بررسی روش‌های تشخیص هرزنامه‌ها در شبکه‌های اجتماعی با استفاده از داده‌کاوی

صالح بیت شیخ احمد[۱] و مهناز رفیعی[۲]
[۱]دانشجوی کارشناسی ارشد، گروه کامپیوتر، موسسه آموزش عالی غیرانتفاعی اروندان خرمشهر، salehste@gmail.com
[۲]استادیار، گروه کامپیوتر، واحد رامهرمز، دانشگاه آزاد اسلامی، رامهرمز، ایران، m.rafie@srbiau.ac.ir

چکیده - شبکه‌های اجتماعی به مجموعه‌ای از افراد گفته می‌شود که به صورت گروهی با یکدیگر ارتباط داشته و مواردی مانند اطلاعات، نیازمندی‌ها، فعالیت‌ها و افکار خود را به اشتراک می‌گذارند. ولی به وجودآمدن پیام‌های ناخواسته موجب آزار کاربران و پایین آمدن کارایی شده است. امروزه این پیام‌های ناخواسته که به عنوان هرزنامه شناخته می‌شوند به مشکل عمده ای تبدیل شده‌اند. هرزنامه‌ها منجر به اتلاف منابع شبکه، کامپیوترها و زمان می‌گردند. لذا جهت شناسایی و جلوگیری از هرزنامه‌ها تلاش‌های زیادی صورت گرفته است، ولی با پیشرفت این تلاش‌ها هرزنامه‌ها باز هم گسترش و پیشرفت می‌کنند. بنابراین هرزنامه به یک مشکل جدی تبدیل شده است. بدین منظور روش‌ها و الگوریتم‌های بسیاری جهت جلوگیری و فیلتر هرزنامه پیشنهاد شده است که هدف اصلی افزایش دقت روش تشخیص هرزنامه در پیام‌ها است. این مقاله، فعالیت ضد هرزنامه و کاربردهای آن در شبکه‌های اجتماعی را با استفاده از الگوریتم های داده کاوی مورد بررسی قرار داده است.

کلید واژه- هرزنامه، ماشین بردار پشتیبان، ویژگی کاربران، ارتباطات

۱. مقدمه

شبکه‌های اجتماعی به مجموعه‌ای از افراد گفته می‌شود که به‌صورت گروهی با یکدیگر ارتباط داشته و مواردی مانند اطلاعات، نیازمندی ها، فعالیت ها و افکار خود را به اشتراک می‌گذارند. امروزه، شبکه‌های اجتماعی برخط، ابزار بسیار محبوبی برای همکاری و ارتباط به‌شمار می‌روند که میلیون ها نفر از کاربران اینترنت را به خود جلب کرده‌اند. شبکه‌های اجتماعی برخط مانند فیسبوک، لینکداین و توییتر از پرطرفدارترین این برنامه‌ها به‌حساب می‌آیند. این شبکه‌ها، به‌ویژه آن‌هایی که کاربرد های معمولی و غیرتجاری دار ند، مکان هایی در دنیای مجازی هستند که مردم خود را به‌طور خلاصه معرفی کرده و امکان برقراری ارتباط بین خود و هم‌فکرانشان را در زمینه‌های مورد علاقه فراهم می‌کنند [۱]. اما زندگی در چنین فضایی آدابی دارد که باید به آن‌ها توجه ویژه‌ای نمود تا از مشکلات جدی جلوگیری شود. در سال های اخیر، نظرات جعلی و هرزنامه مشکلی است که به‌شدت در حال گسترش و افزایش است. هرزنامه‌ها، پیام‌های ناخواسته‌ای هستند که به‌طور گسترده به تعداد زیادی از کاربران با هدف کلاهبرداری، نشر اکاذیب، ایجاد شایعات، رعب و وحشت و همچنین تبلیغات ارسال می‌گردد. امروزه در حدود ۵۰ درصد نامه‌های کاری الکترونیکی، هرزنامه هستند [۲]. هرزنامه سبب مشکلاتی مانند اشغال منابع، اتلاف پهنای‌باند و طولانی‌شدن زمان ارتباط می‌شود [۳]. افرادی که اقدام به ارسال پیام‌های هرزنامه می‌نمایند، هرزنامهر نامیده می شوند. توییتر به‌عنوان یکی از بزرگ‌ترین شبکه‌های اجتماعی محسوب می شود که روزانه پیغام‌های متنی زیادی در آن تولید و ارسال می‌شود. به این پیغام‌ها توییت گفته می‌شود. توییت‌ها معمولاً دارای ساختار متفاوتی با پیغام‌های سایر شبکه‌های اجتماعی هستند به‌طوری‌که تشخیص هرزنامه در آنها پیچیده‌تر است. طبق مطالعاتی که انجام شده، حدود ۱۵ درصد کاربران توییتر بات‌ها هستند و معمولاً از هر ۲۰ توییت، یکی از آن‌ها هرزنامه است. لذا شبکه اجتماعی مزبور در این مقاله بیشتر مورد توجه واقع شده است. شایان ذکر است که تشخیص هرزنامه‌ها با استفاده از تکنیک‌های داده کاوی یکی از دغدغه‌های اصلی

