فصلنامه پژوهش در علوم، مهندسی و فناوری
دوره ۳، شماره ۳، پاییز ۱۳۹۶
صفحات ۳۶-۴۲
ISSN: 2476-5082
www.uctjournals.com

تحلیل تصاویر روند در شبکه های اجتماعی

نسترن نادری بلداجی

دانشجوی کارشناسی ارشد مهندسی کامپیوتر-نرم افزار دانشگاه غیرانتفاعی اروندان خرمشهر
Nastarannaderi1@gmail.com

چکیده

امروزه حجم زیادی از تصاویر با تگ های مشابه توسط کاربران در شبکه های اجتماعی آنلاین به اشتراک گذاشته میشود و کاربران دیگری این تصاویر را مبادله می کنند، بدون آنکه سندیت آن را بررسی کنند؛ در مورد اینکه کدام تصویر واقعی است و کدام تصویر جعلی شبهاتی بوجود می آید و در مواقع حساسی که ما مجبور به استناد به اطلاعات منتشر شده میشویم باید بتوانیم جعلی یا حقیقی بودن آن را بررسی کنیم.

در این مقاله روشهای کشف روندهای مشابه بررسی شده و با توجه به اینکه محیط شبکه های اجتماعی ماهیتی ناهمگن دارد ، برای کشف روندهای مشابه از الگوریتم های هوشمند محاسباتی بهینه سازی ازدحام ذره استفاده شده و همچنین برای کشف تگ های توصیفی مشترک، الگوریتمی به نام الگوریتم جستجوی مشترک به کار گرفته شده است. پس از آنکه تصاویر ترند شده در شبکه اجتماعی شناسایی شدند باید اعتبارسنجی شوند که کدام تصاویر واقعی و کدامیک جعلی هستند و این فرآیند، با استفاده از تکنیک های ماتریس هم رخداد که برای تشخیص تصویر حقیقی وجود دارد، جعلی یا حقیقی بودن تصاویر به اشتراک گذاشته شده بررسی خواهند شد . و در پایان به نتیجه گیری در مورد الگوریتم های مختلف خواهیم پرداخت.

واژه‌های کلیدی: : فضای مجازی، شبکه های اجتماعی، الگوریتم جستجوی مشترک، بهینه سازی ازدحام ذره، جعل تصویر، ماتریس هم رخداد

۱- مقدمه

با رواج پیدا کردن شبکه های اجتماعی آنلاین تعداد زیادی از افراد با این رسانه های اجتماعی (تلگرام، واتس آپ، توئیتر) به صورت آنلاین در هر زمان و از هر مکانی با یکدیگر در ارتباطند. کاربران در محیط شبکه های اجتماعی ، حجم زیادی از محتواهای رسانه ای را مبادله می کنند و به اشتراک میگذارند و اخبار جدید و وقایع اخیر را دنبال می کنند. مثلاً هر دقیقه هزاران توئیت روی توئیتر فرستاده می شود و چند هزار تصویر ویدئویی روی یوتوب آپلود میشود. تعداد بیشماری از این پست های منتشر شونده در رواج یافتن یک موضوع یا شناخته شدن آن تاثیر می
گذارند که اصطلاحاً به آن روال شدن یا ترند شدن گفته می شود. خصوصیت بارز شبکه های اجتماعی آنلاین این است که به سرعت افزایش می یابد و کاربران در تولید محصولات جدید شرکت داده میشوند. در این مقاله، هدف بررسی تکنیک های شناسایی و کشف منابع و محتواهای یکسان و جعلی بودن یا نبودن آنهاست. این مقاله از نوع مروری و از نظر روش جمع آوری داده ها اسنادی یا کتابخانه ای محسوب میشود. سؤالهای اساسی مقاله عبارتند از:

- ✓ انتخاب اولیه برای شناسایی Fake در شبکه های مختلف بر چه اساسی باشد؟
- ✓ استخراج داده های مشکوک چگونه انجام شود؟
- ✓ طبقه بندی داده های مورد بررسی با چه راهکاری انجام شود؟
- ✓ تکنیک تشخیص جعل یا حقیقی بودن سند چه باشد؟

