سال سیزدهم، شماره‌های ۴۷ و ۴۸، بهار و تابستان ۱۴۰۰

صص: ۱۳۶_۱۵۹

دو فصلنامه علمی

فناوری اطلاعات و ارتباطات ایران

استفاده از شبکه مولد متخاصم شرطی برای تولید داده مصنوعی با هدف بهبود کلاس‌بندی کاربران منتشرکننده اخبار جعلی

عارفه اسمعیلی* سعید فرضی**

*کارشناس ارشد نرم افزار، گروه مهندسی کامپیوتر، دانشگاه صنعتی خواجه نصیر الدین طوسی

** استادیار گروه مهندسی کامپیوتر، دانشگاه صنعتی خواجه نصیر الدین طوسی

تاریخ دریافت:۱۳۹۹/۰۷/۱۹ تاریخ پذیرش:۱۳۹۹/۱۲/۰۲

نوع مقاله: پژوهشی

چکیده

سالیان درازی است که اخبار و پیام‌های جعلی در جوامع انسانی منتشر می‌گردد و امروزه با فراگیرشدن شبکه‌های اجتماعی در بین مردم، امکان نشر اطلاعات نادرست بیشتر از قبل شده است. بنابراین، شناسایی اخبار و پیام‌های جعلی به موضوع برجسته‌ای در جوامع تحقیقاتی تبدیل شده است. ضمناً، شناسایی کاربرانی که این اطلاعات نادرست را ایجاد می‌کنند و در شبکه نشر می‌دهند، از اهمیت بالایی برخوردار است. این مقاله، به شناسایی کاربرانی که با زبان فارسی اقدام به انتشار اطلاعات نادرست در شبکه اجتماعی توئیتر می‌کنند، پرداخته است. در این راستا، سیستمی بر مبنای ترکیب ویژگی‌های بافتار-کاربر و بافتار-شبکه با کمک شبکه مولد متخاصم شرطی برای متوازن‌سازی مجموعه داده پایه‌ریزی شده است. هم‌چنین، این سیستم با مدل‌کردن شبکه اجتماعی توئیتر به گراف تعاملات کاربران و تعبیه گره به بردار ویژگی توسط Node۲vec ، کاربران منتشرکننده اخبار جعلی را شناسایی می‌کند. علاوه بر این، با انجام آزمایشات متعدد، سیستم پیشنهادی تا حدود ۱۱٪ ، ۱۳ ٪ ،۱۲ ٪ و ۱۲ ٪ به ترتیب در معیارهای دقت، فراخوانی، معیار اف و صحت نسبت به رقبایش بهبود داشته است و توانسته است دقتی در حدود ۹۹٪ در شناسایی کاربران منتشرکننده اخبار جعلی ایجاد کند.

کلید واژگان: شناسایی کاربر منتشرکننده اخبار جعلی، مجموعه داده‌های نامتوازن، شبکه مولد متخاصم، گراف تعاملات کاربران، تعبیه گره.

۱ مقدمه

از زمان‌های گذشته تاکنون اطلاعات و پیام‌های جعلی همواره وجود داشته است [۱]، که برای جوامع بشری مشکلات فراوانی ایجاد کرده است [۲].

در نتیجه، همواره اطمینان از درستی خبر درجوامع بشری احساس شده است [۳]. امروزه نیز با پیشرفت و گسترش شبکه‌های اجتماعی و دسترسی آسان به آن‌ها، شبکه‌های اجتماعی به پلتفرم مناسبی برای دنبال‌کردن رخداد و اخبار جهان تبدیل شده‌اند [۴]. علاوه بر این، در این شبکه‌ها اجازه انتشار اطلاعات متنوع و زیاد، بدون چک کردن اعتبار[۱] آن‌ها داده می‌شود [۵]. بنابراین، کاربران می‌توانند با

نویسنده مسئول: عارفه اسمعیلی

arefehesmaili@email.kntu.ac.ir

[۱] Credibility

ایجاد حساب جعلی[۲]، انواع جدیدی از اطلاعات مخرب[۳] و نادرست را در شبکه‌های اجتماعی تولید و منتشر کنند. به طور مثال، هرزنامه‌ها[۴] نوعی فعالیت مخرب هستند که کاربران جعلی[۵] پیام‌های ناخواسته‌ای را به صورت پیام‌های کلاهبرداری، پیام‌هایی شامل ویروس و غیره ازطریق آن‌ها ارسال می‌کنند [۶]. اکثر اخبار جعلی در زمینه‌های مسائل اعتقادی، اقتصادی و سیاسی وجود دارد [۷]. برای اشاره به نمونه‌ای از این نوع فعالیت‌ها، می‌توان به انتخابات آمریکا در سال ۲۰۱۶ اشاره کرد که مطالعه منابع خبری جعلی در آخرین هفته انتخابات توسط مردم، روی نتایج انتخابات اثرگذار بوده است [۸]. همانطور که مشهود است، این پیام‌ها اعتبار شبکه‌های اجتماعی را کاهش می‌دهد و امنیت کاربران و حریم شخصی آن‌ها را نیز تحت تاثیر خود قرار می‌دهد [۱۰]. بنابراین، شناسایی اخبار و پیام‌های جعلی در بین جوامع تحقیقاتی به موضوع برجسته‌ای تبدیل شده است. شبکه‌های اجتماعی آنلاین[۶] مانند توئیتر، فیس‌بوک و لینکدین و غیره به دلیل فراگیری و استفاده بیشتر از آن‌ها در بین مردم نسبت به گذشته تبدیل به بستری برای انتشار اطلاعات و اخبار نادرست شده است [۱۱]. ضمناً، توئیتر به علت تبدیل‌شدن به مجرایی برای انتشار اخبار بلادرنگ در بین دولتمردان و افراد تحصیل‌کرده، پلتفرم مناسبی برای انتشار اخبار جعلی شده است[۷]. ضمناً، چون اکثرکاربران توئیتر اقدام به تبادل اطلاعات با زبان انگلیسی می‌کنند، بیشتر تحقیقات بر روی این زبان صورت گرفته است [۱۲] و از توجه به زبان‌های مهم دیگری مانند فارسی که منابع زبان‌شناسی‌کمتری برای آن‌ها وجود دارد، غفلت شده است.

علی‌رغم اینکه مطالعات انجام‌شده در حوزه شناسایی اخبار جعلی معمولاً بر روی متن خبر انجام شده است، شناسایی کاربر منتشرکننده این اخبار نیز از اهمیت ویژه‌ای برخوردار است [۶]. کاربران منتشرکننده اخبار جعلی در این مقاله، حساب کاربری هستند که حداقل یک بار پیامی حاوی خبر جعلی در شبکه اجتماعی منتشر کرده‌اند. در این مقاله، یک سیستم پیشنهادی برای شناسایی کاربران منتشرکننده اخبار جعلی مبتنی بر ترکیب ویژگی‌های مبتنی بر کاربر-شبکه پیشنهاد داده شده است. از چالش‌هایی که در این حوزه وجود داشت، می‌توان به ۱) عدم توازن کلاس‌ها در مجموعه داده و ۲) معرفی سیستمی برای شناسایی کاربران جعلی از کاربران عادی اشاره کرد. برای رفع چالش اول، از روش یادگیری عمیق، شبکه مولد متخاصم شرطی[۸] برای متوازن‌سازی مجموعه داده استفاده شده است و برای رفع چالش دوم، با مدل‌کردن شبکه اجتماعی توئیتر به گراف وزن‌دار و جهت‌دار و ترکیب اطلاعات کاربران با ویژگی‌هایی که باتعبیه[۹] گره به بردار ویژگی (Node۲vec) به دست می‌آیند، کلاس-بندی برای دسته‌بندی کاربران منتشرکننده اخبار جعلی از کاربران عادی طراحی شده است. ضمناً، در این مقاله از مجموعه داده جمع-آوری‌شده توئیتر فارسی در بازه دو هفته‌ای مدت وقوع زلزله کرمانشاه ایران در سال ۱۳۹۶ استفاده شده است، که با برچسب‌گذاری دستی[۱۰] داده‌ها، توسعه‌داده شده است. به کمک انجام آزمایشات مختلف و متنوع بر روی مجموعه داده که با اهداف معینی صورت گرفته است، برتری سیستم پیشنهادی در مقایسه با رقبای خود چون روش بیش نمونه‌برداری اقلیت مصنوعی[۱۱]، Borderline-SMOTE، ADASYN، CCR، Cure-SMOTE، Gaussian-SMOTE، Random-SMOTE، Distance-SMOTE، Cluster-SMOTE و غیره در معیارهای ارزیابی چون صحت[۱۲]، فراخوانی[۱۳]، معیار اف[۱۴] و دقت[۱۵] نشان داده شده است.

نوآوری مقاله ما به صورت زیر خواهد بود:

- گسترش مجموعه داده فارسی در شبکه اجتماعی توئیتر، برای شناسایی کاربران منتشرکننده اخبار جعلی
- معرفی سیستمی برای کلاس‌بندی کاربران منتشرکننده اخبار جعلی و کاربران عادی

در بخش بعدی دسته‌بندی برکارهای گذشته در این حوزه انجام شده است. در ادامه، خلاصه‌ای از شبکه مولد متخاصم و شبکه مولد متخاصم شرطی و روش تعبیه گره (Node۲vec) ارائه خواهد شد. هم‌چنین، در بخش ۳ سیستم پیشنهادی مقاله و در بخش ۴ آزمایشات تکمیلی و نتایج ارزیابی نمایش داده شده است. نهایتاً، به ترتیب در بخش ۵ و ۶ نتیجه‌گیری و مراجع استفاده شده، شرح داده شده است.

۲ کارهای مرتبط و پیش‌زمینه

۲,۱ پیش‌زمینه

در این بخش از مقاله، خلاصه‌ای از شبکه مولد متخاصم و شبکه مولد متخاصم شرطی و سپس، روش تعبیه گره به بردار (Node۲vec) به‌طورخلاصه تشریح می‌گردد.

❖ **شبکه مولد متخاصم و شبکه مولد متخاصم شرطی:**

[۲] Fake account
[۳] Malicious
[۴] Spam
[۵] Fake users
[۶] Online Social Network (OSN)
[۷] https://blog.pixelfish.com.au/twitter-vs-facebook-vs-instagram-vs-linkedin
[۸] Conditional generative adversarial network (CGAN)
[۹] Embedding
[۱۰] Manual
[۱۱] Synthetic Minority Oversampling Technique (SMOTE)
[۱۲] Accuracy
[۱۳] Recall
[۱۴] F-measure
[۱۵] Precision

شبکه مولد متخاصم بر مبنای رقابت بین دو مؤلفه تولیدکننده G و تمایزدهنده D پایه‌ریزی شده است. هدف G فریب دادن D است. هدف D ایجاد تمایز بین نمونه‌های تولیدی G و نمونه‌های موجود در مجموعه داده است. هر دو مؤلفه سعی در باهوش کردن یکدیگر دارند. با بازخورد گرفته شده از نمونه‌های تولیدی G توسط D، عملکرد G بهبود می‌یابد. هم‌چنین، اگر D بتواند به راحتی نمونه‌های واقعی را از نمونه‌های تولیدی G تشخیص دهد، G کیفیت نمونه‌های تولیدی خود را کاهش می‌دهد. مؤلفه تولیدکننده G به صورت $G: Z \rightarrow X$ تعریف می‌شود که Z فضای اختلال[۱۶] با بعد دلخواه d_z است و هم‌چنین، X فضای داده است که هدف G به دست آوردن توزیع داده است. مؤلفه تمایزدهنده به صورت $D: X \rightarrow [۰, ۱]$ تعریف می‌شود و احتمال اینکه نمونه از مجموعه داده یا از G می‌آید، را تخمین می‌زند. این دو مؤلفه در یک بازی کمینه-بیشینه[۱۷] مطابق رابطه ۱ با هم به رقابت می‌پردازند:

$$min_G max_D V(D.G) = E_D + E_G \qquad (۱)$$

به طوریکه:

$$E_D = E_{X \sim P_{data(x)}}[log \mathrm{D}(\mathrm{x})]$$

$$E_G = E_{Z \sim P_{z(Z)}}[\log(۱ - \mathrm{D}(\mathrm{G}(\mathrm{z})))]$$

$x \in X$ مقادیری هستند که از توزیع داده $P_{data(x)}$ نمونه گرفته شده‌اند و مقادیر $z \in Z$ از توزیع اختلال $P_{z(Z)}$ می‌آیند. مرحله آموزش شامل k مرحله آموزش D و یک مرحله آموزش G است. D در طول آموزش یاد می‌گیرد که به نمونه‌های داده واقعی برچسب یک و به نمونه‌های تولیدی G برچسب صفر دهد. هم‌چنین، G با دادن برچسب یک به داده‌های تولیدی خود سعی دارد D را فریب دهد. با یادگیری توزیع داده توسط تولیدکننده و رسیدن متمایز-کننده به نصف برای هر مقدار ورودی، این بازی خاتمه می‌یابد [۱۳].

شبکه مولد متخاصم شرطی نوع توسعه داده شده شبکه مولد متخاصم است که یک فضای اضافی Y دارد که در آن اطلاعات اضافی از مجموعه داده آموزشی به ساختار فوق اضافه می‌شود و روی نمونه-های تولیدی G شرط می‌گذارد. در این چارچوب فضای جدید Y به هر دو مؤلفه اضافه می‌گردد. به طوریکه، مؤلفه G، $G: Z*Y \rightarrow X$ و مؤلفه D، $D: X*Y \rightarrow [۰, ۱]$ تعریف خواهد شد.

پارامترهای رابطه ۱ به صورت رابطه ۲ بازنویسی می‌گردد:

$$E_D = E_{X.Y \sim P_{data(x.y)}}[log \mathrm{D}(\mathrm{x.y})]$$

$$E_G = E_{Z \sim P_{z(Z)}.y \sim P_{(y)}}[\log(۱ - \mathrm{D}(\mathrm{G}(\mathrm{z}.\mathrm{y}).y))] \qquad (۲)$$

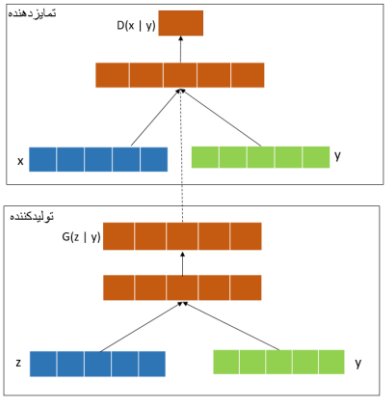
مرحله آموزشی در هر دو شبکه مشابه یکدیگر است و تابع هزینه با m دسته کوچک[۱۸] از نمونه‌های آموزشی و به همین تعداد از نمونه-های فضای اختلال به روز می‌شوند. تابع هزینه مؤلفه تمایزدهنده در رابطه ۳ نشان داده شده است.

$$J_D = -\frac{۱}{۲m}\Big(\sum_{i=۱}^{m} \log \mathrm{D}(x_i.y_i) + \sum_{i=۱}^{m} \log(۱ - D(G(z_i.y_i).y_i))\Big) \qquad (۳)$$

برای جلوگیری از اشباع[۱۹] تمایزدهنده، تابع هزینه تولیدکننده به صورت رابطه ۴ درنظر گرفته خواهد شد [۱۴].

$$J_G = -\frac{۱}{m}\Big(\sum_{i=۱}^{m} \log \mathrm{D}(G(z_i.y_i).y_i)\Big) \qquad (۴)$$

با به‌روزرسانی‌های چرخشی مبتنی بر شیب[۲۰] بین دو رابطه ۳ و ۴ شبکه مولد متخاصم شرطی آموزش می‌بینند. معماری ساده‌ای از شبکه مولد متخاصم شرطی در شکل ۱ قابل مشاهده است [۱۵].

