مبارزه با ریسک‌های مالی نوظهور در عصر داده‌های بزرگ

فاطمه دادبه

سمیه همتی

۱. مقدمه

با گسترش داده‌های بزرگ، زندگی روزمرهٔ مردم به‌طور کامل از جمله فعّالیت‌های مالی آنها ثبت شده است. برای مثال، گوشی‌های هوشمند یا جی‌پی‌اس و تکنولوژی بلوتوث، استفادهٔ مردم را از بانک‌ها و ای‌تی‌ام‌ها، مال‌ها یا ساختمان‌های اداری دنبال می‌کند که به‌طور مستمر فعّالیت‌ها و هر تماس فیزیکی ممکن، زمان سپری شده در شبکه‌های اجتماعی، تحقیق در جداول و عکس‌های ماهواره‌ای از وضعیت ساختمان‌ها و تعداد ماشین‌های پارک شده در پارکینگ در یک زمان خاص را ثبت می‌کند. استفاده و کنترل این داده‌ها می‌تواند به شناخت بهتر و مقابله

با ریسک‌های مالی پنهان کمک کند. در همین حال، گستره‌ای از الگوریتم‌ها برای تکنولوژی‌های نوظهور، به‌طور گسترده در سیستم‌های مربوط به مالی مورد استفاده قرار می‌گیرند. توصیه‌هایی از طریق الگوریتم‌های هوش مصنوعی ایجاد می‌شوند که تصمیم می‌گیرند چه آگهی‌ها، اقلام خرید و چه اخباری باید اولویت‌بندی شود. مدل‌ها و تکنولوژی‌های فرایند زبان طبیعی[1] برای شناخت، از احساسات و عقاید کاربران اتخاذ شده است. شبکه‌های عصبی عمیق و الگوریتم‌های هوشمند در وسایلی برای نظارت به موقع، Q&A هوشمند و دستیار روبات‌ها در میان افراد دیگر گنجانده شده‌اند.

امروزه، مالی به یک حوزهٔ بسیار پیچیده تبدیل شده است؛ دیگر محدود به بانکداری سنتی و سیستم‌های معاملات، سهام و بازارهای آتی نیست امّا اگرچه واحدهای پول دیجیتال نوظهور، خرید آنلاین و سیستم‌های پرداخت را در بر می‌گیرد، حتی با شبکه‌های اجتماعی و شبکه، کمین‌های سیاسی و محدوده سیستم‌های هوشمند شامل اینترنت اشیا، همپوشانی دارد. برای مثال، اخبار شکست می‌توانند پایه‌های سیستم مالی را تکان دهد یا یک سیستم هوشمند در مورد برنامه‌ریزی مسیر می‌توانند تصمیم بگیرد که چگونه موجودی‌ها می‌توانند به‌طور کامل تحویل داده شوند. فین تک، ترکیبی از اینترنت، داده‌های بزرگ و تکنولوژی‌های مالی، کسب و کارهای جدید در شاخه‌های مختلف مثل پرداخت، سپرده، وام، سرمایه‌گذاری و امکانات بازار را افزایش می‌دهد که تأثیر قابل ملاحظه‌ای بر مدل‌های کسب و کار مالی سنتی و خطرات بزرگ پنهان می‌گذارند. بیشتر از این، گردش ارزهای دیجیتال گوناگون، از قبیل بیت‌کوین، شکل جدیدی از ارز خارجی، مدیریت ارز خارجی یک کشور را به چالش می‌کشد. بنابراین، ریسک‌های مالی در عصر داده‌های بزرگ، مرزهای داده‌ها، الگوریتم‌ها و سیستم‌های سنتی را در هم می‌شکند و سیستم‌ها، چالش‌های بیشتری را نسبت به همیشه ایجاد می‌کنند.

۱،۱. ریسک‌های مالی در عصر داده‌های بزرگ

براساس گزارش توسعهٔ مالی اینترنتی چین در سال ۲۰۲۰ که توسط آی ریسرچ[2] منتشر شده است، سرمایه‌گذاری در تکنولوژی‌های مالی ۱۱۲ میلیارد یوآن، دارای افزایشی معادل۱۹,۴درصد بود، در حالی که سرمایه‌گذاری در هوش مصنوعی در بانکداری۱۴,۳ میلیارد یوآن، افزایشی معادل ۲۸,۸درصد داشت. علاوه بر این، سه حوزهٔ سرمایه‌گذاری ممتاز، کنترل ریسک هوشمند، بیمهٔ هوشمند و خدمات هوشمند مشتری هستند که در برگیرندهٔ بیش از ۷۰ درصد از کل سرمایه‌گذاری‌ها هستند. یکی از اثرات جانبی سیاست پذیرش برای شرکت‌های با تکنولوژی بالادر بازار این است که تعداد زیادی از نهادهای مالی غیرسنتی بدون مکانیسم‌های کنترل ریسک، در محل کار، درگیر کسب و کارهای با ریسک بالا هستند. به‌طور مستمر، تقلّب مالی، محدودهٔ گسترده‌ای از سرمایه‌گذاران و اجتماعاتی که ممکن است را تحت تأثیر قرار می‌دهد. متقلّبان مالی مشهور، ایزوبااو[3] حجم معاملات تجمعی هفتاد میلیارد یوآن از سال ۲۰۱۴ تا ۲۰۱۵ را دارد.

ریسک‌های مالی شکل‌های متفاوتی دارند و وابسته به محدوده‌ای از عوامل در عصر داده‌های بزرگ از قبیل سوءاستفاده از الگوریتم و تکنولوژی‌های هوش مصنوعی، واحدهای پول دیجیتال و کسب و کارهای آنلاین هستند. برای بررسی این چالش‌ها برای ریسک‌های مالی نوظهور، باید یک تاکسونومی مشهور را در ریسک‌های مالی دنبال کرد که به‌عنوان ریسک‌های بازار، اعتباری، نقدشوندگی، نوسانات، عملیات و تطبیق- جرایم مالی[4] طبقه‌بندی

می‌شوند. با این حال، بررسی‌ها اساساً بر ریسک‌های اعتباری، نقدینگی، تطبیق- جرایم مالی تمرکز دارد، چرا که مدیریت و کشف آنها همیشه نیازمند تحلیل داده‌ها با گسترهٔ وسیع و ناهمگون هستند. علاوه بر این، هر نوع ریسکی اگر به‌طور نامناسب مدیریت شود، می‌تواند منجر به ریسک سیستمیک شود که می‌تواند منجر به سقوط کل سیستم مالی یا همهٔ بازار شود یا حتی می‌تواند منجر به بحران‌های اقتصادی شود.

چالش‌هایی برای مدیریت ریسک مالی

چالش‌های اصلی مبارزه با ریسک‌های مالی در عصر داده‌های بزرگ می‌تواند به‌صورت زیر خلاصه شود:

- نیاز به استفاده از داده‌های چند شکلی. داده‌های بزرگ، روشی پایدار برای مدیریت ریسک‌های مالی فراهم می‌کنند. داده‌ها معمولاً از چندین منبع و کانال از قبیل گزارش‌های مالی، نمودار داده‌های فروش، اطلاعات غیر سنتی از متن‌ها، تصاویر و ویدئوهای شبکه‌های اجتماعی، عکس‌های ماهواره‌ای و داده‌های تحرک انسانی به‌دست می‌آیند. این داده‌ها معنادار هستند و برخی به‌طور کامل اطلاعاتی را برای دیگری فراهم می‌کنند، بنابراین انعکاس سیگنال زودهنگام برای برآوردهای قابل اعتماد از ریسک‌هایی که قابل مشاهده نیستند وقتی که با تک تک کانال‌ها کار می‌کنند، فراهم می‌کنند. با این حال، تلفیق ناهمگون، داده‌های قطع شده از کانال‌های چندگانه، مدل‌های سنتی و روش‌های آماری در مدیریت ریسک مالی از قبیل رگرسیون خطی، روش‌های بیز و مدل‌های مارکوف کلاسیک چالش برانگیز است.
- ماهیت وابسته به دامنه بلند و ناهمگون. داده‌ها اشیا اغلب مرتبط با هم هستند. برای مثال؛ شرکت‌ها کسب و کار وابسته به هم هستند و پول از یک حساب به حساب دیگر منتقل می‌شود. علاوه بر این، روابط برای برخی از داده‌ها از قبیل شبکه‌های سهامداران و شرکت‌ها، ناهمگن هستند. برخی از مهمترین ریسک‌های مالی مهم می‌توانند تنها از طریق وابسته دامنه بلند شناسایی شود در حالی که روابط وابستگی دامنه کوتاه این ریسک‌ها از قبیل بحران وام‌های بدون پشتوانه در سال ۲۰۰۸ نمی‌توانند منعکس شوند. با این حال، داده کاوی الگوهای وابسته دامنه بلند و ناهمگون برای شناسایی ریسک‌ها، با وجود ده‌ها میلیارد روابط بین داده‌ها، بسیار چالش برانگیز است.
- ویژگی‌های پویا و هم‌زمان. بازار مالی جهانی بسیار پویا است و داده‌ها که از سازمان‌های مالی ایجاد شده‌اند هم‌زمان از نظر محتوا و دارایی تغییر می‌کنند؛ در نتیجه در ردیابی و تحلیل دشواری وجود دارد.
- هماورد. تقلب مالی و پولشویی از ریسک‌های تطبیق – جرایم مالی وابسته به تکنیک‌های پیشرفتهٔ جدید و روش‌های استتار است و تشخیص این موارد دشوار است.

۱،۲. پروژه‌های تحقیقاتی مرتبط

در چندین سال گذشته، پروژه‌های تحقیقاتی مرتبط، به منظور مطالعه دربارهٔ مشکلات چالش برانگیز بیان شده در تحقیقات گذشته، سرمایه‌گذاری شده است.

برنامه داده‌های بزرگ مؤسسهٔ ملی علم[۵] در پروژه‌های تحقیقاتی بسیاری بر مبنای علم داده سرمایه‌گذاری کرده است. «اکوسیستم شناخت بازارهای مالی» مدل‌های رفتاری جدیدی از معاملات مالی با استفاده از تکنیک‌های داده‌های بزرگ و نیز معیارها و داده‌های جدید برای رشتهٔ مالی در اقتصاد معرفی کرده است. «کشف دستکاری بازار مالی: داده‌های یکپارچه شده و رویکرد مدل محور» رویکردهای داده محور خلاقانه را برای بهبود کشف و بازدارندگی از دستکاری بازار به کار می‌برند. تمرکز اصلی پروژه بر به‌کارگیری شبیه‌سازی و بهینه‌سازی برای خلق استراتژی‌های دستکاری از طریق جریان داده‌های بازار است و این رفتارهای دستکاری را از طریق جعل امضاها و فعّالیت‌های فریب‌آمیز مشخص می‌کند. «تجزیه و تحلیل شبکه بر مبنای جریان داده‌های اقتصادی برای نظارت ثبات مالی» بر تعیین و پیش‌بینی شرکت‌کنندگان بازار تمرکز دارد که می‌تواند سیستم‌های کلی مالی را با بالا بردن سرمایه از طریق طیف وسیعی از جریان داده‌های مالی کمی متنوع، متاداده و آگهی‌های بازار، در معرض خطر قرار دهد.

علوم طبیعی ملی چین[۶] طرح تحقیقاتی بزرگی را با عنوان «تحقیق در مورد مدیریت داده‌های بزرگ محور و تصمیم‌گیری» را ارائه داد که یک سری پروژه‌های تحقیقاتی پایه‌ای مربوط به تأمین مالی اینترنت را پشتیبانی می‌کند. «نظارت مالی دادهٔ بزرگ محور و پلتفرم خدمات و کاربرد نمایش» مطالعاتی از قبیل نظارت مالی اینترنت و ایجاد نقشه‌های دانش برای مالی اینترنت را انجام داد. «مکانیسم خلق تحلیل، کشف و هماهنگی ارزش بر مبنای داده‌های بزرگ مالی انجمن دانش» بر کشف ارتباط در دانش و ساخت یک نمودار دانش[۲۱] در مقایس بزرگ با استفاده از رویکرد خلاقانه و سپس کاربرد این روش‌ها و نمودار دانش در مشکلات جهان واقعی، تمرکز دارد. اجرای این پروژه‌ها به شناخت بهتر تأمین مالی اینترنت کمک می‌کند و مبانی نظری و تکنولوژی‌ها برای نظارت و خدمات تأمین مالی اینترنت داده بزرگ محور فراهم می‌کند.

