پژوهش‌های پولی-بانکی
سال ششم، شماره ۱۸، زمستان ۱۳۹۲
صفحات ۹۱-۱۰۸

تشخیص تقلب در تراکنش‌های کارت‌های اعتباری با الگوریتم سلول‌های دندریتیک

سمانه سرورنژاد [*]
رضا ابراهیمی آتانی [†]
سید امیرحسن منجمی [‡]

تاریخ دریافت: ۱۳۹۱/۱۰/۰۳ تاریخ پذیرش: ۱۳۹۲/۰۴/۰۱

چکیده

عدم امنیت تراکنش‌ها یکی از مهم‌ترین موانع برای استفاده و ترویج بانکداری الکترونیکی است و تشخیص تقلب از مسایل مهم در مؤسسات مالی و به‌ویژه بانک‌ها به‌شمار می‌رود. از آنجا که روشی مطمئن همراه با نرخ کشف بالا جهت تشخیص تقلب در کارت‌های اعتباری وجود ندارد، بنابراین در پژوهش حاضر با استفاده از الگوریتم مبتنی بر سیستم ایمنی بدن انسان و به‌طور خاص نظریه خطر موسوم به الگوریتم سلول‌های دندریتیک برای کشف تقلب در تراکنش‌های کارت‌های اعتباری تمرکز کرده‌ایم. این الگوریتم به‌دلیل نداشتن فاز آموزش برای پایگاه داده‌های آنلاین (پرکاربرد در تشخیص تقلب‌های آنلاین) بسیار مناسب است. همچنین، به‌علت دارابودن ساختار ساده دارای سرعت پردازش بالاست. با توجه به نتایج به‌دست‌آمده می‌توان این‌گونه بیان کرد که الگوریتم سلول‌های دندریتیک، الگوریتمی مناسب با دقت مناسب (۹۰ درصد) جهت تشخیص تقلب در تراکنش‌های کارت‌های اعتباری است. این الگوریتم به‌دلیل اعمال علامت امن تعداد شناسایی اشتباه تراکنش‌های نرمال به‌عنوان متقلب را کاهش می‌دهد.

واژه‌های کلیدی: سیستم ایمنی، امنیت، بانکداری الکترونیک
طبقه‌بندی JEL: L81، A12، L86، F47

[*] کارشناس ارشد، دانشکده اطلاعات و تجارت، دانشگاه گیلان، رشت؛ soroumejad@msc.guilan.ac.ir
[†] استادیار، دانشکده مهندسی کامپیوتر، دانشگاه گیلان، رشت؛ rebrahimi@guilan.ac.ir (نویسنده مسئول)
[‡] دانشیار، دانشکده مهندسی کامپیوتر، دانشگاه اصفهان، اصفهان؛ monadjemi@eng.ui.ac.ir

۱ مقدمه

امروزه شناخت فعالیت‌های کلاهبرداران و متقلبان دارای تأثیر مستقیم بر خدمت‌رسانی به مشتریان، کاهش هزینه‌های عملیاتی و باقی ماندن به‌عنوان یک ارائه‌دهنده خدمات مالی معتبر و قابل اطمینان است. بنابراین، مؤسسات مالی و پولی به‌ویژه بانک‌ها به‌شدت به‌دنبال سرعت عمل و همچنین صحت و دقت در شناخت فعالیت‌های متقلبان هستند. به‌کارگیری تکنیک‌های شناسایی تقلب به‌منظور کشف تقلب و جلوگیری از اقدامات متقلبانه در سیستم‌های بانکداری الکترونیک، به‌ویژه کارت‌های اعتباری که جزء مهمی از نظام بانکداری الکترونیک محسوب می‌گردند، اجتناب‌ناپذیر است.

برای واژه تقلب در مقالات و منابع علمی، معانی مختلفی بیان شده است، اما آنچه در تمامی این تعاریف مشترک است، این است که تقلب، نوعی سوءاستفاده از منابع، در جهت منافع شخصی، به‌عمد و کاملاً غیرقانونی است.[۱] کشف تقلب عبارت است از عملیاتی که برای تصمیم‌گیری قطعی در رابطه با یک رفتار مشکوک صورت می‌گیرد. این عملیات باید تاحدامکان سریع و پیش از اتمام تراکنش باشد، ضمن اینکه موارد تکراری باید شناسایی شوند.[۲] بنابراین در پژوهش حاضر بر روی کشف تقلب در تراکنش‌های کارت‌های اعتباری با استفاده از روش مبتنی بر سیستم ایمنی بدن انسان تمرکز شده است.[۳]

توانایی‌های سیستم ایمنی بدن انسان به‌ویژه ایمنی اکتسابی الهام‌بخش سیستم‌های ایمنی مصنوعی بوده است. پژوهش‌هایی که در موضوعاتی نظیر تشخیص الگو، تشخیص ناهنجاری در شبکه‌ها، داده‌کاوی و تشخیص تقلب انجام گرفته، نشان‌دهنده پتانسیل بالای این الگوریتم‌هاست. این الگوریتم‌ها عموماً در پنج دسته مغز استخوان[۴]، انتخاب منفی[۵]، انتخاب کلونال[۶]، شبکه ایمنی[۷] و نظریه خطر[۸] (الگوریتم سلول‌های دندریتیک) جای می‌گیرند. در این پژوهش بر جدیدترین بخش یعنی نظریه خطر تمرکز می‌شود.

---

[1] Leung, Yan & Fong (2004)

[2] Ravisankar, Ravi, Rao & Bose, (2011)

[۳] توضیحاتی در رابطه با نحوه عمل سیستم ایمنی بدن انسان در پیوست ارائه شده است.

[4] bone marrow

[5] negative selection

[6] clonal selection

[7] immune network

[8] danger theory

در این پژوهش، در بخش ۲ به مروری بر ادبیات حوزه تشخیص تقلب در تراکنش‌های کارت‌های اعتباری با استفاده از تکنیک‌های مختلف خواهیم پرداخت. در بخش ۳ نظریه خطر به‌همراه الگوریتم مبتنی بر آن مورد بحث و بررسی قرار خواهد گرفت. در بخش ۴ مجموعه داده به‌کاررفته معرفی‌شده و نتایج به‌دست‌آمده بیان می‌شود. در بخش ۵ مدل سلول‌های دندریتیک ارائه می‌شود و درنهایت، در بخش ۶ نتیجه‌گیری، محدودیت‌ها و راهکارهایی برای مطالعات آینده ارائه خواهد شد.