کاربران و مدیران شبکه‌های اجتماعی است، لذا تاثیر الگوریتم هایی همچون خوشه بندی، ماشین بردار پشتیبان، درخت تصمیم، بیزین و K-NN نیز مورد بررسی قرار گرفته است.

ساختار کلی مقاله به این ترتیب ا ست که در بخش دوم مبانی و پیشینه تحقیق بیان می شود و به برر سی کارهای انجام شده در این زمینه، معایب و مزایای آن می پردازد. بعد از برر سی کارهای انجام شده، بخش سوم به معرفی جزئیات روش پیشنهادی می‌پردازد. در نهایت در بخش چهارم، نتایج نهایی بیان می شود و راهکارهایی برای ادامه تحقیق و بهبود آن در آینده معرفی می‌گردد.

۲. روش‌های تشخیص هرزنامه‌ها

تحقیقات انجام گرفته نشان می‌دهد که توییت‌های جعلی با همان سرعت توییت‌های نرمال منتشر می‌شوند. ربات‌ها نقش زیادی در سرعت بخشیدن به گسترش هرزنامه‌ها دارند و کاربران انسانی هم به افزایش حجم آن‌ها کمک می‌کنند. برای تشخیص هرزنامه‌ها کارهای متعددی انجام شده است که در ادامه به معرفی چند مورد می‌پردازیم.

کار انجام شده توسط وانگ را می‌توان پیشگامی در تشخیص هرزنامه‌ها در توییتر بر اساس شناسایی ارتباط بین کاربران دانست. در این روش با استفاده از مدل گراف مستقیم وجود ارتباطات دنبال کننده-دوستی بین کاربران شناسایی می‌شود و با استفاده از ویژگی‌های متنی استخراج شده، هرزنامه‌ها شناسایی می‌شوند. به صورت کلی می‌توان کارهای انجام گرفته برای تشخیص هرزنامه‌ها در توییتر را به چند دسته‌ی کلی تقسیم‌بندی کرد: تحلیل ساختار گراف شبکه‌های اجتماعی مانند روش‌های [۶-۴]، تحلیل ساختار متن و استخراج الگوها مانند [۷]، تحلیل متا دیتای مربوط به پروفایل کاربران و بررسی URLهای استفاده شده مانند [۱۰-۸] و تحلیل رفتارهای تعاملی کاربران [۱۳-۱۱].

۲-۱ ویژگی‌های آماری مبتنی بر تشخیص زمان واقعی اسپم توییتر

در تحقیق [۱۴]، بر روی داده‌های توییتر با پنج هدف: ارائه فیلترینگ بلادرنگ، افزایش مقیاس‌پذیری، تصمیم‌گیری دقیق، قابلیت آموزش مجدد مدل با داده‌های جدید و طبقه‌بندی مستقل از متن پژوهشی انجام شده است. در این پژوهش از یک مجموعه داده‌ی ۵۰۰۰۰۰تایی استفاده شده است که ۴۰۰۰۰۰ عدد از آن‌ها برای داده‌های آموزشی و ۱۰۰۰۰۰ عدد از آن‌ها برای آزمایش مورداستفاده قرار گرفت. داده‌های آموزشی با ترکیب یک‌به‌یک (نیمی نرمال و نیمی هرزنامه) یک‌به‌چهار و یک‌به‌ده موردبررسی قرار گرفتند. دقت حاصل ۹۱ درصد است. مزیت این پژوهش این است که سیستمی بلادرنگ است که می‌تواند با تأخیر کوتاهی عمل فیلترینگ را انجام دهد. مقیاس‌پذیری بالایی دارد و می‌تواند برای داده‌های حجیم به‌درستی عمل کند. عیب این پژوهش عدم دسترسی تمامی مجموعه داده‌ها است و الگوریتم‌های دیگر را موردبررسی قرار نداده است. ازآنجایی‌که در مجموعه داده‌های موجود، تعداد کاربران اسپم نسبت به کاربران نرمال بسیار کم هستند برای ایجاد مدل و ارزیابی آن نمی‌توان از تمامی رکوردهای موجود در مجموعه داده‌ها استفاده نمود، زیرا مدل تنها ویژگی کاربران نرمال را در نظر گرفته و کاربران اسپم را نادیده می‌گیرد. برای رفع این مشکلات باید نسبت نمونه‌های مختلف از کلاس‌های مختلف متناسب در نظر گرفته شود.

