تشخیص موثر تقلب ها و کلاهبرداری های بانکی پیشرفته آنلاین در داده های به شدت نامتوازن

چکیده :

تقلب در بانکداری آنلاین نشان دهنده منابع یکپارچه اجتماعی ، سایبری و دنیای فیزیکی است. این تشخیص نوعی استفاده از اینترنت و موارد گسترده با روش (W2T) است. با این حال، اطلاعات بسیار محدودی برای تشخیص تقلب پویا از رفتار مشتری واقعی در چنین محیط اطلاعاتی بسیار پراکنده و نامتوازن در دسترس می باشد، که باعث می شود تشخیص فوری و موثر بیشتر مهم و چالش برانگیز شود. در این مقاله، ما یک چارچوب آنلاین تشخیص تقلب بانکی موثر داریم که از ترکیب منابع مربوطه و شامل چندین تکنیک پیشرفته داده کاوی است. با ساخت یک بردار برای هر معامله بر اساس توالی رفتار تاریخی مشتری ، ما نرخ افتراق هر معامله موجود در برابر مشخصات رفتار مشتری را بدست می آوریم . یک الگوریتم ، کانترست ماینبر برای کاوش موثر الگوهای کانترست و تفکیک رفتار های جعلی از اصیلی معرفی شده است ، به دنبال انتخاب الگوی موثر که ترکیبی از پیش بینی مدل های مختلف و خطر است. نتایج حاصل از آزمایشات واقعی داده های بانکی آنلاین در مقیاس بزرگ نشان می دهد که سیستم ما می تواند به دقت بالاتر و حجم هشدار پایین تر از سیستم تشخیص تقلب ، معیار ترکیب دانش تخصصی و روش های تشخیص تقلب سنتی دست یابد.

کلمات کلیدی : تشخیص تقلب ، بانکداری آنلاین ، الگوی مقابل ، شبکه عصبی ، داده کاوی

1. مقدمه

با استفاده گسترده از تکنولوژی اینترنت به طور فزاینده [15، 47، 60]، بانکداری آنلاین (اگرچه بانکداری اینترنتی نامیده می شود) به عنوان یک کانال اصلی برای خرده فروشی و تجارت در حال ظهور است. در مقابل، فعالیت های بانکداری آنلاین جعلی بیشتر و بیشتر پیچیده می شوند، و به طور جدی امنیت و اعتماد در تجارت با بانکداری آنلاین تهدید شده است. تقلب در بانکداری آنلاین یک مسئله جدی در مدیریت جرائم مالی برای همه بانک ها تبدیل شده

است. این اتفاق سبب به چالش کشیدن و منجر به ضرر و زیان های عظیم، که ناشی از ظهور و تکامل تقلب در بانکداری آنلاین ، از قبیل کلاهبرداری فیشینگ، الودگی بد افزار و وب سایت های شبح است. تشخیص موثر و کارآمد تقلب بانکی اینترنتی به عنوان یک چالش عمده برای تمام بانک ها در نظر گرفته شده ، و از علل افزایش نگرانی است. یک سیستم تشخیش کلاهبرداری انلاین بانکی می تواند شامل استفاده از روش اینترنت (W2T) و چیز های گسترده باشد [66-63]. به این علت موقع جمع آوری داده ها چند جنبه از مشتریان بانکداری آنلاین خواسته می شود ، از جمله اطلاعات دموگرافیک، داده های معامله بانکی آنلاین ، اطلاعات تراکنش کارت اعتباری و انواع دیگر از داده های معامله . این داده ها از طریق اینترنت WWW / و SEA به شبکه بانکداری آنلاین و مرکز داده های مشتری منتقل می شود. مرکز داده یک پلت فرم برای کل فرایند تشخیص تقلب در بانکداری آنلاین را فراهم می کند. این یک چرخه اطلاعات کامل از اطلاعاتی ناهمگن ، و دانش در دنیای فیزیکی برای ارائه خدمات فعال در جهان مجازی به مشتریان اجتماعی است. مشتریان بانکداری آنلاین (در جهان اجتماعی)، همه چیز (در جهان فیزیکی)، و سیستم های کامپیوتری (در دنیای سایبر) یک نهاد یکپارچه برای تحقق بخشیدن به هماهنگی و همزیستی با استفاده از یک چرخه داده W2Tهستند. در این چرخه، فرآیند تشخیص کلاهبرداری یک وظیفه مهم است.

تقلب در بانکداری اینترنتی نمایانگر ویژگی های پیچیده خاص است (به بحث های مفصل در بخش 2.1 نگاه کنید) :

- مشتریان مشکوک فعال و هوشمند در انجام فعالیت های بانکی جعلی ،
- رفتار جعلی بسیار پویا ،
- تقلب پنهان و متنوع در رفتار مشتری ، مربوط به تقلب در معاملات بسیاری از مجموعه داده های بزرگ پراکنده و نامتوازن، و
- وقوع تقلب در یک زمان بسیار محدود که نیاز به تشخیص زمان واقعی دارند.

تقلب در بانکداری آنلاین نیاز به تشخیص فوری دارد ، دلیل آن بازیابی موارد از دست رفته در صورت انجام کلاهبرداری است. اکثر مشتریان معمولا به ندرت تاریخ بانکی آنلاین خود را به طور منظم چک می کنند و بنابراین قادر به کشف

و گزارش معاملات جعلی بلافاصله پس از وقوع یک کلاهبرداری نیستند. این باعث می شود امکان بازیابی موارد از دست رفته بسیار کم شود. علاوه بر این، تمام هشدارهای تولید شده از سیستم تشخیص باید به صورت دستی بررسی شود، که بسیار وقت گیر است. سیستم های تشخیص بانکداری آنلاین دارای دقت بالا، نرخ تشخیص بالا و کنترل تعداد هشدار در تجارت پیچیده بانکداری آنلاین هستند.

ویژگی های بالا و الزامات تجارت تا حد زیادی روش های تشخیص تقلب و مدل های داده کاوی موجود را برای حفاظت از معاملات کارت اعتباری، تجارت الکترونیک، بیمه، خرده فروشی، ارتباطات راه دور، کامپیوتر، و غیره به چالش می کشد ، این روش ها عملکرد ضعیف در بهره وری و / یا دقت در زمانی که برای تشخیص تقلب در بانکداری آنلاین استفاده می شود را نشان می دهد [35]. به عنوان مثال، کارت اعتباری و یا تشخیص تقلب مخابراتی اغلب در کشف الگوهای رفتاری خاص از یک مشتری خاص و یا گروه رخ می دهد، اما معاملات بانکی آنلاین مربوط به سرقت تمرکز بسیار پویا است و بسیار شبیه به رفتار مشتری واقعی به نظر می رسد. برخی از روش های تشخیص نفوذ در یک محیط کامپیوتر پویا به خوبی انجام شده، اما آن ها نیاز به مقدار زیادی از داده های آموزشی با سیاهه های مربوط به عنوان شواهد دارند. با این حال، هیچ مدرکی وجود ندارد که آشکار نشان دهد که یک تراکنش بانکی آنلاین جعلی است.