با محبوبیت گسترده و استفاده وسیع از شبکه های اجتماعی در سالهای اخیر کاربران می توانستند محتوای رسانه ای را از طریق این رسانه ها به اشتراک بگذارند و با افزایش تقاضا بین افراد ، شناسایی تصاویر و تگ های به اشتراک گذاشته شده بین افراد، اهمیت پیدا کرد. محققان مختلفی تاکنون روشهای مختلفی برای کشف تصاویر و محتوای ترند بررسی کرده اند: یکی از این روشها، تکنیک Social Tagging بوده که بهبود قابل توجهی در کشف آیتم های مربوط و نامربوط برای پروفایل هر کاربر دارد. محققان شروع به استفاده از برچسب زدن اجتماعی به نام Folksonomy کردند تا یافتن منابع مناسب برای هر پرس و جو براساس سلیقه و تنظیمات خود کاربر باشد. کاربران Folksonomy به سلسله مراتب یا ساختار خاصی وابسته نیستند، آنها مجاز به استفاده از هر کلمه یا ترکیبی از کلمات چندگانه برای توصیف و حاشیه نویسی منابع هستند[8]. اما نظریه دیگر این بود که اکثر کاربران دامنه های مختلف ممکن است دیدگاه و درک مختلفی برای یک مورد یا برچسب داشته باشند (یانگ و همکاران)
Balakrishnan و همکاران. [۴] مدل های چهره را برای تگ عکس ها مطالعه کرده و سپس عکس های برچسب شده را جستجو کردند. هات و همکاران [۱۰] یک طرح جدید برای افزایش دقت بازیابی تصویر مبتنی بر برچسب پیشنهاد می کند. تصاویر در این طرح با شباهت معنایی و اطلاعات تولید شده گروه بندی می شوند. ژائو و همکاران [۲۹] آیتم برچسب گذاری شده و نمایه کاربر را به عنوان رتبه بندی مبتنی بر گرافیک از چند نوع آیتم های مرتبط مدل سازی کردند. در این مقاله استدلال این است که کاربرانی که تصاویر بیشتری به اشتراک میگذارند بین خودشان ارتباطات قویتری وجود دارد. تحلیل تصاویر مشترک یا به اشتراک گذاشته شده می تواند مبنایی برای تحلیل ارتباطات بین انتشار دهندگان باشد. از آنجا که شبکه های اجتماعی آنلاین، پویا و ناهمگن هستند ؛ اخیراً ، برای توسعه راه حل های پیشنهاد شده از الگوریتم های هوشمند محاسباتی الهام گرفته از طبیعت CI (الگوریتم تکامل ژنتیک ، ازدحام ذره و ...) استفاده میشود. Zhao Xing و همکارانش از الگوریتم PSO استفاده کردند.[24]سلیمانی پور و همکارانش از ترکیب الگوریتم های PSO , ACO برای پیدا کردن روابط در شبکه های اجتماعی بهره بردند.[27]
الگوریتم PSO از روش‌های فرا ابتکاری است که با الهام گیری از رفتار اجتماعی گروهی از پرندگان مهاجر که در تلاش برای دستیابی به مقصد ناشناخته ای هستند، توسط کندی و ابرهارت در سال ۱۹۹۵ میلادی توسعه داده شده است. در الگوریتم CPSO ذرات در زیرگروههایی قرار میگیرند و الگوریتم PSO استاندارد در هر زیرگروه به طور جداگانه اجرا میشود. البته لازم

به ذکر است که در به روز رسانی ذرات در یک گروه، از اطلاعات بهینه گروههای دیگر نیز استفاده میشود و تبادل اطلاعات از نوع همکاری است.

ما در این مقاله فقط علاقه خاصی به تمرکز روی تصاویر روند داریم و بررسی می کنیم که آیا تصویر ترند شده توصیفی واقعی از ادعایی را که ارائه می کنند، به ما نشان می دهند یا خیر؟ در تمامی بررسی های صورت گرفته مدل با جمع آوری تصاویر روند بر اساس علائم و پروفایل های کاربران شروع می شود.