۲ مرور ادبیات

در حوزه تشخیص تقلب در تراکنش‌های کارت‌های اعتباری پژوهش‌های متعددی انجام گرفته است که در ادامه برخی از آنها معرفی می‌شود.

کورچادو و آلونسو[۱] (۲۰۰۲)، سیستمی برای کشف تقلب تلفن‌همراه ارائه نموده‌اند که مبتنی بر شبکه عصبی مصنوعی دوطرفه است. هسته سیستم یک Bi-ANN[۲] است که رفتار مشتری تلفن‌همراه را پیش‌بینی می‌کند. این کار از طریق نظارت بر رفتار مشتری و پیش‌بینی رفتار مشتریان بر اساس سری‌های زمانی صورت می‌پذیرد. مزیت کلیدی چنین سیستمی توانایی آن در کشف تقلب با استفاده از رکوردهای جزئیات تماس و به‌شکل بی‌درنگ است. همان‌طور که اشاره شد، ارزیابی نتایج در این پژوهش جهت کشف تقلب، تنها بر اساس جزئیات رکوردهای تماس بوده و روی مشتریان تلفن همراه انجام می‌گیرد.

در مطالعه کرنکر، ولک، سدلار، بستر و کش[۳] (۲۰۰۹) از الگوریتم ماشین بردار پشتیبان[۴] جهت تشخیص تقلب استفاده شده است. الگوریتم ماشین بردار پشتیبان که به اختصار به آن بردار ماشین گفته می‌شود، یک الگوریتم رایانه‌ای است که به کمک مثال یاد می‌گیرد چگونه به اشیای مختلف برچسب‌های مرتبط به آنها را اختصاص دهد. در این پژوهش از تحلیل مؤلفه‌های اصلی[۵] جهت مشخص‌کردن ویژگی‌های مهم تراکنش ورودی از روی ویژگی‌های تراکنش‌ها با ابعاد مختلف و تعداد زیاد استفاده می‌شود. در این پژوهش، از روش تالاب‌های ساخته‌شده انباشته[۶] جهت وزن‌دادن به ویژگی‌های انتخاب‌شده مرحله پیش

[1] Corchado & Alonso
[2] Bidirectional Artificial Neural Network
[3] Krenker, Volk, Sedlar, Bester & Kosh
[4] support vector machine
[5] principal component analysis
[6] integrated constructed wetlands

استفاده می‌شود. اگرچه استفاده از تحلیل مؤلفه‌های اصلی باعث بهبود سرعت این الگوریتم شده، اما کاهش میزان دقت این الگوریتم را نیز به‌همراه داشته است.

اوسونا، فروند و گیروسیت[1] (۱۹۹۷) الگوریتمی پیشنهاد کردند که پایه و اساس آن الگوریتم ژنتیک بود. الگوریتم ژنتیک نوع خاصی از الگوریتم‌های تکامل است که از تکنیک‌های زیست‌شناسی فراگشتی مانند وراثت و جهش استفاده می‌کند. این الگوریتم بر اساس اصول انتخاب طبیعی داروین[2] برای یافتن فرمول بهینه جهت پیش‌بینی یا تطبیق الگو است. در پژوهش مذکور بر پایه الگوریتم ژنتیک و قابلیت منطق تخمین‌زنی، تراکنش‌های کارت‌های اعتباری به دو دسته مشکوک و غیرمشکوک طبقه‌بندی می‌شوند. اساساً، این روش فرآیند امتیازدهی را دنبال می‌کند.

دومان و حمدی[3] (۲۰۱۱) از روشی مبتنی بر تطبیق دنباله‌ها جهت تشخیص تقلب در کارت‌های اعتباری استفاده نموده‌اند. تطبیق دنباله‌ها در بایوانفورماتیک[4] برای شباهت بین رشته‌های ژنی استفاده می‌شود و یک ابزار برای اندازه‌گیری و ارزیابی شباهت بین دو یا چند دنباله است. در پردازش تراکنش‌های کارت‌های اعتباری، دنبالهٔ استفاده (رفتارهای قبلی مشتری) شامل اطلاعاتی راجع به مبلغ تراکنش، فاصله زمانی از آخرین خرید، روز، هفته و غیره، برای صادرکننده کارت دردسترس است. هر انحراف دنبالهٔ استفاده از دنبالهٔ موجود (تراکنش مورد بررسی) می‌تواند از طریق تطبیق دنباله‌ها محاسبه شود. در این پژوهش به‌علت عمل تطبیق و مقایسه تک‌تک دنباله‌ها با یکدیگر، عمل پردازش وقت‌گیر است، ضمن اینکه در این پژوهش به نکته تقلید رفتار فرد متقلب از رفتار صاحب اصلی کارت به‌هنگام عمل تقلب،که عموماً یکی از ترفندهای متقلبان است توجهی نشده است.

در کوندو، سورال و مجومدار[5] (۲۰۰۹) از روشی مبتنی بر مدل مخفی مارکوف[6] جهت تشخیص تقلب استفاده شده است. این روش نیز پارامترهای مربوط به تراکنش‌ها را به‌کار می‌برد و با استفاده از رفتارهای دریافتی مشتریان الگوهایی را می‌سازد و سپس با استفاده از امتیازدهی سریع به هر تراکنش وارده بر اساس پارامترهای از پیش‌مشخص‌شده، تصمیم

[1] Osuna, Freund & Girosit
[2] Darwin
[3] Duman & Hamdi
[4] Bioinformatics
[5] Kundu, Sural & Majumdar
[6] Markov

می‌گیرد که تراکنش پذیرفته و یا رد شود. این روش یک روش آماری است که از مقادیر پارامترها در یک محدوده خاص استفاده کرده و با محاسبه امتیاز تراکنش بر اساس مقادیر منصوب‌شده به پارامترها، احتمال متقلبانه‌بودن آن را تعیین می‌کند. این مقدار بین صفر و یک است. هر چه عدد به‌دست‌آمده نزدیک به یک باشد، مشکوک‌تر بوده و می‌تواند مؤلفه هشدار سیستم تشخیص تقلب را راه‌اندازی نماید.

