مطالعات مهندسی صنایع و مدیریت تولید

دوره ۳، شماره ۲، تابستان ۱۳۹۶

صفحات ۳۸- ۲۷

Online ISSN: ۲۴۷۶-۴۴۵۰

Print ISSN: ۲۵۳۸-۲۱۰۱

www.irijournals.com

تحلیل و کشف تقلب در خرید آنلاین با استفاده از شبکه عصبی یادگیری بیزی

سید علی حسینی [۱]، حامد فضل الله تبار [۲]

[۱] دانشجوی کارشناسی‌ارشد مهندسی فناوری اطلاعات، دانشگاه علوم و فنون مازندران، مازندران، ایران

[۲] استادیار مهندسی صنایع، دانشگاه علوم و فنون مازندران، مازندران، ایران

چکیده

هدف از انجام این پژوهش تحلیل و کشف تقلب هایی است که مشتریان در خرید آنلاین از طریق فیشینگ دچار آن می شوند که در این راستا از الگوریتم شبکه عصبی یادگیری بیزی استفاده نمودیم. شکی نیست که استفاده از اینترنت روز به روز بیشتر می شود و ما ناگزیریم که برخی از کارهای مالی خود را نیز با اینترنت انجام دهیم؛ بنابراین باید راه مقابله با این کلاهبرداری ها را بدانیم. کلاه برداری اینترنتی یعنی استفاده از سرویس های اینترنت به منظور فریب اشخاص یا شرکت ها و انتقال و دزدی پول هایشان. این کلاهبرداری ممکن است در اتاق های گفتگوی اینترنتی، نامه های الکترونیک یا وب سایت ها رخ دهد. در این تحقیق از داده های یک بانک بزرگ برزیلی بهره گرفته شده است. تراکنش های صورت گرفته در بازه زمانی جولای ۲۰۰۴ تا سپتامبر ۲۰۰۴ می باشد. ارزیابی این مدل پس از یادگیری به وسیله داده های آموزشی، نشان می دهد که مدل پیشنهادی تمامی تراکنش های تقلبی داده های آزمایشی را تشخیص داده و اعلان مناسب صادر می کند.

واژه‌های کلیدی: تقلب، خرید آنلاین، تجارت الکترونیک، کلاه برداری، شبکه عصبی یادگیری بیزی.

۱- مقدمه

در عصر حاضر فضای مجازی، در تمامی ابعاد زندگی بشر همچون ابعاد اقتصادی، اجتماعی، سیاسی و فرهنگی نفوذ کرده و سبب به وجود آوردن شیوه‌های جدید تولید، ابزار، کالاها و خدمات جدید و روش‌های مختلف ارتباطات شده است .تاثیر این عرصه در بخش اقتصاد، بیش از دیگر بخش‌های جامعه رواج یافته و در نهایت، سبب پیدایش اقتصاد دیجیتالی در عصر حاضر شده است. از جمله مباحث مهم در اقتصاد دیجیتالی، تجارت الکترونیکی و یا همان خرید و فروش الکترونیکی یا خرید و فروش آنلاین است.

ظهور اینترنت به توسعه شکوفایی تجارت الکترونیک منجر شده است. تجارت الکترونیک را می‌توان به عنوان "هرنوع معامله کسب و کاری که در آن مجموعه ای از تراکنش‌های الکترونیکی به جای تبادل فیزیکی یا تماس مستقیم می‌باشند" تعریف کرد (وانگ[1] و همکاران،۲۰۱۵). پس تجارت الکترونیک، خرید و فروش محصولات یا سرویس‌ها از طریق رسانه‌های الکترونیکی، مانند اینترنت و دیگر شبکه های کامپیوتری است (کارمونا[2] و همکاران،۲۰۱۲). از طرفی به وب سایتی که شامل خدمات و کالاهایی برای عرضه به کاربران اینترنتی می باشد، فروشگاه اینترنتی، الکترونیکی یا مجازی گفته می شود. فروشگاه های اینترنتی با توجه به نوع اجناسی که در آنها به فروش می رسد به دو گروه اصلی تقسیم بندی می شوند: business-to-consumer (B2C) online shopping و business-to-business (B2B) online shopping.

تجارت الکترونیکی یا خرید و فروش آنلاین صرفنظر از مزایایی که دارد همچون عدم محدودیت جغرافیایی، عدم تفاوت با فروشگاه فیزیکی، فروش شبانه روزی، تبلیغات وسیعتر و کم هزینه تر، قابلیت پیگیری سریع روند خرید تر، مشکلات فراوانی را هم برای کاربران اینترنت ایجاد کرده است.

کلاهبرداری های اینترنتی در تمامی کشورهای جهان کم و بیش وجود دارد، اما در کشور ما جدای کلاهبرداری های فروشگاه های آنلاین غیرمعتبر، عدم آشنایی عموم کاربران با بانکداری الکترونیک و روش های حفظ اطلاعات آن و... و از طرف دیگر ضعف‌های اطلاعاتی برخی خدمات الکترونیک بانک ها موجب بروز کلاهبرداری هایی شده است.

