کشف اسپمرها در شبکه های اجتماعی با استفاده از ماشین یادگیری رویکردی مبتنی بر بازخوردکاربران

محمد نوراللهی[1]،حسن صانعی فر[2] سهیل افراز[3]

1 .کارشناسی ارشد مهندسی کامپیوتر نرم افزار،گروه مهندسی کامپیوتر،پردیس علوم تحقیقات اردبیل، دانشگاه آزاد اسلامی اردبیل،ایران.

گروه مهندسی کامپیوتر،واحد اردبیل، دانشگاه آزاد اسلامی اردبیل،ایران.

Mr.Norullahi@yahoo.com

2.دانشیار گروه مهندسی کامپیوتر، دانشکده مهندسی کامپیوتر، موسسه آموزش عالی رجا، قزوین .

hassan.saneifar@gmail.com

3.دانشیار مهندسی کامپیوتر، دانشکده مهندسی کامپیوتر، عضو هیات علمی دانشگاه آزاد اسلامی واحد اردبیل

soheilafraz@gmail.com

چکیده

امروزه شبکه های اجتماعی به یکی از رایج ترین ابزارهای ارتباطی در زندگی روزمره بشر تبدیل شده است. ارتباط حاصل در این شبکه ها میتواند یک مکالمه ساده دوستانه باشد و یا یک موضوع مهم تجاری و غیره. متاسفانه همین عمومیت و سادگی باعث ایجاد هرزنامه نویسان و کلاهبرداران اینترنتی شده است ، مسئله ای که روزی کم اهمیت جلوه می نمود ، امروزه به معضلی جدی برای میلیون ها کاربر بدل شده است . هم زمان با پیدایش هرزنامه محققین نیز روشهایی برای برخورد با این پدیده معرفی کرده اند تنوع روش های پیشنهادی بسیار زیاد است ، از تکنیک های ساده مبتنی بر قوانین گرفته تا تکنیک های پیچیده در هوش مصنوعی . در میان کارهای انجام شده روشهای مبتنی بر الگوریتم یادگیری ماشینی حجم گسترده ای در کارهای مرتبط در این ضمینه را به خود اختصاص داده است و نتایج بسیار خوبی از این روش ها حاصل شده است.

در این مقاله، یک راه حل نظارتی مبتنی بر یادگیری ماشین وجود دارد که برای تشخیص کآرای اسپمر ، ارائه شده است. ابتدا، یک مجموعه داده از گوگل پلاس جمع آوری شده است که شامل 30116 کاربرو بیش از 16 میلیون پیام است. پس از آن، مجموعۀ دادۀ نشاندار از کاربران ساخته می شود . کاربران به صورت دستی به اسپمرها و غیر- اسپمرها طبقه بندی شده. سپس، مجموعه ای از ویژگی هایی را از محتوای پیام ها و رفتار اجتماعی کاربران استخراج می کند،در ماشین بردار پشتیبانی می کرده بر اساس الگوریتم تشخیص اسپمر اعمال می کند.

در نهایت آزمایش ما نشان می دهد که راه حل پیشنهادی قادر به ارائه عملکرد عالی با میزان مثبت واقعی اسپمر 99.5٪ و غیر اسپمر 99.9٪ است.

کلمات کلیدی : شبکه های اجتماعی ، اسپمر، ماشین یادگیری ،هرزنامه، غیر اسپم

مقدمه

در چند سال گذشته، شبکه های اجتماعی آنلاین مانند فیس بوک، توییتر، گوگل پلاس و غیره تبدیل به یکی از راه های اصلی برای کاربران اینترنت برای حفظ ارتباطات با دوستان خود شده اند. [1-3] بر اساس گزارش Statista ، تعداد کاربران شبکه های اجتماعی به 1.61 میلیارد تا اواخر سال 2013 رسیده، و تا پایان سال جاری یعنی 2017 ،کاربران جهان، حدود 3.49 میلیارد برسد. [4] با این حال، همراه با موفقیت های بزرگ فنی و بازرگانی، پلت فرم شبکه های اجتماعی نیز مقدار زیادی از فرصت ها را برای اسپمر رادیو و تلویزیون فراهم می آورد، که پیام ها و رفتار مخرب را گسترش می دهد .بر اساس گزارش Nexgate [5]، در طول نیمه اول سال 2013، رشد اسپم اجتماعی ، 355٪ درصد بسیار سریع تر از میزان رشد حساب و پیام در اجتماعی ترین شبکه های مارک دار و نشان دار شده است.تاثیرات پیام های اسپم اجتماعی از اهمیت بالایی برخوردار است .یک پیام اسپم اجتماعی به طور بالقوه توسط همه دنبال کنندگان و دوستان دریافت کننده دیده می شود. واین شاید سبب تفسیر اشتباه و سوء تفاهم در مبحث های خاص و روند عمومی شود .به عنوان مثال، روند موضوعات پر طرفدار، همیشه توسط اسپمرها برای انتشار نظرات با آدرس ها URL مورد سوء استفاده قرار گرفته اند ، کلیه کاربران با وب سایت های کاملا نامربوط مورد راهنمایی غلط قرار گرفته اند. از آنجا که اکثر شبکه های اجتماعی، خدمات مربوط به صفحات وب ها را در داخل پیام ارائه می دهند، بدون مراجعه به سایت شناسایی محتوای آنها بسیار مشکل است. چند طرح مطرح شده از حوزه های صنعت و دانشگاه وجود دارند که در مورد راه حل های ممکن برای تشخیص اسپم و فیلتر کردن بحث می کند (توصیف شده در بخش 2).با این حال، آنها هم یا بی اثر هستند یا بر اساس شرایط بیش از حد سختگیرانه کمتر مورد استفاده قرار می گیرند .به عنوان مثال، بسیاری از مطالب و ویژگی رفتاری و غیره. این مقاله ، محتوای اسپم اجتماعی و مسائل مربوط به رفتار کاربران را بررسی می کند، و یک مدل یادگیری ماشینی موثر برای تشخیص اسپم پیشنهاد می کند .

