مروری بر روش‌های تشخیص شایعات در شبکه‌های اجتماعی

فریباصادقی[1] و امیرجلالی بیدگلی[2]*

[1]دانشجوی کارشناسی ارشد مهندسی فناوری اطلاعات، دانشکده فنی مهندسی، دانشگاه قم، قم
F.Sadeghi@stu.qom.ac.ir

[2]استادیار، دانشکده فنی مهندسی، دانشگاه قم، قم
jalaly@qom.ac.ir

چکیده

شایعات، اخبار تأییدنشده و اغلب اشتباهی هستند که به‌صورت وسیع در سطح جامعه منتشر و موجب سلب اعتماد یا افزایش کاذب اعتماد گره‌های یک شبکه به یک نهاد یا موضوع می‌شوند. با فراگیرشدن شبکه‌های اجتماعی در سال‌های اخیر، به‌رغم کاربردهای مثبت آنها، انتشار شایعات ساده‌تر و شایع‌تر شده است. شایعات یک چالش امنیتی در شبکه‌های اجتماعی محسوب می‌شوند، چون یک گره بدخواه می‌تواند با انتشار یک شایعه به‌سهولت، اهداف خود را بدنام و یا منزوی کند. ازاین‌رو تشخیص شایعات چالش مهمی در سازوکارهای امنیت نرم مانند اعتماد و شهرت است. پژوهش‌گران تاکنون روش‌های مختلفی را جهت مدل‌سازی، تشخیص و پیش‌گیری از شایعه ارائه داده‌اند. در این پژوهش روش‌های تشخیص شایعه در شبکه‌های اجتماعی مرور خواهند شد. در ابتدا به‌صورت اجمالی ویژگی‌های مورد استفاده در پژوهش‌های پیشین را مورد بررسی قرار می‌دهیم؛ سپس رویکردهای مورد استفاده را بررسی خواهیم کرد و مجموعه داده‌هایی را که بیشتر مورد استفاده قرار گرفته‌اند، معرفی می‌کنیم. در پایان، چالش‌هایی که برای پژوهش‌های آینده در زمینهٔ کاوش در رسانه‌های اجتماعی برای تشخیص و حل شایعات وجود دارد، معرفی شده است.

واژگان کلیدی: تشخیص شایعه، شبکهٔ اجتماعی، یادگیری ماشین، شبکهٔ عصبی

۱- مقدمه

رشد سریع فناوری اطلاعات و رسانه‌های ارتباطی در بستر اینترنت و نفوذ در زندگی روزمرهٔ مردم باعث شده شیوه‌های ارتباطی از بستر سنتی خود به محیط‌های مجازی انتقال پیدا کنند. امروزه شبکه‌های اجتماعی یک مفهوم حیاتی هستند که از طریق آن‌ها اخبار و اطلاعات منتشر می‌شوند. شبکه‌های اجتماعی این امکان را برای کاربران فراهم می‌کنند تا به‌طور مداوم با یکدیگر در ارتباط باشند و از وقایع و اتفاقات جاری مطلع شوند. برخلاف تمام جنبه‌های مثبت این نوع شبکه‌ها، به‌دلیل سهولت در انتشار اخبار و اطلاعات، اخبار غلط و شایعات نیز به‌سادگی در این شبکه‌ها منتشر می‌شوند، و به‌آسانی در بین طیف کثیری از مردم گسترش می‌یابند، حتی گاهی اوقات در این رسانه‌ها، خبر زودتر از رسانه‌های حرفه‌ای و رسمی منتشر می‌شود. شایعه یک چالش برای امنیت شبکه‌های اجتماعی و سازوکارهای امنیت نرم (مانند اعتماد و شهرت) است. شایعات گاهی فقط اطلاعات کاذبی هستند که در بین افراد منتشر می‌شوند و گاهی به‌صورت هدفمند انتشار می‌یابند و می‌توانند خسارات جبران‌ناپذیری برای ارگان‌ها، سازمان‌ها، افراد حقیقی، دولت و حتی احاد مردم به‌وجود بیاورند، باعث افزایش اضطراب اجتماعی شوند، کاهش میزان بهره‌وری و تولید را به همراه داشته باشند و باعث فلج‌کردن چرخهٔ اقتصادی شوند، و بی‌اعتمادی، بدبینی و سوءظن را در جامعه افزایش دهند. برای مثال در سال ۱۳۹۶ پس از زلزلهٔ کرمانشاه و خسارت به مناطق غرب ایران، زلزلهٔ کوچکی در تهران اتفاق افتاد و پس از آن بازار داغ شایعات در فضای مجازی به راه افتاد که هر صد سال یک‌بار زلزلهٔ مهیبی در تهران رخ می‌دهد که زمین کل تهران را در خود خواهد بلعید، این خبر باعث ایجاد رعب و وحشت زیادی در بین مردم شد و رونق اقتصادی و فضای بورس و معاملات ارزی با نوسانات شدیدی روبه‌رو شد [۱].

*نویسندهٔ مسؤول

اطلاعاتی را که در فضای مجازی انتشار می‌یابند، می‌توان به اطلاعات صحیح، اطلاعات غلط و اطلاعات تأییدنشده تقسیم‌بندی کرد. در مقالات مختلف تعاریف زیادی برای شایعه آورده شده که از بین آن‌ها تعاریف زیر بیشتر از سایر تعاریف تکرار شده‌اند:

۱) شایعه، اطلاعات تأییدنشده هنگام رخداد وقایع است، که با ابهام و عدم قطعیت همراه است و به‌سرعت در جامعه منتشر می‌شود. ویژگی اصلی شایعه قابلیت انتشار کنترل ناپذیر و سرعت بالای نشر آن است [۲, ۳].

۲) شایعه را می‌توان بیانیه‌ای بحث‌برانگیز و واقعه‌ای قابل بررسی دانست و نوعی آلودگی ذهنی، که مانند بیماری‌های همه‌گیر به‌صورت ویروسی انتشار می‌یابد [۴].

۳) شایعه، خبری است که شایع شود، و از فردی به فرد دیگر سرایت می‌کند؛ اما از آن جهت که به‌طورمعمول حرف‌های غیردقیق، اشتباه و دروغ هیجان‌انگیزتر از واقعیت هستند، تعداد شایعه‌های دروغ و نادرست بیشتر است، در حدی که درعمل شایعه به معنای مطالب دروغین نیز به‌کار می‌رود [۵-۸].

امروزه شایعات در رسانه‌های اجتماعی به یک نگرانی جدی تبدیل شده‌اند؛ به‌ویژه هنگامی که مردم از توانایی‌هایشان برای تأثیرگذاری بر جامعه آگاهی دارند، در سراسر دنیا نیز غول‌های تجاری و مقامات دولتی و پژوهش‌گران در تلاش برای شکست تأثیرات منفی شایعات هستند. می‌توان کارهای انجام‌شده در زمینهٔ شایعات را در چهار گروه دسته‌بندی کرد:

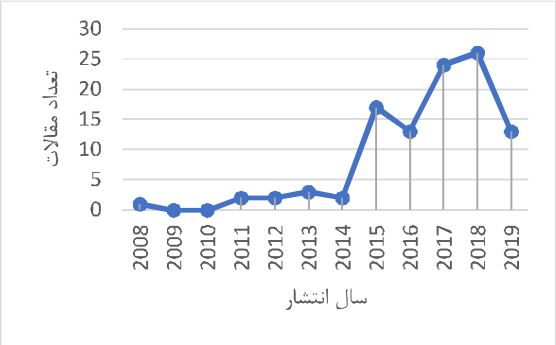
۱) شناسایی گره‌های منشاء شایعه

۲) تشخیص شایعه

۳) تحلیل و مدل‌سازی انتشار شایعه

۴) کنترل انتشار شایعه

همهٔ موارد بالا به نوعی وابسته به تشخیص درست و به‌هنگام شایعات است؛ به همین علت سامانه‌های تشخیص شایعات در رسانه‌های اجتماعی با استقبال زیادی روبه‌رو شده‌اند. در بسیاری از پژوهش‌های انجام‌شده، سه رسانهٔ کلیدی مورد مطالعه در حوزهٔ شایعات توئیتر[1]، ویبو[2] و فیسبوک[3] هستند. بیش‌تر روش‌های انجام‌شده برای تشخیص شایعات به کاربران رسانه‌های اجتماعی نیاز دارند تا از طریق آن‌ها بتوانند شایعات را شناسایی کنند [۸]. این روش‌ها اغلب بر تشخیص شایعات در طی مراحل پخش آن‌ها در رسانه‌های اجتماعی تمرکز دارند و نه تشخیص شایعات در حال ظهور در مراحل اولیه. روش‌های یادگیری ماشین به اطلاعات برچسب‌خورده جهت آموزش نیاز دارند؛ اما بررسی دستی اطلاعات موجود در شبکه‌های اجتماعی بسیار سنگین بوده و با چالش‌های زیادی روبه‌رو است و زمان زیادی برای برچسب‌گذاری اطلاعات به‌صورت دستی نیاز است [۹]. ممکن است به‌دلیل کمبود اطلاعاتی که در مورد موضوع مورد نظر وجود دارد، برچسب‌هایی با کیفیت پایین تولید شود [۱۰]، بنابراین، این روش‌ها نمی‌توانند در تشخیص شایعات به‌طور مؤثر کارا واقع شوند. با توجه به این چالش‌ها، در حال حاضر سامانه‌های موجود برای تشخیص شایعات به عوامل انسانی و رایانه‌ای نیاز دارند. یکی از مهم‌ترین چالش‌های تشخیص شایعات در رسانه‌های اجتماعی داده‌های بسیار زیاد این شبکه‌ها است که در هر لحظه به‌سرعت افزایش می‌یابد. ماهیت این داده‌های بدون ساختار باعث می‌شود پردازش این داده‌ها بسیار چالش‌برانگیز باشد. تاکنون تلاش‌های زیادی برای تشخیص شایعات در شبکه‌های اجتماعی صورت گرفته که می‌توان آن‌ها را به‌تفکیک سال‌های مختلف در شکل (۱) مشاهده کرد که روند رو به رشد پژوهش‌ها و توجه به حوزه تشخیص شایعه را در سال‌های اخیر نشان می‌دهد. این آمار شامل کلیه پژوهش‌هایی با واژه کلیدی تشخیص[4] یا شناسایی[5] شایعات و مشتقات آن هستند.