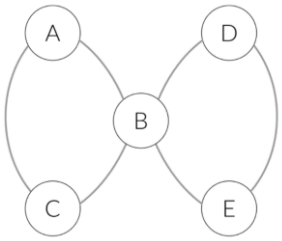


شکل ۱. شبکه مولد متخاصم شرطی [۱۵]

❖ **Node۲vec:** Node۲vec روش یادگیری نیمه‌ناظر برای تعبیه[۲۱] گره به نقاطی در فضای برداری با بعد کمتر با حفظ بیشترین همسایگی است. این روش دو معادله[۲۲] هموفیلی و ساختاری را در نظر می‌گیرد. در معادلات هموفیلی[۲۳] گره‌ها می‌توانند مبتنی بر جامعه‌ای[۲۴] که به آن تعلق دارند، سازماندهی شوند و در معادله ساختاری[۲۵] گره‌ها می‌توانند براساس نقش ساختاری خود در شبکه، سازماندهی شوند. به طور مثال، در شکل۲ گره E , C در

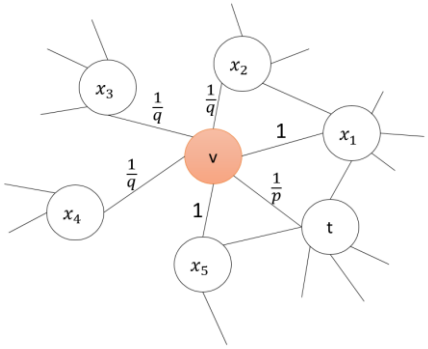
[۱۶] Noise
[۱۷] Min-Max
[۱۸] Mini-batch
[۱۹] Saturation
[۲۰] Gradient-Based
[۲۱] Embedding
[۲۲] Equivalence
[۲۳] Hemophilia
[۲۴] Community
[۲۵] Structural

دو جامعه مجزا قرار دارند ولی نقش ساختاری یکسانی دارند. گره‌هایC , A به یک جامعه تعلق دارند.



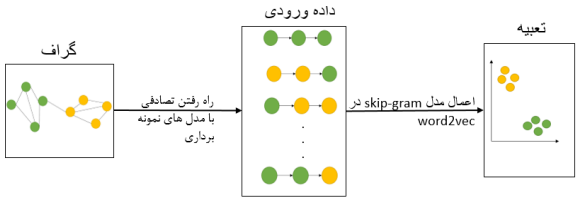
شکل ۲. نمونه گراف

این روش بر روی دو هدف تمرکز دارد. هدف اول آن، تعبیه گره‌هایی که به یک جامعه تعلق دارند، در نزدیکی یکدیگر است و هم‌چنین، هدف دوم آن، تعبیه گره‌ها با نقش ساختاری یکسان در گراف در نزدیکی یکدیگر است. بنابراین، برای تحقق این دو هدف، Node۲vec با وزن‌دهی یال‌های گراف به صورت شکل ۳، و با پیاده‌روی تصادفی[۲۶] روی گراف و با ترکیب جستجوی اول سطح[۲۷] برای معادلات ساختاری و با جستجوی اول عمق[۲۸] برای معادلات هموفیلی، دنباله‌ای از گره‌ها در گراف ایجاد می‌کند که شبیه به دنباله‌ای از کلمات در جمله خواهد بود. سپس، همانطور که در شکل۴ مشهود است با کمک ابزار Word۲vec و بهره‌گیری از Skip-gram دنباله ایجادشده را به بردار ویژگی تبدیل می‌کند [۱۶].



$$\alpha_{pq}(t,x)=\begin{cases}\frac{1}{p} & \text{اگر } d_{tx}=0 \text{ آنگاه}\\ 1 & \text{اگر } d_{tx}=1 \text{ آنگاه}\\ \frac{1}{q} & \text{اگر } d_{tx}=2 \text{ آنگاه}\end{cases}$$

شکل ۳. نحوه وزن‌دهی به یال‌ها. فرض شده است که در پیاده‌روی تصادفی از گره t به v رفته شده است، حال باید مشخص شود از گره v به کدام گره خواهد رفت که طبق معادله وزن‌دهی می‌شود و جهت حرکت مشخص خواهد شد [۱۶].



شکل ۴. مراحل Node۲vec

۲,۲ کارهای مرتبط

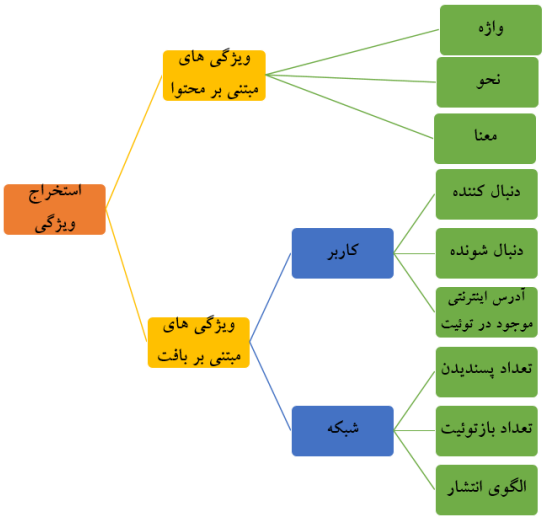
اولین سایت اجتماعی با نام Six degree.com در سال ۱۹۹۷ میلادی شروع به کارکرد ولی خیلی زود کنار گذاشته شد [۱۰]. بعد از آن شبکه‌های اجتماعی چون فیس‌بوک، لینکدین، اینستاگرام، توئیتر و غیره برای برقراری ارتباط کاربران سراسر جهان با یکدیگر، یافتن اخبار و به اشتراک‌گذاری رویدادها به صورت تصویر، متن، ویدئو و غیره پا به عرصه ظهور گذاشتند. از طرفی با گسترش و فراگیری این شبکه‌ها در بین مردم، شبکه‌های اجتماعی نوظهور تبدیل به پلتفرم مناسبی برای انتشار اطلاعات غلط، لینک‌های هرزنامه، پیام‌های ناخواسته و ساخت حساب‌های جعلی شده‌اند [۷].

اخبار جعلی عمداً برای فریب‌دادن خواننده نوشته می‌شوند، که نادرستی آن‌ها توسط منابع موثق قابل اثبات است [۱۷]. اما شایعات اطلاعاتی هستند که درستی آن‌ها توسط منبع رسمی تایید نشده است و در حال پخش‌شدن در بین افراد هستند [۱۸]. کاربران مخرب، به دنبال نقض حریم خصوصی کاربران دیگر یا سوء استفاده از نام و اعتبار آن‌ها با ایجاد حساب جعلی هستند [۱۹]. توئیتر یکی از رایج‌ترین وب سایت‌هایی است که میکروبلاگینگ رایگان شامل ارسال تصویر، ویدئو، متن و غیره را در اختیار کاربرانش قرار داده است [۲۰]. کاربران توئیتر برای تبادل اطلاعات می‌توانند از پیام‌های کوتاهی شامل حداکثر ۲۸۰ کاراکتر که به آن‌ها توئیت[۲۹] گفته می‌شود، استفاده کنند [۲۱]. ضمناً، ارتباطات جهت‌دار خواهد بود، یعنی هر کاربر دنبال‌کننده[۳۰] و دنبال‌شونده[۳۱] خود را دارد. هم‌چنین، توئیت می‌تواند در شبکه بازنشر شود که به آن ریتوئیت[۳۲] گویند. ضمناً، می‌توان در جواب توئیتی پاسخی گذاشت. کاربران توئیتر معمولاً از هشتگ برای مشخص‌کردن موضوع خاص در توئیت خود استفاده می‌کنند. هشتگ‌های مشهور به موضوعات روز[۳۳] تبدیل می‌شوند [۲۲].

مطالعات گسترده‌ای در شبکه اجتماعی توئیتر برای شناسایی اقدامات فریبکارانه مبتنی بر آدرس اینترنتی، محتوای جعلی، شناسایی کاربر جعلی، استخراج هرزنامه در موضوعات روز انجام شده است [۲۳].

[۲۶] Random walk
[۲۷] Breadth First Search (DFS)
[۲۸] Depth First Search (BFS)
[۲۹] Tweet
[۳۰] Follower
[۳۱] Following
[۳۲] Retweet
[۳۳] Trending Topic

محققان در ابتدا مدلی پیشنهاد دادند که هرزنامه‌ها را از طریق آدرس اینترنتی آن‌ها فیلتر می‌کرد. به همین منظور، توئیتر توسط Bot Maker امکان حذف هرزنامه‌ها توسط آدرس اینترنتی آن‌ها را فراهم آورد. اما محققان دریافتند که ۹۰ درصد هرزنامه‌ها با استفاده از آدرس اینترنتی جدید از فیلتر گذر می‌کردند که باعث شکست این ایده شد. اخیراً، محققان به دنبال روش‌هایی برای استفاده از الگوریتم‌های مبتنی بر یادگیری ماشین هستند [۱۰]. شناسایی کاربران و اخبار جعلی یک کلاس‌بندی شامل دو کلاس است که شامل دسته‌بندی کاربران و اخبار، به جعلی و عادی است. کارهای انجام شده درحوزه استخراج ویژگی‌های موردنیاز برای کلاس‌بندها به دو دسته ۱) مبتنی بر محتوا[۳۴] و ۲) مبتنی بر بافتار[۳۵] تقسیم می‌گردد. در ویژگی‌های مبتنی بر محتوا به قواعد زبان‌شناسی مانند نحو[۳۶]، معنا[۳۷]، واژه[۳۸] در متن توجه می‌شود. از آنجا که ساختار زبان-شناسی متن عادی می‌تواند توسط متن جعلی تقلید شود [۲۴]، ویژگی‌های مبتنی بر بافتار به روی کار آمدند. هم‌چنین، این ویژگی نیز شامل دو دسته مبتنی بر شبکه[۳۹] و مبتنی بر کاربر[۴۰] است. در ویژگی‌های مبتنی بر بافتار-کاربر به ویژگی‌های آماری چون شماره حساب، آدرس اینترنتی موجود در توئیت، عکس پروفایل کاربران، تعداد پست ایجاد شده توسط کاربر، تعداد دنبال‌کننده و دنبال‌شونده، سن و غیره توجه می‌شود. امکان تقلید و جعل در این نوع ویژگی‌ها نیز به کمک ایجاد پست و خرید دنبال‌کننده و غیره وجود دارد. برای جلوگیری از این موضوع، از ویژگی‌های مبتنی بر بافتار - شبکه مانند الگوی انتشار، چگالی، ضریب خوشه‌بندی[۴۱]، تعداد ریتوئیت، دفعات انتشار پست، پسندیدن[۴۲] یک پست و تعاملات کاربر با خبر و غیره می‌توان استفاده‌کرد. دسته‌بندی انواع ویژگی‌ها برای شناسایی اخبار و کاربر جعلی در شکل۵ آمده است.



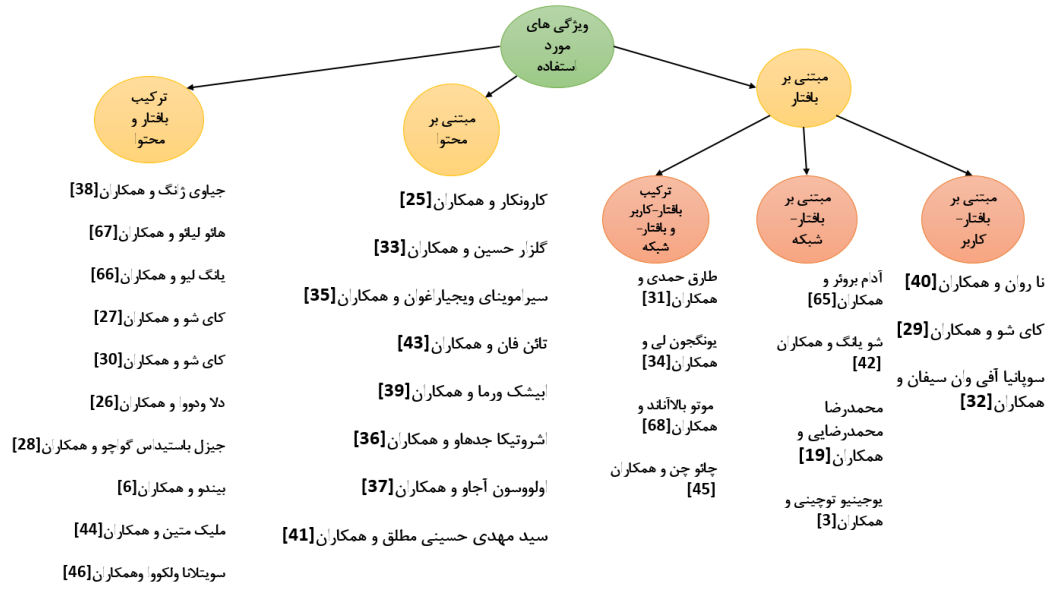
شکل ۵ . گونه شناسی انواع ویژگی‌ها و مثال‌هایی از هر گروه برای شناسایی اخبار و کاربر جعلی

کارونکار و همکارانش [۲۵]، برای شناسایی پروفایل کاربران جعلی در فیس‌بوک از روش‌های زبان‌شناسی استفاده کرده‌اند که امکان تقلید این نوع ویژگی توسط کاربران جعلی وجود دارد. برای جلوگیری از این موضوع، در این مقاله، به ویژگی‌های مبتنی بر شبکه و بافتار-کاربر توجه شده است. دلا ودووا و همکارانش [۲۶]، کای شو و همکارانش [۲۷]، جیزل باستیداس گواچو و همکارانش [۲۸] از ترکیب ویژگی‌های مبتنی بر محتوا و شبکه برای تشخیص اخبار جعلی استفاده کرده‌اند، که به دلیل استفاده از ویژگی‌های مبتنی بر محتوا بر مشکل شروع سرد[۴۳] غلبه پیدا کرده‌اند؛ شروع سرد به معنای ایجاد و انتشار پستی به تازگی در شبکه است که کاربری آن را نپسندیده یا بازنشر نکرده است و الگوی گسترش آن در شبکه هنوز تکمیل نشده است. سویتلانا ولکووا و همکارانش [۴۶]، بر روی اخبار فریب، تبلیغات، هجو و غیره در زمان حمله تروریستی بروکسل در سال ۲۰۱۶ و با در نظر گرفتن متن توئیت و تعاملات کاربران در شبکه توئیتر کار کرده‌اند؛ این نویسندگان از ایده آموزش شبکه عصبی بر روی مجموعه داده متوازن استفاده کرده‌اند. ضمناً، هائو لیائو و همکارانش [۶۷]، با ساخت گرافی بین کاربر و نظرات کاربر در شبکه اجتماعی به دنبال استخراج ویژگی‌های محتوا و شبکه برای شناسایی اخبار جعلی بوده‌اند. ضمناً، این نویسندگان برای تعبیه اطلاعات به بردار از روش‌های مبتنی بر مکانیزم توجه[۴۴] که باعث حذف زیادی از اطلاعات نامرتبط می‌شود، استفاده کرده‌اند. کای شو و همکارانش [۳۰]، گرافی بین پاسخ کاربر بر روی خبر، کاربر دریافت‌کننده و ارسال‌کننده خبر تشکیل داده است و از ترکیب

[۳۴] Content-based
[۳۵] Context-based
[۳۶] Syntax
[۳۷] Semantic
[۳۸] Lexical
[۳۹] Network-based
[۴۰] User-based
[۴۱] Clustering Coefficient
[۴۲] Like
[۴۳] Cold start
[۴۴] Attention mechanism-based methods

