مطالعات علوم کاربردی در مهندسی

دوره ۳، شماره ۳، پاییز ۱۳۹۶

صفحات ۱۳- ۱

Online ISSN: ۲٤۷٦-٤٥۰۷

Print ISSN: ۲٥۳۸-۲۰٥٥

www.irijournals.com

تشخیص هرزنامه در شبکه های اجتماعی براساس رفتار کاربران مبتنی بر نظریه فازی

فرزانه پرگنه[۱]، محسن فیروزبخت [۲]

[۱] کارشناسی ارشد، گروه مهندسی کامپیوتر دانشکده فنی و مهندسی دانشگاه آزاد واحد الکترونیکی

[۲] دکتری، استادیار گروه مهندسی کامپیوتر دانشگاه آزاد اسلامی واحد تهران جنوب

چکیده

وجود حجم عظیم اسپم ها و حملات مرتبط در شبکه های اجتماعی مثل فیس بوک، توئیتر و غیره، این شبکه ها را با مشکلات زیادی روبرو کرده است. از طرفی ارائه روشی بهینه که بتوان اسپم ها را با دقت قابل قبولی تشخیص داده و تا حدودی مشکلات رسانه های اجتماعی را کاهش دهد پیچیده است؛ بنابراین وجود یک روش مطمئن و کارآمد در این عرصه بسیار ضروری و حائز اهمیت است. در این پژوهش با بکارگیری یک سیستم فازی، روشی مطرح شده است که تا حد بسیار مطلوبی می تواند اسپم ها را در شبکه اجتماعی شناسایی نموده و به میزان قابل توجهی از این سربار اطلاعاتی فارغ گردد. بطور کلی روش مطرح شده دارای مراحلی است که عبارتند از: ۱) ورود دیتاست مربوط به شبکه اجتماعی ۲) اعمال پیش پردازش بر روی داده ها و حذف داده های بلا استفاده ۳) تفکیک داده های آزمایشی و آموزشی ۴) اعمال سیستم فازی FCM جهت تولید قوانین ۵) اعمال سیستم فازی تولید شده بر روی داده ها و تشخیص اسپم. در نهایت پس از شبیه سازی روش پیشنهادی مشاهده گردید که میزان دقت روش ارائه شده ۹۹.۹٪ بوده است که نسبت به روشهایی که تا کنون مطرح شده است بهبود قابل توجهی داشته است.

واژه‌های کلیدی: شبکه اجتماعی، تشخیص هرزنامه، نظریه فازی.

مقدمه

شبکه های اجتماعی مجازی نسل جدیدی از فضای روابط اجتماعی هستند که با اینکه عمر خیلی زیادی ندارند، توانسته اند به خوبی در زندگی مردم جا باز کنند. مردم بسیاری در سنین مختلف و از گروه های اجتماعی متفاوت در شبکه های اجتماعی مجازی کنار هم آمده اند و از فاصله های بسیار دور در دنیای واقعی، از طریق شبکه های اجتماعی با هم ارتباط برقرار می کنند. شبکه های اجتماعی نقش پررنگی در دنیای امروز دارند و نمی توان آن ها را نادیده گرفت. این سایت ها بر ابعاد مختلف زندگی فردی و اجتماعی افراد و در سطح کشورها و حتی بین الملل تاثیرگذارند و به همین دلیل در حال گسترش هستند و در آینده نقش به مراتب بیشتر و مهم تری را در زندگی بازی خواهند کرد.

هرزنامه

اینترنت، امکان استفاده از سرویس ها و خدمات متعددی را در اختیار کاربران قرار می دهد. ارسال و دریافت نامه های الکترونیکی، یکی از متداولترین سرویس ارائه شده بر روی نت است. با وجود تمام مزیت های سرویس فوق، در چند سال گذشته و همزمان با رشد و گسترش استفاده از اینترنت شاهد مسائل و مشکلات جانبی در این رابطه می باشم. توزیع نامه های آلوده به ویروس ها و یا کرم ها، ارسال و یا دریافت نامه های الکترونیکی ناخواسته که از آنان با نام هرزنامه یا هرزنامه یاد می شود، نمونه هایی در این زمینه می باشد (فیرت و همکاران[1]، ۲۰۱۰).

روش مبتنی بر محتوا

در این روش، محتویات پست الکترونیکی مانند ضمیمه مورد بررسی قرار می‌گیرند و اگر کلماتی در آن منطبق با فرهنگ لغات مورد نظر بیابند، آن را فیلتر خواهند کرد. از انواع آن می‌توان روش‌هایی مانند تحلیل مبتنی بر امضا، تحلیل لغوی، تحلیل‌های اکتشافیو پردازش زبان طبیعی را نام برد که همگی به این نوع عملیات فیلتر تعلق دارند (اسپیرین و هان[2]، ۲۰۱۲). برای مقابله با مشکلات روش های قبل، فیلترهایی بر مبنای تطابق رشته هم در سرآیند و هم در بدنه از طریق پیداکردن رشته های واقعی و استخراج آن به وجود آمده است. این که هرزنامه‌نویس‌ها کلمه‌ای مانند Viagra را به طور درست در هرزنامه قرار دهند، فیلتر به راحتی آن را پیدا خواهد کرد، اما هنگامی که آن را به صورت‌های مختلف به نمایش قرار می دهند، تشخیص آن سخت خواهد شد (وانگ و همکاران[3]، ۲۰۱۲).