کلاهبرداران و شیادان همچنان از موهبت ناشناس ماندن در فضای اینترنت نهایت سواستفاده را انجام می دهند و همه این ها هشداری است برای کاربران اینترنت که بیش از همیشه مراقب اطلاعات و پولشان در اینترنت باشند تا از کلاهبرداری اینترنتی در امان بمانند .کلاهبرداری اینترنتی یعنی استفاده از سرویس های اینترنت به منظور فریب اشخاص یا شرکت ها و انتقال و دزدی پول هایشان. این کلاهبرداری ممکن است در اتاق های گفتگوی اینترنتی، نامه های الکترونیک یا وب سایت ها رخ دهد. این مساله امروزه به یکی از مسایل مهم دنیای فناوری اطلاعات تبدیل شده است چرا که هر ساله پول زیادی از این راه رد و بدل شده و قربانیان زیادی به دام این شیادی ها می افتند. شکی نیست که استفاده از اینترنت روز به روز بیشتر می شود و ما ناگزیریم که برخی از کارهای مالی خود را نیز با اینترنت انجام دهیم؛ بنابراین باید راه مقابله با این کلاهبرداری ها را بدانیم.
روش های متعددی برای تشخیص کلاه برداری اجرا شده است. هدف ما ازین پژوهش کشف وتحلیل کلاه برداری هایی است که در خرید آنلاین مشتریان با آن مواجه می شوند چون همزمان خرید آنلاین و کشف تقلب باهم بررسی نشده است که در این راستا از ابزار شبکه عصبی یادگیری بیزی استفاده می شود که یک نوآوری است و تابه حال انجام نشده است.

۲- ادبیات تحقیق

[1] Wang

[2] Carmona

واژه تقلب مفاهیمی نظیر استفاده بدون مجوز، استفاده از هویت جعلی، سرقت الکترونیکی وجه و غیره را شامل می شود. لذا تقلب مفهوم گسترده ای دارد؛ اما تعاریف مختلف ارائه شده برای آن در "انجام عمل بدون مجوز برای بدست آوردن منافع" اتفاق نظر دارند. به عبارتی مفهوم کلی تقلب، استفاده یا دسترسی بدون مجوز به سرویس در جهت بدست آوردن سود است (آلبرچت[1] و همکاران،۲۰۰۸)

کشف تقلب به عملیاتی گویند که برای تصمیم گیری قطعی در رابطه با یک رفتار مشکوک صورت می گیرد. این عملیات باید حتی الامکان سریع و قبل از اتمام تراکنش باشد، ضمن اینکه موارد تکراری باید شناسایی شوند (دیو آبراهام[2]،۲۰۰۹). ابزارهای متفاوتی برای شناسایی تقلب وجود داشته که اغلب الگوریتم های بکار رفته در آنها یادگیری ماشین یا به طور خاص تر داده کاوی می باشد؛ مانند بر پایه قوانین، درخت تصمیم (وانگ[3] و همکاران،۲۰۰۳)، شبکه عصبی (کرنکر[4] و همکاران،۲۰۰۹) سیستم ایمنی مصنوعی (ندا سلطانی، ۱۳۹۰)، شبکه بیزین (سریواستاوا[5] و همکاران، ۲۰۰۸)، خوشه بندی، روش های آماری (دلوگوز و میلر فانک[6]،۲۰۰۹) و منطق فازی.

داده‌کاوی بخشی از فرآیند بزرگ‌تر استخراج دانش از پایگاه‌های داده می‌باشد که در فرآیند تصمیم‌گیری در کاربردهای متعددی مورداستفاده قرار می‌گیرد (سیم[7] و همکاران،۲۰۱۴) و شامل مراحل زیر می‌باشد:

پاک‌سازی داده‌ها: حذف داده‌های نا ایستا و مزاحم

یکپارچه‌سازی داده‌ها: ترکیب منابع داده متعدد و پراکنده و احیاناً ناهمگن

انتخاب داده‌ها: بازیابی داده‌های مربوط به عمل کاوش از پایگاه داده‌ها

تبدیل داده‌ها: تبدیل یا تلفیق داده‌ها به اشکالی مناسب برای بکار بردن روش‌های مختلف آماری

داده‌کاوی: از مرحله‌های ضروری فرآیند KDD است که در آن از روش‌های مختلف آماری برای استخراج الگوها و مدل‌ها استفاده می‌شود.