یک مورد امیدوار کننده به تازگی پدید آمده است که به دقت تفاوت بین رفتار های جعلی و واقعی را بررسی می کند، و روش های مربوطه و الگوهای مقابل را توسعه می دهد، به عنوان مثال، در مقابل مجموعه [6] و الگوهای در حال ظهور [24، 25، 52]. با این حال، آزمایش از روش های کلاسیک بر روی داده های بانکداری آنلاین واقعی نشان داده است که دقت و صحت آنها به دلیل چالش در تشخیص تقلب بانکداری آنلاین بسیار بالا است. علاوه بر این، با توجه به پژوهش [61]، الگوی کنتراست به معنی یک مشکل سخت NP است، زمان زیادی هنگامی که تعداد ویژگی ها آن زیاد است صرف آن می شود ، و آستانه سرعت کشف و شناسایی آن کوچک است. بر اساس آزمایش های ما، روش الگوی کنتراست [24] به طور موثر طرح بانکداری آنلاین را انجام نمی دهد.

تنها چند مقاله در مورد کنترل تقلب در بانکداری آنلاین [35، 37، 44] وجود دارد. جریان اصلی سیستم های تشخیص تقلب بانکی آنلاین برای ایجاد قوانین برای فیلتر معاملات مشکوک که مشکلات مهمی ایجاد می کنند، از جمله نرخ

مثبت کاذب بسیار بالا و نرخ تشخیص کم ، به کارشناسان تکیه می کنند. از همه مهمتر، اقتباس قوانین به پویایی تقلب و به تخصص دامنه به طور کامل وابسته هستند. کیفیت تشخیص تقلب بدون کنترل پایدار بسیار وقت گیر است ، و از طریق پیام رسان از قوانین پشتیبانی نمی کند.

بسیاری از حوادث و رفتارهای قبلی در زمان های مختلف مستقل بوده و اطلاعات ثبت شده در توالی رویداد را نادیده می گیرد. در بانکداری آنلاین، توالی فعالیت برای افتراق رفتار جعلی از رفتار واقعی مفید هستند. به عنوان مثال در جدول 1 و 2 نشان داده شده است. جدول 1 دنباله دسترسی به صفحه وب توسط یک تروجان است، در حالی که جدول 2 یک معامله واقعی از طریق یک مرورگر وب است. دو ویژگی متضاد بین این دو توالی وجود دارد. یکی این که تقلب کنار برخی از صفحات وب مانند homepage.aspx بعد از ورود و صفحه چاپ پس از تایید انتقال ناچیز برای ارسال معامله می باشد. دیگر این که معامله در کمتر از 3 ثانیه پس از ورود به سیستم به پایان رسیده ، که بیش از حد برای یک کاربر معمول بانکداری آنلاین برای دستیابی از طریق یک مرورگر وب سریع است.

Time	PageLink
21:55:42.190	Login.aspx
21:55:43.260	BalanceCheck.aspx
21:55:43.890	PayForm.aspx
21:55:44.121	PayConfirm.aspx
21:55:45.091	HomePage.aspx

جدول1. توالی رفتار تقلبی.

Time	PageLink
21:58:06.190	Login.aspx
21:58:07.391	HomePage.aspx
21:58:15.260	BalanceCheck.aspx
21:58:27.890	PayForm.aspx
21:59:22.121	PayConfirm.aspx
21:59:27.091	Print.aspx
21:59:32.091	HomePage.aspx

جدول 2. توالی رفتار اصلی.

با استفاده از داده ها و ویژگی های تجارت فوق، این مقاله یک چارچوب موثر در شناسایی پیچیده تقلب بانکی اینترنتی پیشنهاد می کند. ایده اصلی، مزایا و کمک های حاصل از این چارچوب به شرح زیر است :

- این است که توسط تئوری مهندسی meta-synthetic [11] ، M-Computing [12] و حکمت وب از چیزهایی که [66] الهام گرفته، یک راه حل سیستماتیک با ترکیب دانش دامنه، تجربه در سیستم تشخیص مبتنی بر قواعد ، مزایای استفاده از مدل های مختلف، و پالایش توسط کارشناسان فراهم می کند.

- این تعبیه ماژول سیستماتیک با انتخاب ویژگی ها بر اساس کسب اطلاعات، استخراج رفتار مقابل طبقه بندی ساختمان ، تولید نمره خطر کلی برای هر معامله در بانکداری آنلاین، و شناسایی الگوهای رفتاری جعلی را انجام می دهد. سیستم تشخیص تقلب بانکی آنلاین با هر سیستم بانکداری آنلاین و یا خدمات ارتباط برقرار نمی کند.

- ما نه تنها دنبال این اطلاعات رفتاری برای شناسایی الگوهای مقابل هستم، بلکه این روش جدید، برای ادغام رفتار های متوالی پایگاه داده برای استخراج الگوهای مقابل موثر است.

- این سیستم شامل ادغام چندین مدل داده کاوی ، هزینه شبکه های عصبی حساس [67]، الگوی کنتراست ، و جنگل تصمیم است. از آنجا که مدل های مختلف کشف تقلب و الگوهای رفتاری واقعی از زوایای مختلف مورد برسی قرار می گیرد ، ترکیب آنها [13] الگوهای رفتاری جامع تر ارائه می دهد.

- هر مدل را می تواند به راحتی در طول زمان از تغییرات در رفتار جعلی حفظ کنید.

- آزمایش های عظیم در یک بانک در استرالیا که سیستم و مدل های تشخیص نرخ بالاتر و سرعت کاذب پایین تر دارند از هر مدل داده کاوی کلاسیک سیستم مبتنی بر قواعد موجود مورد استفاده در تمام بانک های بزرگ استرالیا بهتر عملکرده است. علاوه بر این، سیستم ما عملکرد تشخیصی نسبتا خوب بر روی مجموعه داده بسیار نامتوازن و مدل الگوی کنتراست ایجاد می کند و به اصلاح بر روی داده های زمان واقعی کارآمد است. همچنین الگوهای رفتاری توالی کشف ، و اطلاعات بیشتری در مورد شواهد پزشکی قانونی برای تشخیص تقلب فراهم می کند.