(شکل-۱): تعداد پژوهش‌های منتشرشده در حوزهٔ تشخیص شایعات به‌تفکیک سال‌های مختلف

در این مقاله هدف ما ارائهٔ یک پیشینهٔ جامع برای روش‌های تشخیص شایعات و ویژگی‌های مورد استفاده در آنها است. مقاله به‌صورت زیر ادامه پیدا می‌کند. در بخش بعدی ابتدا ویژگی‌های مورد استفاده در مقالات را مورد بررسی قرار می‌گیرد؛ سپس در بخش سوم به مرور روش‌های موجود در

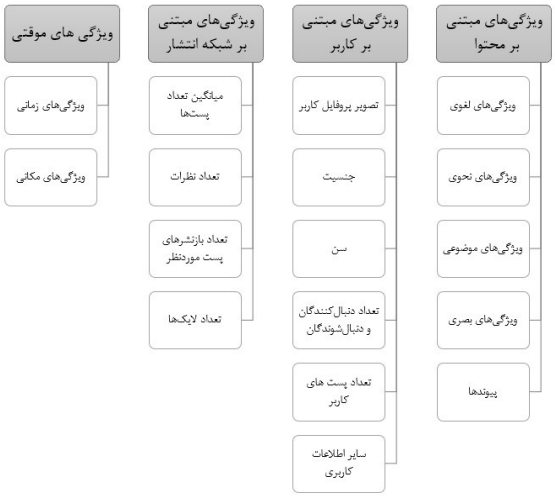
[1] Twitter
[2] Weibo
[3] Facebook
[4] Detection
[5] Identification

مقالات در دو دسته کلی روش‌های باناظر و بدون ناظر می‌پردازیم. بخش چهارم مجموعه داده‌های دردسترس را برای پژوهش‌های در حوزه تشخیص شایعات معرفی می‌کند و درنهایت در بخش پنجم مقاله با جمع‌بندی و معرفی چالش‌های این حوزه پایان می‌یابد.

۲- ویژگی‌های مورداستفاده در تشخیص شایعات

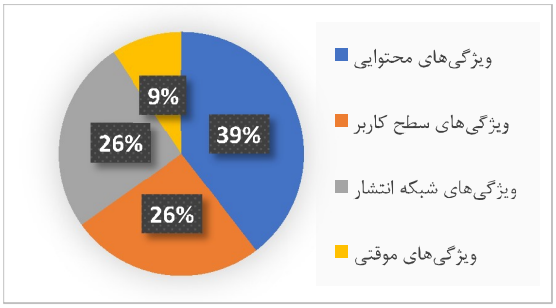
در پژوهش‌های مورد بررسی از ویژگی‌های مختلفی برای تشخیص شایعات استفاده شده که در هر روش به سطوح مختلفی از مهندسی ویژگی نیاز است که برخی روش‌ها نیازمند استخراج دستی کلیهٔ بردارهای ویژگی و برخی دیگر تنها نیازمند سطحی از استخراج دستی این بردارها هستند؛ که با توجه به این‌که بسیاری از ویژگی‌ها به‌هم مرتبط هستند بسیار مهم و ضروری است که این ویژگی‌ها به‌درستی استخراج شوند و این امر یک گام اساسی در تشخیص شایعات است. براساس شکل (۲) ویژگی‌ها را از چهار بُعد می‌توان موردبررسی قرارداد:

۱) ویژگی‌های مبتنی بر محتوا
۲) ویژگی‌های مبتنی بر کاربر
۳) ویژگی‌های مبتنی بر شبکه انتشار
۴) ویژگی‌های موقت زمانی و مکانی



(شکل-۲): ویژگی‌های مورد استفاده در تشخیص شایعات

در شکل (۳) نمودار آماری استفاده از هر دسته از ویژگی‌ها برحسب تعداد مقالات آورده شده است. همان‌طورکه در شکل مشاهده می‌شود، ویژگی‌های محتوایی، بیش از سایر ویژگی‌ها در پژوهش‌های حوزهٔ تشخیص مورد استفاده پژوهش‌گران قرارگرفته‌اند. در ادامهٔ این بخش ویژگی‌های هر دسته را به همراه نمونه پژوهش‌هایی از هر دسته معرفی و مرور می‌کنیم.



(شکل-۳): نمودار ویژگی‌های مورداستفاده برحسب تعداد مقالات

۲-۱- ویژگی‌های مبتنی بر محتوا

ویژگی‌های محتوایی ویژگی‌های استخراج‌شده از محتوای متنی و محتوای بصری داده‌ها است، و شامل ویژگی‌های لغوی، ویژگی‌های نحوی، ویژگی‌های موضوعی، بصری و پیوندها می‌شود، ویژگی‌های لغوی شامل میانگین طول پست‌های شبکه اجتماعی، آمار واژگان، الگوهای شایعات در سطح لغات و ضمایر مورد استفاده، واژگان احساسی منفی و مثبت منتقل‌شده از طریق پست موردنظر و علائم نگارشی مانند علامت سؤال و تعجب است. ویژگی‌های نحوی برگرفته از دستورزبان جملات و شایعات در سطح جمله است. برای مثال تعداد واژگان کلیدی، نمره احساسات و یا قطبیت جمله و ویژگی‌های موضوعی از سطح مجموعهٔ پیام‌ها استخراج می‌شوند که هدف آن‌ها درک پیام‌ها و روابط بنیادین آن‌ها در یک پیکرهٔ زبانی است. ویژگی‌های بصری شامل ویژگی‌های محتوایی و آماری داده‌های بصری و شکلک‌ها است و پیوندها از طریق گذاشتن هشتگ یا نشانی‌های اینترنتی[1] و مخاطب قراردادن یک فرد خاص[2] تشکیل می‌شوند. در پژوهش‌های [۴, ۷, ۹, ۱۱-۲۴] از این ویژگی‌ها بهره گرفته شده است. به‌عنوان نمونه *وو*[3] و *همکارانش* [۱۳] با استفاده ویژگی‌های موضوعی، لغوی و نحوی و استفاده از یک مدل توزیعی، هیجده ویژگی موضوعی را با استفاده از یک ماشین بردار پشتیبان مبتنی بر گراف در تمام پیام‌ها آموزش دادند که هر پیام می‌تواند به یک یا چند موضوع تعلق داشته باشد و بدین ترتیب شایعات را در کار خود شناسایی کردند.

[1] URL
[2] Mention
[3] Wu

[DOR: 20.1001.1.24763047.1398.8.1.1.7]

۲-۲- ویژگی‌های مبتنی بر کاربر

ویژگی‌های مبتنی بر کاربر از پروفایل شبکهٔ اجتماعی کاربر گرفته شده است. یکی از ویژگی‌های کلیدی رسانه‌های اجتماعی در مقایسه با رسانه‌های سنتی، تعاملی‌بودن آن‌ها است. مانند تعاملات میان کاربران، مثل «افزودن دوست[1]» و «دنبال‌کردن[2]» این نوع تعاملات، شبکه‌ای گسترده و پیچیده را تشکیل می‌دهند که در آن اطلاعات زیادی جریان می‌یابد. تجزیه و تحلیل ویژگی‌های کاربران می‌تواند سرنخ‌های مهمی را برای تشخیص شایعات فراهم کند. ویژگی‌های کاربر می‌تواند خصوصیات یک کاربر یا یک گروه کاربری را که شامل چندین کاربر است که با هم در ارتباط هستند، توصیف کند. ویژگی‌های فردی که از یک کاربر استخراج می‌شود، مانند «زمان ثبت نام»، «تصویر پروفایل کاربر»، «سن»، «جنسیت»، «شغل»، «تعداد دنبال‌کنندگان»، «تعداد دنبال‌شوندگان» و «تعداد پست‌های ارسال‌شده»، نوع کاربر که آیا جزو افراد مشهور[3] و شناخته‌شده است یا خیر؟! و سایر اطلاعاتی که کاربر در زمان ثبت‌نام خود وارد کرده است. در پژوهش‌های [۴, ۵, ۱۱, ۱۵, ۱۷, ۱۹-۲۱, ۲۵-۲۷] می‌توان نمونه‌هایی را از استفاده از این ویژگی‌ها مشاهده کرد. به‌عنوان مثال *یانگ*[4] *و همکارانش* [۱۷] در بخشی از کار خود با استفاده از ویژگی‌های سطح کاربر یعنی تاریخ ایجاد پروفایل کاربری، تعداد دنبال‌کنندگان و دنبال‌شوندگان، جنسیت کاربر، نام کاربری فرد و مشخصات موجود در صفحهٔ بیوگرافی او به تشخیص شایعات پرداختند.