۲. داده‌ها

۲،۱. مجموعه داده‌ها

در حال حاضر، جمع‌آوری داده‌های بزرگ مالی نیازمند مدیریت تحصیل اشیا، کانال‌ها، فرکانس‌ها و نیز سفارشی‌سازی هدف محور است که معمولاً با استفاده از کاوشگران وب در پایگاه‌های داده یا بر روی منابع ابری اجاره شده به اجرا در می‌آید. با تعیین برخی اهداف یا سایت‌های مرتبط به‌عنوان نودهای اولیه کاوشگران وب با دنبال کردن استراتژی اول که از طریق داده‌های تکراری به همراه لینک‌های وب است آنها را بازیابی می‌کنند. به‌علاوه، لاگ‌های سیستم از قبیل لاگ‌های رویدادهای فرایند درخواست وام در یک بانک، فعّالیت مؤسسات مالی مختلف را در مقادیر زیادی ثبت می‌کنند که می‌تواند در تصمیم‌گیری‌های تجاری دقیق در بسیاری از حوزه‌ها شامل مدیریت ریسک، استفاده شود. در حالی که داده‌های ساختار یافته اغلب برای الگوریتم‌های داده‌های بزرگ آماده می‌شوند، داده‌های ساختار نیافته نیازمند پردازش بیشتر هستند. بسیاری از مدل‌های یادگیری عمیق به‌تازگی به وظایف خاص برای استخراج اطلاعات از داده‌های ساختار نیافته در مالی کمک می‌کنند.

به‌علاوه، چهارچوب‌های معمولی وجود دارد که دانش ساختار یافته از انبوهی از داده‌های ساختار نیافته را از قبیل تکست کیوب[۷] توسعه می‌دهند.

۲،۲. سازمان داده‌ها

نمودارهای دانش به‌عنوان یک تکنیک غنی و شهودی برای بیان دانش، توجه

جدول شماره ۱- مجموعه داده‌های عمومی			
مجموعه داده‌ها	نوع شبکه	اندازه داده	توضیحات
داده‌های مالی زچ Czech Financial Data	تراکنش‌های بانکی	05M .1	چالش کشف انتقال داده‌های ناشناس منتشر شده از بانک زچ
داده‌های الیپتیک Elliptic Data	بیت‌کوین	234K	نمودار تراکنش جمع‌آوری شده از بلاکچین بیت‌کوین
داده‌های تاریخی بیت‌کوین	بیت‌کوین	86M .4	داده‌های بیت‌کوین در فواصل یک دقیقه‌ای از صرافی‌های منتخب از ژانویه ۲۰۱۲ تا مارچ ۲۰۲۱
داده‌های مالی مصنوعی برای کشف تقلب	تراکنش‌های بانکی موبایل	36M .6	مجموعه داده‌ها مصنوعی ایجاد شده از طریق شبیه‌سازی پول موبایل پی سیم(PaySim)
داده‌های عظیم بازار سهام	سهام و صندوق‌های قابل معامله در بورس	5M .17	قیمت‌ها و حجم‌های روزانه تاریخی همه سهم‌ها و صندوق‌های قابل معامله در بورس آمریکا
داده‌های کارت اعتباری	تراکنش‌های کارت اعتباری	284K	تراکنش‌های کارت اعتباری ناشناس با برچسب تقلبی یا واقعی
داده‌های وبسایت فیشینگ	منبع‌یاب یکنواخت Uniform resource locator(URL)	52K .3	۲۱۱۹ سایت فیشینگ از فیش تانک(PhishTank) و ۱۴۰۷ سایت قانونی از پایگاه داده الکسا.

گستره‌ای را از دیدگاه سازمان داده‌های بزرگ به خود جلب کرده‌اند. این اساساً یک شبکهٔ معنایی است که در آن نودها واحدهای اقتصادی یا مفاهیم را بیان می‌کنند و اج‌ها[8] (یال‌ها) ارتباط معنایی مختلفی را بین واحدهای اقتصادی/ مفاهیم بیان می‌کنند. برای مثال، «ایلان ماسک» کارآفرین رؤیاپرداز است که در آن «ایلان ماسک» یک واحد اقتصادی (موجودیت) است و کارآفرین یک مفهوم است. مطابق منابع این داده‌ها، نمودارهای دانش می‌تواند به دو دسته تقسیم شود: صفحات وب معمول مثل نمودار دانش گوگل و مایکروسافت بینگ و دایرةالمعارف‌های آنلاین ساختاریافته مرتبط از قبیل یاگو[9] و دی‌بی‌پدیا[10]. رویدادی یکی دیگر از مهمترین انواع دانش علاوه بر واحدهای اقتصادی/ مفاهیم و ارتباطات آنان است. یک نمودار دانش متشکل از رویدادها و شرکت‌کنندگانشان است که اغلب نمودار رویداد نامیده می‌شود. نمودارهای رویداد موجود شامل آی سی ای دبلیو اس۵[11]، جی دلت۶[12] و نمودار رویداد مالی مؤسسه تکنولوژی هاربین[13] است.

نمودارهای دانش به ابزاری مهم برای شناسایی ریسک مالی تبدیل شده است. با استفاده از ارتباط بین سهامداری و مؤسسات مالی ال وی و همکاران[14] یک نمودار دانش برای حقوق مالکانهٔ مالی در چین ایجاد کردند که شامل بیش از ۴۵ میلیون واحد اقتصادی(موجودیت)/ مفهوم و ۱۴۵ میلیون ارتباط است. از طریق نمودار حقوق مالکانه، ذینفعان و افراد درست را که در عمل بازی می‌کنند می‌توان برای هدایت توسعهٔ پایدار صنعت مالی نمایش داد.

۲،۳. ماهیت(ویژگی) داده‌ها

بعد از این که داده‌های بزرگ به دقّت ساختاربندی و ارائه شدند، کاوش ارزش آنها به یک کار اصلی تبدیل می‌شود. در سال‌های اخیر، صنعت مالی فراتر از داده‌های سنتی رفته است مانند بورس اوراق بهادار و انتشار اطلاعات مربوط به شرکت‌ها و توجه زیادی که به ثبت‌های فروش شرکت، شبکه‌های اجتماعی، اطلاعات تراکنش‌های کارت اعتباری، اطلاعات موقعیت‌یابی و عکس‌های ماهواره‌ای و سایر موارد داشته است؛ چرا که آنها شامل اطلاعات ارزشمند در بسیاری از موضوعات مثل شناخت ریسک در بازارهای مالی است. با این حال، استفاده از داده‌ها از روش‌های مختلف، چالش‌هایی را در کشف ارزش داده‌ها به ارمغان آورده است. به خاطر الگوریتم‌های داده‌های بزرگ، این دیدگاه‌ها در دسترس محقّقان و سرمایه‌گذاران هستند.

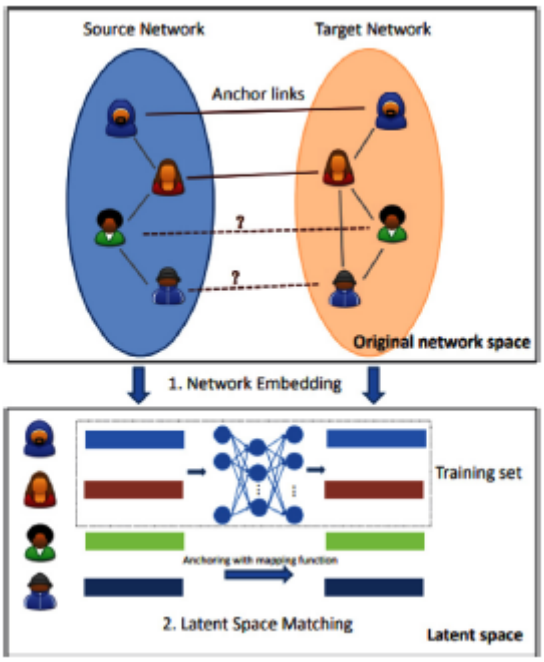
داده‌های متنی، که به‌طور خاص از

رسانه‌های اجتماعی هستند، به‌طور گسترده در پیش‌بینی، کشف و مدیریت ریسک‌های مالی، مورد استفاده قرار می‌گیرند. برای مثال، فرم ۱۰-k نقش مهمی برای شرکت‌های پذیرفته شده در بورس اوراق بهادار در افشاییات متنی بازی می‌کند.

در همین حال، جستجو در اخبار وال استریت ژورنال، حساسیت بازار سهام به شایعات را نشان می‌دهد. عنوان گرایش‌ها و احساسات در رسانه‌های اجتماعی نیز به‌عنوان منابع مهم برای کارهای مرتبط بیشتر، مورد مطالعه قرار گرفته است.

در همین حال، داده‌های تصویری و ویدئویی می‌توانند نقش غالب در کشف تقلب مالی بازی کنند. برای مثال ویدئوهای خروجی بیش از ۱۰،۰۰۰ ساعت رسوایی‌های تقلب لاکین کافه در سال ۲۰۲۰ را نشان می‌دهد و تصاویر ماهواره‌ای برای شناسایی تقلب گروه شانگزیدائو (یکی از شرکت‌های پذیرفته شده در بورس اوراق بهادار چین) از طریق ردیابی قایق‌های ماهیگیری مورد استفاده قرار گرفت. داده‌های صوتی نیز برای جلسات سرمایه‌گذار و درآمد تماس‌های کنفرانسی مهم هستند. آنها نه تنها شامل اطلاعات متنی هستند، بلکه شامل سرنخ‌های احساسی هستند که از آن ویژگی‌ها می‌توان برای پیش‌بینی نوسانات قیمت سهام استفاده کرد. علاوه بر این، داده‌های ساختار یافته از زنجیرهٔ تأمین از قبیل جریان اپلیکیشن بانکی متفاوت با هوش اینترنت اشیا از طریق تحلیل داده‌های کاربران می‌توانند مشتریان را قبل از این که کارت‌های اعتباریشان دزدیده شود، آگاه سازد. با این حال، انواع مختلف داده‌های جایگزین مکملی برای هر کدام از حل وظایف خاص هستند. با توسعه یادگیری عمیق، ترکیب اطلاعات چندوجهی، از طریق یک مدل واحد برای پیش‌بینی ریسک‌های مالی امکان پذیر می‌شود.

شکل ۱: پیل: شناسایی لینک‌ها در سراسر شبکه ها



۲,۴. داده‌های عمومی برای تحقیقات مالی

جدول شماره یک داده‌های عمومی در دسترس استفاده شده در تحقیقات ریسک مالی را به‌صورت خلاصه نشان می‌دهد.

۳. فناوری‌ها و روش‌های نوظهور

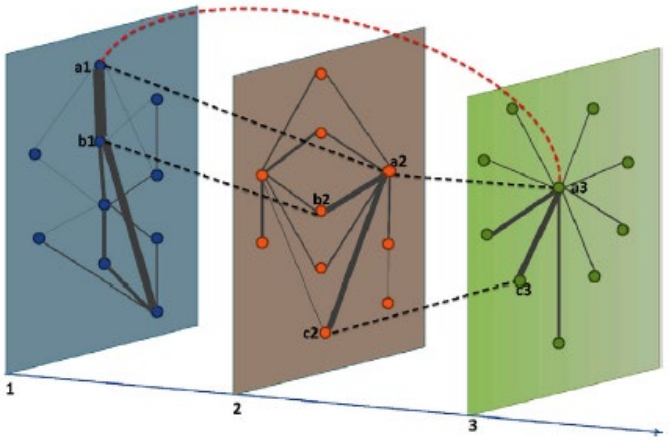
در این بخش فناوری‌ها و روش‌های نوظهوری را که از داده‌های بزرگ یا ناهنجاری‌ها، الگوها و روندها در داده‌های بزرگ استخراج می‌کنند که ابزارهای مؤثری برای یادگیری مدیریت ریسک‌های مالی در عصر داده‌های بزرگ هستند، مرور می‌شود.