ارزیابی الگوها: شناسایی الگوهای جذاب ارائه دانش، بر اساس معیارهای جذابیت

ارائه دانش: ارائه دانش استخراج‌شده با استفاده از تکنیک‌های نمایش اطلاعات

در این میان تئوری‌های یادگیری از هوش مصنوعی و یادگیری ماشین از حوزه‌هایی هستند که در تکنیک‌های داده‌کاوی مورداستفاده قرار می‌گیرند. شبکه های عصبی سیستم ها و روش‌های محاسباتی نوینی هستند برای یادگیری ماشینی، نمایش دانش و در انتها اعمال دانش به‌دست‌آمده در جهت پیش‌بینی پاسخ‌های خروجی از سامانه های پیچیده (لانکاشیر[8] و همکاران،۲۰۰۹)

[1] Albrecht

[2] Dave Abraham

[3] Wang

[4] Krenker

[5] Srivastava

[6] Dlugosz and Miller-Funk

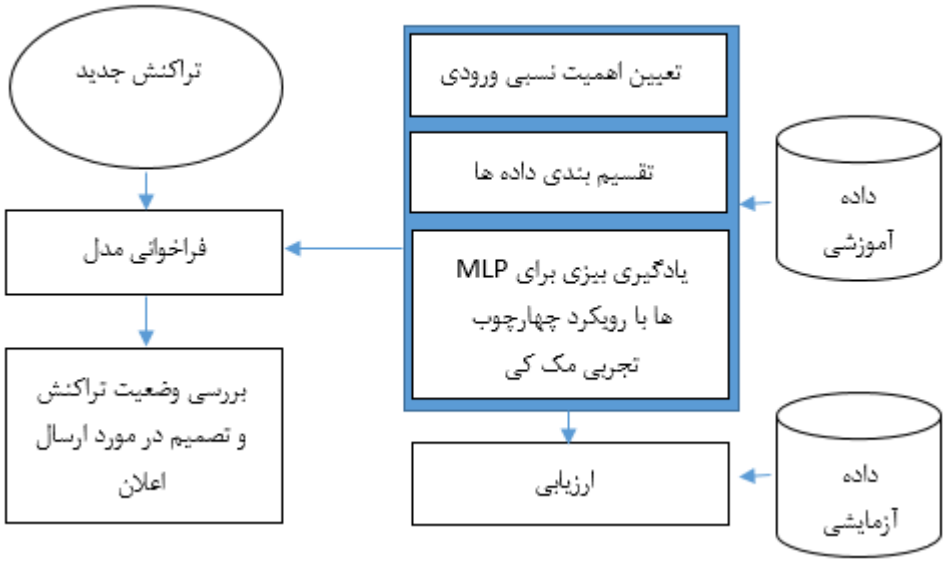
[7] Sim

[8] Lancashire

۳- مدل تحقیق

شکل یک مدل پیشنهادی تحقیق ما را نشان می دهد. این پژوهش به بررسی ظرفیت‌های توضیحی طبقه‌بندی کننده‌های شبکه عصبی با استفاده از تنظیم وزن تعیین ارتباط خودکار پرداخته و یافته‌های حاصل از بکارگیری این شبکه‌ها را برای کشف و شناسایی کلاهبرداری در خرید آنلاین، گزارش می‌دهد.

طرح تابع هدفِ تعیین ارتباط خودکار، روشی جهت تعیین آموزنده‌ترین داده‌ها برای مدل شبکه عصبی آموزش دیده، ارائه می-دهد. پیاده‌سازی رویکرد چارچوب استدلالی Mackay برای یادگیری بیزی به عنوان یک روش عملی برای آموزش این نوع شبکه‌ها پیشنهاد می‌شود.



شکل ۱- مدل تحقیق

۴- روش تحقیق

۴-۱- داده ها

در این تحقیق از داده های به کار گرفته شده در مقاله گدی[1] و همکاران (۲۰۰۸) استفاده شده است. این داده ها از یک بانک بزرگ برزیلی گرفته شده است. تراکنش های صورت گرفته در بازه زمانی جولای ۲۰۰۴ تا سپتامبر ۲۰۰۴ می باشد. سپس با استفاده از قانون زیر داده ها به عنوان تقلبی و یا نرمال برچسب گذاری شده اند:

اگر طی دو ماه پس از وقوع یک تراکنش که به آن دوره کارایی گفته می شود، مشتری آن را به عنوان یک تراکنش تقلبی گزارش داده باشد و یا بانک به آن تراکنش مظنون شده و تایید کرده باشد که آن تراکنش متعلق به مشتری مورد نظر نبوده است، آن تراکنش تقلبی است و در غیر اینصورت تراکنش به عنوان یک تراکنش نرمال در نظر گرفته می شود. اگر تراکنشی به عنوان تقلبی برچسب بخورد بانک اطمینان ۱۰۰٪ در مورد این ادعا دارد اما اگر تراکنشی برچسب نرمال بخورد نمی توان مطمئن بود که این تراکنش تقلبی نیست و تنها می توان گفت که تراکنش در این بازه زمانی هنوز به عنوان تقلبی تشخیص داده نشده است. با این وجود می توان گفت که حداقل ۸۰٪ از تقلب ها طی دو ماه قابل تشخیص هستند.

