فصل ۱۵

تکنیک‌های کشف اخبار جعلی در رسانه‌های اجتماعی

اکراتی ساکسِنا، پراتیشدا ساکسنا، و هَریتا رِددی[۱]

چکیده؛ ظهور وب‌سایت‌های متعددِ رسانه‌های اجتماعی در قرن بیست‌ویک، عرضه‌گاهی را فراهم آورده‌است تا مردم سراسر جهان به‌سادگی از طریق دستگاه‌هایی به‌طورگسترده دردسترس، مانند: گوشی‌های هوشمند (اِسمارت‌فون‌ها) در آن فعالیت کنند. رسانه‌های خبری سنتی کارشناسانی را در هر حوزه دارند که می‌توانند محتوای ارائه‌شده در اخبار را راستی‌آزمایی کنند. این امر مردم از اقشار مختلف جامعه را قادر ساخته‌است تا به ارسال محتوا درباره‌ی موضوعات متنوع، از امور جاری گرفته تا تاریخی، بپردازند، البته ثابت کردن صحت این محتواها کار ساده‌ای نیست. هرچند، باتوجه‌به حجم زیاد پُست‌های رسانه‌های اجتماعی در هر روز، یک انسان عادی در مواجهه با محتوای این پست‌ها نمی‌تواند به‌سادگی اطلاعات نادرست را از اطلاعات حقیقی تمیز دهد. این امر موجب شده‌است تا پژوهشگران به کشف خودکار اخبار جعلی علاقه‌مند شوند. در این فصل، ما به بررسی شاخصه‌های مورد استفاده در شناسایی اخبار جعلی و اقسام مختلف تکنیک‌های کشف اخبار جعلی خواهیم

[۱]A. Saxena (✉) Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands
e-mail: a.saxena@tue.nl
P. Saxena G. L. Bajaj Institute of Technology and Management, Greater Noida, India
e-mail: pratistha.saxena@glbitm.ac.in
H. Reddy
Surat, Gujarat, India

Principles of Social Networking, Smart Innovation, Systems and Technologies ۲٤٦,
https://doi.org/۱۰,۱۰۰۷/۹۷۸-۹۸۱-۱٦-۳۳۹۸-۰_۱۵

پرداخت. هم‌چنین، خلاصه‌ای از مجموعه‌داده‌های دردسترس برای کشف اخبار جعلی و راهنمایی‌هایی را برای مطالعات آینده ارائه می‌کنیم.

۱۵.۱ مقدمه

رسانه‌های اجتماعی شیوه‌ی ارتباط کاربران را تغییر داده‌است؛ کاربران را قادر ساخته‌است محتوایی را ارسال کنند که تنها برای کاربران گزینش‌شده قابل مشاهده باشد یا اینکه کل دنیا بتوانند آن را ببینند. وب‌سایت‌های رسانه‌های اجتماعی، مانند: توئیتر، میلیون‌ها کاربر دارند که انواع محتواها را روزانه ارسال می‌کنند. درحقیقت، مِندوزا[۲] و همکارانش [۱] توانستند حدود ۴.۷ میلیون پُست بومی را درخصوص زلزله‌ی ۲۰۱۰ شیلی براساس هشتگ‌های توییتر فهرست‌بندی کنند.

سهولت استفاده از رسانه‌های اجتماعی می‌تواند جدانشدنی انتشار اطلاعات در موقعیت‌هایی مانند واکنش به فاجعه باشد، سؤالی که در این‌جا مطرح می‌شود این است که محتوای ارسال‌شده به‌صورت آنلاین تاچه‌حد مورد تأیید و حقیقتاً درست است. اطلاعات غلط می‌تواند خروجی‌های منفی شدیدی، مانند: ایجاد وحشت، را در پی داشته‌باشد.

اخبار جعلی معمولاً چنین تعریف می‌شوند: محتوای غیرواقعی که با قصد گمراه کردن مردم یا قانع کردن آن‌ها به باور آن محتوا ایجاد شده‌است [۲]. شو[۳] و همکارانش [۳] دو شاخصه‌ی اخبار جعلی را مشخص می‌کنند: ۱. اخباری حاوی محتواهایی که می‌توان غلط بودنشان را تأیید کرد و ۲. اخباری که عمداً برای گمراه کردن مردم ساخته شده‌اند. به‌طورکلی، اطلاعات غلط می‌تواند به دو دسته تقسیم شود: ۱. مبتنی بر نظر: در این دسته هیچ حقیقت منفرد مبنایی‌ای وجود ندارد که بگوید فلان شخص نظرات غلطی را ارائه داده‌است، و این دسته عمدتاً در مورد نشریه‌های [مروری] آنلاین رخ می‌دهد، و ۲. مبتنی بر واقعیت: در این دسته حقیقتی مبنایی وجود دارد که اطلاعات آن را تکذیب می‌کند و این اخبار جعلی، شایعات و اطلاعات نادرست را پوشش می‌دهد [۴].

در قرن نوزدهم، دسترسی به اخبار چاپ‌شده با قیمت ارزان مسیری را برای انتشار اخبار جانبدارانه گشود. هرچند، امروزه با ظهور رسانه‌های اجتماعی، می‌توان به‌سادگی محتوا را بدون قضاوت سردبیر، تأیید واقعیت، یا بررسی توسط شخصی ثالث به اشتراک گذاشت [۵]. این امر علاوه بر آن‌که موجب می‌شود اخبار بیشتر از رسانه‌های اجتماعی مصرف شود [دریافت شود]، احتمال قرار گرفتن مردم در معرض اطلاعات غلط را نیز افزایش می‌دهد. مطالعه‌ای در حدود ۱۰ سال پیش روی کاربران کانادایی شبکه‌ی اجتماعی انجام شد، این مطالعه نشان داد؛ دو پنجم کاربران اخبار را از کسانی مصرف [دریافت] می‌کنند که در پلتفورم‌هایی چون فیس‌بوک دنبالشان می‌کنند تا [از طریق این افراد] در جریان انواع اخبار باشند [۶].

[۲] Mendoza
[۳] Shu

مطالعه‌ی دیگری[۴] در سال ۲۰۱۷ نشان داد؛ حدود دو سوم بزرگسالان ایالات‌متحده اخبار موجود در پلتفورم‌های رسانه‌های اجتماعی را دنبال می‌کنند.

با آنلاین شدن منابع خبری، درآمد حاصل از بازدیدها و کلیک‌های خوانندگان روی مقالات خبری مشوقی برای به‌اشتراک-گذاری سریع اخبار می‌شود، که احتمالاً این امر موجب کنار گذاشتن بررسی دقیق صحت اطلاعات می‌شود [۷]. تیترهای خبری که «تله‌ی کلیک» هستند، تیترهایی چشم‌گیرند وبا این نیت برای مقالات خبری ساخته‌شده‌اند که کاربران را اغوا کنند تا روی لینک یا بر روی گزینه «بیشتر بخوانید» کلیک کنند [۳]. بعضی از انواع تیترهای خبریِ «تله‌ی کلیک» درباره‌ی محتوای مقاله‌های خبری اغراق می‌کنند یا حتی مربوط به مقالاتی هستند که در واقعیت صحت ندارد. انگیزه‌ی دیگر برای تولید اخبار جعلی ترویج یک ایدئولوژی، عمدتاً طی انتخابات و با استفاده از انتشار اخبار به‌نفع کاندیدا یا حزبی مشخص است. ربات‌ها (Bots) یا حساب‌های خودکارشده‌ی رسانه‌های اجتماعی می‌توانند سرعت انتشار اخبار جعلی را افزایش دهند. در طول انتخابات ریاست‌جمهوری ۲۰۱۶ ایالات‌متحده و انتخابات ۲۰۱۷ فرانسه از ربات‌ها برای به‌اشتراک-گذاری حجم قابل‌توجه محتواهای مربوط به سیاست استفاده شد [۸ و ۹].

در این فصل، ما به بحث درباره‌ی تکنیک‌های کشف اخبار جعلی می‌پردازیم که بر اساس سه بُعد، دسته‌بندی می‌شوند. بُعد اول استفاده از اطلاعات موجود در خبر و منبع خبر برای تشخیص جعلی بودن یا نبودن خبر است. روش‌های متعددی وجود دارد، شامل: اثبات کردن اینکه از کدام منابع خبر یا کدام کاربران در رسانه‌های اجتماعی بیشتر انتظار می‌رود که اخبار جعلی یا جهت‌دار را ارسال کنند [۱۰ و ۱۱]؛ یافتن حساب‌های جعلی در شبکه‌های اجتماعی [۱۲]؛ و تحلیل کردن تفاوت‌های زبان‌شناختی [۱۳ و ۱۴] و چندرسانه‌ای [۱۵ و ۱۶] میان محتوای اخبار جعلی و محتوای اخبار حقیقی. دومین بُعد پی بردن به واکنشی‌است که به خبرها نشان داده می‌شود، شامل: تردیدهایی که خوانندگان اظهار می‌کنند و و اینکه اخبار جعلی تا چقدر متفاوت از اخبار موثق پخش می‌شوند. بُعد سوم آن است که بااستفاده از منابع دانش ادعاها را بررسی و صحت‌وسقم آن‌ها را معلوم کرد. کشف اخبار جعلی به‌جهات مختلف می‌تواند امری چالش‌برانگیز باشد. پُست‌های رسانه‌های اجتماعی یا وب‌سایت‌های خبری ممکن است همراه با تصاویر و محتواهای چندرسانه‌ای باشند و باید توجه داشت که دستکاری و فوتوشاپ تصاویر کار ساده‌ای است [۲ و ۷]. در رسانه‌های اجتماعی، این امکان وجود دارد که با جعل هویت کاربران انسانی [حقیقی] حساب‌های جعلی ساخت.

ساختار فصل بدین شرح است: در بخش ۱۵.۲ ما به شرح تکنیک‌های شناسایی شاخصه‌ها و کشف اخبار جعلی می‌پردازیم، ازجمله: مجموعه داده‌ها و مدل‌های دسته‌بندی. در ادامه‌ی این بخش نتیجه‌گیری و راهنمایی برای پژوهش‌های آتی در این حوزه آمده‌است. در بخش ۱۵.۴ ما به خوانندگان علاقه‌مند به پژوهش در این حوزه مطالب بیشتر را پیشنهاد می‌دهیم.

[۴] https://www.reuters.com/article/us-usa-internet-socialmedia/two-thirds-of-american-adultsget-news-from-social-media-survey-idUSKCN۱BJ۲A۸.

۱۵.۲ کشف اخبار جعلی

روبین[۵] و همکارانش [۱۷] بر اساس سه نوع اخبار جعلی، سه دسته‌بندی از کشف اخبار جعلی را پیشنهاد داده‌اند: ۱. گزارش‌های ناصادقانه‌ای که روزنامه‌نگاران ازخود می‌سازند و در نشریات زرد و روزنامه‌نگاری تَبلوید[۶] دیده می‌شود؛ ۲. حقه‌های اینترنت؛ و ۳. اخبار جعلی ساخته‌شده با هدف طنز و هجو که اغلب در قالب اخبار پَرودی[۷] است. از آن‌جاکه ساخت طنز متفاوت از ساختار جدی است و خوانندگان باید قصد هجو را تشخیص دهند، لذا تشخیص آنلاین و در اسرع وقت سایر اخبار گمراه‌کننده ضروری است تا بتوان مانع از گمراه‌شدن مردم به باور این اخبار به‌عنوان حقیقت، شد. یک نمونه از نظارت خودکارشده‌ی آنلاین بر اطلاعات غلط «هوآکسی (Hoaxy)» است که در میان داده‌های وب‌سایت‌های راستی‌آزمایی می‌خزد و توییتر را برای استخراج مصداق‌های به‌اشتراک‌گذاری این اخبار دنبال می‌کند [۱۸]. در بخش بعدی؛ ما به بحث درباره‌ی دسته‌بندی‌های شاخصه‌های به‌کارگرفته‌شده در تشخیص اخبار غلط از اخبار موثق می‌پردازیم.

۱۵.۲.۱ شاخصه‌های کشف اخبار جعلی

یکی از مهم‌ترین شاخصه‌ها برای تشخیص اعتبار اخبار محتوای آمده در متن اصلی خبر و واکنش‌های کاربران رسانه‌ی اجتماعی به آن است. ویژگی‌های منبع خبر، شامل: وب‌سایت منتشرکننده‌ی خبر یا کاربر رسانه‌ی اجتماعی‌ای که آن خبر را پُست کرده‌است، به‌علاوه‌ی ویژگی‌های کاربرانی که آن خبر را پخش می‌کنند از دیگر شاخصه‌های مفید برای کشف اخبار جعلی است.

۱۵.۲.۱.۱. شاخصه‌های مبتنی بر محتوا

یک مقاله‌ی خبری دارای یک تیتر است که حتی باوجود کوتاه بودن ممکن است حاوی سرنخ‌هایی باشد که نشان دهند محتوای خبر گمراه‌کننده است. تیتر ممکن است به‌صورت جذب‌کننده یا برانگیزنده‌ی احساسات و در قالب یک تله‌ی کلیک نوشته شود تا کاربران را به آن وب‌سایت جذب کند [۱۹]. بدنه‌ی اصلی مقاله‌ی خبری شامل یک شرح مفصل از خبر یا ادعای خبر است و متن طولانی‌تری از تیتر را به‌جهت تجزیه‌وتحلیل زبان‌شناختی ارائه می‌دهد. یک مطلب خبری هم‌چنین می‌تواند حاوی محتواهای چندرسانه‌ای مانند تصاویر باشد، و کاربران رسانه‌ی اجتماعی می‌توانند به پُست‌های میکرو بلاگینگ (بلاگ‌نویسی کوچک) خود محتواهای چندرسانه‌ای الصاق کنند. مشاهده شده‌است که در خصوص اطلاعات غلط، کاربران معمولاً از تصاویری استفاده می‌کنند که دقیقاً مربوط به آن رویداد نیست [۲۰]. مقاله‌های خبری از طریق لینک‌های خارجی در رسانه‌ی اجتماعی به‌اشتراک گذاشته می‌شوند یا ممکن است محتوای خبر مستقیماً در یک میکروبلاگ به‌صورت

---

[۵] Rubin

۶ سبک پرطرفدار روزنامه‌نگاری است که در بین عموم به نشریات زرد و شایعه‌پراکن شهرت دارد [مترجم].

۷ پَرودی یا نقیضه به تقلید طنزگونه از یک اثر هنری دیگر می‌گویند که به قصد استهزا یا نقد آن اثر صورت می‌گیرد [مترجم].

اجمالی ارسال [پُست] شود. به‌اشتراک‌گذاری اطلاعات در رسانه‌های اجتماعی بدان‌جهت‌که کاربر می‌تواند اطلاعات را در قالب پُست‌های کوتاه به‌اشتراک بگذارد و ممکن است لینک ارجاع به مقاله‌ی چاپ‌شده را نیاورد، چالش‌برانگیزتر می‌شود. البته، ویژگی‌های محتوای میکروبلاگ می‌تواند با تجزیه‌وتحلیل نظرات [کامنت‌ها] و پاسخ‌های سایر کاربران تکمیل شود، بدین صورت که نظرات موافق یا ضد ارائه‌شده نسبت به خبر به اشتراک‌گذاشته‌شده را استخراج کرد.

شاخصه‌های زبان‌شناسی عبارتند از: ویژگی‌های متنی مبتی بر کاراکترها [حروف و اشکال]، لغت‌ها، جمله‌ها و درمجموع مطلب به‌نگارش درآمده. شاخصه‌های زبانی متداول در پردازش زبان‌های طبیعی (NLP) که عبارتند از: ویژگی‌های لغوی و ویژگی‌های نحوی [۳]. برای ویژگی‌های زبانی عمومی می‌توان؛ فراوانی واژه‌های پالایشی[۸]، تعداد علامت‌های تعجب و سؤال، برچسب‌گذاری جزء کلام[۹]، خوانایی متن، استفاده از کلمات دستوری، و غیره را مثال زد. شاخصه‌های منحصری برای پلتفورم هر شبکه‌ی اجتماعی وجود دارد، ازجمله: شمار یا تناسب مقدار لینک‌های خارجی (URLها) و هشتگ‌ها (عنوان موضوعات پرطرفدار [تِرندشده] که بعد از پیشوند «#» نوشته می‌شوند). سایر ویژگی‌های متنی عبارتند از: قطبیت [تمایل متن] که همان احساس مثبت یا منفی موجود در متن است؛ عینیت؛ و مخالفت. تفاوت‌های زبانی مشهودی در سبک نگارش محتوای فریبنده [۲۱] وجود دارد، ضمن‌آنکه اخبار جعلی نیز می‌توانند فاقد عینیت باشند [۲۲]. برای تأیید میزان اعتماد در یک مطلب خبری از نشانگرهای گفتمان[۱۰] استفاده می‌شود [۲۳]. در جدول ۱۵.۱؛ ما بعضی از شاخصه‌های زبانی مورداستفاده در پژوهش طبقه‌بندی اسناد را بیان کرده‌ایم.

۱۵.۲.۱.۲ شاخصه‌های مبتنی بر کاربر

شاخصه‌های مبتنی‌بر-کاربر با سطح-حساب کاربر عبارتند از: ویژگی‌های کاربری که یک پُست مشخص را می‌سازد و ویژگی‌های کاربرانی که با گذاشتن لینک، به‌اشتراک‌گذاری، گذاشتن نظر [کامنت] یا پاسخ به آن پُست مشخص واکنش بیشتری را به آن نشان می‌دهند. برای نمونه، این ویژگی‌ها عبارتند از: مدت زمانی که کاربر در آن پلتفورم ثبت شده‌است، تعداد میکروبلاگ‌های پُست‌شده توسط کاربر، شمار دنبال‌کنندگان [فالوورها] و دوستان، آیا کاربر تأییدشده‌است، آیا کاربر بیو دارد، و غیره [۲۸]. وثوقی[۱۱] و همکارانش [۳۰] شش ویژگی کاربر را برای کاربرانی که در شایعه‌پراکنی دخالت دارند برمی‌شمرند: ۱. کاربر چقدر جنجالی است، ۲. اصل بودن توییت‌های کاربر، ۳. حساب کاربر تأییدشده‌است، ۴. مشارکت

---

[۸] واژه پالایشی یا واژه تصفیه‌شونده یا واژه ایستاده‌شده کلماتی هستند که قبل یا بعد از پردازش داده‌های زبان طبیعی پالایش (تصفیه) می‌شوند. معمولا واژه‌های پالایشی به رایج‌ترین کلمات در یک زبان اشاره دارد، اما هیچ فهرست جامعی از این واژه‌ها، که در تمام ابزارهای پردازش زبان طبیعی استفاده شوند، موجود نیست. در واقع ابزارهای موجود هم از چنین فهرستی جامع و یکسانی استفاده نمی‌کنند (ویکی پدیا). [مترجم].

[۹] در زبان‌شناسی پیکره‌ای، برچسب‌گذاری جزء کلام برچسب‌گذاری دستوری یا ابهام‌زدایی رده واژه، فرایند برچسب‌گذاری یک واژه در یک متن است، که آن برچسب متناظر با رده جزء کلامی خاص آن واژه می‌باشد. (ویکی‌پدیا) [مترجم].

[۱۰] قش‌نمای کلامی یا نشانگر گفتمان یا عبارت اشاره‌ای اصطلاحی در زبان‌شناسی است که به کلمات و گروه‌هایی اشاره دارد که به تنهایی بدون معنی و فاقد نقش نحوی هستند و نقش‌شان در جمله معمولاً به صورت رابطی میان قطعات کلامی و نشان دهنده‌ی وجود یک رابطه‌ی کلامی است. (ویکی‌پدیا) [مترجم].

[۱۱] Vosoughi

کاربر، ۵. نقش کاربر، و ۶. تأثیرگذاری کاربر. ویژگی‌های سطح-کاربر هم‌چنین برای شناسایی گروه‌های کسانی که اطلاعات غلط در پلتفورم پخش می‌کنند مورد استفاده قرا می‌گیرد. ویژگی‌های کاربران سطح- گروه شامل ویژگی‌های سطح- فردی گردآمده در یک گروه از کاربران می‌شود و پیش‌فرض آن است که می‌توان با تشخیص بعضی از ویژگی‌های منحصر، گروه‌های کاربرانی که اخبار جعلی را پخش می‌کنند را شناسایی کرد [۳].

۱۵.۲.۱.۳ شاخصه‌های مبتنی‌بر شبکه

شاخصه‌های مبتنی‌بر شبکه یا گراف عبارتند از: اتصالاتِ شبکه‌ی کاربرانی که در اخبار شبکه‌های اجتماعی دخالت دارند، و هم‌چنین شبکه‌ی انتشار که الگوی پخش اخبار در شبکه‌های اجتماعی را نشان می‌دهد [۳۱].

جدول ۱۵.۱ شاخصه‌های مبتنی‌بر زبان‌شناسی

مجموعه ویژگی	هدف	ویژگی‌ها
ژو[۱۲] و همکاران [۲۴]	کشف فریب در تعامل به‌واسطه‌ی رایانه	**کمیّت:** کلمه، جمله، معرف، فعل، عبارت اسمی **پیچیدگی:** شمار متوسط عبارت[۱۳] و نقطه‌گذاری؛ کلمه، جمله و عبارت با طول متوسط **غیرمستقیم بودن:** به‌کاربردن حالت مجهول، افعال معین شرطی، عینیت‌بخشی، اصطلاحات کلی، عدم قطعیت، ارجاع‌به خود، ارجاع‌به گروه، و ارجاع به غیر. **حالت بیان:** احساسی [۲۵] **تنوع:** تنوع در واژگان محتوا و دایره‌ی لغات، حشو. **خاص بودن:** محتوای ادراکی و زمانی- مکانی، اثر مثبت و منفی **غیررسمی بودن:** نسبت خطای املایی
برِنان و گرین-استادت[۱۴] [۲۶]	انتساب نویسندگی	شمار واژگان منحصر، شاخص خوانایی گانینگ فاکس[۱۵]، یک شاخص[آزمون] خوانایی جایگزین دیگر، تعداد کاراکترها بدون احتساب فضاهای خالی، تراکم واژگانی، شمار جمله، میانگین تعداد هجاها در هر کلمه، میانگین طول جمله
مجموعه ویژگی رایت‌پرینتش (WritePrints) [۲۷]	نویسنده‌ی منسوب به پیام‌های آنلاین	**خصوصیت‌های واژگانی:** خصوصیات در سطح کاراکتر، مانند: تعداد کاراکترها، کاراکترهای درج‌شده به‌صورت بزرگ، رقم‌ها، فاصله‌ی ایجادشده توسط کلید تَب (Tab spaces)، فاصله‌های خالی، و فراوانی الفبایی حروف/ کاراکترهای خاص. خصوصیات در سطح واژه عبارتند از: تعداد واژه‌ها، تعداد واژه‌های کوتاه، طول جمله و میانگین واژه، واژه‌ای که تنها یک‌بار در متن آمده‌است (hapax Legomena) و کلمه‌ای که تنها

[۱۲] Zhou

[۱۳] مجموعه‌ای از چند واژه که دارای فاعل و فعل باشد. (آبادیس)

[۱۴] Brennan and Greenstadt

[۱۵] Gunning Fox readability index: نوعی آزمون در تشخیص میزان خوانایی متن نگارش‌شده به زبان انگلیسی

مجموعه ویژگی	هدف	ویژگی‌ها
		دوبار در متن آمده‌است (dis legomena)، مقیاس‌های غنای واژگان، توزیع طول واژه‌ها. **خصوصیت‌های معنایی:** خصوصیات در سطح جمله، شامل: فراوانی نقطه‌گذاری‌ها، فراوانی واژه‌های دستوری و غیره. **خصوصیت‌های ساختاری:** شمار سطرها، شمار جمله‌ها، شمار پاراگراف‌ها، نسبت جمله‌ها/ کاراکترها/ واژه‌ها در هر پارگراف، درودها، جداکننده‌ها، محتوای نقل‌شده و موقعیت قرارگیری آن در متن، تورفتگی، امضا **خصوصیت‌های خاصِ محتوا:** فراوانی واژگان خاصِ دامنه‌ی انتخاب‌شده
کاستیلو[۱۶] و همکاران [۲۸]	اعتبارسنجی اطلاعات رسانه‌های اجتماعی	شکلک‌ها (لبخند، اخم) گنجانده‌شده، تعداد واژه‌های مثبت، تعداد واژه‌های منفی، امتیاز تعلق‌گرفته به عواطف، طول واژه-ها/ کاراکترها، ضمیرهای اول شخص و سوم شخص، علامت‌های سؤال و تعجب، هشتگ‌ها (#)، URLها، و مِنشن‌ها (@)
هورن[۱۷] و همکاران [۲۹]	کشف اخبار جعلی	**پیچیدگی:** شاخص‌های خوانایی گانینگ فاگ (Gunning Fog)، فِلش- کینکِید (Flesch-Kincaid)، اِس‌اِم‌اوجی (SMOG)، عمق معنا، درخت‌های عبارت فعلی و اسمی، میانگین طول کلمه، تنوع واژگان **روان‌شناسی**: قدرت عواطف؛ تعداد واژگان تحلیلی، گویای بصیرت و سببی؛ تعداد واژگان گویای اختلاف‌نظر، قطعیت، آزمون‌وخطا، تمایز، وابستگی، قدرت، پاداش، ریسک، احساس و دغدغه‌ی شخصی. **سبک:** تعداد واژه‌ها، اسم‌ها، ضمیرهای ملکی، تعداد واژه‌های در زمان گذشته و در زمان آینده، تعداد ادوات استفهام، و غیره.

۱۵.۲.۲ دسته‌بندی‌های کشف اخبار جعلی

تکنیک‌های کشف اخبار جعلی را می‌توان از لحاظ به‌کارگیری شاخصه‌های فوق‌الذکر به دسته‌ی کلی تقسیم کرد.

۱. **ویژگی‌های محتوا و منبع خبر:**

a. زبان‌شناسی عمومی: ویژگی‌های زبان‌شناسیِ اخبار یا ادعاهای جعلی متفاوت از ویژگی‌های زبان‌شناسی اخبار حقیقی هستند[۱۳، ۱۴، ۲۹، ۳۲-۳۴].

b. فریب: اخبار جعلی احتمالاً حاوی زبان فریب‌آمیزی است که نویسنده با آن زبان سعی دارد سبک نگارش را پنهان نگه دارد، بی‌طرفی را نشان دهد، یا زحمت مازادی برای تأکید بر واقعی بودن خبر بکشد [۲۱، ۲۴، ۳۵].

c. عینیت [بی‌طرفی]: اخبار جعلی ممکن است فاقد عینیت [بی‌طرفی] باشد و یک‌جانبه با گرایش به یک ایدئولوژی نوشته شود [۲۲].

[۱۶] Castillo
[۱۷] Horne

d. تصاویر و سایر محتواهای چندرسانه‌ای: اخبار جعلی ممکن است حاوی تصاویر نامربوط، باکیفیت پایین، تکراری یا دستکاری‌شده باشند [۱۶، ۲۰، ۳۶، ۳۷].
e. اعتبار منبع: ممکن است بعضی از منابع در گذشته محتوای غلط تولید کرده باشند، هم‌چنین شاید بعضی از حساب‌های رسانه‌های اجتماعی درحقیقت ربات (bots) باشند [۱۰، ۳۸].

۲. **بستر اجتماعی:**

a. تردید و مخالفت: احتمالاً اخبار جعلی نظرات [کامنت‌های] سایر کاربران رسانه‌های اجتماعی که تردید، ناباوری یا حتی مخالفت دارند را به خود جلب می‌کند [۳۹- ۴۱].
b. الگوهای پخش: اخبار جعلی ممکن است سریه‌تر، بیشتر و عمقی‌تر پخش شوند [۴۲].
c. کاربرانی که اخبار جعلی را پخش می‌کنند احتمالاً کاربرانی کم‌اعتبارتر و به‌شدت جنجالی‌تراند [۳۰، ۴۳، ۴۴].

۳. **راستی‌آزمایی:** یک ادعا را با استفاده از منابع دانش معتبری که حقیقت مسلم محسوب می‌شوند، تأیید می‌کند [۴۵، ۴۶].

۱۵.۲.۲.۱ ویژگی‌های محتوا و منبع خبر

اخبار در وب‌سایت‌های شبکه‌های اجتماعی در قالب میکروبلاگ‌ها، ارسال [پُست] می‌شوند و ممکن است در چنین محتوایی یک لینک خارجی داده‌شود که به یک وب‌سایت خبری ارجاع داشته‌باشد. یکی از شاخصه‌های اصلی برای تشخیص اعتبار اخبار رسانه‌ی اجتماعی؛ ویژگی‌های محتوای مقاله‌ی خبری‌ای است که در آن توییت یا خود توییت منبع به‌اشتراک گذاشته‌شده‌است. به‌علاوه، اعتبار منبعی که آن مقاله‌ی خبری را منتشر کرده یا آن خبر را در رسانه‌های اجتماعی ارسال [پُست] کرده‌است خود یک مؤلفه در تشخیص میزان اعتبار خبر است. در این بخش؛ ما به بحث کشف اخبار جعلی بر اساس محتوای خبر و اعتبار منبع خبر می‌پردازیم.