سیرواستاوا، کاندو و سورال[1] (۲۰۰۸) از روش منطق فازی جهت شناسایی رفتار مشکوک در بانکداری اینترنتی استفاده کرداند. منطق فازی برای ذخیره و استفاده از تجارب خبرگان به‌صورت مجموعه قواعد بروز تقلب به‌کار می‌رود. در این پژوهش برای استفاده از قابلیت سیستم فازی در شناسایی داده‌های متقلب، نخست کلیه رفتارهای گذشته مشتریان را به‌عنوان پایگاه دانش در چند سطح مختلف دسته‌بندی و به سیستم آموزش داده سپس سیستم خبره فازی را برای استنتاج این خروجی‌ها طراحی کرده است. اگر چه این سیستم حیطهٔ وسیعی از عوامل شناسایی‌کننده رفتار و عملکرد کاربر را دربرمی‌گیرد، زمان زیادی صرف پردازش عوامل و فاکتورهای ورودی می‌کند.

۳ نظریه خطر و الگوریتم مبتنی بر آن

در سال‌های گذشته ایمنی‌شناسان برای توضیح نحوه عملکرد سیستم ایمنی بدن انسان در برابر حمله آنتی‌ژن‌ها از مدل تمایز خودی و غیرخودی استفاده می‌کردند. بر اساس این مدل سیستم ایمنی بدن انسان همواره در حال شناسایی و مقابله با سلول‌هایی است که جزئی از بدن نیستند و به‌اصطلاح خارجی (غیرخودی) هستند. پیشرفت‌های بعدی پرسش‌هایی را مطرح کردند که با این مدل در تناقض بودند. از جمله این پرسش‌ها می‌توان به مواردی که در ادامه می‌آیند اشاره کرد.

- چرا بدن انسان به غذایی که می‌خوریم و غیرخودی است، واکنش نشان نمی‌دهد؟
- دلایل وجود یک پیوند اعضا موفقیت‌آمیز در بدن، با وجود اینکه آن عضو خارجی است چیست؟
- بدن انسان در طول عمر در حال تغییر است و بنابراین سلول خودی نیز تغییر می‌کند. وقتی سلول خودی تغییر می‌کند چگونه همچنان به‌عنوان سلول خودی در نظر گرفته می‌شوند؟

[1] Srivastava, Kundu & Sural

در سال‌های بعدی ماتزینگر[1] (۱۹۹۴) نظریه‌ای مطرح کرد که بر اساس آن هر پاسخ ایمنی نتیجه احساس خطر همراه با وجود الگوی آنتی‌ژنی ناشناس است. در این فرضیه تنها شرط لازم برای آغاز واکنش ایمنی وجود سلول‌های غیرخودی نیست، بلکه این فرایند را ناشی از وجود هر دو شرط لازم یعنی سلول‌های غیرخودی و خطر می‌داند. منظور از وجود خطر در این فرضیه در ادامه ارائه شده است. به‌طورکلی، دو نوع مرگ سلولی وجود دارد که عبارت‌اند از:[2]

- مرگ طبیعی[3] یا برنامه‌ریزی‌شده: نوعی مرگ سلولی است که در شرایط فیزیولوژیک (طی فرآیندی طبیعی) اتفاق می‌افتد و به‌وسیله خود سلول در حال مرگ کنترل می‌شود.
- مرگ غیرطبیعی[4]: در این نوع مرگ، سلول‌ها بر اثر عوامل غیرطبیعی همچون حمله ویروس‌ها از بین می‌رود. در این حالت مواد شیمیایی[5] داخل سلول در محیط اطراف آن منتشر می‌شوند.

وقتی سلولی به‌صورت غیرطبیعی می‌میرد، مواد شیمیایی داخل آن در محیط اطراف آن منتشر می‌شود. ترشح این مواد سبب ارسال علامت‌های خطر شده و در محدوده کوچکی حول سلول مرده پراکنده می‌شود. این محدوده را محدوده خطر می‌نامند. در این ناحیه سلول‌های دندریتیک[6] (DC) که جزئی از سیستم ایمنی هستند، فعال شده و شروع به ازبین‌بردن آنتی‌ژن‌هایی می‌کنند که در این ناحیه هستند. به‌همین دلیل نظریه خطر را الگوریتم سلول‌های دندریتیک (DCA) نیز می‌نامند. شکل ۱ نمایی از نظریه خطر است.[7]

۴ داده‌ها و روش‌های مورد استفاده

در این بخش ابتدا به بررسی داده‌های مورداستفاده و روش انتخاب علامت‌ها و آماده‌سازی داده‌ها و توصیف ویژگی‌های آنها پرداخته می‌شود و سپس تکنیک‌های به‌کاررفته برای تحلیل داده‌ها بحث و بررسی خواهند شد.

---

[1] Matzinger

[2] Matzinger & Gallucci (2001)

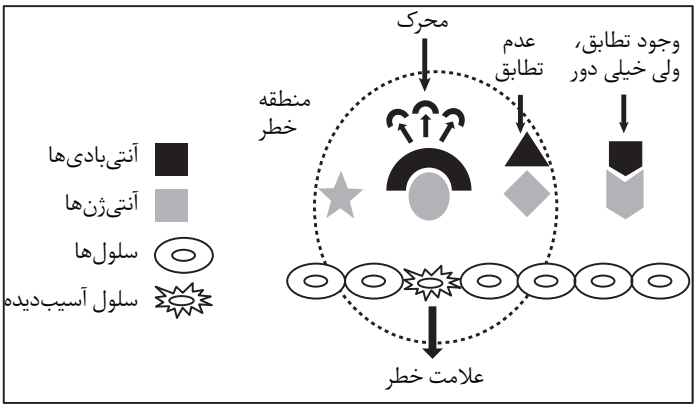
[3] apoptosis

[4] necrosis

[5] cytokine

[6] Denderitic Cell

[7] Greensmith (2007)



شکل ۱. نمایی از نظریه خطر. برگرفته از گرین‌اسمیت[1] (۲۰۰۷).