ادامه مقاله به شرح زیر تدوین شده است. بخش 2 ویژگی های تقلب بانکداری آنلاین در جزئیات و ارائه یک نمای کلی از کار مرتبط در کشف تقلب. بخش 3 بیانیه مشکل و تعریف اصطلاحات را ارائه می دهد ، در حالی که بخش 4 دقت

و چارچوب تشخیص تقلب بانکی آنلاین در جزئیات را توضیح می دهد. روش مقابل استخراج الگوی با بردار مقابل در بخش 5 معرفی و روش امتیازدهی خطر بر اساس مدل های ترکیب شده در بخش 6 است. ارزیابی آزمایش ارائه شده در بخش 7 مورد بحث است و 8 نتایج و جهت تحقیقات آینده را نشان می دهد.

2. ویژگی های تقلب بانکی آنلاین و کارهای مرتبط

در این بخش، ما برای اولین بار به طور خلاصه ویژگی های اصلی تقلب در بانکداری آنلاین، و سپس به بحث های مربوط به کار در مناطق مختلف از کشف تقلب می پردازیم. کار منتشر شده در مورد کشف تقلب به حوزه تقلب در کارت اعتباری، نفوذ کامپیوتر و تقلب مخابراتی مربوط می شود. بنابراین ما هر یک از این بحث ها و توضیح محدودیت کار های موجود در هنگام اعمال برای کشف تقلب بانکداری آنلاین را انجام می دهیم.

1.2. ویژگی های تقلب در بانکداری آنلاین

از نقطه نظر سیستم، جوهر از کلاهبرداری های آنلاین نشان دهنده دهید مصنوعی از تعامل بین منابع در سه جهان : سوء استفاده از هوش متقلبانه در جهان اجتماعی، سوء استفاده از فن آوری وب و منابع بانکداری اینترنتی در دنیای سایبری، و سوء استفاده از ابزار تجارت و منابع در جهان فیزیکی. این یک نمونه از مشکلات در اینترنت (W2T) است. یک تحقیق در ویژگی های آن برای توسعه راه حل موثر است، که پس از آن برای حل دیگر مسائل در W2T مفید خواهد بود.

تحقیقات ما در یکی از بزرگترین بانک ها در استرالیا نشان می دهد که در دنیای واقعی بانکداری آنلاین مجموعه داده های معامله و بیشتر تقلب بانکداری آنلاین دارای ویژگی ها و چالش های زیر است: (1) مجموعه بزرگی از داده های بسیار نامتوازن ؛ (2) تشخیص زمان واقعی؛ (3) رفتار تقلبی پویا. (4) شواهد فورنزیک ضعیف ؛ و (5) الگوهای رفتاری متنوع واقعی.

(1) مجموعه بزرگی از داده های بسیار نامتوازن. توجه به مطالعه و انجام آن بر روی داده های بانکداری آنلاین یک بانک در استرالیا، موجب تشخیص تقلب بانکی آنلاین شامل تعداد زیادی از معاملات، معمولا میلیونی شد. با این حال، تعدادی از تقلب ها روزانه است که معمولا بسیار کوچک هستند. به عنوان مثال، تنها 5 تقلب در میان بیش از

300،000 معامله در یک روز. این نتایج در وظیفه تشخیص تقلب در میان تعداد گسترده ای از معاملات واقعی. بسیار نادر و پراکنده است.

(2) تشخیص زمان واقعی تقلب. در بانکداری آنلاین، فاصله زمانی بین پرداخت یک مشتری و واریز به حساب مقصد معمولا بسیار کوتاه است. برای جلوگیری از ، از دست دادن فوری پول ، یک هشدار تشخیص تقلب باید با بیشترین سرعت ممکن ایجاد شود. که نیاز به یک سطح بالا از بهره وری در کشف تقلب در داده های بزرگ و نامتوازن دارد.

(3) رفتار تقلبی پویا . کلاهبرداران به طور مستمر تکنیک های خود را برای شکست بانکداری آنلاین تغییر می دهند است. نرم افزارهای مخرب، که برای بخش بیشتری از تقلب بانکداری آنلاین گزارش شده است ، که بیش از 55،000 برنامه ی مخرب جدید روزانه ساخته می شود [5]. این نهاد تشخیص تقلب نیاز به دفاع در برابر یک مجموعه در حال رشد حملات دارد. این فراتر از توانایی هر یک از مدل های تشخیص تقلب واحد است، و نیاز به قابلیت تطبیقی از مدل ها و امکان درگیر شدن چند مدل [13] برای اعمال نفوذ به چالش هایی که نمی تواند توسط هر مدل انجام شود.

(4) شواهد فورنزیک ضعیف برای تشخیص تقلب. برای معاملات بانکی آنلاین، تنها دانستن حساب های منبع، حساب های مقصد و ارزش دلار در ارتباط با هر معامله ممکن است، اما سایر اطلاعات خارجی، به عنوان مثال، هدف از هزینه، در دسترس نیست. علاوه بر این، به استثنای سرقت ID ، تقلب بانکداری آنلاین ربودن یک سیستم بانکداری آنلاین است اما با حمله به رایانه های مشتریان ایجاد نمی شود. در تشخیص تقلب، تنها فعالیت های بانکداری آنلاین ثبت شده در سیستم های بانکی می تواند دیده شود، نه روند سازش طیف و شواهد پزشکی قانونی (از جمله برچسب که نشان دهد آیا یک معامله جعلی است) که می تواند برای درک ماهیت فریب بسیار مفید باشد. این باعث چالش برای شناسایی تقلب پیچیده با اطلاعات بسیار محدود است.

(5) الگوهای رفتاری متنوع واقعی مشتری. رابط کاربری بانکداری آنلاین را فراهم می کند، یک ورودی یک مرحله ای برای مشتریان جهت دسترسی به بسیاری از خدمات بانکی و حساب های متعدد است. در انجام تجارت بانکداری آنلاین، هر مشتری ممکن است تراکنش های بسیار متفاوت برای مقاصد مختلف انجام دهد. این امر منجر به تنوع معاملات واقعی مشتری می شود. علاوه بر این، کلاهبرداران با شبیه سازی رفتار مشتری واقعی و تغییر رفتار خود غالبا به رقابت

با پیشرفت در تشخیص تقلب می پردازند. این باعث دشواری در توصیف تقلب و حتی سخت تر از آن متمایز از رفتار واقعی می شود.

(6) سیستم بانکی آنلاین ثابت است. روند بانکداری آنلاین و سیستم بانکی ثابت هستند. هر مشتری به سیستم بانکی دسترسی دارد و تنها می تواند از خدمات از یک مسیر از پیش تعیین شده استفاده کند. این امر منجر به مراجع خوب برای توصیف توالی معمول در رفتار واقعی، و برای شناسایی سوء ظن کوچک در بانکداری آنلاین جعلی است.