این مقاله شامل بخش های اصلی زیر است:
ویژگی اسپمر را برای تشخیص اسپم و تست نتایج کل درگوگل پلاس تصویب و بروز رسانی می شود. برای تجزیه و تحلیل دادها در Google+ API ، یک مجموعه داده های خاص بمنظور استخراج پیامهای عمومی تمام کاربران غیر مجاز، در داخل پلت فرم Google+ توسعه یافته است. .
نوآوری مهم این مقاله، بررسی مجموعه ای از مهم ترین ویژگی های مربوط به محتوای پیام و رفتار کاربران و اعمال آنها بر روی ماشین بردار پشتیبان بر اساس الگوریتم طبقه بندی برای تشخیص اسپمر است.

آزمایش و کار مقایسه نشان می دهد که راه حل پیشنهادی قادر به ارائه دقت و صحت بالاتر است. از طریق الگوریتم های انتخاب ویژگی و تست کردن آزمایش، ده تا از مهم ترین ویژگی و ارزش و وزن این ویژگی ها شناخته شده است. نتایج آزمایش بیشتر به ویژگی اسپم انتخاب شده اعتبار می دهد (طبقه بندی دستی) و همچنین توضیح می دهد که چرا راه حل پیشنهادی می تواند عملکرد عالی به دست آورد .

با روابط کاربری دوستانه، کارآمد و نتیجه طبقه بندی دقیق ، کاربران عادی قادر به تشخیص هر کاربر **گوگل پلاس** با عملیات ساده هستند . [6]

بخش 2- پس زمینه ای از شبکه های اجتماعی **گوگل پلاس** ارائه می دهد و برخی از آثار مرتبط در مورد تشخیص اسپمر را نمایش می دهد .

بخش 3 - نحوه جمع آوری داده های مجموعه داده و ویژگی های استخراج کردن را معرفی می کند .

بخش 4 - مدل تشخیص اسپمر آزمایشات و ارزیابی مربوطه را توصیف می کند .در نهایت، نتیجه گیری و کارهای آینده در بخش 5 داده شده است.

2. کارهای مرتبط

2.1 شبکه های اجتماعی گوگل پلاس

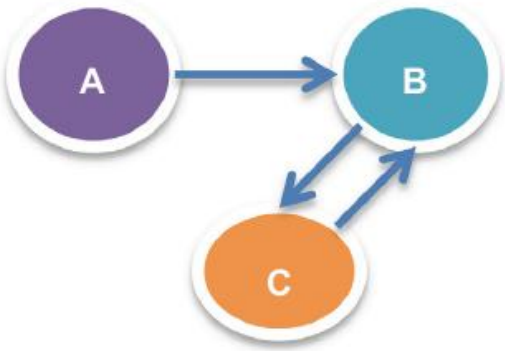
با توجه به آمار[3]، تعداد کاربران سایت گوگل پلاس به بیش از 500 میلیون رسیده است .آمار نشان می دهد که **گوگل پلاس** مداوما در میان 25 مورد برتر، بیشترین وب سایت باز دیدشده در طول چند سال گذشته قرار دارد [7] .

نرم افزار **گوگل پلاس** ، شبیه به فیس بوک هست، که در آن کاربران پیام ارسال می کنند، با دوستان ارتباط برقرار می کنند، و در مورد اخبار و به اشتراک گذاری موضوعات جالب از طریق خدمات شبکه های اجتماعی صحبت می کنند پیغام های ارسال شده به دنبال کنندگان یا اصطلاحا فالو کنندگان بلافاصله تحویل داده خواهد شد .هر کاربر ، توسط یک نام کاربری یا یوزرنیم منحصر به فرد شناسایی می شود .کاربری که دنبال یا فالو می شود می تواند درخواست را قبول ، و یا فقط رد کند . شکل . 1 یک مثال نمودار دنبال کنندهٔ ساده را توصیف می کند که در آن کاربر A کاربر B را دنبال می کند ، و کاربر B و C یک دیگر را دنبال می کنند .تعدادی عبارات در +Google وجود دارد که به کاربران برای ارتباط برقرار کردن با دیگران با یک روش بهتر، از جمله اشاره، ریپست یا ارسال مجدد و هشتگ اجازه می دهد.

2.1.1. اشاره کردن

یک پیام **گوگل پلاس** شامل یک سری از کلمات کلیدی مانند @ نام کاربری است، به این معنی که فرستنده پیام مایل است تا با کاربران ذکر شده چیزی را به اشتراک بگذارد در نتیجه، **گوگل پلاس** به طور خودکار به کاربران با ارسال پیام اشاره شده یا ذکر شده در صفحه اصلیش اطلاع خواهد داد.

به عنوان مثال ? @ Mohammad wann'a go for a Tea : @ Mohammad wann'a



شکل 1. مثال نمودار دنبال کننده ساده

2.1.2 . ارسال مجدد

ارسال مجدد راه دیگری برای ارسال پیام است.یک کاربر همیشه پیام های کاربران دیگر را که مورد علاقه اوست ارسال مجدد می کند پیام اعلان مجدد شده یا Repost شده نیز توسط دنبال کنندگان یا فالو کنندگان کاربر، دریافت می شود.