۲-۳- ویژگی‌های مبتنی بر شبکه‌ی انتشار

ویژگی‌های مبتنی بر شبکهٔ انتشار از ویژگی‌های شبکه‌ای که در آن شایعات پخش می‌شوند و توپولوژی شبکه به‌دست می‌آید. به‌عنوان نمونه‌های از این ویژگی‌ها می‌توان از «گذاشتن یک محتوای جدید[5]»، «گذاشتن نظر برای محتواهای مختلف[6]»، «ارسال مجدد محتوای فردی دیگر به‌عنوان پست جدید[7]» «تعداد بازخوردهای مثبت و منفی[8]»، منفی[8]»، شبکهٔ ارتباطی بین کاربران، تعداد گره‌ها، پیوندها و متوسط درجهٔ تراکم شبکه نام برد. در پژوهش‌های [۴, ۵, ۷, ۹, ۱۲, ۱۷, ۱۹, ۲۱, ۲۵-۲۷] این ویژگی‌ها برای تشخیص شایعات مورد استفاده قرارگرفته‌اند. به‌طور مثال *یانگ و همکارانش* [۱۷] در بخش دیگری از کار خود از این دسته ویژگی‌ها برای تشخیص شایعات استفاده کردند، به این صورت که در شبکه اجتماعی ویبو، هر کاربر می‌تواند یک پیام را بخواند و درمورد آن اظهار نظر کند، که شرایط خوبی را برای سازندگان شایعه به‌منظور به‌اشتباه‌انداختن مردم فراهم می‌کند. به‌عنوان مثال، برخی از بررسی‌ها نشان می‌دهد که گروهی از افراد که شایعات را منتشر می‌کنند با قراردادن نظرات و بازخوردهای مثبت در زیرپست‌ها برای تبلیغ اعتبار و گسترش نفوذ از یکدیگر پشتیبانی می‌کنند و همچنین، کاربر بیشتر مایل به اعتماد به اطلاعات ارائه‌شده توسط کاربرانی است که در فهرست دوستان او قرارگرفته‌اند؛ که همه این موارد می‌توانند تأثیر مهمی برروی بررسی‌ها داشته باشند که نباید نادیده گرفته شوند.

۲-۴- ویژگی‌های موقتی

ویژگی‌های موقتی شامل ویژگی‌هایی هستند که نسبت به زمان متغیر هستند. مانند موقعیت مکانی کاربر در زمان ارسال نخستین پست شایعه و تقسیم رویدادهای شایعه به فواصل زمانی مختلف و استخراج ویژگی‌های آن‌ها، که ویژگی‌های زمانی به عملکرد بلندمدت پست‌های شایعه بستگی دارد و در مراحل ابتدایی انتشار شایعه مناسب نیستند؛ زیرا تعداد آن‌ها در نخستین ساعات انتشار شایعه محدود است و برای تشخیص شایعه مناسب نیستند. در پژوهش‌های [۱۲, ۱۶, ۲۵, ۲۸] این دسته از ویژگی‌ها مورداستفاده قرارگرفته‌اند. به‌عنوان نمونه در پژوهش [۲۸] کوان و همکارانش در طی ۵۶ روز از لحظه انتشار پیام‌ها تا روز ۵۶ام روند انتشار را جمع آوری و از این ویژگی‌ها جهت شناسایی شایعات استفاده کردند. ایشان دریافتند برخی از ویژگی در فواصل کوتاه‌مدت مانند چند روز یا بلافاصله پس از انتشار می‌توانند جهت تشخیص شایعات مورد استفاده قرار گیرند؛ درحالی‌که برای برخی دیگر از ویژگی‌ها مدت زمان بیشتری جهت استفاده مورد نیاز است. به‌عنوان نمونه میزان گسترش پیام‌ها و ویژگی‌های زمانی تا روز چهاردهم را در الگوریتم‌های دسته‌بندی دقت مناسبی که بتوان به آن استناد کرد، در اختیار قرار نمی‌دهند و نیاز به بازهٔ زمانی طولانی‌تری از زمان انتشار دارند.

۳- رویکردهای مورداستفاده در تشخیص شایعات

روش‌های ارائه‌شده را برای تشخیص شایعات می‌توان به دو دسته اصلی تقسیم کرد: ۱) رویکردهای بدون ناظر ۲) رویکردهای با ناظر. دستهٔ نخست این روش‌ها برای تشخیص

[1] Add friend
[2] Following
[3] Celeberity
[4] Yang
[5] Posting
[6] Commenting
[7] Reposting
[8] Likes & Dislike

[DOR: 20.1001.1.24763047.1398.8.1.1.7]

شایعات از داده‌های بدون برچسب استفاده می‌کنند. این دسته در برابر دسته دوم شامل تعداد بسیار محدودی از مقالات هستند. در مقابل، دستهٔ دوم رویکردها که بخش اعظم مقالات را در بر می‌گیرند از داده‌های دارای برچسب و به‌طورعمومی روش‌های طبقه‌بندی استفاده می‌کنند. برچسب‌ها در این رویکرد به‌طورمعمول توسط افراد متخصص یا سایت‌های حقیقت‌سنجی[1] به داده‌ها زده می‌شوند.

۳-۱- رویکردهای بدون ناظر

تهیه داده یکی از چالش‌های حوزهٔ پژوهش‌های شناسایی شایعه است. از آنجا که تهیهٔ داده‌های برچسب‌دار برای تشخیص شایعات کاری دشوار و پرهزینه است، برخی از رویکردها برای تشخیص شایعات از داده‌های بدون برچسب استفاده می‌کنند و برای آموزش سامانه نیازی به داده‌های برچسب‌دار ندارند که به آن‌ها رویکردهای بدون ناظر می‌گویند. در این قسمت نمونه‌هایی از این پژوهش‌ها آورده شده‌اند.

ژائو[2] *و همکارانش* در سال ۲۰۱۵ با پیداکردن الگوی متنی خاص در پیام‌های شایعه که به‌منظور جلب توجّه افکار عمومی مورد استفاده قرارگرفته‌اند، درابتدا پست‌های مربوطه را که حاوی این الگوهای متنی هستند، جمع‌آوری و سپس آن‌ها را خوشه‌بندی کردند. ایشان برای رتبه‌بندی و دسته‌بندی خوشه‌ها از الگوریتم‌های تشخیص شهرت، درخت تصمیم و ماشین بردار پشتیبان استفاده کردند و توانستند برای پیام‌هایی که درطول روز منتشر می‌شوند، خوشه‌های با رتبه‌های بالاتر را با دقت زیادی شناسایی کنند. این خوشه‌ها نشان‌دهندهٔ احتمال وجود شایعه هستند و درنهایت می‌توان برای بررسی بیشتر، این خوشه‌ها را در اختیار کارشناس انسانی قرار داد [۱۴].

چن[3] *و همکارانش* در سال ۲۰۱۸ در کار خود از یک مدل یادگیری بدون ناظر[4] با استفاده از شبکه‌های عصبی بازگشتی و کدگذار خودکار[5] استفاده کردند که نتایج تجربی حاصل از کار نشان می‌دهد که آن‌ها به‌دقّت بالای ۹۲٫۴۹٪ دست یافتند. آن‌ها در کار خود بر رفتار کاربر و ویژگی‌های حاصل از آن تمرکز می‌کنند و شایعات را به‌عنوان بی‌نظمی و ناهنجاری در پست‌های منتشرشده توسط کاربران در نظر می‌گیرند [۲۷].