۱.۴ مجموعه داده به‌کاررفته

مجموعه داده مورداستفاده، مربوط به یک صادرکننده کارت اعتباری است و دارای ۴۱۶۴۷ رکورد با ۳/۷۴ درصد تقلب است که در بازه زمانی ۱۴ جولای ۲۰۰۴ تا ۱۲ سپتامبر ۲۰۰۴ ثبت‌شده و داده‌ها به‌عنوان تقلبی و یا نرمال برچسب‌گذاری شده‌اند. این مجموعه داده برای آزمون در مرجع استفاده شده است.[2] در جمع‌آوری داده‌ها، حجم بالایی از اطلاعات از منابع متنوعی جمع‌آوری می‌شود که معمولاً ناکامل و ناسازگارند. پیش‌پردازش داده‌ها یک مرحله کلیدی است که برای اطمینان از کیفیت داده‌های جمع‌آوری شده و سازمان‌دهی داده‌ها به‌شکل مناسب برای اعمال الگوریتم مذکور انجام می‌شود. این مرحله عبارت است از حذف داده‌های پرت و ناهنجار[3] و نرمال‌سازی داده‌ها که در ادامه هرکدام توضیح داده می‌شود.

۱.۱.۴ حذف داده‌های پرت و ناهنجار

در ابتدا می‌بایست داده‌های پرت حذف شوند. یک داده پرت به مشاهده‌ای اطلاق می‌شود که از میانگین متغیری مربوط به مجموعه داده خیلی دور است. این فاصله یا دوربودن بر حسب یک آستانه تعریف می‌شود (معمولاً دو برابر انحراف معیار).

[1] Greensmith
[2] Gadi, Wang & Lago (2008)
[3] noise

۲.۱.۴ نرمال‌سازی داده‌ها

با توجه به این موضوع که، هر یک از داده‌ها با معیارهای متفاوت اندازه‌گیری شده و بازه اندازه‌گیری مختلفی دارند، بنابراین داده‌ها نرمال‌سازی شده و مقادیر آنها به بازه صفر تا یک نگاشت می‌شوند. روشی که در این پژوهش برای این منظور استفاده شده است روش نرمال‌سازی کمینه-بیشینه[1] است. این روش نوعی نرمال‌سازی خطی است و در رابطه ۱ مشخص شده است.

$$V' = \frac{V - \min_A}{\max_A - \min_A} \qquad (۱)$$

۲.۴ الگوریتم سلول‌های دندریتیک

با الهام‌گرفتن از فرضیه نظریه خطر، الگوریتمی تحت عنوان الگوریتم سلول‌های دندریتیک طراحی شده است.[2] در ادامه به به بررسی ورودی‌ها، خروجی‌ها و همچنین نحوه عملکرد این الگوریتم خواهیم پرداخت.

۱.۲.۴ انتخاب علامت‌ها

همان‌طور که قبلاً اشاره شد، سلول‌های آسیب‌دیده از خود علامت‌هایی ساطع می‌کنند که در این الگوریتم به‌عنوان ورودی لحاظ می‌گردند و عبارت‌اند از: علامت خطر، علامت PAMP[3]، علامت ایمنی و علامت التهاب. از طرفی، با توجه به قابلیت‌های داده‌کاوی در کشف روابط بین داده‌ها، در این مقاله ما از این تکنیک‌های هوشمندانه داده‌کاوی برای مشخص‌کردن رابطه بین ویژگی‌های پایگاه داده و تقلب رخداده در داده‌ها بهره برده، علامت‌های مذکور را انتخاب کردیم.

داده‌کاوی نامی است عمومی برای دامنه گسترده‌ای از الگوریتم‌ها مانند یادگیری درخت تصمیم، روش‌های بیزین، شبکه‌های عصبی و غیره[4] که از این بین یادگیری درخت تصمیم از معدود الگوریتم‌هایی است که علاوه بر کشف روابط حاکم بر داده‌ها، توصیفی از این روابط را نیز فراهم می‌کند. در این پژوهش درخت تصمیم با استفاده از الگوریتم CART[5]

---

[1] min-max normalization

[2] Desmet, Eeckhout & Bosschere (2005)

[3] Pathogen-Associated Molecular Pattern

[4] Desmet et al. (2005)

[5] Classification and Regression Trees

به‌دست‌آمده است. به‌طور کلی، روش CART، برای ساخت یک درخت تصمیم، داده‌ها را به‌طور متناوب به زیرمجموعه‌های مشابه افراز می‌کند تا آنجا که هر زیرمجموعه دارای تعداد مشخصی نمونه شود. سپس، شاخه‌های درخت ساخته‌شده تا تحقق معیار توقف یا رسیدن به سطح پیچیدگی خواسته‌شده، هرس می‌گردند.[1] به بیان دیگر، CART ابتدا تفاوت‌های هر نمونه را با سایر نمونه‌ها می‌یابد و درخت مورد نظر را تولید می‌کند. سپس هرس‌کردن درخت از طریق یافتن تفاوت‌های مشابه انجام می‌شود. برخورداری از خصوصیاتی چون سادگی، دقت در پیش‌بینی، قابلیت دسترسی و خروجی قابل فهم موجب شده تا در پژوهش حاضر این الگوریتم انتخاب شود.

همچنین، نرم‌افزاری که در این پژوهش در مرحله انتخاب علامت‌ها با استفاده از الگوریتم CART مورداستفاده قرار گرفت، SPSS Clementine 11 است. این نرم‌افزار برای داده‌کاوی طراحی شده است و از قابلیت‌هایی جهت پیش‌پردازش، مدل‌سازی، ارزیابی و نمایش داده‌ها برخوردار است و توسط شرکت SPSS منتشر شده است. بخش مدل‌سازی Clementine که در این پژوهش مورد استفاده قرار گرفته است بسیار قوی و جامع است. در ادامه به ذکر عملکرد هر یک از علامت‌های مذکور خواهیم پرداخت.