2.2. تحقیقات پیشین صورت گرفته

در ده سال گذشته، تشخیص اسپم ایمیل و مکانیزم فیلتر به طور گسترده ای اجرا شده است .هدف اصلی برای این کار را می توان به دو دسته تقسیم کرد . مدل -مبتنی بر محتوا و مدل مبتنی بر هویت.در مدل مبتنی بر محتوا، مجموعه ای از روش های یادگیری ماشینی [8.9] برای تجزیه محتواست که با توجه به کلمات کلیدی و الگوهایی اجرا شده است که هرزنامه ها یا اسپم احتمالی هستند. در مدل مبتنی بر – هویت ، بیشترین رویکرد مورد استفاده ، بدین شکل است که هر کاربر، یک لیست سفید و لیست سیاهی از آدرس های ایمیل که باید و نباید توسط مکانیسم ضد اسپم [10.11] مسدود شود را حفظ می کند .

در فیس بوک یک الگوریتم EdgeRank [13] پیشنهاد می شود با این روش که امتیاز هر پست با توجه به چند ویژگی (به عنوان مثال، تعدادی از امثال، تعداد نظرات، یا ارسال مجدد ها ، و غیره) اختصاص داده

شده و محاسبه می شود .بنابراین، هرچه EdgeRank بالاتر باشد، امکان کمتری برای یک یک اسپمر دارد .عیب این روش این است که اسپمر ها می توانند به شبکه های آنها بپیوندند و به طور مداوم دوست و لایک داشته باشند و به یکدیگر به منظور دستیابی به نمره EdgeRank بالا اظهار نظر کنند.

در دانشگاه ها، ساریتا یاردی و همکارانش [14] ، رفتار بخش کوچکی از اسپمر ها را در توییتر مطالعه کردند، و دریافتند که رفتار اسپمر، از کاربران قانونی در زمینه ارسال توییت، پیروان یا فالو کنندگان، دوستان فالو کننده و غیره متفاوت است .

استرینگینی و همکارانش [15] بیشتر به بررسی ویژگی های اسپم از طریق ایجاد تعدادی از پروفیل های- هانی در سه سایت شبکه اجتماعی بزرگ پرداخته اند (فیس بوک، توییتر و مای اسپیس) و چند ویژگی های مشترک (تعقیب -با -دنبال کننده، نسبت URL ، شباهت پیام، ارسال مجدد پیام ، تعداد دوست ، و غیره) را به طور بالقوه برای تشخیص اسپمر شناسایی کردند.

با این حال، اگر چه هر دوی این دو رویکرد ،چارچوب convincible را برای تشخیص اسپمر معرفی کرد ، اما آنها فاقد خصوصیات روش های دقیق و ارزیابی نمونه اولیه هستند.

وانگ [16] یک دسته بندی کننده بیزی ساده مبتنی بر الگوریتم طبقه بندی اسپمر را برای تشخیص رفتار مشکوک از رفتار طبیعی، درشبکه توییتر ، با نتیجه دقیق (مقدار اندازه گیری-F) از 89٪ درصدی پیشنهاد کرد .

گائو و همکارانش .[17] مجموعه ای از ویژگی های جدید را برای بازسازی موثر پیام های اسپم در مبارزات انتخاباتی به جای بررسی آنها به صورت جداگانه با میزان دقت بیش از 80٪ را تصویب و به روز کردند . نقطه ضعف این دو روش این است که آنها به اندازه کافی دقیق نیستند.

بنیونتو و همکارانش [18] مجموعه داده های بزرگی از توییتر جمع آوری کردند و 62 ویژگی مربوط به رفتار اجتماعی کاربران را شناسایی کردند.

این ویژگی ها به عنوان ویژگی هایی در یک فرایند یادگیری ماشینی برای طبقه بندی کاربران به صورت اسپمر و یا غیر-اسپمردر نظر گرفته می شوند .

با این حال، این روش مبتنی بر مقدار زیادی از ویژگی های انتخاب شده هست که ممکن است محاسبات سنگینی داشته و زمان زیادی در آموزش مبتنی بر مدل صرف کند.

به طور کلی، این مقاله به شرح زیر مفهوم مشابه با آثار قبلی ، اما، با چند نقطه برجسته را دنبال می کند:

1 . مدل طبقه بندی مبتنی بر- SVM پیشنهادی ما چند مورد از ویژگی ها را بررسی می کند و می توان ملاحظه کرد که به بهترین نتیجه عملکرد، با میزان و ارزش اندازه گیری-F رسیده که بیش از 99٪ است .

هر چند مجموعه داده های مختلف جمع آوری شده با محتویات مختلف ممکن است در نتیجه محاسبه باعث میزان اندکی انحراف شوند ، ولی این بهترین نتیجه ای است تا کنون به دست آمده است.

2. اهمیت هر یک از ویژگی های انتخاب شده ، از طریق weka [21] ، یک نرم افزار داده کاوی بر ابزار جاوا مورد مطالعه و تایید است.

استفاده ترکیبی از این ویژگی همچنین توضیح میدهد که چرا روش پیشنهادی قادر به دستیابی به میزان دقت بسیار بالاتر از دیگر آثار موجود است.