۳-۲- رویکردهای باناظر

رویکردهای باناظر برای کار تشخیص شایعات و آموزش سامانه، به برچسب داده‌ها نیازمند هستند و براساس داده‌های برچسب‌دار می‌توانند روند شناسایی شایعات را در داده‌های آموزشی یاد بگیرند و شایعات را در داده‌های آزمون شناسایی کنند. بر همین اساس پژوهش‌هایی که به این مدل‌ها اشاره دارند، در این قسمت مورد بررسی قرارگرفته‌اند. به‌طورکلی رویکردهای مورد استفاده در این قسمت به روش‌های یادگیری ماشین کلاسیک که نیازمند استخراج دستی بردار ویژگی‌ها و روش‌های یادگیری کم‌عمق و یادگیری عمیق که در بعضی از پژوهش‌ها به سطحی از مهندسی ویژگی نیازمند هستند، تقسیم می‌شوند. شبکه‌های عصبی عمیق[6] می‌توانند بسیاری از مشکلات موجود در یادگیری ماشین کلاسیک را بهتر و با دقت بالاتری حل کنند و کارایی سامانه را بهبود ببخشند. در رویکردهای شبکه عصبی عمیق، دو شبکه بیش از سایرین مورد استفاده قرار گرفته‌اند؛ یکی از این شبکه‌ها، شبکه‌های عصبی کانولوشنی[7] و دیگری شبکه‌های عصبی بازگشتی[8] هستند. شبکه‌های عصبی کانولوشنی از لایه‌های کانولوشنی تشکیل شده و ساختار آن به مدل‌سازی ویژگی‌های معنایی و سراسری کمک می‌کند. این شبکه‌ها در ابتدا جهت پردازش تصویر مورد استفاده قرار گرفتند، اما اکنون کارآیی خود را در بسیاری از حوزه مانند متن نیز اثبات کرده‌اند. در مقابل، شبکه‌های عصبی بازگشتی بیشتر جهت بررسی یک دنباله[9] از داده کارآیی دارند. در این شبکه‌ها به‌طورعمومی داده‌های شایعه به‌عنوان داده‌های متوالی در نظر گرفته شده‌اند. ساختار این نوع شبکه‌ها، می‌تواند ویژگی‌های دنباله‌ای را که یکی از مشخصه‌های اصلی شایعه است، مدل‌سازی کند. در این بخش در ابتدا روش‌های کلاسیک مورد استفاده در تشخیص شایعه را مرور می‌کنیم. درادامه روش‌های نوین در یادگیری ماشین مانند شبکه‌های عمیق بررسی شده‌اند. درنهایت این بخش با جمع‌بندی کارهای بررسی‌شده در حوزه تشخیص شایعه و مقایسه آن‌ها در جدول (۱) پایان می‌پذیرد.

کاستیلو[10] *و همکارانش* در سال ۲۰۱۱ با استفاده از دسته‌بندهای مختلف مانند ماشین بردار پشتیبان[11]، درخت

[1] مانند سایت‌های politifact.com و snopes.com

[2] Zhao
[3] Chen
[4] Unsupervised Learning
[5] Autoencoders

[6] Deep Neural Network
[7] Convolutional neural network (CNN)
[8] Recurrent neural network (RNN)
[9] Sequence
[10] Castillo
[11] SVM

[DOR: 20.1001.1.24763047.1398.8.1.1.7]

تصمیم، شبکه‌های بیزی[1] برای تشخیص شایعات به دقت ۸۶٪ رسیدند [۱۱].

برای نشان‌دادن یک سند براساس واژه‌هایی که در آن وجود دارد، به‌طورمعمول از مدل بسته واژگان[2] استفاده می‌شود. ما[3] و همکارانش در سال ۲۰۱۷ با استفاده از مدل بسته واژگان و یک مدل زبانی شبکهٔ عصبی[4] برای تولید بردارهای متنی از محتوای متون شایعه با الگوریتم‌های تعبیه معنایی[5] برروی داده‌های شبکه‌ی ویبو استفاده کردند و به دقت بیش از ۹۰٪ با استفاده از مدل بسته واژگان و دقت بیش ۶۰٪ با استفاده از شبکه عصبی در تشخیص شایعات دست‌یافتند [۱۸].

وو[6] *و همکارانش* در سال ۲۰۱۵ با آموزش یک ماشین بردار پشتیبان مبتنی بر گراف نشان دادند هر پیام می‌تواند به یک یا چند موضوع تعلق داشته باشد، و درنهایت شایعات را با دقت ۹۰,۴٪ شناسایی کردند. آن‌ها توانستند یک ساختار دقیق برای توصیف فرایند انتشار برای یک پیام پیشنهاد کنند [۱۳].

جین[7] *و همکارانش* در سال ۲۰۱۵ برای تشخیص شایعات از یک روش دسته‌بندی دوسطحی با استفاده از الگوریتم جنگل تصادفی و درخت تصمیم‌گیری استفاده کردند. آن‌ها فرض می‌کنند که پیام‌های تحت یک موضوع به‌احتمال دارای مقادیر اعتبار مشابهی هستند. با این فرض، آن‌ها پیام‌ها را به مباحث مختلف دسته‌بندی می‌کنند و ویژگی‌های سطح موضوع را با جمع‌آوری ویژگی‌های سطح پیام به‌دست می‌آورند؛ سپس نتایج حاصل از دسته‌بندی در سطح موضوع با ویژگی‌های سطح پیام تلفیق می‌شوند، تا دسته‌بند بهتری را آموزش دهند. نتایج نشان می‌دهند که رویکرد دوسطحی نتایج بهتری نسبت به روش یک‌سطحی دارد. آنها ادعا می‌کنند که این نوع ویژگی‌های موضوعی می‌تواند تأثیر داده‌های پرت[8] را کاهش دهد، درحالی‌که بیشترین جزئیات مربوط به سطح پیام را حفظ می‌کند. دقت روش پیشنهادی آن‌ها به ۹۳٪ رسید [۱۵].

گوپتا[9] *و همکارانش* در سال ۲۰۱۳ نخستین تلاش را برای درک الگوهای زمانی، شهرت اجتماعی و الگوهای نفوذ برای انتشار تصاویر جعلی در توئیتر ایجاد کردند. آن‌ها یک مدل دسته‌بندی برای شناسایی تصاویر جعلی را در توئیتر درطول طوفان سندی پیشنهاد دادند که با استفاده از الگوریتم درخت تصمیم به دقت ۹۷٪ رسیدند [۱۲].

ژانگ[10] *و همکارانش* در سال ۲۰۱۵ رابطهٔ بین تصاویر و شایعات مربوط به حوزه‌ی سلامت را موردمطالعه قرار دادند. در این پژوهش موتور جستجوی بایدو[11] برای پیداکردن تصویر اصلی، جهت محاسبهٔ فاصلهٔ زمانی بین تصویر اصلی و تصویر فعلی به‌کارگرفته شده است. آن‌ها در کار خود علاوه‌بر تشخیص شایعات به بررسی صحت و درستی شایعات نیز می‌پردازند و معتقدند شایعاتی که حاوی آمار، ارقام، مرجع و منبع هستند با احتمال بیشتری در زمرهٔ شایعات صحیح قرار می‌گیرند. ویژگی‌های آماری بصری در این پژوهش از سه جنبه مورد بررسی قرارگرفته است:

۱) تعداد تصاویر: کاربران می‌توانند صفر، یک یا چند تصویر را همراه با محتوای متنی در توئیت خود ارسال کنند. برای نشان‌دادن وقایع تصاویر در پیام‌های شایعه، کل تصاویر در یک رویداد شایعه و نسبت پیام‌هایی که دست‌کم یک یا چند تصویر دارند، شمارش می‌شود.

۲) محبوبیت تصاویر: بعضی از تصاویر بسیار محبوب هستند و نظرات و توجهات بیشتری را به سمت خود جلب می‌کنند. نسبت محبوب‌ترین تصاویر برای نشان‌دادن این ویژگی محاسبه می‌شود.

۳) نوع تصویر: برخی از تصاویر دارای سبک خاصی هستند. به‌عنوان‌مثال، تصاویر عریض، تصاویر با نسبت طول به عرض بسیار بزرگ هستند. نسبت این نوع تصاویر نیز به‌عنوان یک ویژگی آماری محسوب می‌شود. در این پژوهش، نویسندگان دریافتند که تصاویر در شایعات و غیر شایعات از نظر توزیع بصری متمایز هستند. ارزیابی این ویژگی‌های بصری بر روی مجموعه‌ای از ۲۵۵۱۳ تصویر از ۱۴۶ رویداد نشان می‌دهد که تصاویر در شایعات بیشتر متقارن و تنوع کمتری دارند و در مقایسه با غیر شایعات، کمتر تشکیل خوشه می‌دهند؛ درنتیجه، این ویژگی می‌تواند برای تشخیص وقایع شایعه مفید باشد. آن‌ها در کار خود برای دسته‌بندی شایعات از الگوریتم رگرسیون لجستیک[12] استفاده کردند و به دقتی در حدود ۷۵٪ دست یافتند [۱۶].

[1] Bayasian network
[2] Bag of Words (BoW)
[3] Ma
[4] Neural Network Language Model
[5] Semantic Embedding Algorithm
[6] Wu
[7] Jin
[8] outlier
[9] Gupta
[10] Zhang
[11] Baidu
[12] Logistic regression

در پژوهش دیگری که توسط *زمانی*[1] *و همکارانش* در دانشگاه تهران در سال ۲۰۱۷ با هدف تشخیص شایعه در مجموعه توئیت‌های فارسی انجام شد، با استفاده از نتایجی که از الگوریتم‌های درخت تصمیم و بیز ساده[2] گرفته شد، نشان دادند که چه واژگانی بیشتر در شایعات فارسی مورداستفاده قرارگرفته و کاربران مایل به درج و پخش چه نوع شایعاتی هستند. آن‌ها در کار خود به دقت بیش از ۸۰٪ برای تشخیص شایعات فارسی دست یافتند [۱۹].