۳,۱. یادگیری عمیق و یادگیری ویژگی[۱۵]

استفادهٔ مؤثر از داده‌های مالی چند وجهی، استخراج ویژگی‌های اطلاعاتی از آن است که می‌توان با استفاده از قدرت یادگیری عمیق به آن دست یافت. کلید موفقیت یادگیری عمیق در این زمینه، توانایی آن در یادگیری خودکار ویژگی‌های با کیفیت بالا از داده‌های در مقیاس بزرگ است.

یادگیری ویژگی، فرایند تبدیل داده‌های خام به فرم مناسب‌تری برای یادگیری ماشینی است. می‌توانند الگوهای کلیدی را از داده‌ها استخراج کند و نقش مهمی در برنامه‌های یادگیری ماشینی ایفا کند که در آن عملکرد ممکن است به شدت ویژگی بستگی داشته باشد.

به‌عنوان مثال، به‌عنوان نقطه عطفی در چشم‌انداز کامپیوتر، شبکه عصبی



شکل ۲: شماتیک یک شبکهٔ تجاری چند تکه. سه برش، که هر کدام مربوط به یک شبکهٔ معاملاتی روزانه است، نشان داده شده است. یک معامله‌گر در روزهای مختلف معاملاتی با اتّصالات بین برش‌ها متّصل می‌شود. خطوط یکپارچه نشان دهندهٔ اتّصالات درون‌برشی و خطوط نقطه‌چین نشان دهندهٔ اتّصالات بین‌برشی هستند.

کانولوشن[۱۶]، هسته‌های کانولوشن را به‌طور خودکار از داده‌ها یاد می‌گیرد و از روش‌های استخراج ویژگی دستی معمولی بهتر عمل می‌کند. برای داده‌های نمودار، شبکه‌های عصبی نمودار که ساختار نمودار را در فرایند یادگیری ترکیب می‌کنند، می‌توانند نمایش‌هایی با کیفیت بالا برای وظایف پایین دستی آموزش دهند. با کمک چهارچوب‌های پیش آموزشی، داده‌های چندرسانه‌ای می‌توانند به‌طور مشترک برای یادگیری ویژگی در یک مدل یکپارچه مورد بهره‌برداری قرار گیرند. چنین مدل یکپارچه آموزش داده شده بر روی داده‌های چند وجهی می‌تواند اطلاعات غنی‌تری را در مقایسه با مدل‌های مستقل ارائه دهد.

۳،۲. پروفایل کاربری و مدل‌سازی رفتار

در یک سیستم بزرگ با عوامل متقابل متعدد، علی‌رغم تصادفی بودن و غیر قابل پیش‌بینی بودن رفتار هر فرد، رفتار گروهی از نظم قوی برخوردار است. این امر، چهارچوبی را فراهم می‌کند که می‌تواند برای بررسی رفتارهای گروهی در چنین سیستم‌هایی استفاده شود. علاوه بر این، پروفایل کاربر، که هدف آن درک و تجزیه و تحلیل رفتار کاربر است، می‌تواند برای کشف اطلاعات ارزشمند بسیار مفید باشد.

روش‌های سنتی مدل‌سازی رفتار کاربر بر مدل‌سازی کلاسیک مارکوف برای شبیه‌سازی رفتارهای کاربر تکیه دارند. کارهای اخیر بر ایدهٔ یادگیری ویژگی فشرده (با پیروی از روش رمزگذار- رمزگشا) متمرکز شده است که می‌تواند برای بازسازی توالی‌های خام مورد استفاده قرار گیرد. در شبکه‌های عصبی جریان مجدد[۱۷]، ساختارهای بازگشتی مانند حافظهٔ بلندمدت کوتاه مدت برای مدل‌سازی توالی‌ها استفاده می‌شود. علاوه بر معماری جدید، معماری‌های کانولوشن[۱۹] و توجه نیز معرفی شده‌اند.

شناسایی هویت کاربر. یک کاربر ممکن است چندین حساب در شبکه‌های مختلف داشته باشد. بنابراین، شناسایی کاربران بین پلتفرمی برای پروفایل کاربر و مدیریت ریسک مالی مهم است. دو نوع اصلی از رویکردها برای این امر روش‌های نظارت شده و روش‌های بدون نظارت هستند. در مقایسه با روش‌های قبلی، روش‌های بدون نظارت کمتر مستعد تغییرات ظریف در ساختار شبکه هستند و قوی‌تر هستند برای مثال پیل[۲۰] تعبیه‌های قوی را از ساختارهای شبکه به شیوه‌ای بدون نظارت می‌آموزد و سپس جفت‌های منطبق را با توجه به همسایگان نزدیک پیدا می‌کند. (به شکل ۱ نگاه کنید).

۳،۳. نمودار دانش

همان‌طور که در بخش ۲. ۲ ذکر شد، نمودارهای دانش اطلاعات غنی موجودیت‌ها و روابط بین آنها را نشان می‌دهد و آنها را برای کاربردهای مالی بسیار مفید می‌کند. نمودار دانش یک مدل داده ساختار یافته برای نمایش دانش است و استدلال نمودار دانش(یعنی استنتاج حقایق گمشده در نمودار دانش) یک کار اساسی است. اخیرا مدل‌های مبتنی بر ترجمه به‌طور گسترده مورد بررسی قرار گرفته‌اند. این مدل‌ها اغلب با ترجمه یک موجودیت سر به موجودیت دم از طریق این رابطه، جاسازی‌ها را یاد می‌گیرند.

نمودارهای دانش همچنین می‌توانند برای ارائه دنباله‌ای از رویدادها استفاده شوند. هنگامی که فقط رابطه زمانی بین رویدادها را در نظر می‌گیریم، نمودارهای رویداد اغلب به دنباله‌ای از زیر نمودارهای رویداد سازماندهی می‌شوند. مدل‌های

موجود ابتدا زیر نمودار را در یک مهر زمانی رمز گذاری می‌کنند و سپس اطلاعات ترتیب زمانی را از طریق یک مدل توالی مدل می‌کنند.

۳,۴. فناوری‌های NLP

در بازارهای مالی، داده‌های متنی می‌توانند اطلاعات غنی را برای تجزیه و تحلیل ریسک‌ها فراهم کنند. بنابراین، NLP از جمله استخراج موجودیت، تجزیه و تحلیل احساسات، ترجمهٔ ماشینی و یادگیری ویژگی (مدل زبان) مجموعهٔ بسیار مهمی از فناوری‌ها برای درک و استخراج اطلاعات و دانش مالی است. علاوه بر این، موفقیت فناوری‌های ترجمهٔ ماشینی چند زبانه مبتنی بر شرکت‌های بزرگ، بازار مالی باز را در سراسر کشورها تقریباً یکپارچه کرده است.

مدل زبان از پیش آموزش دیده در سال‌های اخیر، مدل‌های زبان از پیش آموزش دیده در مقیاس بزرگ مبتنی بر معماری تبدیل کننده، پیشرفت‌های برجسته‌ای در وظایف NLP، مانند BERT و GPT-3 ایجاد کرده‌اند. آنها ابزار قوی‌تری برای کسب اطلاعات از متون در تحلیل مالی، مانند تحلیل احساسات متون مالی، ارائه می‌کنند. علاوه بر این، GPT-3 پتانسیل زیادی برای تولید زبان طبیعی (NLG) نشان می‌دهد. GPT-3 که بر روی پیکره‌های با مقیاس بزرگ جمع‌آوری شده از وب آموزش دیده است، قدرت چشمگیری در تولید متون با کیفیت بالا از خود نشان داده است که از متون نوشته شده توسط انسان قابل تشخیص نیستند.

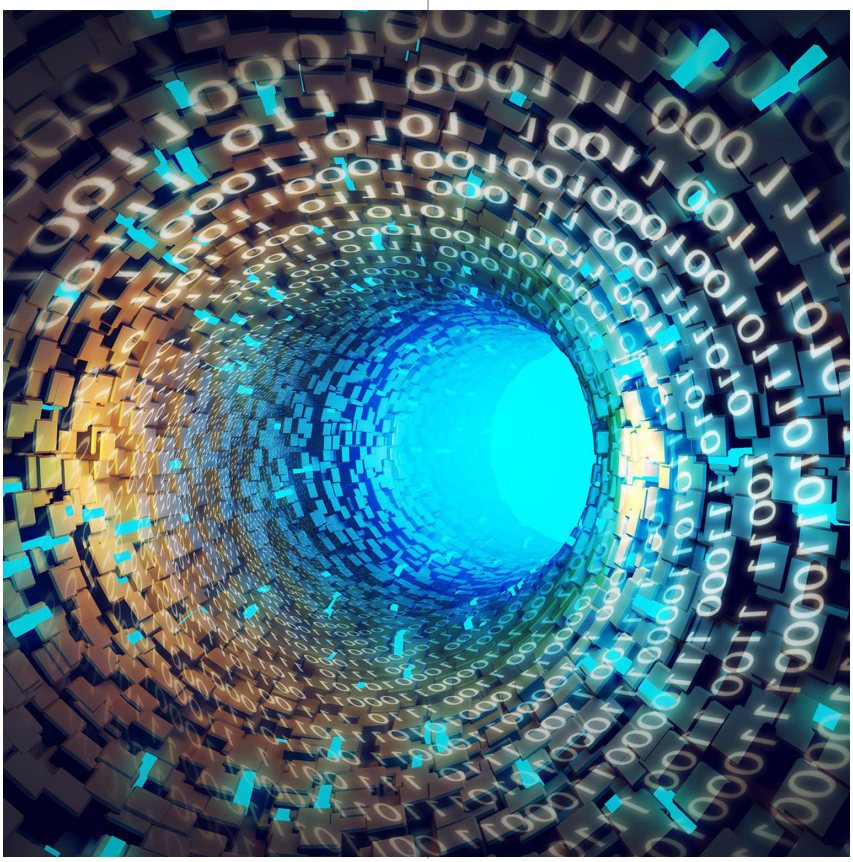
۳,۵. تجزیه و تحلیل نمودار و شبکه

در مالی مدرن، نه تنها تک‌تک واحدهای اقتصادی و ویژگی‌های آنها را مدّنظر قرار می‌دهد بلکه روابط بین آنها نیز مهم است. برخی داده‌ها در حال حاضر در شبکه‌هایی مانند شبکه‌های سرمایه‌گذاری و معاملات وجود دارند، در حالی که سایر داده‌ها به‌طور ضمنی و ناهمگون مرتبط هستند، به‌عنوان مثال، داده‌های مربوط به بنیانگذاران کسب و کار، تجارت، رفتارهای سرمایه‌گذاری و بورس‌ها. در این بررسی، از این موضوع پیروی می‌شود که یک نمودار نمایشی ریاضی از رئوس (واحدهای اقتصادی/ موجودیت‌ها) و لبه‌های آنها (روابط) است و از آن به جای واژهٔ شبکه استفاده می‌شود که به نمونه‌های خاصی از یک گراف اشاره دارد.