علامت PAMP، معرف وجود خطر با احتمال قوی است. در پژوهش مذکور، علامت PAMP، به‌عنوان اولین نشانه تقلب محسوب می‌شود. بدین منظور عواملی که مهم‌ترین نشانه رفتار غیرطبیعی مشتری در زمان وقوع تقلب است به‌عنوان علامت PAMP در نظر گرفته می‌شود. در گروه علامت‌های PAMP دو مورد که نشانه وجود تقلب هستند به‌عنوان PAMP1 و PAMP2 در نظر گرفته می‌شود. در این پژوهش، اولین نشانه بروز تقلب تجاوز از مقدار عمومی برداشت مشتری (با توجه به بررسی رفتار گذشته وی) در نظر گرفته شده است که این خصیصه به‌عنوان علامت PAMP1 در نظر گرفته می‌شود. علامت PAMP2، انحراف از محدوده مکانی است که مشتری در گذشته از کارت خود استفاده می‌کرده است.

در سیستم ایمنی، علامت خطر در اثر مرگ ناگهانی و غیرطبیعی سلولی توسط بافت تولید می‌شود. این علامت در اثر خارج‌شدن مواد داخل سلول و افزایش غلظت مواد در بافت ایجاد می‌شود و باعث تغییر وضعیت سلول از حالت نابالغ به بالغ می‌شود. علامت خطر، دیگر علامت نشانه تقلب است. با این تفاوت که میزان اهمیت کمتری نسبت به علامت PAMP دارد و تنها مقادیر زیاد این علامت خطرناک است. تغییر خصیصه‌ای به نام

[1] Jing & Chiu (2001)

کد رده تجاری[1] به‌عنوان علامت Danger1 در نظرگرفته می‌شود. این خصیصه کد اجناس خریده‌شده توسط مشتری است. همچنین Danger2، به‌عنوان دومین نشانه از گروه خطر، متناظر با تغییر رفتار مشتری در زمان‌های استفاده از کارت اعتباری خود است. بنابراین، تغییر استفاده از کارت اعتباری در زمان‌هایی که در بازه زمانی گذشته مشتری نیست به‌عنوان علامت Danger2 در نظرگرفته می‌شود.

در سیستم ایمنی طبیعی، بر اثر فعالیت طبیعی بافت، علامت‌هایی آزاد می‌شود که علامت ایمنی نامیده می‌شود. در مدل، هر رفتاری از سیستم که نشان‌دهنده وضعیت طبیعی سیستم باشد به‌عنوان علامت ایمنی در نظرگرفته می‌شود. این علامت باعث کاهش اثر علامت PAMP و خطر می‌شود و در نتیجه میزان شناسایی اشتباه تراکنش‌های نرمال به‌عنوان متقلب را کاهش می‌دهد. برای این علامت یک مقدار در نظرگرفته شده است و آن نسبت تعداد تراکنش‌های موفق هر مشتری به کل تراکنش‌های همان مشتری است.

علامت التهاب، علامتی است که به‌تنهایی برای تغییر وضعیت سلول‌های دندریتیک کافی نیست، اما باعث تشدید تأثیر علامت‌های دیگر می‌شود و مقدار علامت خروجی را افزایش می‌دهد. در این پژوهش برای این علامت نیز یک مقدار در نظرگرفته شده است و آن نسبت تعداد تراکنش‌های هر مشتری به ماکزیمم تعداد تراکنش‌های همه مشتریان است. همچنین سلول‌های دندریتیک به هنگام مواجه‌شدن با آنتی‌ژن‌ها در سه حالت قرار می‌گیرند که این سه حالت نیز به‌عنوان بخشی از خروجی این الگوریتم به حساب می‌آیند و عبارت‌اند از: نابالغ، نیمه بالغ و بالغ.

۲.۲.۴ آنتی‌ژن

منظور از آنتی‌ژن در این الگوریتم داده‌های مورد نظر برای طبقه‌بندی است. در این پژوهش هر سطر از مجموعه داده، به‌عنوان یک آنتی‌ژن در نظرگرفته شده است.

۳.۲.۴ نحوه اعمال الگوریتم سلول‌های دندریتیک

پس از مشخص‌کردن علامت‌های ورودی این الگوریتم می‌بایست به پردازش این علامت‌ها و آنتی‌ژن‌های نمونه‌برداری شده توسط این الگوریتم بپردازیم. شبه‌کد و نحوه عملکرد این الگوریتم در ادامه در شکل ۲ ارائه شده است.

همان‌طور که در شکل ۲ مشهود است، در این الگوریتم ابتدا یک DC مقداردهی اولیه شده و زمان نمونه‌برداری آنتی‌ژن و علامت‌ها مشخص می‌شود. پس از آن، چنانچه میزان علامت

[1] merchant category code

خروجی CSM کمتر از حد آستانه[1] باشد، DC می‌تواند به نمونه‌برداری و ذخیره آنتی‌ژن بپردازد و سپس، علامت‌های خروجی را محاسبه نماید. خط ۸ مشخص‌کننده اتمام بازه نمونه‌برداری آنتی‌ژن و علامت‌ها و شروع بررسی وضعیت DC است. بر اساس مقایسه انجام گرفته در خطوط ۱۰ تا ۱۴، مقدار علامت‌های خروجی M و SM مقایسه شده و بر اساس این مقایسه محتوای آنتی‌ژن (تراکنش‌ها در این پژوهش) به‌عنوان نرمال یا ناهنجار مشخص می‌شود. در نهایت، DC فعلی حذف شده و DC جدیدی جایگزین آن خواهد شد. محاسبه علامت‌های خروجی (CSM, SM, M) از طریق رابطه ۲ انجام می‌گیرد.

```
ورودی‌ها: علامت‌هایی از تمام گروه‌ها و آنتی‌ژن
خروجی‌ها: آنتی‌ژن و ارزش‌های بافت
۱   DC را شروع‌کن؛
۲   تازمانی‌که سیگنال خروجی CSM<آستانه مهاجرت است، انجام‌ده
۳       آنتی‌ژن را دریافت کن؛
۴       آنتی‌ژن را ذخیره کن؛
۵       علامت‌ها را دریافت کن؛
۶       علامت‌های خروجی موقت را محاسبه کن؛
۷       علامت‌های خروجی تجمعی را به‌روزرسانی کن؛
۸   پایان
۹   موقعیت مکانی سلول را به گره لنفاوی به روز کن؛
۱۰  اگر خروجی نیمه‌بالغ < خروجی بالغ، آنگاه
۱۱      ارزش ۰ را به بافت سلول اختصاص ده؛
۱۲  درغیراین‌صورت
۱۳      ارزش ۱ را به بافت سلول اختصاص ده؛
۱۴  پایان
۱۵  سلول را نابود کن؛
۱۶  سلول را در جامعه جایگزین کن؛
۱۷  پایان
۱۸  ارزش MCAV را محاسبه کن و بافت را برای هر نوع آنتی‌ژن ثبت کن؛
```