ویژگی‌های زبان‌شناسی عمومی: ویژگی‌های زبان‌شناسی رایج در محتوای خبر قرین با ویژگی‌های متنی‌ای است که منحصر به هر پلتفورم است، برای مثال: در توییتر؛ مِنشِن‌ها[۱۸] یا هشتگ‌ها در پی بردن به میزان اعتبار توییت‌ها مفید هستند [۳۴، ۴۷]. قزوینیان[۱۹] و همکارانش [۳۲] درخصوص الگوهای واژگانی و ویژگی‌های رده‌ی جزء کلام محتواهای شایعات شناسایی‌شده، به این دو مورد پی بردند: استفاده زیاد از این الگوها و ویژگی‌ها، و دقت بالای آن‌ها. برعکس، درخصوص ویژگی‌های انحصاری پلتفورم، مانند: هشتگ‌ها یا URLهای توییتر، دقت بالا، ولی میزان استفاده از آن‌ها معدود

[۱۸] کامنت گذاشتن به توییت ها در توییتر را منشن میگویند در واقع پاسخ دادن و یا واکنش نشان دادن به توییت ها را منشن می‌گویند. [مترجم]
[۱۹] Qazvinian

است، احتمالاً علتش آن است که حجم بسیار از توییت‌ها فاقد هشتگ‌ها یا URLها هستند. گوپتا[۲۰] و همکارانش به بررسی ویژگی‌های توییت‌ها، مانند: شمار ناسزاها، شمار واژگانی که بیان‌گر احساسات و عواطف مثبت/ منفی هستند؛ و بررسی ویژگی‌های کاربری که توییت را ارسال [پُست] کرده‌است، مانند: سن کاربر توییتر، تعداد دنبال‌کنندگان [فالوورها]/ دوستان او، پرداختند و با استفاده از الگوریتم‌های رتبه‌بندی ماشین بردار پشتیبانی (SVM) و مکانیزم بازخورد شبه- مرتبط[۲۱] توییت‌های مربوط به حوادث مهمی چون شورش‌های سال ۲۰۱۱ بریتانیا را بر اساس میزان اعتبارشان رتبه‌بندی کردند [۳۳ و ۴۸]. به‌عقیده‌ی آن‌ها؛ شمار کاراکترهای خاص شاخص مناسبی در تشخیص میزان اعتبار است، احتمالاً بدین‌سبب که توییت‌هایی که آگاهی‌دهنده و پیوسته هستند و حاوی هشتگ‌ها، URLها و مِنشن‌ها می‌باشند از تعداد کاراکترهای بیشتری برخوردارند. بویدیدو[۲۲] و همکارانش از هردوی ویژگی‌های توییت و ویژگی‌های کاربر استفاده کردند و اظهار داشتند که عملکرد مدل‌هایشان، حتی بدون به‌کارگیری ویژگی‌های منحصر-زبانی، برای چندین زبان عالی است. بااین‌حال، موفقیت اندک مدل‌هایشان درخصوص زبان فرانسه گویا آن بود که لازم است در جست‌وجوی ویژگی‌های منحصر- زبانی نیز بود. آن‌ها درخصوص هر دو مجموعه داده‌ی MediEval [قرون وسطایی] ۲۰۵ و ۲۰۱۶ به نمره‌ی F۱ بالای ۰.۹۳ دست یافتند [۳۴، ۳۶].

در موقعیت‌های اضطراری، مانند شورش‌ها با فجایع طبیعی، ارزیابی صحت‌وسقم اخبارِ رسانه‌های اجتماعی به امری حیاتی تبدیل می‌شود تا بتوان مانع از ایجاد وحشت به‌سبب اطلاعات غلط شود. شیا[۲۳] و همکارانش [۴۹] برای کشف خودکار سناریوهای این‌چنینی در مواقع اضطراری، از طبقه‌بندی شبکه بیزین [۵۰] استفاده کردند که برای تشخیص اعتبار توییت‌ها، ویژگی‌های مربوط به محتوا را هم‌سو با ویژگی‌های مربوط به نویسنده، مربوط به موضوع، مربوط به انتشار توییت‌ها را بررسی می‌کرد، مدل آن‌ها هم‌چنین قادر بود توییت‌های معتبر را بهتر از توییت‌های نامعتبر دسته‌بندی کند. ویژگی‌های زبان‌شناسی توییت‌ها نیز در تشخیص تصاویر جعلیِ پخش‌شده در توییتر در ارتباط با طوفان سندی سال ۲۰۱۲ مفید واقع شدند [۴۷]. نویسندگان ۹۷ درصد دقت پیش‌بینی در تشخیص توییت‌های حاوی URLهای تصاویر جعلی از تصاویر واقعی رسیدند و این کار را با استفاده از طبقه‌بندی‌های درخت تصمیم J۴۸ روی ویژگی‌های زبان‌شناسی توییت‌ها، شامل: طول متن؛ شمار علامت‌های سؤال؛ شمار علامت‌های تعجب؛ شمار واژگان؛ شمار کاراکترها [حروف] درج‌شده به‌صورت بزرگ، انواع ضمیرها؛ و هم‌چنین شمار هشتگ‌ها، URLها و مِنشِن‌ها، انجام دادند. گوپتا و همکارانش برای کشف بی‌درنگ اعتبار اخبار، سیستمی به‌نام توییت‌کِرِد (TweetCred) را ایجاد کردند که به اعتبار توییت‌ها براساس یک مقیاس- نقطه‌ای در زمان حقیقی نمره می‌داد و این کار را با استفاده از محتوای آن توییت و چند ویژگی دیگر انجام می‌داد [۵۱]. بعضی از این ویژگی‌ها عبارتند از: فراداده‌های (متادِیتاهای) توییت مانند: مختصات جغرافیایی و زمانی دروقت توییت؛ فراداده‌های

---

[۲۰] Gupta

[۲۱] بازخورد شبه-مرتبط یکی از روش های بهبود نتایج موتورهای جستجو است .با استخراج خودکار اطلاعات از یک نتیجه جستجوی قبلی، یک پرس و جو جدید به عنوان تعمیم درخواست اصلی مطرح می شود و سپس دوباره جستجو می شود.

[۲۲] Boididou

[۲۳] Xia

نویسنده مانند:تعداد دنبال‌کنندگان/ دوستان او و سن کاربر توییتر؛ اطلاعات شبکه‌ی توییتر مانند: تعداد ری‌توییت‌ها؛ و اطلاعات مربوط‌به لینک‌های داخل توییت مانند: امتیاز WOT[۲۴] برای URLهایی که لینکشان در توییت است. طبق بازخورد کاربران، چه بازخوردهای موافق و چه مخالف؛ اختلاف نمره‌ای که کاربران می‌دادند نسبت به نمره‌ای که در بازه‌ی ۷-۱ به آن‌ها داده‌شده بود در ۶۳ درصد موارد تنها ۱-۲ نمره بود.

تجزیه‌وتحلیل ویژگی‌های زبان‌شناسی- روان‌شناختی روی اخبار با استفاده از استعلام زبان‌شناسی و تعداد واژگان (LIWC) [۵۲] نشان داده‌است؛ درحالی‌که اخبار جعلی، شمار بسیاری واژگان ادراکی، اجتماعی و مثبت را دربردارند، اخبار موثق حاوی واژگانی‌اند که بیان‌کننده‌ی بصیرت و سایر فرایندهای شناختی هستند [۵۳]. نویسندگان با استفاده از یک مجموعه‌ی کامل از ویژگی LIWC و یک طبقه‌بندی خطی SVM توانستند به دقت ۷۰ درصدی در یک مجموعه‌داده‌ی ایجادشده از طریق جمع‌سپاری دست یابند. اوبرِین[۲۵] و همکارانش [۱۴] کارکرد واژگان قوی در اخبار جعلی برای جلب توجه مردم و مبالغه را بررسی کردند. راشکین[۲۶] و همکارانش [۵۴] نشان دادند؛ ۱. تمایل به استفاده از ضمیرهای اول شخص و دوم شخص در اخبار جعلی بیشتر است، که احتمالاً این امر نشان‌دهنده‌ی نویسندگی توأم با خیال‌پردازی است، هم‌چنین میزان استفاده از کلمات اغراق‌آمیز مانند صفت‌های عالی در اخبار جعلی بیشتر است. ۲. اخبار قابل اعتماد با ارائه ارقام ملموس و ابهام کمتر دارای مدارک بیشتری در پشتیبانی از خود هستند. تلاش‌های راشکین و همکارانش بیشتر معطوف به ایجاد تمایز میان فریب‌ها، طنز و تبلیغات [پروپاگاندا] در دسته‌ی اخبار غیرقابل اطمینان بود. براساس مشاهده‌ی آن‌ها؛ این گروه‌ها هرکدام خصوصیت‌های متفاوتی دارند، برای مثال: در اخبار تبلیغاتی به نسبت اخبار فریب‌آمیز صفت‌های عالی بیشتری به‌کار برده می‌شود.

هورن[۲۷] و همکارانش [۲۹] مشاهده کردند که میان طنز و اخبار جعلی ازلحاظ روان‌شناسی، سبک، و پیچیدگی شباهت-های بیشتری وجود دارد تا میان طنز و اخبار حقیقی. آن‌ها توانستند با دقت ۹۱ درصد طنز را از اخبار واقعی تمیز دهند، درحالی‌که دقت کارشان در تفکیک طنز از اخبار جعلی ۶۷ درصد بود. هم‌چنین تاکنون این‌طور معلوم شده‌است که اخبار جعلی حاوی محتوای حشوآمیزتر و نقل‌قول کمتر است، احتمالاً به‌این دلیل که محتوای حقیقی یا نقل‌قول‌هایی وجود ندارند که به‌عنوان مدرک از آن‌ها نوشت [۲۹، ۵۵]. اخبار جعلی هم‌چنین می‌توانند بر اساس تیترهای خبری‌شان متمایز از یکدیگر شوند. تیترهای اخبار جعلی ممکن است دربردانده‌ی مقدار زیادی محتوا که در یک تک جمله گنجانده شده‌است، استفاده‌ی وافر از حروف بزرگ، و از قلم افتادن واژگان پالایشی به‌جهت اشاره به هر تعداد نهاد که ممکن است، باشند [۲۹]. هورن و همکارانش [۲۹] از یک SVM هسته (کِرنِل) خطی روی مجموعه‌داده‌ی مربوط به اخبار سیاسی استفاده کردند و دقتشان در بهره‌گیری از محتوای متنِ عنوان خبر به ۷۸ درصد ارتقا یافت، درحالی‌که دقت آن‌ها در بهره‌گیری از متن

[۲۴] امتیاز میزان اعتبار در وب‌سایت تراست: https://www.mywot.com/

[۲۵] O'Brien
[۲۶] Rashkin
[۲۷] Horne

بدنه‌ی خبر ۷۱ درصد بود. این مسأله می‌تواند برای هدف گرفتن افرادی به‌کار رود که از تیتر اخبار فراتر نمی‌روند و به‌دنبال مدرکی مستدل، استدلال و منطق در محتوای واقعی مقاله نمی‌گردند.

در اسناد متنی؛ محتوا می‌تواند در قالب یک ماتریس تعبیه‌ی کلمه که با استفاده از تکنیک‌های واژه- برداری ساخته می‌شود نشان داده شود. یکی از ساده‌ترین مدل‌ها، مدل بسته‌ی کلمات (BOW) است. در این مدل از تعداد دفعات پدیدار شدن کلمه‌ها در یک سند متنی برای ترسیم یک بردار استفاده می‌شود. فراوانی اصطلاح- معکوس فراوانی متن (فراوانی وزنی تی‌اِف- آی‌دی‌اِف) روش دیگری برای تشخیص اهمیت یک نشانه باتوجه به سند درعوض تعداد دفعات پدیدار شدن آن نشانه است. سایر مدل‌های پیچیده عبارتند از: الگوریتم Word۲Vec (شامل: Skip-gram و بسته‌ی کلمات پیوسته) [۵۶]، و FastText [۵۷]. ایده‌ی پس الگوریتم Word۲Vec یادگیری نمایش کلمات، مانند: کلمات مشابه‌ای که در نمایش فضای برداری در مجاورت واقع می‌شوند. با استفاده از شبکه‌های عصبی ترسیم‌شده براساس رابطه‌ی کلمه‌ها با کلمه‌های هم‌جوارشان در سند، کلمه‌ها به‌صورت بردارها کدگذاری می‌شوند. در FastText، کلمه‌ها قبل از تبدیل شدن به بردارها طبق اِن-گرم‌ها تفکیک می‌شوند.

طی چند سال گذشته، رویکردهای یادگیری عمیق نیز برای کشف اخبار جعلی مورد استفاده واقع شده‌اند. این رویکردها عمدتاً همراه با نمایش‌های برداری کلمات متن خبر استفاده می‌شوند. می‌توان از شبکه‌های عصبی برای استخراج شاخصه-های قوی متن استفاده کرد، بی‌آنکه نیاز به شاخصه‌های دست‌ساز باشد [۵۸]. کریمی[۲۸] و همکارانش درخصوص مجموعه داده‌ی LIAR [دروغ‌گو] موفق شدند با استفاده از انجام طبقه‌بندی چندطبقه‌ای در دقت‌شان روی روش‌های خط مبنا ارتقا یابند. آن‌ها این کار را با ضمیمه‌ساختن منابع متعدد با استفاده از شبکه‌های عصبی پیچشی (CNNها) انجام دادند، این شبکه‌ها عبارتند از: بیانات واقعی‌ای که اعتبار آن‌ها باید مشخص شود، پروفایل و تاریخچه‌ی بیانات شخص گوینده و گزارش‌های منابع موثق. سینقانیا[۲۹] و همکارانش [۵۹] مدل HAN۳ را ارائه کردند؛ یک مدل شبکه‌ی عصبی با داشتن سلسله‌مراتب ۳- سطحی دارای لایه‌های توجه (Attention) برای تعیین میزان اهمیت هر بخش مقاله‌ی خبری ورودی که میزانی متفاوت دارد. این سه سطح مربوط‌به رمزگذاری دنباله‌های کلمه، جملات، و بدنه‌ی تیتر خبر است و از طریق این شبکه‌ی عصبی یک بردار برای خبر ترسیم می‌شود که مقاله‌ها را دسته‌بندی می‌کند. نویسندگان با درنظر گرفتن سایت‌هایی که PolitiFact آن‌ها را جعلی خوانده‌است و سایت‌هایی که Forbes اصل می‌داندشان، به دقت ۹۶.۷۷ درصد روی یک مجموعه‌داده دست یافتند. یکی از پایه‌های این مدل، کلمات و جملات مشخصی است که برای تعیین طبقه‌ی سند از مابقی کلمات و جملاتی که مهم‌تر هستند. برای داده‌های ورودی، جای‌گیری‌های کلمه در متن با استفاده از مدل GloVe (بردار جهانی) به‌دست می‌آید که یک روش یادگیری ماشینی بدون نظارت برای ترسیم نمایش‌های بردار- کلمه است [۶۰]. راشکین و همکارانش [۵۴] از شبکه‌ی حافظه‌ی طولانی کوتاه‌مدت (LSTM) استفاده کردند، بدین‌صورت‌که؛

[۲۸] Karimi
[۲۹] Singhania

دنباله‌ی کلمه را به‌عنوان داده‌ی ورودی در مجموعه‌داده‌ی PolitiFact[۳۰] درنظر گرفتند و LSTM در این‌مورد از سایر مدل‌ها مانند مدل بیز ساده برای دسته‌بندی دو-طبقه‌ای، بهتر عمل کرد. هرچه شاخصه‌های LIWC بیشتر افزوده شوند، عملکرد سایر طبقه‌بندی‌کننده‌ها نیز ارتقا می‌یابد. بااین‌حال، معضل مربوط به یادگیری عمیق، ماهیت جعبه سیاه آن است که نیاز به شفافیت بیشتر را از لحاظ تصویرسازی الگوهای متنی‌ای که برای طبقه‌بندی کاربردی‌تر باشند، تشدید می‌کند [۱۴].

از تکنیک‌های خلاصه‌سازی سند برای تولید خودکار یک خلاصه استفاده می‌شود که عبارت است از وصفی کوتاه و دربردارنده‌ی نکات اصلی سند. شیم[۳۱] و همکارانش عملکرد مدل‌های کشف اخبار جعلی که با تکنیک‌های خلاصه‌سازی سند استخراجی تکمیل شده بود را بررسی کردند. آن‌ها از Lexrankr، یک سیستم خلاصه‌سازی، برای استخراج خلاصه-هایی که دربردارنده‌ی سه جمله از مقاله‌های خبری بودند استفاده کردند. براساس تعبیه‌های کلمات TF-IDF و با استفاده از مدل‌های یادگیری ماشینی؛ عملکرد هردو مورد متن کامل و متن خلاصه‌شده برای کشف اخبار جعلی مورد ارزیابی قرار گرفت. درحالی‌که عملکرد مدل‌های مبتنی بر چکیده در مقایسه با مدل‌های مبتنی بر متن کامل به‌طور کلی اختلاف چندانی نداشتند، عملکرد مبتنی بر متن مدل رگرسیون لجستیک بهتر بود. طبقه‌بندی‌کننده SVM با بهترین عملکرد به دقت اعتبارسنجی برابر ۷۴ درصد، برای هردو مدل متن کامل و چکیده، رسید.

روش‌های کشف فریب: کشف اخبار جعلی ارتباط تنگاتنگی با شناسایی زبان فریب‌آمیز دارد. زبان فریب‌آمیز تمایزات قابل‌توجهی از لحاظ شاخصه‌های زبان‌شناسی دارد، برای مثال؛ در مقاله‌های مروری آنلاین درباره‌ی سفر که فریب‌آمیز هستند به عواطف مثبت‌تر، پیچیدگی و استفاده از نام‌های برند می‌توان برخورد [۱۳]. پیغام‌های فریب‌آمیز به‌طور عمدی ارسال می‌شوند تا به نتیجه‌گیری‌های غیرحقیقی منجر شوند. ولی اگر ارسال‌کننده‌ی پیغام، آن را ندانسته و بدون قصد فریب، ارسال کرده باشد آنگاه ممکن است فریب تلقی نشود [۲۴]. ژو[۳۲] و همکارانش [۲۴] در پژوهش‌های اولیه خود درخصوص کشف فریب خودکارشده، تجزیه‌وتحلیلی را روی سرنخ‌های مبتنی بر زبان‌شناسی در ارتباطات با واسطه‌ی رایانه انجام دادند که هر دوی پیغام‌های صادقانه و فریب‌آمیز بین داوطلبان را شامل می‌شد. طبق مشاهدات انجام‌شده؛ ارتباط فریب‌آمیز حاوی تعداد بیشتر کلمات، افعال و جملات بود، احتمالاً بدان سبب که باید دریافت‌کننده پیام را فریب می‌داد که اطلاعات واقعی هستند. محتوای فریب‌آمیز هم‌چنین حاوی تنوع واژگان کمتر از لحاظ محتوایی و لغوی است و از زبانی غیرفوری جهت وجود ارجاع‌به‌خودهای کمتری در آن، استفاده می‌کند، [۲۴، ۶۲]. ژانگ[۳۳] و همکارانش [۶۳] از شاخصه‌های زبان‌شناسی مشابهی استفاده کردند و با گزینش شاخصه‌ها به متمایزترین شاخصه‌ها درخصوص متون چینی فربینده و غیرفریبنده رسیدند. آن‌ها در پژوهشی دیگر از شرکت‌کنندگان انسانی خواستند تا بیاناتی صادقانه و بیاناتی فریب‌آمیز

[۳۰] www.politifact.com/.
[۳۱] Shim
[۳۲] Zhou
[۳۳] Zhang

داشته‌باشند؛ سپس براساس توزیع‌های طبقه‌ی کلمه‌ی LIWC مشخص شد که شمار کلمات، بیانگر قطعیت بیشتر در بیانات فریبنده بودند، دلیل آن می‌تواند این باشد که گوینده لازم می‌بیند که روی حقیقی بودن بیاناتش تأکید کند؛ درعین‌حال شمار کلماتی که بیانگر نظر و بینش فرد باشد در بیانات فریبنده، کمتر بود [۶۴]. هم‌چنین مشخص شد که در بیانات فریب‌آمیز گوینده‌ی (فاعل) انسانی سعی دارد تا با بکار بردن کلماتی که بیشتر به خودش مربوط می‌شوند خود را از دروغ‌های این بیانات مجزا کند.

فریب و جعل در متن را می‌توان با نگاه به سبک نگارش تشخیص داد. زمانی‌که نویسنده‌ی مقاله‌ای سعی دارد سبک نگارش خود را پنهان سازد، در شاخصه‌های زبان‌شناختی مشخصی از متن، تضاد مشاهده می‌شود. افروز[۳۴] و همکارانش [۲۱] مشخص کردند که صحت یک پُست آنلاین به صحت منبع آن پُست بستگی دارد. کوتاهی طول برخی از اخبار آنلاین در مقایسه با اسناد بلند مرسوم، موجب می‌شود شناسایی شاخصه‌های غنی در ارتباط با سبک نگارش چالش-برانگیز باشد [۶۵، ۶۶]. بااین‌حال، از طیف وسیعی از ویژگی‌های مربوط به سبک به‌کار گرفته می‌شوند تا اسناد کوتاه آنلاین را به نویسنده‌ای نسبت داد. ویژگی‌های لغوی متن شامل ویژگی‌های سطح- کاراکتر و سطح- کلمه است، درحالی‌که ویژگی‌های دستوری می‌تواند میان سبک نگارش نویسندگان بر اساس اینکه هر نویسنده چگونه جملاتش را مرتب کرده‌است یا به‌عبارت دیگر بر اساس ویژگی‌های سطح- جمله متن، تفکیک قائل شود [۲۷]. ویژگی‌های ساختاری، ویژگی‌های سطح- بالاتری هستند که به سازمان‌دهی کلی یا طرح‌بندی یک قسمت از متن می‌پردازند.

می‌توان برای شناسایی نویسنده‌ی پیام‌های آنلاین از ترکیب ویژگی‌های لغوی، دستوری، ساختاری و خاص- محتوا (مانند: واژگان کلیدی) استفاده کرد. افروز و همکارانش [۲۱] از مجموعه‌های ویژگی‌هایی که منسوب به نویسنده هستند استفاده کردند، مانند: مجموعه ویژگی رایت‌پرینتس (چاپ‌های نوشتاری) [۲۷]، مجموعه ویژگی برِننان و گرین‌استادت [۲۶]، و هم‌چنین سایر ویژگی‌های مربوطبه کشف دروغ [۶۷، ۶۸]. آن‌ها موفق شدند با استفاده از SVM روی مجموعه ویژگی رایت‌پرینتس، فریب در اسناد آنلاین را با بهترین نمره-F۱ مجموع یعنی ۹۶.۶درصد تشخیص دهند. هم‌چنین مشخص شد که مجموعه ویژگی رایت‌پرینتس برای کشف اخبار جعلی نیز کارآمد است [۵۵]. فِنگ[۳۵] و همکارانش [۶۹] از ویژگی‌های دستوری عمیق در قالب گرامر مستقل از متن تصادفی (PCFG) برای کشف مقالات مروری جعلی استفاده کردند. آن‌ها به دقت ۹۱.۲ درصد درخصوص یک مجموعه‌داده از مرورهای سفر رسیدند که این دقت بهتر از آن درصدی بود که طی ویژگی‌های دستوری و لغوی سطحی به‌دست آمد.

در یک متن منسجم، برخلاف مجموعه‌ای از جملات منفرد، ارتباطی کاربردی میان بخش‌های مختلف متن وجود دارد، که می‌تواند به‌صورت یک ساختار بلاغی طراحی شود. روبین[۳۶] و همکارانش از نظریه‌ی ساختار بلاغی [۷۰، ۷۱] توأم با

[۳۴] Afroz
[۳۵] Feng
[۳۶] Rubin

مدل‌سازی در فضای بردار استفاده کردند تا متن را در دسته صادقانه یا فریب‌آمیز طبقه‌بندی کنند[۳۵]. نخست، دو خوشه برای اخبار فریبنده و غیرفریبنده محاسبه شدند، سپس براساس محاسبه فاصله، به اخبار دریافتی، برچسب‌هایی اختصاص داده‌شد. بااستفاده‌از این رویکرد، این پژوهشگران توانستند به دقت ۶۷ درصدی دست یابند. البته، بااستفاده‌از مدل‌سازی تخمینی، آن‌ها به دقت ۵۶ درصد درخصوص مجموعه‌ی مورد آزمون رسیدند. این رویکرد غربالگری می‌تواند به شناسایی کاندیداها جهت راستی‌آزمایی بیشتر کمک کند.

طنز نوع دیگری از اخبار فریبنده است و می‌تواند حاوی سرنخ‌هایی باشد که مشخص می‌کنند که آن خبر غیرواقعی است. روبین و همکارانش [۷۲] از طبقه‌بندی‌کننده‌ی SVM استفاده کردند تا با به‌کارگیری ویژگی‌های ذیل اخبار طنزآمیز را طبقه‌بندی کنند: ۱. چرند بودن (پوچی): درصورتی که آخرین جمله‌ی خبر موجودیت‌های مشخصی را معرفی کند که کاملاً به مابقی خبر بی‌ربط هستند آنگاه میزان چرند بودن آن خبر بیشتر است. ۲. شوخ‌طبعی: درصورت وجود کمینه ارتباط بین اولین و آخرین جمله‌های خبر می‌توان به شوخ‌طبعی پی برد. ۳. ویژگی‌های مربوط به گرامر شامل برچسب‌گذاری جزء کلام (POS)، ۴. استفاده‌ی بیشتر از نقطه‌گذاری‌ها در طنز به‌سبب وجود جملات پیچیده‌ی دارای بندهای (جمله-واره‌های) متعدد، و ۵. کلمات دارای تأثیر منفی.

کشف عینیت (بی‌طرفی): یکی از ابعاد تجزیه‌وتحلیل سبک نگارش که می‌توان برای طبقه‌بندی متن آن را لحاظ کرد. مطالعه‌ی قطبیت متن برای پی بردن به این است که آیا یک‌جانبه یا متعصبانه است یا خیر. سرنخ‌های زبان‌شناسی کمک می‌کنند تا تعصبات در مقاله‌های ویکی‌پدیا فاش شوند؛ تعصبات قالب‌بندی که از طریق استفاده از کلمات فاعلی و اصطلاحات یک جانبه، شناسایی می‌شوند؛ و تعصبات دانش‌شناختی که ازطریق ویژگی‌هایی مانند استفاده از افعال فاعلی[۳۷] و قاطعانه شناسایی می‌شوند [۷۳]. پاتهَست[۳۸] و همکارانش [۲۲] با تجزیه‌وتحلیل اخبار دریافتی از منابع حزب راست و حزب چپ درمقایسه با اخبار منابع جریان‌های اصلی در مجموعه‌داده‌ی بازفید (Buzfeed)[۳۹]، به بررسی رابطه‌ی بین اخبار جعلی و اخبار فراحزبی پرداختند. آن‌ها از مفهوم نقاب‌برداری که پیش‌تر با هدف تأیید نویسندگی متن ارائه شده‌بود استفاده کردند [۷۴]. زمانی‌که دو اثر متفاوت از یک نویسنده دارای مجموعه کوچکی از ویژگی‌ها هستند که این ویژگی‌ها به‌اقتضای ژانر یا موضوع متفاوت در هر اثر یا به‌سبب تعامد نویسنده برای پنهان کردن سبک نگارش در یکی از آثار، متفاوت هستند، آنگاه نقاب‌زدایی می‌تواند در رفع معضل تأیید نویسنده کمک‌کننده باشد. ایده‌ی پس نقاب‌زدایی چنین است؛ اگر این

---

[۳۷] افعال فاعلی (Factive verbs) فعل‌هایی هستند که پیش‌فرض صدق یک گزاره را در بطن خود دارند، برای مثال: جمله‌ی « من می‌دانم او قاتل است» داری فعل فاعلی « دانستن» و گزاره‌ی « او قاتل است» می‌باشد که به‌طور پیش‌فرض صادق است. از نمونه‌های افعال فاعلی: دانستن، پشیمان بودن، فهمیدن، به یادآوردن، و فراموش کردن است. [مترجم].

[۳۸] Potthast

[۳۹] رجوع شود به https://www.buzzfeednews.com/article/craigsilverman/partisan-fb-pages-analysis و https://github.com/BuzzFeedNews/۲۰۱۶-۱۰-facebook-fact-check.

مجموعه‌های کوچک از ویژگی‌های متفاوت کنار گذاشته شوند، دشوار خواهد بود که میان متون یک نویسنده‌ی واحد از لحاظ ویژگی‌ها تفکیک قائل شد.