ویژگی های بالا جهت تشخیص تقلب بانکداری آنلاین و تشخیص تقلب بانکی آنلاین با چندین چالش عمده در پژوهش WWW و W2T روبرو است، به ویژه برای جامعه داده کاوی : داده ها بسیار نامتوازن، داده های بزرگ، کارایی مدل در برخورد با داده های پیچیده ، داده کاوی پویا و استخراج الگو با برچسب محدود و یا بدون برچسب محدود، و تفکیک تجزیه و تحلیل داده ها بدون تمایز است. علاوه بر این، به منظور توسعه یک مدل واحد برای مقابله با تمام جنبه های بالا بسیار چالش برانگیز است، که تا حد زیادی در چالش تشخیص تقلب کارامد می باشد.

2.2. کار عمومی در تشخیص تقلب

بسیاری از تکنیک های آماری و یادگیری ماشین برای مقابله با تقلب توسعه یافته اند [54]، به عنوان مثال، شبکه عصبی، درخت تصمیم [48]، رگرسیون لجستیک [4] و قانون مبتنی بر سیستم های خبره [22]. آن ها برای تشخیص فعالیت های غیر طبیعی و برای تشخیص تقلب در بسیاری از زمینه ها، از جمله پول شویی، تقلب در کارت اعتباری، کامپیوتر [29]، و غیره استفاده شده اند. آنها را می توان به عنوان روش بدون نظارت و تحت نظارت طبقه بندی کرد. روش بدون نظارت، مانند مدل پنهان مارکف [46، 56]، به طور عمده در تشخیص استفاده می شود و سنبل تشخیص زمانی نمونه آموزشی بدون برچسب است. بر اساس داده های تاریخی و دامنه دانش، بانکداری آنلاین می تواند به وضوح نمونه داده ها را برای گزارش از قربانیان یا سازمان کنترل جرم مرتبط با برچسب جمع آوری کند. روش بدون نظارت از اطلاعات برچسب استفاده نمی کند، و دقت پایین تری از روش نظارتی دارد. برخی از روش های نظارت، مانند شبکه های عصبی و جنگل تصادفی [10]، به خوبی در انجام بسیاری از برنامه های کاربردی طبقه بندی شده اند، از جمله برنامه های تشخیص تقلب، حتی در حالات طبقه بندی نامتعادل خاص [2، 9، 14، 42، 50، 67]. با این حال، آن ها

با اطلاعات بسیار نامتوازن توانایی مقابله ندارند، و یا قادر به برخورد با پیچیدگی های جامع به عنوان داده های بانکی آنلاین و تجارت نمی باشند.

همچنین درک تضاد پیچیدگی ها بین رفتار های جعلی و رفتار واقعی می تواند الگوهای ضروری باشد که، زمانی که در یک طبقه بندی به ثبت رسید، منجر به دقت بالا و قدرت پیش بینی شود. چنین درک باعث ظهور استخراج الگوی مقابل، مانند الگوی [24، 25] در حال ظهور، پریدن الگوهای در حال ظهور [41]، و مجموعه های متقابل می شود[6]. با این حال، آزمایش های ما نشان می دهد که این روش برای تشخیص تقلب نادر در میان تعداد بسیار زیادی از معاملات واقعی، موثر نیست.

3.2. تشخیص تقلب در بانکداری آنلاین

تعداد بسیار کمی از مقالات در مورد تشخیص تقلب در بانکداری آنلاین [49] وجود دارد. بسیاری از آن ها نگرانی از پیشگیری تقلب دارند، که با استفاده از اقدامات امنیتی کارآمد برای جلوگیری از معاملات مالی متقلبانه توسط کاربران غیر مجاز و برای اطمینان از درستی معامله انجام شده [8, 20, 28, 33, 39] . Aggelis [1] یک سیستم تشخیص تقلب بانکداری آنلاین برای پردازش آنلاین است. یکی دیگر از سیستم های ارائه شده که [37] به خوبی کار می کند اما نیاز به یک جزء دارد که باید دانلود شود و در دستگاه مشتری نصب گردد، که برای استقرار ناخوشایند است.

در عمل، سیستم های تشخیص تقلب بانکی آنلاین بر پایه قوانین در معاملات ایجاد می شوند. قوانین عمدتا با توجه به دامنه دانش تولید شده؛ در نتیجه، این سیستم ها معمولا میزان کاذب بالا دارند، اما نرخ تشخیص تقلب کم است. نکته مهم، قوانین تطبیقی به تغییرات در نوع تقلب است.

4.2. تشخیص کارت اعتباری تقلبی

تقلب آنلاین و کلاهبرداری آنلاین: تقلب در کارت اعتباری به دو نوع تقسیم می شود. تقلب آفلاین با استفاده از یک کارت فیزیکی به سرقت رفته در یک مرکز فروشگاه. در اغلب موارد، موسسه صدور کارت می تواند آن را قبل از استفاده جعلی قفل کند. کلاهبرداری آنلاین از طریق وب، خرید تلفنی و یا مواردی که وجود کارت نیاز نیست. تنها جزئیات کارت مورد نیاز است، و امضا دستی و مشخصات صاحب کارت در زمان خرید لازم نیست. کارت اعتباری آنلاین تقلب

در معامله را افزایش داده است. در مقایسه با تشخیص تقلب بانکداری آنلاین، بسیاری از بحث پژوهش در دسترس و راه حل های مربوط به کارت اعتباری تشخیص تقلب [3، 23، 43] وجود دارد.