اعتبارسنجی آدرس فرستنده

در این روش بر اساس یک سری از الگوریتم‌ها، میزان اعتبار آدرس فرستنده، مورد بررسی قرار می‌گیرد. در واقع هر آدرس پست الکترونیکی یک قالب مشخص دارد، لذا از طریق قالب و روش‌های دیگر میزان اعتبار آن بررسی خواهد شد (استاربوک و همکاران[4]، ۲۰۱۶)

منطق فازی

منطق فازی تکنولوژی نسبتا جدیدی است که در مقابل روشهای مرسوم برای طراحی و مدلسازی سیستمی که نیازمند ریاضیات و احتمالاتپیشرفته و نسبتا پیچیده می باشد، به کار می رود. از مقادیر و قوانین مبتنی بر متغیرهای زبانی و یا به عبارتی از دانش فرد خبره با هدف ساده، دقیق وکارآمد کردن طراحی سیستم استفاده میکنند (اسمیتسون[5]، ۲۰۱۶).

[1] Firte et al

[2] Spirin & Han

[3] Wang et al

[4] Starbuck et al

[5] Smithson

در سال ۲۰۱۱ ابو نیمه و همکارانش[1] نرم افزار دیفنسیو در فیسبوک را مورد مطالعه قرار دادند. این نرم افزار با استفاده از دسته بندی ماشین بردار پشتیبان به متن امتیاز دهی نموده و همزمان توسط میان اعتبار اکتشافی امتیازی به فرستنده می دهد. پس از آن میانگین امتیازهای ماشین بردار پشتیبان و میزان اعتبار قواعد اکتشافی پست ها را دسته بندی نموده و تشخیص می دهد که پست ارسال شده، هرزنامه یا مجاز است. نتایج این بر روی فیسبوک نشان می دهد، کسر بیشتری از پست ها هرزنامه و مقدار کمی از آن ها مخرب هستند که ۹٪ پست های کاربران فیسبوک هرزنامه و ۰.۰۳٪ پست ها لینک های مخرب هستند.

در تحقیقی دیگر که در سال ۲۰۰۷ توسط یولام و همکارانش[2] انجام شد از یک رویکرد یادگیری ماشین بر اساس ویژگی های استخراج شده از شبکه های اجتماعی در تبادل ایمیل های سیاه استفاده می کنند. نمونه برداری در این مقاله از مجموعه داده های Enronاستفاده شده و در آن ایمیل های موجود در صندوق پیام ۱۵۰ کاربر Enron (فرستندگان و گیرندگانی که با آدرس enron.com@هستند) مورد بررسی قرار گرفته است. نتایج این تحقیق نشان می دهد با انتخاب کمترین تعداد دروازه بان ها فیسبوک بالای ۴۰٪ و فلیکر بالای ۴۵٪ قابلیت دسترسی دارند.

وانگ[3] در سال ۲۰۱۰ تحقیقی بر روی توییتر انجام داد که سیستمی را طراحی می کرد که پیام های هرزنامه را تشخیص می دهد روابط دنبال کنندگان و دوستان در این شبکه با استفاده از مدل گراف اجتماعی مورد بررسی قرار گرفته است. در این سیستم از سیاست هرزنامه در توییتر، استفاده از سیستم های مبتنی بر محتوا و مبتنی بر گراف کمک گرفته شده است. طبق نتایج بدست امده از دسته بندی بیزین، ۳۹۲ کاربر به عنوان ارسال کننده هرزنامه شناخته شده اند که در میان ۳۴۸ کاربر به درستی تشخیص داده شده و این سیستم تشخیص هرزنامه، ۸۹٪ درست عمل نموده است. همچنین نتایج این تحقیق نشان می دهد هرزنامه در توییتر دربرگیرنده لینک می باشد.

در سال ۲۰۱۱ اواد و همکارانش[4] مشهورترین روش های یادگیری ماشین (مانند دسته بندی بیزین و ماشین بردار پشتیبان و سیستم ایمنی مصنوعی ماشین) و مساله طبقه بندی ایمیل های هرزنامه را مورد بررسی قرار داده اند و دقت آن ها و هرزنامه هایی که فراخوانی شده اند بررسی و طبقه بندی شده است، همچنین نشان داده شده کارایی الگوریتم بیزین بهتر بوده است.

در سال ۲۰۱۰ در مقاله ای یک روش جدید را برای تشخیص هرزنامه ها ارائه نمودند. به طوری که یک فرم افلاین را توسعه داده و از الگوریتم نزدیکترین همسایگی استفاده کرده است و مجموعه داده های ایمیل را از قبل برای برای فرایند یادگیری رده بندی می نماید و از این طریق برای شناسایی هرزنامه اقدام می نماید (فیرت و همکاران، ۲۰۱۰).

تحقیقی دیگر در سال ۲۰۱۱ با هدف تشخیص هرزنامه بر اساس داده کاوی برای امنیت شبکه های اجتماعی توسط زین جی[5] و همکارانش انجام شده است. در این تحقیق از الگوریتم خوشه بندی برای خوشه در مقیاس بزرگ و مجتمع کردن آن با طراحی الگوریتم یادگیری فعال در برخورد با چالش مقیاس پذیری و تشخیص زمان واقعی بکار گرفته شده است. با توجه به این که هرزنامه ها به طور گسترده مجموعه داده ها را از فیسبوک جمع آوری کرده فرایند فیلترینگ هرزنامه به اندازه کافی موثر نمی باشد. ولی سیستم ارائه شده قادر به اداره کردن در برخورد با تعداد زیادی از پست ها و نظارت بر زمان واقعی فعالیت های اجتماعی در شبکه های اجتماعی در شناسایی هرزنامه ها می باشد.