3 . جمع آوری و تجزیه و تحلیل مجموعه داده ها

3.1. .مجموعه داده ها و ویژگی های مجموعه

مانند به بسیاری از رسانه های اجتماعی، توسعه دهندهٔ عمومی Google+ API تنها قابلیت دانلود بر روی پیام های اخیر از کاربران مجاز را فراهم می کنددرواقع .این به عنوان مانعی بر سر راه نحوه جمع آوری داده ها است .برای حل این مشکل، داده های خاص crawler و مکانیزم جمع آوری ویژگی ها توسعه می یابند ، که در مراحل زیر (شکل 2). توصیف می شود:

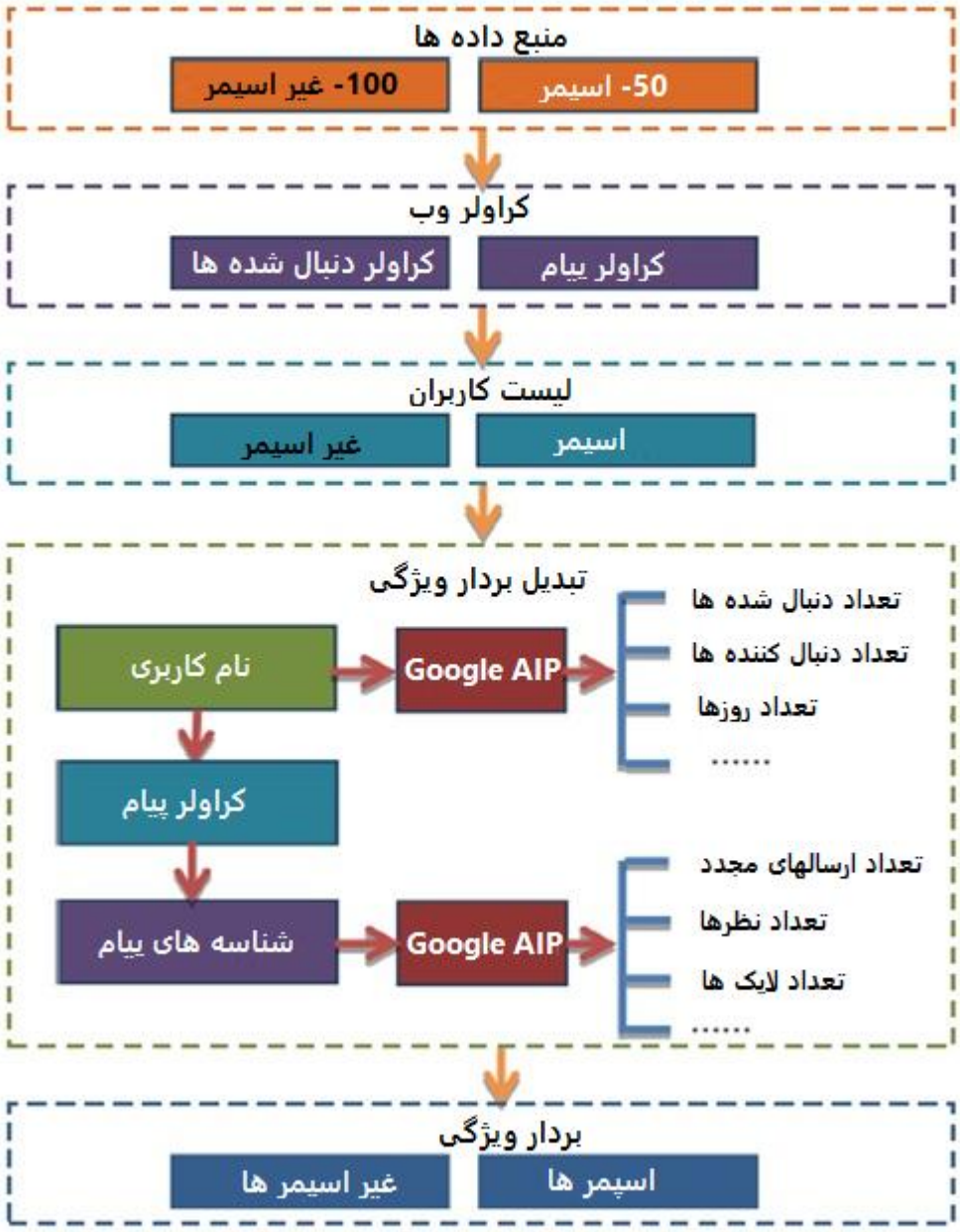
1 . 100 کاربر عادی (از کاربران مشهور، شرکت ها، و دولت که اغلب پست ارسال مجدد دارند) و 50 کاربر اسپمر که اغلب در معرض رفتار مخرب هستند به صورت دستی به عنوان یک منبع داده انتخاب شده اند.

2 .دو نوع کراولر داده ها برای کاربر عادی و اسپمر به ترتیب توسعه یافته اند. کراولرکاربر عادی که برای استخراج لیست کاربر عادی از دنبال شده ها یا فالویی ها است، که به عنوان کاربران نرمال نیز در نظر گرفته شده اند چرا که بسیاری از کاربران عادی بعید است که به دنبال اسپمر ها باشند .کراولر اسپمر (یا خزندهٔ اسپمر) ، برای استخراج لیست اسپمرها پشت پیام های اعلان مجدد خاص اسپمر است . در نهایت، 30116 کاربر Google+ استخراج شده اند.

3 .برای هر کاربر، ما اطلاعات را در داخل 500 پیام های اخیر مربوطه کراول کرده ایم،

قدم اول : اطلاعات کاربر عمومی (به عنوان مثال، تعداد دنبال ها، تعداد فالو کننده ها، روز ایجاد شده، و غیره) می تواند از طریق Google+ API به دست آید .

قدم دوم: از طریق نام کاربری، قادر به کراول کردن مجموعه ای از پیام از طریق ویژگی های آن. به عنوان مثال، تعداد مانند ارسال های مجدد، تعداد نظرات، تعداد امثال لایک ها، و غیره خواهیم بود که می تواند با کمک Google+ API به دست آید .