۳,۵,۱. تجزیه و تحلیل ساختار

شبکه‌ها در سطح میکروسکوپی، تصادفی و بی نظم هستند و اغلب در سطح ماکروسکوپی منظّم و مرتب به نظر می‌رسند. این ویژگی، محقّقان را به کاوش و تحلیل ساختار میانی شبکه (زیر نمودار) که سطوح خرد و کلان را به‌هم متّصل می‌کند، جذب کرده است. به‌عنوان مثال هلوسکوپ[۲۲] موجودیت‌ها و نمودارهای متقلّبانه را در شبکه‌ها براساس توپولوژی و نمودارهای میله‌ای شناسایی می‌کند. در حال حاضر، تجزیه و تحلیل میان ساختار شبکه عمدتاً بر ساختار جامعه متمرکز است. موتیف‌ها، که به‌عنوان الگوهای فرعی مکرّر کوچک در شبکه‌ها شناخته می‌شوند، برای درک ساختار شبکه‌های بزرگ نیز مهم هستند. شکل ۲ یکپارچگی تمام شبکه‌های معاملاتی روزانه سهام را در یک شبکه تجاری چند تکه پیشنهاد می‌کند.

جوامع و زیر نمودارها، جوامع در تعداد



زیـادی از شـبکه‌های پیچیـده در دنیـای واقعـی وجـود دارنـد. به‌طـور کلـی، سـاختار جامعـه بـه گروه‌هایـی از گره‌هـا اشـاره می‌کنـد کـه «در داخـل به‌طـور محکـم و به‌صـورت پراکنـده از بیـرون بـه هـم متّصـل هسـتند.»

در گذشـته، معیارهـای یکسـانی بـرای جامعـه وجـود نداشـت. متعاقبـاً، نیومـن بخشـی بـودن[۲۳] را براسـاس درک شـهودی از سـاختار جامعـه پیشـنهاد کـرد کـه به‌عنـوان تراکـم بـالای ارتباطـات در یـک جامعـه و ارتباطـات نسـبتاً پراکنـده بیـن جوامـع مختلف مشـخص می‌شـود. از آنجایـی کـه بخشـی بـودن تعریـف رسـمی خوبـی از پارتیشـن‌بندی جامعـه براسـاس تراکـم لبـه ارائـه می‌دهـد، بسـیاری از روش‌هـا بـرای کشـف جامعـه براسـاس بخشـی بـودن بهینه‌سـازی پیشـنهاد شـده اسـت. انعطاف‌پذیـری آن همچنیـن بـه محقّقـان اجـازه می‌دهـد تـا به‌راحتـی آن را به‌گونـه‌ای گسـترش دهنـد کـه بـا انـواع مختلـف شـبکه‌ها، ماننـد نمودارهـای دو بخشـی، مطابقـت داشـته باشـد. نکتـهٔ قابـل توجـه ایـن اسـت کـه نظریـهٔ گـراف (نمـودار) طیفـی کـه نقـش مهمـی در بهینه‌سـازی بخشـی بـودن ایفـا کـرده اسـت، به‌عنـوان ابـزار قدرتمنـدی بـرای وظایـف برنامه‌محـور فراتـر از کشـف جامعـه ظهـور کـرد.

تکامـل جوامـع. تکامـل سـاختارهای جامعـه یکـی دیگـر از موضوعـات مهـم اسـت. تکامـل یـک ویژگـی اساسـی شـبکه‌های دنیـای واقعـی اسـت و تعامـلات بیـن سـاختارهای شـبکه را آشـکار می‌کنـد. پـالا و همـکاران[۲۴] تکامـل جامعـه را براسـاس روش اکتشـاف جامعـهٔ پیشـنهادی آنهـا بـا اسـتفاده از نفـوذ کامـل زیـر نمـودار مطالعـه کـرد. آنهـا بـه ایـن نتیجـه رسـیدند کـه ثبـات جوامـع کوچـک پیش‌نیـازی بـرای تضمیـن وجـود آنهـا اسـت، در حالـی کـه پویایـی جوامـع بـزرگ اسـاس وجـود آنهـا اسـت.

رابطـهٔ بیـن سـاختار جامعـه و پویایـی شـبکه ارتبـاط نزدیکـی بـا تکامـل جوامـع دارد. روش اصلـی تحقیـق، شـامل یادگیـری خصوصیـات دینامیـک شـبکه از طریـق فرایندهـای همگام‌سـازی و انتشـار اسـت. رابطـهٔ بیـن فراینـد انتشـار و سـاختارهای جامعـه مطالعـه شـد و اشـاره شـد کـه حالت‌هـای تعـادل محلـی ناشـی از فراینـد انتشـار می‌توانـد سـاختار اجتماعـی شـبکه‌ها را منعکـس کنـد. بـر ایـن اسـاس، آنهـا هدایـت شـبکه را بـرای توصیـف رابطـه بیـن جامعـه شـبکه و حالت‌هـای تعـادل محلـی پیشـنهاد کردنـد و یـک روش کشـف جامعـه مبتنـی بـر بهینه‌سـازی هدایـت شـبکه، بـا عملکـرد قابـل توجـه در مقایسـه بـا روش‌هـای بهینه‌سـازی بخش‌بنـدی ارائـه کردنـد.

۳،۵،۲. نمودار شبکه‌های عصبی

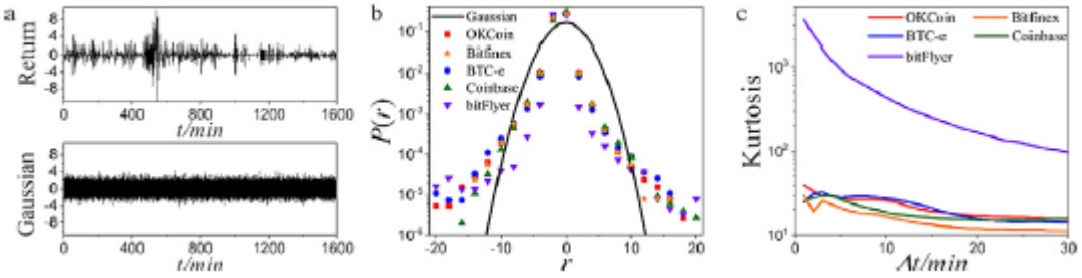
بـا توجـه بـه قابلیـت نمایـش قدرتمنـد نمودارهـا، یادگیـری نمایـش نمـودار، توجـه تحقیقـات گسـترده‌ای را بـه خـود جلـب کـرده اسـت، عـلاوه برایـن، در سـال‌های اخیـر، نمـودار شـبکه‌های عصبـی(GNN) در یادگیـری نمایـش نمـودار و کارهـای بعـدی، ماننـد طبقه‌بنـدی گـره و پیش‌بینـی پیونـد، بـه موفقیـت زیـادی دسـت یافته‌انـد. نمـودار شـبکه‌های عصبـی از یـک طـرح تجمـع همسـایگی پیـروی می‌کننـد، جایـی کـه نمایـش یـک گـره بـا تجمیـع بازگشـتی و تبدیـل نمایـش گره‌هـای همسـایهٔ آن به‌دسـت می‌آیـد. موفقیـت نمـودار شـبکه‌های عصبـی را می‌تـوان بـه قـدرت بیـان بـالای آنهـا در یادگیـری بازنمایـی گره‌هـا و نمودارهـا نسـبت داد. هـدف روش‌هـای موجـود طراحـی شـبکه‌های عصبـی بـرای داده‌هـای گـراف اسـت. GWNN عمل‌گـر پیچیدگـی نمـودار را از طریـق تبدیـل موجـک نمـودار پیاده‌سـازی می‌کنـد.

GNNهـا در تشـخیص ریسـک مالـی بـه پیشـرفته‌های زیـادی دسـت یافته‌انـد. HGN یـک شـبکه ناهمگـن از دسـتگاه‌های حسـاب اسـت و یـک GCN بـرای شناسـایی حسـاب‌های جعلـی ایجـاد می‌کنـد. عـلاوه بـر ایـن کشـف پـول نقـد مـورد بررسـی قـرار گرفـت. نویسـندگان ویژگی‌هـای همسـایه را براسـاس متـا مسـیرها بـرای به‌دسـت آوردن عبـارات جمـع‌آوری کردنـد و توجـه را از هـر متـا مسـیر بـرای به‌دسـت آوردن نمایـش گره‌هـا بـرای طبقه‌بنـدی گره‌هـا آموختنـد. GRC انـواع مختلفـی از روابـط را از طریـق مکانیسـم توجـه بـه خـود مشـخص می‌کنـد و از یـک زمینـهٔ تصادفـی شـرطی بـرای شناسـایی تقلـب در وام اسـتفاده می‌کنـد.

۳،۵،۳. اسـتخراج نمـودار در مقیـاس بـزرگ

بـا رشـد انفجـاری در حجـم داده‌هـا، تعـداد شـبکه‌های دنیـای واقعـی در حـال افزایـش اسـت. ایـن امـر، چالش‌هـای بزرگـی را بـرای اسـتخراج، تجزیـه و تحلیـل و پـردازش نمـودار ایجـاد می‌کنـد. بنابرایـن، چندیـن روش اسـتخراج نمـودار در مقیـاس بـزرگ پیشـنهاد شـده اسـت. بـرای مثـال اسـپک گریـدی[۲۵] به‌طـور مؤثـر انـواع زیـر نمودارهـای متراکـم مشـکوک را در شـبکه‌های بـزرگ براسـاس تئـوری گـراف طیفـی تشـخیص می‌دهـد، کـه بـرای تشـخیص ناهنجاری‌هـای مالـی مفیـد اسـت، ماننـد زیـر نمودارهایـی کـه ناگهـان در یـک شـبکه تراکنـش متراکـم می‌شـوند. در مقابـل، تکنیـک خلاصه‌سـازی نمـودار، کـه نمـودار ورودی را بـه یـک نمـودار

جدول ۲- خلاصه‌ای از کارهای تحقیقاتی انجام شده در کاربردهای مختلف						
مقاله	موضوع	موضوع فرعی	سال	یادگیری ماشینی سنتی	یادگیری عمیق	روش ها
Ha	ریسک اعتباری		۲۰۱۶		*	شبکه عصبی
Moradi	ریسک اعتباری		۲۰۱۹	*		شبکه فازی تطبیقی
Zhang	ریسک اعتباری		۲۰۲۰		*	GBDT، شبکه عصبی
Yang	ریسک اعتباری		۲۰۱۸		*	Time aware LSTM
Tavana	ریسک نقدینگی		۲۰۱۸		*	ANN-BN
Guijarro	ریسک نقدینگی		۲۰۱۹	*		رگرسیون خطی
Sahin Askari	ریسک تطبیق- جرایم مالی	تقلب کارت	۲۰۱۳و۱۷	*		درخت تصمیم‌گیری
Mubalaik	ریسک تطبیق- جرایم مالی	تقلب کارت	۲۰۱۷		*	MLP
Malini	ریسک تطبیق- جرایم مالی	تقلب کارت	۲۰۱۷	*		KNN
Mardnez	ریسک تطبیق- جرایم مالی	دستکاری بازار	۲۰۱۶			یادگیری تقویتی
Shi	ریسک تطبیق- جرایم مالی	دستکاری بازار	۲۰۱۹			استخراج نمودار
Humpherys	ریسک تطبیق- جرایم مالی	تقلب صورت‌های مالی	۲۰۱۱	*		درخت تصمیم‌گیری
Li Sun	ریسک تطبیق- جرایم مالی	مبارزه با پولشویی	۲۰۲۰ و ۲۱			استخراج نمودار
Shi	ریسک تطبیق- جرایم مالی	مبارزه با پولشویی	۲۰۱۹			استخراج متناوب
Nyman	ریسک سیستمیک		۲۰۲۱	*		بخش‌بندی
Zhou	ریسک سیستمیک		۲۰۲۰		*	CNN، BIGRU
Catullo	ریسک سیستمیک		۲۰۱۵	*		یادگیری تقویتی
Yu	ریسک سیستمیک		۲۰۲۰	*		مدل شبکه، GBDT
Ahelegbey	ریسک سیستمیک		۲۰۲۱	*		VAR
Bianchi	ریسک سیستمیک		۲۰۱۹	*		MCMC
O'Halloran	ریسک سیستمیک		۲۰۱۹	*		مدل مارکوف



شکل ۳. دامنه توزیع بازده قیمتی. شکل a بالا نمونه‌ای از بازده قیمت نرمال شده در پلتفرم OKCoin از ساعت ۱۴:۰۴ در ۳ مارچ ۲۰۱۷ تا ساعت ۲۲:۴۵ در ۴ مارچ ۲۰۱۷ است و صفحه پایینی شکل a سیگنال نویز نمونه‌برداری شده از توزیع گاوسی است. شکل b توزیع بازده قیمت نرمال شده در پلتفرم‌های مختلف بیت‌کوین در مقایسه با توزیع گاوسی زمانی که Δt=۲ دقیقه است. شکل c کشیدگی k در مقایسه با Δt در مقایسه با پلتفرم‌های بیت‌کوین است.