این مقاله در ادامه به این شکل سازمان دهی شده است: بخش ۲ فضای مدل را ترسیم می کند. بخش ۳ روش های اخیر برای کشف تصاویر روند بررسی می کند. بخش ۴ روش کشف تصاویر جعلی را ارائه می کند. بخش ۵ نتایج مقاله را بازگو می کند.

۲- ترسیم فضای مدل

ما فرض میکنیم هر محتوایی که توسط کاربر در شبکه قرار داده شده برای شناخت پروفایل کاربر ارزشمند است. و علاوه بر محتوای قرار داده شده، علائم ایجاد شده توسط کاربر را هم درنظر میگیریم؛ ابتدا علائمی که برای توصیف تصاویر استفاده شده اند را جستجو می کنیم. معمولاً کاربران با به اشتراک گذاشتن برخی محتواها یا تصاویر تعاملات را ایجاد میکنند. مثلا یک کاربر تصویری را روی توئیتر قرار میدهد. و زیر آن جمله ای مینویسد و علائم توصیفی بیشتری را برای برجسته کردن اطلاعات ایجاد میکند. بر اساس جمع آوری این اطلاعات میتوان مشخصه های تعاملات خطی بین کاربران مانند تگ ها، تعداد درخواست های کاربران، نوشته های روی دیوار پروفایل ها را توسعه داد. الگوریتم های زیادی برای یافتن اسناد مشابه توسط محققین مورد مطالعه قرار گرفتند ما در این مقاله تکنیک های سالهای اخیر را بررسی میکنیم. از جمله الگوریتم های استفاده شده برای یافتن روابط در شبکه های اجتماعی در سال های اخیر عبارتند از الگوریتم جستجوی مشترک، الگوریتم پیچیده درهم سازی حساس به مکان ، LSH الگوریتم های هوشمند محاسباتی ACO , PSO که در کوتاهترین زمان و با کاهش فضای محاسبات پاسخ بهینه را به ما خواهند داد.

۳- بررسی الگوریتم های کشف روابط در شبکه های اجتماعی

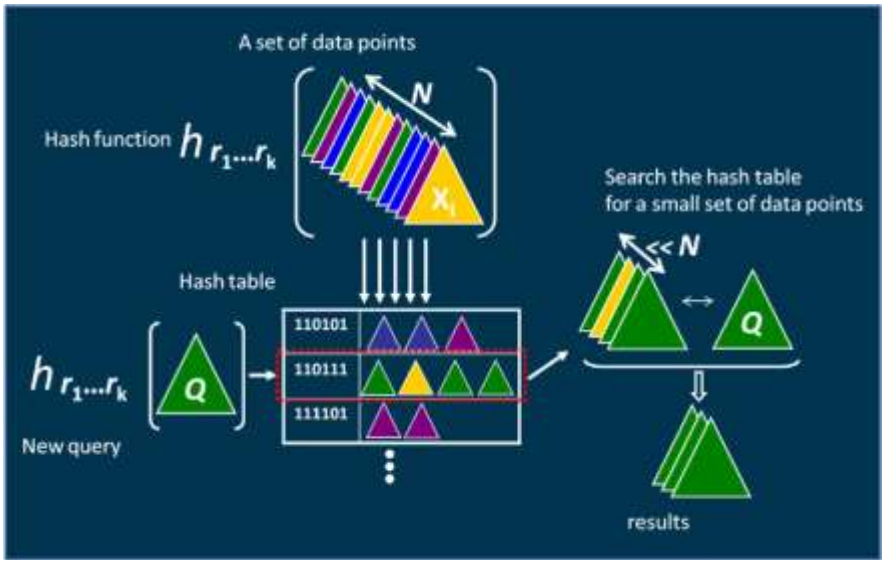
با توجه به رشد سریع و روزافزون شبکه های اجتماعی و گسترش حجم اطلاعاتی که روی این رسانه ها ردو بدل میشوند، تشخیص اجتماع در شبکه های به منظور افشاسازی رویدادها اهمیت ویژه ای می یابد و از آنجائیکه داده های زیادی برای یافتن روندها در این نوع از شبکه ها نداریم ، ارتباطات میان کاربران مهم می شود. در شبکه های اجتماعی کاربران برای داشتن تعامل با یکدیگر برخی محتواها مانند تصاویر را به اشتراک می گذارند و برای توصیف احساسات عاطفی خود ، از علائم در کنار محتوای رسانه ای پست شده استفاده می کنند. بنابراین ما برای بدست آوردن روابط، بانک اطلاعات کلیه لغات احساسی و کلیدی و تگ هایی که کاربران بکاربرده اند ، را باید ایجاد کنیم.