۲-۲- استفاده از یک روش داده کاوی جهت تشخیص هرزنامه در فیسبوک

در [۱۵] با تکنیک درخت تصمیم J48، با استفاده از متن پیام‌های فیسبوک، تعداد کلمات کلیدی، میانگین تعداد کلمات، طول متن و تعداد لینک‌ها هرزنامه‌ها شناسایی می‌شوند. نرخ دقت و فراخوانی به دست آمده در مقاله به ترتیب ۶۱ و ۶۳ درصد است.

۲-۳ تشخیص اسپم‌های توییتر با استفاده از یادگیری عمیق

وو و همکاران در [۱۶] از تکنیک‌های یادگیری عمیق با استفاده از تبدیل کلمه به بردار برای تشخیص و شناسایی اسپم‌ها استفاده کرده‌اند. آن‌ها بر این عقیده هستند که تنها استفاده از ویژگی‌های آماری متون نمی‌تواند دقت کافی را برای یادگیری دسته‌بندها فراهم کند. آن‌ها از تعدادی ویژگی‌های خودساخته مربوط به کاربران و ارتباط آن‌ها در توییتر استفاده کرده‌اند. این ویژگی‌ها به نحوی کدگذاری شده‌اند که بتوان آن‌ها را به‌عنوان ورودی دسته‌بند در نظر گرفت. در ادامه از شبکه عصبی عمیق برای یادگیری و تشخیص هرزنامه‌ها استفاده شده است. ویژگی‌های مربوط به کاربران به صورت خودساخته و تجربی هستند و از ساختار درونی شبکه به درستی استفاده نشده است. در روش مزبور، ویژگی‌های ارتباط بین کاربران به شکلی پیدا می‌گردد که بتوان کاربران را از

لحاظ نزدیکی به هم رتبه‌بندی نمود و حتی بتوان گفت اگر کاربری با دو نفر رابطه دوستی دارد به کدام نزدیک‌تر است.

۲-۴ پرداختن به مشکل عدم تعادل کلاس در تشخیص اسپم‌های توییتر با استفاده از یادگیری دسته جمعی

لی‌یو و همکاران در [۱۷]، یک ساختار متفاوت برای تشخیص پست‌های هرزنامه در توییتر معرفی کردند. تاکید اصلی آن‌ها بر روی ساختار داده‌ها متمرکز بود. در این تحقیق بیان شده است که ترکیب نامتوازن داده‌های اسپم و غیر اسپم باعث بروز خطا توسط دسته‌بندهای یادگیری ماشین می‌شود. به‌طوری که در واقعیت معمولاً چیزی حدود ۵ الی ۱۰ درصد پست‌ها اسپم هستند، ولی در مجموعه داده‌های استفاده شده گاهاً داده‌های اسپم و غیر اسپم به‌نسبت ۵۰ به ۵۰ تقسیم بندی می‌شوند که این می‌تواند دقت کار را کاهش دهد. آن‌ها یک فرآیند سه مرحله‌ای برای تشخیص اسپم‌ها در توییتر معرفی کرده‌اند. در ابتدا با استفاده از سه روش مختلف نمونه‌برداری از داده‌ها انجام می‌گیرد و تعدادی نمونه متشکل از هرزنامه و غیر هرزنامه برای دسته‌بند مشخص می‌شوند. یک روش بدون جایگذاری، روش دیگر با جایگذاری و روش سوم با استفاده از تکنیک نمونه‌سازی فازی انجام شده است. در مرحله دوم برای هر کدام از سه روش مطرح شده با استفاده از دسته-بندهای یادگیری ماشین مرحله دسته‌بندی انجام گرفته و پست-های هرزنامه و غیر هرزنامه شناسایی می‌شوند. در این مقاله از دسته‌بندهای ماشین بردار پشتیبان، درخت‌تصمیم و بیزین ساده استفاده شده است. در نهایت در مرحله سوم بر اساس قانون رای اکثریت، تصمیم نهایی در مورد هرزنامه یا غیر هرزنامه بودن گرفته می‌شود. این روش از لحاظ دسته‌بند ساختار قدرتمندی دارد اما ویژگی‌های استفاده شده برای دسته‌بندها فقط اطلاعات آماری خود متن است که باعث می‌شود دقت کار پایین بیاید.