خلاصه بسیار کوچک‌تر کاهش می‌دهد، می‌توانند به مدیریت نمودارهای مقیاس بزرگ کمک کند. برای مثال SWeG یک نمودار خلاصه فشرده با اصلاحات لبه و جستارهای نمودار تقویت شده را پیدا کرد. DPGS مدل پیکربندی را به‌عنوان یک مدل پوچ در خلاصه‌سازی نمودار اتّخاذ می‌کند و هم در حافظه و هم در زمان برای آموزش GNN صرفه‌جویی می‌کند که وظایف استخراج نمودار را در شبکه‌های مالی در مقیاس بزرگ تسهیل می‌کند.

۴. کاربردها

در این قسمت مطالعاتی بررسی شده است که ریسک‌های مالی خاص را با استفاده از داده‌های بزرگ و روش‌های نوظهور حل می‌کنند. کارهای انجام شده در این حوزه در جدول ۲ خلاصه شده است.

۴,۱. ریسک اعتباری

ارزیابی ریسک اعتباری اساس و کلید مدیریت ریسک است. در حالی که اطلاعات اعتباری مالی سنتی اغلب فقط دارای ویژگی‌های مالی قوی است (مانند کارت اعتباری، ارز خارجی، وام‌های خصوصی و سایر داده‌های تراکنش مالی)، ارزیابی اعتبار براساس الگوریتم‌های داده‌های بزرگ به‌طور جامع داده‌های مالی، داده‌های خدمات عمومی دولت را در نظر می‌گیرد. ها و همکاران[۲۶] یک مدل امتیازدهی اعتباری براساس یادگیری عمیق و انتخاب ویژگی ایجاد کرد و رتبه‌بندی اعتباری کاربران را با توجه به ویژگی‌های ورودی کاربران ارزیابی کرد. این آزمون نشان داد که یادگیری عمیق روشی مؤثر برای مدیریت داده‌های ویژگی اعتباری با ابعاد بالا است.

مرادی و رفیعی یک سیستم استنتاج فازی مبتنی بر شبکه تطبیقی را پیشنهاد کردند که هم داده‌های مشخصات مشتری و هم عوامل سیاسی اقتصادی نوسانی را برای ارزیابی ریسک اعتباری در خود جای می‌دهد. سیستم پویا می‌توانست جدولی از مشتریان بد را به صورت ماهانه تولید کند و نتایج با موقعیت‌های واقعی بسیار سازگار بود.

ژانگ و همکاران[۲۷] یک مدل امتیازدهی اعتباری یکپارچه آنلاین را طراحی کرد که می‌تواند در زمان براساس یک چهارچوب یادگیری ابتکاری به نام DeepGBM به‌روز شود که مزایای درخت تصمیم تقویت‌کننده گرادیان (GBDT) و شبکه عصبی را برای مدیریت هر دو مقوله‌ای و طبقه‌ای عددی به‌کار می‌گیرد.

به‌عنوان نمونه‌ای از داده‌های غیرسنتی، یک مدل DeepCredit عمدتاً براساس LSTM آگاه از زمان، با استفاده از مجموعه داده‌های جریان، جریان کلیک کاربر کامل از یک پلتفرم وام‌دهی P2P که بیش از ۴ میلیون فعّالیت مالی ۱۰ هزار کاربر را ثبت کرده بود، ساخت. این مدل هم انحرافات فردی و هم پیش‌فرض‌ها را با دقّت بالا پیش‌بینی کرد.

۴,۲. ریسک نقدینگی

تأمین نقدینگی کلید تمام تئوری‌های واسطه‌گری مالی است. در مدل‌های سنتی، ریسک نقدینگی با در نظر گرفتن تغییرات در برخی شاخص‌ها در یک افق زمانی خاص برای مقایسه اندازه‌گیری می‌شود. با این حال، کثرت، تعدد و تنوع حساب‌ها، محاسبهٔ جریان‌های نقدی را بسیار دشوار و زمان‌بر می‌کند و بنابراین به‌دست آوردن این داده‌ها در مدت زمان کوتاه چالش برانگیز است.

امروزه می‌توان از الگوریتم‌های داده‌های بزرگ برای ارزیابی بهتر ریسک نقدینگی و تحلیل عوامل کلیدی و ارتباطات متقابل

آنها استفاده کرد. تاوانا و همکاران[28] از شبکه‌های عصبی برای تحلیل و ارزیابی ریسک نقدینگی و عوامل کلیدی استفاده کرد. مدل دو مرحله‌ای پیشنهادی ANN-BN اساساً با هدف تسهیل تجزیه و تحلیل سیستماتیک اقدامات خاص بانک براساس نسبت‌های ترازنامه انجام شد و ثابت شد که به اندازهٔ کافی انعطاف‌پذیر است تا در هر سناریو مبتنی بر وام اعمال شود.

گیجارو و همکاران[29] تأثیر رسانه‌های اجتماعی بر نقدینگی بازار مالی را با انجام یک تحلیل رگرسیون با شاخص اس‌اندپی ۳۰۵۰۰[30] پس از امتیازدهی آن براساس حساسیت‌های محتوای توییتر، تحلیل کرد. نتایج نشان داد که پس از افزودن حجم متوسط متحرک دو روزه، تمایل سرمایه‌گذاران به‌طور معناداری و مثبت با انتشار مؤثر نقدینگی همبستگی دارد.

۴،۳. ریسک نوسانات

ریسک نوسان در بازارهای مالی، احتمال نوسانات قیمت پرتفوی تحت تأثیر تغییرات برخی عوامل ریسک است. در حال حاضر، یکی از معیارهای اصلی برای تعیین کمیّت ریسک مالی، ارزش در معرض ریسک[31] است که در آن پیش‌بینی نوسانات بازار نقش مهمی ایفا می‌کند. از آنجایی که بازده قیمت، کلید درک و مدل‌سازی نوسانات بازار است، تحقیقات در مورد ریسک نوسان عمدتاً بر بازده قیمت متمرکز است؛ جایی که داده‌های متوالی رایج و گسترده هستند.

در مطالعات قبلی، محقّقان دریافته‌اند که توزیع بازده قیمت، یک دم پهن[32] را نشان می‌دهد (شکل ۳).

بازده قیمت دارای همبستگی خودکار خطی نیست، در حالی که قدر مطلق بازده قیمت، یک حافظهٔ وسیع را نشان می‌دهد. در عصر داده‌های بزرگ، روش‌های تحلیل سنتی برای داده‌های متوالی به‌طور کلی به سه دسته تقسیم می‌شوند: شبیه‌سازی تاریخی، روش تحلیلی و شبیه‌سازی مونت کارلو. با این حال، این روش‌ها اغلب به نتایج دقیقی دست نمی‌یابند؛ برعکس، الگوریتم‌های داده‌های بزرگ می‌توانند بر این مشکل غلبه کنند. فیشر و کراوس[33] با استفاده از شبکه LSTM به پیش‌بینی بهتری از حرکات جهتی سهام اس‌اندپی[34] ۵۰۰، در یک بازهٔ زمانی طولانی در مقایسه با رگرسیون لجستیک، جنگل تصادفی و حتی شبکه‌های عصبی عمیق استاندارد دست یافتند. آنها همچنین نشان دادند که نوسان پرتفوی هدایت شده توسط شبکه‌های LSTM به‌دلیل عوامل خطر مشترک به‌طور قابل توجهی کاهش یافته است.

۴،۴. دستکاری بازار

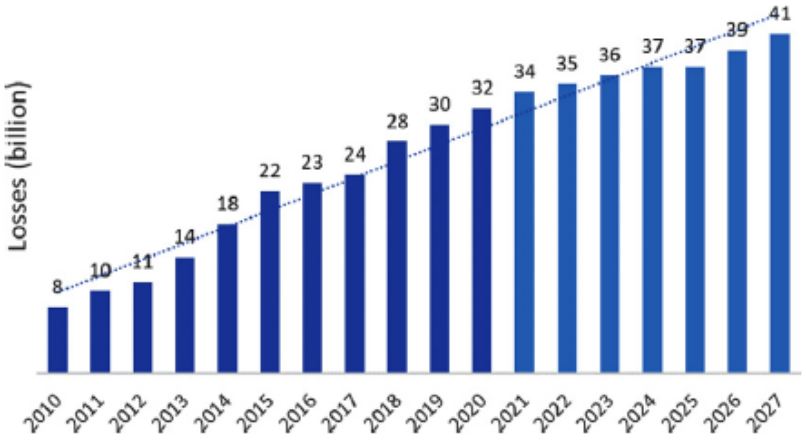
بازارهای مالی، بستری را برای شرکت‌ها فراهم می‌کنند تا با اجازه دادن به سرمایه‌گذاران برای معامله سهام، سرمایه خود را افزایش دهند و در شرایط عادی، قیمت محصولات مالی، منعکس کنندهٔ قضاوت‌های رایج در مورد ارزش شرکت‌ها است. با این حال، قیمت‌های بازار را می‌توان با اطلاعات نادرست و معاملات متقلّبانه دستکاری کرد و بر اطلاعات سرمایه‌گذاران تأثیر گذاشت و بازار را به ضرر توسعهٔ بازار مالی مختل کرد.

مارتینز و همکاران[35] از یک چهارچوب یادگیری تقویتی در قابلیت مشاهدهٔ کامل و جزئی فرایندهای تصمیم مارکوف استفاده کرد و با یافتن دلایلی که این معامله‌گران را تشویق به انجام فعّالیت‌های متقلّبانه می‌کند، رفتار اساسی مجرمان را تحلیل کرد. علاوه بر روش‌های یادگیری ماشینی، برخی از روش‌های تحلیل شبکه برای تشخیص دستکاری بازار استفاده شده‌اند. برخی مطالعات تلاش کرده‌اند تا با تجزیه و تحلیل شبکه‌های تجاری، دستکاری مبتنی بر معامله را در بازار سهام چین شناسایی کنند و سهام دستکاری شده به‌طور مؤثر با رویکرد همبستگی درجه- قدرت متمایز شدند. علاوه بر این، با ایجاد یک شبکه معاملاتی چند بخشی، آنها با موفقیت معامله‌گران غیرعادی را در بازار سهام شناسایی کردند. علاوه بر این، آنها دسته شبکه تجاری را برای تشخیص تبانی در دستکاری سهام تجزیه و تحلیل کردند.

۴،۵. ریسک مالی سیستمی

انواع مختلفی از ریسک‌هایی که قبلاً مورد بحث قرار گرفت، می‌توانند باعث سقوط در یک صنعت یا اقتصاد شود. بنابراین، مدیریت ریسک مناسب برای درمان ریسک سیستمیک، به‌ویژه از منظر نظارتی مورد نیاز است. مدیریت ریسک سیستمی سنتی عمدتاً بر شاخص‌های خرد سازمانی و شاخص‌های کلان اقتصادی، مانند کفایت دارایی، کیفیت دارایی و نقدینگی تکیه می‌کند که پایه و اساس روش‌های نظارت بر ریسک، از جمله میانگین موزون یا مدل‌های ساختاری را ایجاد می‌کند. با این حال، محدود به مکانیسم به‌روزرسانی یا ویژگی ذاتی، شاخص‌های سنتی اغلب از پسماند، فرکانس پایین و سایر معایب هستند و ترکیب داده‌های جایگزین می‌تواند مشکل را کاهش دهد.