[1] Gadi

- با توجه به کم بودن تعداد تراکنش های تقلبی به نسبت با تراکنش های نرمال، داده های نرمال گلچین شده صرفا ۱۰٪ از تراکنش های نرمال داده های اولیه بانک می باشد.
- مجموعه داده بدست آمده شامل ۴۱۶۴۷ تراکنش است که ۳/۷۴ ٪ از آنها یعنی ۱۵۵۹ تراکنش تقلبی می باشد.
- از ۴۱۶۴۷ تراکنش گلچین شده، تعداد ۱۶۴۷ تراکنش را برای بکارگیری در نرم افزار کشف تقلب جداسازی کرده ایم.
- از ۴۰۰۰۰ تراکنش، ۶۸/۷۵ ٪ از داده ها (۲۷۵۰۰ تراکنش) را آموزشی قرار داده و ۳۱/۲۵ ٪ (۱۲۵۰۰ تراکنش) مابقی برای تست جدا شده است. جزییات بیشتر داده ها به شرح جدول ۱ است:

جدول ۱ - جزئیات تقسیم داده ها به دو فایل مجزا

	تعداد کل	تعداد تراکنش قانونی	تعداد تراکنش تقلبی	درصد تراکنش تقلبی
تراکنش مربوط به فایل داده آموزشی و آزمایشی	۴۰۰۰۰	۳۸۵۰۰	۱۵۰۰	٪ ۳/۷۵
تراکنش های مربوط به بخش نرم افزار	۱۶۴۷	۱۵۸۸	۵۹	٪ ۳/۵۸
مجموع	۴۱۶۴۷	۴۰۰۸۸	۱۵۵۹	٪ ۳/۷۴

۴-۲- بکارگیری مدل

برای پیاده سازی مدل، ابتدا با استفاده از تابع هدف ARD بهینه شد (وین[1] و همکاران، ۲۰۰۵). اهمیت نسبی ورودی های مدل آموزش دیده را تعیین می نماییم. این تابع به ما امکان می دهد اندازه وزن های مرتبط به هر ورودی را به طور مجزا کنترل نماییم. ما از لحاظ آماری تمامی ورودی ها را نرمال سازی نموده ایم. ما باید بدانیم که کدام داده ها بیشترین سهم را در شناسایی کلاه برداری های فیشینگ در خرید آنلاین دارا هستند. همچنین این اطلاعات باعث بروز راه حل های موثرتر و کم هزینه تر می شود؛ بنابراین هدف نهایی انتخاب ورودی (داده)، انتخاب حداقل ورودی مورد نیاز جهت رسیدن به ساختار مورد نظر داده هاست. در ادامه داده ها را به سه گروه طبقه می کنیم. گروه شماره یک بیشتر شامل تراکنش های تقلبی و گروه شماره دو بیشتر شامل تراکنش های قانونی می باشد. در گروه شماره سه تراکنش های نویز یا پرت قرار گرفته است. در ادامه جزئیات بیشتر هر یک از گروه ها در جدول ۲ ارائه شده است.

جدول ۲ - جزئیات دسته بندی داده ها به چند گروه

	گروه ۱	گروه ۲	گروه ۳	مجموع
تراکنش قانونی	۶	۳۸۴۲۸	۶۶	۳۸۵۰۰
تراکنش های	۱۴۲۰	۶	۷۴	۱۵۰۰

[1] Viaene

تقلبی				
مجموع	۱۴۲۶	۳۸۴۳۴	۱۴۰	۴۰۰۰۰

هدف نهایی یادگیری، تولید مدلی است که بر روی اهداف داده ای جدید خوب اجرا شود. اگر این مورد مد نظر باشد، می توانیم بگوییم که این مدل به خوبی تعمیم دهی می شود. هدف ارزشیابی عملکرد، برآورد میزان خوب اجرا شدن یک مدل بر روی اهداف داده هایی است که در مجموعه آموزشی وجود داشته اند.

هدف یادگیری بیزی، توسعهٔ مدل‌های احتمالی است که داده‌ها را پردازش کرده و پیش‌گویی‌های بهینه ارائه می‌دهند. تفاوت مفهومی بینِ برآورد بیزی و برآورد احتمال حداکثر این است که ما دیگر پارامترهای مدل را ثابت درنظر نمی‌گیریم، بلکه آنها را متغیرهای تصادفی درنظر می‌گیریم که ویژگی‌شان یک مدل احتمال مشترک است. این امر بر اهمیت توجه به عدم قطعیت ذاتی در مورد نگاشت تابع صحیح که از یک نمونه آموزشی معین آموزش می‌بیند، تأکید دارد. دانش پیشین یعنی اینکه عقیدهٔ ما در مورد پارامترهای مدل قبل از مشاهدهٔ داده‌ها، به شکل یک چگالی احتمال پیشین رمزگذاری می‌شود. زمانیکه داده‌ها مشاهده شوند، دانش پیشین را می‌توان با استفاده از قضیه بیزی، به یک چگالی احتمالی پسین تبدیل نمود. این دانش پسین را می‌توان برای انجام پیش‌بینی‌ها استفاده نمود.