بیشتر کار در پیشگیری و کشف تقلب کارت اعتباری با شبکه های عصبی انجام شده است [36] CARDWATCH . [2] ویژگی های یک شبکه عصبی آموزش دیده با داده های گذشته یک مشتری خاص است و باعث می شود که شبکه برای پردازش الگوهای جاری تشخیص ناهنجاری ممکن شود. برایوس و لانگسدوف یک سیستم ارتباط مبتنی بر قواعد ترکیب با رویکرد عصبی تطبیقی ارائه کرده اند [9] . فالکون، توسطHNC ، با استفاده از شبکه های عصبی مصنوعی آموزش دیده در یک نوع از الگوریتم آموزش پس انتشار توسعه یافته [32]. یادگیری ماشین، تشخیص الگو تطبیقی، شبکه های عصبی، و مدل سازی آماری به منظور توسعه مدل های فالکون پیش بینی شده برای ارائه یک اندازه اطمینان در مورد اینکه آیا یک معامله خاص جعلی است به کار گرفته می شود. یک طبقه بندی عصبیMLP یک مثال دیگر از یک سیستم است که با استفاده از شبکه های عصبی عمل می کند [26]. تنها بر روی اطلاعات عمل خود و سابقه فوری آن عمل می کند، اما نه در پایگاه داده های تاریخی از فعالیت های گذشته دارنده کارت. روش موازی شبکه عصبی گرانول (GNN) یک روش فازی مبتنی بر قواعد شبکه عصبی است [57]. سیستم عصبی به صورت موازی با استفاده از مجموعه داده های آموزشی آموزش دیده، و شبکه عصبی فازی موازی آموزش دیده و به کشف قوانین فازی برای پیش بینی آینده می پردازد. معرفی یک مدل ترکیبی، ترکیب یک سیستم خبره با یک شبکه عصبی برای افزایش مدل سازی آمار و کاهش تعداد خطا "نادرست" [19]. همچنین برخی از روش های بدون نظارت، مانند HMM [56] و خوشه [46]، هدف قرار دادن مجموعه داده های بدون برچسب است.

همه روش های تشخیص تقلب کارت اعتباری به دنبال کشف الگوهای مخارج بر اساس داده های تاریخی از فعالیت های گذشته مشتری خاص است. این برای بانکداری آنلاین به دلیل تنوع فعالیت های مشتریان بانکداری آنلاین و داده های تاریخی محدود در دسترس برای یک مشتری مناسب است.

5.2. تشخیص نفوذ به کامپیوتر

بسیاری از سیستم های تشخیص نفوذ پایه عملیات خود را در تجزیه و تحلیل داده های ممیزی تولید شده توسط سیستم عملیات قرار می دهند. سوء استفاده و تشخیص ناهنجاری: روش تشخیص نفوذ در کامپیوتر به طور گسترده بر اساس یک مدل از رسوخ به دو دسته تقسیم می شود. تلاش برای تشخیص سوء استفاده و به رسمیت شناختن حملات در قالب یک الگو یا امضا، و سپس نظارت بر چنین حوادثی [30، 34، 38]. روش سوء استفاده عبارتند از سیستم های خبره، استدلال مبتنی بر مدل، انتقال حالت تجزیه و تحلیل و ضربه زدن به کلید نظارت دینامیک [58]. تشخیص سوء استفاده ساده و سریع است. نقطه ضعف اصلی آن این است که ممکن است پیش بینی همه حملات مختلف به دلیل الگوهای شناخته شده امکان پذیر نباشد. تشخیص ناهنجاری برای ایجاد یک مشخصات طبیعی تاریخی برای هر کاربر تلاش می کند و سپس با استفاده از یک انحراف بزرگ از مشخصات برای نفوذ استفاده می کند [30، 55]. روش تشخیص ناهنجاری شامل روش های آماری، نسل پیش بینی، و شبکه های عصبی است. استفاده از تشخیص ناهنجاری ممکن است برای تشخیص حملات باشد؛ ضعف آن این است که به احتمال زیاد در مقابل نرخ بالای هشدار اشتباه کند.

روش داده کاوی را می توان برای تشخیص نفوذ استفاده کرد. مدل های طبقه بندی با الگوریتم قوانین انجمن و اپیزودهای مکرر برای تشخیص نفوذ ناهنجاری توسعه یافته است [40]. این رویکرد به طور خودکار می تواند از یک مقدار زیادی از داده های ممیزی و مدل های تشخیص مختصر و دقیق تولید شود. با این حال، نیاز به مقدار زیادی از داده های ممیزی به منظور محاسبه مجموعه مشخصات دارد. از آنجا که اکثر شواهد فورنزیک برای تقلب در رایانه های مشتریان است و بازیابی آن بسیار دشوار می باشد، روش های تشخیص نفوذ می تواند به طور مستقیم برای بانکداری آنلاین استفاده شود.

6.2. تشخیص تقلب از راه دور

انواع مختلف از تقلب های مخابراتی را می توان به دو دسته طبقه بندی کرد . تقلب اشتراکی و تقلب سوار: تقلب اشتراک زمانی که اشتراک در یک سرویس به دست آمده است رخ می دهد ، اغلب با جزئیات هویت کاذب و به هیچ وجه قصد پرداخت. تقلب سوار زمانی که یک سرویس بدون اختیار لازم استفاده می شود رخ می دهد و معمولا با ظهور

تماس های ناشناخته در لایحه شناسایی شده است. کار پژوهش در تشخیص تقلب مخابراتی به طور عمده در شناسایی تقلب سوار متمرکز شده است. بیشتر تکنیک های استفاده از جزئیات داده های ثبت تماس برای ایجاد پروفایل رفتار مشتریان، و تشخیص انحراف این پروفایل است.

روش پیشنهادی عبارتند از رویکرد مبتنی بر قواعد [53]، شبکه های عصبی [45، 59]، روش تجسم [18]، و غیره. در میان آن ها، شبکه های عصبی در واقع می توانند پروفایل کاربر در شیوه ای مستقل را محاسبه کنند. شبکه های عصبی برای کاهش قابل توجه هزینه های بهره برداری هستند. همانطور که با تشخیص کارت اعتباری تقلبی، روش های تشخیص تقلب مخابراتی برای توصیف الگوهای رفتار مشتریان بانکداری آنلاین دشوار است.

واضح است که، هیچ روش واحدی نمی تواند مشکل تشخیص تقلب بانکی آنلاین را به راحتی حل کند. از آنجا که روش های مختلف مزایایی در ابعاد مختلف دارند، این باور وجود دارد که یک راه حل ترکیبی از هر گونه راه حل واحد بهتر است. شبکه های عصبی با موفقیت در هر سه نوع تشخیص تقلب به تصویب رسیده اند و اعتقاد بر یک مدل پایدار است.در بانکداری آنلاین رفتار توالی داده های در دسترس آنلاین به سیستم رابط بانکی ورود کرده اند و بین فعالیت های غیر طبیعی و نرمال تفاوتی است، الگوی رفتار پی در پی باید برای تشخیص تقلب گنجانده شود.

3. بیان مسئله

در این بخش، مفاهیم و نمادهای که در این مقاله مورد استفاده قرار گرفته است تعریف می کنیم.

تعریف 1 (معامله) یک معامله τ ، یک تاپل $\tau = \{a_1, a_2, \cdots, a_L\}$ است، که از ارزش ها همه صفات L تشکیل شده $\{A_1, A_2, \cdots, A_L\}$ و در بانکداری تراکنشی داده ها موجود است . طول یک معامله | τ | با تعدادی از ویژگی های درگیر است، به عنوان مثال، | τ | = L .