کوان[3] *و همکارانش* در سال ۲۰۱۳ با استفاده از ویژگی‌های مبتنی بر کاربر، شبکه انتشار و ویژگی‌های موقتی مانند درنظرگرفتن نوسان انتشار شایعات و استفاده از الگوریتم‌های ماشین بردار پشتیبان، درخت تصمیم و جنگل تصادفی توانستند شایعات را با دقت بیش از ۸۱٪ در کار خود شناسایی کنند [۲۵].

یانگ[4] *و همکارانش* نیز در سال ۲۰۱۵ با استفاده از الگوریتم بیز ساده و یک دسته‌بند دوکلاسه و ویژگی‌های سطح انتشار و کاربری و ویژگی‌های محتوایی در شبکهٔ ویبو شایعات را با دقت ۹۲٪ شناسایی کردند [۱۷].

کوان[5] و همکارانش در سال ۲۰۱۷ ثبات ویژگی‌ها را در شبکهٔ توئیتر درطول زمان مورد مطالعه قرار دادند. آن‌ها سطوح عملکردی دسته‌بندی شایعات را درطول سه روز نخست انتشار تا نزدیک به دو ماه از زمان انتشار بررسی می‌کنند و با استفاده از یک الگوریتم دسته‌بندی ابداعی به دقتی در حدود ٪۸۹ در کار خود رسیدند [۲۸].

دهقانی[6] *و همکارانش* نیز در سال ۲۰۱۸ با استفاده از الگوریتم‌های مختلف یادگیری ماشین، تشخیص شایعات در توئیت‌های فارسی را بررسی کردند. آنها به این نتیجه رسیدند که روش جنگل تصادفی و RandomSubSpace نسبت به سایر روش‌ها شایعات را با دقت بهتری شناسایی می‌کنند و با استفاده از این دو الگوریتم به‌ترتیب به دقت ۹۵٪ و ۹۶٪ رسیدند [۲۰].

سیکیلیا[7] *و همکارانش* در سال ۲۰۱۷ با تمرکز بر موضوع سلامتی و بهداشت، اخبار مربوط به حوزهٔ سلامت را از توئیتر جمع‌آوری کردند و توانستند با استفاده از یک الگوریتم جنگل تصادفی شایعات را در پیام‌های خود با دقت ۷۱٪ شناسایی کنند [۲۶].

[1] Zamani
[2] Naiive Bayes
[3] Kwon
[4] Yang
[5] Kwon
[6] Dehghni
[7] Sicilia

گوپتا[8] و همکارانش در سال ۲۰۱۲ درابتدا روش انتشار پیام‌ها را در توئیتر تفسیرکردند. آن‌ها یک شبکه متشکل از کاربران، پیام‌ها و رویدادها را با مشاهدات خود ساختند که شامل دو جزء است: ۱) کاربران قابل‌اعتماد، که به‌طور کلی معتقد به رویدادهای شایعه نیستند. ۲) پیوندهای بین پیام‌های معتبر وزن‌های بزرگ‌تری نسبت به پیام‌های شایعه دارند؛ زیرا پیام‌هایی که در یک رویداد شایعه هستند ادعای منسجمی ندارند. درنهایت با درنظرگرفتن گراف شبکه توانستند شایعات را با استفاده از یک الگوریتم دسته‌بندی ابداعی بر پایهٔ درخت تصمیم به دقت ۸۶٪ شناسایی کنند [۲۱].

جین[9] *و همکارانش* در سال ۲۰۱۴ یک شبکه اعتبار سلسله‌مراتبی سه‌لایه را که از سطوح مختلف معنایی یک رویداد ساخته شده است، پیشنهاد کردند. براساس مشاهداتشان، بسیاری از کاربران به‌طور ناخواسته در رسانه‌های اجتماعی شایعات را پخش می‌کنند و حتّی کاربران معتبر، به‌اشتباه در گسترش شایعات سهیم می‌شوند. آن‌ها دریافتند که به‌طور کلی یک رویداد خبری، در بسیاری از موارد حاوی اطلاعات واقعی و جعلی است؛ بنابراین، بدون تجزیه و تحلیل عمیق‌تر از اجزای آن، یک ارزیابی قانع‌کننده برای این رویداد، دشوار است. هدف آن‌ها به کمینه‌رساندن نفوذ کاربران و تمرکز بر روابط معنایی عمیق‌تر رویدادها با پیشنهاد یک شبکه انتشار اعتبار بود که با استفاده از یک دسته‌بند ابداعی بازگشتی[10] به دقتی در حدود ۸۵٪ رسیدند [۴].

در پژوهش دیگری که توسط جین و همکارانش در سال ۲۰۱۶ انجام شد، وجود دو نوع رابطه بین پیام‌ها در میکروبلاگ‌ها بیان شد. نخستین رابطه پشتیبانی است، یعنی جایی‌که پیام‌هایی که از دیدگاه مشابه استفاده می‌کنند، اعتبار یکدیگر را تأیید می‌کنند. رابطهٔ دیگر تضاد است، زمانی‌که پیام‌ها دیدگاه‌های متضاد دارند، اعتبار یکدیگر را کاهش می‌دهند. ازآن‌جایی‌که میکروبلاگ‌ها محیط‌های چندرسانه‌ای باز هستند، مردم پس از خواندن یک رویداد خبری می‌توانند واکنش‌های افراد دیرباور و حتّی مخالف خود را مطالعه کنند. این صداهای مخالف در کنار اخبار، همراه با صدای حمایت‌کننده‌های اصلی که در شایعات مطرح می‌شوند، اجزای بسیار مهمی برای ارزیابی صحت رویدادهای خبری هستند و با استفاده از این روابط و الگوریتم پیشنهادی مقاله، آن‌ها

[8] Gupta
[9] Jin
[10] Iterative

[DOR: 20.1001.1.24763047.1398.8.1.1.7]

توانستند شایعات را سریع‌تر و با دقت ۸۴٪ شناسایی کنند [۵].

زو[1] *و همکارانش* در سال ۲۰۱۸ در پژوهش خود بین پست‌های اصلی و پیام‌های بازنشرشده تمایز قائل شدند و پیغام‌های حاوی شایعه را از سه جنبه ویژگی‌های محتوایی، ویژگی‌های کاربری و سطح انتشار مورد بررسی قرار دادند. آن‌ها با استفاده از LSTM و سازوکار اهرم توجه[2] به محتوا، سعی دارند با فراهم‌آوردن تمرکز بر واژگان کلیدی منتشرشده در پست‌های اصلی، بازخوردهای مهم را درطول روند انتشار به‌دست آورند و به دقت ۹۴٪ برای تشخیص شایعه در کار خود رسیدند [۲۹].

ما[3] *و همکارانش* در سال ۲۰۱۶ از شبکهٔ عصبی بازگشتی برای تشخیص شایعات استفاده کردند. هدف پژوهش آن‌ها یادگیری تضمینی سری‌های زمانی و متنی از داده‌های برچسب‌دار شایعه و آزمایش‌های گسترده روی آن‌ها توسط این نوع شبکه‌ها بوده است. نتایج پژوهش آن‌ها نشان می‌دهد که دقت مدل آن‌ها به بیش از ۸۸٪ رسیده است [۹].

ما و همکارانش در پژوهشی دیگر که در سال ۲۰۱۸ به انجام رسید، معتقدند تعیین موضع[4] پست‌های موردنظر می‌تواند به تشخیص موفق شایعات مربوط باشد و برعکس. بااین‌حال، بیش‌تر مطالعات موجود به تشخیص شایعه و تعیین موضع به‌عنوان وظایف جداگانه پرداخته‌اند. اما آن‌ها در پژوهش خود، با توجه به همبستگی و ارتباط قوی بین صحت ادعا و موضع بیان‌شده، یک قالب یادگیری چندوظیفه‌ای[5] براساس شبکه‌های عصبی عمیق طراحی کردند و این دو بخش را به‌عنوان یک کار واحد در نظر گرفتند، که شایعات را با دقت بیش از ۷۵٪ شناسایی می‌کند [۳۰].

چن[6] *و همکارانش* در سال ۲۰۱۷ برای نشان‌دادن اهمیت بعضی لغات در پیام‌های شایعه، در کار خود، سازوکار اهرم توجه را در کنار شبکهٔ عصبی بازگشتی به کار بردند. یکی از فرضیات کار آن‌ها این است که ویژگی‌های متنی داده‌های شایعه در گذر زمان، اهمیت خود را از دست می‌دهند. مدل پیشنهادی آن‌ها توانست شایعات را با سرعت و دقت بیش از ۸۸٪ شناسایی کند [۲۲].

ما و همکارانش در سال ۲۰۱۸ در کار خود از ویژگی‌های محتوایی توئیت‌ها با دنبال‌کردن ساختار انتشار نامتوالی آن‌ها و ایجاد نمایش‌های قدرتمند برای شناسایی انواع مختلف شایعات استفاده کردند. آن‌ها از دو مدل شبکهٔ عصبی بازگشتی مبتنی بر درخت تصمیم‌گیری با رویکرد از بالا به پایین و از پایین به بالا برای یادگیری ویژگی‌ها و دسته‌بندی شایعات استفاده می‌کنند، که با الگوی انتشار شایعات هم‌خوانی دارد. مدل‌های شبکهٔ عصبی بازگشتی عملکرد بهتری نسبت به رویکردهای قبلی دارند و بهتر می‌توانند شایعات را در مراحل اولیهٔ پس از انتشار با دقت بیش از ۷۲٪ شناسایی کنند [۲۳].