نایمن و همکاران[36] داده‌های مبتنی بر متن



شکل ۴. زیان جهانی ناشی از تقلب پرداختی از ۹. ۸۴ میلیارد دلار در سال ۲۰۱۱ به ۳۲. ۳۹ میلیارد دلار در سال ۲۰۲۰ رسید که سه برابر شده است. پیش‌بینی می‌شود تقلب در پرداخت‌ها همچنان در حال افزایش باشد و پیش‌بینی می‌شود در سال ۲۰۲۷ هزینه‌ای معادل ۴۰. ۶۲ میلیارد دلار داشته باشد. داده‌های پس از سال ۲۰۲۰ مقادیر پیش‌بینی شده است.

بازار مالی را برای ارزیابی تأثیر محتوای احساسی بر توسعهٔ یک سیستم مالی تجزیه و تحلیل کرد و شکل‌گیری شور و نشاط را قبل از بحران مالی جهانی و همچنین فروپاشی متعاقب آن آشکار کرد. با توسعهٔ الگوریتم‌های داده‌های بزرگ با قدرت محاسباتی بالا، یک سری روش‌های جدید مبتنی بر داده پدیدار شده‌اند.

روش‌های غیرخطی مانند SVM و شبکه‌های عصبی دقّت پیش‌بینی بالاتری را برای ریسک سیستمی به عنوان یک مسألهٔ غیرخطی ارائه می‌کنند. ژو و همکاران[۳۷] یک مدل یادگیری عمیق ترکیبی مبتنی بر CNN و BiGRU برای پیش‌بینی ریسک مالی سیستمی، دستیابی به عملکرد برتر در یادگیری ویژگی پیشنهاد کرد. پس از بحران مالی ۲۰۰۸، تعداد زیادی مدل پویا مبتنی بر تئوری‌های اقتصادی و مالی در حال ظهور، سیر تحوّل ریسک سیستمی را توصیف کردند. کاتولو و همکاران[۳۸] پویایی اقتصاد کلان حاصل از تعامل بین بانک‌ها و شرکت‌های ناهمگون در شبکه اعتباری درون زا را باز تولید کرد، که براساس آن نویسندگان شاخص‌های هشدار اولیه بحران‌ها را تعریف کردند. یو و همکاران[۳۹] یک شبکه بین بانکی را با استفاده از یک رویکرد شبکه پیچیده برای پیش‌بینی سرایت ریسک سیستمی شبیه‌سازی کرد و روشی را برای انتخاب سیاست مدیریت بهینه برای مدیریت ریسک سیستمی ارائه کرد. علاوه بر این، محقّقان در سایر شاخه‌های اصلی روش‌های یادگیری آماری مبتنی بر داده را برای ساخت مدل‌های تکامل ریسک اتخاذ کرده‌اند. شاخص‌های ریسک سیستمی و متغیرهای کلان اقتصادی به‌عنوان متغیرهای حالت استفاده می‌شوند و رگرسیون خودکار برداری[۴۰] و زنجیری مارکوف به‌عنوان شکل کلی‌تری از رگرسیون خودکار برداری برای به دست آوردن یک ماتریس انتقال وضعیت براساس داده‌های تاریخی استفاده می‌شوند.

اهلگبی و همکاران[۴۱] یک مدل رگرسیون خودکار برداری شبکه را برای ارزیابی تأثیر مالی همه گیری کووید ۱۹ پیشنهاد کردند. اخیرا، بسیاری از رویکردهای هوش مصنوعی در تحقیقات تکامل ریسک سیستمی به کار گرفته شده است.

اوهالوران و نواسیک[۴۲] یک موتور ریسک سیستمی براساس موتور ریسک منبع باز طراحی کردند و معیارهای ریسک مختلفی را در خود جای دادند. آنها همچنین یک فناوری شبیه‌سازی را برای ارزیابی تأثیر مقررات بر سیستم مالی به‌طور کلی اقتباس کردند.

۵. ریسک تطبیق- جرایم مالی

ریسک تطبیق- جرایم مالی به خساراتی مربوط می‌شودکه ممکن است زمانی ایجاد شود که یک شرکت یا مؤسسه از قوانین و مقررات مربوط به جرایم مالی در حوزه قضایی مربوطه خود پیروی نکند. جرایم مالی معمولی شامل سرقت هویت، دستکاری بازار، پولشویی و تقلب در صورت‌های مالی است. در این بخش، رویکردهای نوظهور داده‌های بزرگ برای مدیریت ریسک تطبیق- جرایم مالی معرفی شده است.

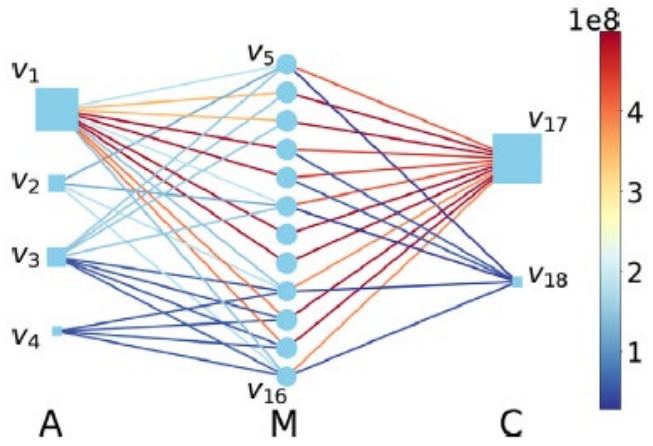
۵,۱. شناسایی دزد

سرقت هویت به مجرمانی اطلاعات می‌شود که از هویت شخص دیگری و سایر اطلاعات مرتبط به روش‌های غیر مجاز استفاده می‌کنند. اغلب منجر به زیان مالی فوری می‌شود و قربانیان ممکن است از یک سری اعتبار و همچنین نیز مشکلات دیگر رنج ببرند. برای کاهش ریسک، وانگ و همکاران[۴۳] از یک مولد احتمالی برای تشخیص سرقت

هویت در شبکه‌های اجتماعی تلفن همراه استفاده کرد. چنین جرایمی که از طریق اینترنت انجام می‌شود " فیشینگ" نامیده می‌شود. رائو و همکاران[۴۴] وبسایت‌های فیشینگ را با استفاده از طبقه‌بندی جنگل تصادفی براساس ویژگی‌های اکتشافی استخراج شده از URL، کدهای منبع و خدمات شخص ثالث شناسایی کرد. در پژوهش بناویدس و همکاران[۴۵] شرح مفصلی از رویکردهای یادگیری عمیق برای مقابله با حملات فیشینگ ارائه کرد. یکی از شکل‌های سرقت هویت، کلاهبرداری از کارت اعتباری است که معمولاً تنها بر یک یا چند حساب کارت اعتباری باز قربانی تأثیر می‌گذارد. همان‌طور که در شکل ۴ نشان داده شده است، تراکنش‌های جعلی از طریق کارتهای اعتباری می‌تواند تا سال ۲۰۲۷ به ۴۱ میلیارد دلار برسد. رویکردهای بسیاری برای چنین فوریت‌های ضد تقلب طراحی شده است. مالینی و پوشپا[۴۶] الگوریتم KNN و روش‌های تشخیص پرت را همراه با روش‌های نمونه‌برداری بیش از حد و PCA، برای بهینه‌سازی نرخ کشف تقلب برای شناسایی تقلب در کارت اعتباری بانکی اتخاذ کردند. عسکری و حسین[۴۷] یک الگوریتم FzzITree را براساس درخت تصمیم ID3 با استفاده از منطق فازی برای کشف تراکنش‌های جعلی پیشنهاد کردند.

۵،۲. تقلب در صورت‌های مالی و تقلب بیمه‌ای

تقلب در صورت‌های مالی تقلب در صورت‌های مالی به ارائه اشتباه عمدی از وضعیت مالی یک واحد تجاری اشاره دارد که با تحریف عمدی یا حذف مبالغ یا افشا در صورت‌های مالی انجام می‌شود.



شکل ۵. نمونه‌ای از نقل و انتقالات پولشویی در یک بانک، ایجاد یک زیرگراف سه جانبهٔ متراکم. رنگ لبه و اندازه گره نشان دهنده مقدار پول منتقل شده است.

زیان‌های ناشی از تقلب در صورت‌های مالی اغلب بسیار بزرگ و گاهی اوقات حتی برای مؤسسات فاجعه بار است. بنابراین، تقلب در صورت‌های مالی یک نگرانی جدی برای حسابرسان، سرمایه‌گذاران و قانونگذاران بوده است. الگوریتم‌های مختلف داده‌های بزرگ برای مقابله با این خطر توسعه یافته‌اند. یک گردش کار رایج شامل انتخاب ویژگی از داده‌های متنوع (شامل متن گزارش و شاخص‌های مالی) و طبقه‌بندی برای پیش‌بینی تقلب است. پس از به‌کارگیری طیف وسیعی از الگوریتم‌های داده‌های بزرگ برای انتخاب ویژگی و طبقه‌بندی برای شناسایی تقلب، هاجک و هنریکس[۴۸] دریافتند که روش‌های مجموعه برای شرکت‌های متقلّب و شبکه‌های باور بیزی (BBN) برای شرکت‌های غیر متقلّب بهترین عملکرد را دارند. با ملاحظات مشابه، یائو و همکاران[۴۹] چندین روش ترکیبی را مقایسه کرد و بهترین ترکیب XGBoost و جنگل تصادفی را برای تشخیص تقلب در بیانیه یافت. با کمک یادگیری عمیق، کراجا و همکاران[۵۰] با اتخاذ یک شبکه توجه سلسله مراتبی (HAN) برای استخراج ویژگی برای بازتاب بهتر ساختارهای سند و تمرکز بر محتوا و زمینه متن، نتایج بهتری را در مقایسه با بسیاری از رویکردهای فعلی به دست آورد. از تکنیک‌های داده کاوی متعدد برای شناسایی شرکت‌هایی که به تقلب در صورت‌های مالی متوسل می‌شوند، استفاده می‌شود. علاوه بر این، چندین روش برای پیش‌بینی تقلب در صورت‌های مالی ارزیابی شد.

تقلب بیمه‌ای تقلب بیمه‌ای یک فریب عمدی است که علیه یا توسط یک شرکت بیمه یا نماینده برای منافع مالی انجام می‌شود. طبق برآورد FBI تقلب در بیمه‌های غیر زندگی، بیش از ۴۰ میلیارد دلار در سال هزینه دارد و بار حق بیمه بیشتری را بر خانواده ایالات متحده وارد می‌کند. با شیوع همه گیری کووید ۱۹، تعداد ادعاهای تقلب احتمالی دو برابر شده است. با این حال، روش‌های سنتی معمولاً

در رسیدگی به اطلاعات متنی در دعاوی شکست می‌خورند، که میتواندمرجع ارزشمندی برای کشف تقلب بیمه باشد. وانگ و زو[51] با استفاده از یادگیری عمیق، ویژگی‌های عددی سنتی را با ویژگی‌های متن استخراج شده توسط LDA برای آموزش شبکه‌های عصبی برای تشخیص تقلب بیمه خودرو ترکیب کردند و نتایج بهتری از مدل پر کاربرد مانند SVM و جنگل تصادفی بود.