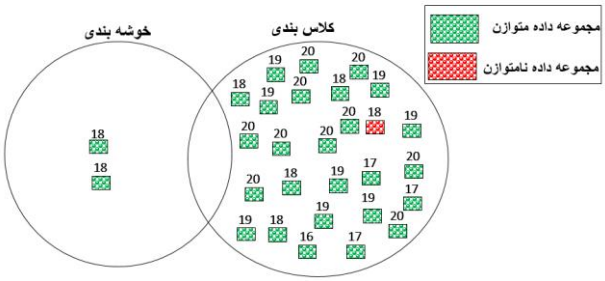
ویژگی‌های مبتنی بر محتوا و شبکه برای تشخیص اخبار جعلی استفاده کرده است و به ویژگی‌های بافتار-کاربر توجه نکرده‌اند؛ با این تفاوت که این نویسنده در مقاله دیگری [۲۹]، از ترکیب دو ویژگی بافتار-کاربر و بافتار-شبکه برای شناسایی کاربر جعلی استفاده کرده‌اند. طارق حمدی و همکارانش [۳۱]، از ترکیب ویژگی‌های کاربر و شبکه با کمک تعبیه گره به بردار (Node۲vec) برای شناسایی منبع فرستنده اخبار جعلی استفاده کرده‌اند ولی در این پژوهش، ترکیب این نوع ویژگی‌ها بر روی مجموعه داده نامتوازن و در زبان فارسی بررسی شده است. موتو بالاآناند و همکارانش [۶۸]، با زیر نظر گرفتن رفتار کاربر در بازه زمانی طولانی و ترکیب ویژگی‌های مبتنی بر بافتار-کاربر و بافتار-شبکه کاربران جعلی را شناسایی کرده‌اند. سوپانیا آفی وان سیفان و همکارانش [۳۲]، بررسی‌هایی بر روی اخبار سیل تایلند با اعمال ویژگی‌های مبتنی بر کاربر انجام داده است. گلزار حسین و همکارانش [۳۳]، برای شناسایی خبر جعلی در زبان بنگلادشی از ویژگی‌های مبتنی بر محتوا کمک گرفته‌اند؛ در صورتیکه، این ویژگی به تنهایی می‌تواند جعل شود و نیاز به استفاده از دیگر ویژگی‌ها وجود دارد. از مزایای پژوهش این نویسندگان می‌توان به ایجاد مجموعه داده جدید در زبان بنگلادشی که منابع زبان شناسی کمتری برای آن وجود دارد، اشاره کرد. یونگجون لی و همکارانش [۳۴]، از ویژگی‌های مبتنی بر کاربر استفاده کرده است و با ایجاد گراف به صورت پیشنهاد افراد شبیه به یکدیگر، به دنبال شناسایی کاربران جعلی هستند اما داده‌های مورد استفاده در آزمایشات آن‌ها متوازن هستند و به مشکل عدم توازن در مجموعه داده اشاره‌ای نداشته‌اند. سیرامومینای ویجیاراغوان و همکارانش [۳۵]، با اعمال روش‌های زبان‌شناسی بر روی متن خبر با تعبیه متن با کمک بردار فراوانی اصطلاح- معکوس فراوانی متن[۴۵] و Word۲vec و غیره به دنبال شناسایی اخبار جعلی است. اشروتیکا جدهاو و همکارانش [۳۶]، برای اثبات عملکرد بهتر روش‌های یادگیری عمیق در شناسایی اخبار جعلی از شبکه‌های عصبی بازگشتی[۴۶] و مدل معنایی ساختاریافته عمیق[۴۷] استفاده کرده است و هم‌چنین، اولووسون آجاو و همکارانش [۳۷]، با تمرکز بر ویژگی‌های محتوایی به دنبال شناسایی اخبار جعلی است اما این نویسندگان به دلیل استفاده شبکه عصبی بازگشتی و شبکه عصبی پیچشی[۴۸] در کارهای آتی خود اشاره داشته‌اند که به مجموعه داده بزرگتری نیاز دارند تا نتایج بهتری به دست آوردند. ژانگ و همکارانش [۳۸]، با ایجادگرافی بین نویسنده و اخبار و موضوع اخبار و ترکیب با اطلاعات محتوایی به دنبال شناسایی اخبار جعلی است. ابیشک ورما و همکارانش [۳۹]، با ایجاد مجموعه داده‌ای در اخبار هند و اعمال ویژگی‌های محتوایی با کمک روش‌های یادگیری عمیق سعی در شناسایی خبر جعلی دارند. ضمناً، این نویسندگان بر روی مجموعه داده متوازن کار کرده‌اند. نا روان و همکارانش [۴۰]، از ویژگی‌های مبتنی بر کاربر از جمله موقعیت جغرافیایی برای شناسایی بازبین‌گر جعلی[۴۹] استفاده کرده‌اند. بیندو و همکارانش [۶]، معتقد است کاربران جعلی با یکدیگر تشکیل جامعه می‌دهند. به همین منظور، از الگوریتم‌های خوشه‌بندی برای شناسایی جامعه کاربران جعلی استفاده کرده است. یوجینیو توچینی و همکارانش [۳]، تنها بر روی گراف کاربرانی که در فیس‌بوک، پست یکدیگر را پسندیدن کار کرده است و هم‌چنین، آدام بروئر و همکارانش [۶۵]، برای شناسایی حساب کاربران جعلی فقط ازگراف اتصالات در شبکه استفاده کرده‌اند که این ویژگی در زمان‌هایی که شروع سرد در شبکه وجود دارد، کارایی ضعیفی از خود نشان می‌دهد، به همین دلیل، در این مقاله از ویژگی بافتار-کاربر هم استفاده شده است. سید مهدی حسینی مطلق و همکارانش [۴۱]، از الگوریتم خوشه‌بندی براساس ویژگی‌های مبتنی بر محتوا برای شناسایی اخبار جعلی استفاده کرده است. شو یانگ و همکارانش [۴۲]، از ویژگی مبتنی بر شبکه برای شناسایی کاربر جعلی استفاده کرده‌اند که مشکل شروع سرد در کار آن‌ها نیز دیده می‌شود. تائن فان و همکارانش [۴۳]، از نحوه نگارش کاربر و با تعبیه متن نگارش شده به بردار ویژگی، حساب کاربران جعلی را شناسایی می‌کند. محمدرضا محمدرضایی و همکارانش [۱۹]، با ایجاد گراف دوستی بین کاربران و محاسبه معیارهای شباهت مانند جاکارد و کسینوس و غیره اقدام به شناسایی کاربران جعلی می‌کند، هم‌چنین، آن‌ها از روش بیش نمونه‌برداری اقلیت مصنوعی برای ایجاد توازن در مجموعه داده استفاده کرده‌اند. اما در این مقاله، علاوه بر ویژگی شبکه بر روی ویژگی‌های کاربر هم کار شده است و برای متوازن‌سازی داده از روش‌های مبتنی بر یادگیری عمیق به کمک شبکه مولد متخاصم شرطی استفاده شده است. هم‌چنین، ملیک متین و همکارانش [۴۴]، برای شناسایی کاربرانی که در شبکه توئیتر هرزنامه ایجاد می‌کنند، از ترکیب سه ویژگی یعنی مبتنی بر محتوا، بافتار-کاربر و بافتار-شبکه استفاده کرده است، اما باید اثرگذاری مدل آن‌ها در شرایط نامتوازن بودن مجموعه داده نیز بررسی گردد. چائو چن و همکارانش [۴۵]، برای شناسایی هرزنامه‌های موجود در توئیتر از ویژگی‌های مبتنی بر بافتار استفاده کرده است. در این مقالات نیز مشکل عدم توازن داده مطرح نیست. یانگ لیو و همکارانش [۶۶]، با اعمال ویژگی‌های مبتنی بر بافتار-کاربر و محتوا روی پاسخ کاربران، اخبار جعلی را شناسایی می‌کنند. ضمناً، آن‌ها با کمک شبکه عصبی بر مشکل شروع سرد غلبه کرده‌اند. در شکل ۶ دسته‌بندی از مطالب گفته‌شده براساس ویژگی مورد استفاده مقالات نشان داده شده است.

[۴۵] Term Frequency - Inverse Document Frequency (TF-IDF)
[۴۶] RNN
[۴۷] Deep Structured Semantic Model (DSSM)
[۴۸] Convolutional Neural Network (CNN)
[۴۹] Fake reviewer

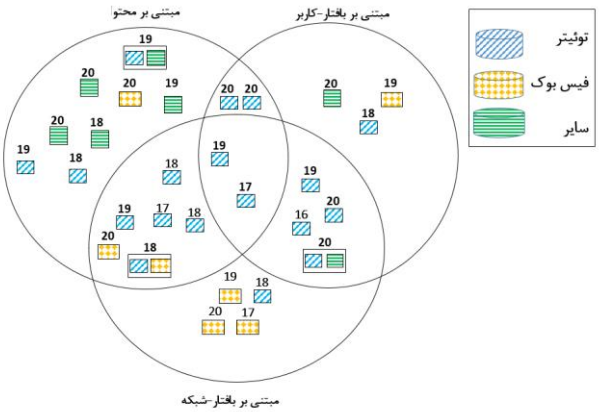


شکل ۶ . گونه شناسی مقالات براساس ویژگی‌های مورد استفاده در آن‌ها

در شکل۷ و۸ سال انتشار مقالات با دو عدد آخر آن سال نشان داده شده است. به طور مثال، سال ۲۰۲۰ با ۲۰ نمایش داده شده است. در شکل۷ دسته‌بندی مقالات از دیدگاه استفاده از الگوریتم‌های کلاس‌بندی و خوشه‌بندی و وجود توازن در مجموعه داده بررسی شده است. در شکل ۸ مقالات از منظر سال انتشار و ویژگی مورد استفاده و مجموعه داده استفاده شده، دسته‌بندی شده‌اند.



شکل ۷. دسته‌بندی مقالات از منظر خوشه‌بندی/کلاس‌بندی و توازن در مجموعه داده



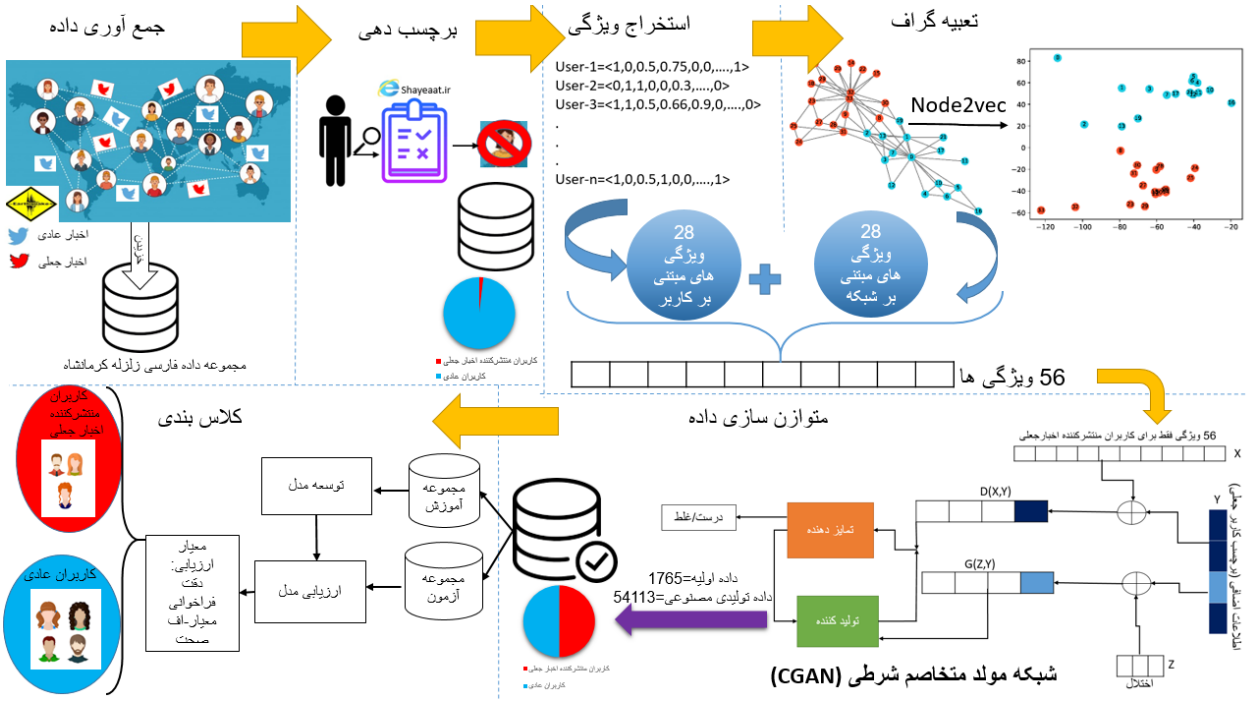
شکل ۸ . دسته‌بندی مقالات براساس ویژگی‌ها و مجموعه داده مورد استفاده

با توجه به مطالعات انجام گرفته مشخص شد اکثر تحقیقات در این حوزه بر روی مجموعه داده متوازن صورت گرفته است و به مجموعه داده‌های واقعی که عدم توازن داده در آن‌ها وجود دارد، توجه اندکی شده است. بنابراین در این پژوهش، سیستمی برای شناسایی کاربران منتشرکننده اخبار جعلی با بهره‌گیری از ویژگی‌های مبتنی بر بافتار شامل ترکیب ویژگی‌های شبکه با ویژگی‌های مبتنی بر کاربر پیشنهاد داده شده است. علاوه بر این، در این سیستم، مدلی برای حل عدم توازن مجموعه داده واقعی به کمک شبکه مولد متخاصم شرطی ارائه شده است که با توجه به بررسی‌های انجام شده نسبت به کارهای پیشین نوآوری به همراه دارد.

۳ سیستم پیشنهادی

در این بخش جزئیات سیستم پیشنهادی به طور کامل شرح داده می‌شود. مراحل شناسایی کاربران منتشرکننده اخبار جعلی با

یادگیری ماشین در این مقاله شامل شش مرحله می‌باشد که به طور خلاصه در شکل۹ نمایش داده شده است:



شکل ۹. مراحل سیستم پیشنهادی

مرحله اول شامل گردآوری مجموعه داده مناسب است. به همین منظور، در این مقاله از مجموعه داده شبکه توئیتر در زبان فارسی استفاده شده است. به همین منظور، از مجموعه داده "RumorTwitterKNTU" که به کمک رابط برنامه‌نویسی نرم-افزار[۵۰] تعبیه شده توسط وب سایت توئیتر و twitter4j جمع‌آوری شده بود، استفاده شده است[۵۱]. این مجموعه داده شامل ۳۵۹۸۰۴۹ توئیت است که توسط ۱۱۱۹۸۱ کاربر که با زبان فارسی در بازه زمانی دو هفته‌ای از ۳ آذر ماه سال ۱۳۹۶ تا ۱۷ آذر ماه سال ۱۳۹۶ در مدت وقوع زلزله کرمانشاه ایران در توئیتر انتشار یافته است [۴۷].

در مرحله دوم برچسب‌دهی داده‌ها به دو کلاس کاربران منتشرکننده اخبار جعلی وکاربران عادی صورت می‌گیرد، که در این مرحله از بین ۴۳۴۵ توئیت که از قبل در مجموعه داده با برچسب شایعه نام‌گذاری شده بود، طی فرآیند انسانی توسط نگارنده این مقاله، متن توئیت‌ها با اطلاعات سایت شایعات[۵۲] بازبینی شده است. نهایتاً، ۲۸۷۸ توئیت با برچسب پیام جعلی نام‌گذاری شده است. در نتیجه، اگر حساب کاربری حداقل یک بار اخبار و اطلاعات جعلی در این مدت پست کرده باشد، آن حساب کاربری با عنوان "کاربر منتشرکننده اخبار جعلی" برچسب‌گذاری شده است. در نتیجه، ۲۱۲۹ کاربر با برچسب "کاربر منتشرکننده اخبار جعلی" و ۱۰۹۸۵۲ کاربر با برچسب "کاربر عادی" نام‌گذاری شده است. در نهایت، مجموعه داده استاندارد به نام "FakeUser_KNTU (FU_KNTU)" ایجاد گشت.