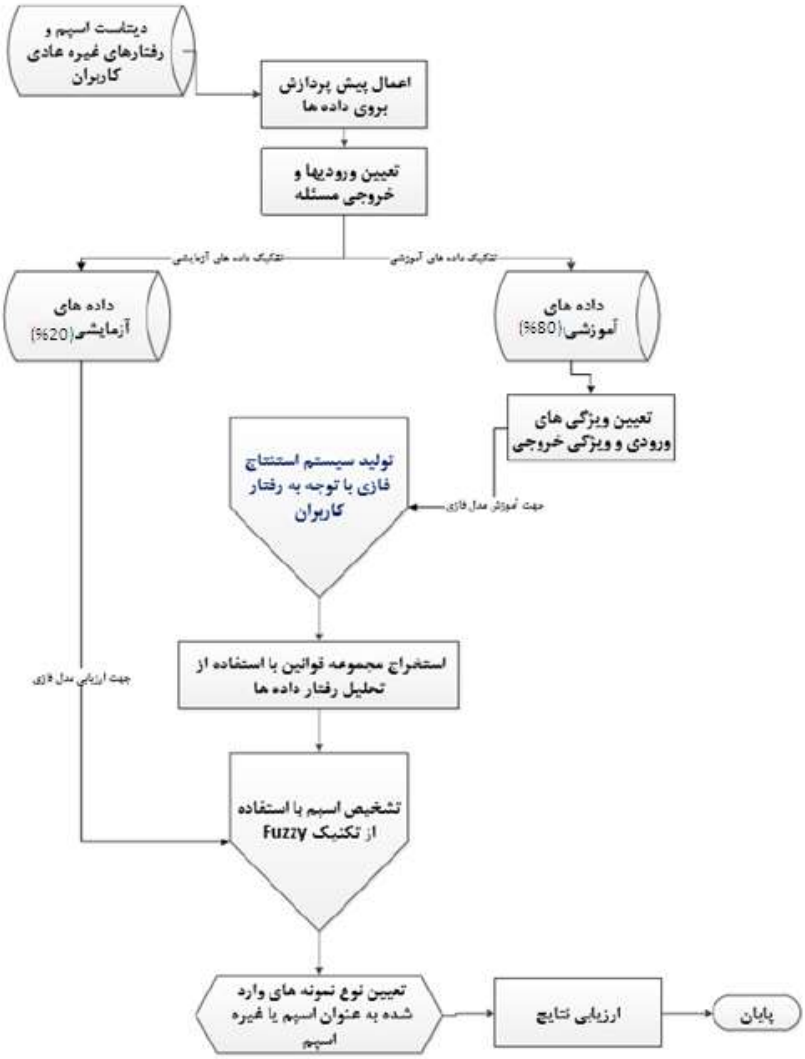
[1] Abu-Nimeh et al
[2] Yu Lam et al
[3] Wang
[4] Awad et al
[5] xin Jin et al

دوره ۳، شماره ۳، پاییز ۱۳۹۶، صفحات ۱۳-۱

لیانگ و همکارانش[1] بر اساس اعتبار بدست آمده از روابط اجتماعی کاربر، سیستمی جهت مسدود نمودن هرزنامه طراحی نموده اند. این سیستم با توجه به واکنش و پاسخ کاربران، خود را به روز می نمایدتا در آینده پاسخ بهتری در مقابل نامه های ناخواسته داشته باشد. این تحقیق بر روی نمونه های فیسبوک پیاده سازی شده است.

فلوچارت روش پیشنهادی

همانطور که از فلوچارت اصلی روش پیشنهادی مشخص است، ابتدا داده های مربوط به رفتارهای غیره عادی کاربران که به عنوان اسپ در شبکه های اجتماعی شناخته می شود به عنوان منبع داده خام به سیستم پیشنهادی وارد می شود. سپس بر روی داده های مربوطه عملیات پیش پردازش انجام شده و نمونه های بلا استفاده و پرت حذف می شوند. عملیات پاکسازی داده نیز در همین مرحله انجام می شود. پس از اعمال پیش پردازش بر روی داده ها، ورودی ها و خروجی مسئله مشخص می گردد. به عبارتی فیلدهایی که به عنوان ووردی مسئله در نظر گرفته می شوند در این مرحله مشخص شده و در نهایت فیلد نهایی که تنها یک فیلد است به عنوان خروجی نمونه ها تعیین می گردد.



شکل ۱: فلوچارت روش پیشنهادی

[1] Liang et al

تجزیه و تحلیل اطلاعات:

با توجه به شبیه سازی های انجام شده، ابزار مورد استفاده جهت پیاده سازی روش پیشنهادی در این تحقیق زبان برنامه نویسی متلب و رپیدماینر است. ابتدا منبع داده مربوطه معرفی و ویژگی های مربوط به آن تشریح می گردد. پس از شبیه سازی روش پیشنهادی، کلیه نتایج و یافته های بدست آمده تشریح شده و در قالب نمودار های مختلف ارائه می گردد. در نهایت نیز نتایج بدست آمده با سایر روشهایی که تا کنون مطرح شده اند ارزیابی و مقایسه می شود.