رویکرد عملی مورد نظر ما در مورد یادگیری بیزی برای MLPها که به چارچوب تجربی معروف است، شامل یک تقریب گوآسی محلی نسبت به چگالی احتمال پسین در فضای وزنی است.

فرض کنید، مجموعه‌ای از داده‌های آموزشی $D=\{(x^i,t_i)\}_i^N=1$ مفروض است. در چارچوب بیزی، یک مدل MLP آموزش دیده از لحاظ چگالی احتمال پسین بر روی وزن‌های $P(w|D)$ تعریف می‌شود. با توجه به مورد پسین، بواسطه انتگرال‌گیری آن می‌توانیم استنباط کنیم. برای مثال، به منظور پیش‌بنی یک طبقه‌بندی با یک بردار ورودی مشخص x، ما به احتمالی نیاز داریم که x متعلق به طبقه t=1 باشد که به صورت زیر بدست می‌آید:

$$p(t=1|\mathbf{x},D)=\int p(t=1|\mathbf{x},\mathbf{w})p(\mathbf{w}|D)\mathrm{d}\mathbf{w},$$

در اینجا $p(t=1|x,w)$ بواسطه تابع شبکه عصبی y(x) بدست می‌آید. زمانیکه داده‌های آموزشی D مشاهده می‌شوند، ما می‌توانیم چگالی احتمال پیشین را به یک چگالی احتمال پسین $P(w|D,\alpha)$ اصلاح کنیم (با استفاده از قضیهٔ بیزی). برای محاسبه انتگرال در معادله بالا، مک کی چند تقریب ساده‌کننده پیشنهاد می‌دهد. با استفاده از این تقریب ها و قضیه بیز، فرمول نهایی مک کی به صورت زیر است:

$$p(t=1|\mathbf{x},D)\approx \mathrm{sigm}(\kappa(s)a^{\mathrm{MP}}),$$

در اینجا α^{MP}، فعالسازی واحد خروجی حلقوی لوژستیک MLP است که با استفاده از پارامترهای شبکه عصبی بهینه محاسبه می‌شود و

$$\kappa(s)=\left(1+\frac{\pi s^2}{8}\right)^{-1/2},$$

در اینجا s^2 واریانس تقریب گوآسی محلی به $p(\alpha|x,D)$ متمرکز در α^{MP} می‌باشد. برای اطلاعات بیشتر در مورد اجرای دقیق چارچوب تجربی مک کی، به کد منبع جعبه ابزار Netlab برای Matlab و اسناد ارائه شده در مطالعه بی شاپ[1] (۱۹۹۵) و نابنی[2] (۲۰۰۱) رجوع شود.

حال ما با رویکرد چهارچوب تجربی مک کی در مورد یادگیری بیزی برای MLP ها، مدل یادگیری تقلب هر یک از گروه های ذکر شده را ایجاد کردیم.

۴-۳- ارزیابی

برای ارزیابی بهتر کارایی مدل، لازم است پیش بینی ها ونتایج به صورت جداگانه در نظر گرفته شود. در جدول ۳ تصویری از متریک های مختلف دسته بندی برای داده های دو کلاس قانونی و تقلبی را مشاهده خواهید کرد.

جدول ۳- ماتریس تداخل

وضعیت نمونه	پیش بینی (مثبت)	پیش بینی (منفی)
واقعیت (مثبت)	هشدار صحیح (تقلب) TP(True Positive)	عدم هشدار غلط (تقلب) FN(False Negative)
واقعیت (منفی)	هشدار غلط (قانونی) FP (False Positive)	عدم هشدار صحیح (قانونی) TN (True Negative)

برای درک بهتر T را صحیح و F را به عنوان اشتباه در نظر می گیریم. همچنین P نشان دهنده هشدار (تقلب) و N نشان دهنده عدم هشدار (مجاز) است. با این تعریف هریک از این اصطلاحات به صورت زیر تفسیر می شوند.