شکل ۲. شبه‌کد الگوریتم سلول‌های دندریتیک. برگرفته از گرین‌اسمیت و ایکلین[2] (۲۰۰۷)

[1] migration threshold

[2] Greensmith & Aickelin

در این رابطه p_i، d_i و s_i به ترتیب مقادیر علامت‌های ورودی PAMP، خطر و ایمنی هستند. i تنها نشان‌دهنده وجود چندمین علامت در یک گروه است و از نظر درجه اهمیت تفاوتی در علامت‌های یک گروه ایجاد نمی‌کند. p_w وزن مرتبط با علامت PAMP، d_w وزن مرتبط با علامت خطر و s_w وزن مرتبط با علامت ایمنی است. همچنین در این رابطه I نشان‌دهنده علامت التهاب است. این رابطه برای هر علامت خروجی (CSM، SM و M) یک‌بار محاسبه می‌شود. وزن‌های مذکور برای اعمال الگوریتم سلول‌های دندریتیک جهت تشخیص تقلب کارت‌های اعتباری با استفاده از روش‌های تجربی (آزمون صحیح و خطا) در جدول ۱ ارائه شده است. لازم به ذکر است که این مقادیر می‌تواند بر اساس کاربرد مورد نظر تغییر یافته تا نتایج بهتری حاصل شود.

جدول ۱
مقادیر وزن‌ها

علامت‌ها	PAMP	خطر	ایمنی
CSM	۱	۰/۵	۱/۵
SM	۰/۰۰	۰/۰۰	۱
M	-۱/۵	۰/۷۵	۱/۵

برای تشخیص تقلب، طبقه‌بندی آنتی‌ژن بر اساس ترکیب‌کردن تحلیل جمعیت DC در دوره آزمون صورت می‌گیرد. آنتی‌ژن‌ها در DCهای مختلف شرکت می‌نمایند، بنابراین پس از محاسبه مقدار صفر یا یک برای هر DC و نسبت‌دادن آن به همه آنتی‌ژن‌های موجود در آن DC، نیاز به محاسبه متغیری تحت عنوان MCAV[1] برای هر آنتی‌ژن داریم. جهت محاسبه این متغیر می‌بایست نسبت تعداد آنتی‌ژن‌های آن نوع خاص که منجر به ناهنجاری شده‌اند به کل تعداد آنتی‌ژن‌های همان نوع، را به دست آورد. هر چقدر که این مقدار به یک نزدیک‌تر باشد آنتی‌ژن مذکور ناهنجار بوده و بر عکس. البته برای محاسبه دقیق‌تر، آستانه ناهنجاری در نظر گرفته می‌شود و در صورتی که مقدار MCAV از حد آستانه بیشتر باشد به‌عنوان ناهنجار در نظرگرفته می‌شود و در غیر این صورت، به‌عنوان نرمال بازه ۰/۴ تا ۱/۱ از طریق محاسبات تجربی به‌عنوان بهترین بازه برای آستانه ناهنجاری محاسبه شده است.

[1] Mature Context Antigen Value

جدول ۲

مقادیر مناسب جهت اعمال الگوریتم DCA برای تشخیص تقلب در کارت‌های اعتباری

پارامترها	مقادیر مناسب
تعداد DC در هر بار اجرای الگوریتم	۱۰۰
تعداد آنتی‌ژن نمونه‌برداری‌شده	۱
بازه مناسب حد آستانه	۰/۴ تا ۱/۱

همچنین، برای اعمال الگوریتم سلول‌های دندریتیک نیازمند مشخص‌کردن پارامترهای دیگر از قبیل تعداد DCهای مورداستفاده در هر بار اجرای الگوریتم، تعداد آنتی‌ژن نمونه‌برداری‌شده توسط DC و بازه مناسب آستانه مجاز هستیم. جدول ۲ مقادیر مناسب هر یک از پارامترهای مذکور را ارائه می‌کند. ارزیابی پارامترهای مذکور بر اساس چهار پارامتر و روابط بین آنها انجام گردیده که در ادامه به بررسی هر یک خواهیم پرداخت.[1]

- هشدار منفی نادرست[2]: یک تراکنش متقلبانه، مشروع تشخیص داده می‌شود.
- هشدار مثبت نادرست[3] : یک تراکنش مشروع، متقلبانه تشخیص داده می‌شود.
- هشدار منفی درست[4]: یک تراکنش مشروع، مشروع تشخیص داده می‌شود.
- هشدار مثبت درست[5]: یک تراکنش متقلبانه، متقلب تشخیص داده می‌شود.

میزان شناسایی درست داده‌های تقلب و نرمال از طریق نسبتی تحت عنوان دقت الگوریتم ارزیابی می‌شود. نحوه محاسبه این نسبت در رابطه ۳ مشخص شده است.