۲-۵ تشخیص اسپم توییتر با استفاده از خوشه‌بندی

میلر و همکاران در سال ۲۰۱۵ از خوشه‌بندی برای تشخیص هرزنامه‌ها در توییتر استفاده کرده‌اند. در این تحقیق تشخیص هرزنامه به عنوان یک مسئله کلاس‌بندی در نظر گرفته نمی‌شود بلکه به عنوان یک مسئله تشخیص ناهنجاری[۱] در نظر گرفته شده است. در این تحقیق در کنار استفاده از ۱۲ ویژگی مربوط به متن و ساختار ارتباطی کاربران از ۹۵ کاراکتر خاص از بین کاراکترهای کد اسکی تعداد عنوان 95 one-gram استفاده شده است. بنابراین در کنار ۱۲ ویژگی اولیه، ۹۵ ویژگی دیگر اضافه شده است و خوشه‌بندی با این ۱۰۷ ویژگی انجام گرفته است. فرآیند خوشه-بندی به‌صورت جداگانه با استفاده از دو روش StreamKM++ و DenStream انجام گرفته است که توسط هر دو روش، نمونه‌ها به دو خوشه‌ی هرزنامه و غیر هرزنامه تقسیم‌بندی می‌شوند. معیار فراخوانی[۲] برای این دو روش حدود ۹۹ درصد است. یعنی در واقع ۹۹ درصد پست‌هایی که هرزنامه هستند را شناسایی می‌کند. اما مقدار False positive آن به نسبت زیاد است. یعنی تعداد زیادی از پست‌هایی که هرزنامه نیستند را نیز هرزنامه تشخیص می‌دهد و این باعث بروز مشکل خواهد شد. همچنین مجموعه داده‌ی مورد استفاده بسیار ساده بوده و توازن مناسبی بین تعداد هرزنامه و غیر هرزنامه وجود ندارد [۱۸].

۲-۶ استنتاج ماشین‌های بردار پشتیبان با استفاده از الگوریتم بهینه‌سازی وال جهت تشخیص پروفایل-های اسپم در شبکه‌های اجتماعی برخط در متن-های زبانی مختلف

الزوبی و همکاران در [۱۹]، یک فرآیند خودکار برای شناسایی کاربران اسپم معرفی کرده‌اند. با توجه به اینکه نسبت کاربران اسپم در توییتر بیشتر از نسبت توییت‌های اسپم به کل توییت‌ها است، بنابراین شناسایی کاربران اسپم می‌تواند در تشخیص هرزنامه‌ها و حتی جلوگیری از ارسال هرزنامه‌ها نقش موثری داشته باشد. در این روش از ترکیب دسته‌بند ماشین بردار پشتیبان و الگوریتم فرامکاشفه‌ای بهینه‌سازی وال استفاده شده است. در این ساختار ابتدا داده‌ها پیش‌پردازش شده و تعدادی ویژگی بر اساس ساختار متن و ارتباط کاربران شناسایی می‌شوند. داده‌های ورودی به دسته‌بند ماشین بردار پشتیبان داده می‌شود و فرایند یادگیری آغاز می‌گردد. در هر مرحله توسط الگوریتم بهینه‌سازی وال مشخص می‌گردد که آیا مقدار بهینه‌سراسری بدست آمده یا خیر. در نهایت بعد از اتمام مراحل، فرآیند یادگیری به اتمام رسیده و بر اساس داده‌های آزمایش نتیجه ارزیابی محاسبه می‌شود. در این روش از ویژگی‌های ساده استفاده شده است و همچنین پیچیدگی زیاد این روش باعث می‌شود که دقت و کارایی نمونه‌های آموزش بسیار بالا

[۱] anomaly

[۲] Recall

و نتایج آزمایش به نسبت کمتر باشد، به عبارتی وقوع بیش‌برازش[۳] در این روش بسیار محتمل است.