منبع داده (دیتاست)

برای شبیه سازی راهکارهای ارایه شده از مجموعه داده MCFU[1] استفاده شده است این مجموعه داده نسخه های متفاوت با حجم های مختلفی ارایه شده است که سعی شده است تا از آخرین نسخه برای شبیه سازی جهت تشخیص اسپم استفاده گردد. بدین منظور از نسخه ۱۰ استفاده شده است. این مجموعه داده دارای حجمی معادل5GB می‌باشد که بعلت عدم قابلیتهای سخت افزاری برای پردازش داده مزبور، از ۱۰ درصد مجموعه داده استفاد شد.

این مجموعه داده دارای ۱۵ ویژگی با مقادیر اکثرا متنی است که در مرحله پیش پردازش داده‌ها لازم به تغییر به مقادیر عددی دارد که بدین منظور ابتدا مقادیر کلیه ویژگی‌ها شناسایی و مقادیر عددی نسبت داده شد. در جدول (۱-۱) و (۱-۲) تعدادی ویژگی‌های تغییر یافته ارایه شده است.

جدول۱: تبدیل مقادیر ویژگی پروتکل به اعداد

	نام پروتکل	معادل فارسی
مشخصه پروتکل	Arp	1
	Esp	2
	Icmp	3
	Tcp	4
	Udp	5

با توجه به جدول فوق مقادیر مشخصه های پورتکل نوع پروتکل استفاده شده را نشان می دهد؛ مثلا پروتکل tcp یا پروتکل udp و غیره. در کل این مقادیر بیانگر این است که اسپ از طریق چه پروتکلی در شبکه ارسال شده است.
بطور کلی در جدول فوق اسامی برخی از پروتکل های محبوبی ارائه شده است که اسپم ها از طریق آنها به سیستم اعمال می شوند. لازم به ذکر است که در موارد کاربرد مثل شبکه های اجتماعی فیس بوک، توئیتر و غیره بر اساس همین پورت ها می توان بسیار از اسپم ها را فیلتر نموده یا از بروز آین موارد جلوگیری به عمل آید.

[1] [http://social.technet.microsoft.com/wiki/contents/articles/4399.private-cloud-reference-model.aspx.

دوره ۳، شماره ۳، پاییز ۱۳۹۶، صفحات ۱۳-۱

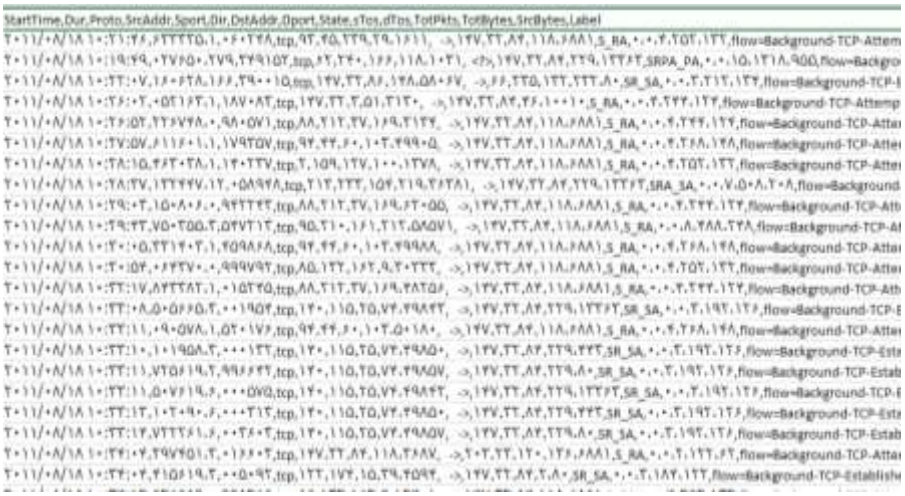
جدول۲: تبدیل مقادیر ویژگی State به عدد۹

	وضعیت	معادل فارسی
مشخصه های وضعیت	A	0
	CON	1
	DNP	2
	ECO	3
	ECR	4
	FA	5
	FSA	6
	INT	7
	PA	8
	RED	9
	ROB	10
	S	11
	TRC	12

همانطور که از جدول ۲ مشخص است مقادیر ویژگی هایی که به صورت رشته ای هستند در قالب عدد خلاصه سازی می شوند تا عملیات پردازش بهتر صورت گیرد. مشخصه های پروتکل و مشخصه های وضعیت برخی از ویژگی هایی هستند که بیانگر نوع اسپم و مقادیر اختصاص داده شده به هر اسپم یا رفتار غیره عادی در شبکه اجتماعی است. بطور کلی میزان خرابی اسپم ها و خطرناک بودن آنها با توجه به اعداد ۰ تا ۱۲ طبقه بندی شده اند. هر چه اعداد بزرگتر باشند نوع اسپم مربوطه از درجه بالاتری از لحاظ میزان خرابی برخوردار است.

همچنین در جدول ۳ نمونه ای از مجموعه داده بصورت متنی ارائه شده است.

جدول۳: نمونه ای از مجموعه داده با مقادیر متنی



همانطور که در جدول ۳ دیده می شود داده های اصلی دارای مقادیر متنی است که جهت پردازش و اجرای روش پیشنهادی میبایست تبدیل به مقادیر عددی شده و در نهایت نیز نرمال سازی شوند. با ایجاد نرمال سازی بر روی داده های عملیات خوشه بندی و اجرای الگوریتم رقابت استعمار با دقت و سرعت بهتری صورت می گیرد.
در قسمت زیر متغیر های مربوط به مجموعه داده نشان داده شده است.