(۳) کل هشدارها/ هشدارهای درست=نرخ شناسایی

۵ مدل الگوریتم سلول‌های دندریتیک

همان‌طور که اشاره شد، هدف اصلی ما در این پژوهش تشخیص تقلب با استفاده از الگوریتم مبتنی بر نظریه خطر تحت عنوان الگوریتم سلول‌های دندریتیک است. شکل ۳ کلیه مراحل طی شده جهت حصول نتایج را ارائه می‌دهد.

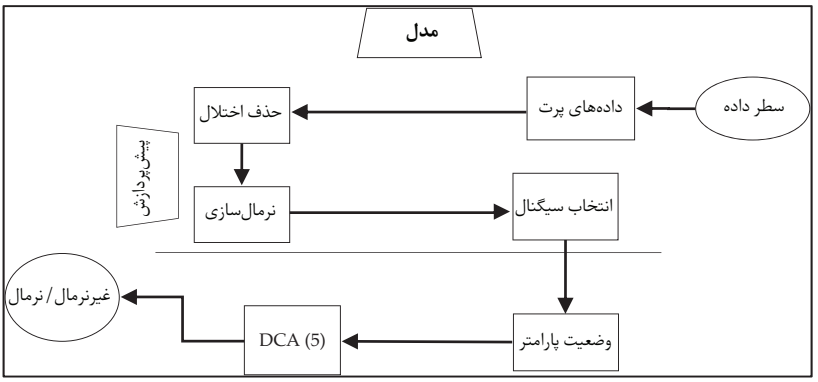
[1] Assumpta & Christian (2002)

[2] false negative

[3] false positive

[4] true negative

[5] true positive



شکل ۳. نمایی از کلیه مراحل انجام‌شده

ابتدا در مرحله یک مجموعه داده مورد آزمون به سیستم ارائه شده و در مرحله بعدی عملیات پیش‌پردازش داده (حذف داده پرت، حذف نویز و نرمال‌سازی داده‌ها) روی آن انجام خواهد گرفت. سپس، در مرحله سوم با استفاده از تکنیک‌های پیشرفته هوش مصنوعی و داده‌کاوی و به‌طور خاص الگوریتم طبقه‌بندی CART به انتخاب علامت‌های مناسب می‌پردازیم. در مرحله چهارم، مقدار مجاز برای پارامترهای این الگوریتم از جمله آستانه مهاجرت، تعداد DC در هر بار اجرای این الگوریتم و تعداد آنتی‌ژن نمونه‌برداری‌شده تعیین می‌شود. در مرحله بعدی نیز الگوریتم سلول‌های دندریتیک روی مجموعه داده مذکور اعمال می‌شود. درنهایت، جهت ارزیابی نتایج به‌دست‌آمده در مرحله قبل، اقدام به مقایسه نتایج الگوریتم سلول‌های دندریتیک با برچسب‌های وضعیت تراکنش‌ها می‌نماییم. شکل ۲ نمایی کلی از مراحل مذکور را ارائه می‌دهد.

پس از مشخص‌کردن مقادیر مجاز برای پارامترها و انتخاب علامت‌های مناسب با استفاده از الگوریتم هوشمندانه CART، الگوریتم سلول‌های دندریتیک بر روی مجموعه داده مذکور اعمال شده و نتایج حاصل از اعمال این الگوریتم در جدول ۳ ارائه شده است.

هدف اصلی سیستم‌های تشخیص تقلب ماکزیمم‌کردن پیش‌بینی‌های درست و نگه‌داشتن پیش‌بینی‌های نادرست در یک سطح قابل قبول است. درحقیقت، سیستمی مطلوب است که تعداد کمینه هشدار مثبت و منفی نادرست و تعداد بیشینه هشدار مثبت و منفی درست و البته، درصد موفقیت بالایی داشته باشد. بنابراین، همان‌طور که در جدول ۳ مشاهده می‌شود، الگوریتم سلول‌های دندریتیک، الگوریتمی مناسب جهت تشخیص تقلب در تراکنش‌های کارت‌های اعتباری است. این الگوریتم به‌دلیل اعمال علامت امن که باعث

کاهش اثر علامت PAMP شده و تعداد شناسایی اشتباه تراکنش‌های نرمال به‌عنوان متقلب را کاهش می‌دهد میزان هشدار مثبت نادرست پایینی دارد. این الگوریتم همچنین به‌دلیل نداشتن فاز آموزش[1] برای پایگاه داده‌های آنلاین[2] (پرکاربرد در تشخیص تقلب آنلاین) نیز بسیار مناسب است. به علاوه، به‌علت دارابودن ساختار ساده دارای سرعت پردازش بالاست. اگرچه این الگوریتم به‌علت ماهیت نامتوازن بودن پایگاه داده‌های تقلب[3] میزان هشدار منفی نادرست نسبتاً بالایی را تولید می‌کند، دارای حد شناسایی قابل‌قبولی است.

جدول ۳

نتایج حاصل از اعمال الگوریتم سلول‌های دندریتیک بر روی مجموعه داده (درصد)

پارامترها / الگوریتم	هشدار مثبت نادرست	هشدار مثبت درست	هشدار منفی نادرست	هشدار منفی درست	نرخ شناسایی
DCA	۷	۹۱	۱۳	۸۹	۹۰

۶ نتیجه‌گیری

هدف اصلی در این مقاله تشخیص تقلب در تراکنش‌های کارت‌های اعتباری با استفاده از الگوریتم مبتنی بر نظریه خطر تحت عنوان الگوریتم سلول‌های دندریتیک روی مجموعه داده حقیقی است. این نتایج سبب تعیین راهبردهای بانکداری در جهت جلب رضایت مشتری می‌شود.

با توجه به نتایج به‌دست‌آمده می‌توان این‌گونه بیان کرد که الگوریتم سلول‌های دندریتیک، الگوریتمی مناسب با دقت مناسب (۹۰ درصد) جهت تشخیص تقلب در تراکنش‌های کارت‌های اعتباری است. این الگوریتم به‌دلیل اعمال علامت امن که باعث کاهش اثر علامت PAMP و خطر شده و تعداد شناسایی اشتباه تراکنش‌های نرمال به‌عنوان متقلب را کاهش می‌دهد، میزان هشدار مثبت نادرست پایینی دارد. به‌دلیل نداشتن فاز آموزش برای پایگاه داده‌های آنلاین (پرکاربرد در تشخیص تقلب آنلاین) نیز بسیار مناسب است. به‌علاوه، به علت داشتن ساختار ساده دارای سرعت پردازش بالا است.