شکل 2 .مجموعه داده ها و روش های جمع آوری ویژگی های

4 .در شکل قبل برای هر کاربر، یک بردار ویژگی ، با توجه به کاربر کراول شده و اطلاعات پیام توضیح داده شده ساخته شده است.پس از آن، کاربران جمع شده به عنوان اسپمرها یا غیر-اسپمرها نامیده شده اند.

در مجموع، 11488 اسپمر و 17646 غیر- اسپمر برچسب.زده شده و نامگذاری شدند در نهایت :

80% اسپمر و غیر- اسپمر از مجموعه دادهٔ با برچسب و اسم ها به صورت تصادفی به عنوان داده آموزشی انتخاب شده اند و، بقیه به عنوان تست داده کنار گذاشته شده اند.

4 .تشخیص اسپمر

بر اساس مجموعه داده ها و ویژگی های مجموعه شرح داده شده در بخش قبل، یک مدل یادگیری ماشینی، برای شناسایی اسپمرها معرفی شده است .

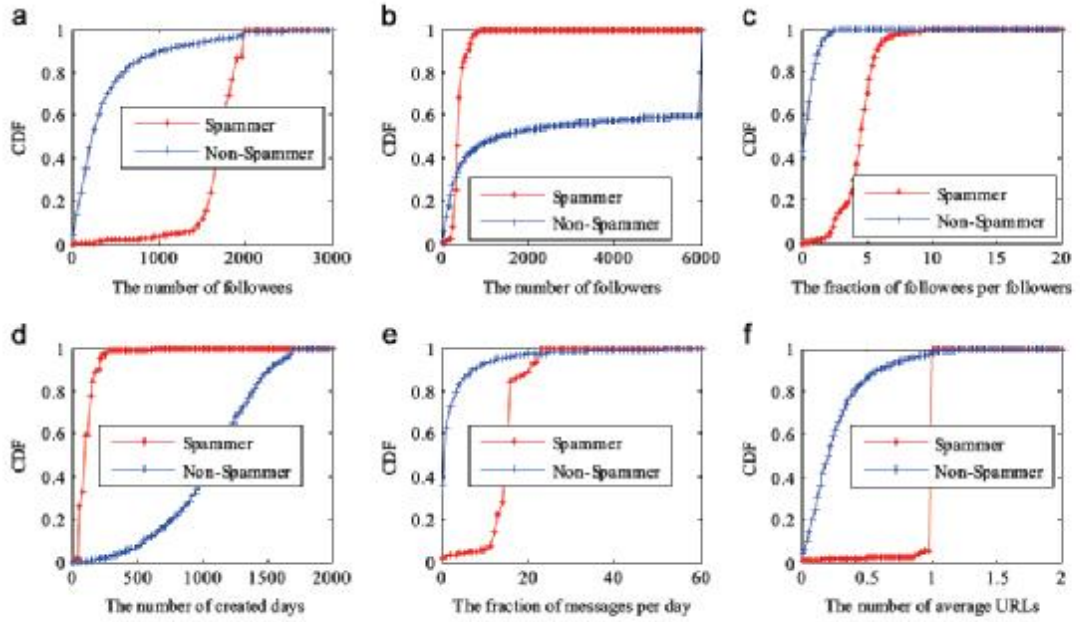
یادگیری نظارت شده [22]، وظیفه یادگیری ماشینی است که یک تابع و عملکردی را از داده های آموزشی برچسب زده شده و نامگذاری شده استنتاج میکند که متشکل از مجموعه ای از نمونه های آموزشی است .در یادگیری نظارت شده، هر مثال، متشکل از یک شی ورودی (معمولا یک بردار) و یک مقدار خروجی مورد نظر (که سیگنال نظارت نیز نامیده می شود) است .

از طریق تجزیه و تحلیل داده های آموزشی، راه حل یادگیری نظارت شده، یک مدل طبقه بندی را برای پیش بینی نمونه های جدید تولید می کند.

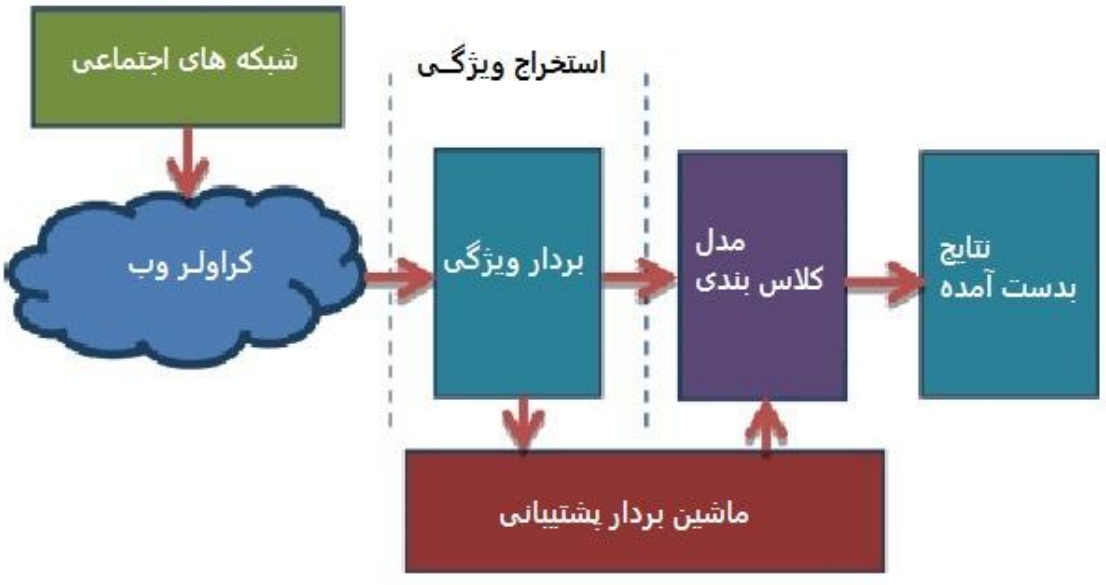
4.1 . ماشین بردار پشتیبان مبتنی بر مدل تشخیص اسپمر