- زمان شروع: این ویژگی زمان وقوع عمل انجام شده که اسپم یا غیره اسپم بوده است را نشان می دهد. این ویژگی مشخص می کند که اسپم در چه زمانی بر میلی ثانیه اتفاق افتاده است.
- مدت زمان فعالیت: این ویژگی تعیین می کند که عمل انجام شده (اسپم) چه مدت زمان به طول انجامیده است. به عبارتی زمان ماندگاری اسپم در رسانه اجتماعی را بیان می کند.
- نوع: اسپمی که اتفاق می افتد یا عملیاتی که صورت می گیرند انواع مختلفی دارند که به صورت شماره گذاری شده مشخص می شوند.
- آدرس مبدا: آدرس ارسال کننده یا مبدا را نشان می دهد.
- پورت مبدا: آدرس دریافت کننده یا مقصد را نشان می دهد.
- مسیر: مسیری مورد نظر را تعیین می کند.
- آدرس مقصد: آدرس مقصد را مشخص می کند.
- پورت مقصد: پورتی که عمل مورد نظر از مقصد در آن مستقر می شود را مشخص می کند.
- وضعیت: وضعیت بسته یا اسپم را مشخص می کند. این مشخصه در جدول ۴-۲ به صورت کامل تشریح شده است.
- وضعیت مبدا: وضعیت مبدا را مشخص می کند.
- وضعیت مقصد: این ویژگی وضعیت مقصد را مشخص می کند.
- تعدادبسته ها: تعداد بسته هایی که در عمل مورد نظر وجود دارد را مشخص می کند.
- مجموع بایتها: این ویژگی مجموع بایتهای کل بسته را مشخص می کند.
- تعدادبایتهای مبدا: تعداد بایتهای پیام مبدا را مشخص می کند.
- برچسب: این ویژگی بیان کننده اسپم است و رفتارهای عادی.

دریافت داده ها

در این پژوهش از راهکار فازی استفاده خواهد شد که در این راهکارها جهت ارزیابی نتایج بایستی علاوه بر داده های آموزش، داده هایی آزمایش نیز لازم می‌باشد. بدین منظور کل داده پیش پردازش شده به دو قسمت داده های آموزش و داده های آزمایش تقسیم شده اند که داده های آموزش شامل ۷۰ درصد داده ها و داده های آزمایش شامل ۳۰ درصد می‌باشد. با توجه به نرمال سازی داده ها، همه فیلدها دارای مقادیر معتبر بوده و قابلیت شرکت در شبیه سازی را داشتند.

مشخصات سیستم مورد استفاده

روش پیشنهادی در این پژوهش با استفاده از شبیه ساز متلب پیاده سازی شده است. همچنین در جدول زیر مشخصات مربوط به سیستمی که پیاده سازی روش پیشنهادی و ارزیابی نتایج در آن انجام شده نشان داده می شود.

جدول۲: مشخصات سیستم جهت شبیه سازی و ارزیابی نتایج

سخت افزار / نرم افزار	مشخصات
سیستم عامل	ویندوز ۷
نوع سیستم عامل	سیستم عامل ۳۲ بیتی و ۶۴ بیتی
حافظه RAM	4 گیگابایت –۳/۰۶ گیگ قابل استفاده
پردازنده	پردازنده اینتل – تعداد هسته ها 7(Core™)i7 CPU – Q 720 @ 1.60GHz 1.60 GHz

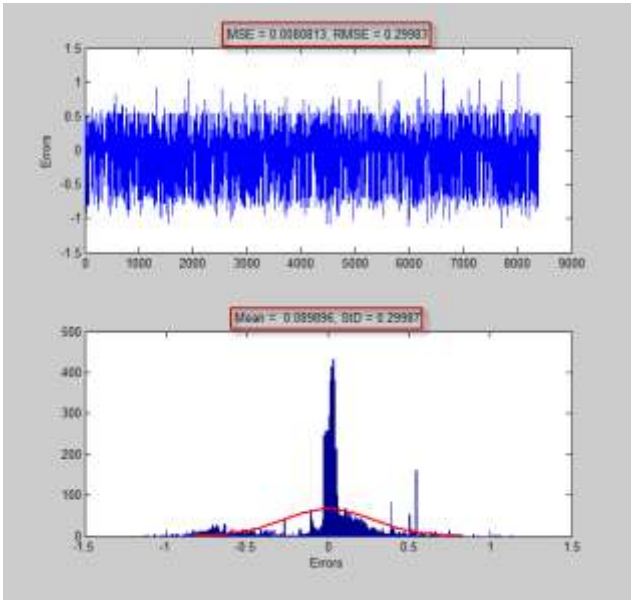
بنابراین با توجه به یک سیستم با مشخصات فوق شبیه سازی مربوطه انجام شده و نتایج ارزیابی شده است. لذا در همین بخش به صورت کامل نتایج بدست آمده تشریح می گردد.