به عنوان مثال، یک مجموعه $\tau_0 = \{age = 25, gender = 'F', career = 'student'\}$ یک معامله با طول $|\tau_0| = 3$ است. مفهوم معامله، منبع داده اصلی یک مجموعه داده های کمتر از مشخصات بالا است. تمام معاملات به صورت یک معامله مجموعه T و هر معامله دارای یک کلاس $C\ (C \in \{Fraud, Genuine\})$ است.

تعریف 2 (الگو) الگوی P ترکیبی از ویژگی های، مقادیر آستانه، و اپراتورهای اتصال (∧) است. برای مثال، $P_0 = (age \in [25,35]) \wedge (gender = M) \wedge (career = IT)$ یک نمونه با طول $|P_0| = 3$ می باشد.

تعریف 3 (هشدار) اگر معامله τ یک الگوی P را ارضا کند، رابطه τ بر این اساس $\tau \models P$، باعث یک هشدار در τ خواهد شد.

معامله مجموعه T را می توان به دو گروه از نظر هشدار تقسیم کرد، یعنی $T_{\models}^{(\mathbf{P})}$ و $T_{\models}^{(\mathbf{P})}$، که در آن $T_{\models}^{(\mathbf{P})}$ مجموعه ای از معاملات با الگوی P است و $T_{\models}^{(\mathbf{P})}$ مجموعه ای از معاملات باقی مانده است. بنابراین، ما داریم:

$$TP^{(\mathbf{P})} = T_{\models}^{(\mathbf{P})} \cap T_+, \; FN^{(\mathbf{P})} = T_{\not\models}^{(\mathbf{P})} \cap T_+ \qquad (3.1)$$

$$FP^{(\mathbf{P})} = T_{\models}^{(\mathbf{P})} \cap T_-, \; TN^{(\mathbf{P})} = T_{\not\models}^{(\mathbf{P})} \cap T_- \qquad (3.2)$$

در اینجا، TP، عدد مثبت است، نشان دهنده تعداد تقلب گرفتار شده توسط الگوی P؛ FP، تعداد مثبت ، نشان دهنده هشدار نادرست موجب شده توسط P؛ FN، تعداد منفی کاذب، نشان دهنده تقلب از دست رفته توسط P؛ TN، عدد منفی است، مخفف تعداد معامله واقعی پیش بینی شده توسط T + است معامله تقلب مجموعه در حالی که تی مجموعه معامله واقعی است، که در آن T = T-+ T + است. هدف از تشخیص تقلب برای بانکداری آنلاین برای رسیدن به یک TP بالاتر و FP پایین تر است.

تعریف 4 (جلسه) جلسه دوره رفتار یک مشتری بین ورود و خروج سیستم از سیستم بانکی آنلاین است.

تعریف 5 (رویداد) فرض کنید $R = \{A_1, .., A_M\}$ مجموعه ای کامل از ویژگی رویداد می شود، تعداد تمام ویژگی های درگیر با حوادث. سپس رویداد در زمان t یک $(M + 3)$تاپل است: $e_t = ([a_{1t}, \ldots, a_{Mt}], c, t, s)$، که در آن a_{mt} که مقدار در زمان t است ، C نوع رویداد است، و بازدید کنندگان یک عدد صحیح برای شناسایی تعداد جلسه که در آن همکاران بوده اند مشخص کرده اند. در بانکداری آنلاین، انواع رویداد وجود دارد، و معامله یک نوع رویداد می باشد.

در یک جلسه، معمولا رویداد های متعدد وجود دارد.

تعریف 6 (دنباله رویداد) یک توالی رویداد S مجموعه ای از حوادث $\langle e_1, e_2, \ldots, e_N \rangle$ است که در آن N تعداد حوادث در یک توالی است، $e_n.t < e_{n+1}.t$ ($e.t$ نشان دهنده زمان یک رویداد)، و حوادث به ترتیب صعودی به عنوان زمان وقوع آنها مرتب شده اند. اگر حوادث e_i, e_j و e_k در همان جلسه رخ دهد، پس از آن $e_i.s = e_j.s = e_k.s$ ($e.s$ نشان دهنده جلسه یک رویداد). طول توالی توسط S نشان داده شده است $|\mathbb{S}| = N$.

در بانکداری آنلاین، توالی منعکس کننده رفتار مشتری است از آنجایی که آن ها حساب خود را در یک بانک افتتاح کرده اند. رفتار مشتری توسط یکی توالی ارائه شده است.

جدول 3 توالی فعالیت های مشتری است. دو نوع رویداد وجود دارد: دریافت و پرداخت. هر رویداد دارای ویژگی های خود است. E1، E4 و E7 رویداد دریافتی هستند، و آن ها تنها شامل یک ویژگی A3 می باشند، در حالی که بقیه رویداد های پرداخت شامل ویژگی های A1 و A2 می باشند. برخی رویداد ها در یک جلسه رخ می دهد ، S1 جلسه متشکل از حوادث E1، E2 و E3، و S2 جلسه شامل E4، E5 و E6 می باشد.

تعریف 7 (توالی پایگاه داده) توالی پایگاه داده A یک مجموعه از توالی است.

به منظور اندازه گیری تضاد بین معامله حاضر و تاریخ معامله مشتریان خود، بردار مقابل را تعریف می کنیم.

Event id	A_1	A_2	A_3	Event type	Time stamp	Session
e_1	–	–	A	Login	t_1	s_1
e_2	2	10	–	Pay	t_2	s_1
e_3	2	10	–	Pay	t_3	s_1
e_4	–	–	A	Login	t_4	s_2
e_5	2	10	–	Pay	t_5	s_2
e_6	1	5	–	Pay	t_6	s_2
e_7	–	–	A	Login	t_7	s_3
e_8	2	5	–	Pay	t_8	s_3

جدول 3 توالی های مشتری.

تعریف 8 (بردار کنتراست) با توجه به یک رویداد e' ، و دنباله ای از S مشتری فعلی، ما می توانیم بردار $V(e') = \{v_1, \ldots, v_M\}$ را رسم کنیم ، که در آن

$$v_i = 1 - \frac{\left|\{e|e \in \mathbb{S}, e.c = e'.c, e.a_i = e'.a_i\}\right|}{|\{e|e \in \mathbb{S}, e.c = e'.c\}|} \qquad (3.3)$$

در اینجا M تعداد تمام ویژگی های درگیر با رویداد است، و $a_i (1 \leq i \leq M)$ مقدار ویژگی هوش مصنوعی A_i است. پاسخ $V(e')$ بردار مقابل e' است . بردار مقابل مجموعه ای از معیارهای بررسی پشتیبانی از ویژگی های این رویداد در حال حاضر در میان فعالیت های تاریخی مشتری دارد.