شکل5. مفهوم اساسی مدل پیشنهادی تشخیص اسپمررا نشان می دهد. در این راه حل، داده های آموزشی ، به یک سری از بردار های ویژگی تبدیل می شوند که از مجموعه ای از ارزش ها برای صفات و نگرش ها تشکیل شده است . این بردارها ، ورودی الگوریتم یادگیری ماشین تحت نظارت را می سازد .پس از آموزش، مدل طبقه بندی برای تشخیص اینکه آیا کاربر خاص متعلق به دسته بندی کاربر عادی و یا اسپمر است این اعمال می شود. از آنجا که اسپمرها و غیر- اسپمرها، رفتار های مختلف اجتماعی ، از طریق تجزیه و تحلیل ویژگی های محتوا و رفتار کاربر دارند ، مدل قادر به تشخیص رفتار غیر طبیعی از موارد قانونی ، است .در این مقاله، ما چند ویژگی های لیست شده را در زیر تنظیم کردیم:

تعداد فالویی ها، تعداد فالوکننده ها، تعداد پیام ها، تعدادی دوستان فالو کننده یا دنبال کنندهٔ یکدیگر، تعداد علاقه مندی ها، تعداد روز ایجاد شده ، تعداد پیام های روزانه، تعداد متوسط ارسال های مجدد، متوسط تعداد نظرات، متوسط تعداد لایک ها، متوسط تعداد و بخشی از پیام های حاوی URL.



شکل 3.مقایسه ویژگی های لیست شده



شکل 4. بررسی اجمالی مدل تشخیص اسپمر.

4.2 .طبقه بندی کنندهٔ ماشین بردار پشتیبان

راه حل تشخیص اسپمر، بر اساس یک طبقه بندی ماشین بردار پشتیبان غیر خطی [23] با هسته تابع پایه شعاعی RBF است .

این حالت می تواند از طریق اجرای ارائه شده توسط LIBSVM ، یک نرم افزار یکپارچه برای حمایت از طبقه بندی ، رگرسیون و برآورد توزیع بردار، به دست آید.

ماشین بردار پشتیبان با عملکرد هسته RBF ، دو پارامتر های آموزشی دارد: C ، مدل را متناسب سازی می کند .و گاما درجه غیرخطی را کنترل می کند .

در این آزمایش، ما یک ابزار انتخاب پارامتر ارائه شده توسط LIBSVM[13] را برای انتخاب پارامترها به طور خودکار با 5 -برابر اعتبار –متقاطع اعمال کردیم.این ابزار از سیاست جستجوی شبکه برای پیدا کردن بالاترین دقت طبقه بندی از طریق محاسبهٔ مقادیر مختلف C و جفت گاما استفاده می کند. در نهایت، مناسب ترین جفت که C و گاما به ترتیب برابر با 128 و 0.03125 تولید شده است و برای مجموعه داده آموزش خاص انتخاب شده است.

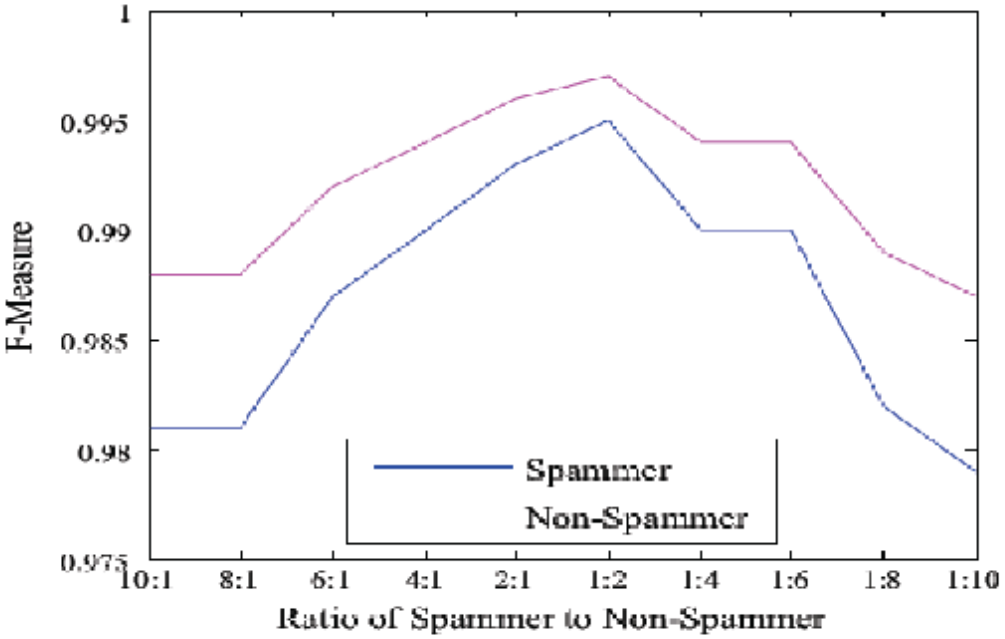
اندازه گیری یا سنجش-F . میانگین هارمونیک بین دقت و یادآوری (مانعیت و جامعیت) است، و به عنوان $F = 2PR/(P + R)$ تعریف شده است .دقت (P) ، نسبت تعداد موارد به درستی طبقه بندی شده به تعداد کل موارد است و توسط فرمول $P = a/(a + c)$ بیان شده است .