نتایج شبیه سازی روش پیشنهادی

در این قسمت با توجه به شبیه سازی انجام شده بر روی منبع داده معرفی شده دقت و خطای تشخیص اسپم محاسبه گردیده است که در نهایت با سایر روشها مورد مقایسه قرار می گیرد. بطور کلی در طی دو مرحله داده ها شبیه سازی شده و نتایجی محاسبه گردیده است:

- یکی در حالت آموزش
- در حالت آزمایش

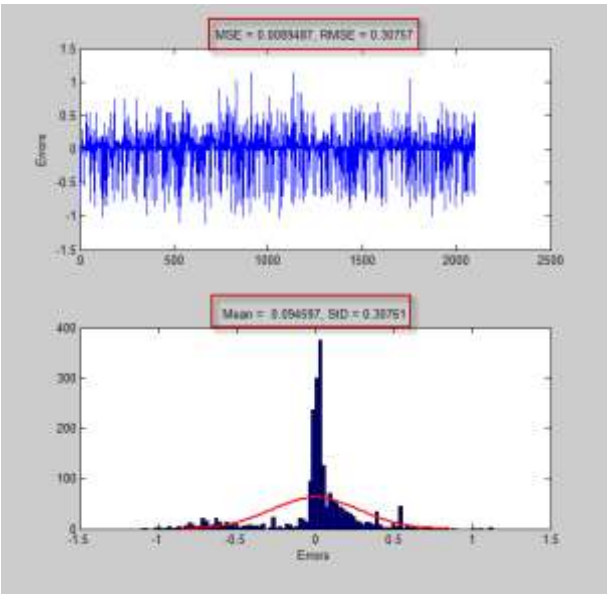
در شکل زیر نتایج مربوط به اجرای روش پیشنهادی در مرحله آموزش نشان داده شده است. در شکل زیر نیز مرحله آموزش روش پیشنهادی و تست بر روی ۸.۰۰۰ داده صورت گرفته است.



شکل ۲: مرحله آموزش روش پیشنهادی و محاسبه نتایج (الف)

شکل (۲) دارای دو بخش بوده که بطور کلی میزان خطای آموزش را نشان می دهد. در بخش بالا میزان خطای تشخیص اسپم برای ۸۰۰۰ نمونه محاسبه می گردد که محور افقی بیانگر تعداد نمونه های آموزش و محور عمودی نیز میزان خطای تشخیص به ازای هر نمونه را نشان می دهد.

همانطور که از شکل بالا دیده می شود میزان خطای روش پیشنهادی در مرحله آموزش بر روی داده های اسپم و غیره اسپم در حدود ۰.۰۸۹۸ است. دقت تولید مدل مربوطه جهت تشخیص اسپم در مرحله آموزش تقریبا برابر با ۹۹.۹٪ است که این دقت نسبت به سایر روشها در حدود ۰.۸ تقریبا بهتر می باشد. از طرفی معیار های RMSE و MSE نیز محاسبه شده اند که از این معیار ها نیز می توان نتایج مختلفی استنباط گردد. انحراف معیار روش پیشنهادی در مرحله آموزش نیز برابر با ۰.۲۹۹ است که نشانگر عملکرد بسیار خوب روش پیشنهادی در مرحله آموزش با استفاده از منطق فازی بر روی داده های اولیه موجود در شبکه اجتماعی است. در شکل زیر نیز مراحل آزمایش بر روی مجموعه داده های ۲.۰۰۰ نمونه ای صورت گرفته است.

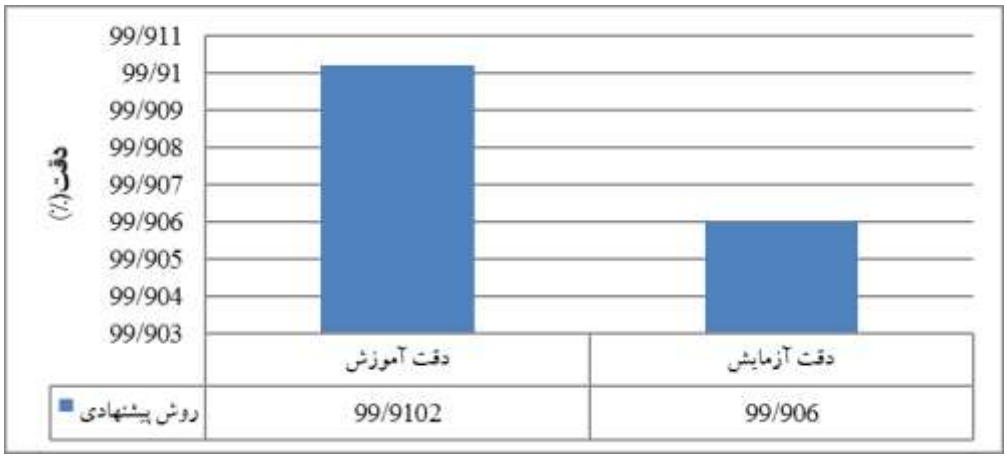


شکل ۳: مرحله آزمایش روش پیشنهادی و محاسبه نتایج (ب)

شکل (۳) دارای دو بخش بوده که بطور کلی میزان خطای آزمایش را نشان می دهد. در بخش بالا میزان خطای تشخیص اسپم برای ۲۰۰۰ نمونه محاسبه می گردد که محور افقی بیانگر تعداد نمونه های آزمایشی و محور عمودی نیز میزان خطای تشخیص به ازای هر نمونه را نشان می دهد.