پاتهَست و همکارانش [۲۲] هم از ویژگی‌های زبان‌شناسی رایج، مانند: واژه‌های پالایشی و اِن-گرم‌ها، و هم چندین ویژگی خاص- دامنه استفاده کردند. آن‌ها ذکر کردند که تفکیک اخبار فراحزبی از اخبار متوازن امری میسر است. هم‌چنین، طبق مشاهدات آن‌ها؛ اخبار فراحزبی، حزب چپ و حزب راست مشترکات مشخصی در سبک‌شان با یکدیگر داشتند. بااین‌حال، نمره F۱ صرفاً با به‌کارگیری ویژگی‌های مربوط‌به سبک برای کشف اخبار جعلی، مطلوب درنمی‌آمد، پژوهشگران فوق پیشنهاد دادند که نتایج تجربی‌شان می‌تواند برای غربالگری اولیه برای کشف اخبار جعلی مورد استفاده قرار گیرد. تشخیص اخبار دارای ته‌مایه‌ی جانب‌دارانه، هم‌چنین می‌تواند به‌کمک تکنیک‌های نشانه‌گذاری منابع خبر یا کاربران، براساس هم-سویی یا وابستگی سیاسی آن‌ها صورت گیرد. این کار از طریق تجزیه‌وتحلیلِ به‌کارگیری هشتگ‌های سیاسی خاص توسط کاربران توییتر انجام می‌شود، زیرا هشتگ‌ها می‌توانند گویای وابستگی‌های سیاسی کاربران باشند [۷۵]. یکی از ویژگی‌ها در سایر آثار، تمایل ناشران دارای دیدگاه‌های جانب‌دارانه به ایجاد مقاله‌های حاوی اخبار جعلی است [۴۴].

تحلیل معنایی گزینه‌های احتمالی واقعیت یا به‌عبارت دیگر تحلیل اظهاراتی که صحت آن‌ها باید مشخص شود در فَکت‌چِکِر (FactChecker) انجام شد تا مشخص شود متن موردنظر عینی است یا صاحب‌نظرانه [۷۶]. ناکاشول و میتچِل[۴۰] گزینه-های احتمالی برای راستی‌آزمایی را از لحاظ سه‌گانه‌ی فاعل- فعل- مفعول (SVO) مدل‌سازی کردند. ارزیابی معنایی سه‌گانه‌ی SVO روی گزینه‌های احتمالی واقعیت، شامل این موارد است: ۱. یافتن نوع فاعل (S) و مفعولی (O) که برای فعل (یا عبارت فعلی) موردنظر خواهد آمد، ۲. تشخیص جوهره‌ی رابطه‌ی بین فاعل (S) و مفعول (O) برای فعل موردنظر (V). ۳. یافتن سایر افعال مترادف که می‌توانند در جای فعل موردنظر (V) بیایند. این تحلیل برای ساخت گزینه‌های احتمالی واقعیت که بتوانند به‌طور بالقوه جایگزین شوند استفاده می‌شوند، گزینه‌ی احتمالی موردنظر برای واقعیت را می‌توان در مقایسه با این گزینه‌های جایگزین رتبه‌بندی کرد. اگر دقت چنین تعریف شود که نمره‌ی باورپذیری یک ادعای صادقانه بیشتر از نمره‌ی باورپذیری یک ادعای غلط باشد، فکت‌چِکِر (FactChecker) پیشنهادی به دقت ۰.۹۰ درمورد گزینه‌های احتمالی بازیابی‌شده از ویکی‌پدیا رسیدند.

[۴۰] Nakashole and Mitchell

تکنیک‌های بصری و چندوجهی: اخبار می‌توانند حاوی محتواهای چندرسانه‌ای، مانند: تصاویر، فیلم و صوت باشد. ویژگی‌های تصاویر آمده در اخبار می‌تواند به ارتقای بیشتر دقتِ کشف اخبار جعلی مبتنی بر متن کمک کند. اخبار جعلی می‌تواند دربردارنده‌ی تصاویری باشد که مربوط به رویداد مورد شرحِ خبر نیستند، یا تصاویری که حتی ممکن است با افزودن یا حذف موجودیت‌هایی، اتصال و روش‌های دیگر دستکاری شده‌باشد [۳۶]. تکنیک‌های یادگیری عمیق در یادگیری ویژگی‌های نهفته‌ی هم تصاویر و هم متن موفق بوده‌اند. جین[۴۱] و همکارانش [۷۷] از مدل مبتنی بر شبکه-ی عصبی بازگشتی و مکانیزم توجه (attention) استفاده کردند تا هم‌بستگی میان تصاویر و متن را دریابند. مدل پیشنهادی بر اساس ورودی چندوجهی محتوای خبر، تصاویر الصاق‌شده و چندین محتوای اجتماعی پرورانده شده‌است. درخصوص مجموعه‌داده‌ی وِیبو (Weibo) که پژوهشگران فوق مطرح کردند، این مدل به دقت ۷۸.۸ درصد برای شناسایی شایعات از غیرشایعات دست یافت. خطار و سایر همکاران [۷۸] یک مدل تشخیص اخبار جعلی چندوجهی مبتنی بر شبکه عصبی را با استفاده از ویژگی‌های بصری و متنی پیشنهاد کرد. براساس داده‌های ارائه شده‌ی حاصل از خود رمزگذار[۴۲] مدل پیشنهادی از مدل‌های تک روشی برای تشخیص اخبار جعلی در مجموعه داده توییتر [۱۵] بهتر عمل می‌کند ودقت بدست‌آمده ۷۴.۵ درصد ار نشان می‌دهد در حالی‌که دقت بدست آمده از طریق مدل‌های متنی و بصری به طور مستقل ۵۲.۶به ترتیب ۵۲.۶ درصد و ۵۹.۶ درصد بوده است.

اخبار جعلی ممکن است حاوی تصاویر نادرست باشد که می تواند به تصاویری با کیفیت پایین یا گمراه کننده منجر شود که مرتبط نیستند و احتمالاً از برخی اخبار دیگر گرفته شده‌اند. چی و همکاران [۲۰] یک مدل هم‌جوشانی مبتنی بر شبکه عصبی را برای ایجاد فشار بر روی دامنه فرکانس (ویژگی‌های فیزیکی) و دامنه پیکسل (ویژگی‌های معنایی) از تصاویر مرتبط با اخبار برای تشخیص اخبار جعلی را پیشنهاد کردند. این روش استخراج ویژگی بصری، زمانی که با مدل‌های چندوجهی پیشنهادی قبلی استفاده شد، دقت بهبود یافته‌ای را نشان داد. تصاویر در اخبار واقعی به این سو گرایش دارند تا وجه‌های مختلفی را از خود نشان دهند، در حالی که اخبار جعلی ممکن است حاوی تصاویر نامرتبط باشند. یکی دیگر از ویژگی‌های مهم، وضوح تصویراست، زیرا یک خبر واقعی حاوی تصاویری با وضوح بیشتری نسبت به اخبار جعلی می‌باشد. یانگ و همکاران [۳۷]مدل TI-CNN را پیشنهاد کرد که از CNN که هم از ویژگی های آشکار و هم از ویژگی‌های پنهان تصاویر خام و محتوای متنی استفاده می‌نماید. مدل پیشنهادی با دقت بالای ۹۲ درصد از سایر روش‌های پایه در مجموعه داده Kaggle [۷۹] بهتر عمل می‌کند. جین و همکاران [۱۶] خاطر نشان کردند محتوای میکروبلاگ اخبار جعلی دربردارنده‌ی تصاویر تکراری است. در مقابل، محتوای مرتبط با اخبار واقعی دارای تنوع بیشتری در تصاویر است، زیرا در مورد یک رویداد واقعی، بسیاری از تصاویر مشروع در دسترس خواهد بود .اخبار واقعی نیز در مقایسه با همان میزان اخبار

[۴۱] Jin

[۴۲] خودرمزگذار (به انگلیسی: autoencoder) یک شبکه عصبی مصنوعی است که برای کدینگ از آن استفاده می‌شود. از خود رمزگذارها برای استخراج ویژگی و فشرده سازی نمایش داده‌های با ابعاد بالا، یا به عبارت دیگر برای کاهش ابعاد استفاده می‌شود.

جعلی توییت شده حاوی تصاویر بیشتری است. بر همین اساس به منظور تأیید اخبار از ویژگی‌های بصری در کنار مؤلفه‌های آماری بهره می‌برند. یکی از ویژگی‌های امتیاز شفافیت است که نشان می‌دهد توزیع تصاویر بین یک رویداد خبری خاص و مجموعه کامل رویدادهای خبری چقدر با یکدیگر متفاوت است.

تصاویر موجود در تاروپود خبری واقعی از منابع مختلف هستند، بر خلاف اخبار جعلی، که تصاویر را از منابع کمتر متنوع دریافت می‌کنند. بنابراین، امتیاز پست‌های خبری واقعی در خصوص این مؤلفه کمتر است.آن‌ها توانستند دقت را برای تأیید اخبار با استفاده از ویژگی‌های غیرتصویری و تصویری در مقایسه با دقت به دست آمده و با استفاده از ۱۱ ویژگی برتر مورد استفاده در [۳۹]، تا بیش از ۱۴ درصد بهبود بخشند. موضوع دیگر در رسانه‌های اجتماعی این است که تصاویر قدیمی، نادرست و دستکاری شده اغلب در رسانه‌های اجتماعی دست به دست می شوند. در سایر مطالعات، جین و همکاران [۸۰] از یک مجموعه داده واژگانی کمکی گسترده با برچسب ضعیف در کنار مجموعه آموزشی برای یادگیری بازنمایی تصویر با استفاده از CNN برای متمایز کردن تصاویر معتبر از تصاویر جعلی استفاده کرده‌اند.

استخراج مؤلفه‌ها از کانال‌هایRGB (قرمز، سبز، آبی) دامنه پیکسل در تصاویر با استفاده ازCNN ها، نتایج نوید دهنده‌ای را برای شناسایی تصاویر جعلی که با استفاده از شبکه‌های مولد، متخاصم[۴۳] (مترجم: مولد رقابتی) تولیدی ایجاد شده‌اند را نشان داده‌اند. [۸۱،۸۲].

اعتبار منبع: اعتبار اخبار وابسته‌به اعتبار منبع آن‌هاست. برای مثال؛ منبع ممکن است در گذشته محتوایی را تولید کرده باشد که در واقعیت غلط بوده‌است، حالا این منبع می‌تواند وب-سایتی باشد که خبر را منتشر کرده‌است یا مدیریت رسانه‌ی اجتماعی‌ای که خبر را پُست کرده‌است [۱۰]. شواهد نشان می‌دهند ویژگی‌های متنی معیارهای مطلوبی برای تشخیص حقیقت در روند کشف اخبار جعلی موجود در مقالات وب‌سایت‌ها می‌باشد. وب‌سایت‌های معتبرتر از مقالات ویکی‌پدیا استفاده می‌کنند و ویژگی‌های متنی مقالاتی که یک وب‌سایت منتشر می‌کند گویای تعصبات و حقانیت آن وب‌سایت است. در [۱۰]، به‌کارگیری خصوصیات مربوطبه ویکی‌پدیا منجر به رسیدن به دقت ۶۲.۲۹ درصد برای واقعیت شد و حذف خصوصیات مربوطبه ویکی‌پدیا منجر به شدیدترین افت در عملکرد پژوهش شد. شایان ذکر است؛ توییت‌های معتبرتر حاوی URLهایی هستند که به دامنه‌های مشهورتر در وب ارجاع دارند [۲۸].

با بررسی ویکی‌پدیا چنین مشاهده شده‌است که؛ خالقان مقاله‌های موثق عمدتاً کاربران باسابقه هستند، حال آن‌که مقاله‌هایی که حقه‌وفریب هستند عمدتاً توسط حساب‌های کاربری نسبتاً تازه خلق شده‌اند [۸۳]. اعتبار حساب‌های کاربری-ای در رسانه‌های اجتماعی که اخبار را ایجاد یا پُست می‌کنند از مؤلفه‌های مهم در تشخیص میزان اعتبار اخبار است.

[۴۳] شبکه‌های مولد رقابتی یا زایای دشمنگونه (به انگلیسی: Generative Adversarial Networks) یک کلاس از چارچوب‌های یادگیری ماشین است که ایان گودفلو و همکارانش در سال ۲۰۱۴ آن را پیشنهاد کردند. در این کلاس، دو شبکه عصبی در یک بازی روبروی یکدیگر قرار می‌گیرند (در چارچوب یک بازی با گردایش صفر، که آن را به نام بازی با مجموع صفر نیز در حوزه ی نظریه بازی ها می شناسیم، در چنین بازی هایی سود یک بازیکن به ضرر بازیکن دیگر است و هر گاه بازیکنی یک امتیاز می گیرد در واقع امتیازی از بازیکن مقابل کم می شود در نتیجه همواره مجموع امتیازات صفر است).

ویژگی‌های مربوط به حساب کاربر مانند: اطلاعات پروفایل حساب ازلحاظ موقعیت جغرافیایی آن، تاریخی که حساب ایجاد شده‌است، و تعداد توییت‌ها همگی می‌توانند برای تشخیص اینکه آیا حساب مشکوک است یا یک ربات (bot) است مفید واقع شوند [۳۸، ۸۴، ۸۵]. گوراجالا[۴۴] و همکارانش [۸۶] داده‌ای را شامل حساب‌های توییتری تجزیه‌وتحلیل کردند و سپس روشی را برای کشف پروفایل‌های جعلی بااستفاده‌از مطابقت الگویی (pattern-matching) با اسامی مورد نمایش و ارزیابی اوقات به‌روزرسانی توییت‌ها پیشنهاد دادند. هرچند که ربات‌ها (bots) به‌جهت منافع اجتماعی مورد استفاده قرار می‌گیرند، ولی می‌توان از آن‌ها برای اهدافی چون پخش کردن اطلاعات غلط و نیز اثرگذاری بر مناظرات سیاسی استفاده کرد [۳۸]. توییت‌هایی که توسط ربات‌ها (botها) پُست می‌شوند عمدتاً از طریق ابزار مبتنی بر API ایجاد شدند، برعکس؛ انسان‌هایی که خودشان در توییتر تولید محتوا می‌کنند از وب یا برنامه‌های کاربردی (اپلیکیشن‌های) موبایل استفاده می‌کنند [۸۷]. الگوهای رفتاری موقت ربات‌ها (botهای) خودکارشده متفاوت از الگوهای رفتاری کاربران معتبر است. هم‌چنین می‌توان با بررسی پراکندگی فاصله‌های زمانی بین ری‌توییت‌ها و بررسی تعداد دفعاتی که حساب، یک URL خاص را ری‌توییت کرده‌است به فعالیت ربات (bot) پی برد. ربات‌ها (botها) معمولاً سطوح فعالیت یکسانی را در طول یک هفته دارند، درصورتی‌که درمورد کاربران انسانی معمول این‌طور نیست [۸۷]. به‌علاوه، درخصوص ربات‌ها (botها) برخلاف افراد غیرمشهور (غیر سلبریتی) که تعداد دوستانشان به تعداد دنبال‌کنندگانشان نزدیک است، چنین دیده می‌شود که تعداد بسیاری از دوستان را اضافه (اَد) می‌کنند ولی درنهایت دنبال‌کنندگان کمتری دارند، زیرا «آبروی حسابشان» کمتر است [۸۷]. خصوصیات مربوط به عواطف نشان داده‌است؛ کاربران انسان عواطف منفی و مثبت قوی‌تری را در مقایسه با ربات‌ها (botها) ابزار می‌کنند، هم‌چنین کمتر به‌نظر می‌رسد ربات‌ها (botها) عواطفشان را در مورد موضوع مشخصی تغییر دهند [۸۹]. هو۴۵ و همکارانش [۱۲] نشان دادند که اطلاعات شبکه‌ی اجتماعی کاربران که در قالب ماتریس‌های مجاورت نمایش داده می‌شوند در تشخیص اسپمِرها در میکروبلاک‌ها مفید واقع می‌شوند.

هو و همکارانش هم‌چنین روی شناسایی منابع معتبرِ اطلاعات درخصوص موضوعات مشخص در توییتر کار کردند. نتایج نشان داده‌است؛ این اعتبار به ساختار شبکه‌ی اجتماعی حساب منبع و اینکه محتوای منبع تا چه حد به دامنه مربوط است بستگی دارد [۹۰]. می‌توان اعتبار یک منبع یا ناشر خبری مشخص را با بررسی دیدگاه‌هایشان (گرایششان به حزب چپ یا راست)، تخصص آن منبع در حوزه‌ی آن موضوع، و قالب (فرمت) خبرها (پژوهشی، سرمقاله ، و غیره) تشخیص داد [۲۳]. لانگ[۴۶] و همکارانش، به ۱۴.۵ درصد ارتقا در دقت خود درخصوص اعمال روش‌های مبنایی روی مجموعه داده‌ی لایِر (LIAR) [دروغگو] رسیدند و این کار را با الحاق اطلاعات سخنگو، ازجمله؛ حزبی که سخنگو بدان وابسته است، شغلش، عنوان و سِمتش و سابقه‌ای از هرگونه ادعای غیرواقعی‌ای که تاکنون داشته‌است، انجام دادند [۱۱، ۹۱]. یانگ[۴۷] و همکارانش،

---

[۴۴] Gurajala

[۴۵] Hu

[۴۶] Long

[۴۷] Yang

یک خصوصیت جدید به‌نام برنامه مشتری (Client Program) که در ساخت میکروبلاگ استفاده می‌شود را به خصوصیات موجود محتوایی، کاربر، انتشار اضافه کردند تا شایعات را در سینا وِیبو (Sina Weibo) شناسایی کنند. آن‌ها دریافتند که اگر از یک برنامه‌ی مشتری غیر موبایلی برای ساخت یک میکروبلاگ درباره‌ی رویدادی که در کشور دیگری اتفاق افتاده‌است استفاده شود، آنگاه احتمال اینکه شایعه باشد بالاست. افزودن خصوصیت‌های برنامه‌ی مشتری و خصوصیت‌های موقعیت جغرافیایی رویداد به خصوصیت‌های متداول کنونی مربوط‌به حساب، مانند اینکه حساب موثق است یا خیر؛ تعداد دنبال-کنندگان؛ و غیره، موجب شد تا دقت از ۷۲.۶ به ۷۷.۴ درصد ارتقا یابد. مشخص شده‌است؛ عملکرد به‌کارگیری خصوصیت-های کاربرانی که توییت‌ها را پُست می‌کنند برای تشخیص توییت‌های حاوی تصاویر جعلی، ضعیف بوده‌است [۴۷].

۱۵.۲.۲.۲ محیط اجتماعی

اخباری که به عنوان یک میکروبلاگ در یک پلتفرمِ شبکه‌بندی اجتماعی به اشتراک گذاشته می‌شوند، زمینه را برای مشارکت کاربران دیگر با اخبار فراهم می‌سازد. ممکن است کاربران دوست داشته باشند لایک کنند، نظر بدهند یا پست را با "دنبال کنندگان" یا "دوستان" خود به اشتراک بگذارند. مشخصات کاربرانی که به اخبار ارسال شده پاسخ می‌دهند و در شبکه‌ی تعاملات میان کاربران، نوع پاسخ داده شده از سوی کاربران و الگوی انتشار اخبار در رسانه‌ی اجتماعی مشارکت می‌کنند، برای شناسایی تفاوت‌ها در محیط اجتماعیِ اخبار جعلی و اخبار واقعی مورد استفاده قرار می‌گیرند. پست‌های اخبار جعلی می‌تواند پاسخ‌هایی را دربرداشته باشد که بیانگر شک‌اندیشی باشد [۳۹، ۴۱، ۹۳]. مشاهده شده است که اخبار جعلی به گونه‌ای متفاوت منتشر می‌شوند و با سرعت بالاتر و دسترسیِ گسترده‌تری انتشار می‌یابند [۴۲]. وب سایت‌های رسانه‌ی اجتماعی با بهره‌گیری از جمع‌سپاری، اخبار جعلی را از طریق درخواست دادن به کاربران، جهت تشخیص و نشانه‌گذاریِ صحیحِ اخبار جعلی شناسایی می‌کنند. در این مورد، اعتبار کاربرانی که اخبار را نشانه‌گذاری می‌کنند نیز ضروری بوده و می‌تواند بر اساس سابقه‌ی کاربر در نشانه‌گذاری اخبار تعیین گردد [۹۴].

در این بخش، ما به طور مفصل به جنبه‌های مختلف محیط اجتماعیِ اخبار جعلی می‌پردازیم و بحث خود را با شناسایی اخبار جعلی طبق الگوهای انتشار آن‌ها آغاز می‌کنیم.

انتشار اخبار. ویژگی‌های اخبار با محوریت انتشار آن‌ها که تعداد ریتوییت‌ها، رشته‌اتفاقات زندگی، تعداد نظرات پست‌ها و پیچیدگیِ درخت ریتوییت، را شامل می‌شود، برای تعیین اعتبار اخبار در رسانه‌ی اجتماعی مورد استفاده قرار می‌گیرند [۲۸، ۹۲، ۹۳]. مشخصات انتشار که برگرفته از درخت ریتوییت می‌باشد، نرخ بالایی از مثبت صحیح[۴۸] را برای اخبار جعلی بدست آورده‌اند، بنابراین اهمیت مشخصات گراف-محور را برای شناسایی اخبار باورنکردنی برجسته می‌کند [۳۹]. همچنین لازم به ذکر است، در حالی که درخت‌های انتشارِ رویدادها/ موضوعات مهم دارای پیچیدگی هستند، توییت‌هایی که در سطح بخصوصی دارای گستردگی بالایی هستند، یعنی همان تعداد زیادی از پست‌های بازارسال شده در سطح خاصی در درخت انتشار، احتمال بیشتری وجود دارد که قابل اعتماد باشند.

سوزوکی[۴۹]، پیشنهادی را برای محاسبه‌ی اعتبار پیام‌ها در رسانه‌ی اجتماعی مانند توییتر با استفاده از بازارسال یا ریتوویت کردن پیام دیگر کاربران ارائه داد [۵۹]. این پیشنهاد بر اساس این فرض است که یک پیام بسیار معتبر اغلب تنها با اصلاحات اندکی بازارسال می‌شود تا پیام اصلی صحیح و سالم بماند. در مقابل، بعید است که پیام‌هایی با اعتبار کمتر بازارسال شوند و حتی اگر آن‌ها دوباره ارسال شوند، کاربران تمایل پیدا می‌کنند که نظرات خود را به آن بیفزایند. کوان و همکاران[۵۰] [۳۱] به این نکته اشاره کردند که شایعه‌های غلط سهم زیادی در توییت‌هایی با الگوی یگانه دارند و در نتیجه

[۴۸] true positive rate
[۴۹] Suzuki
[۵۰] Kwon et al.

چنین توییت‌هایی از سوی دیگر کاربران نادیده گرفته می‌شوند. آن‌ها در کنار ویژگی‌های موقتی و زبان‌شناسی، از الگوهای ساختاری استفاده کردند تا شایعه‌های غلط/ تأییدنشده را شناسایی کنند و طبقه‌بند جنگل تصادفی[۵۱] بهترین امتیاز ۸۹.۳٪ از F۱ را بدست آورد.

انتشار یک توییت، همراه با توییتِ منبع که همان پایه‌ی اصلی بوده و پاسخ دیگر کاربران به توییت که همان نُودهای متصل به منبع با یال‌های هدایت شده است، می‌تواند نشان دهنده یک ساختار درختی باشد. یک رویکرد بالا به پایین، جهتِ انتشارِ اطلاعات را برای یال‌ها در نظر می‌گیرد، در حالی که رویکرد پایین به بالا جهت پاسخ‌ها را لحاظ می‌کند [۴۱]. اطلاعات مبتنی بر ساختار شامل تشخیص شایعه با استفاده از توابع کرنل و از طریق محاسبه‌ی شباهت‌ها میان درخت‌ها می‌شود [۹۶]. برای بهبود فرایند تشخیص شایعه‌ها، RNN ها همراه با ساختارهای درختِ انتشار، استفاده شده‌اند [۴۱]. مونتی و همکاران[۵۲] [۹۷] مشخصات ناهمگن که شامل مشخصات مربوط به محتوا، ساختار شبکه اجتماعی، پروفایل کاربر و انتشار اطلاعات با استفاده از یادگیری پیچیده‌ی هندسی است، را با یکدیگر ترکیب کردند. آن‌ها انتشار اطلاعات را در قالب یک درخت واپخش[۵۳] مدل سازی کردند و توانستند به دقت بالایی در شناسایی اخبار جعلی دست پیدا کنند که نشان می‌داد چنین اخباری می‌توانست در مراحل ابتداییِ انتشار نشانه‌گذاری گردند. یک مدل SVM که مبتنی بر گراف کرنل است و از درختان واپخش، محتوا و ویژگی‌های کاربر-محور استفاده می‌کند، نشان داد که دقت در طبقه‌بندی شایعه‌ها در Sina Weibo ارتقا پیدا کرده است [۹۸]. پژوهشی دیگر، سرعت اخبار جعلی و حقیقی را مقایسه کرد و مشخص نمود که احتمال ریتوییت کردن اخبار جعلی در مقایسه با اخبار واقعی بیشتر است، زیرا اخبار جعلی ظاهراً دارای محتوای "نوین" هستند. اخبار جعلی به صورت آبشاری ریتوییت می‌شوند، در نتیجه پیچیدگی و عمق آن (فاصله از گره منبع) بیشتر بوده و توسط افراد بیشتری در هر سطحی ریتوییت می گردند [۴۲].

انتشار شایعه از سمت کاربران درجه پایین[۵۴] (این درجه بر اساس تعداد دنبال کنندگان است) به کاربران درجه بالاتر[۵۵] یکی از ویژگی‌های مهمی است که برای تأیید شایعه‌ها مورد استفاده قرار گرفته است. کوان و همکاران [۳۱] به مطالعه‌ی طبقه‌بندی شایعه‌ها در توییتر پرداختند و به این نتیجه رسیدند که بخشی از فرایند انتشار اطلاعات از کاربران درجه پایین به کاربران درجه بالا دارای قدرتی پیش‌گویانه می‌باشد. وثوقی و همکاران [۳۰] به این نکته اشاره کردند که هنگامی که یک شایعه درست است، میزان انتشار آن از کاربران درجه پایین به کاربران درجه بالا بسیار زیاد است. دلیل این امر این است که کاربران درجه بالا یا کاربرانی با نفوذ بالا در ریتوییت کردن یک شایعه از طرف کاربرانی با نفوذ کمتر و هم‌چنین بدون دلیلی محکم که بیانگر صحت اطلاعات باشد، ریسک نمی‌کنند.

---

[۵۱] Random Forest classifier : یک روش یادگیری ترکیبی برای دسته بندی رگرسیون (یک نوع مدل اماری است برای پیش بینی یک متغیر از روی چند متغیر دیگر) می باشد.

[۵۲] Monti et al.

[۵۳] diffusion tree

[۵۴] low-degree

[۵۵] higher degree

مدل سازی زمانی. ما و همکاران[۵۶] [۹۹] بر روی تشخیص شایعه‌ها کار کردند: آن‌ها پست‌های میکروبلاگ فردی را طبقه بندی نمی‌کنند اما پست‌های میکروبلاگِ رویدادهایی را بدست می‌آورند که شایعه بودن یک رویداد را اثبات می‌کند. این پست‌ها به صورت سریِ زمانی مدل سازی شدند و مدل شبکه‌ی عصبی بازگشتی (RNN) با استفاده از ارزش‌های TF-IDF توسعه یافته است، و این ارزش‌ها، ارزش‌هایی هستند که در پست‌هایی با بازه‌های زمانی دسته بندی شده‌اند. این مدل پیشنهادی در پایگاه داده Weibo به بالاترین دقت ۰.۹۱ دست پیدا کرد. یو و همکاران[۵۷] [۱۰۰] در خطِ سیر پاراگرافِ مربوط به پست‌های میکروبلاگی، از شبکه عصبی پیچشی[۵۸] استفاده کردند و این پست‌ها مربوط به رویدادی هستند که به پنجره‌های زمانی جهت شناسایی اطلاعات نادرست در مقابلِ اطلاعات درست تقسیم شده و در داده‌های Weibo به دقت ۰.۹۳۳ ارتقا پیدا کرده‌اند.