سپس لازم است تا ویژگی‌های موردنیاز برای شناسایی کاربران منتشرکننده اخبار جعلی استخراج شود. این مقاله تمرکز خود را بر روی تأثیر ویژگی‌های مبتنی بر بافتار که ترکیب ویژگی‌های کاربر و شبکه است، گذاشته است. به همین منظور، در مرحله سوم برای استخراج ویژگی‌های مبتنی بر کاربر، ۲۸ ویژگی از اطلاعات کاربران در نظر گرفته شده است؛ در جدول ۱ اطلاعات و تعاریف آن‌ها قابل مشاهده است. نهایتاً، این ویژگی‌ها به بردار ویژگی تبدیل شده است.

جدول ۱. ویژگی‌های مبتنی بر بافتار-کاربر و تعاریف آن‌ها

نام ویژگی	تعریف ویژگی
۱-Userid	عدد صحیحی است که نشان‌دهنده شناسه منحصر به فرد هر کاربر است.
۲- uisDefaultProfileImage	آیا کاربر از عکس پیش‌فرض استفاده کرده است؟
۳- ugetFollowersCount	تعداد افرادی که کاربر را دنبال می‌کنند.

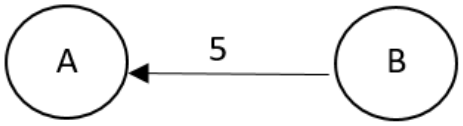
[۵۰] Application Programming Interface (API)
[۵۱] https://trlab.ir/res.php?resource_id=۳
[۵۲] http://shayeaat.ir/

٤- uisProfileUseBackgroundImage	آیا کاربر از عکس پس‌زمینه‌ی پیش‌فرض استفاده کرده است؟
٥- uisDefaultProfile	آیا کاربر تم پروفایلش را عوض کرده است؟
٦- uisShowAllInlineMedia	آیا کاربر لینک ویدیو را گذاشته است و یا خود ویدئو را آپلود کرده است؟
٧- ugetFriendsCount	تعداد افرادی که کاربر آن‌ها را دنبال می‌کند.
Uyear-٨	سال ایجاد حساب کاربری
Umonth-٩	ماه ایجاد حساب کاربری
Uday-١٠	روز ایجاد حساب کاربری
Uhour-١١	ساعت ایجاد حساب کاربری
Uminute-١٢	دقیقه ایجاد حساب کاربری
Usecond-١٣	ثانیه ایجاد حساب کاربری
١٤- ugetFavouritesCount	تعداد توئیت‌هایی که این کاربر پسندیده است.
١٥- UgetUtcOffset	منطقه زمانی کاربر را نشان می‌دهد.
١٦- uisProfileBackgroundTiled	آیا پس‌زمینه‌ی کاربر قالب کاشی‌کاری دارد؟
١٧- ugetStatusesCount	تعداد توئیت و ریتوئیت‌هایی که توسط کاربر ایجاد شده است.
١٨- uisGeoEnabled	آیا منطقه زمانی کاربر فعال است؟
uisVerified-١٩	این ویژگی نشان می‌دهد آیا کاربر مورد علاقه عموم مردم است؟ کاربرانی که تیک آبی کنار پروفایل خود دارند، یعنی مورد علاقه عموم مردم هستند .
٢٠- uisTranslator	آیا کاربر مترجم است؟
٢١- ugetListedCount	تعداد لیست‌های عمومی که کاربر عضو آن است.
٢٢- uTweetCountIn١٥Days	تعداد توئیت ایجاد شده توسط کاربر در بازه دو هفته‌ای جمع‌آوری اطلاعات
٢٣- uRetweetCountIn١٥Days	تعداد ریتوئیت ایجاد شده توسط کاربر در بازه دو هفته‌ای جمع‌آوری اطلاعات
AlphaForSI-٢٤	مقدار آن، از طریق رابطه ۵ محاسبه شده است. $alpha\ for\ si = \frac{\text{تعداد بازتوئیت}}{\text{تعداد توئیت}}$ (۵)
SI-٢٥	برای محاسبه اثرگذاری اجتماعی کاربر[53] از رابطه ۶ به دست آمده است. $user_{si} = \frac{\text{(تعداد بازتوئیت—}}{\text{تعداد توئیت}}$ (تعداد توئیت) ∗ (تعداد دنبال‌کننده) (۶)
AlphaForSI٢-٢٦	مقدار آن، از طریق رابطه ۷ محاسبه شده است. $alpha\ for\ si^{٢}$ = تعداد بازتوئیت − تعداد توئیت (۷)
SI٢-٢٧	برای محاسبه اثرگذاری اجتماعی کاربر از رابطه ۸ به دست آمده است. $user_{si^{٢}}$ = (تعداد دنبال‌کننده) ∗ (تعداد بازتوئیت − تعداد توئیت) (۸)
userAge-٢٨	تفاوت زمان ساخت حساب کاربری و زمان گرفتن داده از توئیتر را نشان می‌دهد.

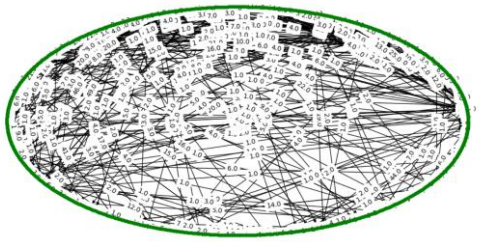
در مرحله چهارم، برای استخراج ویژگی مبتنی بر بافتار-شبکه، شبکه توئیتر به گراف وزن‌دار و جهت‌دار مدل شد. این گراف با G=(V,E) نمایش داده شده است که V نشان‌دهنده کاربران شبکه اجتماعی توئیتر و E روابط و تعاملات بین کاربران را نشان می‌دهد. به طور مثال، اگر کاربر A در زمان جمع‌آوری این مجموعه داده، بیست توئیت ایجاد کرده باشد و در این مدت، کاربرB بر روی پنج توئیت کاربر A پاسخی[54] گذاشته باشد، همانند شکل ۱۰ جهت یال از سمت کاربر B به سمت کاربر A و وزن این یال ۵ خواهد بود. بخش کوچکی از گراف ایجاد شده با ۱۰۰۰ گره در شکل۱۱ قابل مشاهده است.

[53] User's social influence

[54] Reply



شکل ۱۰. ساختار گراف



شکل ۱۱. بخشی از گراف ایجاد شده با ۱۰۰۰ گره

بعد از ایجاد گراف، در این مرحله برای استخراج ویژگی‌های مبتنی بر بافتار-شبکه، از روش تعبیه Node۲vec استفاده شده است که اطلاعات گراف را به بردار ویژگی با بعد دلخواه تبدیل می‌کند. با استخراج ۲۸ ویژگی مبتنی بر شبکه به کمک این روش و ترکیب این ویژگی‌ها با ۲۸ ویژگی مبتنی بر کاربر از مرحله قبل، در نهایت ۵۶ ویژگی برای شناسایی کاربران منتشرکننده اخبار جعلی انتخاب شده است. با وجود ۲۸ ویژگی عددی در اطلاعات کاربران، در جهت برتری پیدا نکردن ویژگی‌های شبکه‌ای بر ویژگی اطلاعات کاربران، دقیقاً تعداد ویژگی هر دسته برابر انتخاب شده است. به عبارت دیگر، برای خنثی کردن اثر یک دسته خاص بر کل نتیجه این تصمیم گرفته شد تا هردو دسته به یک اندازه سیستم نهایی را تحت تأثیر خود قرار دهند.

باید توجه شود که الگوریتم‌های یادگیری ماشین بر روی مجموعه داده‌های متوازن، به خوبی عمل می‌کنند. در نتیجه، در مجموعه - داده‌های نامتوازن عملکرد مناسبی از خود نشان نمی‌دهند. علاوه بر این، هزینه کلاس‌بندی اشتباه نمونه کلاس اقلیت خیلی بیشتر از هزینه کلاس‌بندی اشتباه نمونه کلاس اکثریت است [۴۸]. در مجموعه داده ایجاد شده عدم توازن مشهود است؛ به این معنا که تعداد نمونه‌های کلاس اکثریت[۵۵] در اینجا منظور کاربران عادی از تعداد نمونه‌های کلاس اقلیت یعنی کاربران منتشرکننده اخبار جعلی خیلی بیشتر است. در مجموعه داده‌های نامتوازن، به نسبت نمونه کلاس اقلیت به نمونه کلاس اکثریت نرخ نامتوازنی[۵۶] گویند، که در مجموعه داده "FU_KNTU " عدم توازنی با ۱۷۶۵ نمونه داده کلاس اقلیت و ۵۵۸۷۷ نمونه داده کلاس اکثریت با نرخی در حدود ۰/۰۳ وجود داشت. در نتیجه، مدیریت این نوع داده‌های نامتوازن امروزه به چالش بزرگی در حوزه‌های مختلفی چون بانکداری، امنیت، پزشکی، بازیابی اطلاعات، تشخیص کلاه‌برداری و شایعه و اخبار جعلی تبدیل شده است [۴۹]. از این جهت سه رویکرد برای حل این مشکل وجود دارد.

۱) تغییر در سطح داده : در این دسته بر روی نمونه مجدد[۵۷] تمرکز می‌شود و روش‌های نمونه مجدد شامل افزودن نمونه[۵۸]، کاهش نمونه[۵۹] و یا ترکیب هر دو است. در روش-های افزودن نمونه، سعی می‌شود تا با تولید داده مصنوعی و اضافه کردن داده مصنوعی به کلاس اقلیت، مجموعه داده متوازن شود. در روش کاهش نمونه برخلاف روش قبل سعی می‌شود تا با کاهش تعداد نمونه‌های کلاس اکثریت، مجموعه داده به توازن برسد. یکی از روش‌های مشهور اضافه‌کردن داده، روش بیش نمونه‌برداری اقلیت مصنوعی است. ایده روش بیش نمونه‌برداری اقلیت مصنوعی شامل یافتن داده نزدیک به کلاس اقلیت و تولید داده تصادفی بین محدوده‌ی داده‌های کلاس اقلیت است؛ در این روش به علت توجه به توزیع داده محلی[۶۰] یادگیری به خوبی صورت نمی‌گیرد و ممکن است ایجاد همپوشانی[۶۱] یا داده پرت[۶۲] کند.

۲) تغییر در سطح الگوریتم: روش‌های موجود در این دسته یادگیری را به سمت کلاس اقلیت می‌برند.

۳) تغییر در روش‌های حساس به هزینه[۶۳]: این روش به دنبال کاهش خطا در سطح داده یا الگوریتم است [۱۴] و [۴۸].

در این مقاله برای رفع مشکل عدم توازن از روش تغییر در سطح داده با بهره‌گیری از روش جدیدی به نام شبکه مولد متخاصم استفاده شده است که می‌تواند توزیع‌داده سراسری[۶۴]را یاد بگیرد. با یادگیری توزیع داده توسط این شبکه امکان تولید داده مصنوعی فراهم می‌آید. شبکه مولد متخاصم یک روش یادگیری بدون ناظر است که از یادگیری عمیق برای تولید داده مصنوعی استفاده می‌کند. در یادگیری عمیق می‌توان به یادگیری خودکار ویژگی‌ها توسط شبکه بدون نیاز به دخالت انسانی اشاره داشت، که از مزیت استفاده از این روش است. از این روش در تولید تصویر، صدا، متن، شناسایی نفوذ و کلاه‌برداری و غیره می‌توان استفاده کرد [۵۰] تا [۵۲].

در مرحله پنجم این پژوهش برای ایجاد توازن در مجموعه داده از شبکه مولد متخاصم شرطی برای تولید داده مصنوعی استفاده شده است. برای این منظور، این شبکه تنها با برچسب کاربران منتشرکننده اخبار جعلی که شامل ۱۷۶۵ داده و ۵۶ ویژگی

[۵۵] Majority
[۵۶] Imbalanced Ratio (IR)
[۵۷] Resampling
[۵۸] Oversampling
[۵۹] Under-sampling
[۶۰] Local
[۶۱] Overlapping
[۶۲] Outlier
[۶۳] Cost-sensitive
[۶۴] Global

استخراج‌شده است، آموزش می‌بیند و شبکه مولد متخاصم شرطی بعد از مرحله آموزش، ۵۴۱۱۳ داده مصنوعی با برچسب کلاس اقلیت تولید می‌کند و به مجموعه داده اضافه می‌گردد تا نمونه‌های دو کلاس متوازن شود.

مرحله آخرشامل آموزش کلاس‌بند با ویژگی‌های استخراج‌شده و آزمون کلاس‌بند با معیارهای ارزیابی مناسب مانند صحت، فراخوانی، معیار اف و دقت وغیره است تا کارایی سیستم پیشنهادی مشخص شود.

۴ آزمایشات تجربی

برای ارزیابی سیستم پیشنهادی، دو سناریو با اهداف معین دنبال شده است. هدف سناریو اول تحلیل حساسیت پارامترهای سیستم است. هدف از سناریو دوم مقایسه سیستم پیشنهادی با رقبای شناخته شده در این زمینه با توجه به معیارهای ارزیابی است. در تمامی آزمایشات، از روش Cross Validation استفاده شده است. ضمناً، شبکه مولد متخاصم شرطی با شبکه عصبی پیچیده عمیق پیاده‌سازی شده است [۵۳]. الگوریتم‌های دیگر متوازن‌سازی داده با کمک کتاب‌خانه تعبیه‌شده در پایتون پیاده‌سازی شده است [۵۴].

۴,۱ معرفی روش کار رقبا در متوازن‌سازی داده

در جدول ۲ روش کار تعدادی از الگوریتم‌های متوازن‌سازی داده شرح داده شده‌است.

جدول ۲ . الگوریتم‌های متوازن‌سازی داده

نام الگوریتم	روش کار
SMOTE-۱	این روش، بین کِی نزدیک‌ترین همسایه نمونه کلاس اقلیت و خود نمونه کلاس اقلیت داده مصنوعی تولید می‌کند [۵۵].
Tomek links-۲	این روش، داده‌های کلاس اکثریت که در توزیع کلاس اقلیت هستند و به صورت خطی نمی‌توان دو کلاس را از هم جدا کرد را حذف می‌کند و سپس داده مصنوعی تولید می‌کند [۵۶].
ENN-۳	این روش، اگر حداقل دو برچسب از سه همسایه داده‌های موجود در هر دو کلاس اقلیت و اکثریت شبیه برچسب خود نمونه نباشد، نمونه را از مجموعه داده حذف می‌کند. نهایتاً، بعد از اعمال تمیزی[۶۵] داده مصنوعی تولید می‌کند [۵۶].
Borderline-SMOTE-٤ (Borderline۱,Borderline۲)	این روش، به دنبال رسم خط جداکننده دقیق‌تر بین دو کلاس اقلیت و اکثریت است. سپس، نمونه-های نزدیک خط مرزی را با ایجاد داده مصنوعی بیشتر می‌کند [۵۷] .
Cluster-Smote-٥	این روش، ابتدا با اجرای الگوریتم k-means خوشه‌های کلاس اقلیت را پیدا می‌کند و بعد الگوریتم SMOTE را روی هر خوشه ایجاد-شده اعمال می‌کند، تا داده مصنوعی تولید کند [۵۸].
Distance-SMOTE-٦	این روش، ابتدا میانگین کِی نزدیک-ترین همسایه را پیدا می‌کند و سپس فاصله نمونه میانگین را با نمونه اصلی می‌سنجد و این فاصله را در عددی تصادفی بین صفر و یک ضرب می‌کند و نهایتاً شروع به تولید داده مصنوعی می‌کند [۵۹].
ADASYN-۷	این روش، توزیع وزن‌دار بین نمونه-های کلاس اقلیت در نظر می‌گیرد و داده‌های مصنوعی بیشتری برای نمونه‌هایی که یادگیری آن‌ها سخت‌تر است، ایجاد می‌کند و برای نمونه‌هایی که یادگیری آن‌ها آسان‌تر است، داده‌های کمتری تولید می‌کند [۶۰].
Random-SMOTE-٨	این روش، با انتخاب دو نقطه به صورت تصادفی در فضای داده کلاس اقلیت، مثلثی بین نمونه کلاس اقلیت و دو نقطه انتخابی شکل می‌دهد. سپس، در مثلث ایجادشده به هر تعداد که لازم است تا مجموعه داده متوازن شود، داده مصنوعی تولید - می‌کند [۶۱].
Gaussian- SMOTE-٩	این روش، برخلاف SMOTE که از توزیع احتمال یکنواخت و الگوریتم کِی نزدیک‌ترین همسایه[۶۶] برای تولید داده مصنوعی برای کلاس اقلیت استفاده می‌کند، در این روش از ترکیب الگوریتم کِی نزدیک‌ترین همسایه و توزیع احتمال گوسی استفاده می‌شود [۶۲].
CURE-SMOTE-١٠	این روش، خوشه‌بندی کلاس اقلیت را با خوشه‌بندی توسط بازنمایی[۶۷] انجام می‌دهد و بعد از حذف داده پرت، داده مصنوعی تولید می‌کند [۶۳].