۲-۷ تشخیص هرزنامه به صورت نیمه نظارتی در توییتر

سدهای و همکاران در [۲۰]، از یک روش نیمه نظارتی برای تشخیص هرزنامه‌ها در توییتر استفاده کرده‌اند. در این روش یک چهارچوب بلادرنگ به نام S^3D برای تشخیص توییت‌های اسپم معرفی شده است. این چهارچوب شامل دو بخش کلی است. بخش اول به نام تشخیص هرزنامه و بخش دوم به نام بروزرسانی مدل نام‌گذاری شده است. بخش شناسایی هرزنامه خود شامل ۴ زیر بخش مختلف است: ۱- شناسایی لیست‌های سیاه و افراد گزارش شده توسط سایر کاربران. ۲- شناسایی توییت‌هایی که به صورت انبوه و تکراری ارسال می‌گردند. ۳- شناسایی کاربران معتبر، این افراد عموماً تایید شده هستند و پست‌های ارسال شده از طرف آن‌ها به صورت غیر هرزنامه دسته‌بندی می‌شوند. ۴- استفاده از یک دسته‌بند ترکیبی شامل سه دسته‌بند: بیزین ساده، لجستیک رگرسیون و جنگل تصادفی[۴]، فرآیند یادگیری و دسته‌بندی به صورت ترکیبی توسط این سه دسته‌بند انجام می‌گیرد. اما در بخش دوم بر اساس برچسب‌گذاری‌های انجام شده در بخش اول، مقادیر مربوط به پارامترها بروزرسانی می‌شوند تا فرآیند دسته‌بندی با دقت بیشتری انجام گیرد. امکان تبانی در این روش زیاد است به صورتی که اگر گروهی یک فرد را به عنوان هرزنامه گزارش کنند، چهارچوب استفاده شده، فرد را به عنوان هرزنامه شناسایی می‌کند.

۲-۸ رویکرد انتخاب ویژگی جهت شناسایی هرزنامه در شبکه اجتماعی فیس بوک

این مقاله در مورد تشخیص نظرات اسپم در شبکه اجتماعی فیس بوک است. با مرور پست‌ها، نظرات و بررسی ویژگی‌های آنها، یک سیستم فیلترشکن اسپم برخط در این مقاله طراحی شده است. سیستم فیلتر پیشنهادی قادر است از روش‌های مختلف اکتشافی و الگوریتم‌های بهینه سازی مانند بهینه‌سازی ذرات، بهینه سازی کلونی مورچه‌ها و تکامل تفاضلی برای کشف و فیلتر نمودن مطالب مخرب و جلوگیری از انتشار نظرات اسپم‌ها ، محیطی مطمئن را برای کاربران فراهم کند. این شبکه اجتماعی محبوب علاوه بر این از روش‌های یادگیری ماشینی، تکنیک‌های خوشه‌بندی و درخت تصمیم‌گیری استفاده نموده تا عملکردی دقیق و سرعتی مناسب برای سیستم فیلتر پیشنهادی ارائه دهد [۲۱].

۲-۹ سیستم تشخیص هرزنامه مبتنی بر داده کاوی برای شبکه‌های اجتماعی

در [۲۲]، یک سیستم تشخیص هرزنامه مقیاس‌پذیر بر اساس داده‌کاوی برای امنیت شبکه‌های اجتماعی پیشنهاد شده است. از الگوریتم خوشه‌بندی GAD برای خوشه‌بندی در مقیاس بزرگ و ادغام آن با الگوریتم یادگیری فعال استفاده شده است. از این الگوریتم یادگیری جهت مقابله با چالش‌های تشخیص زمان واقعی و مقیاس‌پذیری استفاده می‌شود.