همانطور که از شکل بالا دیده می شود میزان خطای روش پیشنهادی در مرحله آموزش بر روی داده های اسپم و غیره اسپم در حدود ۰.۰۹۴۵ است. دقت تولید مدل مربوطه جهت تشخیص اسپم در مرحله آزمایش تقریبا برابر با ۹۹.۹٪ است که این دقت نسبت به سایر روشها در حدود ۰.۸ تقریبا بهتر می باشد. از طرفی معیار های RMSE و MSE نیز محاسبه شده اند که از این معیار ها نیز می توان نتایج مختلفی استنباط گردد. انحراف معیار روش پیشنهادی در مرحله آموزش نیز برابر با ۰.۳۰۷ است که نشانگر عملکرد بسیار خوب روش پیشنهادی در مرحله آزمایشبا استفاده از منطق فازی بر روی داده های اولیه موجود در شبکه اجتماعی است. از نکات برجسته اجرای این الگوریتم‌ها تعداد زیاد رکوردهای بخش آموزش و آزمایش بود که منجر به ارایه خروجی دقیق‌تری گردیده است.

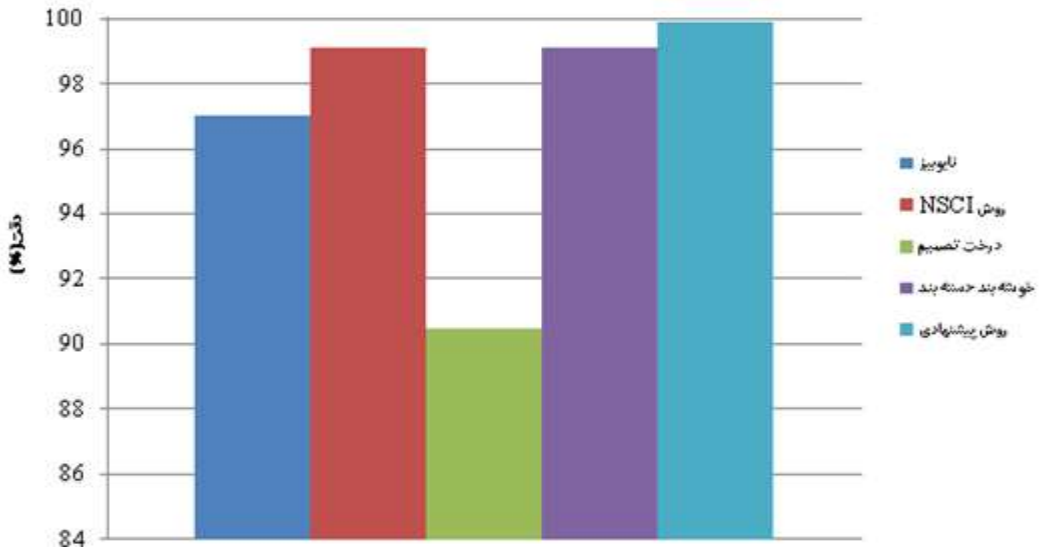
با توجه به شبیه سازی روش پیشنهادی و مشاهده نتایج بدست آمده، قسمت هایی از نتایج روش پیشنهادی از لحاظ میزان خطا و دقت در مرحله تست نسبت به سایر روشها در قسمت زیر بحث می گردد. در نمودار زیر مقایسه دقت مرحله آزمایش و آموزش داده های اسپم و غیره اسپم در روش پیشنهادی نشان داده شده است.



شکل ۴: مقایسه دقت مرحله آزمایش و آموزش داده های اسپم و غیره اسپم

همانطور که دیده می شود میزان دقت در مرحله تست مطرح است که در حدود ۰.۰۰۴ با هم تفاوت دارند که این میزان قابل مطرح نیست؛ بنابراین با این میزان دقت در تشخیص اسپم مراحل مقایسه با سایر روشها صورت می گرد. با توجه به مقایسه دقت روش پیشنهادی با مقاله پایه (گوپتا[1]، ۲۰۱۵) در جدول زیر میزان بهبود روش مطرح شده در این پژوهش نشان داده شده است.

مقایسه دقت تشخیص اسپم در روش پیشنهادی نسبت به سایر روشها



شکل ۵: مقایسه میزان دقت روش پیشنهادی نسبت به سایر روشها

[1] Gupta

بنابراین با توجه به شبیه سازی مسئله مطرح شده در این پژوهش با کمک منطق فازی مشاهده گردید که دقت روش پیشنهادی جهت تشخیص اسپ در رسانه های اجتماعی مثل توئیتر در حدود ۹۹.۹٪ بوده است. همچنین با بررسی مقاله پایه دیده شده است که دقت آنها در حدود ۹۹.۱ در بهترین حالت نیز بوده است. بطور کلی در مقایسه با روشهای بحث شده، میزان بهبود دقت روش پیشنهادی نسبت به روشها خوشه بندی، نایو بیز، درخت تصمیم و روش NSCI[1] به ترتیب در حدود ۱.۰۰۸٪، ۱.۰۲۹٪، ۱.۱۰۳ و ۱.۰۰۸٪ بهبود داشته است.