[۶۵] Cleaning
[۶۶] K- Nearest Neighbor (KNN)
[۶۷] Clustering Using Representatives (CURE)

CCR-۱۱	این روش، ابتدا همسایه‌های نمونه کلاس اقلیت اگر شامل نمونه کلاس اکثریت باشد را حذف می‌کند و سپس، داده مصنوعی بیشتری برای نمونه‌هایی که یادگیری آن‌ها سخت-تراست، تولید می‌کند [۶۴].

۴,۲ داده

همانطور که در بخش ۳ توضیح داده شد، مجموعه داده توئیتر در زبان فارسی در مدت وقوع زلزله کرمانشاه در بازه زمانی دو هفته‌ای از ۳ آذر ماه سال ۱۳۹۶ تا ۱۷ آذر ماه سال ۱۳۹۶ به نام " FU_KNTU" برای شناسایی کاربران منتشرکننده اخبار جعلی توسط نگارنده جمع‌آوری و استفاده شده است. همانطور که در قسمت مقدمه اشاره شد، مجموعه داده در زبان فارسی در شبکه توئیتر به منظور استفاده در شناسایی کاربران منتشرکننده اخبار جعلی وجود نداشته است؛ بنابراین، از دستاوردهای این پژوهش می‌توان به ایجاد این مجموعه داده و گسترش آن برای استفاده در پژوهش‌ها اشاره کرد؛ این مجموعه داده از طریق لینک زیر قابل دسترسی است[۶۸]. خلاصه‌ای از اطلاعات این مجموعه داده در جدول ۳ آمده است.

جدول ۳ . خلاصه‌ای از اطلاعات مجموعه داده

تعدادکاربر	۱۱۱۹۸۱
تعداد توئیت	۳۵۹۸۰۴۹
تعداد خبر شایعه	۴۳۴۵
تعداد خبر جعلی	۲۸۷۸
تعداد کاربران منتشرکننده اخبار جعلی	۲۱۲۹
تعداد کاربرانی که روی توئیت آن‌ها پاسخ گذاشته شده است.	۵۵۸۷۷
تعداد کاربران منتشرکننده اخبار جعلی که روی توئیت آن‌ها پاسخ گذاشته شده است.	۱۷۶۵

۴,۳ معیارهای ارزیابی

برای ارزیابی عملکرد روش‌های مبتنی بر یادگیری ماشین، از معیار-هایی استفاده می‌شود، که خلاصه‌ای از تعاریف آن‌ها در ادامه آمده-است. به همین منظور، در ادامه برای ارزیابی سیستم پیشنهادی و مقایسه سیستم با رقبا از این معیارها استفاده شده است.

- فراخوانی یا نرخ مثبت درست[۶۹] مشخص می‌کند چه تعداد از نمونه‌های مرتبط بازیابی شده‌اند.
- دقت مشخص می‌کند چه تعداد از نمونه‌های بازیابی‌شده واقعا مرتبط هستند.
- صحت مشخص می‌کند چه نسبتی از نمونه‌ها به طور صحیح کلاس‌بندی شده‌اند.
- نرخ مثبت کاذب[۷۰] مشخص می‌کند چه تعداد از نمونه‌های نامرتبط بازیابی شده‌اند.
- AUC_ROC[۷۱] نشان می‌دهد چه مقدار سیستم توانسته بین کلاس‌ها تمایز ایجاد کند. مقدار آن از سطح زیر نموداری که در محور افقی آن، نرخ مثبت کاذب و در محور عمودی آن، نرخ مثبت درست قرار می‌گیرد، به دست می-آید. اگر این مقدار بین ۰/۵ و ۱ باشد، نشان‌دهنده احتمال بالای سیستم در ایجاد تمایز بین کلاس‌ها است. اگر این مقدار برابر ۰/۵ باشد، به این معنا است که کلاس‌بند نمی-تواند بین دو کلاس تمایز ایجاد کند. تعیین مقدار آستانه[۷۲] برای این نمودار به میزان اهمیت دادن به نرخ مثبت کاذب و نرخ مثبت درست در کلاس‌بندی بستگی دارد. در این مقاله، به جهت وجود اهمیت یکسان در بین دو کلاس، مقدار آستانه برابر مقدار ۰/۵ درنظر گرفته شده است [۶۹].
- زیان[۷۳] نشان‌می‌دهد چه قدر پیش‌بینی سیستم روی یک نمونه خاص بد بوده است. اگر پیش‌بینی سیستم عالی باشد، مقدار آن برابر صفر است و بر عکس.
- معیار اف نیز که ترکیبی از دقت و فراخوانی است، به صورت رابطه ۹ تعریف می‌شود.

$$\text{معیار-اف} = \frac{۲\times(\text{دقت}\times\text{فراخوانی})}{(\text{دقت}+\text{فراخوانی})} \qquad (۹)$$

۴,۴ پیکربندی

آزمایشات بر روی سیستم ۶۴ bit, core i۷, NVIDIA GeForce graphic card, ۱TB internal storage, ۸ gig RAM به زبان برنامه‌نویسی پایتون نسخه ۳/۶، با کتابخانه Scikit-learn Keras ,networkx, node۲vec و غیره بر روی سیستم عامل ویندوز ۱۰ انجام شده است.

۴,۵ تحلیل حساسیت

در این بخش، تحلیل حساسیت پارامترهای اثرگذار که شامل سه بخش پارامترهای مؤثر بر شبکه مولد متخاصم شرطی، پارامترهای مؤثر در تعیین ویژگی‌های شبکه‌ای و تأثیر ویژگی‌ها روی سیستم

[۶۸] https://trlab.ir/res.php?resource_id=۷
[۶۹] True Positive Rate (TPR)
[۷۰] False Positive Rate (FPR)
[۷۱] Area Under the Curve(AUC)- Receiver Operating Characteristics (ROC)
[۷۲] Threshold
[۷۳] Loss

پیشنهادی، مورد بررسی قرار می‌گیرد تا بهترین مقدار برای پارامترها با توجه به مجموعه داده مورد استفاده، انتخاب شود.

۴،۵،۱ پارامترهای مؤثر بر شبکه مولد متخاصم شرطی

این بخش شامل دو قسمت محاسبه ماتریس ورودی شبکه مولد متخاصم شرطی و تعیین شرط پایان آموزش شبکه مولد متخاصم شرطی است. سایر پارامترهای مهم شبکه مولد متخاصم شرطی در جدول ۴ مشخص شده است.

جدول۴ . تعدادی از پارامترهای مهم شبکه مولد متخاصم شرطی

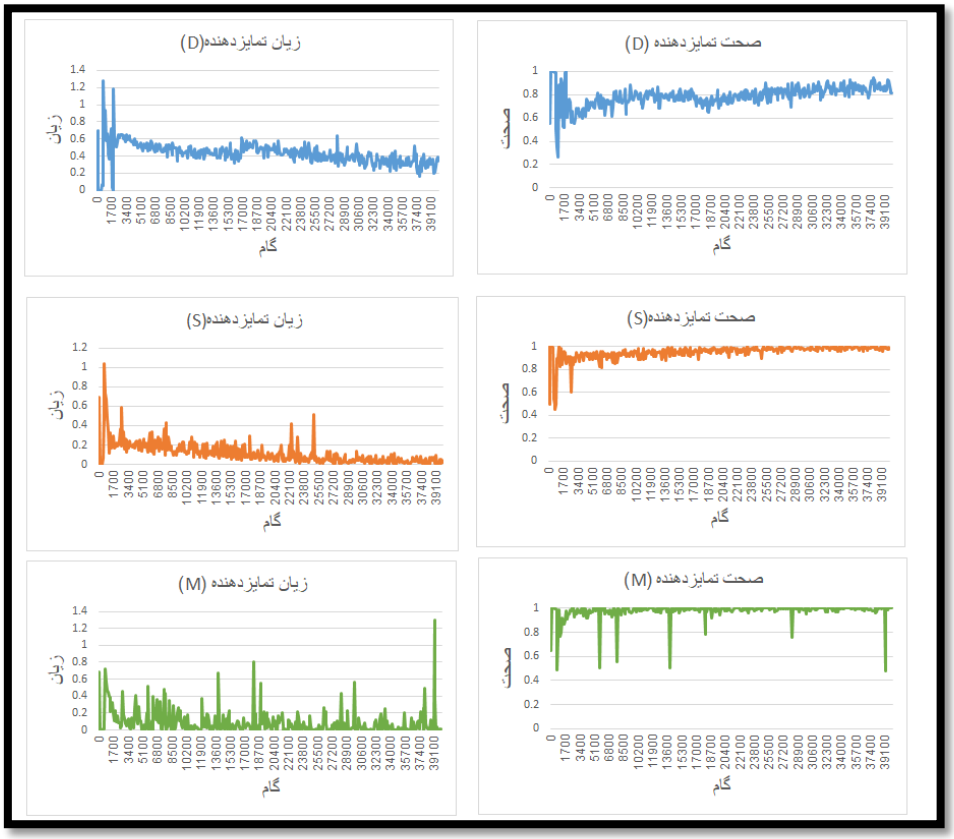
پارامتر	مقدار
اندازه هسته[۷۴]	۵
تعداد لایه‌های فیلتر(تولیدکننده)[۷۵]	۱ ,۳۲ ,۶۴ ,۱۲۸
تعداد لایه‌های فیلتر(تمایزدهنده)[۷۶]	۲۵۶ ,۱۲۸ ,۶۴ ,۳۲
تعداد گام حرکت[۷۷]	۱،۲
اندازه فضای پنهان[۷۸]	۱۰۰
اندازه دسته[۷۹]	۶۴

❖ محاسبه ماتریس ورودی شبکه مولد متخاصم شرطی

شبکه مولد متخاصم از شبکه عصبی پیچشی در مؤلفه‌های خود استفاده می‌کند. بنابراین، ورودی شبکه مولد متخاصم شرطی یک ماتریس دوبعدی n*n است؛ لازم است تا اطلاعات از شکل بردار یک بعدی n تایی به ماتریسی به ابعاد n*n تبدیل شود. برای تحقق این هدف سه ایده بررسی شده است. در ایده قطری(D)[۸۰] ویژگی‌ها بر روی قطر اصلی قرار می‌گیرند و سایر سلول‌ها صفر خواهند بود. در ایده جمع(S)[۸۱] ویژگی‌ها در درایه‌های نظیر به نظیر با یکدیگر جمع می‌شوند. در ایده ضرب (M)[۸۲] ویژگی‌ها در درایه‌های نظیر به نظیر ضرب می‌شوند.

هر سه ایده با ۴۰۰۰۰ بار تکرار گام در شبکه مولد متخاصم شرطی آزمایش شده است. نمودارهای صحت و زیان مؤلفه تمایزدهنده شبکه مولد متخاصم شرطی برای ایده قطری، جمعی، ضربی در شکل ۱۲ رسم شده است. سطر اول ایده قطری و سطر دوم ایده جمعی و سطر سوم ایده ضربی را نشان می‌دهد. محور افقی نمودارها نشان‌دهنده تعداد گام اجرا شبکه مولد متخاصم شرطی و محور عمودی به تفکیک در هر نمودار از راست به چپ، نشان‌دهنده صحت و زیان تمایزدهنده است.

[۷۴] Kernel-size
[۷۵] layer-filters(generator)
[۷۶] layer-filters(discriminator)
[۷۷] strides
[۷۸] latent-size
[۷۹] batch-size
[۸۰] Diagonal
[۸۱] Sum
[۸۲] Multiply



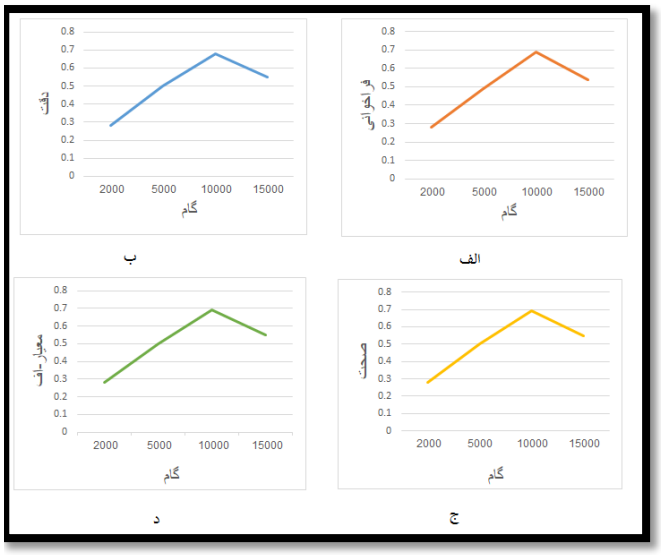
شکل ۱۲. سطر اول نمودارهای صحت و زیان تمایزدهنده برای ایده قطری، سطر دوم نمودارهای صحت و زیان تمایزدهنده برای ایده جمع، سطر سوم نمودارهای صحت و زیان تمایزدهنده برای ایده ضرب را نشان می‌دهد.

تمایزدهنده قصد دارد مقدار زیان خود را کاهش دهد، در صورتیکه تولیدکننده باید با فریب‌دادن تمایزدهنده مانع ازکاهش زیان تمایزدهنده شود و باید بتواند صحت تمایزدهنده را نیز کاهش دهد. همانطور که در شکل ۱۲ نشان‌داده شده است، این اتفاق به درستی در ایده قطری رخ داده است و در دو ایده جمعی و ضربی تولیدکننده نتوانسته تمایزدهنده را فریب دهد. در نتیجه، ورودی قطری برای آموزش شبکه مولد متخاصم شرطی و ادامه آزمایشات در نظر گرفته شده است.