قبل از محاسبه بردار کنتراست e'، ما معمولا برای اولین بار تمام ویژگی های عددی گسسته را برسی می کنیم. برای برخی ویژگی های اسمی که بیش از حد ارزش های متمایز دارد، ما نیز آن ها را به چند گروه طبقه بندی می کنیم. در مرحله استخراج الگو، بردار مقابل می تواند به عنوان ویژگی های مشتق باشد.

فرض کنید e_8 در جدول 3 رویداد مورد نظر را پیش بینی کند. از آنجا که e_8 و e_7 در s_3 همان جلسه رخ می دهد و آن ها دو نوع رویداد های مختلف با ویژگی های مختلف ایجاد می شود، ما با ادغام e_7و e_8 بردار کنتراست را محاسبه می کنیم. با توجه به تعریف بردار مقابل، ما باید 3 عناصر در $V(e_8)$ داشته باشم، که در آن v_1و v_2 برای ارزیابی کنتراست به ترتیب درصفات A_1وA_2 می باشد. بنابراین $v_2 = 1 - 1/4 = 3/4$، و $v_1 = 1 - 3/4 = 1/4,$ است. v_3 کنتراست و e_7رویداد ورودی است، بنابراین $v_3 = 1 - 2/2 = 0$. سپس ما از $V(e_8) = \{1/4, 3/4, 0\}$. در مرحله بعد، $V(e_8)$ را می توان با ویژگی های اساسی از e_8 با هم ادغام کرد که استخراج الگوه ای کنتراست را منجر می شود.

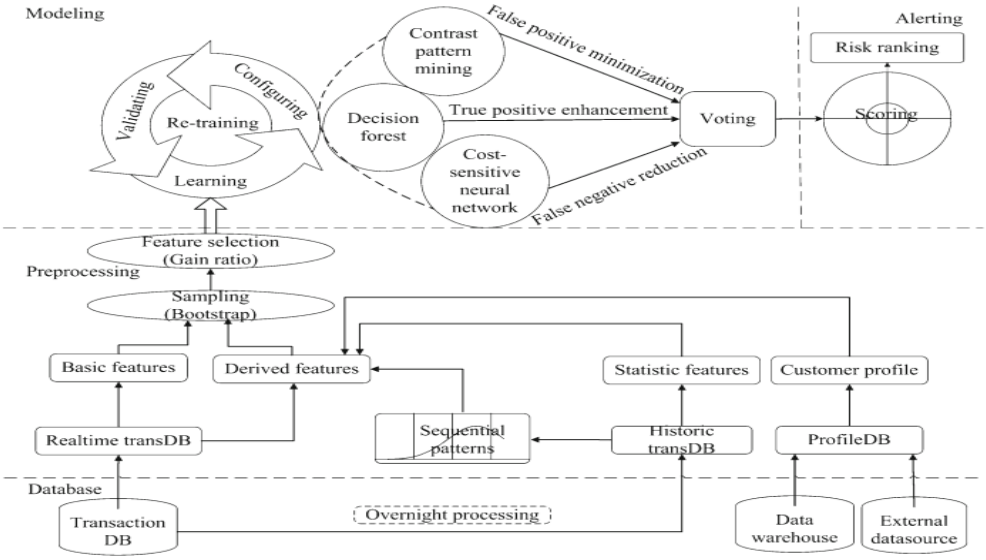
مشکل اساسی در تشخیص تقلب بانکداری آنلاین برای استخراج توالی رویداد است، به صورت یک پایگاه داده به ترتیب، ساخت یک بردار مقابل برای هر مشتری با لینک دادن به / توالی رویداد خواهد بود، و سپس شناسایی الگوهای افتراقی بود. چنین الگوهایی نشان می دهد که خطر ابتلا منجر به تقلب در بانکداری آنلاین است. در بخش های زیر، ما چارچوب سیستم و الگوریتمی برای حمایت از تشخیص تقلب در بانکداری آنلاین معرفی می کنیم.

4 . چارچوب سیستم

ما یک سیستم مدیریت ریسک بانکداری آنلاین I-Alertor پیاده سازی می کنیم .I-Alertor : ادغام ویژگی های مختلف و مدل های داده کاوی با هدف تحکیم منابع مختلف منابع برای نظام حل مسئله است.

شکل 1 سیستم تشخیص تقلب پیشنهادی ما برای معماری بانکداری آنلاین ، I-Alertor را نشان می دهد. I-Alertor متشکل از چهار ردیف است: پایگاه داده ها، داده ها قبل از پردازش، مدل سازی و هشدار، بر اساس فرآیند استخراج. ردیف پایگاه داده در واقع منابع داده است و آن ها به بازیابی اطلاعات مرتبط متصل است . مربوطه آنلاین داده تشخیص تقلب بانکی از منابع داده ناهمگن، از جمله بانکداری اینترنتی زمان واقعی معامله سیاهه های مربوط، داده های اخیر و تاریخی معامله، مشتری اطلاعات دموگرافیک، و سایر منابع خارجی جمع آوری شده. نوع و فرمت داده ها از منابع مختلف متفاوت است. به منظور کاهش حجم بالایی از داده های منبع، ما استخراج اطلاعات مربوطه را از داده های خام و تبدیل آن به فرمت های مورد نیاز برای مدل های. داده های مربوط شامل زمان واقعی معاملات بانکی آنلاین، توالی رفتار مشتری بانکی، داده های تاریخی، و پروفیل مشتری می باشد.

ردیف قبل از پردازش در زمان واقعی تجمع معامله، تعمیر و نگهداری داده های تاریخی، و آماده سازی داده ها برای آموزش مدل و پیش بینی است. همچنین شامل تابع برای انتخاب ویژگی های اساسی و استخراج ویژگی ها می باشد. دو وظیفه اصلی در مرحله پیش پردازش وجود دارد: نمونه برداری داده ها و انتخاب ویژگی. به عنوان داده های معامله بانکی آنلاین بسیار نامتعادل هستند، و نمونه برداری قبل از استفاده از هر مدل داده لازم است [16]. از آنجا که تعداد معاملات واقعی بزرگتر است، ما به منظور کاهش حجم داده ها و توزیع کلاس عدم تعادل از نمونه برداری استفاده می کنیم [27]. در سیستم ما، نمونه های خودگردان [17، 21] برای حفظ توزیع آماری داده ها اعمال می شوند. همچنین انتخاب ویژگی برای مدل بسیار مهم است [31]. بسیاری از ویژگی های استفاده شده در مدل بیش از حد کار آمد هستند، که برای تشخیص زمان واقعی کلاه برداری بسیار مهم و تاثیر گذار است. در سیستم ، ما ارتباط یک ویژگی با هر کلاس از نظر نسبت افزایش اطلاعات خود را بین دو طبقه محاسبه می کنیم ، ویژگی آن ها بیشترین تبعیض را دارند.



