴/۴	-
[۷]	Yelp.com	ابتکاری	۸۶/۵	-
[۳]	آمازون	جمع‌سپاری	۸۹/۸	-
[۱۶]	TripAdvisor	جمع‌سپاری	۹۱/۲	-
[۱۲]	آمازون	دستی	۷۷/۴	-
[۲۶]	آمازون	جمع‌سپاری	۸۹/۳	-
[۱۷]	آمازون	جمع‌سپاری	۷۶/۳	-
[۱۱]	آمازون	ابتکاری	۷۷/۶	-
روش پیشنهادی	آمازون	ابتکاری	۸۴	۸۰
روش پیشنهادی	Epinions	ابتکاری	۹۱/۴	۷۸

این مشاهدات در جدول ۴ آمده است. نتایج نشان می‌دهد که نظرات تشخیصی به‌عنوان جعلی توسط روش پیشنهادی، کم‌تر از نظرات غیرجعلی بازخورد مثبت دریافت کرده‌اند. نتایج نشان می‌دهد که نظرات نخستین یا یکه درمورد یک محصول با احتمال بیشتری مشکوک به تقلب هستند، چراکه متقلبان به محض آماده شدن یک محصول نظرات خود را ارائه می‌دهند تا بیشترین تأثیر را برکاربران بعدی داشته باشند. وجود درصد بالاتری از این نوع نظرات بین نظرات جعلی نسبت به موارد صحیح در نتایج روش پیشنهادی نشان‌دهنده‌ی صحت نتایج است. امتیاز بازخورد نظرات که عددی بین ۱ تا ۵ است، به‌طور متوسط برای نظرات جعلی به‌طور چشم‌گیری پایین‌تر از نظرات واقعی است.

در جدول ۵ نتایج حاصل از مطالعات پیشین که معیار AUC و ACC برای آن‌ها محاسبه شده ارائه شده است. این نتایج گویای آن است که روش پیشنهادی در معیار ACC بهبود قابل توجهی نسبت به سایرین دارد. اگرچه روش معرفی شده در دیگر مطالعات[۸] در معیار AUC نتایج بهتری از روش پیشنهادی به دست آورده، این روش فعالیت متقلبانه‌ی خاصی را که به‌صورت گروهی رخ داده بررسی می‌کند. این روش برکشف گروه‌های متقلب تمرکز دارد درحالی که برای کشف متقلب‌های تکی و همچنین تشخیص نظرات جعلی کاربردی نیست.

www.SID.ir

Archive of SID

با استفاده از نتایج این تحقیق، نظرات مشتریان الکترونیکی قبل از خلاصه‌سازی و نتیجه‌گیری از متن نظرات پالایش می‌شود که با کمک به مشتریان برای اخذ تصمیمات بهتر، منجر به جلب و حفظ اعتماد مشتریان الکترونیکی با حذف نظرات جعلی می‌شود. از سوی دیگر، نتایج سیستم‌های کاوش نظرات مشتریان با حذف نظرات جعلی مؤثرتر است. شناسایی و مقابله با حملات و خسارت‌های احتمالی از جانب رقبا برای فروشندگان برخط از دیگر کاربردهای این تحقیق است.

درادامه‌ی این مطالعه می‌توان از روش‌های طبقه‌بندی جمعی همچون استکینگ[14] و بگینگ[15] برای ترکیب ویژگی‌ها و روابط و نیز ترکیب چندین رابطه استفاده کرد. همچنین تحلیل متن بازخوردها در توسعه‌ی سیستم تشخیص نظرات جعلی می‌تواند بسیار امیدبخش باشد. سیستم‌های ضد تقلب با راهبرد تشخیص، به‌دنبال الگوهای تقلب در داده‌ها پس از وقوع آن هستند. درحالی که راهبرد پیش‌گیری از ورود متقلبان به سیستم قبل از انتشار محتوای متقلبانه ممانعت می‌کند. توسعه‌ی روش‌های پیش‌گیری که اختلالی در سیستم‌های نظرات برخط ایجاد نکند، کم‌تر مورد بحث مطالعات پیشین بوده است و می‌تواند موضوع تحقیقات آتی باشد.

پانوشت‌ها

1. online
2. review spam
3. relational dependency network (RDN)
4. frequent itemset mining (FIM)
5. temporal pattern discovery
6. multi dimensional time series
7. loopy belief propagation (LBP)
8. n-gram
9. part of speech (POS)
10. Mechanical Turk
11. available at https://kdl.cs.umass.edu/display/public/Proximity
12. available at www.cs.uic.edu/ liub/FBS/fake-reviews.html
13. available at http://www.sfu.ca/ sam39/Datasets/EpinionsReviews/
14. Stacking
15. Bagging

منابع (References)