۲-۱۰ دسته‌بند کارآمد جهت تشخیص هرزنامه در شبکه-های اجتماعی

در [۲۳] بر روی شناسایی عملکرد اسپم با استفاده از تجزیه و تحلیل ویژگی و استفاده از طبقه بندی کارآمد متمرکز شده است. هدف اصلی در این مقاله، یافتن رابطه بین ویژگی‌ها و الگوهای طبقه‌بندی جهت تشخیص پیام اسپم از سایت‌های ناخواسته است. این سیستم از الگوریتم‌های طبقه‌بندی کارآمد پس از تجزیه و تحلیل ارتباط ویژگی‌ها جهت شناسایی هرزنامه استفاده می‌کند. در این راستا، یک چارچوب طبقه‌بندی کارآمد برای پیش‌بینی و نظارت بر اسپم ارایه شده است. نتیجه این پژوهش برای کاربران شرکت کننده در شبکه‌های اجتماعی برای اهدافی همچون تجارت، بازاریابی و برقراری ارتباط بین مخاطبین می‌باشد. لذا عملکرد الگوریتم‌های طبقه‌بندی همچون بیزین، ID3، جنگل تصادفی، K-NN و ماشین بردار پشتیبان مقایسه شده است. کار آتی در این راستا ارایه مدلی جهت ادغام یک سیستم هوشمند در این چارچوب طبقه بندی جهت نظارت و تشخیص مستمر بر روی هرزنامه قابل انجام می باشد.

۲-۱۱ روشی ترکیبی برای شناسایی اسپم در شبکه‌های اجتماعی با استفاده از تجزیه و تحلیل گراف و رفتار کاربر

در [۲۴]، روشی مبتنی بر تحلیل گراف جهت تشخیص اسپم از طریق تجزیه و تحلیل رفتارهای آنها و ارتباط آنها با کاربران ارائه شده است. در ضمن، راه‌حلی جهت تسهیل روند تشخیص ارائه

[۳] Overfitting

[۴] Random Forest (RF)

شده است. هدف از این مقاله، استفاده از روش تحلیل گراف ترکیبی و تحلیل رفتاری، افزایش دقت تشخیص و میزان تشخیص با کمک الگوریتم‌های طبقه‌بندی مناسب و ویژگی‌های مؤثرتر است. بنابراین، دو سناریو جهت دستیابی به سطح دقت بالاتر و مثبت کاذب پایین‌تر استفاده شد. سناریوی اول مبتنی بر استفاده از کل داده‌ها برای ساخت و ارزیابی مدل است. نتایج نشان می دهد که علی‌رغم دقت بالای این رویکرد، به دلیل سطح بالای مثبت کاذب، این رویکرد مناسب نیست. در سناریوی دوم، نسبت کاربران عادی به اسپم‌ها ۲ به ۱ در نظر گرفته شد که منجر به نتایج رضایتبخش شد. پس از بررسی ماتریس سردرگمی و مثبت کاذب در الگوریتم‌های مختلف، الگوریتم لجستیک به عنوان یک الگوریتم مناسب انتخاب شده است که اهداف این مطالعه را برآورده می‌کند.

۳- راهکار پیشنهادی

تشخیص هرزنامه‌ها با استفاده از تکنیک‌های داده کاوی و یادگیری ماشین یکی از دغدغه‌های اصلی کاربران و مدیران شبکه‌های اجتماعی است. تابه‌حال کارهای زیادی در مورد تشخیص هرزنامه-ها در شبکه‌های اجتماعی انجام شده است اما برای تشخیص هرزنامه‌ها در توییتر کمتر کار شده است. ساختار اصلی اکثر روش-ها یکسان است. لذا با توجه به راهکارهای انجام شده پیشین و در راستای ارایه روشی موثر قصد داریم مکانیزمی جدید به‌منظور شناسایی توییت‌های اسپم درتوییتر معرفی نماییم [۲۵]. در این راستا نیاز است در ابتدا پیش‌پردازش‌هایی روی متون انجام گرفته و بعد از پالایش متن، تعدادی ویژگی استخراج شده و با استفاده از تکنیک‌های یادگیری ماشین مشخص می شود که کدام پیغام هرزنامه و کدام غیر هرزنامه است. شایان ذکر است که در روش‌های پیشین غالبا تنها از ویژگی‌های داخل متن برای این کار استفاده می‌شد. مثلا استفاده از کلمات خاص، یا فرمت‌های خاص پیغام، یا ارسال همزمان پیغام به تعداد زیادی از کاربران. اما این ویژگی‌ها به تنهایی موثر نخواهند بود. ما علاوه بر این ویژگی‌ها می‌خواهیم از ساختار شبکه‌های اجتماعی، اطلاعات کاربران، وجود رابطه بین گیرنده و فرستنده پیغام و... استفاده کنیم. به‌عنوان مثال ارسال پیغام از یک کاربر به کاربری ناشناس ممکن است هرزنامه تشخیص داده شود، اما در عین حال همین پیغام برای یک کاربر دوست یا هم‌گروه ممکن است هرزنامه نباشد. بنابراین ما سعی می‌کنیم از این ارتباطات و ساختار شبکه اجتماعی برای تشخیص هرزنامه یا عدم وجود آن استفاده کنیم به شکلی که بتوانیم انواع ارتباط بین افراد را پیدا کنیم. به عنوان مثال مشخص شود کاربری که با دو کاربر دیگر ارتباط دوستی دارد به کدام‌یک نزدیک‌تر است. یا اگر کاربری یک کاربر دیگر را در شبکه بلاک کرده است به‌عنوان یک ویژگی مورد استفاده قرار گیرد. همچنین بر اساس سابقه قبلی کاربران رتبه‌بندی می‌شوند و از قبل مشخص می‌شود که هر کاربر چقدر احتمال ارسال پیغام هرزنامه را دارد. که این ویژگی هم می‌تواند کمک‌کننده باشد. در پایان به کمک الگوریتم ماشین بردار پشتیبان (SVM) پیغام‌ها را با استفاده از این ویژگی‌های بیان شده به دو دسته هرزنامه و غیر هرزنامه طبقه بندی می‌شوند. لازم به ذکر است که SVM دسته‌بند قدرتمندی است که با داده‌های زیاد و وجود نویز هم به‌خوبی می‌تواند کار کند و برای مسائل پردازش متن، دسته‌بند مناسبی است.