۳-۱.الگوریتم درهم سازی حساس به مکان یا موضع

یکی از تکنیک ها برای یافتن اسناد مشابه عملیات درهم سازی حساس به مکان یا موضع است که ابتدا توسط Broder[44] معرفی شد. آنها از تکنیک LSH استفاده کردند که روش approximate nearest neighbot نامیده می شود.

ایده اصلی در LSH اعمال توابع درهم سازی در حالتی است که احتمال تصادم یا برخورد برای اسناد مشابه نسبت به آنها که مشابه نیستند بیشتر است. فرض کنید n مجموعه ای از نقاط داده شده در فضای S باشد؛ ابتدا نقاط داده شده با استفاده از تابع درهم سازی H پیش پردازش شده و در جدول درهم سازی ذخیره می شوند. مقادیر درهم سازی از نقاط داده ، به عنوان شاخصی از جدول درهم سازی عمل می کنند.برای تعیین شاخص از جدول درهم سازی ، از طریق مجموعه ای از نقاط داده که با هم ارتباط نزدیک دارند ؛استفاده می شود.

۳-۲.الگوریتم مبتنی بر هوش محاسباتی



در این مدل برای رسیدن به اهداف ذکر شده ابتدا در شبکه های اجتماعی، علائمی که برای توصیف تصاویر استفاده شده اند جستجو می شوند. لیستی از لغات کلیدی احساسی / عاطفی برای جلوگیری از فهم اشتباه الگوریتم و بهبود عملکرد آن ایجاد میشود. برای توصیف الگوریتم ابتدا لیستی از مفاهیم در رابطه با توصیف فرآیند به صورت زیر در نظر گرفته می شوند:

U = { u1 , u2 , u3, … ,u|U|} مجموعه ای از کاربران.

C = { c1, c2 ,c3, … ,c|C|} مجموع همه محتواهای رسانه ای.

T ={ t1, t2, t3, …, t|T|} مجموعه ای از همه علائم (تگ).

ما سه ماتریس اصلی داریم:

ماتریس X ، ماتریس content-tag است و $X_{c,t}$، id هر محتوا و علائم تعیین شده است. $X_{c,t}$مقدار Boolean دارد و نشان میدهد تا چه حدی یک محتوا c با علامت t در ارتباط است.

$= \sum_{t=1}^{|T|}(t_u)y_{u,t}$

فرض کنید y ماتریس user- tag باشد. پس $y_{u,t}$ ، id هر کاربر و علائم استفاده شده را نشان میدهد.در اینجا برای پیدا کردن شباهت بین توئیت ها و شباهت بین کاربران از الگوریتم جستجوی مشترک استفاده شده است که با محاسبه مشابهت های کسینوسی زوجی مانند معادله زیر است:

$Simxi,xj := \frac{xi.xj}{||xi||2.||xj||2}$

اینجا $i, j \in C$ هستند و x_i and x_j بردارهایی از علائم حاشیه ای شان هستند

$Simxi,xj := \frac{xi.xj}{||xi||2.||xj||2}$ => $cos\ \theta = \frac{d1.d2}{||d1||2.||d2||2}$

$$Sim_{t_i,t_j} := \frac{t_i.t_j}{||t_i||^2.||t_j||^2}$$

در اینجا i و j دو تا تگ هستند و x_i and x_j بردارهایی از کاربران که از آن تگ ها استفاده میکنند.

از آنجا که اکثر کاربران در شبکه های اجتماعی سعی میکنند توئیت های خود را با داشتن نشان های بیشتر که در توئیت هایشان تعبیه میشوند ترویج کنند از الگوریتم CPSO (دسته بندی ذرات بهینه) و از الگوریتم معروف مبتنی بر هوش محاسباتی (الگوریتم بهینه سازی گروه ذرات) PSO استفاده شده است.