❖ تعیین شرط پایان آموزش شبکه مولد متخاصم شرطی

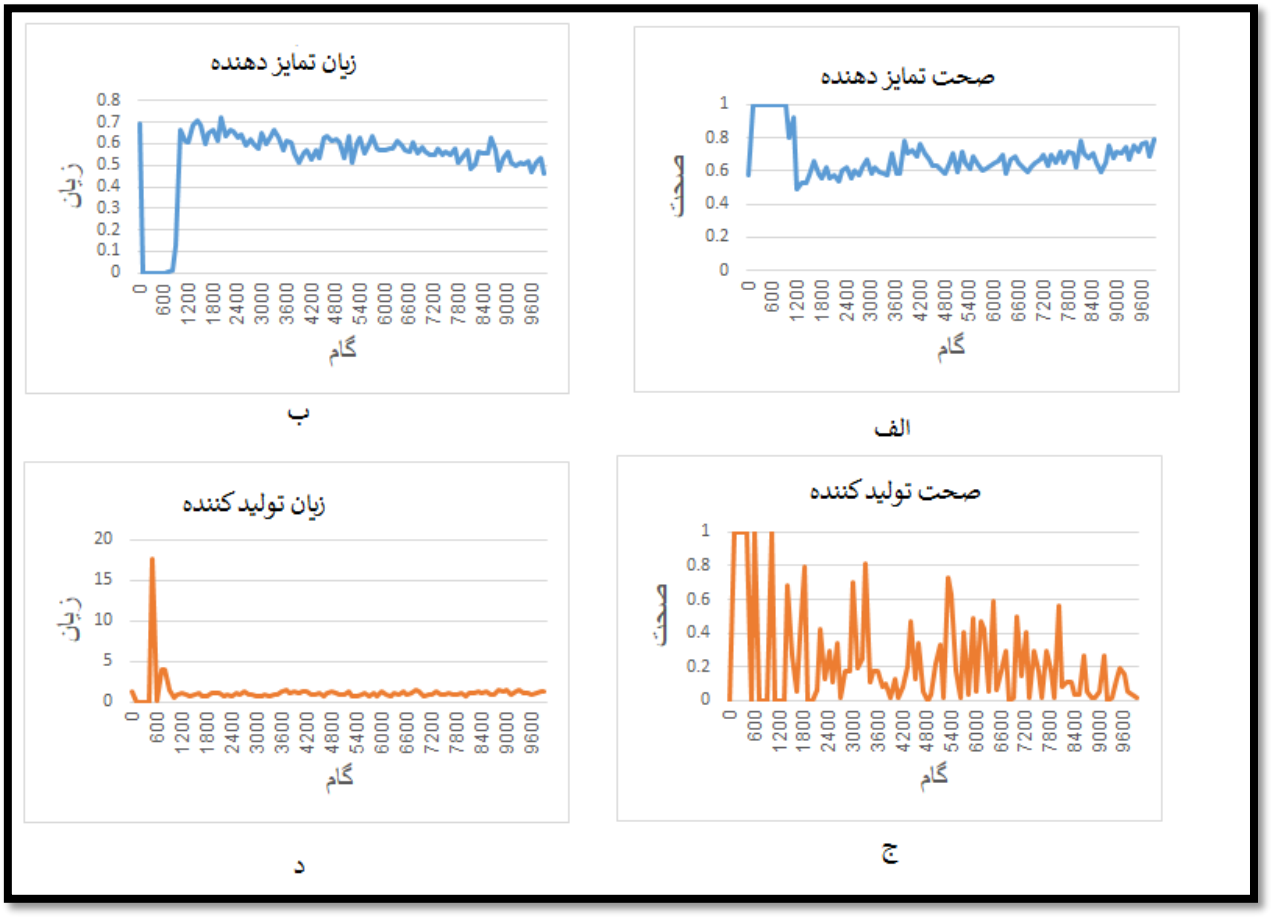
شرط پایان آموزش شبکه مولد متخاصم شرطی براساس تعداد گام تکرار تعیین شده است. به همین منظور، برای نشان‌دادن کیفیت داده‌های مصنوعی تولید شده، این داده‌ها همراه با داده‌های واقعی به کلاس‌بند، ماشین بردار پشتیبان[۸۳] داده شده است. بدیهی است، هرچه معیارهای ارزیابی کلاس‌بند بالاتر باشد، نشان می‌دهد که داده‌های مصنوعی تولید شده توسط شبکه مولد متخاصم شرطی به خوبی ویژگی‌های داده‌های واقعی را یادگرفته‌اند. در شکل ۱۳ محور افقی تعداد گام‌ها و محور عمودی معیارهای ارزیابی دقت، فراخوانی، معیار-اف و صحت هر گام را نشان می‌دهد. همانطور که در شکل ۱۳ مشهود است ، طبق فرضیات در نظر گرفته شده، ۱۰۰۰۰ گام برای آموزش شبکه بر روی ترکیب ویژگی‌های مبتنی بر بافتار-کاربر و بافتار-شبکه نتایج بهتری تولید کرده است.

[۸۳] Support Vector Machine (SVM)

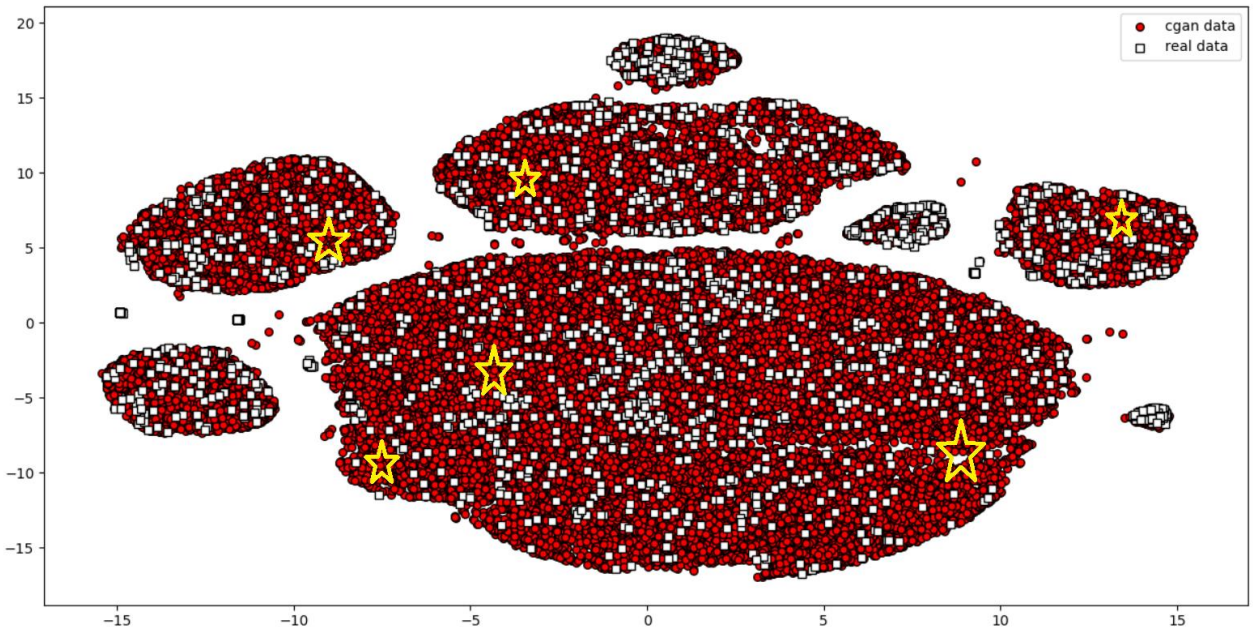


شکل ۱۳. نمودار "الف" تعداد گام و نتایج "فراخوانی " هر گام ، نمودار "ب" تعداد گام و نتایج "دقت" هر گام ، نمودار "ج" تعداد گام و نتایج "صحت" هر گام و نمودار "د" تعداد گام و نتایج "معیار-اف" هر گام در کلاس‌بند svm را نشان می‌دهد.

نمودار صحت و زیان برای ۱۰۰۰۰ گام در شکل ۱۴ رسم شده است. محور افقی نمودارها نشان‌دهنده تعداد گام اجرا شبکه مولد متخاصم شرطی و محور عمودی به تفکیک در هر نمودار نشان‌دهنده صحت و زیان هر مؤلفه است. هم‌چنین، با کمک ابزار کاهش بعد T-SNE داده‌های مصنوعی تولیدی برای کاربر منتشرکننده اخبار جعلی و داده‌های کاربر منتشرکننده اخبار جعلی موجود در مجموعه داده در شکل ۱۵ رسم شده است. در این شکل، داده‌های تولیدی توسط شبکه مولد متخاصم شرطی برای کاربر منتشرکننده اخبار جعلی با دایره قرمزرنگ در نمودار نشان داده شده است و هم‌چنین، داده‌های اصلی مجموعه داده برای کاربر منتشرکننده اخبار جعلی با مربع سفیدرنگ در نمودار نشان داده شده است. در ابزار T-SNE هرچه دو داده به هم شبیه‌تر باشند، با فاصله کمتری کنار هم رسم می‌شوند [۹]. همانطور که در شکل ۱۵ با علامت ستاره مشخص شده است، شبکه مولد متخاصم شرطی هم توانسته به خوبی توزیع داده را یاد بگیرد و هم داده‌های تولیدی مصنوعی جدید تولید کند.



شکل ۱۴. "الف ، ب"به ترتیب، نمودار صحت و زیان تمایز دهنده و "ج ، د"به ترتیب، نمودار صحت و زیان تولیدکننده در ۱۰۰۰۰ گام را نشان می‌دهد.



شکل ۱۵. نمودار T-SNE، داده تولیدی مصنوعی و داده موجود در مجموعه داده برای کاربر منتشرکننده اخبار جعلی را نشان می‌دهد.

بعد از آموزش با ۱۰۰۰۰ گام، ۵۴۱۱۲ داده مصنوعی با برچسب کاربر منتشرکننده اخبار جعلی به فرمت بردار ویژگی اولیه درآمد و به مجموعه داده قبلی اضافه گشت تا در مجموعه داده توازن ایجاد شود. نهایتاً، در مجموعه داده ۵۵۸۷۸ کاربر منتشرکننده اخبار جعلی و ۵۵۸۷۷ کاربر عادی با ترکیب ویژگی‌های مبتنی بر بافتار-کاربر و بافتار-شبکه وجود دارد.

۴،۵،۲ پارامترهای مؤثر در محاسبه ویژگی‌های شبکه‌ای

در روش Node۲vec همانطور که در بخش پیش‌زمینه معرفی شد، تعیین دو پارامتر p,q اهمیت ویژه‌ای دارد. به طوریکه، اگر p<۱ باشد، معادلات ساختاری با جستجوی اول سطح در گراف درنظر گرفته می‌شود و دید محلی از گراف ایجاد می‌کند. اما اگر q<۱ باشد، معادلات هموفیلی با جستجوی اول عمق در گراف در نظر گرفته می‌شود و دید سراسری از گراف ایجاد می‌کند. درنهایت، اگر p=۱,q=۱ باشد، یعنی هر دو معادلات ساختاری و هموفیلی به یک اندازه در نظر گرفته می‌شود. نتیجه بررسی هرسه حالت روی گراف تعاملات بین کاربران در جدول ۵ آمده است. حالت p=۱,q=۰,۵ با توجه به عملکرد بهتر انتخاب می‌شود. این نشان دهنده‌ی این است که کاربران منتشرکننده اخبار جعلی با یکدیگر تشکیل جامعه داده‌اند.

جدول۵ . بررسی تأثیر p,q

معیار-اف	فراخوانی	دقت	
۰/۳۵	۰/۳۵	۰/۳۵	P=۰/۵,q=۱
۰/۶۹	۰/۶۹	۰/۶۸	P=۱,q=۰/۵
۰/۵۹	۰/۵۹	۰/۵۹	P=۱,q=۱

سایر پارامترهای مهم Node۲vec در جدول ۶ مشخص شده است.

جدول ۶ . تعدادی از پارامترهای مهم Node۲vec

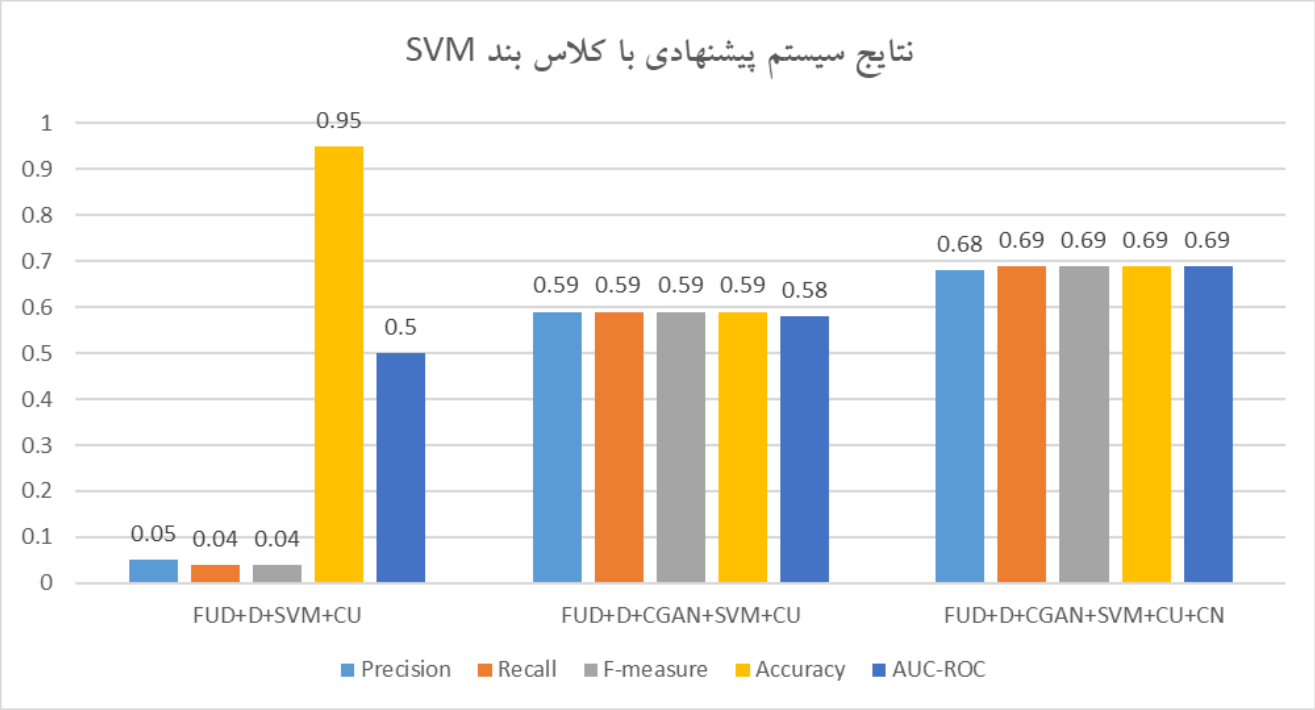
پارامتر	مقدار
تعداد پیاده‌روی به ازای هر گره[۸۴]	۱۰
طول گام‌های[۸۵] پیاده‌روی تصادفی	۸۰
اندازه پنجره skip-gram	۱۰

۴،۵،۳ بررسی تأثیر ویژگی‌ها روی سیستم پیشنهادی

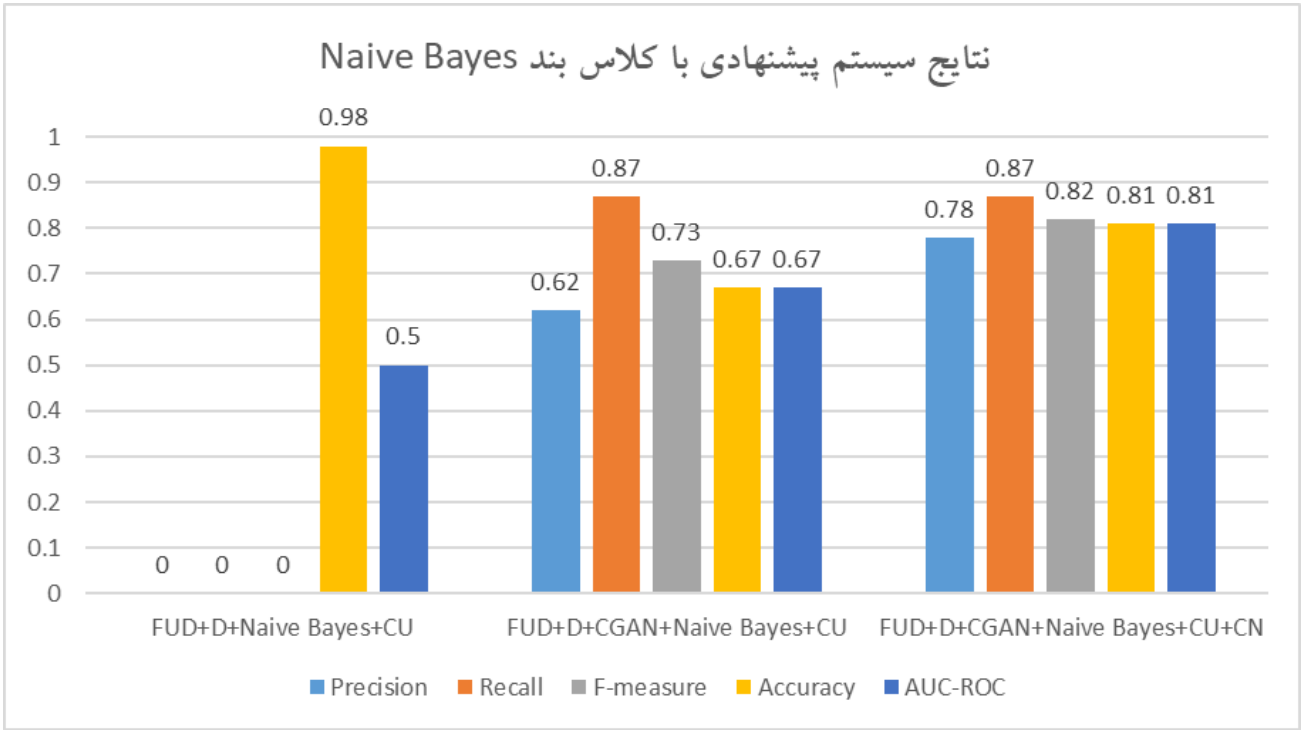
علائم اختصاری برای کاربر منتشرکننده اخبار جعلی با FUD، شبکه مولد متخاصم شرطی با CGAN، ورودی قطری با D و ویژگی مبتنی بر بافتار-کاربر[۸۶] با CU و ویژگی مبتنی بر بافتار-شبکه[۸۷] با CN نام گذاری شده‌اند و از کلاس‌بند ماشین بردار پشتیبان و نیو بیز[۸۸] و کِی نزدیک‌ترین همسایه برای این آزمایش استفاده شده است.