مشارکت ویژگی های کاربران. هنگامی که یک کاربر تصمیم می گیرد با محتوایی ارتباط برقرار کند که با داستان‌ها و پیام‌های او هماهنگ است، داشتن تحلیل از کاربرانی که با یک پست در رسانه‌ی اجتماعی تعامل برقرار می‌کنند، می‌تواند به تشخیص حقه و فریب کمک کند. تاکینی و همکاران[۵۹] [۴۳] با استفاده از رگرسیون لجستیک و جمع‌سپاری برچسب بولیِ همساز[۶۰] که بر پایه‌ی کاربرانی است که پست‌های فیسبوک را "لایک" کرده‌اند، در مورد طبقه‌بندی حقه و فریب به دقتی بیشتر از ۹۹٪ رسیدند. این مقاله با تکمیل ویژگی‌های اجتماعی و ویژگی‌های محتوا-محور که در آن‌ها مشارکت‌های اجتماعیِ کمی در خصوص برخی از اخبار دیده می‌شود، توسعه داده شده است [۱۰۱]. احتمال زیادی وجود دارد که توییت‌هایی که از سوی برخی از کاربران ریتوییت شده‌اند و این کاربران در به اشتراک گذاری یا ارسال شایعه‌ها درگذشته دست داشته‌اند، شایعه باشند [۳۲]. شو و همکاران[۶۱] [۴۴] مدل TriFN را بر اساس ارتباط سه گانه‌ی موجود میان بخش‌ها – مقالات خبری، منتشرکنندگان و کاربرانی که مقالات را در رسانه اجتماعی به اشتراک می‌گذارند، پیشنهاد دادند. یکی از مبناهای این مدل، تمایل کاربران به اشتراک گذاریِ اخبار جعلی با اعتبار کمتر و تشکیل خوشه‌ها با کاربرانی مشابه می‌باشد. وثوقی و همکاران [۳۰] مشاهده کردند که احتمال اینکه کاربرانِ به شدت چالش‌برانگیز، در انتشار شایعات جعلی دست داشته باشند، بیشتر است، در حالی که شایعات حقیقی توسط کاربرانی با اعتبار نسبتاً بالا منتشر شده‌اند. با این حال، انتظار می‌رود که اخبار جعلی معمولاً توسط کاربرانی با اعتبار پایین منتشر شوند، اگرچه در برخی از موارد اضطراری، به عنوان مثال، در انفجارهای ماراتن بوستون، مشخص شد که کاربرانِ تأیید شده با دنبال کنندگانِ بیشتر نیز به دلیل دشواری در تأیید اطلاعات، اخبار جعلی را در مراحل اولیه به اشتراک گذاشتند [۱۰۲].

[۵۶] Ma et al.
[۵۷] Yu et al.
[۵۸] CNNs
[۵۹] Tacchini et al.
[۶۰] Logistic Regression and Harmonic Boolean Label Crowdsourcing
[۶۱] Shu et al.

واکنش کاربران. اخبار جعلی می‌تواند واکنش‌هایی مانند شک و تردید، حیرت و کنجکاوی را در میان کاربران به همراه داشته باشد [۴۰]. ممکن است کاربران نسبت به اطلاعات اخبار، دودلی و شک خود را با پرسیدن سؤال ابراز کنند [۳۹]. حجم پاسخ منفی یا انکاری که از سوی کاربران در توییت‌های مربوط به یک شایعه‌ی خاص اعلام می‌شود، مقیاس خوبی برای تأیید درستیِ شایعه است [۳۰]. واکنش کاربران از لحاظ ویژگی‌های زبان‌شناسی مانند حجم بالای کلمات نفی کننده که در پاسخ به یک شایعه استفاده می‌شود، همراه با دیگر ویژگی‌های انتشار مانند سهمِ انتشار از کاربرانِ درجه پایین به کاربرانِ درجه بالا در یک شبکه اجتماعی، و هم‌چنین ویژگی‌های زمانی مانند تناوب شوک‌های خارجی برای ابهام‌زدایی و دسته‌بندی دقیقِ شایعه در نظر گرفته می‌شوند [۳۱]. مراحل ابتداییِ چرخش اخبار در رسانه اجتماعی تنها اطلاعات محدودی از محیط اجتماعی را ارائه می‌دهند. با این حال، لیو و وو[۶۲] [۱۰۳] تمرکز خود را بر روی شناسایی اخبار جعلی در مراحل اولیه قرار دادند و توانستند در پایان پنج دقیقه از زمان آغاز انتشار اخبار، به دقت ۹۰٪ در طبقه‌بندی دست یابند. آن‌ها از چارچوب مبتنی بر CNN بهره بردند که این چارچوب از ویژگی‌های محتوای واکنش کاربران و ویژگی‌های واکنشی آن‌ها استفاده می‌کند. علاوه بر این، آن‌ها برای دادن اهمیت بیشتر به واکنش‌هایی خاص که برای متمایز شدنِ نوع خبر، حیاتی هستند، از مکانیزم توجهِ جایگاه-آگاه[۶۳] استفاده کردند.

در شناساییِ بی درنگِ شایعات، لیو و همکاران [۱۰۴] این نکته را ذکر کردند که استفاده از ویژگی‌های اعتقادی یعنی ویژگی‌هایی برای فهمیدن اینکه آیا کاربران از طریق یک الگوریتمِ قانون محور از اخبار حمایت می‌کنند، آن را انکار می‌کنند یا در مورد آن می‌پرسند، در شناساییِ اولیه مفید و حتی در شناساییِ مراحل بعدی نیز بهتر خواهد بود. جین و همکاران [۱۰۵] از ایده‌ها و نظرات متناقضی استفاده کردند که توسط کاربران در اخبار رسانه اجتماعی برای تعیین اعتبار آن از طریق واکاوی نظرات، پیشنهاد شده بود، و این نظرات از یادگیری بی‌نظارت[۶۴] بهره بردند و همراه با حمایت و مخالفت لینک‌ها در توییت‌ها، یک شبکه‌ی اعتبار ایجاد کردند. با مدل سازی فرایندِ انتشارِ اعتبار به عنوان یک مسئله‌ی بهینه سازیِ گراف، آن‌ها به دقت ۸۴٪ از مجموع داده‌های Sina Weibo دست یافتند. در نتیجه، با استفاده از پردازش زبان طبیعی[۶۵] در واکنش کاربر، می‌توان به بحث و گفتگوی ناشی از اخبار پی برد.

اطلاع یافتن از موضع دیگر کاربران، بر اساس پست‌های آن‌ها در مورد اخبار و هم‌چنین موضع دیگر منابع خبری نسبت به این ادعا، مسئله‌ای است که به خوبی در مورد آن مطالعه انجام شده است. طبق نظرات کاربران در رابطه با اخبار یا ادعاها، تلاش شده است تا موضع کاربران در این رویداد، شناسایی گردد. اولین مرحله از *چالش اخبار جعلی*[۶۶] تشخیص موضع بود، مثلاً ارائه‌ی یک تیتر (ادعا) و یک مقاله، که موضعِ متنِ مقاله را با توجه به تیتر بدین طریق تعیین می‌کرد: که

---

[۶۲] Liu and Wu
[۶۳] position-aware attention mechanism
[۶۴] unsupervised learning
[۶۵] NLP
[۶۶] http://www.fakenewschallenge.org.

آیا مقاله با تیتر هم‌خوانی دارد یا مخالف آن است، اینکه در مورد تیتر صحبت می‌کند یا اینکه به تیتر مربوط نمی‌شود. یک رویکرد، از ویژگی‌های (مدل) بسته‌ی کلمات[۶۷] استفاده می‌کند، مانند فراوانی کلمه[۶۸] و فراوانی وزنی کلمه‌ی کلیدی[۶۹] برای نشان دادن متن و پرسپترونِ[۷۰] چندلایه‌ای به عنوان یک دسته بند[۱۰۶]. یک رویکرد دیگر از روش ماشین بردار پشتیبان[۷۱] در مورد ویژگی‌های فراوانی وزنیِ کلمه‌ی کلیدی استفاده می‌کند تا تعیین کند که آیا تیتر مرتبط بوده است یا غیرمرتبط، در صورت مرتبط بودن، یک ساختار منطبق با حافظه‌ی طولانیِ کوتاه مدت[۷۲] می‌تواند اخبار-تیترها را در دسته‌های زیر طبقه بندی کند: موافقت‌ها/ مخالفت‌ها/ بحث شده‌ها [۱۰۷]. رویکردی مشابه در تعیین ارتباط/ عدم ارتباط و سپس انجام طبقه بندی‌ای با ساختار ریز و تفکیک شده نیز استفاده شده بود [۱۰۸].

تشخیص موضع در ابتدا در حالت یک تیتر مقاله و با توجه به یک ادعای مشخص جهت تعیین اینکه آیا مقاله از این ادعا حمایت می‌کرد، با آن مخالفت می‌کرد یا تنها آن را گزارش می‌داد، انجام شده بود [۱۰۹]. از نگاه رسانه‌ی اجتماعی، توییت‌هایی وجود دارند که برخلاف شایعات دروغین (بیانگر موضعی مخالف) می‌باشند و توییت‌هایی نیز هستند که با حمایت از اخبار، آن‌ها را تأیید می‌کنند [۱۱۰]. تأیید چنین موضع‌هایی نیز در کارگروه RumourEval که هدف آن تشخیص شایعات است، انجام می‌گیرد. این کارگروه در ابتدا، موضع نظرات کاربران را در مورد یک پست شناسایی می کند و سپس حقیقت و درستیِ این شایعه را که در یک توییت ارائه شده است، تعیین می‌نماید، توییتی که در لحظه‌ی ارسال آن اثبات نشده است [۱۱۳-۱۱۱]. در یکی از رویکردهای این کارگروه‌ها، چِن و همکاران[۷۳][۱۱۳] از یک چارچوب منطبق با سی‌ان‌ان استفاده کردند، به طوری که این توییت‌ها به عنوان ماتریس‌های بردارِ کلمه، رمزی شده بودند. در اولین کارگروه، توییت‌ها با عنوان‌های "حمایت کردن"، "پرسیدن"، "نظر دادن" و "انکار کردن" و در کارگروه دوم با عنوان‌های "شایعه" و "غیرشایعه" نامگذاری شدند [۱۱۳].

دیگر رویکردها. یک رویداد قابل انتشار، جزئیات مخصوصی دارد و این جزئیات یا موضوعات فرعی می‌توانند به رویدادهای فرعی اشاره کنند. جین و همکاران[۷۴][۱۱۴] یک شبکه‌ی سلسله مراتبی از اعتبار همراه با سه لایه‌ی متشکل از رویداد، رویداد فرعی و پیام‌های انفرادی ایجاد کردند. رویدادهای فرعی و پست‌های مربوط با استفاده از رویکردِ خوشه‌بندی تعیین شده بودند. آن‌ها ارزیابی اعتبار را به شکل یک مسئله‌ی بهینه‌سازی گراف درآوردند، به طوری که این اعتبار را بر روی گراف تکثیر کردند. این واحدها در سه لایه از طریق سه لینک وزنی به یکدیگر متصل شده بودند و عمل بهینه‌سازی بر

---

[۶۷] Bag of Words (BOW)

[۶۸] term frequency

[۶۹] TF-IDF

[۷۰] دسته ای از شبکه های عصبی مصنوعی پیشخور

[۷۱] Support Vector Machine

[۷۲] LSTM

[۷۳] Chen et al.

[۷۴] Jin et al.

اساس این محتوا انجام شده بود که این واحدهای متصل شده از طریق لینک‌های وزنی، دارای اعتبار مشابهی هستند. روچانسکی و همکاران[۷۵] [۱۱۵] از طریق ارتقای ویژگی‌های مربوط به متن اخبار، پاسخ به اخبار و ارتقای اخبار توسط کاربران و بدون استفاده از ویژگی‌های دستچین شده، یک مدل ترکیبی پیشنهاد دادند. این مدل دارای سه واحد است که اولین واحد شامل ویژگی‌های متنی و پاسخی می‌شود، و این ویژگی‌ها از طریق شبکه‌های عصبی بازگشتی[۷۶] بدست آمده‌اند که بر پایه‌ی داده‌های موقتیِ تعاملات کاربران با یک مقاله هستند. از طریق یک واحد جداگانه به ویژگی‌های منبع پی برده شده است و طبقه‌بندی این مقاله از طریق واحد سوم انجام گرفته است.

[۷۵] Ruchansky et al.

[۷۶] Recurrent Neural Networks

۱۵.۲.۲.۳ راستی‌آزمایی

راستی‌آزمایی شامل ارزیابی حقیقتِ ادعاها یا اظهاراتی می‌شود که توسط مردم ارسال شده‌اند، مانند سلبریتی‌ها و سیاستمداران [۱۱۶]. راستی‌آزمایی کنفرانس‌ها و آزمایشگاه‌های انجمن ارزیابی[۷۷] [۱۱۷-۱۱۹] شامل تشخیص اظهارات دروغین بیان شده طی گفتگوهای انتخابات ریاست جمهوری در ایالات متحده می‌شد. یکی از روش‌های مشهور برای ادعاهای راستی‌آزمایی این است که مقایسه‌ای میان شباهت معنایی [۱۲۰] اظهارات بیان شده علیه مجموعه داده‌ای از اظهارات راستی‌آزمایی شده که از قبل ایجاد شده است، انجام شود [۱۱۶]. با این حال، این روش نمی‌تواند جهت تأیید کامل ادعاهای جدید مورد استفاده قرار گیرد و لذا در چنین مواردی، رویکرد بهتر این است که یک پایگاه دانشی مانند ویکیپدیا در نظر گرفته شود تا منبعی از حقیقت باشد. با این حال، ولاشوس[۷۸] و ریدل[۷۹] به این موضوع توجه کردند که این رویکرد در مواردی که ادعاها شامل محاسباتی مبنی بر داده‌های موجود شوند، مؤثر نخواهد بود [۱۱۶].

کیامپاگلیا و همکاران[۸۰] [۴۶] ادعاها را با استفاده از گراف‌های دانشی ویکیپدیا انجام دادند تا اسناد پشتیبان دریافت کنند. چالش "استخراج حقیقت و تأیید[۸۱]" شامل طبقه‌بندی ادعاهایی چون "حمایت شده"، "تکذیب شده" یا "اطلاعات ناکافی" مبنی بر اطلاعات ویکیپدیا می‌شود [۴۵]. اولین گام در بیشتر محتواها، بازیابی اسناد مربوطه می‌باشد. این امر اغلب از طریق استخراج موارد نامگذاری شده یا اسامی از ادعاها انجام می‌گیرد و تلاش می‌کند تا مقالات منطبق در ویکیپدیا را از طریق جستجو با استفاده از این کلمات کلیدی بالا بیاورد. گام بعدی انتخاب جمله بود که شامل انتخاب جملات محدود از اسناد بازیابی شده می‌شد، اسنادی که به نوعی مدارک مربوط به ادعا را ارائه می‌داد. رویکردهای این گام شامل محاسبه‌ی شباهت جملات، موارد منطبق و طبقه‌بندی‌های بررسی شده می‌شد. مرحله‌ی نهایی با طبقه‌بندی سروکار داشته و رویکردهای مختلفی برای ادغام این اسناد با یکدیگر استفاده می‌شدند.

امتیاز نهایی‌ایی که در این چالش به این تیم داده شد هم بر اساس نتایج طبقه‌بندی بود و هم بر اساس اسناد بدست آمده. بالاترین امتیاز تیم (UNC-NLP) ۶۴.۲۱ درصد از امتیاز FEVER را کسب کردند و این مجموعه داده نیز در کار نهایی استفاده شد. هانسلوسکی و همکاران[۸۲] [۱۲۱] مرحله‌ی بازیابی اسناد را برای راستی‌آزمایی به عنوان مسئله‌ی پیونددهی موجودیت[۸۳] مدل‌سازی کردند و از مدل استنتاج ترتیبی پیشرفته‌ی[۸۴] اصلاح شده [۱۲۲] استفاده نمودند تا

---

[۷۷] CLEF Fact Checking

[۷۸] Vlachos

[۷۹] Riedel

[۸۰] Ciampaglia et al.

[۸۱] FEVER

[۸۲] Hanselowski et al.

[۸۳] entity linking problem

[۸۴] Enhanced Sequential Inference Model (ESIM)

جملات را برای گزینش در این اسناد بازیابی شده رتبه‌بندی کنند و ادعا را تأیید نمایند. یوندا و همکاران[۸۵] [۱۲۳] نیز از مدل استنتاجی ترتیبی برای یادگیری استفاده کردند تا تعیین کنند که آیا یک جمله از یک گزاره حمایت می‌کند یا آن را تکذیب می‌نماید. یین و روث[۸۶] [۱۲۴] سیستمی را جهت شناسایی اسناد مربوط به ادعایی که برپایه‌ی صفحات ویکی توسعه دادند و هم‌چنین حقیقت و درستی این ادعا بر روی مجموعه داده‌ی FEVER تعیین نمودند [۱۲۵]. در حالی که رویکردهای اولیه عمدتاً رویکردهای مبتنی بر آبراهه داده بودند، که در آن‌ها اسناد در ابتدا شناسایی و سپس ادعا تأیید می‌شود، چارچوب TWOWINGOS کارهای فرعی را در یک حالت یکپارچه مدل سازی می‌کند. یک مدل به صورت مشترک برای این دو کار فرعی توسعه داده شده است تا بتوانند یکدیگر را کامل کنند. ایده این است که بر اساس یک ادعا، اسناد شناسایی گردند و سپس اسناد درست، حقیقت این ادعا را تقویت کنند [۱۲۶].

تکنیکی دیگر برای راستی‌آزمایی ادعاها، بررسی استدلال‌های مخالف این ادعاهاست که در اخبار آمده‌اند [۱۲۷]. راستی‌آزماییِ یک خبر نیز می‌تواند با کمک شخص سوم انجام گیرد اما راستی‌آزمایی هر خبری پرهزینه است. کیم و همکاران[۸۷] [۱۲۸] جهت انتخاب اینکه از میان مجموعه داستان‌های جعلیِ مشخص شده توسط کاربران، کدام خبرها برای راستی‌آزماییِ بیشتر ارسال شوند، الگوریتم "CURB " را پیشنهاد دادند. در کتابی دیگر، حسن و همکاران [۱۲۹] اظهاراتی را مشخص کردند که برای راستی‌آزمایی ارزشمند بود و این فرایند راستی‌آزمایی از مدلی بهره برده بود که بر مبنای یک مجموعه داده‌یِ نشانه‌گذاری شده از اظهاراتِ بیان شده در مناظرات پیشین ریاست جمهوری امریکا توسعه یافته بود. ویژگی‌های زبان‌شناسیِ این اظهارات مانند عواطف، طول جمله، مفهوم TF-IDF، واحد واژگانی و انواع عناصر در این مدل استفاده شده‌اند.

۱۵.۲.۳ مدل‌های طبقه‌بندی

مدل‌های شناسایی اخبار جعلی معمولاً مدل‌های طبقه‌بندی دوگانه هستند، به این معنا که این مدل‌ها با پیش بینی یک ارزش دوگانه نشان می‌دهند که آیا این خبر جعلی است یا خیر. با این حال، طبقه‌بندی دوگانه ممکن است همیشه عملی نباشد، زیرا برخی از بخش‌های مقاله ممکن است واقعاً درست باشند در حالی که برخی نیز ممکن است جعلی باشند. برای چنین مواردی که اخبار اندکی جعلی است، برای رسیدن به هدف طبقه‌بندی می‌توان طبقه‌بندی‌های چندگانه را معرفی کرد تا حدود جعلی‌بودن خبر را نشان دهند [۵۸]. مجموعه داده‌های در دسترس همراه با طبقات چندگانه که نشان دهنده‌ی سطح درستیِ خبر است، شامل LIAR و Vlachos۱٤ می‌شود [۱۱۶]. به همین نحو، برای تشخیص شایعات، طبقات چندگانه نیز می‌توانند وجود داشته باشند، زیرا در برخی موارد، یک شایعه ممکن است حتی بعد از مدت زمانی، تأیید نشده باقی بماند. در این موارد و شایعات تأیید نشده، طبقاتی که دارای ریزساختار هستند، مورد استفاده قرار می‌گیرند،

[۸۵] Yoneda et al.

[۸۶] Yin and Roth

[۸۷] Kim et al.

به عبارت دیگر، شایعات نادرست، شایعات درست، غیرشایعات و شایعات تأیید نشده [۱۳۰، ۹۶، ۴۱]. مسئله‌ی شناسایی اخبار جعلی می‌تواند به جای یک مسئله‌ی طبقه‌بندی به عنوان یک مسئله‌ی رگرسیون طبقه‌بندی شود، در جایی که مدل رگرسیون امتیازی را برای نشان دادن سطح حقیقت در اخبار اختصاص می‌دهد [۷۶]. با این حال، اکثر مجموعه داده‌ها در این دامنه برای طبقه بندی طراحی شده‌اند، یا با عنوان‌های دوگانه یا با عنوان‌های چند دسته‌ای، که این امر باعث می‌شود در تبدیل این برچسب‌ها به امتیازهای پیوسته مشکل ایجاد شود [۱۳۱].

اکثر رویکردهای شناسایی اخبار جعلی بر توسعه‌ی مدل‌های مبتنی بر مجموعه داده‌های از قبل نشانه گذاری شده تکیه دارند [۹۲، ۶۱، ۳۹، ۳۳]. رویکردهای یادگیری عمیق قبلاً برای استخراج مشخصات پنهان در نمونه‌های تصویربرداری شده‌ی متن و تصاویر کاربرد داشتند[۳۷]. شبکه‌ی عصبی پیچشی[۸۸] عمدتاً برای شناسایی اخبار جعلی به کار می‌رفته که مبتنی بر متون و تصاویر خبری بوده است [۵۸، ۳۷، ۲۰، ۱۴]. شبکه‌های عصبی بازگشتی برای داده‌های سری زمانی[۸۹] استفاده می‌شده است [۹۹]. روش یادگیریِ نیمه نظارتی برای TweetCred استفاده می‌شود [۵۱] که امتیاز اعتبار توییت‌ها را در زمان واقعی پیش‌بینی می‌کند.. علاوه بر روش دسته‌بند مبنایی[۹۰]، روش‌های کلی مانند روش گردآوری خودراه انداز نیز مورد استفاده قرار گرفته‌اند و موجب ارتقای دقت در خصوص ویژگی‌هایی می‌شود که با ویژگی‌های مبتنی بر کاربر و مبتنی بر توییت هماهنگ شده‌اند [۵۵، ۳۴]. برای رتبه‌بندی اعتبار، از الگوریتم‌های رتبه‌بندی مانند رتبه‌بندی SVM و ADARANK [۱۳۲] استفاده شده است [۵۱، ۳۳]. رویکردهای خوشه‌بندی برای ایجاد خوشه‌ی خبرنامه‌ها در یک فضای بُرداری استفاده شده‌اند و این فضا بر پایه‌ی شباهت ویژگی آنها و ارزیابی داستان‌های جدید با مبنای فاصله‌ی آنها می‌باشد، به عنوان مثال، فاصله‌ی اقلی‌دوسی با مراکز خوشه [۳۵]. پژوهشگران چندین روش برای مقابله با اخبار جعلی به محض شناسایی آن بر روی یک شبکه‌ی اجتماعی پیشنهاد داده‌اند [۱۳۵-۱۳۳].

جدول ۱۵.۲ مجموع داده‌های شناسایی اخبار جعلی

نام (زبان)	واحد	طبقه	منابع
ولاشوس ۱۴ (انگلیس)	بیانیه	درست، عمدتاً درست، نیمه درست، نیمه غلط، عمدتاً غلط، غلط	[۱۱۶]
توییتر ۱۵ (انگلیس)	رشته	غیرشایعات، شایعات غلط، شایعات درست و شایعات تأیید نشده	[۹۶]

[۸۸] Convolutional Neural Networks (CNNs)

[۸۹] time-series data

[۹۰] base classifier

نام (زبان)	واحد	طبقه	منابع
کِردبانک ۱۵(انگلیس)	رویداد	قطعاً دقیق، احتمالا دقیق، غیرقطعی،احتمالا غیردقیق، قطعا غیردقیق	[۱۳۶]
مدیااِوال ۱۶ (انگلیس)	پست، تصویر	جعلی، واقعی	[۱۵]
کاگل (انگلیس): واقعیت را از اخبار جعلی گرفتن	ادعاها	تعصب، ناخوشایند، توطئه، جعلی، نفرت، junksci ، هجونامه، حالت، نوع	[۷۹]
FNC۱۶ (انگلیس) چالش اخبار جعلی	(ادعا، مقاله)	موافق، مخالف، بحث، غیرمربوط	[۱۳۷]
PHEME (انگلیس)	داستان شایعه توییتری	درست، غلط، تأیید نشده	[۱۳۰]
توییتر ۱۷ (انگلیس)	رشته	غیرشایعه، شایعه غلط، شایعه درست، و شایعه تأیید نشده	[۹۶, ۹۹]
جین ۱۷ (انگلیس)	توییت‌های تصویردار (وِی بوا)	شایعه، غیرشایعه	[۷۷]
Ma۱۷ (انگلیس)	توییت منبع (درخت تکثیر)	درست، غلط، تأیید نشده، غیرشایعه	[۹۶]
Liar۱۷ (انگلیس)	پست	دروغ وحشتناک، غلط، با حداقل ترین درجه از درستی، نیمه درست، بسیار درست و درست	[۱۱]
Horne۱۷ (انگلیس)	خبرنامه (وب سایت)	واقعی، جعلی، هجوآمیز	[۲۹]
SemEval۱۷ Task B (انگلیس)	رشته	صحت	[۱۱۱]
شبکه اخبار جعلی (انگلیس)	ادعاها، کاربران، مشارکت ، لینک‌های اجتماعی	جعلی، درست	[۳, ۱۳۸]
Pratiwi۱۷ (هندی)	صفحه مقاله	معتبر، حقه	[۱۳۹]
Arabic FN(عربی)	(ادعا، مقاله)	موافق، مخالف، بحث، غیرمرتبط	[۱۴۰]
Fever۱۸ (انگلیس)	(ادعا، ویکیپدیا)	حمایت، تکذیب، اطلاعات ناکافی	[۴۵]
Vosoughi۱۸	توییت‌ها	درست، غلط	[۴۲]
CLEF۱۸ (انگلیسی، عربی)	ادعاها	نیمه درست، غلط، درست	[۱۴۱]
Rosas۱۸ (a)(انگلیسی): FakeNewsAMT	مقاله‌ی خبری	جعلی، قانونی	[۵۳]
Rosas۱۸ (b) (انگلیسی): سلبریتی	مقاله‌ی خبری	جعلی، قانونی	[۵۳]

۱۵.۲.۴ مجموع داده‌ها برای شناسایی اخبار جعلی

در جدول ۱۵.۲، ما مجموع داده‌هایی که در مطالعات مربوط به شناسایی اخبار جعلی استفاده شده‌اند را خلاصه کرده‌ایم. جمع سپاری یک رویکرد معمولی است که برای نشانه گذاری مجموع داده‌ها مورد استفاده قرار می‌گیرد. یک مثال از جمع سپاری، جمع آوری پست‌های رسانه‌ی اجتماعی مانند توییت‌ها از طریق Twitter API و درخواست از کارکنان تُرک مکانیکی آمازون[۹۱] جهت ارزیابی اعتبار حوادث برای تفسیر مجموع داده‌ها [۱۳۶] می‌باشد. رویکرد دیگر این است که اخبار قانونی را از وب سایت‌های خبری دریافت کنیم و از کارکنان AMT بخواهیم تا با ایجاد یک ورژن جعلی از آنها، نمونه‌ی کافی از اخبار جعلی بدست آورند.

۱۵.۳ نتیجه گیری

نشان داده شده است که اخبار جعلی به طور بالقوه برای اکوسیستم آنلاین تهدیدآمیز هستند، زیرا می‌توانند بر بسیاری از رویدادهای مهمِ خارج از آن تأثیر بگذارند. تکثیر اخبار جعلی یک مسئله‌ی جدی است که بسیاری از جوامع را تحت تأثیر قرار داده و ما می‌توانیم انتظار داشته باشیم که کارهای بسیار بیشتری در این حوزه در آینده‌ی نزدیک انجام گیرد. در این فصل، ما مروری بر بهترین تکنیک‌ها و رویکردهای فعلی مربوط به جنبه‌های مختلف شناسایی اخبار جعلی خواهیم کرد. ما در مورد تکنیک‌های شناسایی اخبار جعلی درخصوص محتوای اخبار و مشخصات منبع، محیط اجتماعی و راستی‌آزمایی صحبت کرده‌ایم.

شبکه‌های اجتماعیِ جهان واقعیت[۹۲] به شدت پویا بوده و به سرعت رشد می‌کنند. در این شبکه‌ها، اطلاعات به سرعت از یک کاربر به کاربر دیگری منتقل می‌شود. پژوهشگران هنوز در حال طراحی راه حل‌هایی هستند که بتوانند در زندگی واقعی برای کنترل مؤثر محتوای انتقالی اخبار جعلی در زندگی بشر، کاربردی باشند. حجم محتوا در رسانه‌های اجتماعی بسیار زیاد است. بنابراین الگوریتم‌ها باید بهینه‌سازی شوند تا برای داده‌های جریان بلادرنگ[۹۳] مؤثر واقع شوند. در شرایط ضروری مانند یک حمله تروریستی، خبرهای جعلی می‌توانند آسیب‌های جبران ناپذیری وارد کنند و لذا الگوریتم‌ها برای شناسایی اخبار جعلی باید بتوانند نظارت و پیام‌رسانی بلادرنگ را هرچه سریعتر انجام دهند.