[1] training phase

[2] data stream

[3] imbalance data

اشاره به این نکته نیز حائز اهمیت است که باوجود تلاش‌های انجام‌شده جهت کاهش میزان هشدار منفی نادرست الگوریتم سلول‌های دندریتیک به علت ماهیت نامتوازن بودن پایگاه داده‌های تقلب میزان هشدار منفی نادرست نسبتاً بالایی را تولید می‌کند؛ بنابراین برای پژوهش‌های آینده در نظر داریم الگوریتم مذکور را با الگوریتم‌های دودسته‌ای ادغام نماییم. همچنین، از راهکارهای آینده می‌توان به مشخص‌کردن پارامترهای موردنیاز الگوریتم با استفاده از تکنیک‌های هوشمندانه اشاره کرد.

فهرست منابع

ساروخانی، ل.، و منتظر، غ. ع. (۱۳۸۷). طراحی و پیاده‌سازی سیستم هوشمند شناسایی رفتار مشکوک در بانکداری اینترنتی، *فصلنامه علمی پژوهشی فناوری اطلاعات و ارتباطات ایران*، ۱(۱-۲)، ۹-۱۸.

Assumpta, M., & Christian, N. (2002). Determination of e-quality. *Journal of Quality and Reliability Managemen*t, 19(3), 246-258.

Corchado, J. M., & Alonso, L. (2002). Artificial immune systems: A novel paradigm to pattern recognition. In J. M. Corchado, L. Alonso, & C. Fyfe (Eds.) *Artificial Neural Networks in Pattern Recognition (*pp. 67-84*)*. University of Paisley, UK.

Desmet, V., Eeckhout, L., & Bosschere, K. D. (2005). Using decision trees to improve program-based and profile-based static branch prediction. *Lecture Nots in Computer Science*, 3740, 336-352.

Duman, E., & Hamdi, M. (2011). Detecting credit card fraud by genetic algorithm and scatter search. *Journal of Expert Systems with Applications*, 38, 13057-13063.

Gadi, M., Wang, X., & Lago, A. (2008). Credit card fraud detection with artificial immune system. *Artificial Immune Systems Lecture Notes in Computer Science*, 5132, 119-131.

Greensmith, J. (2007). *The Dendritic Cell Algorithm* (PhD thesis). School of Computer Science, University of Nottingham.

Jing, K. L., & Chiu, C. C. (2001). Mining the customer credit by using the neural network model with classification and regression tree approach. *IEEE Joint 9th IFSA World Congress and 20th NAFIPS International Conference*, 2, 923 - 928.

Krenker, A., Volk, M., Sedlar, U., Bester, J., & Kosh, A. (2009). Bidirectional artificial neural networks for mobile-phone fraud detection, *Journal of Artificial Neural Networks*, 31(1), 92-98.

Kundu, A., Panigrahi, S., Sural, S., & Majumdar, A. K. (2009). BLAST-SSAHA hybridization for credit card fraud detection, *IEEE Transactions on Dependable and Secure Computing*, 6(4), 309-315.

Leung, A., Yan, Z., & Fong, S. (2004). *On designing a flexible e-payment system with fraud detection capability*. IEEE Conference on E-Commerce Technology, 127-132.

Matzinger, P., & Gallucci, S. (2001). Danger signals: SOS to the immune system, *Journal of Current Opinion in Immunology*, 13, 114-119.

Matzinger, P. (1994). Tolerance, danger, and the extended family, *Annual Review of immunology*, 12, 991-1045.

Osuna, E., Freund, R., & Girosit, F. (1997). *Training support vector machines: an application to face detection*. International Conference on Computer Vision and Pattern Recognition, 130-136.

Ravisankar, P., Ravi, V., Raghava, R. G., & Bose, I. (2011). Detection of financial statement fraud and feature selection using data mining techniques, *International Journal of Decision Support Systems*, 50, 491-502.

Srivastava, A., Kundu, A., & Sural, Sh. (2008). Credit card fraud detection using hidden Markov model, *International Journal of IEEE Transaction on Dependable and Secure Computing*, 5(1), 67-77.

پیوست

سیستم ایمنی بدن انسان

سیستم ایمنی از سلول‌ها، مولکول‌ها و قوانینی تشکیل شده است که از آسیب‌رساندن عواملی همچون آنتی‌ژن[1] به بدن میزبان جلوگیری می‌کند.[2] سیستم ایمنی به دو دسته اصلی ایمنی ذاتی و ایمنی اکتسابی تقسیم می‌شود. ایمنی ذاتی نوعی دفاع اولیه بدن نسبت به

[1] هر مولکول که توسط بدن به‌عنوان ذره خارجی شناخته شود، آنتی‌ژن (antigen) نامیده می‌شود.

[2] Corchado & Alonso, (2002)

مهاجمان محسوب شده و قادر به تغییرپذیری و یادگیری نیست. در حالی که ایمنی اکتسابی نوعی سد دفاعی قوی‌تر در برابر عوامل بیماری‌زاست. ویژگی‌های مشخص ایمنی اکتسابی، اختصاصی‌بودن برای هر ماکرومولکول و توانایی به‌خاطرسپردن آن و ایجاد پاسخ‌های قوی در برخوردهای مکرر است. قسمتی از کار سیستم ایمنی اکتسابی رهاسازی لنفوسیت‌ها در خون است. لنفوسیت‌ها به دو نوع تقسیم‌بندی می‌شوند: نوع B و T. این سلول‌ها در محیط‌های محافظت‌شده‌ای از قبیل مغز استخوان و تیموس ایجادشده و از طریق جریان خون در کل بدن منتشر می‌شوند و در این حین از طریق ترشح پادتن[1] (ماده‌ای از جنس پروتئین) اقدام به ازبین‌بردن آنتی‌ژن‌های مضر می‌کنند.

[1] antibody