روچانسکی[7] *و همکارانش* در سال ۲۰۱۷ برروی سه ویژگی داده‌های شایعه تمرکز کرده‌اند: ۱) متن یک پست که منتشر می‌شود، ۲) پاسخ کاربران دریافت‌کننده متن و ۳) کاربرانی که آن را منتشر می‌کنند. این ویژگی‌ها جنبه‌های مختلف داده‌های شایعه را نشان می‌دهد. ازآنجایی‌که شناسایی شایعات مبنی‌بر تنها یکی از این ویژگی‌ها کار دشواری است، لذا آن‌ها در کار خود از یک مدل ترکیبی (CSI) استفاده می‌کنند که هر سه ویژگی را برای پیش‌بینی دقیق‌تر و خودکار شایعات ترکیب می‌کند و به دقت بیش از ۸۹٪ رسیدند [۷].

چن و همکارانش در سال ۲۰۱۸ برای تشخیص شایعات به این چالش پرداخته‌اند که مدل‌های یادگیری عمیق برای یادگیری الگوهای کافی به داده‌های زیادی نیاز دارند و با توجه به این‌که شایعات در میان انبوهی از داده‌ها پراکنده هستند، باید ترکیب داده‌های متنوع، سازگار باشند. در این پژوهش از ترکیب RNN و LSTM استفاده شده است و به دقت بیش از ۶۵٪ دست‌یافته‌اند [۳۱].

جین[8] *و همکارانش* در سال ۲۰۱۷ باتوجه به اطلاعات متنی، اطلاعات بصری و ویژگی‌های شبکه انتشار، یک مدل ترکیبی با استفاده از LSTM و شبکه عصبی بازگشتی و سازوکار اهرم توجه ارائه می‌دهند. آزمایش‌های گستردهٔ آن‌ها برروی مجموعه داده‌های مختلف از داده‌های توئیتر و ویبو نشان می‌دهد که مدل آن‌ها می‌تواند شایعات را در داده‌های حاوی محتوای چندرسانه‌ای، بهتر از روش‌های مبتنی بر ویژگی‌های موجود تشخیص دهد. آن‌ها به دقت ۶۸٫۲٪ بر روی داده‌های توئیتر و ۷۸/۸٪ بر روی داده‌های ویبو رسیدند [۲۴].

یو[9] *و همکارانش* در سال ۲۰۱۷ رویکردی مبتنی بر شبکه عصبی کانولوشنی برای تشخیص شایعات پیشنهاد داده‌اند. مدل آن‌ها می‌تواند ویژگی‌های قابل‌توجهی از یک نمونهٔ ورودی را استخراج و ارتباط بین این ویژگی‌ها را

[1] Xu
[2] Attention Mechanism
[3] Ma
[4] Stance Detection
[5] Multi-Task Learning
[6] Chen
[7] Ruchansky
[8] Jin
[9] Yu

کانولوشنی در این مدل از بازنمایی کلمات در فضای تعبیه‌شده استفاده می‌کند تا بازنمایی‌های پنهان از توئیت‌های مرتبط با شایعات را به‌مرور درطول آموزش مدل یاد بگیرد؛ سپس بخش بازگشتی برای پردازش سری زمانی به‌دست‌آمده از شبکهٔ عصبی کانولوشنی، استفاده می‌شود. آزمایش‌های گسترده، عملکرد خوب و دقت بالای ۸۲٪ مدل را در طی ساعات اولیهٔ انتشار شایعه که داده‌ها و اطلاعات درمورد شایعه کم هستند، نشان می‌دهند [۳۳].

(جدول-۱): جدول مقایسهٔ روش‌های موجود در مقالات بررسی شده

مقاله	سال انتشار	شبکه اجتماعی موردمطالعه	روش مورد استفاده	Accuracy	Precision	Recall	F1
[۱۴]	۲۰۱۵	توئیتر	Clustering- SVM- decision tree- Popularity-RT Ratio	-	٪۵۲	-	-
[۲۷]	۲۰۱۸	ویبو	RNN+ AutoEncoder	٪۹۲/۴۹	٪۹۰/۳۶	٪۸۷/۹۹	٪۸۹/۱۶
[۱۱]	۲۰۱۱	توئیتر	SVM - decision tree – Bayesian network	٪۸۶	٪۷۹/۲	٪۷۸/۸	٪۷۸/۷
[۱۸]	۲۰۱۷	ویبو	Bow	٪۹۰<	-	-	-
			Neural network	٪۶۰<	-	-	-
[۱۳]	۲۰۱۵	ویبو	Graph based SVM	٪۹۰/۴	٪۸۸/۶	٪۹۲/۹	٪۹۰/۷
[۱۵]	۲۰۱۵	توئیتر	Random forest – decision tree	-	٪۹۳/۷۴	٪۹۲/۲	٪۹۴
[۱۲]	۲۰۱۳	توئیتر	decision tree	۹۶٪/۶۵	-	-	-
			naiive bayes	٪۹۱/۵۲	-	-	-
[۱۶]	۲۰۱۵	ویبو	Logestic regression	٪۷۵	-	-	-
[۱۹]	۲۰۱۷	توئیتر	naiive bayes - decision tree	-	٪۸۰<	٪۸۱<	
[۲۵]	۲۰۱۳	توئیتر	decision tree	٪۸۲/۱	٪۸۵/۳	٪۸۴/۳	٪۸۲/۲
			random forest	٪۸۹/۷	٪۹۲/۳	٪۸۸/۳	٪۷۸/۸
			Svm	٪۸۷/۳	٪۹۰/۳	٪۸۷/۳	٪۸۶/۷
[۱۷]	۲۰۱۵	ویبو	Naiive bayes – heuristic algoritm	-	٪۹۲/۴	٪۹۳/۶	٪۹۳
[۲۸]	۲۰۱۷	توئیتر	Heuristic algoritm	٪۸۹	٪۸۹	٪۹۱	٪۹۰
[۲۳]	۲۰۱۸	توئیتر	TD-RvNN	٪۷۲<	-	-	٪۶۸/۲
[۲۰]	۲۰۱۸	توئیتر	RandomSubSpace	-	٪۹۶/۴	٪۹۵/۶	٪۹۶
			Naiive bayes	-	٪۵۸/۴	٪۹۱/۱	٪۷۱/۲
			Random forest	-	٪۹۵/۹	٪۹۶	٪۹۶
[۲۶]	۲۰۱۷	توئیتر	Random forest	٪۷۱/۴	٪۷۰/۳	٪۶۹/۹	٪۸۶/۸
[۲۱]	۲۰۱۲	توئیتر	Heuristic algoritm	٪۸۶	-	-	-
[۴]	۲۰۱۴	ویبو	iterative algorithm	۸۵٪/۱	٪۴۵	٪۷۲	٪۵۵
[۵]	۲۰۱۶	ویبو	Heuristic algoritm	٪۸۴	٪۷۸/۶	٪۹۳/۳	٪۸۵/۳
[۹]	۲۰۱۶	ویبو و توئیتر	RNN	٪۸۸<	٪۸۵	٪۹۵	٪۸۹
[۳۰]	۲۰۱۸	توئیتر	RNN+ Multitask learning	-	-	-	٪۷۵<
[۲۲]	۲۰۱۷	توئیتر	CallAttention+RNN	-	٪۸۸/۶۳	٪۸۵/۷۱	٪۸۷/۱۵
		ویبو		-	٪۸۷/۱۰	٪۸۶/۳۴	٪۸۶/۷۲
[۷]	۲۰۱۷	توئیتر	CSI (Huristic Algoritm)	٪۸۹/۲	-	-	٪۸۹/۴
		ویبو		٪۹۵/۳	-	-	٪۹۵/۴
[۲۴]	۲۰۱۷	ویبو	RNN-Lstm+Attention	٪۷۸/۸	٪۸۶/۲	٪۶۸/۶	٪۷۶/۴
		توئیتر		٪۶۸/۲	٪۷۸	٪۶۱/۵	٪۶۸/۹
[۳۲]	۲۰۱۷	توئیتر	CNN	٪۷۷/۷	٪۷۴/۴	٪۷۰/۵	٪۷۹/۳
		ویبو		۹۳٪/۳	٪۹۲/۱	٪۹۲/۱	٪۹۳
[۲۹]	۲۰۱۸	ویبو	MNRD (lstm+attention)	۹۴٪/۴	٪۹۴/۹	٪۹۴	٪۹۴/۳
[۳۱]	۲۰۱۸	توئیتر	RNN+Lstm	-	٪۶۵<	٪۶۷<	٪۶۶<
[۳۳]	۲۰۱۷	توئیتر	RNN+CNN	٪۸۲<	-	-	-

[1] Nguyen

[DOR: 20.1001.1.24763047.1398.8.1.1.7]

۴- معرفی مجموعه‌داده‌ها

۱) مجموعه‌داده PHEME که نخستین نسخه آن در سال ۲۰۱۶ منتشر شد و شامل پنج واقعه خبری است [۳۴]. نسخهٔ دیگری از این مجموعه داده در سال ۲۰۱۸ در تکمیل نسخه قبلی منتشر شد که شامل مجموعه‌ای از ۲۴۰۲ شایعه و ۴۰۲۳ غیر شایعه و اطلاعات صحت سنجی[1] در ارتباط با ۹ واقعه خبری است [۳۵].