۱۵.۴ مطالعات بیشتر

[۹۱] Amazon Mechanical Turk (AMT)

[۹۲] Real-world social networks

[۹۳] real-time streaming data

ما پیشنهاد می‌دهیم تا با بررسی مطالعات ذکر شده در مورد موضوع این تحقیق در قسمت پایین، اطلاعات بیشتری را از دیدگاه‌های مختلف بدست آورید. منابع [۱۴۵-۱۴۲] مقدمه‌ای کوتاه در مورد اخبار جعلی و اطلاعات نادرست ارائه می‌دهد. پژوهش زانتو و همکاران[۹۴] [۱۴۶] یک نظرسنجی در مورد اخبار جعلی است که به بررسی اینکه چگونه مردم اخبار جعلی یا اطلاعات غلط را در مرحله‌ی سیاسی، انتشار و شناسایی آن، دریافت و درک می‌کنند، می‌پردازد. همانطور که صحبت کردیم، شناسایی اخبار جعلی یک حوزه‌ی مطالعاتی خوب است؛ مقالات زیر شامل نظرسنجی‌های انجام شده در مورد شناسایی اخبار جعلی می‌باشد [۱۴۹-۱۴۷، ۴، ۲]. شو و همکاران[۹۵] [۱۵۰] در مورد شناسایی و کاهش اخبار جعلی که از تکنیک‌های تحلیل شبکه‌ها استفاده می‌کند، صحبت کرده‌اند. دیگر مقالات شامل نظرسنجی در خصوص شناسایی خودکارِ شایعه در میکروبلاگ ها نوشته‌ی کائو و همکاران[۹۶] [۱۵۱]، نظرسنجی در مورد انتشار شایعه‌ها در توییتر نوشته‌ی سرانو و همکاران[۹۷] [۱۵۲]، و مقاله‌ای دیگر در مورد شناسایی و رفع شایعه نوشته‌ی زوبیاگاه و همکاران[۹۸] [۱۵۳] می‌باشد. شِلکِه[۹۹] و آتار[۱۰۰] به مروری بر رویکردهای شناسایی منبع شایعه‌ها و اطلاعات نادرست پرداختند؛ اطلاعات بیشتر در [۱۵۴] در دسترس می‌باشد. رِم[۱۰۱] [۱۵۵] برای قدرت دادن به کاربران در مقابله با اخبار جعلی از طریق ارزیابی خودکارِ محتوا و گزینه‌های جایگزینِ مربوط به مصرف رسانه‌ای، یک زیرساخت فناوری ترکیبی را پیشنهاد داد. نی دی و همکاران[۱۰۲] [۱۵۶] به بررسی مدل‌های تکثیر برای انتشار شایعات پرداختند؛ اغلب مدل‌های پیشنهادی در خصوص انتشار بر پایه‌ی مدل‌های اپیدمی هستند. وانگ و همکاران[۱۰۳] [۱۵۷] با چشم‌اندازی وسیع به بررسی ربات‌های اجتماعی و نقش آن‌ها در انتشار اخبار جعلی پرداختند. آلِمانو[۱۰۴] [۱۵۸] در مورد رویکردهای اخبار واقعی از چشم‌انداز سیاستگذاری صحبت کرد. سولیوان[۱۰۵] [۱۵۹] مطالعه‌ای در مورد رویکرد علم کتابخانه و اطلاعات و نقایص آن‌ها در کنترل اخبار جعلی انجام داد. دیگر مطالعات کوتاه در خصوص اخبار جعلی را می توان در [۱۶۳، ۱۶۰، ۵] مشاهده نمود. تعدادی از مقالات مشهور در مورد اخبار جعلی بعد از انتخابات امریکا نیز در [۱۶۵-۱۶۴] موجود می‌باشد.

---

[۹۴] Zannettou et al.

[۹۵] Shu et al.

[۹۶] Cao et al.

[۹۷] Serrano et al.

[۹۸] Zubiaga et al.

[۹۹] Shelke

[۱۰۰] Attar

[۱۰۱] Rehm

[۱۰۲] Ndii et al.

[۱۰۳] Wang et al.

[۱۰۴] Alemanno

[۱۰۵] Sullivan

فصل ۱۶

تکنیک های کاهش و انتشار

اخبار جعلی: نظرسنجی

آکراتی ساکسِنا، پراتیشتا ساکسِنا و هریتا رِدی[۱۰۶]

چکیده: امروزه، اکثر وب سایت‌های شبکه‌های آنلاین اجتماعی میزبان میلیون‌ها حساب کاربری است. این وب سایت‌ها پلتفرمی مناسب برای به اشتراک گذاشتن اطلاعات و نظرات در قالب میکروبلاگ ها را فراهم می‌سازد. با این حال، اشتراک گذاری آسان نیز چندشاخگی در قالب اخبار جعلی، اطلاعات نادرست و شایعات را به همراه دارد، که اخیراً به شدت متداول شده است. تأثیر انتشار اخبار جعلی در اکثر رویدادهای سیاسی مانند انتخابات ایالات متحده و انتخابات جاکارتا و هم‌چنین تخریب شهرت سلبریتی‌ها و شرکت‌ها مشاهده شد. پژوهشگران به مطالعه‌ی انتشار اخبار جعلی در وب سایت‌های رسانه‌ی اجتماعی پرداخته‌اند و تکنیک‌های گوناگونی برای مقابله با این اخبار پیشنهاد داده‌اند. در این فصل ما در مورد مدل‌های انتشار اطلاعات نادرست بحث می‌کنیم و مروری بر تکنیک‌های کاهش اخبار جعلی خواهیم داشت. هم‌چنین لیستی از مجموع داده‌های استفاده شده در مطالعات مربوط به اخبار جعلی را ارائه می‌دهیم. نتیجه‌گیری این فصل با سوالات باز پژوهشی همراه خواهد شد.

۱۶.۱ مقدمه

از زمان پیدایش وب جهان گستر (WWW) بهره‌برداری از اخبار از منابع خبری مرسوم مانند روزنامه‌ها به منابع آنلاین خبری تغییر یافته است و این امر به مردم این امکان را می‌دهد که از بخش‌های مختلف دنیا در جریان اخبار و گرایش‌های متفاوت قرار گیرند. رایج‌ترین منبع آنلاین در حال حاضر شامل دو مدل می‌شود: (۱) وب‌سایت‌های مستقلی که دربردارنده‌ی وب‌سایت‌های خبری و وب‌سایت‌های رسانه‌ای هستند و (۲) رسانه‌های اجتماعی که پلتفرمی برای به اشتراک‌گذاری اخبار و ایده‌ها فراهم می‌سازد و قابلیت به اشتراک‌گذاری در آینده توسط دیگر کاربران را دارد [۱]. رسانه‌ی اجتماعی، شکل این مسئله را که چگونه مردم اطلاعات و نظرات خود را به یکدیگر مخابره می‌کنند، دگرگون ساخته است. شبکه‌های آنلاین اجتماعی[۱۰۷] مانند گوگل پلاس، فیسبوک و توییتر، نه تنها به اشتراک‌گذاری محتوای متنی بلکه به اشتراک‌گذاری لینک‌های

[۱۰۶] A. Saxena (✉) Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands
e-mail: a.saxena@tue.nl
P. Saxena G. L. Bajaj Institute of Technology and Management, Greater Noida, India
e-mail: pratistha.saxena@glbitm.ac.in
H. Reddy
Surat, Gujarat, India

https://doi.org/۱۰,۱۰۰۷/۹۷۸-۹۸۱-۱٦-۳۳۹۸-۰ ۱۵
[۱۰۷] OSN

چندرسانه‌ای و URL با وب سایت‌های خارجی همراه با مخاطبی گسترده را آسان ساخته است. دلیل اصلی اینکه چرا بهره‌برداری از اخبار در رسانه‌های اجتماعی افزایش یافته است شامل دو موضوع می‌شود: (۱) هزینه‌ی پایین دسترسی اینترنت و (۲) اشتراک گذاری آسان یا بحث در مورد اخبار با افراد دیگر در رسانه‌های اجتماعی [۲]. با این حال، دسترسی راحت تر به اینترنت و انتشار سریع محتوا بر روی رسانه‌های اجتماعی نیز به مسئله‌ی انتشار "اخبار جعلی" دامن زده است.

اثرات منفی این موضوع در انتخابات ریاست جمهوری ایالات متحده سال ۲۰۱۶ مشاهده شده است؛ مطالعه‌ای ده روزه در مورد توییتر نشان می‌دهد که کاربران در سراسر کشور بیش از آنکه محتواهای موثق تولید شده توسط افراد حرفه‌ای را بدست آورند، سراغ محتواهایی رفتند که دارای اطلاعات نادرست، توطئه و دوگانگی بود [۳]. یک تئوری مشهور توطئه که همان "رسوایی پیتزاگیت" بود، به طور گسترده در طول مدت انتخابات در حدود یک میلیون توییت در توییتر منتشر کرد و جعلی بودن آن نیز اثبات شد [۲]. مشخص کردن صحت اطلاعات برای انسان کاری دشوار است، به ویژه اینکه اگر این اطلاعات با این هدف ایجاد شده باشند که شکل قانونی داشته باشند. این مورد در آزمایشی بر روی مقالات فریب آمیز ویکیپدیا نشان داده شد، به طوری که حتی کاربران آموزش دیده نیز برخی مقالات فریب‌آمیز را با ارجاع به درست بودن آنها اشتباه تعبیر کرده بودند [۴،۵]. اخبار غیرقابل اعتماد ممکن است شامل حقه‌هایی باشند که سعی دارند خوانندگان را به دو طریق گمراه کنند: (۱) با بدگمانی و تبلیغات که برای ارتقای یک دستورکاری مشخص ایجاد شده است و (۲) هجونویسی که برای ایجاد طنز در اخبار جعلی ایجاد شده است [۶]. بنابراین، احتمالاً خوانندگان بتوانند هدف اخبار طنزآمیز را شناسایی کنند و آن‌ها را جدی نگیرند.

اخبار جعلی ارتباط بسیار نزدیکی با شایعات در رسانه‌ی اجتماعی دارند. دی فونزو[۱۰۸] و بوردیا[۱۰۹] [۷] شایعه را اینگونه تعریف کرده‌اند: اطلاعات تأیید نشده‌ای که در محیط‌های خطرناک، مبهم یا تهدیدات بالقوه جریان دارد. اگرچه شایعه در یک زمان مشخصی تأیید نشده است اما ممکن است به صورت غلط یا درست منتشر شود و یا حتی در آینده تأیید نشده باقی بماند. با این حال، برخی از آثار تحقیقاتی، شایعه را به عنوان اظهاراتی در نظر می‌گیرند که در نهایت غلط بودن آن اثبات می‌شود [۸]. پست‌های رسانه‌ی اجتماعی که مربوط به یک رویداد شایعه شده است ممکن است انواع مختلفی از واکنش‌ها را در کاربران برانگیزد؛ امکان دارد کاربران موافقت کنند؛ مخالفت کنند یا حتی خواستار اسناد بیشتری در حمایت از این ادعاهای ذکر شده در پست باشند [۸]. در این پژوهش، ما در مورد آثار پژوهشی مرتبط با اخبار جعلی، شایعات و اطلاعات نادرست صحبت می‌کنیم؛ و از این اصطلاحات در طول پژوهش به طور متناوب استفاده خواهد شد.

تأثیر منفیِ اخبار جعلی، توجه پژوهشگران را به مطالعه در مورد انتشار، شناسایی و کاهش این اخبار جلب کرده است. درک این موضوع که چگونه اخبار جعلی در مقایسه با اخبار واقعی انتشار می‌یابند، دربردارنده ارتباطات رسانه‌ی اجتماعی و به اشتراک گذاری اطلاعات به عنوان یک شبکه می‌باشد. به‌منظور مطالعه در خصوص انتشار اطلاعات نادرست که در

[۱۰۸] DiFonzo

[۱۰۹] Bordia

آینده در این فصل در مورد آن صحبت خواهیم کرد، چهار نوع اصلی از مدل‌های انتشار وجود دارد. در وب سایت‌های شبکه سازیِ اجتماعی مانند فیسبوک، برخوردی گزینشی با اخبار از نظر ایدئولوژی وجود دارد، زیرا کاربران تمایل دارند تا دیگر کاربرانی که دیدگاه‌های مشابهی دارند را دنبال کنند و این امر منجر به ایجاد پدیده‌ی اتاق پژواک یا حباب فیلتر می‌شود [۹-۱۱]. به دلیل وجود اتاق پژواک، کاربران مکرراً در معرض اخباری هستند که برای یک ایدئولوژی خاصی تولید شده‌اند، ایدئولوژی‌ایی که مطلوب افراد مجاور بوده و احتمال باور آن می‌رود. لذا، وجود اتاق پژواک مسئله‌ی مقابله با اخبار جعلی را چالشی‌تر می‌سازد.

پژوهشگران چندین روش برای کاهش اخبار جعلی پیشنهاد داده‌اند؛ دسته‌های اصلی این رویکردها عبارتند از (۱) مسدودسازی نفوذ [۱۲] و (ii) پویش حقیقت[۱۱۰] [۱۳]. هدف در رویکردهای مسدودسازی نفوذ، تعیین تعداد حداقلی از کاربرانی است که لازم است برای به حداقل رساندن انتشار اخبار جعلی مصونیت پیدا کنند. در دسته‌ی دوم، یعنی پویش حقیقت، ایده‌ی اصلی این است که اطمینان حاصل گردد که کاربران نسبت به اطلاعات درست آگاهی دارند تا آن را باور کنند. در تکنیک‌های پویش حقیقت، ما مجموعه‌ای از کاربرانی را شناسایی می‌کنیم که "تضعیف کننده" نامیده می‌شوند، کسانی که برای مقابله با تأثیر انتشار اخبار جعلی در شبکه اجتماعی، دست به انتشار اطلاعات درست می‌زنند. علاوه بر این، پژوهشگران چندین ابزار کاهشی را توسعه داده‌اند تا به کاربران در خصوص اعتبار محتوا برای جلوگیری از اشاعه‌ی بیشتر اخبار جعلی آگاهی دهند. برای استفاده از این ابزارها، لازم است حداکثر استفاده از تکنیک‌های شناسایی اخبار جعلی انجام گیرد تا اعتبار سنجیده شود و پست‌ها به درستی نشانه‌گذاری گردند [۱۵،۱۴].

ساختار این فصل در آینده توضیح داده خواهد شد. در بخش ۱۶.۲، ما در مورد مدل‌های انتشار اخبار جعلی صحبت خواهیم کرد. در بخش ۱۶.۳، به موضوع تکنیک‌های کاهش اخبار جعلی پرداخته خواهد شد و در آخر نتیجه‌گیری و ایده‌های بیشتر جهت تحقیق در این زمینه ارائه خواهد شد.

۱۶.۲ مدل‌های انتشار

در شبکه‌های آنلاین اجتماعی، اطلاعات از گروهی از گره‌های منبع که گره‌های آغازگر[۱۱۱] نیز نامیده می‌شوند، شروع به پخش شدن می‌کند. این گره‌ها اطلاعات را در حساب‌های آن‌ها به اشتراک می‌گذارند و گره‌های مجاور خود را تحت تأثیر قرار می‌دهند. به محض اینکه یک گرهِ مجاور، اطلاعات را باور کند، اطلاعات را با نُودهای مجاور خود به اشتراک می‌گذارد. بدین صورت، اطلاعات در شبکه منتشر می‌گردد. در شبکه‌های اجتماعی، هر لینک هدایت شده و هدایت نشده احتمالاً دارای تأثیری است که نشان دهنده‌ی احتمال انتقال اطلاعات به گرهِ هدف از گره منبع می‌باشد. در صورتی که کاربر

[۱۱۰] truth campaigning

[۱۱۱] seed nodes

اطلاعات ارائه شده را باور کند، آن کاربر تحت تأثیر قرار گرفته یا آلوده شده نامیده می‌شود. چندین مدلِ انتشار جهت شبیه‌سازیِ اینکه چگونه اطلاعات در یک شبکه پخش می‌شود، پیشنهاد گردیده است. این مدل‌ها نیز به مدل‌های گسترش اشاره دارند. در آینده، ما در مورد مدل‌هایی که پدیده‌ی انتشار بنیادی را توصیف می‌کنند و هم‌چنین در مورد ضمایم آن‌ها صحبت خواهیم کرد.

۱۶.۲.۱ مدل آبشاری مستقل

در مدل آبشاری مستقل[۱۱۲] [۱۶]، یک گره منبع یا گروهی از گره‌ها شروع به آلوده کردن می‌کنند. زمانی که یک گره آلوده شد، در فرایند تکرار بعدی، سعی دارد تا کلیه نودهای مجاور خود با توجه به احتمال تأثیر ارتباط آن‌ها آلوده سازد و این نودها هیچ یک از گره‌های مجاور خود را در فرایندهای بعدیِ تکرار، آلوده نخواهند کرد. اگر در یک فرایند تکرار، گره‌ای وجود نداشته باشد که جدیداً آلوده شده باشد، فرایند انتشار پایان خواهد یافت. تعداد کلی گره‌های آلوده شده دلالت بر قدرت پخش یا قدرت نفوذ نودهای منبع دارد. در مدل آبشاری مستقل، قدرت نفوذی یک گره به این صورت تعیین می‌شود؛ فرایند انتشار از گرهِ ذکر شده شروع می‌گردد و این روند چندین بار تکرار می‌شود. تعداد میانگین گره‌های آلوده شده در کلیه فرایندهای تکرار بر اساس قدرت پخش یا نفوذ گرهِ منبع سنجیده می‌شود. در صورتی که دو نوع از اطلاعات رقابتی در شبکه منتشر یابد، مدل آبشاری مستقل تبدیل به مدل آبشاری مستقل رقابتی[۱۱۳] می‌گردد. [۱۷]

ساختار شبکه، نقش مهمی در انتشار اطلاعات ایفا می‌کند. ساکسنا و همکاران [۱۸] مدل آبشاری مستقل را توسعه و یک مدل انتشار پنج سطحی را پیشنهاد دادند که بر اساس دو ویژگیِ میان-مقیاس در شبکه می‌باشند؛ (i) ساختار همانندی و (ii) ساختار مرکز-پیرامونی. هر گره، یا متعلق به هسته است یا متعلق به پیرامون شبکه. گره‌های پیرامونی نیز در گروه‌هایی سازمان‌دهی می‌شوند [۹]. در شبکه، یال‌ها بر اساس نوع گره منبع و گره هدف، به پنج دسته تقسیم می‌گردند. احتمال نفوذ هر یال به دسته‌ی آن بستگی دارد. این احتمالات به صورت $P_{cc} > P_{cp} > P_{pp_0} > P_{pp_1} > P_{pc}$ منظم می‌شوند و در اینجا c نشان دهنده‌ی گرهِ هسته و P نشان دهنده‌ی گره پیرامون می‌باشد، هم‌چنین pp_0 نشان می‌دهد که هم هدف و هم منبع از یک گروه هستند، و pp_1 نشان می‌دهد که منبع و هدف از گروه‌های مختلفی می‌باشند. این مدل پیشنهادی با استفاده از مجموع داده توییتری هیگز بوزون[۱۱۴] [۲۰] تأیید شده است و نقش ترتیبِ سلسله مراتبیِ جامعه را در روند به اشتراک گذاری اطلاعات توصیف می‌نماید [۲۱].

۱۶.۲.۲ مدل خطی آستانه‌ای[۱۱۵]

[۱۱۲] Independent Cascade Model (ICM)

[۱۱۳] Competitive Independent Cascade Model (CICM)

[۱۱۴] Higgs-Boson Twitter

[۱۱۵] Linear Threshold Model (LTM)

در جهان واقعیت، مشاهده شده است که اگر تعداد دوستان یک کاربر که اطلاعات را باور می‌کنند بالاتر از یک ارزش مشخص آستانه‌ای باشد، کاربر آن اطلاعات را می‌پذیرد یا با آن موافقت می‌کند. ارزش آستانه‌ای ممکن است برای کاربران مختلف متفاوت باشد. در مدل خطی- آستانه‌ای [۲۲]، اطلاعاتِ در حال انتشار از گروهی از گره‌های منبع آغاز می‌گردد. در هر فرایند تکرار، در صورتی که تعداد نودهای مجاوری که در آن‌ها نفوذ شده است، بیشتر از ارزش آستانه‌ای باشد، گره‌های جدیدی تحت نفوذ قرار می‌گیرند. روند آلوده شدن در صورتی که گره جدیدی در یک فرایند تکرار تحت نفوذ قرار نگیرد، متوقف می‌شود. زمانی که ارزش آستانه‌ای برای هر گره یکسان باشد، این مدل با نام مدل tipping[۱۱۶] نیز شناخته می‌شود [۲۳].

هم‌چنین مدل خطی آستانه‌ای به مدلی توسعه یافته است که مدل انتشار رقابتی اطلاعات نام دارد. در مدل آستانه‌ای خطی رقابتی، انواع مختلفی از اطلاعات هم‌زمان منتشر می‌یابند. در صورتی که اثر اطلاعات محاسبه شده که بر اساس گرهِ مجاور است بیشتر از ارزش آستانه‌ای باشد، کاربر اطلاعات را باور می‌کند. یانگ و همکاران[۱۱۷] [۲۴] مدل خطی آستانه‌ای رقابتی را برای مدل سازی انتشار رقابتی اطلاعات در شبکه‌های هدایت شده توسعه و مدل خطی آستانه‌ای را همراه با یک مدل حالت گذار تک سویه[۱۱۸] پیشنهاد دادند. زمانی که اطلاعات از موضوعات چندگانه به صورت پی در پی در شبکه چیده می‌شوند، فام و همکاران [۲۵] مدل خطی آستانه‌ای را اصلاح کردند که نام آن، مدل چندگانه‌ی به کارگیری خطی آستانه‌ای بوده است.

۱۶.۲.۳ بخش بندی مدل‌های انتشار (اطلاعات)

بخش بندی مدل‌ها در مدل سازی ریاضیِ انتشار آلودگی استفاده می‌شود. [۲۶] در این نوع مدل‌ها، کاربران به بخش‌هایی تقسیم می‌شوند و کاربرانی که متعلق به یک بخش هستند از قانون یکسانی پیروی می‌کنند. مدل آسیب پذیر- آلوده شده- بهبود یافته[۱۱۹] ساده‌ترین مدل بخش‌بندی است که در آن یک گره می‌تواند در هر یک از این سه حالت احتمالی وجود داشته باشد: (i) (آسیب پذیر)، (ii) (آلوده شده)، (iii) (بهبود یافته). برای شروع مدل، قرار است کلیه گره‌ها در حالت آسیب‌پذیر باشند. انتشار آلودگی یا نفوذ از طریق آلوده کردن یک گره (یا گروهی از گره‌ها) آغاز می‌گردد و وضعیت آن‌ها قرار است به وضعیت آلوده شده برسد. یک گرهِ آلوده شده/ تحت نفوذ واقع شده به نام u بر هر یک از نودهای مجاور آسیب پذیر خود همراه با احتمال نفوذ (λ) تأثیر می‌گذارد و در صورتی که گره آلوده شود، وضعیت آن‌ها را به وضعیت آلوده شده می‌رساند. احتمال نفوذ در مدل آسیب پذیر- آلوده شده- بهبود یافته نیز اشاره به احتمال آلودگی دارد. به محض اینکه گره u بر کلیه نودهای مجاور خود نفوذ کند، وضعیت این گره با احتمال μ به وضعیت بهبود یافته می‌رسد. یک گره

---

[۱۱۶] تئوری مشهور رفتارهای اجتماعی که دارای کاربردهای فراوان است.

[۱۱۷] Yang et al.

[۱۱۸] LT۱DT

[۱۱۹] SIR

بهبود یافته در طول روند انتشار وضعیت خود را مجدداً تغییر نخواهد داد. روند انتشار نفوذ زمانی که هیچ گره جدیدی در شبکه آلوده نشده باشد، متوقف می‌گردد.

دیگر گونه‌های مدل بخش‌بندی، مدل‌های آسیب پذیر، آلوده شده[۱۲۰] [۲۷] و مدل آسیب پذیر، آلوده شده، آسیب پذیر[۱۲۱] [۲۸] می‌باشد که اغلب جهت شبیه‌سازی روندِ انتشار در جایی که یک گره بتواند در هر یک از این دو وضعیت باشد، مورد استفاده قرار می‌گیرند. ژائو و همکاران[۱۲۲] [۲۹] رفتار بی توجهانه‌ی کاربران را در جایی که آن‌ها هر یک از این سه وضعیت را داشتند، در نظر گرفتند و اینگونه مدل آسیب پذیر- آلوده شده- بهبود یافته را توسعه دادند: سرکوب‌گر، گسترشگر و ناآموخته. آن‌ها تحلیل‌های شمارشی از مطالعه‌ی تأثیر پارامترهای مختلف بر روی انتشار اطلاعات نادرست انجام دادند. آن‌ها مدل آسیب پذیر- آلوده شده- هایبرنیتر- حذف شده[۱۲۳] را پیشنهاد دادند [۳۰]، جایی که گره‌ها نیز می‌توانند هایبرنیتر باشند. شیونگ و همکاران [۳۱] یک مدل آسیب پذیر، متصل، آلوده شده و تحرک ناپذیر[۱۲۴] را پیشنهاد دادند و نشان دادند که چگالی مبنی بر درجه‌ی گره‌های تحت تأثیر قرار گرفته به طور یکنواخت همراه با درجه‌ی آن‌ها افزایش می‌یابد. نِکووی و همکاران[۱۲۵] [۳۲] مدل SIR را با مدل ماکی-تامسون[۱۲۶] [۳۳] ترکیب کردند تا به مطالعه‌ی انتشار شایعه در خصوص نوع متفاوتی از شبکه‌ها مانند شبکه‌های تصادفی و انواع مختلفی از شبکه‌های بی مقیاس[۱۲۷] بپردازند. جین و همکاران[۱۲۸] [۳۴] در مورد شایعه‌ها و اخباری که در توییتر با استفاده از مدل آسیب پذیر، تحت تأثیر قرار گرفته، آلوده شده، شک‌گرای[۱۲۹] منتشر می‌یابند و دارای چهار وضعیت کاربر هستند، مطالعاتی انجام دادند [۳۵].

تامبوسیو و همکاران [۳۶] مدل SIS را در جایی که کاربرانِ آلوده شده به دو بخش با عنوان‌های راستی‌آزمایان و باورکنندگان تقسیم شدند، توسعه دادند. مدل پیشنهادی بر اساس راستی‌آزمایی است که می‌تواند طبق احتمالات مختلف معین گردد. مؤلفان برای این احتمال، آستانه‌ای را ارائه داده اند که می‌تواند به درک تعداد راستی‌آزمایانی که برای حذف اخبار جعلی از سیستم کافی هستند، کمک کند. آن‌ها به علاوه مدل پیشنهادی را بر روی شبکه‌های بی‌مقیاس، تصادفی و شبکه‌های جهان واقعی با استفاده از پارامترهای مختلف مدل تأیید کردند.

---

[۱۲۰] Susceptible, Infected (SI)

[۱۲۱] Susceptible, Infected, Susceptible (SIS)

[۱۲۲] Zhao et al.

[۱۲۳] Susceptible-Infected-Hibernator-Removed (SIHR)

[۱۲۴] Susceptible, Contacted, Infected, and Refractory (SCIR)

[۱۲۵] Nekovee et al.

[۱۲۶] Maki–Thompson (MK)

[۱۲۷] scale-free networks

[۱۲۸] Jin et al.

[۱۲۹] Susceptible, Exposed, Infected, Skeptic (SEIZ)

۱۶.۲.۴ مدل شکل‌گیری افکار[۱۳۰]

مدل‌های مبتنی بر شکل‌گیری افکار به منظور مدل‌سازی انتشار افکار در خصوص شبکه‌های اجتماعی آنلاین طراحی شده‌اند [۳۷]. در مدل شکل‌گیری افکار، افکار فعلی یک گره به افکار پیشین آن و افکار نودهای مجاور آن بستگی دارد. اِوانز و فنگ[۱۳۱] [۳۸] مدل شکل‌گیری افکار را پیشنهاد دادند که ساختار شبکه را هنگام محاسبه‌ی افکار کاربر در نظرمی-گیرد. در مدل پیشنهادی، افکار کاربر به قدرت افکار وی، سطح توافق کاربر با نودهای مجاور آن و درجه‌ی وزنی آن بستگی دارد. انواع مختلفی از مدل‌های شکل‌گیری افکار وجود دارد که شامل [۳۹-۴۳] می‌شوند.

هم‌چنین مدل‌های شکل‌گیری به مدل انتشارِ دو فکر رقابتی مانند اطلاعات نادرست و عقاید متقابل توسعه یافته‌اند. در [۱۳]، مؤلفان مدل شکل‌گیری افکار را برای کاهش اطلاعات نادرست پیشنهاد دادند، جایی که عقیده‌ی یک کاربر طبق عقیده‌ی مجاوران آن تعیین می‌گردد. یک کاربر می‌تواند در هر یک از این سه حالت باشد: باورکردن اطلاعات نادرست، باورکردن پیام متقابل یا خنثی. به منظور محاسبه‌ی عقیده‌ی گرهِ u، در ابتدا ما گره‌های مجاور آن را که اطلاعات نادرست و پیام‌های متقابل آن را باور کرده است، شناسایی می‌کنیم، و آن‌ها را به ترتیب در دو مجموعه به نام‌های مجاوران منفی و مثبت قرار می‌دهیم. سپس، ما نفوذِ هر دو نوع از گره‌های مجاور را بر روی گره u با استفاده از فرمول پیشنهادی محاسبه می‌کنیم. در صورتی که نفوذ اطلاعات نادرست بیشتر از نفوذ پیام‌های متقابل (که بر اساس افکار مجاوران آن محاسبه شده است) و آستانه باشد، کاربر پیام‌های متقابل را باور می‌کند. درغیر اینصورت، کاربر خنثی باقی می‌ماند. این مدل بر روی شبکه توییتری برای انتشار شایعه تأیید شده است و دقت آن در مدل سازی افکار بهتر از مدل آبشاری مستقل و مدل کلاسیک شکل‌گیری افکار می‌باشد.