- TP: هشدار داده شده و هشدار مزبور صحیح است، یعنی سیستم مورد را به عنوان تقلب شناسایی کرده است و این شناسایی صحیح است. (موارد تقلبی که کشف شده اند).
- FP: هشدار داده شده اما این هشدار صحیح نیست، یعنی سیستم مورد را به عنوان تقلب شناسایی کرده است ولی این شناسایی صحیح نیست. (موارد مجازی که به عنوان تقلب شناسایی شده اند.)
- FN: هشدار داده نشده اما این عدم هشدار صحیح نیست، یعنی سیستم مورد را به عنوان مجاز شناسایی کرده است ولی این شناسایی صحیح نیست. (موارد تقلبی که کشف نشده اند)
- TN: هشدار داده نشده و این عدم هشدار صحیح است، یعنی سیستم مورد را به عنوان تراکنش مجاز شناسایی کرده است و این شناسایی صحیح است. (موارد مجازی که درست شناسایی شده اند)

در ادامه معیار ارزیابی معرفی خواهد شد که رویکرد سودگرایی داشته و بر اساس مبالغ تراکنش ها میزان سود گرایی مدل مورد نظر را بدست خواهد آورد. از جمله مزیت این معیار این است که هر چهار حالت پیش بینی را به صورت زیر دربر خواهد داشت:

[1] Bishop

[2] Nabney

- TP: به علت جلوگیری از انجام تراکنش تقلبی، برای بانک سودآوری داشته است.
- TN: به علت عدم جلوگیری از انجام تراکنش قانونی، برای بانک سودآوری داشته است.
- FP: به علت جلوگیری از انجام تراکنش قانونی، هزینه به بانک تحمیل شده است.
- FN: به علت عدم جلوگیری از انجام تراکنش تقلبی، هزینه به بانک تحمیل شده است.

در سیستم بانکی تراکنش ها دارای ارزش متفاوتی می باشند. برای مثال تراکنش با ارزش مالی صد هزار تومان مهم تر از تراکنش با ارزش مالی هزار تومان خواهد بود. در نتیجه برای اینکه محاسبات ارزش واقعی داشته باشد، مبلغ تراکنش را به عنوان ارزش هر تراکنش در نظر خواهیم گرفت.

همچنین عامل دیگر اثر گذار، ضریب هر نوع متریک خواهد بود. برای مثال تراکنش های از نوع FPو FN به علت اشتباه در پیش بینی درست (هزینه دار بودن) چالش های دیگری برای بانک ایجاد خواهد کرد، مثلا برای FN چالش هایی نظیر هزینه پیگیری و عدم اعتماد مشتری برای بانک وجود خواهد داشت. در ادامه ضرایب اثرگذاری هر یک از متریک ها تشریح شده است:

ضریب α برای TP: این ضریب برای تعیین مقدار سودمندی تراکنش های از نوع TP می باشد.

ضریب β برای TN: این ضریب برای تعیین مقدار سودمندی ناشی از عدم جلوگیری انجام تراکنش های قانونی می باشد.

ضریب γ برای FP: این ضریب برای هزینه ناشی از هشدار غلط می باشد. برای مثال هزینه ناشی از عدم اعتماد مشتری و هزینه پیگیری برای اصلاح تراکنش برای این حالت می باشد.

ضریب δ برای FN: این ضریب برای هزینه های ناشی از عدم هشدار تراکنش های غیرقانونی می باشد. هزینه ناشی از عدم اعتماد مشتری، هزینه پیگیری اشتباه به وجود آمده و زمان صرف شده برای مشتری و مهم تر از آن هزینه مربوط به برداشت شدن غیرقانونی وجه، از جمله هزینه هایی است که در این حالت رخ خواهد داد.

هریک از ضرایب اشاره شده بالا، با توجه به عواملی نظیر کمیت داده، اهمیت کشف و غیره توسط افراد خبره بانکی قابل تعیین می باشد. در نهایت فرمول بدست آمده به صورت زیر خواهد بود:

$$\text{Gain} = \sum_{k=0}^{tpc}(\alpha * V_k) + \sum_{k=0}^{tnc}(\beta * V_k) + \sum_{k=0}^{fpc}(\gamma * V_k) + \sum_{k=0}^{fnc}(\delta * V_k)$$

متغییر V_k اشاره به مبلغ هر تراکنش دارد. حال می توان با کمک ضرایب تعیین شده، تابع دیگری را تعریف کرد که در آن حداکثر میزان سود قابل حصول مشخص می شود. حداکثر میزان سود زمانی بدست می آید که نرم افزار کشف درست کار کند. در واقع این ابزار مانع انجام تراکنش های تقلبی شده و برای کلیه تراکنش های قانونی نیز هشداری صادر نکند.

$$\text{Total Profit} = (\beta * \text{Legitimate TransVal}) + (\alpha * \text{Fraud TransVal})$$

حال با داشتن میزان سودمندی بدست آمده و حداکثر سود قابل حصول، درصد سودمندی را می توان به صورت زیر تعیین کرد.

$$\text{Profit Percent} = \frac{\text{Gain}}{\textit{Total Profit}}$$

در محاسبه میزان سود، ضرایب موجود (میزان اثرپذیری هر حالت) در فرمول تعیین سود را با توجه به نظر فرد خبره بانکی و به صورت زیر مقداردهی کرده ایم:

$\alpha = 2$

$\beta = 1$

$\gamma = -3$

$\delta = -5$

۵- یافته‌های تحقیق

این تحقیق سعی دارد تا با آموزش مدل پیشنهادی به وسیله داده های موجود، سیستمی را طراحی کند که بتواند تقلب های صورت گرفته در خرید آنلاین را شناسایی کند. جدول ۴ نتایج بدست آمده از اجرای الگوریتم شبکه عصبی یادگیری بیزی و جزئیات اجرای مدل پیشنهادی بر روی تراکنش های تست را نشان می دهد.