۲) مجموعه‌داده تولیدشده برای RumourEval ۲۰۱۷ با دارابودن بیش از سیصد شایعه برچسب‌گذاری‌شده برای تعیین حقیقت شایعه و استفاده از برچسب‌های صحیح، غلط و تأییدنشده برای هرکدام از آنها است، که مجموعه آموزشی شامل ۲۹۷ شایعه است که برای هشت رویداد جمع‌آوری شده که شامل ۲۹۷ منبع و ۴۲۲۲ پاسخ توئیت است [۳۶]. آخرین نسخهٔ این مجموعه داده در سال ۲۰۱۹ منتشرشده است که داده‌های آن از توئیتر و تارنمای Reddit جمع‌آوری شده است [۳۷].

۳) مجموعه‌داده مناسب دیگری که برای تشخیص شایعه پیشنهاد می‌شود، داده‌های منتشرشده توسط کوان[2] و همکارانش در سال ۲۰۱۷ است، که شامل ۵۱ رویداد شایعه درست و شصت رویداد شایعه غلط است. هر رویداد شایعه شامل تعداد دست‌کم شصت توئیت مرتبط با آن است [۲۸].

۴) مجموعه‌داده RUMDECT که در سال ۲۰۱۶ منتشر شد، از دو نوع اطلاعات ویبو و توئیتر تشکیل‌شده است. اطلاعات ویبو شامل ۲۳۱۳ شایعه و ۲۳۵۱ غیر شایعه و داده‌های حاصل از توئیتر شامل ۴۹۸ رویداد شایعه و ۴۹۴ رویداد غیر شایعه است [۹].

۵) مجموعه‌دادهٔ KNTUPT که توسط دانشگاه خواجه‌نصیر از ۲۴ نوامبر ۲۰۱۷ تا ۸ دسامبر ۲۰۱۷ جمع‌آوری شده است، این مجموعه شامل ۳۵۹۸۰۴۹ توئیت فارسی جمع‌آوری‌شده از شبکه اجتماعی توئیتر است که در آن ۴۳۴۵ توئیت شایعه وجود دارد [۲۰].

۶) مجموعه‌دادهٔ FakeNewsNet که توسط آزمایشگاه داده‌کاوی دانشگاه آریزونا در سال ۲۰۱۹ منتشرشده و شامل ۵۷۵۵ داده شایعه و ۱۷۴۴۱ غیر شایعه و طبق ادعای منتشرکنندگان کامل‌ترین مجموعه‌داده موجود در زمینه تشخیص اخبار جعلی و شایعات است که شامل اطلاعات شبکه اجتماعی توئیتر، سایت‌های خبری، اطلاعات مکانی و زمانی داده‌ها است [۳۸].

[1] Veracity Detection
[2] Kwon

۵- نتیجه‌گیری

در سال‌های اخیر پژوهش در زمینهٔ توسعهٔ ابزارهای تشخیص و اعتبارسنجی شایعات به‌علت محبوبیت و اهمیت روزافزون رسانه‌های اجتماعی در زندگی مردم، به‌طور قابل‌توجهی افزایش یافته است. شبکه‌های اجتماعی، کاربران و پژوهش‌گران را قادر می‌سازد تا اخبار و حقایق را پس از انتشار به‌سرعت جمع‌آوری و سپس با استفاده از روش‌های یادگیری ماشین و سایر روش‌ها در مورد اعتبار اخبار منتشرشده تصمیم‌گیری و اظهارنظر کنند. در این مقاله، به‌طورخلاصه مطالعاتی را که به‌منظور توسعهٔ سامانه‌های دسته‌بندی شایعات انجام شده‌اند، بیان شد، طیف وسیعی از مطالعات، روش‌های مختلفی را برای درک و مشخص‌کردن شایعات اجتماعی در نظر گرفته‌اند و این تنوع کمک می‌کند تا چالش‌های پیش رو برای ساخت سامانه‌های کارآمد در تشخیص شایعات در آینده روشن‌تر شود. این حوزه همچنان به‌عنوان یک موضوع باز پژوهشی نیاز به مطالعات بیشتری دارد. چالش‌هایی در پژوهش‌های انجام‌شده در حوزهٔ تشخیص شایعات وجود دارد که بعضی از آن‌ها عبارت‌اند از:

۱) **تشخیص سریع**: چرخهٔ زندگی یک داستان در شبکه‌های اجتماعی بسیار کوتاه است. بیشتر شایعات در عرض چند ثانیه یا چند دقیقه به‌طور ویروسی منتشر می‌شوند. مهم است که شایعات را در مراحل اولیهٔ خود تشخیص داده شوند؛ اما به‌دلیل محدودبودن منابع در آغاز یک شایعه، تشخیص شایعات در مراحل اولیه بسیار دشوار است [۳].

۲) **تشخیص شایعه در متون طولانی**: در حال‌حاضر روش‌های تشخیص شایعات در رسانه‌های اجتماعی برروی متون کوتاه متمرکز است. بااین‌حال، بیشتر اخبار شامل متون طولانی می‌شوند، که نیاز به تأیید و تصدیق دارند. برخلاف شایعات با متن کوتاه، شایعات طولانی، اطلاعات معنادار زیادی دارند که مانع فهم و درک آسان آن می‌شود؛ علاوه‌براین، در اغلب موارد، تنها بخشی از یک شایعهٔ متنی طولانی حاوی اطلاعات غلط است، درحالی‌که بقیهٔ متن درست است؛ بنابراین غیرمنصفانه است که تمام متن به‌عنوان یک شایعه دسته‌بندی شود. بنابراین یک سامانه خوب در تشخیص شایعه در متون طولانی، سامانه‌ای است که علاوه‌بر بهبود عملکرد سامانه‌های گذشته، موقعیت دقیق اطلاعات نادرست متن را نیز تشخیص دهد [۸].

۳) **تشخیص شایعات در داده‌های چندوجهی**[3]: بیش‌تر شایعاتی که در رسانه‌های اجتماعی منتشر می‌شوند، از ترکیب داده‌های متنی، تصویری و ویدئویی تشکیل‌شده‌اند.

[3] MultiModal

[DOR: 20.1001.1.24763047.1398.8.1.1.7]

این نوع شایعات مشکلاتی را برای روش‌های تشخیص به ارمغان می‌آورند. بنابراین، تجزیه و تحلیل روابط بین داده‌ها با روش‌های متعدد و توسعه مدل‌های پیشرفته مبتنی بر ترکیب این داده‌ها می‌تواند کلید شناسایی شایعات در سناریوهای پیچیده‌تر باشد [۲۴].

۴) **عدم استنتاج معانی توسط روش‌های موجود**: به علّت اتّکا به روش‌های آماری و عدم توجه به معانی نهفته در متون و جملات، سامانه‌های تشخیص شایعهٔ موجود در این زمینه کارا نیستند و دقت بالایی برای تشخیص ندارند هرچند این روش‌ها می‌توانند بُردار مفهومی از واژگان و جملات با توجه به جایگاه آن‌ها در متن آموزشی ایجاد کنند؛ اما در اینجا منظور از درک واژگان استفاده از روش‌های تعبیه واژگان نیست. روش‌های تعبیه واژگان قادر به استنتاج مفاهیم نیستند. به‌عنوان نمونه واژگان خوب و بد دارای بردارهای شبیه هم خواهند شد، چون در داده آموزشی جایگاه استفاده مشابهی دارند. منظور از چالش یادشده استفاده از روش‌هایی است که قادر به درک بالاتر مفاهیم و استنتاج نتیجه باشد. یکی از موارد مطرح در این روش‌ها، روش‌های استنتاج در زبان طبیعی[1] است، که قادر است درستی یک جمله (تالی[2]) را با استفاده از جملهٔ دیگری که به اصلاح مقدم[3] نامیده می‌شود استنتاج کند. هرچند این روش‌ها نیز در مراحل اولیه خود از روش‌های تعبیه کلمات استفاده می‌کنند، اما از آن‌ها نمی‌توان به‌عنوان روش‌های برداری‌سازی نام برد. استفاده از چنین روش‌هایی که جز دستاوردهای اخیر در حوزه پردازش زبان طبیعی[4] و یادگیری عمیق است، در شناسایی شایعات مغفول مانده است و سامانه‌های جدید باید به‌گونه‌ای طراحی شوند که بتوان معانی نهفتهٔ درون واژگان و جملات را از آن استنباط کرد [۳, ۸].

۶- مراجع

[1] "earthquake's of tehran." yon.ir/NjPfw (accessed.

[2] N. DiFonzo and P. Bordia, Rumor psychology: Social and organizational approaches. American Psychological Association Washington, DC, 2007.