۴- نتیجه‌گیری

استقبال گسترده از شبکه‌های اجتماعی و امکانات عظیم آنها و فرصت‌های رو به رشد، کاربران و مخاطبان بسیاری را به خود جلب نموده است. اما در کنار پیام‌ها، مباحث جذاب و جالب، محتوای نامناسب و بعضا مجرمانه مانند هرزنامه نیز در این شبکه‌ها منتشر می‌شود. هرزنامه‌های مخرب قصد ارسال مطالب نادرست یا نامربوط برای توزیع اطلاعات نادرست در شبکه‌های اجتماعی برخط را دارند. در این راستا، الگوریتم‌های بسیاری جهت جلوگیری از هرزنامه پیشنهاد شده است که هدف اصلی افزایش دقت روش تشخیص هرزنامه در پیام‌ها می باشد. لذا در این مقاله، فعالیت ضد هرزنامه و کاربردهای آن در شبکه‌های اجتماعی با استفاده از الگوریتم های داده کاوی مورد توجه قرار گرفته است. به عنوان کارهایی که در آینده در این راستا قابل انجام است استفاده از الگوریتم های داده کاوی به صورت تلفیقی را می توان نام برد.

مراجع

[1] N. Jindal و L. Bing., “Opinion spam and analysis”, In Proceedings of the 2008 International Conference on Web Search and Data Mining, 2008.

[2] M Alsaleh, A. Alarifi, F. Al-Quayed, A. S. Al-Salman, Combating Comment Spam with Machine Learning Approaches, in: 14th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 295–300, 2015.

[3] A. Heydari, M. a. Tavakoli, N. Salim and Z. Heydari, “Detection of review spam: A survey," Expert

Systems with Applications, Vol. 42, No. 7, pp. 3634-3642, 2015.

[4] C. Yang, R. Harkreader, J. Zhang, S. Shin, G. Gu, Analyzing Spammers' Social Networks for Fun and Profit: A Case Study of Cyber Criminal Ecosystem on Twitter, in: Proceedings of the 21st International Conference on World Wide Web, WWW '12, ACM, New York, NY, USA, pp. 71–80, 2012.

[5] H. Yu, M. Kaminsky, P. B. Gibbons, A. D. Flaxman, SybilGuard: Defending Against Sybil Attacks via Social Networks, IEEE/ACM Transactions on Networking, Vol. 16, No. 3, pp. 576–589, 2008.

[6] G. Danezis, P. Mittal, SybilInfer: Detecting Sybil Nodes using Social Networks, in: Proceedings of the Network and Distributed System Security Symposium (NDSS), San Diego, California, USA, 2009.

[7] H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, B. Y. Zhao, Detecting and characterizing social spam campaigns, in: Proceedings of the 10th ACM SIGCOMM Internet Measurement Conference (IMC), Melbourne, Australia, pp. 35–47, 2010.