ما به صورت مختصر در خصوص کلیه‌ی مدل‌های انتشار صحبت کرده‌ایم، که عمدتاً برای مدل سازی انتشار اطلاعات غلط و کاهش آن کاربرد دارد. در بخش بعدی، ما به بررسی تکنیک‌های کاهش اخبار جعلی می‌پردازیم.

۱۶.۳ کاهش اخبار جعلی

کاهش اخبار جعلی امری بسیار چالش برانگیز برای محققان و پژوهشگران است. پیشینه‌ی تحقیق در تکنیک‌های کاهش به چهار دسته‌ی اصلی زیر تقسیم بندی می‌شوند:

۱. **مسدود سازی نفوذ:** هدف این رویکرد، شناسایی مجموعه‌ی حداقلی از کاربرانی است که ایمن سازی آن‌ها درصدر انتشار اطلاعات نادرست در شبکه را به حداقل می‌رساند. به منظور انتخاب مجموعه‌ای بهینه از کاربرانی

[۱۳۰] Opinion Formation Model

[۱۳۱] Evans and Feng

که پارامترهای مختلف دریافت کرده‌اند، مانند مجموعه‌ای از نودهای منبع که آغازگر انتشار (اطلاعات) هستند، گره‌های هدف که باید نجات یابند و هم‌چنین مهلت زمانیِ شایعه (بعد از آن زمان، شایعه مؤثر نخواهد بود، به عنوان مثال، شایعه‌های مربوط به انتخابات تنها قبل از انتخابات مؤثر هستند) و غیره، ما به بحث در مورد چندین روش پیشنهادی خواهیم پرداخت.

۲. **پویش حقیقت:** رویکرد دیگر برای مقابله با اثر معکوس اخبار جعلی این است که کاربران را نسبت به اطلاعات درست آگاه کنیم. مطالعات پژوهشی نشان می‌دهند، در صورتی که کاربران هم با اطلاعات جعلی و هم اطلاعات درست مواجه شوند، تصمیم می‌گیرند که اطلاعات درست را باور کنند و میزان به اشتراک گذاری اطلاعات جعلی را کاهش دهند [۴۴،۴۵]. در این بخش، ما به تکنیک‌های تقریبی، حریصانه و ابتکاری محورِ پویش حقیقت می‌پردازیم.

۳. **ابزار کاهش:** پژوهشگران ابزارهایی را طراحی کرده‌اند که کاربر را متوجهِ اعتبار اطلاعات می‌کند تا میزان انتشار اطلاعات جعلی را به حداقل برساند. هدف این ابزارها، کاهش جریان اطلاعات جعلی در سراسر شبکه می‌باشد.

٤. **چندسویگی:** در این بخش، ما به مطالعاتی می‌پردازیم که مشمول هیچ کدام از دسته‌بندی های بالا نمی‌شوند. این بخش عمدتاً شامل پژوهش‌های اجتماعی است که برای رسیدن به پاسخ سوالات زیر انجام گرفته‌اند: چرا یک کاربر اطلاعات نادرست را به اشتراک می‌گذارد، چه شرایط و محیطی باعث می‌شود که کاربر آن را منتشر کند و اینکه چگونه می‌توانیم به مردم این آگاهی را بدهیم که اطلاعات جعلی را به اشتراک نگذارند.

۱۶.۳.۱ مسدودسازی نفوذ

حداکثرسازی نفوذ، مسئله‌ای است که در علوم شبکه در مورد آن به خوبی مطالعه شده است بر شناسایی حداقل مجموعِ سازوارگرهای اولیه جهت به حداکثر رساندن انتشار نفوذ در یک شبکه‌ی معین تمرکز دارد [۲۲،۴۶]. با این حال، در خصوص انتشار اخبار جعلی، هدف ما یافتن حداقل مجموعه‌ای از کاربرانی است که ایمن‌سازی آن‌ها انتشار اخبار جعلی را به حداقل می‌رساند. این امر به مسئله‌ی به حداقل رساندن نفوذ یا مسدودسازی نفوذ اشاره دارد.

۱۶.۳.۱.۱ فرمول سازی مسئله

تصور کنید که گراف$G(V,E)$ یک گراف معین است ، M بیانگر مجموعه‌ای از گره‌هاست که شروع به انتشار اطلاعات غلط می‌نمایند و K تعداد گره‌های مسدود شده/ ایمن شده می‌باشد. در برخی از آثار پژوهشی، هنگامی که هدف آن‌ها معکوس کردن اثر با استفاده از بودجه معین K است، K به بودجه یا هزینه نیز اشاره دارد و برای مسدودسازی یک گره در شبکه، بودجه‌ی ثابتی موجود است. در صورتیکه مجموعه $M(M \subset V)$ به انتشار اطلاعات غلط بپردازد، $\pi G(V,E)$ (M) نشان

دهنده‌ی تعداد گره‌های تحت تأثیر قرار گرفته در گراف معین G(V,E) خواهد بود. در زمینه‌ی حداقل سازی نفوذ، گراف معین G(V,E)، مجموع M و محدودیت بودجه‌ی K، شناسایی زیرمجموعه T از نودهایی با اندازه‌ی K از V-M به نحوی که πG(V,E)(M) به حداقل برسد، هدف ما می باشد.

اولین رویکرد عملی و شهودیِ راه حل، روش حریصانه است. در شیوه‌ی حریصانه، ما گره‌ای را انتخاب می‌کنیم که نفوذ را در شبکه به حداقل برساند و مکرراً گره‌هایی را اضافه کند که بعداً نفوذ را تا زمانی که تعداد مورد نیاز از گره‌ها انتخاب شده‌اند، به حداقل برساند. این روش در الگوریتم ۱ توضیح داده شده است.

یک رویکرد معروف دیگر برای حل چنین مشکلاتی، استفاده از روش‌های ابتکاری است. در پیشینه‌ی تحقیق، چندین مقیاس مرکزیت وجود دارد که جهت شناسایی کاربران بانفوذ در یک شبکه معین مورد استفاده قرار می‌گیرند. به محض اینکه کاربران بانفوذ شناخته می‌شوند، کاربرانی با بالاترین میزان K که دارای بیشترین ارزش مرکزیت هستند، برای ایمن‌سازی انتخاب می‌شوند. آموروسو و همکاران[۱۳۲] [۱۲] یک رویکرد ابتکاری دو مرحله‌ای را ارائه دادند که به این صورت عمل می‌کند: (i) در ابتدا مجموعه‌ای از کاربرانی را شناسایی می‌کند که به احتمال زیاد دارای کاربران منبع انتشار هستند، سپس (ii) تعداد محدودی هشدار دهنده قرار می‌دهد تا عمل انتشار اطلاعات نادرست را در شبکه مسدود کند. اولین گام از شناسایی منبع در یک روش ابتکاری، ان-پی سخت است، و گام دوم از جایگزینی هشدار دهنده، #پی –کامل است. علاوه بر این، کائو و همکاران [۴۸] در مورد به حداکثررساندن مسدودسازی اطلاعات برای تحدید شایعه (CIBM) همراه با بودجه‌ی معین در محیط تجارت الکترونیک مطالعه کردند. مسئله‌ی پیشنهادی، ان پی-سخت همراه با ویژگی‌های زیرمدولی و یکنواخت است. پیشنهاد مؤلفان، جامعه‌ای است که الگوریتم‌هایی را تقسیم می‌کند، این الگوریتم‌ها مجموعه‌ای از گره‌هایی را انتخاب می‌کنند که هم نودهای منفی و هم غیرفعال را مسدود کرده و محدودیت شایعه‌ها را با استفاده از ساختار جامعه بهینه سازی می‌نماید.

الگوریتم ۱: مسدودسازی نفوذ - رویکرد حریصانه (G(V, E), M, k)

داده‌های ورودی: گراف معین G(V, E)

M مجموعه گره (رأس)هایی است که شروع به انتشار اطلاعات نادرست می‌نماید.

مقدار K تعداد گره (رأس‌های) انتخاب شده می‌باشد.

---

[۱۳۲] Amoruso et al.

داده‌های خروجی: T مجموعه ای از گره های انتخاب شده با اندازه k می‌باشد.

$T = \phi$;
for *i in range*$(1, k)$ **do**
  for *each node v in* $\{V - M - T\}$ **do**
    $s_v = \pi_{G(V',E)}(M)$, where $V' = V - T - \{v\}$;
  end
  $T = T \cup argmin_{v \in \{V-M-T\}}\{s_v\}$;
end
Return T;

در شبکه‌های جهان واقعی، کاربران در جوامعی سازماندهی می‌شوند. از جامعه‌ی اطلاعات برای ارائه‌ی روش‌های مؤثرتر جهت محدودسازی شایعه استفاده می‌گردد، زیرا می‌تواند اطلاعاتی را در مورد مجموعه‌ای از گره‌هایی که قرار است تحت تأثیر قرار گیرند، ارائه دهد. در کنار این، در زندگی واقعی، ممکن است شرایطی وجود داشته باشد که هدف آن حفظ جامعه‌ی تعین شده از اثرات منفی شایعه می‌باشد. ژنگ[۱۳۳] و پان[۱۳۴] [۴۹] روش حریصانه‌ی مبتنی بر حداقل پوشش رأسی را جهت محدود کردن شایعه هنگامی که از یک جامعه‌ی معین بیرون آمده باشد، پیشنهاد کردند (CR). هدفِ راه حل‌های پیشنهادی، یافتن زیرمجموعه‌ای از کاربرانی است که باید مسدود شوند تا اثر شایعه در CR به حداقل برسد و هم‌چنین تعداد کلی گره‌های تحت‌تأثیر واقع شده بیشتر از محدودیت معینی نیست. هم‌چنین وو و همکاران[۱۳۵] [۵۰] یک استراتژی پویای مسدود کننده با محوریت جامعه را برای کنترل انتشار شایعه پیشنهاد دادند. این روش پیشنهادی در ابتدا نفوذ هر گره را در جامعه‌ی خود و هم‌چنین در کل شبکه محاسبه کرده و سپس این اطلاعات را جهت انتخاب گره‌هایی با بالاترین میزان K برای مسدودکردن اثر شایعه ادغام می‌کند.

علاوه بر این، فِن و همکاران[۱۳۶] [۵۱] به تحلیل مسئله‌ی حداقل هزینه برای مسدودسازی شایعه پرداختند، که در آن شایعه برخاسته از یک جامعه‌ی معینِ CR بوده و کمترین زیرمجموعه‌ی گره‌ها مسدود شده‌اند تا تعداد افراد آلوده شده در جوامع مجاور CR به حداقل برسد. نویسندگان مجموعه‌ای از رئوس را با نام مجموعه‌ی انتهایی پل[۱۳۷] ایجاد کردند که شامل رئوسی می‌شد که در آن هر رأس حداقل دارای یک گره مجاور در جامعه CR است و می‌توان با آغازگران شایعه به آن دست یافت. آن‌ها نشان دادند که مسئله‌ی LCRB-P که در آن لازم است بخشی از گره‌های انتهایی پل محافظت شوند، زیرمدولی هستند و یک راه حل مبتنی بر رویکرد حریصانه را همراه با تخمین $(1 - 1/e)$- ارائه کردند. فام و همکاران

[۱۳۳] Zheng
[۱۳۴] Pan
[۱۳۵] Wu et al.
[۱۳۶] Fan et al.
[۱۳۷] bridge end set

[۵۲] به مطالعه‌ی مسئله‌ی مسدودسازی اطلاعات نادرست هدفمند[۱۳۸] پرداختند، به طوری که هدف آن‌ها شناسایی مجموعه‌ای بهینه از کاربرانی است که ایمن سازی آن‌ها انتشار اخبار جعلی را از طریق ارزش آستانه‌ای مشخصِ γ به حداقل می‌رساند. آن‌ها اثبات کردند که مسئله، همان مسئله‌ی #پی-سخت از مدل LTM است. علاوه بر این، مؤلفان برای ارائه‌ی راه حلی در بازه‌ی نسبتِ $\ln(\gamma/\epsilon)+1$ ، از راه حل بهینه، یک روش حریصانه را معرفی کردند.

وانگ و همکاران[۱۳۹] [۵۳] در مورد به حداقل رساندن شایعه‌ها در جایی که هر کاربر یک آستانه‌ی زمان تحمل دارد و کارایی شبکه در صورتی که زمان ایمن‌سازی کاربر فراتر از آستانه‌ی تحمل آن باشد، کاهش می‌یابد (زمان ایمن‌سازی، طول زمانی را مشخص می‌کند که کاربر ایمن شده در نظر گرفته می‌شود). نفوذ با استفاده از مدل پویای انتشار آیزینگ[۱۴۰]، منتشر می‌شود و این مدل هم به محبوبیت در سطح جهانی توجه دارد و هم جذب عناوین خاص شایعه. مؤلفان بر اساس نظریه‌ی بقا و اصل حداکثر درست‌نمایی، یک راه حل مسدودسازی پویا و حریصانه را پیشنهاد دادند. یائو و همکاران[۱۴۱] [۵۴] در مورد به حداکثر رساندن نفوذ از چشم‌انداز مدل سازیِ موضوع مطالعاتی انجام دادند، جایی که احتمال نفوذ از یک کاربر به کاربر دیگر بستگی به موضوع دارد. اخباری که دارای اطلاعات نادرستِ آشکار است، می‌توانند موضوعات چندگانه داشته باشند و انتشار آن‌ها می تواند با استفاده از مدل چندگانه‌ی خطی آستانه‌ایِ موضوعی[۱۴۲] مدل سازی گردد [۲۵،۵۵]. مسئله‌ی انتخاب کاربران-K برای به حداقل رساندن اثر اطلاعات نادرست چند موضوعی، ان پی- سخت است. مؤلفان اثبات کرده‌اند که در اینجا تابع هدف، یکنواخت و زیرمدولی است. به علاوه، آن‌ها نشان دادند که یک روش تقریبی همراه با فاکتور تقریبی $(1-1/\sqrt{e})$ موفق‌تر از روش‌های ابتکاری عمل می‌کند.

یانگ و همکاران[۱۴۳] [۵۶] به حل دو تغییر مسدودسازی نفوذ پرداختند که به حداقل رساندن هزینه همراه با اختلال[۱۴۴] و به حداقل رساندن انتشار با هدف ضمانت شده[۱۴۵] با استفاده از برنامه‌ریزی خطیِ عدد صحیح[۱۴۶] نام دارد. در LMD، آن‌ها بر یافتن مجموعه‌ای از سازوارگرهای اولیه تمرکز می‌کنند که هزینه آن‌ها بیشتر از هزینه‌ی مشخص شده است، اما انتشار نفوذ (هزینه‌ی کلی گره‌های فعال) باید به حداقل برسد. آن‌ها برای LMD، روش‌های ابتکاری معرفی کردند که در آن، گره‌های K با حداقل رتبه صفحه یا درجه انتخاب می‌شوند و نشان می‌دهند که روش‌های پیشنهادی در شبکه‌های جهان واقعی بسیار خوب عمل می‌کند. در DMGT، مؤلفان تمرکز خود را بر یافتن کمترین مجموعه‌ای از گسترشگران

---

[۱۳۸] Targeted Misinformation Blocking (TMB)

[۱۳۹] Wang et al.

[۱۴۰] مدل ریاضی از فرومغناطیس در مکانیک آماری.

[۱۴۱] Yao et al.

[۱۴۲] Multiple Topics Linear Threshold model

[۱۴۳] Yang et al.

[۱۴۴] Loss Minimization with Disruption (LMD)

[۱۴۵] Diffusion Minimization with Guaranteed Target (DMGT)

[۱۴۶] Integer Linear Programming (ILP)

اولیه از مجموعه‌ی معین گره‌های ابتدایی قرار می‌دهند، به طوری که کلیه گره‌های هدف علیرغم به حداقل رسیدن عمل انتشار در شبکه، مورد نفوذ قرار می‌گیرند. نویسندگان یک راه حل حریصانه پیشنهاد دادند که در هر مرحله‌ی تکراری، یک گره بر اساس بیشترین بهره‌ی نهایی انتخاب می‌شود. با این حال، هدف مؤلفان در این مقاله، به حداقل رساندن نفوذ است اما این موضوع با سناریوهای انتشار اخبار جعلی در زندگی واقعی فرق دارد؛ سناریوهایی که در آن‌ها علیرغم تلاش ما برای مسدود کردن گره‌ها یا یال‌ها به منظور به حداقل رساندن اثر آن در شبکه، گره‌های آغازگر آلوده شده‌اند و شروع به انتشار اطلاعات جعلی در سراسر شبکه می‌کنند.

در مسائل مربوط به مسدودسازی یال، حداقل مجموعه‌ای از یال‌ها انتخاب می‌شوند که اطلاعات را جهت به حداقل رساندن فرایند انتشار اخبار نادرست گسترش نخواهند داد. از آنجایی که یال‌ها برای انتشار اطلاعات غیرفعال هستند، این مسائل به راه‌حل‌های مبتنی بر مسدودسازی یال یا حذف یال ارجاع داده می‌شوند. وانگ و همکاران [۵۷] در مورد به حداقل رساندن نفوذ هدف با استفاده از مسدودسازی یال مطالعه کردند و آن‌ها بر این نکته تمرکز کردند که مجموعه گره‌هایی را که عمدتاً کاربران هدف هستند، از اطلاعات نادرست حفظ کنند. آن‌ها نشان دادند که با داشتن بودجه‌ای محدود، مشکل همان ان پی-سخت است، به همین دلیل روش حریصانه‌ای را پیشنهاد دادند که دارای تقریب-$(1 - 1/e)$ است. با این حال، در شرایطی که محدودیت بودجه وجود نداشته باشد، مؤلفان یک راه حل بهینه ارائه دادند. هر دو راه‌حل در شبکه‌های اجتماعیِ جهان واقعی از نظر اثربخشی و بازدهی تأیید شده‌اند. هم‌چنین کیمورا و همکاران [۵۸] در مورد روش مسدودسازی لینک نیز برای به حداقل رساندن آلودگیِ متوسط و بدترین آلودگی مطالعه کرده‌اند. میانگین درجه‌ی آلودگیِ شبکه همان میانگین درجات نفوذ، و درجه بدترین آلودگی نیز بیشترین درجات نفوذِ کلیه گره‌ها در یک شبکه‌ی مشخص می‌باشد.

یان و همکاران [۵۹] اثبات کردند که به حداقل رساندن انتشار شایعه[۱۴۷] که یال‌ها را از یک یال مشخصی که قرار است انتشار شایعه را به حداقل برساند، حذف می‌کند، یک زیرمدولی نیست. مؤلفان برای تابع هدف RSM کران پایینی زیرمدولی[۱۴۸] و کران بالایی زیرمدولی[۱۴۹] ارائه دادند و یک راه حل ابتکاری برای تخمین عملکرد هدف مشخص طراحی نمودند. تونگ و همکاران[۱۵۰] [۶۰] روش NetMelt را ارائه کردند که یال‌های K را از شبکه حذف می‌کند تا اثر شایعه را به حداقل برساند. راه حل پیشنهادی، یالی که باید با استفاده از مقدار مشخصه‌ی ماتریسِ مجاورتِ گراف حذف شوند، را شناسایی می‌کند. دیگر مقالات مبتنی بر حذف لینک برای محدودسازی شایعه شامل [۶۱-۶۴] می‌شود.

[۱۴۷] RSM

[۱۴۸] submodular lower bound

[۱۴۹] submodular upper bound

[۱۵۰] Tong et al.

هه و همکاران[۱۵۱] [۶۵] با استفاده از ترکیب مسدودسازی گره و حذف یال به مطالعه‌ی مدل امنیت شبکه‌ی آمیخته‌ی تعمیم یافته[۱۵۲] پرداختند. همکاری مؤلفان شامل موارد زیر می‌شود: (۱) یک راه حل تقریب-$(d + 1)$ با درجه زمان چندجمله‌ای جهت شناسایی مجموعه‌ای بهینه در هنگام انتشار شایعه تا جهش- d ، (۲) مشتق گرفتن تقریب، هنگامی که ∞ برابر با d است، و (۳) تقریب ۳۲ با درجه زمان چندجمله ای در گراف‌های دوبخشی، هنگامی که d برابر با ۱ است. به علاوه، این مقاله دیگر نتایج گراف‌های منتظم و ساختار درختی را نیز دربردارد.

در شبکه‌های پویای جهان واقعی، ممکن است شرایط پیش بینی نشده‌ای نیز وجود داشته باشد. روش‌های مورد بحث که مجموعه‌ای از گره‌های K را شناسایی می‌کنند و این گره‌ها باید براساس اسنپ شات شبکه مسدود شوند، ممکن است برای شبکه‌های پویا عملی نباشد. شی و همکاران[۱۵۳] [۶۶] برای محدودسازی شایعات در شبکه‌های پویا، جایی که گره‌های K در هر فرایند تکرار انتخاب می‌شوند تا زمانی که بودجه به اتمام برسد، یک راه حل تطبیقی پیشنهاد کرده‌اند.

کلیه‌ی روش‌هایی که در بالا ذکر شد فرض را بر این قرار می‌دهند که کاربران اطلاعات نادرست را از مجاوران خود دریافت می‌کنند؛ با این حال، کاربران می‌توانند اطلاعات نادرست را در خود جستجو کنند. ژانگ و همکاران[۱۵۴] [۶۷] هنگام کنترل شایعه، این رفتار کنشگرایانه‌ی کاربران که مسئله‌ی مسدودسازی شایعه با محوریت جستجوی[۱۵۵] کاربر نامیده می‌شود، را نیز در نظر می‌گیرند. مؤلفان، انتشار نفوذ را با استفاده از یک گام تصادفی مدل سازی کردند و نشان دادند که BUK یک زیرمدولی می‌باشد. آن‌ها یک راه حل حریصانه پیشنهاد دادند که BUK را با تقریب -$(1 - 1/e)$ تقریب می‌زند.

جدول ۱۶.۱ اهمیت کار در این مسیر را با پارامترهای لحاظ شده ذکر می‌کند. در جدول ۱۶.۱، "پیچیدگی" بیانگر پیچیدگی مسئله‌ی مطالعه شده است، K (یک رقم ثابت در گره‌های مسدود شده) نشان می‌دهد که تعداد گره‌هایی که قرار است مسدود شوند، ثابت هستند یا در بودجه‌ی مشخص محدود هستند، "گره‌های هدف" نشان می‌دهند که هدف این کار حفظ مجموعه‌ی ثابتی از نودها در برابر آلوده شدن به اطلاعات جعلی می‌باشد، "ضریب رفع آلودگی" نشان می‌دهد که هدف نویسنده این است که $\theta\%$ از نودها نباید آلوده شوند، "محدودیت زمانی" نشان می‌دهد که در برخی موارد محدودیت زمانی قائل شده است، مانند وقتی که آلودگی بعد از زمان t یا زمان لاگین کاربر شناسایی می‌گردد، غیره، "مهلت" اشاره به این دارد که شایعه بعد از مدتی از بین خواهد رفت، "مدل انتشار" بیانگر مدل پایه‌ی انتشار است که در این مقاله مورد استفاده قرار گرفته است، از قبیل مدل آبشاری مستقل (ICM)، مدل خطی آستانه‌ای (LTM)، یا مدل‌های

[۱۵۱] He et al.

[۱۵۲] Mixed Generalized Network Security (MGNS)

[۱۵۳] Shi et al.

[۱۵۴] Zhang et al.

[۱۵۵] Browsing-based rUmor blocK (BUK)

چندبخشی، "روش‌های مرجع" نشان دهنده‌ی روش‌هایی است که قبلا مؤلفان مقاله‌ی خود را با آن مقایسه می‌کردند. نکته: اگر در مقاله‌ای، مؤلفان هیچ کدام از محدودیت‌ها را مطرح نکرده باشند، جدول خالی باقی می‌ماند.

۱۶.۳.۲ پویش حقیقت

در روش‌های جلوگیری از نفوذ، گره‌ها (رأس) و یال از شیوع بیشتر اخبار جعلی در جهت کاهش تأثیرات این اخبار جعلی مسدود می‌شوند. هرچند که تأثیر معکوس ممکن است کاهش یابد.

جدول ۱۶,۱: روش‌های جلوگیری از نفوذ

منابع	پیچیدگی	K	گره های (رأس) هدف	نسبت آلودگی زدایی	محدودیت زمانی	مهلت	مدل ارائه شده	رهیافت آنی خط پایه
کائو و همکاران	ان –پی سخت	*					LTM مدل	حداکثر پراکندگی تعداد رئوس جهت داری که از یک گره خارج می شوند ، تعداد رئوس جهت داری که از یک گره خارج می شوند به طور تصادفی
فان و همکاران	LCR-B: زیرواحد LCRB- D: ان پی		*	*			ICM مدل	حداکثر درجه نزدیکی یا مجاورت
وانگ و همکاران					*		ICM	حریصانه
یائو و همکاران		*					ICMامدل	Top -ICM:آگاهی میانی Top-ICM: آگاهی از تعداد رئوس جهت داری که از یک گره خارج می شوند ، تعداد رئوس جهت داری که از یک گره خارج می شوند
ژنگ و پن				*			ICMامدل	حداکثر درجه، میانی مرکزی
آموروسو و همکاران	ان - پی سخت	*	*				ICMامدل	
ژانگ و همکاران	– #Pپیچیدگی						ICMامدل	
وانگ و همکاران		*					ICM مدل	درجه خروجی، میان مرکزی، رتبه صفحه
یان و همکاران	تابع هدف، زیرمدولی نیست	*					ICM مدل	بیشترین درجه خروجی، میان مرکزی،
تان و همکاران							SIR مدل	مستعد بیشترین مجاورت
گای و همکاران	ان –پی سخت	*					ICM مدل	
ژانگ و همکاران	ان – پی سخت	*					گام تصادفی	بیشترین میزان

جدول ۱۶,۱: روش‌های جلوگیری از نفوذ

منابع	پیچیدگی	K	گره های (رأس) هدف	نسبت آلودگی زدایی	محدودیت زمانی	مهلت	مدل ارائه شده	رهیافت آنی خط پایه
کوئی و همکاران	ان – پی سخت	*			*		ICM مدل	DAVA
وانگ و همکاران	یال مسدود کننده: اِن پی سخت (بودجه محدود)	*					LTMمدل	تصادفی ، بیشترین یال وزن
کیمورا و همکاران	یال مسدود کننده	*					ICM مدل	لینک میان – مرکزی، لینک درجه‌خروجی
یانو و همکاران	یال مسدود کننده	*					ICM مدل	لینک میان – مرکزی، لینک درجه‌خروجی
یان و همکاران	یال مسدود کننده : زیر مدولی نیست	*					ICMمدل	روش تراوش باند حذف K-edgeT، درجه خروجی، رتبه صفحه، تصادفی
نقندی و مدال	یال مسدود کننده	*					SIR مدل	میان- مرکزی، حریصانه، تصادفی
وانگ و همکاران	یال مسدود کننده: ان پی سخت (بودجه محدود)	*	*				LTMمدل	تصادفی بالاترین یال وزن
ووو و همکاران	O(۱+ TMM) ، #۱ از یال ها Tنمایانگر تکرار M بیانگر جامعه ، هر جامعه‌ای Nتعداد گره دارد.	*			*		ICMمدل	انتشار یا تکثیر نرمال، حریصانه، تصادفی، مسدود کننده آبشاری مثبت، مسدود کننده‌ی مدل ICM دینامیکی
ژانگ و همکاران	اِن – پی سخت	*					SIR مدل	تصادفی،درجه‌بندی،رتبه صفحه، نت شیلد
	پی سخت زیر مدل LTM و ان پی سخت زیر مدل ICM	*		*			مدل LTM و مدل ICM	حریصانه،درجه‌بندی،رتبه صفحه، DAVA

ارائه اطلاعات صحیح و درست به کاربران در راستای کمک و یاری فهم اخبار می‌باشد و هم‌چنین رویکرد های مختلفی باعث شکل‌گیری نظرات بی‌طرفانه در مورد عناوین خبری ارائه شده، می‌شود. هم‌چنین تصمیمات کاربران مبنی بر اشتراک-گذاری اطلاعات را تحت تأثیر قرار خواهد داد و کاربران به اشتراک‌گذاری اطلاعات صحیح و درست و نه شایعات، بیشتر مشتاق خواهند بود و این رویکرد در زندگی واقعی عملی‌تر است و البته موجب تقلیل تأثیر اخبار جعلی را بر روی شبکه خواهد شد.