در هر سه نمودار شکل‌های ۱۶، ۱۷ و ۱۸ به ترتیب از سمت چپ نتایج کلاس‌بندها را با استفاده از مجموعه داده نامتوازن در شناسایی کاربران منتشرکننده اخبار جعلی با درنظر گرفتن ویژگی‌های مبتنی بر بافتار-کاربر را نشان می‌دهد. همانطور که مشهود است، الگوریتم-های یادگیری ماشین در صورتیکه در کلاس‌های مجموعه داده توازن وجود نداشته باشد، نتایج خوبی از خود نشان نمی‌دهند. بنابراین، در این مقاله برای ایجاد توازن از شبکه مولد متخاصم شرطی، کمک گرفته شده است. همانطور که مشهود است، صحت نمی‌تواند معیار خوبی برای ارزیابی در مجموعه داده‌های نامتوازن باشد؛ به همین منظور، از معیارهای دیگری نیز استفاده شده است. سپس، نتایج کلاس‌بندها با استفاده از مجموعه داده متوازن‌شده با شبکه مولد متخاصم شرطی با تمرکز بر ویژگی‌های مبتنی بر بافتار-کاربر نشان داده شده است. در آخر، نتایج کلاس‌بندها با استفاده از مجموعه داده‌ی متوازن‌شده با شبکه مولد متخاصم شرطی با تمرکز بر ترکیب ویژگی‌های مبتنی بر بافتار-کاربر و ویژگی‌های مبتنی بر بافتار-شبکه نشان داده شده است.

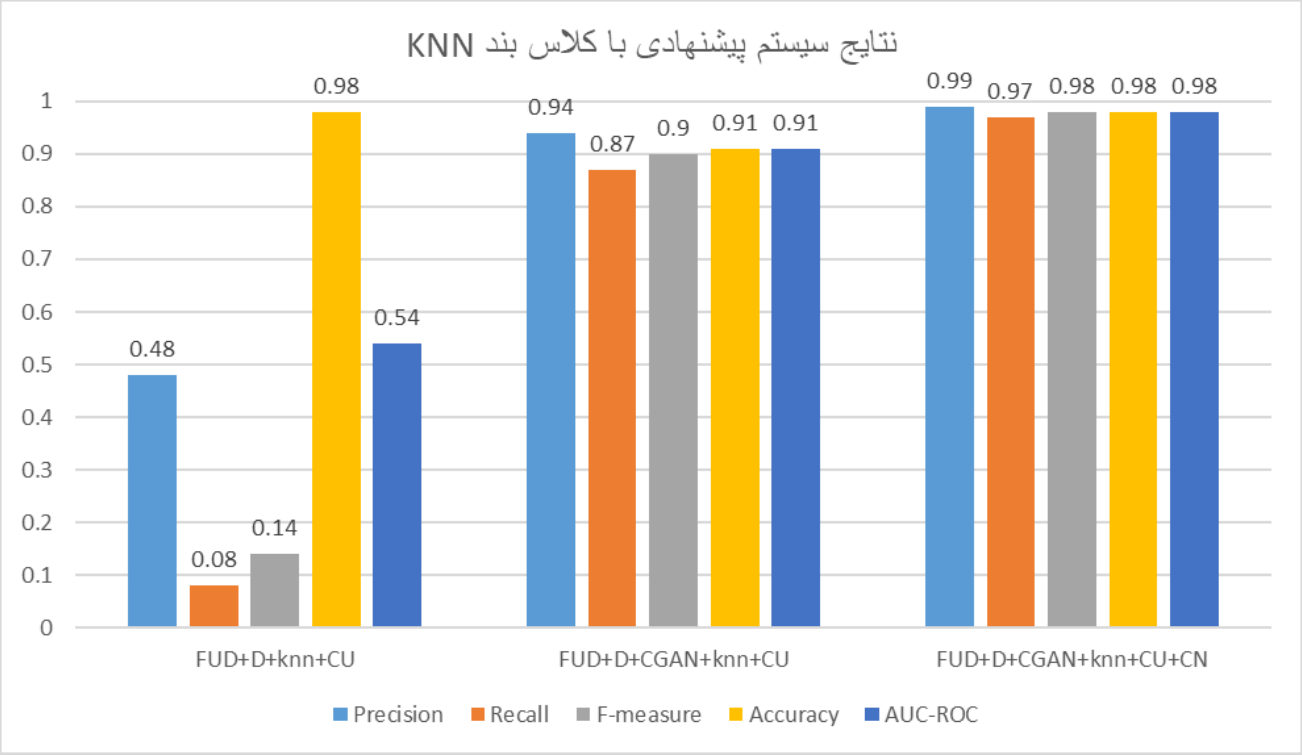
[۸۴] Number of walk per node
[۸۵] Walk Length
[۸۶] Context-User (CU)
[۸۷] Context-Network (CN)
[۸۸] Naive Bayes



شکل ۱۶. به ترتیب از سمت چپ نتایج استفاده از مجموعه داده نامتوازن و سپس، نتایج استفاده ازمجموعه داده متوازن‌شده با CGAN با تمرکز بر ویژگی بافتار-کاربر و در آخر، نتایج استفاده ازمجموعه داده متوازن‌شده با CGAN با تمرکز برترکیب ویژگی‌های بافتار-کاربر و بافتار-شبکه با کلاس‌بند SVM نشان داده شده است.



شکل ۱۷. به ترتیب از سمت چپ نتایج استفاده از مجموعه داده نامتوازن و سپس، نتایج استفاده ازمجموعه داده متوازن‌شده با CGAN با تمرکز بر ویژگی بافتار-کاربر و در آخر، نتایج استفاده ازمجموعه داده متوازن‌شده با CGAN با تمرکز برترکیب ویژگی‌های بافتار-کاربر و بافتار-شبکه با کلاس‌بند Naïve Bayes نشان داده شده است.

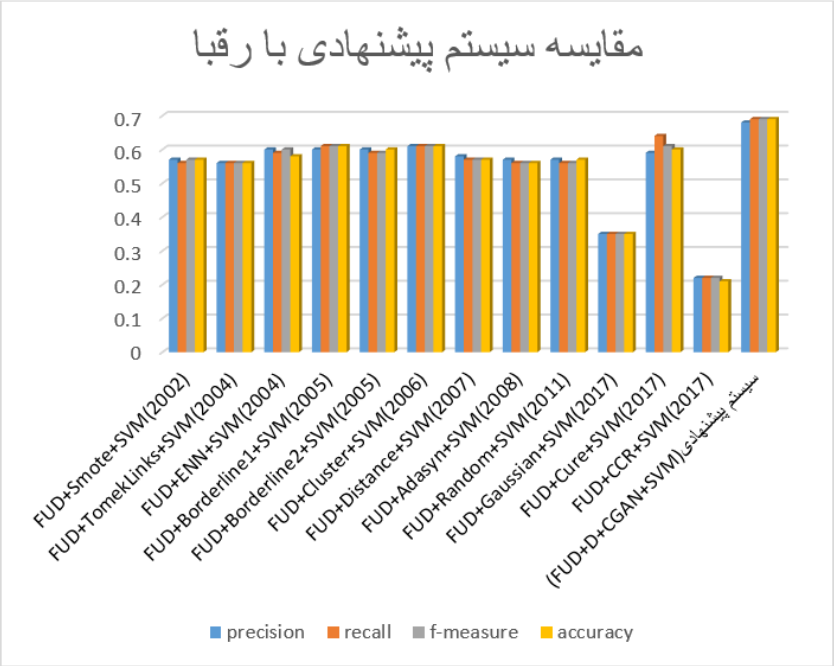


شکل ۱۸. به ترتیب از سمت چپ نتایج استفاده از مجموعه داده نامتوازن و سپس، نتایج استفاده ازمجموعه داده متوازن‌شده با CGAN با تمرکز بر ویژگی بافتار-کاربر و در آخر، نتایج استفاده ازمجموعه داده متوازن‌شده با CGAN با تمرکز برترکیب ویژگی‌های بافتار-کاربر و بافتار-شبکه با کلاس‌بند KNN نشان داده شده است.

همانطور که در نمودارها مشاهده می‌شود، ترکیب دو ویژگی و ایجاد توازن در مجموعه داده با کمک شبکه مولد متخاصم شرطی، در عملکرد سیستم پیشنهادی نهایی (FUD+D+CGAN+CU+CN)، در هر سه کلاس‌بند بهبود ایجاد کرده است. به طور مثال، در شکل ۱۸ سیستم پیشنهادی توانسته در کلاس‌بند KNN به اعداد ۹۹٪، ۹۷٪ ، ۹۸٪، ۹۸٪ و ۹۸٪ به ترتیب در معیارهای ارزیابی دقت، فراخوانی، معیار-اف، صحت و AUC-ROC دست پیدا کند.

۴,۶ مقایسه سیستم پیشنهادی با رقبا

در این مقاله برای تولید داده مصنوعی از شبکه مولد متخاصم شرطی استفاده شده است که بهبود عملکرد آن نسبت به استفاده از روش‌های سایر رقبا برای تولید داده مصنوعی با تمرکز بر ترکیب ویژگی‌های مبتنی بر بافتار-کاربر و مبتنی بر بافتار-شبکه در شکل ۱۹ نشان داده شده است.



شکل ۱۹. مقایسه سیستم پیشنهادی(شبکه مولد متخاصم شرطی) بانسخه‌های متفاوت روش بیش نمونه‌برداری اقلیت مصنوعی در ترکیب ویژگی‌های مبتنی بر بافتار-کاربر و مبتنی بر بافتار-شبکه

همانطور که در نمودار شکل ۱۹ مشاهده می‌شود، سیستم پیشنهادی این مقاله از تمام روش‌های موجود برای متوازن‌سازی داده نام‌برده-شده عملکرد بهتری داشته است. برای مثال، حتی در مقایسه با نزدیک‌ترین الگوریتم یعنی الگوریتم CURE به مقدار۹٪، ۵ ٪، ۸٪، ۹ ٪ و هم‌چنین، به نسبت الگوریتم پایه‌ای مانند SMOTE به مقدار ۱۱ ٪، ۱۳ ٪، ۱۲٪ و ۱۲ ٪ به ترتیب در معیارهای دقت، فراخوانی، معیار اف و صحت بهتر عمل کرده است. باید اشاره داشت که با توجه به عملکرد بهتر نسبت به رقبا، سیستم پیشنهادی به دلیل استفاده از یادگیری عمیق از نظر پیچیدگی و زمان اجرا هزینه بالاتری به نسبت رقبایش دارد که در جدول ۷ این مقایسه نشان داده شده است.

جدول ۷. زمان اجرا (برحسب ثانیه) سیستم پیشنهادی در مقایسه با رقبا

نام الگوریتم	SMOTE	TomekLinks	ENN	Borderline۱	Borderline۲	Cluster	Distance	Adasyn	Random	Gaussian	Cure	CCR	سیستم پیشنهادی (استفاده از CGAN برای متوازن‌سازی داده)
زمان اجرا	۲۹/۹۶	۲۷۷/۸۸	۲۷۱/۱۵	۲۲/۵۶	۲۵/۴۳	۴۷/۹۷	۱۷/۴۸	۲۳/۵۱	۱۷/۳	۲۰/۲۶	۴۶/۲۹	۳۹/۴۲	۹۵۵۹/۳۶

۵ نتیجه‌گیری

با توجه به فراگیری شبکه‌های اجتماعی در بین مردم و امکان انتشار بیشتر اخبار و اطلاعات نادرست نسبت به گذشته و هم‌چنین، اهمیت بالای شناسایی منبع منتشرکننده این اطلاعات نادرست، در این مقاله، یک سیستم برای شناسایی کاربران منتشرکننده اخبار جعلی که اقدام به انتشار نادرست در توئیتر در زبان فارسی کرده‌اند، پیشنهاد شده است. این سیستم بر مبنای استفاده از ویژگی‌های مبتنی بر بافتار یعنی ترکیب ویژگی‌های مبتنی بر کاربر و مبتنی بر شبکه پایه‌ریزی شده است، که برای استخراج ویژگی‌های مبتنی بر بافتار-کاربر از اطلاعات کاربران و برای استخراج ویژگی‌های مبتنی بر بافتار-شبکه از تعبیه گره به بردار(Node۲vec) برای تبدیل گراف تعاملات کاربران به بردار ویژگی کمک گرفته شده است. ضمناً، به دلیل عدم توازن در مجموعه داده از شبکه مولد متخاصم شرطی برای رفع این

چالش استفاده شده است تا با تولید داده مصنوعی مجموعه داده به تعادل برسد. هم‌چنین، عملکرد سیستم پیشنهادی به کمک کلاس-بندها در طی دو سناریو تحلیل پارامتر حساسیت و مقایسه با رقبا بررسی شد. از دستاوردهای دیگر این مقاله می‌توان به ایجاد و گسترش مجموعه داده جدید برای شناسایی کاربران منتشرکننده اخبار جعلی در شبکه توئیتر در زبان فارسی که منابع زبان‌شناسی-کمتری دارد، به نام "FU_KNTU " در مدت وقوع زلزله کرمانشاه سال ۱۳۹۶ ایران اشاره کرد. با توجه به این موضوع که اکثر پژوهش-های اخیر در این حوزه بر روی مجموعه داده متوازن صورت گرفته است، از توجه به مجموعه داده‌های نامتوازن که در دنیای واقعی وجود دارد، غفلت شده است. بنابراین، از مهم‌ترین برتری‌های سیستم پیشنهادی به طور متمایز نسب به پژوهش‌های پیشین، می‌توان به رفع چالش مجموعه داده نامتوازن با ایده‌ای جدید که در واقع، متوازن‌سازی با روش یادگیری عمیق به نام شبکه مولد متخاصم شرطی است، اشاره داشت. در نهایت نشان داده شد، سیستم پیشنهادی با یادگیری توزیع‌داده سراسری تا حدود ۱۱٪ ، ۱۳ ٪ ۱۲، ٪ و ۱۲ ٪ به ترتیب در معیارهای دقت، فراخوانی، معیار اف و صحت نسبت به رقبایش که بر روی یادگیری توزیع داده محلی تمرکز دارند، بهبود داشته است و توانسته است دقتی در حدود ۹۹٪ در شناسایی کاربران منتشرکننده اخبار جعلی ایجاد کند. ضمناً، با ترکیب ویژگی-های مبتنی بر بافتار-کاربر و بافتار-شبکه عملکرد سیستم پیشنهادی افزایش داشته است. هم‌چنین، با ترکیب این دو ویژگی، مشکل شروع سرد در شبکه وجود نخواهد داشت. اما باید به اینکه اشاره داشت که شبکه مولد متخاصم به دلیل استفاده از یادگیری عمیق زمان آموزش و پیچیدگی بیشتری نسبت به رقبایش دارد. در آخر، از کارهای آتی می‌توان به تغییر ورودی شبکه مولد متخاصم شرطی اشاره کرد تا با عدم تبدیل ورودی به ماتریس، مقدار خطای حاصل از این فرض کاهش یابد. علاوه بر این، ترکیب ویژگی‌های مبتنی بر محتوا و ویژگی‌های ذکرشده و هم‌چنین، تنظیم سایر پارامترها با آموزش بر روی مجموعه داده، سیستم دقیق‌تری خواهد ساخت. نهایتاً، روشی برای رفع پیچیدگی وکاهش زمان آموزش شبکه مولد متخاصم شرطی ارائه داد.