۴- تکنیک های کشف جعل در تصویر

عمدتا دو نوع الگوریتم کشف جعل عکس وجود دارند که فعال و منفعل هستند. در الگوریتم های فعال نیاز است عکس اصلی در دسترس باشد و از روی آن تصاویر جعلی کشف می شوند مانند امضای دیجیتال و واترمارک در الگوریتم منفعل تشخیص جعل تصویر شامل تشخیص جعل و موارد کپی شده بدون در نظر گرفتن محتوای تصویر اولیه است؛ در این تکنیک ما برای کشف جعل نیاز به تصویر اصلی نداریم و بهمین علت این روش در فضای مجازی بیشتر کاربرد دارد. Li و همکاران روشی برای شناسایی جعل نسخه برداری در حالت منفعل را پیشنهاد دادند[43] آنها با استفاده از اپراتور تلفن همراه بافت تصویر را نشان دادند و تصویر اصلی را به تصویر سیاه و سفید تغییر دادند، اما با وجود آلاینده های صوتی و فشرده سازیJPEG اجزائ با فرکانس بالا برای چنین تصاویری نمی توانند ثابت بمانند به همین دلیل از فیلتر عبور پایین استفاده کردند و دریافتند که فیلتر کردن اگر بیش از دوبار انجام شود سرعت عملکرد بالا می رود و سپس تصویر را به M Shamim Hossaina وهمکارانش برای کشف تصاویر آزمایشات خود را بر مبنای تبدیل هرمی کشش انجام دادند زیرا تبدیل هرمی کششی میتواند یک تصویر را از جهات مختلف تجزیه کند و با استفاده از ماتریس (هم رخداد) co-occurrence خصوصیات آن کشف شود. ماتریس هم رخداد ماتریسی است که دربردارنده اطلاعاتی در رابطه همرخدادی تصاویر با ارتباط بین مقادیر پیکسل های مجاور در یک تصویر میباشد. اولین بار دانشمندی به نام هارلیک [10] از ماتریس هم رخداد استفاده کرد.

اندازه ماتریس هم رخداد وابسته به شدت رنگ پیکسل های موجود در تصویر است. اگر فاصله دو پیکسل d=1 باشد زاویه های بین دو پیکسل می تواند 135, 90, 45, 0 درجه بیان شوند. بعد از تولید ماتریس هم رخداد ویژگی‌های آماری همچون کنتراست، میانگین، واریانس، عدم شباهت، گشتاور زاویه‌ای دوم، انتروپی، همگن بودن و وابستگی را می‌توان استخراج کرد.دلیل استفاده از این روش کم هزینه بودن محاسبات آن است.

بسیاری از بلوک های هم پوشان تقسیم میکنند و برای پیدا کردن بلوک هایی که مطابقت دارند از فاصله اقلیدسی استفاده کردند.

در مرحله اول عکس به یک قطعه رنگ اصلی (cr) تبدیل میشود زیرا این تبدیل، بی نظمی هایی را برجسته می سازد که در قطعات قرمز، سبز، آبی یا درخشان قابل مشاهده نیستند.

بعد وارد کردن قطعه cr بر SPT می باشد.SPT یک تکنیک تجزیه عکس چند روزولیشن می باشد که در آن یک عکس به چند اندازه و جهت تجزیه می شود.

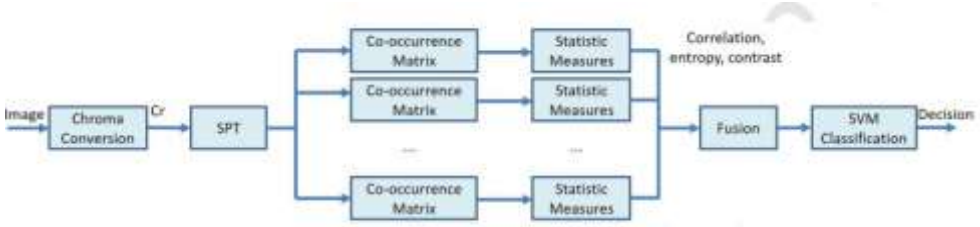


Figure 3: Process flow of the image forgery detection method.