[3] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, and R. Procter, "Detection and resolution of rumours in social media: A survey," *ACM Computing Surveys (CSUR)*, vol. 51, no. 2, pp. 32, 2018.

[4] Z. Jin, J. Cao, Y.-G. Jiang, and Y. Zhang, "News credibility evaluation on microblog with a hierarchical propagation model," *in Data Mining (ICDM), 2014 IEEE International Conference on,* 2014: IEEE, pp. 230-239.

[5] Z. Jin, J. Cao, Y. Zhang, and J. Luo, "News Verification by Exploiting Conflicting Social Viewpoints in Microblogs," in *AAAI*, 2016, pp. 2972-2978.

[6] W. Chen, C. K. Yeo, C. T. Lau, and B. S. Lee, "Behavior deviation: An anomaly detection view of rumor preemption," *in Information Technology, Electronics and Mobile Communication Conference* (IEMCON), 2016 IEEE 7th Annual, 2016: IEEE, pp. 1-7.

[7] N. Ruchansky, S. Seo, and Y. Liu, "Csi: A hybrid deep model for fake news detection," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management,* 2017: ACM, pp. 797-806.

[8] J. Cao, J. Guo, X. Li, Z. Jin, H. Guo, and J. Li, "Automatic Rumor Detection on Microblogs: A Survey," arXiv preprint arXiv:1807.03505, 2018.

[9] J. Ma et al., "Detecting Rumors from Microblogs with Recurrent Neural Networks," in *IJCAI*, 2016, pp. 3818-3824.

[10] M. Wang, B. Ni, X.-S. Hua, and T.-S. J. A. C. S. Chua, "Assistive tagging: A survey of multimedia tagging with human-computer joint exploration," vol. 44, no. 4, p. 25, 2012.

[11] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on twitter," in *Proceedings of the 20th international conference on World wide web*, 2011: ACM, pp. 675-684.

[12] A. Gupta, H. Lamba, P. Kumaraguru, and A. Joshi, "Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy," in *Proceedings of the 22nd international conference on World Wide Web*, 2013: ACM, pp. 729-736.

[13] K. Wu, S. Yang, and K. Q. Zhu, "False rumors detection on sina weibo by propagation structures," in *Data Engineering (ICDE) 2015 IEEE 31st International Conference on*, 2015: IEEE, pp. 651-662.

[14] Z. Zhao, P. Resnick, and Q. Mei, "Enquiring minds: Early detection of rumors in social media from enquiry posts," in *Proceedings of the 24th International Conference on World Wide Web, 2015: International World Wide Web Conferences Steering Committee*, pp. 1395-1405.

[15] Z. Jin, J. Cao, Y. Zhang, and Y. Zhang, "MCG-ICT at MediaEval 2015: Verifying Multimedia Use with a Two-Level Classification Model," in MediaEval, 2015.

[16] Z. Zhang, Z. Zhang, and H. Li, "Predictors of the authenticity of Internet health rumours," *Health Information & Libraries Journal*, vol. 32, no. 3, pp. 195-205, 2015.

[17] Y. Yang, K. Niu, and Z. He, "Exploiting the topology property of social network for rumor detection," in *Computer Science and Software*

[1] Natural Language Inference
[2] Premise
[3] Hypothesis
[4] Natural Language Processing

[DOR: 20.1001.1.24763047.1398.8.1.1.7]

Engineering (JCSSE), 2015 12th International Joint Conference on, 2015: IEEE, pp. 41-46.

[18] B. Ma, D. Lin, and D. Cao, "Content representation for microblog rumor detection," in *Advances in Computational Intelligence Systems*: Springer, 2017, pp. 245-251.

[19] S. Zamani, M. Asadpour, and D. Moazzami, "Rumor detection for Persian Tweets," *in Electrical Engineering (ICEE), 2017 Iranian Conference on*, 2017: IEEE, pp. 1532-1536.

[20] S. D. Mahmoodabad, S. Farzi, and D. B. Bakhtiarvand, "Persian Rumor Detection on Twitter," in *2018 9th International Symposium on Telecommunications (IST)*, 2018: IEEE, pp. 597-602.

[21] M. Gupta, P. Zhao, and J. Han, "Evaluating event credibility on twitter," in *Proceedings of the 2012 SIAM International Conference on Data Mining*, 2012: SIAM, pp. 153-164.

[22] T. Chen, L. Wu, X. Li, J. Zhang, H. Yin, and Y. Wang, "Call Attention to Rumors: Deep Attention Based Recurrent Neural Networks for Early Rumor Detection," arXiv preprint arXiv: 1704.05973, 2017.

[23] J. Ma, W. Gao, and K.-F. Wong, "Rumor detection on twitter with tree-structured recursive neural networks," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2018, vol. 1, pp. 1980-1989.

[24] Z. Jin, J. Cao, H. Guo, Y. Zhang, and J. Luo, "Multimodal fusion with recurrent neural networks for rumor detection on microblogs," in *Proceedings of the 2017 ACM on Multimedia Conference*, 2017: ACM, pp. 795-816.

[25] S. Kwon, M. Cha, K. Jung, W. Chen, and Y. Wang, "Prominent features of rumor propagation in online social media," in *2013 IEEE 13th International Conference on Data Mining*, 2013: IEEE, pp. 1103-1108.

[26] R. Sicilia, S. L. Giudice, Y. Pei, M. Pechenizkiy, and P. Soda, "Health-related rumour detection on Twitter," in *Bioinformatics and Biomedicine (BIBM), 2017 IEEE International Conference on*, 2017: IEEE, pp. 1599-1606.

[27] W. Chen, Y. Zhang, C. K. Yeo, C. T. Lau, and B. S. J. P. R. L. Lee, "Unsupervised rumor detection based on users' behaviors using neural networks," vol. 105, pp. 226-233, 2018.

[28] S. Kwon, M. Cha, and K. Jung, "Rumor detection over varying time windows," *PloS one*, vol. 12, no. 1, p. e0168344, 2017.

[29] N. Xu, G. Chen, and W. Mao, "MNRD: A merged neural model for rumor detection in social media," in *2018 International Joint Conference on Neural Networks (IJCNN)*, 2018: IEEE, pp. 1-7.

[30] J. Ma, W. Gao, and K.-F. Wong, "Detect rumor and stance jointly by neural multi-task learning," in *Companion of the The Web Conference 2018 on The Web Conference 2018*, 2018, pp. 585-593.

[31] T. Chen, H. Chen, and X. Li, "Rumor Detection via Recurrent Neural Networks: A Case Study on Adaptivity with Varied Data Compositions," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2018: Springer, pp. 121-127.

[32] F. Yu, Q. Liu, S. Wu, L. Wang, and T. Tan, "A convolutional approach for misinformation identification," in *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, 2017: AAAI Press, pp. 3901-3907.

[33] T. N. Nguyen, C. Li, and C. Niederée, "On early-stage debunking rumors on twitter: Leveraging the wisdom of weak learners," in *International Conference on Social Informatics*, 2017: Springer, pp. 141-158.

[34] Z. Arkaitz, W. S. H. Geraldine, L. Maria, and P. Rob, PHEME dataset of rumours and non-rumours. 2016.

[35] K. Elena, L. Maria, and Z. Arkaitz, PHEME dataset for Rumour Detection and Veracity Classification. 2018.

[36] L. Derczynski, K. Bontcheva, M. Liakata, R. Procter, G. W. S. Hoi, and A. Zubiaga, "SemEval-2017 Task 8: RumourEval: Determining rumour veracity and support for rumours," arXiv preprint arXiv:1704.05972, 2017.

[37] D. Leon et al., RumourEval 2019 data. 2019.

[38] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "Fakenewsnet: A data repository with news content, social context and dynamic information for studying fake news on social media," arXiv preprint arXiv:1809.01286, 2018.

فریبا صادقی مدرک کارشناسی خود را در رشته مهندسی فناوری اطلاعات در دانشگاه کردستان در سال ۱۳۹۵ اخذ کرده و اکنون دانشجوی ارشد رشته مهندسی فناوری اطلاعات گرایش تجارت الکترونیک در دانشگاه قم است. از جمله زمینه‌های پژوهشی مورد علاقه او می‌توان به امنیت نرم، تحلیل شبکه های اجتماعی، پردازش زبان طبیعی و یادگیری ماشین اشاره کرد.

امیرجلالی بیدگلی تحصیلات خود را در مقاطع کارشناسی ارشد و دکترای مهندسی کامپیوتر (نرم‌افزار) به‌ترتیب در سال ۱۳۸۸ در دانشگاه علم و صنعت ایران و ۱۳۹۴در دانشگاه اصفهان به پایان رساند. وی از سال ۱۳۹۵ به عضویت هیئت علمی دانشگاه قم در آمد و هم اکنون استادیار گروه مهندسی کامپیوتر این دانشگاه است. زمینه‌های پژوهشی مورد علاقه ایشان مشتمل بر اعتماد محاسباتی، امنیت نرم‌افزار، پروتکل‌های رمزنگاری و تحلیل شبکه‌های اجتماعی است.

[DOR: 20.1001.1.24763047.1398.8.1.1.7]