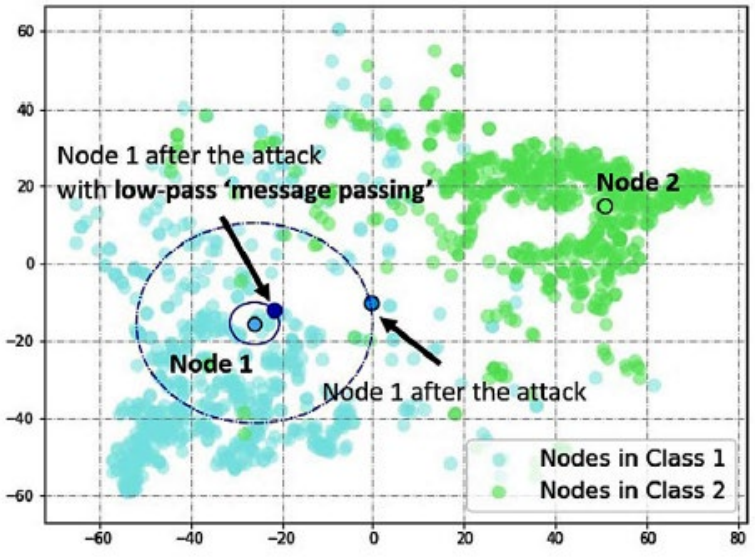
۵،۳. مبارزه با پولشویی

پولشویی رفتار پنهان کردن منبع پولی است که از طریق فعّالیت‌های نامشروع به دست آمده است. براساس گزارش دفتر مبارزه با مواد مخدر و جرایم سازمان ملل، میزان تخمین زده شده پولشویی در سطح جهان در یک سال ۲ تا ۵ درصد تولید ناخالص داخلی جهانی یا ۸۰۰ میلیارد دلار - ۲ تریلیون دلار به دلار آمریکا است. بنابراین، مبارزه با پولشویی (AML) برای ثبات مالی ملی اهمیت حیاتی دارد.

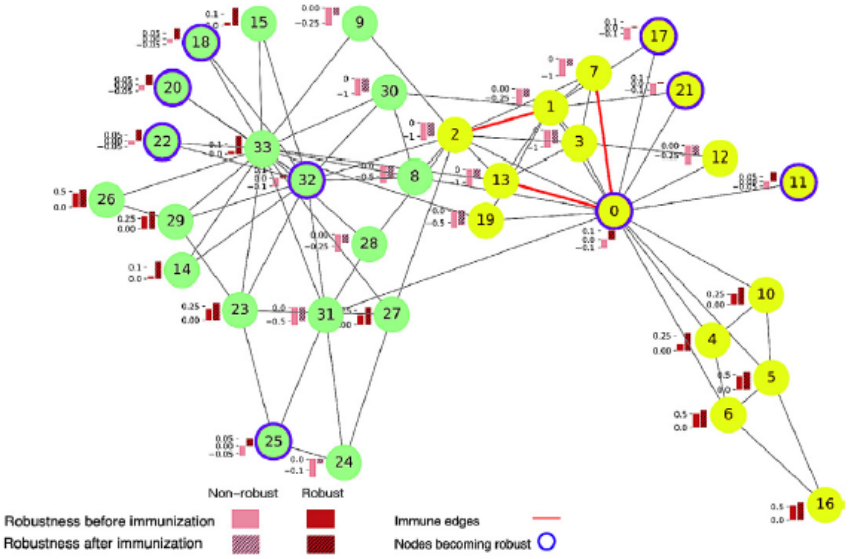
کلاسیک‌ترین رویکرد مورد استفاده در بانک برای AML طبقه‌بندی مبتنی بر قانون است. با استفاده از سفارشات خرید و فروش، بازده قیمت می‌تواند روشی نوآورانه و مؤثر برای تشخیص ناهنجاری در پلتفرم‌های بیت‌کوین ارائه کند. با این حال، این الگوریتم‌های مبتنی بر قواعد به شدت به دانش تخصصی متکی هستند، متقلّبان به راحتی از آنها فرار می‌کنند و نمی‌توانند برای کشف انواع جدیدی از رفتارهای پولشویی استفاده شوند. الگوریتم‌های یادگیری ماشینی نیز برای شناسایی فعّالیت‌های پولشویی استفاده می‌شوند که می‌توانند به‌طور مؤثر در سناریوهای جدید بدون محدودیت‌های قوانین ثابت درون یابی شوند. SVM قبلاً برای پردازش داده‌های بزرگ استفاده می‌شد و دقّت بالایی داشت. یک روش مبتنی بر یادگیری عمیق را پیشنهاد کرد که برای وظایف مبارزه با پولشویی با استفاده از روابط مشتری با مشتری آموزش داده شده بود. با این حال، این الگوریتم‌ها فعّالیت‌های پولشویی را به شیوه‌های نظارت شده یا نیمه‌نظارت شده شناسایی می‌کنند که از طبقات نامتعادل و عدم سازگاری رنج می‌برند. علاوه بر این، روش‌های مبتنی بر خوشه‌بندی برای شناسایی فعّالیت‌های پولشویی با گروه‌بندی معاملات مشکوک به خوشه‌ها به کار گرفته شده‌اند. با این حال، این روش‌ها تعامل بین حساب‌ها را در نظر نمی‌گیرند، که منجر به نرخ بالای تشخیص مثبت کاذب می‌شود. نمودارها و شبکه‌ها نمایش‌های قدرتمندی برای تجزیه و تحلیل وابستگی درونی در میان حساب‌های مشکوک درگیر در پولشویی هستند.

شکل ۵ یک مثال واقعی از انتقال پولشویی در یک بانک را نشان می‌دهد که شامل یک جریان دو مرحله‌ای از حساب‌های مبدأ به حساب‌های میانی به مقصد است، که در آن سه نوع حساب به ترتیب با A، M، C نشان داده می‌شوند.

کولادون و ریموندی[52] از تحلیل شبکه‌های اجتماعی برای آشکار کردن نقش‌های اساسی و ساختارهای سازمانی استفاده کردند. با این حال، این روش‌ها ردیابی جریان را انجام نمی‌دهند یا ضمانت‌های نظری ارائه نمی‌دهند. یک رویکرد مقیاس‌پذیر مبتنی بر جریان، فلوسکوپ[53] برای تشخیص جریان انتقال کامل پول از مبدأ به مقصد، که دارای محدودیت نظری بود، پیشنهاد شد. این تراکنش‌ها را با استفاده از یک نمودار چند بخشی مدل کرد و یک متریک ناهنجاری جدید برای بهینه‌سازی طراحی کرد. آزمایشات گسترده ارائه شده در این



a) طبقه بندی گره با حملات خصمانه



b) ایمن‌سازی دشمن در شبکه باشگاه‌های کاراته[۵۷]

مقاله نشان داد که فلوسکوپ در تشخیص رفتارهای مختلف پولشویی خصمانه مؤثر و قوی است. بر این اساس، سان و همکاران[۵۴] جریان انتقال را به‌عنوان دو تانسور[۵۵] جفت شده، با در نظر گرفتن ویژگی‌های متعدد تراکنش‌ها، مدل کرد.

۵،۴. قاچاق و تحریم

قاچاق سیاست‌های مبارزه با قاچاق اغلب نقش مهمی در سیاست‌گذاری ایفا می‌کنند. در سال‌های اخیر، فناوری‌های مختلفی برای مبارزه با قاچاق در نظر گرفته شده است. داده‌هایی مانند تراکنش‌های مالی، تماس‌های تلفن همراه و پیام‌های متنی به‌طور قابل‌توجهی به کار مبارزه با قاچاق انسان کمک می‌کنند. موستو امکان استفاده از داده‌های بزرگ را برای مؤثرتر کردن مبارزه با قاچاق بررسی کرد.

تحریم‌ها قهقرایی جهانی شدن و تشدید مبارزه ژئوپلیتیکی منجر به اعمال تحریم‌های اقتصادی گاه به گاه بر کشورها می‌شود. به‌عنوان مثال، تحریم‌های ایران و روسیه باعث شده است که مردم شروع به ارزیابی ضررهای بعدی با استفاده از الگوریتم‌ها کنند. ترگاب و همکاران از الگوریتم OLS برای تخمین زیان اقتصادی ناشی از تحریم‌های اعمال شده توسط اتحادیه اروپا علیه فدراسیون روسیه و ضدتحریم‌های روسیه استفاده کرد. علاوه بر این، تحریم‌ها می‌توانند در پیش‌بینی ورشکستگی نقش داشته باشند.

۵،۵. مبارزه با رشوه و فساد (ABC)

فساد سوء استفاده از قدرت عمومی با انگیزهٔ منافع شخصی است، عملی که درآمدهای مالیاتی را از توسعهٔ ملی منحرف می‌کند. نه تنها اعتماد به دولت و نهادها را از بین می‌برد، بلکه اثربخشی و انصاف سیاست‌های عمومی را نیز کاهش می‌دهد. براساس گزارش صندوق بین‌المللی پول (IMF)، سالانه حدود یک تریلیون دلار از درآمدهای دولت جهانی به دلیل فساد از بین می‌رود.

پیشرفت در فناوری اطلاعات امکان استفاده از روش‌های یادگیری ماشینی را برای پیشگیری و کشف فساد فراهم کرده است. در برخی از کشورها، تکنیک‌های مربوطه قبلاً برای مبارزه با فساد دولتی

به‌کار گرفته شده بود. برای مثال، در بریتانیا، اگزیگر و شفافیت بین‌الملل (TI) تلاش کردند تا با تجزیه و تحلیل سوابق عمومی، خطرات فساد را شناسایی کنند. IBM همچنین با دولت کنیا برای شناسایی عوامل کلیدی رشوه‌خواری از طریق الگوریتم‌ها کار کرده است. در سال ۲۰۲۰، مایکروسافت راه‌اندازی فناوری و راه‌حل‌های ضد فساد خود (ACTS) را اعلام کرد که از فناوری‌هایی مانند رایانش ابری، هوش مصنوعی و یادگیری ماشینی برای شناسایی و جلوگیری از فساد در دهه‌های آینده استفاده می‌کند.

۶. رهنمودهای آتی

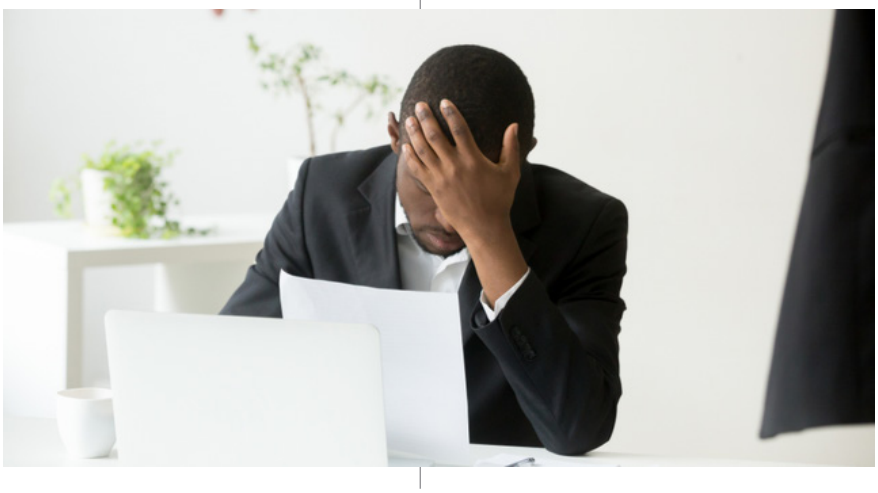
۶,۱. تکنیک‌ها و تشخیص دیپ‌فیک[۵۶]

یادگیری عمیق در زمینه NLP و چشم‌انداز کامپیوتری به موفقیت‌های زیادی دست یافته است. با این حال، از قدرت GPT-3 برای تولید متن با کیفیت بالا می‌توان برای تولید اخبار یا نظرات جعلی استفاده کرد. تکنیک دیپ‌فیک مبتنی بر یادگیری عمیق برای ایجاد تصاویر جعلی و ترکیب ویدیوها یا گفتارهای جعلی استفاده شده است. این تکنیک‌ها که بر روی تصاویر و مجموعه داده‌های ویدیویی بزرگ آموزش داده شده‌اند، می‌توانند اخبار جعلی و شایعات سیاسی را با دستکاری یا جایگزینی اطلاعات چهره ویدیوهای اصلی و ترکیب سخنرانی‌های جعلی تولید کنند. کلاهبرداری مالی با استفاده از تکنیک دیپ‌فیک با موفقیت انجام شده است. در همین حال، افکار عمومی را می‌توان با اخبار و ویدیوهای جعلی مبادله کرد. هنگامی که به اندازهٔ کافی به آن پاسخ داده نشود، می‌تواند بر شهرت شرکت‌ها تأثیر بگذارد، بر رفتار مصرف کنندگان تأثیر بگذارد و حتی بر قیمت سهام تأثیر بگذارد و در نهایت کل بازار مالی را به خطر بیندازد. با این حال، شناسایی و شناسایی تقلبی یک مشکل باز باقی مانده است.