یادآوری (R) ، نسبت تعداد موارد به درستی طبقه بندی شده به تعداد کل موارد پیش بینی شده است و با $R = a/(a + b)$ فرمول بیان شده است.برای ارزیابی عملکرد طبقه بندی کننده ها، مقدار اندازه گیری یا سنجش-F دقیق تر است ،چون یک ترکیب از دقت و یادآوری است.

4.3 .نسبت اسپمر به غیر اسپمر

در مرحله اول، ما از مجموعه داده آموزشی کامل برای تست کار و دستیابی به ارزش و مقدار اندازه گیری -F اسپمر و غیراسپمر را تا 90.6٪ و 92.2٪ استفاده کردیم .این ممکن است نتیجه بهینه نداشته باشد. به منظور دستیابی به دقت تشخیص بالاتر اسپمر، نسبت اسپمر به غیر-اسپمر در مجموعه داده آموزشی به شرح زیر تغییر می کند 1:10 ,1:8 , 1:6, 1:4, 1:2, 2:1, 4:1 6:1, 8:1, 10:1 ، با نتیجه دقیق و صحیح طبقه بندی مربوطه در شکل 6 نشان داده شده است .

این نشان می دهد که ارزش یا مقدار اندازه گیری -F از هر دو اسپمر و غیر اسپمر به طور همزمان رشد می کند که نسبت اسپمر کاهش می یابد، و بالاترین دقت و صحت حدود 99.5٪ و 99.9٪ را به دست می آورد زمانی که نسبت به2: 1 تنظیم شده است. پس از آن، دقت به سرعت افت می کند در حالی که نسبت غیر-اسپمر افزایش می یابد.



شکل5. دقت و صحت طبقه بندی با نسبت های مختلف اسپمر به غیر-اسپمر در مجموعه داده آموزشی.

از سوی دیگر، مشخص است که مقداردهی یک نسبت مناسب به غیر- اسپمر ، زمانی که تفاوت کیفی بزرگ با صحت و دقت پایین تر نتیجه می شود مهم است (مثلا 10:1 یا 1:10).
این به این دلیل است که به احتمال زیاد نسبت بزرگی از اسپمرها، به طبقه بندی نادرست کاربران عادی برای اسپمر و بالعکس نشان داده شده اند. بنابراین، در آزمایش زیر، نسبت اسپمر به غیر-اسپمر 1:2 تنظیم شده است.

4.4 .نتیجه و مقایسهٔ طبقه بندی

جدول 2 ماتریس سردرگم به دست آمده توسط طبقه بندی کنندهٔ SVM را نشان می دهد . این جدول نشان می دهد که راه حل پیشنهادی ما که کاملا کارآمد است، و با 99.1٪ اسپمر و 99.9٪ غیر اسپمر، به طورصحیح طبقه بندی شده است، و تنها بخش کوچکی از اسپمرها و غیر اسپمرها طبقه بندی نادرست شده اند .علاوه بر این، همچنین روش پیشنهادی را با سایر طبقه بندی ها مقایسه می کنیم:
درخت تصمیم، Naïve Bayes و شبکه بیز Bayes ، با اجرای ارائه شده توسط WEKA .

برای هر طبقه بندی کننده، همان عیارهای ارزیابی (دقت، یادآوری یا جامعیت و اندازه گیری-F)، برای هر دو اسپمر و غیر- اسپمر ، با نتیجه نشان داده شده در جدول 4 محاسبه شده است.

بدیهی است که طبقه بندی کنندهٔ، SVM قادر به دستیابی به بهترین دقت است .این نشان می دهد که ابرصفحه محاسبه شده توسط SVM می تواند داده های آموزشی را به دو بخش با حداکثر حاشیه ازهم جدا کند .

علاوه بر این، نشان داده شده است که سه طبقه بندی دیگر نیز دقت خوبی به دست آورده است .این به این دلیل است که از ویژگی های مناسب (از جمله محتوا و رفتار کاربر) انتخاب شده اند که به طور موثر و کارآمدی قادر به تشخیص اسپمر از غیر- اسپمر هستند .

Table 4

Classifier	Precision		Recall		*F*-measure	
	Spammer	Non-spammer	Spammer	Non-spammer	Spammer	Non-spammer
SVM	0.999	0.995	0.991	0.999	0.995	0.997
Decision Tree	0.942	0.95	0.953	0.958	0.947	0.954
Naïve Bayes	0.939	0.96	0.922	0.966	0.93	0.963
Bayes Network	0.946	0.915	0.907	0.956	0.926	0.935

شکل5. مقایسه روش پیشنهادی با سایر طبقه بندی ها.

5 .نتیجه گیری و کارهای آینده

در این مقاله، ما یک راه حل تشخیص اسپمر برای شبکه های اجتماعی براساس ماشین یادگیری معرفی کرده ایم.در این روش ، محتوا و ویژگی های رفتاری کاربران را بررسی شده، و آنها در SVM مبتنی بر الگوریتم برای طبقه بندی اسپمر اعمال شوند.

نشان داده ایم که از طریق تجزیه و تحلیل ها، راه حل پیشنهادی امکان پذیر است و قادر به رسیدن به نتیجه طبقه بندی بسیار بهتر از دیگر روش های موجود است.