جدول ۴ - جزئیات اجرای مدل پیشنهادی بر روی تراکنش های تست

نوع	گروه ۱	گروه ۲	گروه ۳	مجموع
کل تراکنش ها	۱۴۲۶	۳۸۴۳۴	۱۴۰	۴۰۰۰۰
تراکنش های تست (٪۳۱)	۴۴۶	۱۲۰۱۱	۴۳	۱۲۵۰۰
TN	۴	۱۲۰۷	۰	۱۲۰۱۱
FN	۰	۰	۰	۰
TP	۴۴۲	۴	۳۱	۴۷۷
FP	۰	۰	۱۲	۱۲

همانطور که در جدول بالا مشاهده می شود، ما بر روی ۳۱٪ از داده ها یعنی ۱۲۵۰۰ تراکنش مدلمان را جهت تست پیاده کردیم. (مابقی داده ها یعنی ۲۷۵۰۰ تراکنش برای آموزش مدل بوده است).

در گروه یک که بیشتر تراکنش ها در آن تقلبی بوده، از ۴۴۶ تراکنش تست، مدل، ۴ تراکنش را که قانونی بوده اند مجاز دانسته و برای آنها هشدار صادر نکرده و باعث سود دهی شده است. برای ۴۴۲ تراکنشی که تقلبی بوده اند نیز هشدار صادر کرده و ازین طریق باعث سودرسانی شده است. همچنین در این گروه تعداد موارد مجازی که به عنوان تقلب شناسایی شده اند (FP) و موارد تقلبی که کشف نشده اند (FN) نیز صفر است.

در گروه دو که بیشتر تراکنش ها در آن قانونی بوده، از ۱۲۰۱۱ تراکنش تست، مدل، ۱۲۰۷ تراکنش را که قانونی بوده اند مجازدانسته و برای آنها هشدار صادر نکرده و باعث سود دهی شده است. برای ۴ تراکنشی که تقلبی بوده است نیز هشدار صادر

کرده و ازین طریق باعث سودرسانی شده است. همچنین در این گروه تعداد موارد مجازی که به عنوان تقلب شناسایی شده انـد (FP) و موارد تقلبی که کشف نشده اند (FN) نیز صفر است.

در گروه سوم نیز از ۴۳ تراکنش موجود برای تست، مدل پیشنهادی ما تعداد ۳۱ تراکنش را که تقلبی بوده اند کشف کرده و هشدار صادر کرده که منجر به سود دهی شده است اما ۱۲ تراکنشی که قانونی بوده اند را تقلبی تشخیص داده است. ولی تمامی موارد تقلب را شناسایی کرده است.

۶- بحث و نتیجه‌گیری

درک معناشناسیِ زیربنای خروجی مدل‌های شبکه عصبی ثابت کنندۀ یک جنبۀ مهم از پذیرش آنها توسط متخصصـین حـوزه برای تحلیل روتین و اهداف تصمیم‌گیری می‌باشد. لذا در این پایان نامه یک سیستم کشف تقلب در خرید آنلاین بـا اسـتفاده از شبکه های عصبی یادگیری بیزی طراحی شد که از داده های تراکنش های یک بانک برزیلـی اسـتفاده کـرده اسـت. مطالعـات پیشین هیچ کدام تقلب را در حوزه خرید آنلاین و فیشینگ مطرح نکرده بودند که استفاده از تلفیق دو الگوریتم شبکه عصبی و شبکه بیزی یک نوآوری در این زمینه بوده است. در ابتدا قسمتی از داده ها مورد آموزش قرار گرفتـه انـد و پـس از یـادگیری، مدل با سایر داده ها مورد ارزیابی قرار گرفت.

از جمله فاکتور بسیار مهم در سیستم های کشف تقلب، نظیر سیستم پیشنهادی این پایان نامه، به حداقل رسـاندن معیـار FN است. همانطور که مشاهده شد، این مدل توانست معیـار FN را بـه صـفر برسـاند کـه تـاثیر بسـزایی در کـارایی سیسـتم دارد. همچنین از نظر بقیه معیار ها (FP،TP،TN) که با ارسال به موقع اعلان و عدم ارسال اعلان در زمان اشـتباه، باعـث سودرسـانی می شوند، این مدل عملکرد خیلی خوبی داشته است.