۶,۲. حملات خصمانه

محقّقان دریافته‌اند که شبکه‌های عصبی عمیق را می‌توان به‌راحتی توسط نمونه‌های به اصطلاح متخاصم فریب داد که با افزودن اغتشاشات نامحسوس برای انسان به نمونه‌های اصلی به‌دست می‌آیند. چنین تکنیک‌های «حمله خصمانه» از شناسایی فرار می‌کنند و تهدیدات بالقوه‌ای را برای متقاضیان، به‌ویژه در زمینهٔ مالی ایجاد می‌کنند که می‌تواند منجر به عواقب بسیار حیاتی شود. به‌عنوان مثال، در سیستم‌های امتیازدهی، کلاهبرداران می‌توانند برای فرار از مدل‌های تشخیص تقلب، یک رابطهٔ دوستی را با دیگران جعل کنند. در سال‌های اخیر، محقّقان بر طراحی روش‌های دفاعی در برابر حملات متخاصم متمرکز شده‌اند. یک شبکه عصبی قابل اثبات قوی برای طبقه‌بندی گره از طریق ارسال پیام کم‌گذر پیشنهاد شد. از نظر تئوری، تحت حملات متخاصم، با یک ماژول با پلاگین آسان برای GNNها، که به اندازهٔ بودجهٔ حملهٔ خطی قوی و به اندازهٔ شبکه‌های عصبی دقیق است، دارای مرز بالایی است. شکل ۶a نشان می‌دهد که با مکانیسم انتقال پیام پایین‌گذر، جاسازی گره ۱ فقط یک جابجایی جزئی دارد، در حالی که جاسازی از خط پایه دیگر به‌طور قابل توجهی تغییر می‌کند. دفاع تأیید شده بدترین عملکرد را برای یک حمله مشخص ارائه می‌کند و ایمن‌سازی متخاصم برای بهبود استحکام قابل تأیید در برابر هر حملهٔ خصمانهٔ قابل قبول، همان‌طور که در شکل ۶b نشان داده شده است، پیشنهاد می‌شود.

برای مقابله با تهدیدات بالقوهٔ جنایت، مانند گمراه‌کردن قضاوت سرمایه‌گذاری‌ها و توصیه‌ها توسط مدل‌های یادگیری عمیق مورد استفاده قرار گیرند. با این حال، از منظر کاربردی، هیچ‌کس هنوز یک الگوریتم دفاعی قدرتمند طراحی نکرده است که بتواند در برابر طیف وسیعی از الگوریتم‌های حملهٔ متخاصم مقاومت کند.



۶،۳. علیت و تفسیر پذیری

در حال حاضر، الگوریتم‌های داده‌های بزرگ که برای ریسک‌های مالی اعمال می‌شوند، عمدتاً مبتنی بر آماری هستند که بر همبستگی عوامل تأکید دارند. اگر این مدل‌ها براساس همبستگی‌ها و نه روابط علّی اساسی ساخته شوند، نسبت به تغییرات محیطی آسیب‌پذیر خواهند بود. مدل‌سازی روابط علّی برای انجام پیش‌بینی‌ها و تصمیم‌گیری‌های مالی ارزشمند است، که می‌تواند استحکام پیش‌بینی را بهبود بخشد، نتایج قابل تفسیر ارائه کند و استنتاج خلاف واقع را امکان‌پذیر کند.

مطالعات سنتی در مورد علّیت به کارآزمایی‌های تصادفی‌سازی و کنترل‌شده (RCTs) برای تعیین رابطهٔ علّی دقیق نیاز دارند. با این حال، داده‌های بزرگ رویکرد دیگری را برای ما به ارمغان می‌آورد، یعنی کاوش روابط علّی در داخل داده‌های مشاهده‌ای. با پیروی از روش مدل علّی ساختاریافته (SCM)، روابط علّی با یک نمودار غیرچرخه‌ای جهت‌دار (DAG) نشان داده می‌شود، که در آن گره‌ها عامل هستند و یال‌ها روابط علت- معلولی مستقیم هستند. کشف علّی یک تکنیک امیدوارکننده برای استخراج روابط علّی در داده‌های مالی است. روش‌های مؤثر بسیاری برای کشف علّی برای داده‌های ثابت وجود دارد. با این وجود، چالش‌های آینده برای مدل‌سازی روابط علّی در امور مالی عمدتاً در مسائل عوامل پنهان نهفته است.

عوامل پنهان، عواملی حیاتی امّا غیرقابل مشاهده هستند و ممکن است به‌عنوان عوامل مخدوش کنندهٔ سایر عوامل

قابل مشاهده عمل کنند و منجر به همبستگی‌های جعلی شوند.

۶،۴. حریم خصوصی

همان‌طور که قبلاً ذکر شد، داده‌های بزرگ می‌توانند بسیاری از برنامه‌های مالی را تسهیل کنند. با این حال، همچنان نگرانی‌های مربوط به حریم خصوصی، به‌ویژه در زمینهٔ مالی وجود دارد. داده‌های مورد استفاده در برنامه‌های مالی اغلب حاوی اطلاعات حساسی در مورد کاربران هستند، در حالی که شرکت‌ها تمایل دارند داده‌ها را برای خود نگه دارند. بنابراین، شرکت‌هایی که مکان کاربران را از طریق برنامه‌های پیش‌بینی آب و هوا جمع‌آوری می‌کنند، می‌توانند نسبت به دفاتر اعتباری سنتی توانایی بهتری در پیش‌بینی اعتبار کاربران داشته باشند. علاوه بر عادی‌سازی استفاده از اطلاعات مشترک کاربران، هماهنگی جهانی اشتراک‌گذاری داده‌ها از کشورها یا بخش‌های مختلف برای شناسایی خطرات در سراسر تک‌تک مرزها بسیار مهم است. با این حال، به‌دلیل سیاست‌های پیچیدهٔ اشتراک‌گذاری داده‌ها در کشورهای مختلف یا منافع و قانون‌گذاران مختلف در یک کشور، چنین اشتراک‌گذاری بین مرزی همیشه چالش برانگیز است. این منجر به تکه‌تکه شدن بیشتر اقتصاد دیجیتال جهانی می‌شود و مانع از مبارزه با ریسک‌های مالی می‌شود. بنابراین می‌توان این سؤال را مطرح کرد که چگونه می‌توان از داده‌های مربوط به اطلاعات حساس کاربران به درستی و ایمن استفاده کرد و در عین حال پراکندگی داده‌های اقتصادی را از بین برد؟ الگوریتم‌هایی مانند حریم خصوصی متفاوت که توسط MIT Technology Review به‌عنوان یکی از ۱۰ فناوری پیشرفت ۲۰۲۰ فهرست شده است، یکی از دستورالعمل‌های کلیدی برای مبارزه با ریسک‌های مالی در آینده است.

۶،۵. شبیه‌سازی چند عاملی

اکثر مطالعات کلان و خرد در مورد مشکلات مالی بر موضوعات ناسازگاری تمرکز می‌کنند که ارتباطات نادری

بین آنها وجود دارد. داده‌های حسگرهای منفرد یک دید خرد از رفتارهای مالی ارائه می‌دهند، در حالی که آمار یک نمای کلان ارائه می‌دهد. اخیراً، شبیه‌سازی یادگیری تقویتی چند عاملی نتایج شگفت‌آوری را نشان داده است. شبیه‌سازی مبتنی بر عامل، با توجه به محیط‌ها و جوایز مناسب، توانایی یافتن راه‌حل یا پیش‌بینی بهتر از انسان‌ها را نشان می‌دهد.

بنابراین، محیط‌های نزدیک به اقتصاد دنیای واقعی، شبیه‌سازی‌های چند عاملی و یادگیری می‌توانند امیدوارانه راه‌حلی سیستماتیک برای درک عمیق سیستم اقتصاد فیزیکی ما و پرداختن به خطرات در حال ظهور در آن باشند. اگرچه مفهوم «دوقلوهای دیجیتال» برای شبیه‌سازی مدل فیزیکی فضاپیما آغاز شد امّا به موضوعی نوظهور در اقتصاد دیجیتال تبدیل شده است. با این حال، پیچیدگی دنیای واقعی باعث می‌شود که این دوقلو دیجیتال به دور از تحقّق باشد.

۷. نتیجه گیری

یک چهارچوب بازبینی برای طبقه‌بندی آثار مرتبط پیشنهاد شده است:

- از چه داده‌هایی استفاده شود
- چگونه می‌توان داده‌های بزرگ را با ابزارهای نوظهوری که می‌توانند تجزیه و تحلیل کنند یا از آنها بیاموزند، تقویت کنیم
- برجسته کردن میزان موفقیت کار تحقیقاتی در کاربردهای مختلف

در نهایت، علاوه بر بحث در مورد روش‌هایی که می‌توانند مسائل مربوط به حریم خصوصی را برطرف کنند، فهرستی از این که در آینده چه کارهایی را از نظر مدیریت خطرات ناشی از تکنیک‌های دیپ‌فیک، حملات خصمانه به مدل‌های عمیق، روش‌های علت- معلولی و شبیه‌سازی دنیای فیزیکی انجام شود، را ارائه کردیم. هنوز مشکلات باز هستند و شتاب بیشتری می‌گیرند. ■

پی‌نوشت‌ها:

1- Natural Language Processing) NLP)
2- iResearch
3- eZubao
4- Financial Crime-Compliance(FCC)
5- National Science Foundation(NSF)
6- The National Natural Science of China (NSFC)
7- TextCube
8- edges
9- YAGO
10- DBPedia
11- ICEWS 5
12- GDELT6
13- Harbin Institute of Technology
14- LV et al.
15- Deep learning and representation learning
16- convolutional neural network (CNN)
17- recurrent neural networks (RNNs)
18- long short term memory (LSTM)
۱۹-کانولوشن در فارسی پیچشی یا همگشتی نیز ترجمه شده است
20- PALE
21- Knowledge graph (KG)
22-Holoscope
23- modularity
24- Palla et al
25-SpecGreedy2017
26-Ha et al.
27- Zhang et al.
28-Tavana et al.
29-Guijaro et al.
30- S&P500
31- Value at Risk(VaR)
32- Fat tail
33- Fischer and Krauss
34- S&P500
35- Martinez et al.
36- Nyman et al.
37- Zhou et al.
38- Catullo et al.
39- Yu et al.
40- vector auto-regression (VAR)
41- Ahelegbey et al.
42- O'Halloran and Nowaczyk
43- Wang et al.
44- Rao et al.
45- Benavides et al.
46- Malini and Pushpa
47- Askari and Hussain
48- Hajek and Henriques
49- Yao et al
50- Craja et al.
51- Wang and Xu
52- Colladon and Remondi
53- FlowScope
54- Sun et al.
55- tensors
56- Deepfake
57- karate club network

منبع

Cheng, X. , Liu, S. , Sun, X. , Wang, Z. , Zhou, H. , Shao, Y. , & Shen, H. (2021). Combating Emerging Financial Risks in the Big Data Era: A Perspective Review. Fundamental Research.

فاطمه دادبه: دکتری مدیریت مالی
سمیه همتی