تحقیقات گرت [۸۴] نشان‌می‌دهد، کاربران خواندن مقالات خبری با دیدگاه و رویکردهای متفاوت و آنچه آنان بدان باور دارند را رها نمی‌کنند. هم‌چنین مطالعات نشان می‌دهد کاربران زمان زیادی را به جست‌وجوی دیدگاه‌ها در مقالات خبری صرف می‌نمایند. تحقیقات وندر لیندن و همکاران [۸۵] نشان می‌دهد که نگرش عمومی را می‌توان با ارائه حقایق در برابر اطلاعات نادرست در مورد تغییرات آب و هوا تلقیح کرد. یافته‌های کوک و همکاران [۸۶] بیانگر آن است که تلقیح پیام‌هایی که هم استدلال‌های ناقص در اطلاعات غلط و هم اجماع علمی در مورد عناوین خبری را شرح می‌دهند در خنثی کردن اثرات نامطلوب اطلاعات نادرست مؤثرتر هستند. تاناکا و همکاران [۴۴]. در این مورد تحقیقات بیشتری صورت گرفته و به یافته‌هایی مبنی بر این، درصورتی‌که افراد قبل از مواجه با شایعات با اخبار صحیح و درست مواجه شوند این امر موجب کاهش چشمگیر انتشار شایعات خواهد شد. اوزتورک و همکاران. [۴۵] این مشکل به شیوه‌ای واقعی مورد مطالعه قرار گرفت در این تحقیق بر روی نیت افراد در به اشتراک‌گذاری مجدد و یا بازتوییت شایعه‌ها و مقابله با شایعه‌ها بیشتر از تعداد افرادی که به بازنشر شایعه پرداخته بوند، متمرکز شد. آن‌ها عنوان کردن نشان دادن شایعه و غیر شایعه، به طور هم‌زمان، به کاهش میزان شایعه‌پراکنی کمک خواهد کرد و هم‌چنین باید گفت اجرای این امر امکان‌پذیرتر است. همه این مطالعات از این مفهوم حمایت می‌کنند که روش‌های مبارزه با حقیقت برای کنترل اخبار جعلی در زندگی واقعی بهتر عمل می‌کنند.

در ابتدا، به منظور کاهش گره‌های آلوده در موارد سرایت آلودگی و یا ویروس،کمپین‌های مخالفین یا مکانیزم‌های دفاعی مورد مطالعه قرار گرفت. نیکول و لیلجنستام [۸۷] مکانیزم‌های دفاعی اکتیو را در برابر ویروس‌های اینترنتی مورد مطالعه قرار دادند و نتایج حاصل نشان داد و نشان داد که با شروع دفاع با تعداد کافی گره (تبدیل به ضد کرم رایانه‌ای)، هر بخش مورد نظر از گره‌ها می‌تواند از آلوده شدن به کرم رایانه‌ای محافظت شود. تحقیقات در موقعیت دو بازاری که در حال رقابت بر سر محصولات و افزایش نفوذ خود بودند چندین استراتژی رقابتی را مورد مطالعه قرار داده‌اند. [۸۸،۸۹]. کوستکا و همکاران [۹۰] مطالعات نشان داد، انتخاب گره آغازین در یک بازی(تمرین) دو نفره که سطح شایعات در آن به میزان حداکثری است، برای هر دو شرکت کننده در حالت اِن- پی[۱۵۶] کامل است. آن‌ها بعدها عنوان کردند یافتن یک راه‌حل

[۱۵۶] در نظریه پیچیدگی محاسباتی NP یکی از بنیادی‌ترین کلاس‌ها است. NP مخفف عبارت « «Non-Deterministic Polynomialاست که به زمان اجرای آن اشاره دارد.

NP مجموعهٔ کلیه مسائل تصمیم‌گیری است که پیدا کردن جواب بله برای آن‌ها شامل اثبات ساده‌ای است که جواب حقیقتاً باید بله باشد.

تقریبی برای اولین بازیکن نیز در حالت ان-پی کامل است علاوه براین بیان داشت، اولین بازیکن بودن همیشه سودمند نیست زیرا بازیکن دوم می‌تواند گره‌های بیشتری را نسبت به بازیکن اول آلوده کند، حتی اگر از استراتژی‌های بهینه پیروی کنند. رویکردهای مشابه ممکن است به منظور مقابله با اثرات منفی اشاعه‌ی اخبار دروغین در شبکه مورد استفاده قرار گیرند که بعدها در مورد آن بیشتر گفتگو خواهیم کرد.

۱۶.۳.۲.۱ فرمول‌سازی مساله

تصور کنید، گرافی با فرمول G(V,E) در دست داریم، M بیانگر مجموعه‌ای از رأس‌ها (گره) که شروع به انتشار اطلاعات غلط می‌نمایند و K مقدار تعداد رأس‌هایی است که باید انتخاب شوند. در صورتیکه مجموعه M($M \subset V$) به انتشار اطلاعات غلط بپردازد و مجموعه T ($T \subset V - M$) به انتشار اخبار درست بپردازد، $\pi G(V,E)$ (M,T) بیانگر تعداد رأس‌های آسیب دیده از اطلاعات نادرست در نمودار G(V,E) می‌باشد. در اغلب شیوه‌های پویش حقیقت، نمودار G(V,E)، تعیین M و K، هدف ما انتخاب زیر مجموعه T شامل از گره‌ها با اندازه k از مجموعه V-M انتخاب شود به نحوی که πG(V,E)(M, T) به حداقل برسد. روش حریصانه (مترجم: روشی در الگوریتم) در الگوریتم شماره ۲ شرح داده شده است. در شیوه‌ی حریصانه، ما رأسی را برای شروع یک پویش حقیقت انتخاب کرده تا بدین منظور تأثیر اطلاعات نادرست در شبکه را به حداقل رساند، به طور مکرر اضافه کردن رأس‌ها را ادامه می‌دهیم و تا زمانی که اثرات منفی به حداقل برسد تعداد مورد نیاز گره انتخاب شود.

الگوریتم ۲: پویش حقیقت - رویکرد حریصانه (G(V, E), M, k)

داده‌های ورودی: G(V, E) گراف

M مجموعه گره (رأس)هایی است که شروع به انتشار اطلاعات نادرست می‌نماید.

مقدار K تعداد گره (رأس‌های انتخاب شده می‌باشد.

داده‌های خروجی: T مجموعه ای از گره های انتخاب شده با اندازه k برای پویش حقیقت می‌باشد.

$T = \phi$;
for i *in* $range(1, k)$ **do**
  for *each node* v *in* $\{V - M - T\}$ **do**
    $s_v = \pi_{G(V,E)}(M, T \cup v)$;
  end
  $T = T \cup argmin_{v \in \{V-M-T\}}\{s_v\}$;
end
Return T;

محدودیت‌های متعددی در حین مطالعه تکنیک‌های پویش حقیقت توسط آثار مختلف وجود دارند مثلاً بعضی از فعالیت‌ها مبنی بر این تصور است که شایعه پایان و مهلتی دارد و بعد از آن زمان مهلت داده شده آن شایعه دیگر اثرگذار نیست. برخی آثار دیگر را در نظر می‌گیرند، گروهی از کاربران وجود دارند که ممکن است مبارزان آینده نگر حقیقت باشند و k کاربر باید از آن گروه انتخاب شوند. یکی دیگر از تفاوت‌های جالب در جایی است که هدف ما مجموعه معینی از کاربران هدف هستند که باید تا پایان مهلت باید از اطلاعات واقعی آگاه باشند .خلاصه‌ای از محدودیت‌ها برای کارهای مختلف در جدول ۱۶.۲ آورده شده است. در ادامه برخی از این آثار را به تفصیل مورد بررسی قرار می دهیم.

بوداک و همکاران[۹۱] نشان داد که انتخاب حداقل گروهی از کاربران به منظور انتشار اطلاعات "خوب" برای به حداقل رساندن تأثیر اطلاعات "بد" ان_ پی سخت است. نویسندگان ثابت کردند که مشکل زیر مدول است و یک راه حل مبتنی بر رویکرد حریصانه ارائه کردند. آن‌ها هم‌چنین آبشار مستقل چند کمپینی را پیشنهاد کردند، مدل (MCICM) که انتشار دو آبشار (خوب و بد) را مدل‌سازی می‌کند و تصور می‌کند هنگامی‌که اطلاعات خوب و بد سعی می‌کنند تا به طور همزمان روی کاربر تأثیر بگذارند، کاربر اطلاعات خوب را می‌پذیرد. وقتی‌که که یک گره هر یک از این موارد را می پذیرد اطلاعات، وضعیت خود را در آینده هرگز تغییر نخواهد داد. نتایج تجربی نشان داد که در بیشتر موارد، انتخاب گره‌ها بر اساس درجه مرکزیت، یعنی یک معیار مرکزیت با محاسبه آسان، با توجه به رویکرد حریصانه عملکرد خوبی دارد.

تانگ و همکاران[۹۲] تصدیق کردند، مشکل جلوگیری از اطلاعات غلط را نمی‌توان در یک ضریب $\Omega(2^{log^{1-\epsilon}n^4})$ در یک پیچیدگی زمانی تخمین زد مگر به صورت $NP \subseteq DTIME(n^{polylogn})$ دربیاید. آن‌ها اولویت‌های آبشاری چندگانه به نام اولویت آبشاری همگن (هم اطلاعات غلط و هم آبشارهای مثبت دارای اولویت یکسان)، اولویت آبشار غالب M در این آبشار(اطلاعات نادرست در اولویت هستند)، اولویت آبشار غالب P (در این آبشار اطلاعات مثبت دارای اولویت است) پیشنهاد کردند و آن‌ها را بر اساس شبکه جهانی واقعی مورد مطالعه قرار دادند.

جدول ۱۶,۲ روش‌های پویش حقیقت

منابع	پیچیدگی	K	رویکرد کاهش یافته	گره های (رأس) هدف	نسبت آلودگی زدایی	محدودیت زمانی	مهلت	مدل ارائه شده	رهیافت آنی خط پایه
نگویان وهمکاران [۹۳]					*	*		LTM مدل ICM مدل	
یوداک و همکاران [۹۱]	ان –پی سخت		*			*		ICM مدل	درجه، سرایت آغازین، رایت وسیع
تانگ و همکاران [۹۲]		*	*					ICM مدل	وزن زیاد، مجاورت، تصادفی
هی و همکاران [۱۰۲]	ان –پی سخت	*						LTM	درجه،تصادفی، اکتشافی
یانگ و همکاران [۱۰۳]		*						ICMمدل	درجه حداکثر، کاهش درجه[۱۰۴]، میان مرکزی،مجاورت منطقه‌ای
هی و همکاران [۱۰۵]	APXسخت					*		ICM /LTM	پی آی دی اس تصادفی
لیتو و همکاران	ان-پی کامل	*						دینامیک -LTM	درجه، حریصانه
پنگ و پن [۱۷]	ان-پی سخت	*						ICM مدل	درجه، مجاورت،تصادفی ، حریصانه
ژو و همکاران[۱۰۷]	ان-پی سخت	*						ICM مدل	درجه، مجاورت،تصادفی ، حریصانه
ال وی و همکاران [۱۰۸]	ان-پی سخت	*						ICM مدل	درجه، مجاورت،تصادفی
سیمسون و همکاران [۱۰۹]		*						ICM مدل	
فان و همکاران [۱۱۰]	ان – پی سخت	*					*	ICM مدل LTM	هیچ اقدامی را نشان نمی‌دهد

جدول ۱۶,۲ روش‌های پویش حقیقت

منابع	پیچیدگی	K	رویکرد کاهش یافته	گره های (رأس) هدف	نسبت آلودگی زدایی	محدودیت زمانی	مهلت	مدل ارائه شده	رهیافت آنی خط پایه
وو وهانگ [۱۱۱]	ان – پی سخت	*				*		LTM مدل	درجه حداکثر و تصادفی
ژانگ و همکاران [۱۱۲]	اِن پی کامل	*						LTMمدل	تصادفی ، بیشترین درجه، بیشترین حریصانه، حریصانه متوسط
ژانگ و همکاران[۱۱۳]		*					*	LTMمدل	تصادفی ، بیشترین درجه، بیشترین حریصانه، حریصانه متوسط
یانگ و همکاران [۲۴]	یال مسدود کننده		*					LTMIDT	تصادفی، بیشترین درجه
حُسنی و همکاران [۱۱۴]	ان –پی سخت	*		*				ICMمدل	حداکثر،تصادفی
تانگ وهمکاران [۱۱۵]	ان –پی سخت	*						ICMمدل	مجاورت، حریصانه، تصادفی
سانگ و همکاران	ان –پی سخت	*					*	ICMمدل	رتبه صفحه،LSMI، [۱۱۶]، بزرگترین میزان آلودگی یا سرایت
ووو و همکاران [۱۱۷]			*			*		SCTIR	حریصانه، تصادفی، حداکثر درجه، نزدیکترین نقطه
فان و همکاران [۱۱۸]	اِن – پی سخت	*						تراست نامتقارن	تصادفی، حداکثر درجه
ساکسن و همکاران [۱۳]			*	*			*	مدل OF	تصادفی، بیشترین میزان درجه خروجی، بیشترین درجه خروجی، کمترین فاصله، TIB[۹۵] TMB [۸۳]
لین و دای [۱۱۹]		*						مدل ICM	CMIA-Oپنگ و پن [۱۷]، حریصانه،HD، مجاورت و تصادفی

آن‌ها همچنین یک الگوریتم تقریبی را پیشنهاد کردند با استفاده از تکنیک کران پایین بالا و انجام آزمایش‌ها برای نشان دادن نسبت تقریبی تقریباً ثابت آن‌ها روشی پیشنهاد دادند که می‌توان با ارائه یک تقریب بهتر برای کران بالا و پایین آن روش را ارتقا بخشید.

نگویان و همکاران[۹۳] مسئله βTI مورد مطالعه قرار گرفت. در این مطالعه، هدف، شناسایی یک مجموعه حداقلی ازS رئوس تأثیرگذار که اخبار صحیح را منتشر می کنند، می‌باشد. به طوری که نسبت ضد آلودگی مورد انتظار بعد از زمانT ، βباشد، البته با توجه به اینکه کاربران مجموعه I اطلاعات جعلی را منتشر می کنند. نویسندگان روش Greedy Viral Stopper (GVS) را پیشنهاد کردند، که در آن گره‌ها به طور مکرر برای انتشار اطلاعات خوب انتخاب می‌شوند تا تعداد کل گره‌های ایمن سازی به حداکثر برسد. آن‌ها همچنین یک راه حل پیشنهادی کران بالایی با توجه به حل بهینه ریاضی ارائه کردند. مؤلفان فرض کردند که اطلاعات خوب و اطلاعات غلط هر دو با استفاده از مدل انتشار یکسان با احتمال سرایت یکسان منتشر می‌شوند. آن‌ها نیز تصور کردند هنگامی که یک گره هر دو اطلاعات را دریافت کند، گره متکی بر اطلاعات خوب است و آن اخبارخوب را بیشتر منتشر می‌کند. رویکرد پیشنهادی برای به‌دست آوردن یک مجموعه راه‌حل بهتر برای شبکه‌های ساختار یافته جامعه اصلاح می‌شود جایی که گره‌ها از هر جامعه به طور حریصانه برای ایمن‌سازی انتخاب می‌شوند تا زمانی که کسر β گره‌ها از جامعه ایمن‌سازی شود، بنابراین کسر β در کل شبکه به دست می‌آید. راه‌حل پیشنهادی جامعه - محور بهتر و سریعتر عمل می‌کند زیرا یک کاربر به دلیل تأثیر هموفیلی نسبت به سایر جوامع، احتمال بیشتری برای آلوده‌کردن کاربران جامعه خود دارد.[۹۴]

در سناریوهای دنیای واقعی، تأثیر اخبار جعلی پس از مدتی از بین می رود. سونگ و همکاران [۹۵] این را به عنوان یک پارامتر در نظر گرفت و راه‌حلی را برای شناسایی پویش حقیقت مناسب با توجه به شایعه پیشنهاد کرد. راه‌حل پیشنهادی دارای دو مرحله است: (i) مجموعه گره هایی که ممکن است توسط شایعه آلوده شوند را پیدا کنید و میزان تهدید را برآورد کنید (تعداد گره‌هایی که ممکن است توسط گره داده شده تحت تأثیر قرار گیرند) سطح هر گره متعلق به مجموعه است و(ii) استفاده از درختان Weighted Reverse Reachable (WRR) برای انتخاب حریصانه k پویش حقیقت می‌باشد که بیشتر گره‌ها را از اطلاعات نادرست در مهلت مقرر ایمن می‌سازد.

در کارهای فوق الذکر، فرض بر این است که مبارزان پویش حقیقت، اطلاعات واقعی را تبلیغ خواهند کرد. با این حال، در زندگی واقعی، یک کاربر ممکن است مایل به راه‌اندازی یک پویش حقیقت یا ارسال یک پیام متقابل در پروفایل نباشد. رویکرد واقع بینانه‌تر زمانی خواهد بود که تمایل کاربر را نسبت به شروع آن پویش بدانیم. ساکسنا و همکاران [۱۳] این مشکل واقع بینانه را در نظر گرفتند، جایی که مجموعه‌ای از کاربران کاهش‌دهنده احتمالی، ممکن است علاقه‌مند به کاهش اطلاعات نادرست باشند یا می‌توانند توسط مقامات برای انتشار برخی پیام‌های خاص کنترل شوند، از قبل در دسترس باشند. هدف نویسندگان این است که مجموعه‌ای از مبارزان پویش حقیقت k را از مجموعه کاهش‌دهنده‌هایی که نسبت به شروع‌کننده شایعات و زمان به پایان رسیدن شایعات آگاهی دارند را شناسایی کنند. راه‌حل پیشنهادی، ابتدا

(مترجم: مکانیسم کاهش قدرت) هر یک از مبارزان حقیقت مورد انتظار را با توجه به مجموعه اولیه آغازگران شایعه برآورد کنند و سپس گره‌های top-k دارای حداکثر کاهش قدرت هستند را انتخاب کنند. اقدام پیشنهادی از روش‌های پیشرفته و روش‌های اکتشافی به خوبی مطالعه شده بهتر عمل می‌کند. ساکسنا و همکاران [۹۶] هم‌چنین روشی به نام k-TruthScore (مقدار K امتیاز حقیقت) پیشنهاد کرد که مقدار K مبارزان حقیقت را برای به‌حداقل رساندن تأثیر معکوس اطلاعات نادرست را به هنگام سوگیری‌های شدید کاربران، شناسایی می‌کند. خو و همکاران [۹۷] روش جدیدی را برای مبارزه با انتشار اخبار جعلی ارائه کردند که در آن کاربران طعمه را در شبکه‌های اجتماعی آنلاین (OSN) مستقر می‌کنند و آن‌ها را با برخی از کاربران فعال منتخب شبکه مرتبط می‌سازند. برای استقرار کاربر طعمه، نویسندگان ابتدا حداقل مجموعه‌ای از گره‌ها را انتخاب کرده و آن‌ها را نظارت می‌کنند. پس از انتخاب گره ها، ارائه‌دهنده‌ی سرویس آنلاین شبکه اجتماعی با این کاربران مشورت می‌کند. هر یک از کاربران توافق شده به دو گره طعمه متصل خواهند شد. این کاربران طعمه به طور منظم انتشار اطلاعات در شبکه را مشاهده می‌کنند تا کرم یا انتشار اطلاعات نادرست را شناسایی کنند. مؤلفان از این گره‌های فریبنده برای کنترل انتشار کرم استفاده نمی‌کنند، هرچند این کار برای رسیدگی به چنین مشکلاتی قابل تعمیم است. آن‌ها نشان دادند که مشکل بکارگیری گره‌های طعمه به گونه‌ای است که همه گره‌ها در داخل R-hops از کاربران طعمه پوشانده شده‌اند، معادل مجموعه مشکل صعودی تعمیم یافته است که در این حالت NP کامل است. با این حال، می‌توان الگوریتم‌های حریصانه یا اکتشافی را برای شناسایی حداقل مجموعه‌ای از کاربران برای ارتباط با کاربران فریب پیشنهاد کرد.

وایلدر و وروبیچیک [۹۸] از یک رویکرد تئوری بازی برای مقابله با اخبار جعلی که از کانال‌های مختلف در طول انتخابات منتشر می‌شوند، استفاده کردند. آن‌ها استراتژی‌های مهاجم-مدافع را با استفاده از بازی حاصل جمع - صفر مورد مطالعه قرار دادند، جایی که مهاجم قصد دارد با انتشار اخبار جعلی انتخابات را براندازی کند و مدافع برای به حداقل رساندن تأثیر مهاجم هدف قرار می دهد. این مشکل با استفاده از دو ساختار جمعیتی مختلف مورد مطالعه قرار می‌گیرد، (i) جمعیت‌های متمایز که در آن رأی‌دهندگان توسط کانال‌ها تقسیم می‌شوند و (ii) جمعیت‌های غیر متمایز که در آن رای‌دهندگان می‌توانند از طریق کانال‌های متعدد دسترسی داشته باشند. مؤلفان نشان می‌دهند که در مورد جمعیت‌های غیر متمایز، محاسبه یک استراتژی ترکیبی مدافع بهینه APX-hard است. نتایج تجربی نشان می‌دهد که استراتژی‌های مدافع پیشنهادی بازدهی تقریباً بهینه ارائه می‌کنند و نشان می‌دهند که می‌توان با استفاده متوسط از منابع محدود و اطلاعات کافی در مورد ترجیحات یا تمایلات رأی دهندگان، از انتخابات دفاع کرد. تانیمی اس و همکاران [۹۹] هم‌چنین یک رویکرد نظری بازی را مورد بحث قرار می دهد جایی که بازیکن اول برای به حداقل رساندن تأثیر اطلاعات نادرست و بازیکن دوم برای به حداکثر رساندن آن هدف قرار می دهد. مؤلفان دو روش (i) ریاضی و (ii) اکتشافی حریصانه را برای حل به حداقل رساندن اطلاعات نادرست از دیدگاه بازیکن اول پیشنهاد کردند.

فرجتبار و همکاران [۱۰۰] یک روش کاهش مبتنی بر فرآیند نقطه‌ای را با استفاده از چارچوب یادگیری تقویتی ارائه کرد. هدف راه حل ارائه شده، بهینه سازی اقدامات در راستای حداکثر پاداش کل طبق محدودیت های بودجه داده شده می‌باشد. راه حل ارائه شده به موقع در شبکه توییتر برای مبارزه با کمپین اخبار جعلی اعمال شد و با هدف تحقیقاتی آغاز شد و نتایج امیدوارکننده ای به همراه داشت. یان و همکاران [۱۰۱] انتشار بدافزار در شبکه‌های اجتماعی آنلاین را برای تجزیه و تحلیل اثرات آلودگی اولیه، ساختارهای اجتماعی، احتمال کلیک کاربر و الگوهای فعالیت مشاهده کرد. آن‌ها بعدها طرح‌های دفاعی کاربر محور و سرورمحور و اثربخشی آن‌ها در برابر انتشار بدافزار را که می‌توان در برنامه‌های کاربردی واقعی اعمال کرد، مورد مطالعه قرار دادند. در جدول ۱۶.۲، اقدام مختلف را بر اساس پارامترهای مختلفی که در نظر گرفته اند، مقایسه می کنیم.

روش‌های مورد بحث می‌توانند با آگاه کردن کاربران از اخبار صحیح، انتشار اطلاعات نادرست را کاهش دهند، اما امکان استفاده از آن‌ها برای شبکه‌های اجتماعی آنلاین هنوز یک سوال تحقیقاتی باز است. در شبکه اجتماعی آنلاین، چالش متقاعد کردن کاربران برای پویش حقیقت است و این امر در آینده خط جدیدی از تحقیقات را باز می‌کند که فراخوانی برای انجام یک تجزیه و تحلیل عمیق از ویژگی‌ها و رفتار کاربران شبکه اجتماعی آنلاین می‌باشد.

۱۶.۳.۳ ابزارهای کاهش‌دهنده

مطالعات مبتنی بر روان‌شناسی نشان می‌دهد که کاربران تمایل دارند تا اطلاعات درست را در مواجه با اطلاعات درست و جعلی باور کنند. ابزارهای کاهشی متعددی به منظور براورد و نمایش اعتبار اخبار در جهت کمک به تصمیم‌گیری به اشتراک‌گذاری اطلاعات بیشتر از سوی کاربران وجود دارد. پارک وهمکاران[۱۲۰] سرویسی به نام NewsCube[۱۵۷] را طراحی کردند که دیدگاه‌های مختلفی را در مورد یک رویداد خبری در اختیار کاربران قرار می‌دهد. بدین طریق، کاربران قادرند مقالات خبری را از رویکردها مختلف را درک و مطالعه کنند و بر مبنای عقیده خود یک نظر بی‌طرفانه نتیجه‌گیری کنند. به منظور ارزیابی *NewsCube* های مفید، مؤلف بر روی ۳۳ مشارکت‌کننده، شامل دانش‌آموز، محقق، کادر اداری (۱۶ مرد و ۱۷ زن)، آزمایش کنترل شده انجام داد. ۷۲ درصد از مشارکت‌کنندگان اظهار داشتند به خواندن مقالات چندگانه در سرویس *NewsCube* علاقه‌مند هستند؛ هرچند ۷۰ درصد از کاربران واقعا آن مقالات را مطالعه کردند. به جز ۳ نفر از شرکت‌کنندگان، کلیه‌ی شرکت کنندگان اذعان داشتند مطالعه و مقایسه مقالات چندگانه حائز اهمیت است. ۱۶ نفر از شرکت‌کنندگان بیان داشتند سرویس *NewsCube* به آن‌ها کمک کرده تا نقطه نظرات بی‌طرفانه‌ای نسبت به رویدادهای خبری اتخاذ نمایند، دو نفر نظر منفی و باقی افراد به جهت عدم زمان کافی در جلسه آزمایش برای ارزیابی مؤثر محصول، نظر خنثی داشتند. بنابراین، می‌توان اثر بخشی چنین ابزارهایی را در کاهش انتشار اخبار جعلی مشاهده کرد. انالز وهمکاران [۱۲۱] افزونه‌ی تحت عنوان "Dispute Finder" طراحی کردند، در یک برنامه افزودنی مرورگر برای اطلاع دادن به

[۱۵۷] یک سرویس خبری جدید اینترنتی که هدف آن کاهش اثر سوگیری‌های رسانه‌ای است.

کاربران در صورتی که اطلاعاتی که می‌خوانند، توسط منابع دیگر مورد بحث قرار گرفت و هم‌چنین لیستی از مقالات خبری را که دیدگاه‌های دیگر را پشتیبانی می‌کنند، نمایش می دهد. هم‌چنین کاربران قادرند تا ادعاهای چالش برانگیزی را به پایگاه داده اضافه کنند، که بعدها به کاربران با اطلاعات بیشتر کمک خواهد کرد. هم‌چنین نویسندگان با شرکت‌کنندگانی که در طول جلسه‌ی آزمایش از ابزارها استفاده نمودند، مصاحبه کردند و اکثر شرکت‌کنندگان نسبت به این ابزارها نظر مثبت داشتند. حسن و همکاران [۱۲۲] سیستمی را با عنوان "Fact Watcher " پیشنهاد کردند، این سیستم قادر است تا با ارائه حقایقی که به عنوان سرنخ در جریان اخبار عمل می‌کنند به کاربران کمک کند. هم‌چنین سیستم سرویس‌های اضافی هم‌چون factranking، جستجوی حقایق مبتنی بر کلمه کلیدی، سرویس ترجمه fact-to-statement، ارائه می‌دهند. سرویس ClaimBuster [۱۲۳] یک سیستم حقیقت‌یابی (fact-checking) است این سیستم دارای مؤلفه-های مختلف مبتنی بر گام‌های متفاوت حقیقت یابی (fact-checking) از جمله؛ (i) Claim Monitor این سرویس به جمع‌آوری داده‌ها از سایت‌های مختلف و شبکه‌های اجتماعی می‌پردازد، (ii) سرویس Claim Matcher حقایق مشابهی را بر روی وب‌سایت‌های مختلف حقیقت‌یابی می‌یابد. (iii) Claim Checker نتایج تحقیقات در مورد موضوع از سطح وب جمع‌آوری می‌شود و (iv) گزاش سرویس fact-checker گزارش نهایی را برای نمایش دادن به کاربران آماده می‌سازد. راتکیویچ و همکاران [۱۲۴] سرویس وب را تحت‌عنوان Truthy کردند در این سرویس پیام‌های توییتری مربوط به انتخابات سیاسی آمریکا را جمع‌آوری و تبلیغات فریب‌دهنده، تکنیک لکه‌دار کردن و سایر اطلاعات نادرست و رفتارهای آشکار بر اساس مؤلفه‌های توییتری از قبیل؛ هشتگ‌ها، بازتوییت‌ها و یادآوری‌ها را شناسایی می‌نماید.