نتیجه گیری

تشخیص اسپم در شبکه های اجتماعی امروزه تبدیل به یک مشکل اساسی و چالش بزرگ برای هم کاربران و هم مدیران سیستم شده است. وجود روش یا الگوریتم هایی که بتواند به صورت مطلوب این نوع مزاحمت ها و مشکلات را مرتفع نماید بسیار لازم و ضروری است. با توجه با این دیدگاه در این پژوهش از یک سیستم فازی جهت تشخیص اسپم ها استفاده شده است. بوطر کلی روش مطرح شده در این تحقق به مراحلی تقسیم بندی شده است که عبارتند از: ۱) ابتدا دیتاست اصلی که حاوی برخی رفتار های غیره عادی به عنوان اسپم است به سیستم یا چارچوب پیشنهادی وارد می شود. این دیتاست شامل ۱۰.۵۰۰ نمونه است.۲) پس از اینکه داده ها به سیستم وارد شدند، ویژگی های ورودی و ویژگی خروجی شبیه سازی تعیین می گردد. ۳) پس از این کار لازم است داده های استفاده شده به دو دسته آموزشی و آزمایشی تقسیم بندی شوند. از داده های آموزشی جهت تولید قوانین و از داده های آزمایشی به منظور ارزیابی مدل و روش پیشنهادی استفاده می گردد. در این تحقیق ۸۰٪ از داده ها به عنوان داده های آموزشی و ۲۰٪ از داده ها نیز به عنوان داده های آزمایشی منظور می گردد.۴) اجرای روش فازی بر روی داده های آموزشی جهت تولید مجموعه قوانین. سیستم استنتاج فازی جهت استخراج قوانین مربوطه ابتدا از FCM استفاده نموده تا بتواند تعداد قوانین را تعیین کرده و در نهایت با توجه به تابع عضویتش قسمت مقدم و تالی قوانین را تولید می نماید. نوع روش فازی استفاده شده در این پژوهش روش سوگنو (Sugeno) می باشد. ۵) ارزیابی نتایج با اعمال داده های تست.

پس از شبیه سازی روش پیشنهادی مشاهده گردید دقت مدل ما در حدود ۹۹.۹٪ بوده است که نسبت به سایر روشها به صورت میانگین در حدود ۱.۲٪ بهبود داشته است.

منابع:

1. Alex Hai Wang, DON’T FOLLOW ME Spam Detection in Twitter, IEEE, Proceedings of the International Conference on Security and Cryptography (SECRYPT), pp 1-10, 2011
2. Ch. Liang, Y. Chen, G. Liao and B. Cheng. (2010). Anti-spam Email System in Facebook,Proc. IEEE Symp. Email System in Facebook, Proc. IEEE Symp. Computer, pp. 183-186, Dec. 2010, doi: 10.1109/ COMPSYM.5685522.
3. Gupta, R. Kaushal, Improving Spam Detection in Online Social Networks, Springer, pp. 1-2, 2015.
4. HoYu Lam, DitYan Yeung, A Learning Approach to Spam Detection based on SocialNetworks:http://www.cse.ust.hk/~dyyeung/paper/pdf/yeung.ceas%2007.pdf

[1] Non Spammer Correctly Identifierd

5. L. Firte, C. Lemnaru, R. Potolea. (2010).Spam Detection Filter using KNN Algorithm andResampling, IEEE, Intelligent Computer Communication and Processing (ICCP), pp27 -33.
6. Saeed Abu-Nimeh, Thomas M. Chen and Omar Alzubi, (2011).Malicious and Spam Posts in Online Social Networks, IEEE Computer Society, pp 23-28.
7. Smithson, M. (2016). Fuzzy Sets and Fuzzy Logic in the Human Sciences. InFuzzy Logic in Its 50th Year (pp. 175-186). Springer International Publishing.
8. Spirin, N., & Han, J. (2012). Survey on web spam detection: principles and algorithms. ACM SIGKDD Explorations Newsletter, 13(2), 50-64.
9. Starbuck, Bryan T., et al. Advanced spam detection techniques, U.S. Patent No. 9,305,079. 5 Apr. 2016..
10. W.A. Awad,S.M. ELseuofi. (2011).MACHINE LEARNING METHODS FOR SPAM E-MAILCLASSIFICATION, International Journal of Computer Science & Information Technology IJCA (IJCSIT), Vol 16, No 1, pp.39-45.
11. Xie, S., Wang, G., Lin, S., & Yu, P. S. (2012, August). Review spam detection via temporal pattern discovery. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 823-831). ACM.
12. Xin Jin,Cindy Xide Lin. (2011). SocialSpamGuard: A Datamining Based Spam detection System for Social Network, https://netfiles.uiuc.edu/xinjin3/.../VLDB11SocialSpamGuard.pdf, 2011, pp 1458- 1461.

مطالعات علوم کاربردی در مهندسی

دوره ۳، شماره ۳، پاییز ۱۳۹۶، صفحات ۱۳-۱

Detection of The Spam on the Social Networks Based on User Behavior using Fuzzy Theory

Farzaneh Parganeh[1], Mohsen Firuzbakht[2]

1. *MSc, Department of Computer Engineering, Faculty of Engineering, Islamic Azad University, Electronic Department*
2. *Ph.D., Assistant Professor, Department of Computer Engineering, Islamic Azad University, Tehran South Branch*

Abstract

The huge volume of spam and related attacks on social networks such as Facebook, Twitter, etc. has had many problems with these networks. On the other hand, it is complicated to provide an optimal way to detect spam with acceptable accuracy and to some extent reduce social media problems. Therefore, the existence of a reliable and efficient method in this field is very important. In this research, using a fuzzy system, a method has been developed that can identify spam in the social network to a very good degree and can significantly improve this information burden. In general, the proposed method has the following steps: 1) Data entry on the social network 2) Pre-processing on data and removal of unused data 3) Breakdown of experimental data and training 4) Implementation of FCM fuzzy system Generate rules 5) Apply fuzzy system generated on the data and detect spam. Finally, after simulating the proposed method, it was observed that the accuracy of the presented method was 99.9%, which improved significantly compared to the methods that have been proposed so far.

Keywords: Social Network, Spam Detection, Fuzzy Theory