مراجع

[۱] Parikh, S.B. and P.K. Atrey. "Media-rich fake news detection: A survey. " in ۲۰۱۸ IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). ۲۰۱۸. IEEE.

[۲] Kochkina, E., M. Liakata, and A. Zubiaga, "All-in-one: Multi-task learning for rumour verification. " arXiv preprint arXiv:۱۸۰۶,۰۳۷۱۳, ۲۰۱۸.

[۳] Tacchini, E., et al., "Some like it hoax: Automated fake news detection in social networks. " arXiv preprint arXiv:۱۷۰٤,۰۷۵۰٦, ۲۰۱۷.

[٤] Shu, K., et al., "Fake news detection on social media: A data mining perspective. " ACM SIGKDD explorations newsletter, ۲۰۱۷. ۱۹(۱): p. ۲۲-۳٦.

[٥] Inuwa-Dutse, I., M. Liptrott, and I. Korkontzelos, "Detection of spam-posting accounts on Twitter. " Neurocomputing, ۲۰۱۸. ۳۱٥: p. ٤۹٦-٥۱۱.

[٦] Bindu, P., R. Mishra, and P.S. Thilagam, "Discovering spammer communities in Twitter. " Journal of Intelligent Information Systems, ۲۰۱۸. ٥۱(۳): p. ٥۰۳-٥۲۷.

[۷] de Souza, J.V., et al., "A systematic mapping on automatic classification of fake news in social media. " Social Network Analysis and Mining, ۲۰۲۰. ۱۰(۱): p. ۱-۲۱.

[۸] Grinberg, N., et al., " Fake news on Twitter during the ۲۰۱٦ US presidential election. " Science, ۲۰۱۹. ۳٦۳(٦٤۲٥): p. ۳۷٤-۳۷۸.

[۹] Maaten, L.v.d. and G. Hinton, "Visualizing data using t-SNE. " Journal of machine learning research, ۲۰۰۸. ۹(Nov): p. ۲٥۷۹-۲٦۰٥.

[۱۰] Gheewala, S. and R. Patel. "Machine learning based Twitter Spam account detection: a review. " in ۲۰۱۸ Second International Conference on Computing Methodologies and Communication (ICCMC). ۲۰۱۸. IEEE.

[۱۱] Gaonkar, S., et al. " Detection Of Online Fake News: A Survey. " in ۲۰۱۹ International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN). ۲۰۱۹. IEEE.

[۱۲] Hardalov, M., I. Koychev, and P. Nakov. " In search of credible news. " in International Conference on Artificial Intelligence: Methodology, Systems, and Applications. ۲۰۱٦. Springer.

[۱۳] Goodfellow, I., et al. "Generative adversarial nets. " in Advances in neural information processing systems. ۲۰۱٤.

[۱٤] Douzas, G. and F. Bacao, "Effective data generation for imbalanced learning using conditional generative adversarial networks. "

[۱۵] Mirza, M. and S. Osindero, "Conditional generative adversarial nets. " arXiv preprint arXiv:۱۴۱۱,۱۷۸۴, ۲۰۱۴.

[۱۶] Grover, A. and J. Leskovec. "node۲vec: Scalable feature learning for networks. " in Proceedings of the ۲۲nd ACM SIGKDD international conference on Knowledge discovery and data mining. ۲۰۱۶. ACM.

[۱۷] Conroy, N.K., V.L. Rubin, and Y. Chen, "Automatic deception detection: Methods for finding fake news. " Proceedings of the Association for Information Science and Technology, ۲۰۱۵. ۵۲(۱): p. ۱-۴.

[۱۸] Bondielli, A. and F. Marcelloni, "A survey on fake news and rumour detection techniques. " Information Sciences, ۲۰۱۹. ۴۹۷: p. ۳۸-۵۵.

[۱۹] Mohammadrezaei, M., M.E. Shiri, and A.M. Rahmani, "Identifying fake accounts on social networks based on graph analysis and classification algorithms. " Security and Communication Networks, ۲۰۱۸. ۲۰۱۸.

[۲۰] Yang, C., R. Harkreader, and G. Gu, "Empirical evaluation and new design for fighting evolving twitter spammers. " IEEE Transactions on Information Forensics and Security, ۲۰۱۳. ۸(۸): p. ۱۲۸۰-۱۲۹۳.

[۲۱] Wang, A.H. "Don't follow me: Spam detection in twitter. " in ۲۰۱۰ international conference on security and cryptography (SECRYPT). ۲۰۱۰. IEEE.

[۲۲] Benevenuto, F., et al. "Detecting spammers on twitter. " in Collaboration, electronic messaging, anti-abuse and spam conference (CEAS). ۲۰۱۰.

[۲۳] Masood, Faiza, et al. "Spammer detection and fake user identification on social networks." IEEE Access ۷ (۲۰۱۹): ۶۸۱۴۰-۶۸۱۵۲.

[۲۴] Xie, Y., et al. "A Fake News Detection Framework Using Social User Graph. " in Proceedings of the ۲۰۲۰ ۲nd International Conference on Big Data Engineering. ۲۰۲۰.

[۲۵] KARUNAKAR, M.G., et al., " ADAPTIVE DETECTING FAKE PROFILES IN ONLINE SOCIAL NETWORKS. "

[۲۶] Della Vedova, M.L., et al. "Automatic online fake news detection combining content and social signals. " in ۲۰۱۸ ۲۲nd Conference of Open Innovations Association (FRUCT). ۲۰۱۸. IEEE.

[۲۷] Shu, K., et al. "defend: Explainable fake news detection. " in Proceedings of the ۲۵th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ۲۰۱۹.

[۲۸] Guacho, G.B., et al. "Semi-supervised content-based detection of misinformation via tensor embeddings. " in ۲۰۱۸ IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). ۲۰۱۸. IEEE.

[۲۹] Shu, K., et al. "The role of user profiles for fake news detection. " in Proceedings of the ۲۰۱۹ IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. ۲۰۱۹.

[۳۰] Shu, K., S. Wang, and H. Liu. "Beyond news contents: The role of social context for fake news detection. " in Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining. ۲۰۱۹.

[۳۱] Hamdi, T., et al. "A Hybrid Approach for Fake News Detection in Twitter Based on User Features and Graph Embedding. " in International Conference on Distributed Computing and Internet Technology. ۲۰۲۰. Springer.

[۳۲] Aphiwongsophon, S. and P. Chongstitvatana. "Detecting fake news with machine learning method. " in ۲۰۱۸ ۱۵th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON). ۲۰۱۸. IEEE.

[۳۳] Hussain, M.G., et al., "Detection of Bangla Fake News using MNB and SVM Classifier. " arXiv preprint arXiv:۲۰۰۵,۱۴۶۲۷, ۲۰۲۰.

[۳۴] Li, Y., et al., "Exploiting similarities of user friendship networks across social networks for user identification. " Information Sciences, ۲۰۲۰. ۵۰۶: p. ۷۸-۹۸.

[۳۵] Vijayaraghavan, S., et al., "Fake News Detection with Different Models. " arXiv preprint arXiv:۲۰۰۳,۰۴۹۷۸, ۲۰۲۰.

[۳۶] Jadhav, S.S. and S.D. Thepade, "Fake news identification and classification using DSSM and improved recurrent neural network classifier. " Applied Artificial Intelligence, ۲۰۱۹. ۳۳(۱۲): p. ۱۰۵۸-۱۰۶۸.

[۳۷] Ajao, O., D. Bhowmik, and S. Zargari. "Fake news identification on twitter with hybrid cnn and rnn models. " in Proceedings of the ۹th international conference on social media and society. ۲۰۱۸.

[۳۸] Zhang, J., B. Dong, and S.Y. Philip. "Fakedetector: Effective fake news detection with deep diffusive neural network. " in ۲۰۲۰ IEEE ۳۶th International Conference on Data Engineering (ICDE). ۲۰۲۰. IEEE.

[۳۹] Verma, A., V. Mittal, and S. Dawn. "FIND: Fake information and news detections using deep learning. " in ۲۰۱۹ Twelfth International Conference on Contemporary Computing (IC۳). ۲۰۱۹. IEEE.

[۴۰] Ruan, N., R. Deng, and C. Su, "GADM: Manual fake review detection for O۲O commercial platforms. " Computers & Security, ۲۰۲۰. ۸۸: p. ۱۰۱۶۵۷.

[۴۱] Hosseinimotlagh, S. and E.E. Papalexakis. "Unsupervised content-based identification of fake news articles with tensor decomposition ensembles. " in Proceedings of the Workshop on Misinformation and Misbehavior Mining on the Web (MIS۲). ۲۰۱۸.

[۴۲] Yang, S., et al. "Unsupervised fake news detection on social media: A generative approach. " in Proceedings of the AAAI Conference on Artificial Intelligence. ۲۰۱۹.

[۴۳] Phan, T.D. and N. Zincir- Heywood, "User identification via neural network based language models. " International Journal of Network Management, ۲۰۱۹. ۲۹(۳): p. e۲۰۴۹.

[۴۴] Mateen, M., et al. "A hybrid approach for spam detection for Twitter. " in ۲۰۱۷ ۱۴th International Bhurban Conference on Applied Sciences and Technology (IBCAST). ۲۰۱۷. IEEE.

[۴۵] Chen, C., et al., "Statistical features-based real-time detection of drifted twitter spam. " IEEE Transactions on Information Forensics and Security, ۲۰۱۶. ۱۲(۴): p. ۹۱۴-۹۲۵.

[۴۶] Volkova, S., et al. "Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on twitter. " in Proceedings of the ۵۵th Annual Meeting of the Association for Computational Linguistics (Volume ۲: Short Papers). ۲۰۱۷.

[۴۷] Mahmoodabad, S.D., S. Farzi, and D.B. Bakhtiarvand. "Persian rumor detection on twitter. " in ۲۰۱۸ ۹th International Symposium on Telecommunications (IST). ۲۰۱۸. IEEE.

[۴۸] Wang, W., et al. "Global-and-Local Aware Data Generation for the Class Imbalance Problem. " in Proceedings of the ۲۰۲۰ SIAM International Conference on Data Mining. ۲۰۲۰. SIAM.

[۴۹] Rout, N., D. Mishra, and M.K. Mallick, "Handling imbalanced data: A survey", in International Proceedings on Advances in Soft Computing, Intelligent Systems and Applications. ۲۰۱۸, Springer. p. ۴۳۱-۴۴۳.

[۵۰] Chen, H. and L. Jiang, " Efficient GAN-based method for cyber-intrusion detection. " arXiv preprint arXiv:۱۹۰۴,۰۲۴۲۶, ۲۰۱۹.

[۵۱] Lee, J. and K. Park, "GAN-based imbalanced data intrusion detection system. " Personal and Ubiquitous Computing, ۲۰۱۹: p. ۱-۸.

[۵۲] Kim, J.-Y., S.-J. Bu, and S.-B. Cho. "Malware detection using deep transferred generative adversarial networks. " in International Conference on Neural Information Processing. ۲۰۱۷. Springer.

[۵۳] Radford, A., L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks. " arXiv preprint arXiv:۱۵۱۱,۰۶۴۳۴, ۲۰۱۵.

[۵۴] Kovács, G., "An empirical comparison and evaluation of minority oversampling techniques on a large number of imbalanced datasets. " Applied Soft Computing, ۲۰۱۹. ۸۳: p. ۱۰۵۶.۶۲.

[۵۵] Chawla, N.V., et al., "SMOTE: synthetic minority over-sampling technique. " Journal of artificial intelligence research, ۲۰۰۲. ۱۶: p. ۳۲۱-۳۵۷.

[۵۶] Batista, G.E., R.C. Prati, and M.C. Monard, "A study of the behavior of several methods for balancing machine learning training

[۶۵] Breuer, Adam, Roee Eilat, and Udi Weinsberg. "Friend or Faux: Graph-Based Early Detection of Fake Accounts on Social Networks." Proceedings of The Web Conference ۲۰۲۰. ۲۰۲۰.

[۶۶] Liu, Yang, and Yi-Fang Brook Wu. "FNED: A Deep Network for Fake News Early Detection on Social Media." ACM Transactions on Information Systems (TOIS) ۳۸,۳ (۲۰۲۰): ۱-۳۳.

[۶۷] Liao, Hao, Qixin Liu, and Kai Shu. "Incorporating User-Comment Graph for Fake News Detection." arXiv preprint arXiv:۲۰۱۱,۰۱۵۷۹ (۲۰۲۰).

[۶۸] Balaanand, Muthu, et al. "An enhanced graph-based semi-supervised learning algorithm to detect fake users on Twitter." The Journal of Supercomputing ۷۵,۹ (۲۰۱۹): ۶۰۸۵-۶۱۰۵.

[۶۹] Fawcett, Tom. "An introduction to ROC analysis." Pattern recognition letters ۲۷,۸ (۲۰۰۶): ۸۶۱-۸۷۴.

data. " ACM SIGKDD explorations newsletter, ۲۰۰۴. ۶(۱): p. ۲۰-۲۹.

[۵۷] Han, H., W.-Y. Wang, and B.-H. Mao. "Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. " in International conference on intelligent computing. ۲۰۰۵. Springer.

[۵۸] Cieslak, D.A., N.V. Chawla, and A. Striegel. "Combating imbalance in network intrusion datasets. " in GrC. ۲۰۰۶.

[۵۹] De La Calleja, J. and O. Fuentes. "A Distance-Based Over-Sampling Method for Learning from Imbalanced Data Sets. " in FLAIRS Conference. ۲۰۰۷.

[۶۰] He, H., et al. "ADASYN: Adaptive synthetic sampling approach for imbalanced learning. " in ۲۰۰۸ IEEE international joint conference on neural networks (IEEE world congress on computational intelligence). ۲۰۰۸. IEEE.

[۶۱] Dong, Y. and X. Wang. "A new over-sampling approach: random-SMOTE for learning from imbalanced data sets. " in International Conference on Knowledge Science, Engineering and Management. ۲۰۱۱. Springer.

[۶۲] Lee, H., J. Kim, and S. Kim, "Gaussian-Based SMOTE Algorithm for Solving Skewed Class Distributions. " International Journal of Fuzzy Logic and Intelligent Systems, ۲۰۱۷. ۱۷(۴): p. ۲۲۹-۲۳۴.

[۶۳] Ma, L. and S. Fan, "CURE-SMOTE algorithm and hybrid algorithm for feature selection and parameter optimization based on random forests. " BMC bioinformatics, ۲۰۱۷. ۱۸(۱): p. ۱-۱۸.

[۶۴] Koziarski, M. and M. Woźniak, "CCR: A combined cleaning and resampling algorithm for imbalanced data classification. " International Journal of Applied Mathematics and Computer Science, ۲۰۱۷. ۲۷(۴): p. ۷۲۷-۷۳۶.