[8] K. Thomas, C. Grier, J. Ma, V. Paxson, D. Song, Design and Evaluation of a Real-Time URL Spam Filtering Service, in: 32nd IEEE Symposium on Security and Privacy (S&P), Berkeley, California, USA, pp. 447–462, 2011.

[9] S. Lee, J. Kim, WarningBird: Detecting Suspicious URLs in Twitter Stream, in: 19th Annual Network and Distributed System Security Symposium (NDSS), San Diego, California, USA, pp. 183–195, 2012.

[10] F. Benevenuto, G. Magno, T. Rodrigues, V. Almeida, Detecting Spammers on Twitter, in: In Collaboration, Electronic messaging, Anti-Abuse and Spam Conference (CEAS), Vol. 6, 2010.

[11] O. Varol, E. Ferrara, C. A. Davis, F. Menczer, A. Flammini, Online Human-Bot Interactions: Detection, Estimation, and Characterization, in: International AAAI Conference on Web and Social Media, AAAI Press, pp. 280–289, 2017.

[12] K. Lee, B. D. Eoff, J. Caverlee, Seven Months with the Devils: A Long-Term Study of Content Polluters on Twitter, in: Proceedings of the Fifth International Conference on Weblogs and Social Media, pp. 185–192, 2011.

[13] P. N. Howard, B. Kollanyi, Bots, #StrongerIn, and #Brexit: Computational Propaganda during the UK-EU Referendum, Social Science Research Network (SSRN), 2016.

[14] C. Chen, Y. Wang, J. Zhang, Y. Xiang, W. Zhou, G. Min, Statistical features-based real-time detection of drifted twitter spam, IEEE Transactions on Information Forensics and Security, Vol. 12, No. 4, pp. 914–925, 2017.

[15] M. Soiraya, S. Thanalerdmongkol and C. Chantrapornchai, "Using a Data Mining Approach: Spam Detection on Facebook," International Journal of Computer Applications, Vol. 58, No. 13, pp. 26-31, 2012.

[16] T. Wu, S. Liu, J. Zhang, Y. Xiang, Twitter spam detection based on deep learning, in: Proceedings of the Australasian Computer Science Week Multiconference, ACM, 2017.

[17] Sh. Liu, Y. Wang, J. Zhang, Ch. Chen, Y. Xiang, Addressing the Class Imbalance Problem in Twitter Spam Detection Using Ensemble Learning, Computers & Security, pp. 321-330, 2017.

[18] Z. Miller, B. Dickinson, W. Deitrick, Twitter spammer detection using data stream clustering, Journal of Information Sciences, Vol. 260, pp. 64–73, 2015.

[19] A. M. Al-Zoubi, H. Faris, J. Alqatawna, M. A. Hassonah, Evolving Support Vector Machines using Whale Optimization Algorithm for spam profiles detection on online social networks in different lingual contexts, Knowledge-Based Systems, Vol. 65, No. 6, pp. 123-132, 2018.

[20] S. Sedhai, A. Sun, Semi-supervised spam detection in twitter stream, IEEE Transactions on Computational Social Systems, Vol. 5, No. 1, pp. 169–175, 2018.

[21] Mohammad Karim Sohrabi, Firoozeh Karimi, A Feature Selection Approach to Detect Spam in the Facebook Social Network, Arabian Journal for Science and Engineering, Vol. 43, No. 2, pp 949–958, 2018.

[22] Xin Jin, Cindy Xide Lin, Jiebo Luo, Jiawei Han, Social Spam Guard: A Data Mining Based Spam Detection System for Social Media Networks, 37th International Conference on Very Large Data Bases, Vol. 4, No. 12, pp. 1458-1461, 2011.

[23] E.Nalarubiga,, M.Sindhuja, Efficient Classifier for Detecting Spam in Social Networks, International Journal of Innovative Science, Engineering & Technology (IJISET), Vol. 2, No. 10, 2015.

[24] Mona Najafi Sarpiri, Taghi Javdani Gandomani, Mahsa Teymourzadeh, Akram Motamedi, A Hybrid Method for Spammer Detection in Social Networks by Analyzing Graph and User Behavior, Journal of Computers, Vol. 13, No. 7, pp. 823-828, 2018.

[۲۵] Erbs, Nicolai, Keyphrase Extraction using Textual and Visual Features, 52nd Annual Meeting of the Association for Computational Linguistics, 2014.