1. Ma, Y. and Li, F. "Detecting review spam: Challenges and opportunities", in Collaborative Computing, *8th International Conference on Networking, Applications and Worksharing* (2012)
2. Jindal, N., Liu, B. and Lim, E.-P. "Finding unusual review patterns using unexpected rules", *in Proceedings of the 19th ACM International Conference on Information and Knowledge Management* (2010).
3. Ott, M., Choi, Y., Cardie, C. and Hancock, J.T. "Finding deceptive opinion spam by any stretch of the imagination", *in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics* (2011).
4. Akoglu, L., Chandy, R. and Faloutsos, C. "Opinion fraud detection in online reviews by network effects", *In proceeding of seventh International AAAI Conference on weblogs and Social Media* (2013).
5. Jindal, N. and Liu, B. "Opinion spam and analysis", *in Proceedings of the International Conference on Web Search and Web Data Mining* (2008).
6. Li, F., Huang, M.Y. and Zhu, X. "Learning to identify review spam", *in Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, **3**, pp.2488-2493 (2011).
7. Mukherjee, A., Venkataraman, V., Liu, B. and Glance, N. "Fake review detection: Classification and analysis of real and pseudo reviews", Technical Report, Department of computer Scierce (UIC-CS-2013-3) University of Illinois Chicago (2013).
8. Mukherjee, A., Liu, B. and Glance, N. "Spotting fake reviewer groups in consumer reviews", *in Proceedings of the 21st International Conference on World Wide Web* (2012).
9. Feng, S., Xing, L., Gogar, A. and Choi, Y. "Distributional footprints of deceptive product reviews", in ICWSM, The 6th International AAAI Conference on weblogs and Social media (2012).
10. Xie, S., Wang, G., Lin, S. and Yu, P.S. "Review spam detection via temporal pattern discovery", *in Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2012).
11. Fei, G., Mukherjee, A., Liu, B., Hsu, M., Castellanos, M. and Ghosh, R. "Exploiting burstiness in reviews for review spammer detection", *in Seventh International AAAI Conference on Weblogs and Social Media* (2013).
12. Mukherjee, A., Kumar, A., Liu, B., Wang, J., Hsu, M., Castellanos, M. and Ghosh, R. "Spotting opinion spammers using behavioral footprints", *in Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2013).
13. Hu, N., Bose, I., Koh, N.S. and Liu, L. "Manipulation of online reviews: An analysis of ratings, readability, and sentiments", *Decision Support Systems*, **52**(3), pp. 674-684 (2012).
14. Lim, E.-P. Nguyen, V.-A. Jindal, N. Liu B. and Lauw, H.W. "Detecting product review spammers using rating behaviors", *in Proceedings of the 19th ACM International Conference on Information and Knowledge Management* (2010).
15. Jiang, B. and Chen, B. "Detecting product review spammers using activity model", *in International Conference on Advanced Computer Science and Electronics Information* (2013).

www.SID.ir

Archive of SID

16. Feng, S., Banerjee, R. and Choi, Y. "Syntactic stylometry for deception detection", *in Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, Short Papers, **2**, pp.171-175 (2012).
17. Harris, C. "Detecting deceptive opinion spam using human computation", in Workshops at AAAI on Artificial Intelligence (2012).
18. Xi, Y. "Chinese review spam classification using machine learning method", *in Control Engineering and Communication Technology, 2012 International Conference on* (2012).
19. Wang, G., Xie, S., Liu, B. and Yu, P.S. "Identify online store review spammers via social review graph", *ACM Transactions on Intelligent Systems and Technology (TIST)*, **3**(4), pp. 61 (2012).
20. Duan, H. and Zirn, C. "Can we identify manipulative behavior and the corresponding suspects on review websites using supervised learning?", Secure IT Systems, pp. 215-230 (2012).
21. Lai, C., Xu, K., Lau, R.Y., Li, Y. and Song, D. "High-order concept associations mining and inferential language modeling for online review spam detection", *in Data Mining Workshops, 2010 IEEE International Conference on* (2010).
22. Liu L. and Wang, Y. "A method for sorting out the spam from Chinese product reviews", *in Consumer Electronics, Communications and Networks, 2012 2nd International Conference on* (2012).
23. Wu, G., Greene, D., Smyth, B. and Cunningham, P. "Distortion as a validation criterion in the identification of suspicious reviews", *in Proceedings of the First Workshop on Social Media Analytics* (2010).
24. Gambetta, D. "Can we trust trust?", in Trust: Making and Breaking Cooperative Relations, pp. 213-237 (1988).
25. Molavi Kakhki, A., Kliman-Silver, C. and Mislove, A. "Iolaus: Securing online content rating systems", *in Proceedings of the 22nd International Conference on World Wide Web* (2013).
26. Ott, M., Cardie, C. and Hancock, J.T. "Negative deceptive opinion spam", *in Proceedings of NAACL-HLT* (2013).
27. Larue, T.H. "Trust: The connecting link between organizational theory and philosophical ethics", *Academy of Management Review*, **20**(2), pp. 379-403 (1995).
28. Neville J. and Jensen, D. "Relational dependency networks", *The Journal of Machine Learning Research*, **8**, pp. 653-692 (2007).
29. Neville, J. and Jensen, D. "Dependency networks for relational data", *in Proceedings of the Fourth IEEE International Conference on Data Mining* (2004).
30. Huang, J. and Ling, C.X. "Using AUC and accuracy in evaluating learning algorithms", *IEEE Transactions on Knowledge and Data Engineering*, **17**(3), pp. 299-310 (2005).

www.SID.ir