با این حال، دو موضوع باز هنوز هم برای پاسخ فوری در انتظار است .از یک طرف ، اگر چه روش پیشنهادی می تواند نتیجه دقیق طبقه بندی را به دست آورد،اما باید دید آن بیش از یک ساعت در یک روند آموزش مدل طول می کشد یا نه .

بنابراین، یک مسئله باز شامل تشخیص اسپمر آنلاین است و آن این است که توانایی داده های زمان واقعی و مجموعه ویژگی ها، زمان آموزش پایین تری را با دقت بالا در بر دارد.

یادگیری ماشینی فوق العاده، (ELM) [25.26]، یک طرح آموزشی جدید شبکه های عصبی پیشخور است که زمان آموزشی بسیار پایین تر و دقت و صحت مشابهی را ارائه می دهد، که می تواند یک راه حل ممکن باشد.از سوی دیگر، ویژگی های استخراج شده در راه حل پیشنهادی ما (روش موجود نیز) بر اساس تجزیه و تحلیل آماری و انتخاب دستی است .

با این حال، با توجه به قرار گرفتن ما در عصر داده های بزرگ با حجم عظیمی از داده ها و دسترسی راحت [27]، در راه حل ما مکانیسم استخراج ویژگی ممکن است کم تر تطبیقی باشد.

مراجــع

[1] Facebook, ⟨http://www.facebook.com/⟩.
[2] Welcome to Twitter, ⟨http://twitter.com/⟩.
[3] Google plus, ⟨ https://plus.google.com /⟩.
[4] Statista, ⟨http://www.statista.com/⟩.
[5] Nexgate. 2013 State of Social Media Spam, ⟨http://nexgate.com/wp-content/ uploads/2013/09/Nexgate-2013-State-of-Social-Media-Spam-Research-Re port.pdf⟩, 2013.
[6] Google-Plus crawler, ⟨http://googlepluscrawler.sourceforge.net/⟩.
[7] Alexa Top 500 Global Sites, ⟨http://www.alexa.com/topsites⟩.
[8] M. Uemura, T. Tabata, Design and evaluation of a Bayesian-filter-based image spam filtering method, in: Proceedings of the International Conference on Information Security and Assurance (ISA), IEEE, 2008, pp. 46–51.
[9] B. Zhou, Y. Yao, J. Luo, Cost-sensitive three-way email spam filtering, J. Intell. Inf. Syst. 42 (1) (2013) 19–45.
[10] Jung, E. Sit, An empirical study of spam traffic and the use of DNS black Lists, in: Proceedings of the 4th ACM SIGCOMM Conference on Internet Measure-ment, ACM, 2004, pp. 370–375.
[11] M. Antonakakis, R. Perdisci, D. Dagon, W. Lee, N. Feamster, Building a dynamic reputation system for DNS, in: Proceedings of the Third USENIX Workshop on Large-scale Exploits and Emergent Threats (LEET), 2010.
[12] Trust evaluation based content filtering in social interactive data, in: Proceed-ings of the 2013 International Conference on Cloud Computing and Big Data (CloudCom-Asia), IEEE, 2013, pp. 538–542.
[13] J. Kincaird, Edgerank: the secret sauce that makes Facebook's news feed tick, TechCrunch, 2010, ⟨http://techcrunch.com/2010/04/22/facebook-edgeran⟩.
[14] S. Yardi, D. Romero, G. Schoenebeck, Detecting spam in a Twitter network, First Monday 15 (1) (2009).

[15] G. Stringhini, C. Kruegel, G. Vigna, Detecting spammers on social networks, in: Proceedings of the 26th Annual Computer Security Applications Conference, ACM, 2010, pp. 1–9

[16] A.H. Wang, Don't follow me: spam detection in Twitter, Security and Cryptography (SECRYPT), in: Proceedings of the 2010 International Conference on. IEEE, 2010, pp. 1–10

[17] H. Gao, Y. Chen, K. Lee, D. Palsetia, A. Choudhary, Towards online spam filtering in social networks, in: Proceedings of the Symposium on Network and Distributed System Security (NDSS), 2012.

[18] Y. Zhu, X. Wang, E. Zhong, N.N. Liu, H. Li, Q. Yang, Discovering spammers in social networks, in: Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence (AAAI), 2012

[19] Y. Zhu, X. Wang, E. Zhong, N.N. Liu, H. Li, Q. Yang, Discovering spammers in social networks, in: Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence (AAAI), 2012.

[20] X. Hu, J. Tang, Y. Zhang, H. Liu, Social spammer detection in microblogging in: Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, ACM, 2013, pp. 2633–2639.

[21] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The WEKA data mining software: an update, ACM SIGKDD Explor. Newsl. 11 (1) (2009) 10–18.

[22] F. Wang, C. Zhang, Robust self-tuning semi-supervised learning, Neurocom-puting 70 (16) (2007) 2931–2939.

[23] C. Cortes, V. Vapnik, Support-vector networks, Mach. learn. 20 (3) (1995) 273–297.

[24] LIBSVM – A Library for Support Vector Machines, ⟨http://www.csie.ntu.edu. tw/ cjlin/libsvm/⟩.

[25] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: theory and applications, Neurocomputing 70 (1) (2006) 489–501.

[26] G.-B. Huang, H. Zhou, R. Zhang, Extreme learning machine for regression and multiclass classification, IEEE Trans. Syst., Man, Cybern. 42 (2) (2012) 513–529.

[27] X. Zheng, N. Chen, Z. Chen, C. Rong, G. Chen, W. Guo, Mobile cloud based framework for remote-resident multimedia discovery and access, J. Internet Technol. 15 (6) (2014) 1043–1050.