از نظر معیار معرفی شده سود نیز مدل معرفی شده عملکرد قابل ملاحظه ای دارد و اختلاف ناچیزی با حداکثر میزان سود محاسبه شده دارد. این معیار رویکرد جامعی داشته و برای محاسبه سود، هر چهار حالت پیش بینی را در نظر می گیرد. همچنین ضریب اثرپذیری هر چهار حالت در معادله پیشنهادی ثابت و از پیش تعیین شده نبوده و با توجه به داده های بانکی، این ضرایب توسط فرد خبره بانکی قابل تغییر می باشد.

منابع

۱. سلطانی، ندا (۱۳۹۰). ارائه مدلی برای سرویس کشف تقلب در محاسبات ابری با استفاده از سیستم ایمنی مصنوعی، پایان نامه کارشناسی ارشد دانشگاه امیرکبیر، دانشکده کامپیوتر و فناوری اطلاعات.

2. A.Krenker. M. Volk, and et al. "Bidirectional Artificial Neural Networks for Mobile-Phone Fraud Detection". ETRI Journal, Vol. 31, No. 1, pp. 92-94, 2009.
3. A.Srivastava, A. Kundu, and et al, "Credit Card Fraud Detection Using Hidden Markov Model", IEEE Transactions on dependable and secure computing, Vol. 5, No. 1 pp. 37-48, 2008.
4. Bishop, C. M. (1995). Neural networks for pattern recognition. Oxford: Oxford University Press.
5. C.J. Carmona a*, S. Ramirez-Gallego a, F. Torres b, E. Bernal c, M.J. del Jesus a, S. Garcia a. ‘‘Web usage mining to improve the design of an e-commerce website: OrOliveSur.com’’, Expert Systems with Applications 39 (2012) 11243–11249

6. Dave Abraham, CEO, Signify, the book of "Why 2FA in the clouds?" September 2009, Abraham has a degree in Applied Computing from East Anglia University in Norwich.
7. H. Wang, W. Fan, P. Yu, J. Han, Mining Concept Drifting Data Stream Using Ensemble Classifiers, Proc. Of SigKDD, 2003.
8. Lancashire LJ, Lemetre C, Ball GR. An introduction to artificial neural networks in bioinformatics—application to complex microarray and mass spectrometry datasets in cancer studies. Brief Bioinform 2009; 10(3): 315-29.
9. M. Gadi, X. Wang, A. Lago, Comparison with Parametric Optimization in Credit Card Fraud Detection, IEEE, 2008.
10. Nabney, I. T. (2001). Netlab: Algorithms for pattern recognition. New York: Springer.
11. S. Dlugosz, and U. Miller-Funk, "The value of the last digit: statistical fraud detection with digit analysis". Advances in Data Analysis and Classification, vol. 3, No.3, pp. 281—290, 2009.
12. S. Viaene, G. Dedene, R.A. Derrig. "Auto claim fraud detection using Bayesian learning neural networks". Expert Systems with Applications 29 (2005) 653–666.
13. Shan Wang Hasan Cavusoglu Ziliang Deng, "Early Mover Advantage in E-commerce Platforms with Low Entry Barriers: The Role of Customer Relationship Management Capabilities", Information and Management (2015)
14. Sim, J.J. Tan, G.W.H. Wong, H.C.J. Ooi, K.B. Hew, T.S. 2014. Understanding and predicting the motivators of mobile music acceptance – a multi-stage MRAartificial neural network approach. Telematics Inform. 31 (4), 569–584.
15. W. Albrecht, C. Albrecht, and et al, "Current Trends in Fraud and its Detection", Information Security Journal: A Global Perspective, Vol. 17, No. 1, pp.2-12, 2008.

مطالعات مهندسی صنایع و مدیریت تولید

دوره ۳، شماره ۲، تابستان ۱۳۹۶، صفحات ۳۸-۲۷

Fraud Detection Analysis in Online Shopping Process Using Bayesian learning Neural Networks

Seyed Ali Hosseini, Hamed Fazlollahtabar
Mazandaran University of Science and Technology, Babol, Iran.

Abstract

The purpose of this research is to analyze and detect the fraud that customers are having in online purchasing through phishing. In this regard, we use Bayesian learning neural network algorithm. There is no doubt that the use of the Internet is increasing day by day, and we have to do some of our financial work with the Internet, so we need to know how to deal with these scams. Internet fraud is the use of Internet services in order to deceive individuals or companies and transfer and steal their money. This scam may occur in chat rooms, emails, or websites. In this study, we use huge Brazilian database. Transactions conducted between July 2004 and September 2004. The evaluation of this model -after learning by educational data- shows that the proposed model detects all fake transaction data of the test data and issues an appropriate notification.

Keywords: Fraud Detection Analysis, Online Shopping, Bayesian learning Neural Networks.