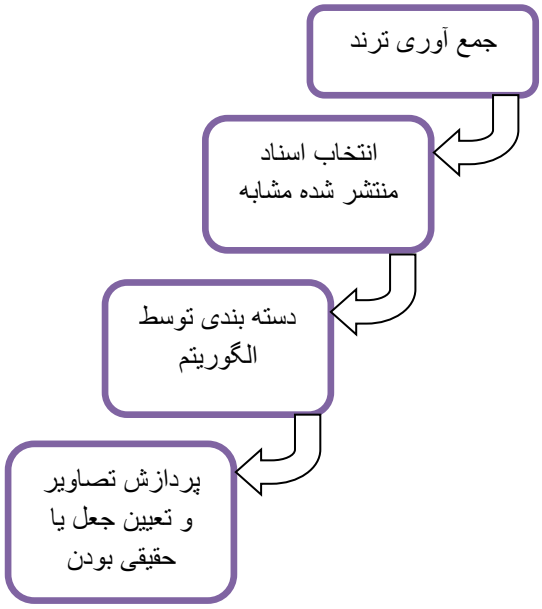
۵- بررسی جعلی یا حقیقی بودن تصاویر

چندین روش برای کشف تصاویر جعلی وجود دارند؛ روشهای مربوطه به دو دسته تقسیم می شوند:

- کشف بدون نام و موضع یابی
- کشف با موضع یابی

در روش کشف بدون نام ، خروجی ، پس از بررسی جعلی بودن یا حقیقی بودن ، دودویی است.اما روش موضع یابی ، فقط برای کشف این نیست که آیا تصویر حقیقی یا جعلی است بلکه در صورت جعلی بودن ، لوکیشن جعل تصویر را هم نشان می دهد.

۶- مقایسه الگوریتم ها



۷- نتیجه گیری

در این تحقیق ما می خواستیم مطمئن شویم یک تصویر واقعی است و توصیف ادعا شده را به حقیقت بیان می کند.
ابتدا یک دیتابیس از تصاویر پروفایل ها و علائمی که برای توصیف تصویر استفاده شده را باید جمع آوری کرد؛ با استفاده از الگوریتم های بررسی شده در بالا می توان تصاویر منتشر شده روند را کشف کرد.

۸- منابع

9- [8] Jiawei Han, Micheline Kamber. Data Mining: Concepts and Techniques, 3rd edition, 2012.
10-
11- [9] www.developers.facebook.com.
12-
13- [۲۷].Oliveira, M., Bastos-Filho, C.J.A. and Menezes, Using network science to assess particle swarm optimizers," Soc. Netw Anal. Min. (2015) 5:3
14- [24] L.Z. Xing, H.L. Le, and Z. Hui, A Novel Social Network Structural Balance Based on the Particle Swarm Optimization
15- Algorithm," Cybernetics and Information Technologies, vol. 15, no. 2,pp23-352015
16- 4. Balakrishnan, S., Chaudhuri, S. and Narasayya, V., (2015). AutoTagn search my photos: leveraging the social graph
17- for photo tagging. In: Proceedings of the 24th international conference on world wide web companion, 163–166
18- 8. Doerfel S, Zoller D, Singer P, Niebler T, Hotho A, Strohmaier M (2016) What users actually do in a social
19- tagging system: a study of user behavior in BibSonomy. ACM Transactions on the Web (TWEB) 10(2):14

20- 10. Ha E, Kim Y, Hwang E (2016) A categorization scheme of tag-based folksonomy images for efficient image
21- retrieval. KIISE Transactions on Computing Practices 22(6):290–295
22- [32] Nor Bakiah Abd Warif, Ainuddin Wahid Abdul Wahab, Mohd
23- Yamani Idna Idris, Roziana Ramli, Rosli Salleh, Shahaboddin
24- Shamshirband, and Kim-Kwang Raymond Choob, Copy-move
25- forgery detection: Survey, challenges and future directions, Journal of Network and Computer Applications, Volume 75, Page 259–278November 2016