فیگویرا و اولیویرا [۱۲۵] به طور خلاصه در مورد سایر الگوریتم‌های fact-checking مانند افزونه مرورگر "FiB: Stop living a lie"[۱۵۸] ، رویکردهای تشخیص اخبار جعلی فیس‌بوک وغیره به بحث‌وگفتگو پرداخته‌اند. کوها افزونه‌ای را تحت عنوان "Related Fact Checks " طراحی کرد، این افزونه حقایق مرتبط با آیتم‌های خبری جست‌وجو شده برای کاربران را به نمایش می‌گذارد [۱۲۶]. دی آلفرو و همکاران [۱۲۷] ابزاری را برای توییتر با عنوان Truth Value (ارزش حقیقت) طراحی کردند[۱۵۹] جایی‌که به هر پست خبری یک امتیاز شهرت اختصاص داده می‌شود. هم‌چنین تسهیل رأی گیری را برای کاربران امکان‌پذیر می‌سازد بدین طریق آن‌ها می‌توانند تا به مقالات خبری به عنوان قابل اعتماد یا جعلی/ گمراه‌کننده رأی دهند، و وزن آراء به شهرت بستگی دارد. هم‌چنین نویسنده ربات توییتری برای پاسخگویی به سؤالات کاربران و یک نشانک مرورگر برای نمایش سریع امتیازهای سایت‌ها در حالی که کاربران در مرورگر به آن‌ها دسترسی دارند. پاولسکا و همکاران [۱۲۸]. به مطالعه‌ی دقت و کارایی سازمان‌های حقیقت‌یاب اروپایی پرداختند؛ نتایج حاصل‌شده نشان می‌دهد آن‌ها همچنان به بهبود جریان کاری و عملکرد نیازمندند، لذا ان‌ها قادرند تا توصیه‌های بهتری را به کاربران

[۱۵۸] گروهی از هزاره ها الگوریتمی را برای شناسایی داده های جعلی در فیس بوک می‌باشد.

[۱۵۹] این پروژه در سایت https://truthvalue.org در دسترس می‌باشد.

ارائه دهند. ابزارهای متعددی به منظور شناسایی اخبار جعلی از قبیل[۱۴] FakeNewsTracker ، PolitiFact[۱۵]، [۱۲۹] FactCheck.org ، snopes.com [۱۳۰] وجود دارد .

وب‌سایت‌های آنلاین از جمع‌سپاری ها به منظورکنترل انتشار اخبار جعلی از طریق کاربران برای نمایان ساختن استوری که حاوی اخبار اشتباه یا جعلی هستند، استفاده می‌نماید. هنگامی که آن استوری توسط افراد کافی نشانه‌گذاری شد، برای بررسی واقعیت به شخص ثالث مورد اعتماد ارسال می‌شود و اگر حاوی اخبار جعلی باشد، استوری به‌عنوان مسئله‌ی مورد بحث علامت‌گذاری می‌شود. هریک از تماس‌های راستی‌آزمایی هزینه‌بردار هستند، بنابراین لازم است تا یک مبادله بهینه بین تعداد کل flagها و تماس‌های راستی‌آزمایی درنظر گرفته شود. در صورتی‌که در آغاز تعداد فلگ‌ها بالا باشد و بعداً مشخص شود که اطلاعات نادرست بوده، آن زمان، اخبار تأثیر خوبی بر روی جمعیت داشته است. به طور مشابه در صورتی‌که از همان آغاز تعداد flagها کم باشد، استوری ممکن است اشتباه باشد یا بالعکس اشتباه نباشد زیرا کاربران هر یک تعصبات در هر موضوعی تعصبات خود را دارند و ممکن است بر اساس تعصبات خود آن را جعلی تلقی نماید، این در حالیست اگر اطلاعات واقعا جعلی نباشند در این‌صورت هزینه راستی‌آزمایی به هدررفته است. کیم و همکاران [۱۳۱] الگوریتمی را تحت عنوان "CURB" ارائه کردند تا بدین طریق توازن را بین تعدادی flag و راستی‌آزمایی ایجاد نماید و اینکه کدام استوری باید به طور بهینه به منظور راستی‌آزمایی ارسال شود تا بدین جهت انتشار اطلاعات نادرست به حداقل برسد. احتمالاً مقالات پیشنهادی بعدها با توجه به مؤلفه‌های واقع‌گرایانه‌ای هم‌چون؛ کاربران به همان میزان در نشانه‌گذاری اطلاعات نادرست خوب نیستند، در نظر گرفتن اعتبار کاربر هنگام محاسبه وزن کل پرچم‌گذاری شده پُست، نفوذ و اثر یک کاربر با توجه به این نکته؛ چه تعداد کاربر ممکن است تحت تأثیر اطلاعات به اشتراک گذاری شده از سوی همان کاربر باشد، ارتقا یابد. هم‌چنین فیسبوک در جایی‌که کاربران قادر باشند مقالات خبری را به بحث‌وگفتگو بگذارند، مؤلفه‌ای را ارائه کرده است و این در صورتی است که آن‌ها تصور کنند اطلاعات صحیح نیستند و محققان تأثیر نشانه‌گذاری را بر روی انتشار هر چه بیشتر مقاله بررسی کرده باشند.

سایتی به نام Newport Buzz مقاله‌ای را با عنوان اینکه چگونه هزاران ایرلندی برای اسارت به ایالات متحده آورده شدند، منتشر کرد و این ماجرا توسط فیس‌بوک به عنوان موضوع مورد بحث نشانه‌گذای شد. [۱۳۲]. هرچند سر دبیر سایت گزارش کرد، ترافیک سایت بعد از هشدار فیس‌بوک افزایش یافت است هرچند نمایانگر این مسئله نبود که افراد زیادی به درستی آن اعتقاد دارند زیرا ممکن است کاربران از روی کنجکاوی از آن مقاله بازدید کرده باشند. نمونه‌ی دیگر مربوط RealNewsRightNews می‌باشد، صاحب آن متذکر شد، قرارگرفتن نشان مناقشه بر روی یکی از مقالات تأثیری بر روی ترافیک ورودی ندارد. این مثال نشان می‌دهد ما به تحصیلات و سیستم‌های آگاهی دهنده‌ی بهتری برای کاربران احتیاج داریم، در نتیجه ترافیک سایت چنین اخباری جعلی کاهش می‌یابد و مطمئناً بر روی رفتارهای کاربرانی که اطلاعات نادرست را منتشر می‌کنند، تأثیر خواهد داشت.

وول و لی [۱۳۳] وجود برخی از کاربران گاردین که در توییتر فعال هستند را نشان دادند این کاربران، افرادی هستند که با reply (پاسخ دادن) و یا با ارائه URL ها، اخبار صحیح و درست را در بحثو گفتگوهای به اشتراک می‌گذارند. افراد فعال در گاردین که به‌طور مستقیم به پاسخگویی و یا تصحیح اطلاعات مشغول هستند، direct guardians می‌نامند، افرادی که به کار بازتوییت اطلاعات تصحیح‌شده می‌پردازند را secondary guardians می‌نامند. این کاربران گاردین می‌توانند از ابزارهای راستی‌آزمایی و انتشار پست‌های صحیح در شبکه‌های اجتماعی اقدام نمایند. هم‌چنین نویسندگان مدل پیشنهادی URL راستی‌آزمایی را ارائه دادند این مدل برای کمک به نگهبانان در راستای اجرای هرچه راحت‌تر فرآیند راستی‌آزمایی می‌باشد.

درحالی‌که هدف اصلی طراحی این ابزارها باید به‌گونه‌ای باشد تا به راحتی در دسترس کاربران قرار گیرند بدین وسیله آن‌ها قادر خواهند بود تا به طور دائم اطلاعات جدیدی را اضافه نمایند و اخبار به‌روز شوند. این ابزارها به کاربران کمک خواهد کرد تا رویکرد بی‌طرفانه‌ای را اتخاذ نمایند، اما اگر به طور هم‌زمان کاربران کورکورانه به این ابزارها اعتماد نمایند. نگوین و همکاران[۱۳۴] بر روی این‌که چگونگه پیش‌بینی درست و یا نادرست ابزارهای حقیقت‌یاب بر روی دقت تصمیم-گیری کاربران تأثیرگذار است، مطالعاتی را انجام داده‌اند. آن‌ها نشان دادند، هر چند کاربران دقت خود را از طریق تعامل با این ابزارها افزایش دهند با این حال پیش‌بینی نادرست ابزارها بر روی دقت قضاوت انسان‌ها تأثیر منفی دارد. بنابراین لازم است تا این ابزارها احتمال درستی اعتبار یک مقاله خبری را نشان دهد. لیز در مورد طراحی ابزارهای راستی‌آزمایی بر مبنای دیدگاه بازیابی اطلاعات (IR) ، الزامات، چالش ها و نتایج مورد انتظار بحثوگفتگو کرده است. جزئیات بیشتر در [۱۳۵] موجود می‌باشد.

۱۶.۳.۴ مطالعات علمی اجتماعی

در این بخش، ما تمام مطالعاتی که به‌منظور درک و فهم جنبه‌های متعددی از انتشار اخبار جعلی انجام شده را بررسی می‌نماییم جنبه‌هایی از قبیل آن‌که چرا افراد اقدام به نشر اخبار جعلی می‌نمایند، چگونه این کار ار انجام می‌دهند، چگونه می‌توان نگرش آن‌ها را در این خصوص تغییر داد. روان‌شناسان به‌منظور درک نحوه‌ی ایمن‌سازی افراد در برابر انتشار اخبار جعلی مطالعاتی را انجام داده‌اند. روزنبیک و ون در لیندن [۱۳۶] بازی را به نام "بازی اخبار جعلی" طراحی کرده‌اند. در این بازی از کاربران خواسته‌شده تا مقالات خبری جعلی در خصوص مسائل مهم سیاسی ایجاد نماید. هر یک ازگروه کاربران می‌توانند مقاله‌ای جعلی بر اساس ۴ رویکرد متفاوت به نام (i) منکر یا انکار کننده در این رویکرد، انگیزه‌ی فرد نسبت به موضوع باید کوچک و بی‌اهمیت به‌نظر آید، (ii) رویکرد هشدار دهنده در این رویکرد تمرکز بر روی این مسئله که موضوع باید بزرگ و مهم به نظر برسد، (iii) تله کلیک، تمرکزش این است که مقاله باید تا حد امکان کلیک بخورد و (iv) نظریه پرداز توطئه، هدفش، بی‌اعتمادی نسبت به روایت‌های رسمی جاری و وادار کردن مخاطب، تا از او تبعیت کند. در این آزمایش، ساختاری به شرکت‌کنندگان که جمعاً تعداد ان‌ها ۹۵ نفر است، داده شد تا مقاله‌ای را از منظر نقشی که انتخاب کرده است ایجاد نماید. نویسندگان بعد از انجام بازی مشاهده کرد، کاربران کمتر تحت‌تأثیر اخبار جعلی قرار گرفته‌اند.

نویسندگان پیشنهاد کردند، آموزش اولیه رسانه‌ای ممکن است به افراد کمک کند تا نسبت به رویکردهای متفاوت اگاهی پیدا کنند و اینکه چگونه به‌وجود می‌آید بنابراین مردم می‌توانند در برابر خطر اطلاعات نادرست مقاومت کنند. جانگ و کیم [۱۳۷] انتشار اخبار جعلی را از منظر ادراک سوم شخص مطالعه کردند (TPP) و نشان دادند کاربرانی با ادراک بیشتر از شیوه‌های منسجم رسانه‌ای حمایت نمی‌کنند، اما آن‌ها حمایت قابل توجهی از روش‌های مداخله‌جویانه‌ی سواد رسانه‌ای نشان می‌دهند. این مطالعه می‌تواند به عنوان شیوه‌ی بنیادین در طراحی تکنیک‌های مداخله‌جویانه‌ی سوادرسانه‌ای در راستای آگاه سازی کاربران نسبت به استنباط‌های متفاوت سوم شخص، استفاده شود. پنی کوک و رند [۱۳۸] نشان دادند افرادی که قبلاً در معرض اخبار جعلی و واقعی بوده‌اند، می‌تواند اخبار جعلی را با دقت بالاتری متمایز کند و با قدرت تفکر تحلیلی کاربر در ارتباط است. کانوح [۱۳۹] به مطالعه‌ی تأثیر عادت‌های خوردن و آشامیدن را بر روی انتشار اخبار جعلی پرداخت و در مطالعات خود نشان داد، افرادی در موقعیت‌های خوردن وآشامیدن بیشتر اقدام به انتشار اخبار جعلی می‌نمایند. هم‌چنین از واقعیت انتشار گسترده اخبار جعلی در شبکه‌های اجتماعی حمایت می‌نماید، به این دلیل که مردم در اوقات فراغت خود از شبکه‌ها بازدید می‌کنند. غنایی و میجوا[۱٤۰] داده‌های توییتری مربوط به درمان بی‌اثر سرطان را جمع‌آوری کرد و ویژگی کاربرانی که دست به چنین اقدامی می‌زنند و در مقابل کاربرانی که واقعا به مسئله‌ی سرطان علاقه‌مند هستند را مورد مطالعه قرار دادند. یک طبقه‌بندی بر اساس آن می‌توان کاربرانی که اقدام به انتشار اطلاعات نادرست در مورد درمان‌های بی‌اثر می‌کنند را شناسایی کرد، ارائه دادند و بیش از ۹۰ درصد مواردی همچون مؤلفه‌هایی که کاربران استفاده می‌کنند، سبک نوشتن، تمایلات کاربران قابل شناسایی است. از این شیوه می‌توان در کشف چنین کاربرانی در شبکه‌های اجتماعی آنلاین، اقدامات بعدی ان‌ها در خصوص انتشار محتوی، استفاده نمود. انتشار اخبار جعلی ممکن است ازطریق تخصیص نمره‌ی اعتباردهی به کاربران کنترل شود، بدین ترتیب سایر کاربران با دقت بیشتری می‌توانند اعتبار پست‌های جدید به اشتراک گذاشته از سوی کاربرانی با اعتبار پایین را تشخیص دهند. بالمائو و همکاران [۱۴۱] با استفاده از تیمی متشکل از حقیقیت‌یاب‌ها و استفاده از این اطلاعات به منظور محاسبه‌ی میزان اعتماد کاربران بر اساس معیارهای اخبار منتشر شده از سوی آنان، اقدام به بررسی میزان اعتماد مؤلفه‌های خبری کردند، سپس بالمائو و همکارانش به جهت محاسبه صحت اخبار آینده و با تکیه بر اعتماد کاربران از مدل بیضوی استفاده می‌کنند. پروژه‌های آتی بر تکنیک‌های بهبود یافته در راستای کنترل انتشار اخبار جعلی از سوی کاربرانی با اعتبار پایین، متمرکز می‌باشد و این هم‌چنان یک تحقیق باز است و در آینده ممکن است به صورت عمیق‌تری مورد بررسی قرار گرفت. در شبکه‌های آنلاین اجتماعی افراد ترجیح می‌دهند تا با کاربرانی هم‌عقیده با خودشان در ارتباط باشند و این موجب می‌شود تا جوامعی در مورد موضوعی به نام اتاق‌های پژواک از همان عقیده مشترک پیروی کنند. نگوین و همکاران[۱۴۲] تکنیکی را به‌منظور مختل ساختن اتاق‌های پژواک ارائه کردند. در این شیوه‌های ارائه شده، مؤلف، ابتدا کاربران را به دو دسته بر مبنای نقطه نظر سیاسی، متشکل از دموکرات‌ها و جمهوری‌خواهان تقسیم می‌شوند. در صورت وجود هر گونه اختلافی درباره هر موضوع در گروه‌ها، پست محبوب یا محتوای محبوب در مورد موضوع مورد نظر از سوی یکی از گروه‌ها انتخاب شده و به گروه دیگر پیشنهاد می‌شود. بنابراین روش توصیه شده کابران را نسبت به نقطه‌نظرات مختلف آگاه خواهد ساخت و مانع از افزایش

اتاق‌های پژواک‌های رسانه‌ای می‌گردد. در جهان واقعی، یک شبکه به دو ساختار جامعه متفاوت بر مبنای موضوع‌های مختلف تقسیم می‌شود و ممکن است دو کاربر بر روی یک موضوع با یکدیگر موافق باشند اما در خصوص موضوعات دیگر نقطه‌نظر و عقیده کاملاً متفاوتی نسبت به یکدیگر داشته باشند. روش پیشنهاد شده بعدها می‌تواند به شبکه‌های جهان واقعی، با استفاده از ساختارهایی مبتنی بر موضوع بکار بسته‌شود، اما پیچیدگی و امکان‌سنجی چنین رویکردی هم‌چنان یک پرسش باز است. بر خلاف این تحقیق ،کالینی و همکاران [۱۴۳] از ان- گرم، فراوانی وزنی تی‌اف-آی‌دی‌اف [۱۴۴] به منظور استفاده از روش یادگیری نظارت شده برای کاربران خوشه‌ای در توییتر هم برای گروه دموکرات و جمهوری‌خواه بهره‌برده‌اند.

بلیر [۱۴۵] مطالعه‌ای را بین ۸ – ۹ ماه می در سال ۲۰۱۷ با ۲۹۹۴ شرکت‌کننده‌ی آموزش‌دیده از تُرک مکانیکی آمازون[۱۶۰] انجام داد، ۵۴ درصد از افراد را زنان، گروه سنی میانه بین ۲۵-۳۴ ، ۵۵ درصد دارنده ی مدرک لیسانس و بالاتر، ۳۲ درصد خود را جمهوری‌خواه یا جمهوری‌خواه میانه معرفی کردند و ۵۸ درصد، خود را دموکرات میانه یا دموکرات معرفی کردند. در این بررسی، نویسنده اخباری را با برچسب "نرخ کاذب" نشانه‌گذاری شده‌اند دقت درک کمتری نسبت به اخبار با برچسب "بحث‌برانگیز" داشت. با این حال، هیچ تأثیری بر روی دقت به‌دست آمده از اخبار بدون برچسب ندارد. تحقیق‌وبررسی در این حوزه بعدها ممکن است با استفاده از منابع مختلف سایت‌های حقیقت یابی (راستی‌آزمایی) گسترش یابد و نتایج را بر روی پرتال به نمایش بگذارد. یکی از مواردی که می‌توان بعدها در مورد آن مطالعه کرد مربوط به ادراک متفاوت افراد از منابع حقیقت‌یابی می‌باشد. جزئیات بیشتر در مورد ادبیات تحقیق و محدودیت‌های راستی‌آزمایی را می‌توان در [۱۴۶] مشاهده کرد.

۱۶.۳.۵ مجموع داده‌ها برای پژوهش‌های مبنی بر کاهش انتشار اطلاعات غلط

تکنیک‌های مسدود کننده‌ی نفوذ و پویش حقیقت عمدتاً در مجموع داده‌های شبکه‌ی موجود تأیید شده‌اند. ما برخی از این مجموع داده‌هایی که مورد استفاده قرار گرفته‌اند را در اینجا به صورت خلاصه آورده‌ایم.

۱. فیس‌بوک: اسنپ‌شات‌های فیس‌بوک که شامل ۶۳۳۹۲ گره و ۸۱۶۸۳۱ یال [۱۴۷] می‌شود، توسط [۱۴۸] استفاده شده‌اند. بوداک و همکاران [۹۱] آزمایشاتی را بر روی چهار اسنپ‌شات فیس‌بوک انجام دادند، دو اسنپ‌شات برای هر یک از (۱) شبکه اصلی سانتا باربارا[۱۶۱] و (۲) شبکه اصلی مونته ری بِی[۱۶۲]. یک اسنپ‌شات فیس‌بوک که دارای ۴۰۳۹ گره و ۸۸۲۳۴ یال است، در [۵۰،۵۴،۷۶] استفاده شده است. مجموع داده‌ی دیگری از فیس‌بوک (۴۳۹۵۳ گره و ۲۶۲۶۳۱ یال) که با بکارگیری وال پست‌ها ایجاد شده است، در [۱۳] استفاده می‌شود.

[۱۶۰] یکی از خدمات وب آمازون است. [مترجم]

[۱۶۱] Santa Barbara

[۱۶۲] Monterey Bay

۲. توییتر: زیرمجموعه‌ی استخراج شده از توییتر به صورت گسترده جهت پژوهش در مورد کاهش اخبار جعلی استفاده شده است. در [۱۴۹]، مؤلفان یک مجموع داده‌ای از ۵۵۴ هزار گره و ۴.۲۹ میلیون یال را جمع‌آوری کردند و این داده‌ها در [۹۵] مورد استفاده قرار گرفته است. یک اسنپ‌شات توییتر [۱۵۰] که دارای ۸۱۳۰۶ گره و ۱۷۶۸۱۴۹ یال است در [۱۳، ۵۷] استفاده شده است. یک زیرگراف توییتر که با به کارگیری توییترهای مربوط به هیگز-بوزون [۲۰] استخراج شده است در [۹۲] استفاده شده است. دیگر اسنپ‌شات‌های توییتر[۷۱،۷۹،۱۵۱] هستند.

۳. گوالا [۱۵۲]: این مجموع داده از گوالا که همان شبکه اجتماعی آنلاین لوکیشن-محور است، استخراج شده است. این مجموع داده در دوره فوریه ۲۰۰۹ تا اکتبر ۲۰۱۰ جمع‌آوری شد و شامل ۱۹۶۵۹۱ گره و ۹۵۰۳۲۷ یال می‌شود. این توسط [۶۷، ۹۵] استفاده شده است.

٤. Weibo: [۱۴۹] weibo.comرا جست‌وجو کرد و مجموع داده‌ای مشتمل بر ۱.۰۲ میلیون گره و ۱۶۶.۷ میلیون یال جمع آوری نمود. این مجموع داده توسط [۹۵] استفاده شده است. اسنپ شات دیگری که دارای ۲۳.۸۶ گره و ۱۸۳۵۴۹ یال است نیز در [۵۳] مورد استفاده قرار گرفت.

۵. فوراسکوئر: [۱۴۹] مجموع داده‌ای از ۴.۹ میلیون گره و ۵۳.۷ میلیون یال را جمع آوری کرد و این داده‌ها در [۹۵] استفاده می‌شود.

٦. ویکی-ووت [۱۵۳]: این شبکه حاوی داده‌هایی از ویکیپدیا می‌باشد که از ابتدای آن تا ژانویه ۲۰۰۸ رأی داده است. این شامل ۷۱۱۵ گره و ۱۰۳۶۸۹ یال است. این مجموع داده در [۱۲، ۵۷، ۵۹، ۷۹، ۱۵۱، ۱۵۴] استفاده شده است.

۷. ناتلا۰۸ [۱۵۵، ۱۵۶]: این مجموع داده از شبکه‌ی به اشتراک گذاریِ همتا به همتای ناتلا[۱۶۳] استخراج شده است، جایی که در آن گره‌ها کاربران میزبان هستند و یال‌ها ارتباط میان آنها می‌باشند. یک اسنپ‌شاتی که دارای ۶۳۰۱ گره و ۲۰۷۷۷ یال است در [۱۲]، و دیگری در [۸۱، ۸۳] استفاده شده است.

۸. اسلش دات: این یک مجموع داده‌ی دوستانه از کاربران حاضر در وب سایت اسلش دات هستند که دارای ۱۳۱۸۲ گره و ۳۰۹۱۴ یال می‌باشد. این داده در [۵۹، ۶۷] استفاده می‌شود.

۹. گوگل پلاس: این یک شبکه‌ی هدایت شده است که از گوگل پلاس استخراج شده و یک یال هدایت شده بین دو گره نشان می‌دهد که یک گره دارای گره دیگری در سیکل خود می‌باشد. این داده شامل ۲۳۶۲۸ گره و ۳۹۲۴۲ یال است. این مجموع داده در [۵۹، ۱۵۱] استفاده شده است.

[۱۶۳] Gnutella peer-to-peer filesharing network

۱۰. ایپینیونز[۱۶۴] [۱۵۷]: این یک شبکه‌ی امانتی است که از Epinions.com، یک سایت برای نقدهای کلی مصرف کننده، استخراج شده است. این شبکه شامل ۷۵۸۷۹ گره و ۵۰۸۸۳۷ یال است. این مجموع داده در [۱۲، ۵۵، ۵۷، ۷۱، ۷۴، ۷۹، ۱۵۴] استفاده شده است.

۱۱. انرون [۱۵۸]: یک شبکه‌ی ارتباطات ایمیلی است. این مجموع داده شامل ۳۶۶۹۲ گره و ۳۶۷۶۶۲ گره است. این در [۵۱، ۶۷، ۶۹] استفاده شده است.

۱۲. نت ساینس [۱۵۹، ۱۶۰]: یک شبکه‌ی هم تألیفی از پژوهشگران علوم شبکه است. شامل ۱۵۸۸ گره و ۲۷۴۲ یال است. این مجموع داده در [۲۴، ۵۲، ۵۶] استفاده شده است.

۱۳. گراف سیستم‌های مستقل [۱۶۱]: این یک شبکه چه کسی با چه کسی صحبت می کند[۱۶۵] است که از جریان ترافیک داده‌های روتر استخراج شده است. این مجموع داده در طول ۷۸۵ روز از ۸ نوامبر ۱۹۹۷ تا ۲ ژانویه ۲۰۰۰ جمع آوری شده و شامل ۶.۴ هزار گره و ۱۲.۵ هزار یال می‌شود. این در [۵۲] استفاده شده است.

۱۴. Hep-Th [۲۲]: این یک شبکه‌ی هم تألیفی است که با به‌کارگیری انتشارات در بخش arXiv از تئوری فیزیک انرژی بالا ایجاد شده است. این شبکه در [۵۲، ۵۶، ۶۶، ۸۳، ۱۰۲، ۱۱۹، ۱۴۸، ۱۵۴] استفاده شده است.

۱۵. Hep-Ph [۱۶۱]: [۹۲] از گراف استنادی Hep-Ph استفاده می‌کند که دارای ۳۴۵۴۶ گره و ۴۲۱۵۷۸ یال می‌باشد. دیگر ورژن ها در [۵۱، ۵۵، ۵۹] استفاده شده‌اند.

۱۶. NetPhy: [۶۶، ۱۰۲] از شبکه‌ی هم تألیفی استفاده کرده است که از مقالاتی در بخش فیزیک استخراج شده است [۱۰۴، ۱۱۹].

۱۷. برایت کایت: دارای ۱۹۷ هزار گره و ۹۵۰ هزار یال است و در [۷۴] استفاده شده است. دیگر ورژن‌های برایت کایت در [۸۱، ۸۳] استفاده شده است.

۱۶.۴ نتیجه‌گیری

با توجه به اثرات اخیر انتشار اخبار جعلی در اکثر رویدادها، این امر به ویژه در رسانه اجتماعی یک تهدید واقعی برای جامعه به‌حساب می‌آید. اخبار جعلی و اطلاعات نادرست تبدیل به بخش جدانشدنی از شبکه سازیِ آنلاین اجتماعی شده‌اند، جایی که کاربران با نظرات، اعتقادات و چشم‌اندازهای خود بر یکدیگر اثر می‌گذارند. در این فصل، ما ابتدا در مورد اینکه چگونه اخبار جعلی در رسانه اجتماعی منتشر می‌شوند، صحبت کردیم. درک بهتر از انتشار اخبار به ما کمک می‌کند تا برای شناسایی اخبار جعلی و کاهش آن تکنیک‌های مؤثری طراحی کنیم. برای کاهش اخبار جعلی، روش‌های مسدودسازیِ نفوذ و پویش حقیقت پیشنهاد شده است که بر شناساییِ مجموعه‌ای بهینه از کاربرانی تمرکز دارد که یا می‌توانند برای

[۱۶۴] Epinions
[۱۶۵] who-talks-to-whom

مسدودسازیِ نفوذ، ایمن سازی شوند و یا اطلاعات حقیقی را در شبکه به ترتیب منتشر کنند. ما در مورد پیشرفته‌ترین تکنیک‌های کاهش اخبار جعلی بحث کردیم. هم‌چنین به بحث در مورد ابزارهایی پرداختیم که برای کمک به کاربران از طریق اینکه آیا اخبار ارائه شده جعلی هستند یا خیر و احتمال آن چقدر است، طراحی شده‌اند.

در حال حاضر، این پیشینه‌ی تحقیق شامل روش‌های تعمیم‌یافته‌ای می‌شود که با کلیه‌ی اطلاعات غلط مقابله می‌کند. با این حال، هایدن و آلتویس [۱۶۲] پیشینه‌ی تحقیق اخبار جعلی را بررسی کردند و سه دسته بندی از اخبار جعلی ارائه دادند: (i) اطلاعات دروغ (اطلاعات غلطی که عمداً به اشتراک گذاشته می‌شوند)، (ii) اطلاعات نادرست (اطلاعات غلطی که غیرعمدی به اشتراک گذاشته می‌شوند)، و (iii) تکمیل کننده (رد کردن اطلاعاتی که شخص با آن مخالفت می‌کند، برای ختم یک مناظره). انتشار انواع مختلفی از اطلاعات غلط و تکنیک‌های کاهش آن‌ها همچنان سؤالی باز برای مطالعات بیشتر است. به علاوه نیاز است که مطالعات آکادمیک و سیاستگذاری‌ها برای مقابله‌ای مؤثر با اثرات معکوسِ اطلاعات غلط همسو شوند. این مقاله نیز یک سوال باز پژوهشی است و با توجه به پیچیدگیِ آن‌ها، در مورد چگونگی اجرای این دو به موازات یکدیگر هنوز بررسیِ کامل و عمیقی انجام نگرفته است.