شکل 1. چارچوبی برای تشخیص تقلب در بانکداری آنلاین.

مدل سازی ردیف نسل مدل را ایجاد می کند ، مانند تشکیل مدل، تنظیم پارامتر، برنامه ریزی کار، مدل و آموزش مجدد، و غیره سه روش داده کاوی در سیستم به تصویب رسیده :

- استخراج الگوی مقابل، که به شناسایی رفتار بانکی با تقلب در بانکداری آنلاین در ارتباط است.

- شبکه های عصبی حساس به هزینه، که بر هزینه بالاتر از ایجاد یک خطا در رده بندی یک تقلب نسبت به یک معامله واقعی است؛

- جنگل تصمیم، که ترکیبی از قدرت درخت های تصمیم گیری فردی به شیوه ای و به اشتراک گذاری برای ساخت گروه درخت تصمیم گیری است؛

ردیف هشدار ترکیب خروجی از این سه مدل با توجه به روش رای گیری در شرایط خاص برای هر معامله است. در نهایت، یک نمره خطر برای هر تراکنش تولید می شود. یک هشدار ممکن است دارای نمره بالاتر از حد آستانه باشد. ترکیبی از سه روش نرخ مثبت کاذب و نرخ منفی کاذب را کاهش می دهد و نرخ مثبت واقعی را افزایش می دهد.

در زیر، ما اول بحثی خلاصه در مورد سه مدل و ترکیب آن ها انجام می دهیم . از آنجا که روش استخراج الگوی مقابل کلاسیک برای تشخیص تقلب بانکی آنلاین مناسب نیست، روش جدیدی به نام کنتراست برای استخراج رفتار بانکی آنلاین در بخش 5 ارائه شده است، و جزئیات خطر بر اساس ترکیب مدل در بخش 6 داده شده است.

1.4. استخراج الگوی کنتراست

تعریف 9 (الگوی کنتراست) با توجه به دو مجموعه داده های معامله، D_f و D_g ، D_f شامل نمونه داده تقلبی، و D_g شامل نمونه داده واقعی هستند. فرض کنید $S_{D_f}(X)$ پشتیبان مجموعه آیتم Xدر D_fو $S_{D_g}(X)$ و حمایت از Xدر D_g را است، و سپس الگوی کنتراست **(CPs)**را می توان به شرح زیر تعریف کرد:

$$CPs = \left\{ X | S_{D_g}(X) \leq \omega * S_{D_f}(X), S_{D_f}(X) \geq \theta \right\} \qquad (4.1)$$

ω> 0 ضریب کنتراست است و θ آستانه حداقل نرخ تشخیص است.

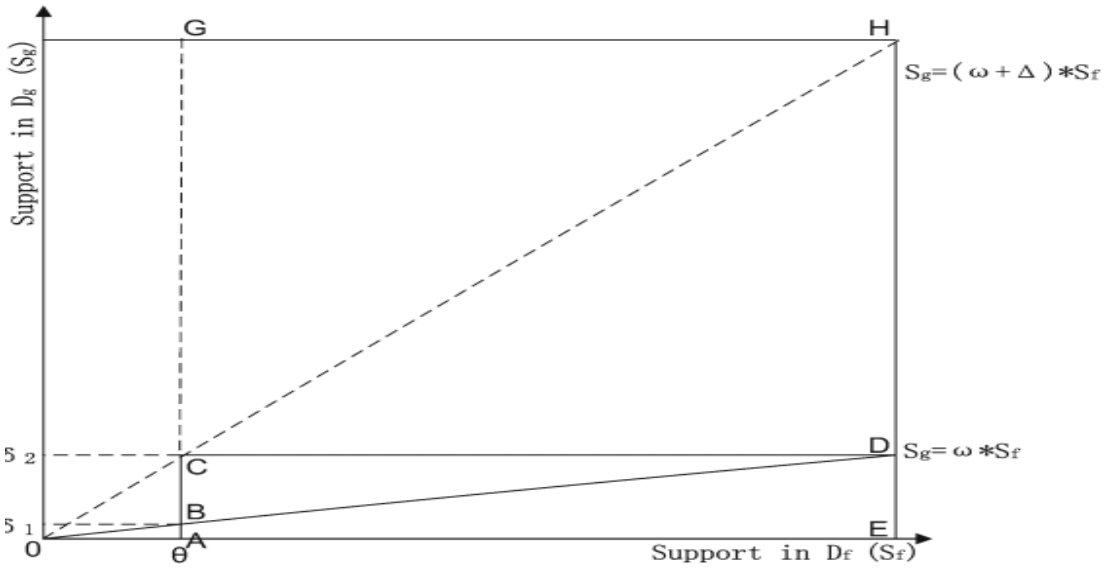
علاوه بر این، عملکرد کنتراست به معنی تفاوت حمایت از یک مجموعه آیتم Xدر دو جدول داده D_fو D_g را به صورت زیر تعریف می کنیم ،

$$F(X) = \frac{S_{D_f}(X)}{S_{D_g}(X)} \qquad (4.2)$$

برای راحتی توضیح، تمام مجموعه آیتم در یک پشتیبانی پیش بینی شده است، همانطور که در شکل 2، که در آن Y محور برای حمایت از یک آیتم تنظیم شده در مجموعه داده واقعی (D_g) است، و محور X پشتیبانی در مجموعه تقلب (D_f) است. الگوهای مقابل همه در ذوزنقه ABDE واقع شده اند.

روش معمولی برای استخراج الگوی کنتراست مقابل در حال ظهور است [24]. یک مورد خاص در حال ظهور است، پریدن در حال استخراج الگوی کنتراست (JEP) برای شناسایی مجموعه آیتم متضاد پیشنهاد می شود ، که پشتیبانی در D_g صفر است، اما D_f برابر یا بزرگتر از یک حد آستانه خاص برایD_f است. در واقع، JEPs به ندرت در کلاس مجموعه داده ها نامتوازن قرار می گیرد، مانند پایگاه داده معاملات بانکی آنلاین، که در آن اقلیت به عنوان هدفD_f برای تنظیم است.

همانطور که قبلا اشاره شد استخراج الگوی مقابل یک مشکل سخت NP است و زمان آن بالا است به ویژه هنگامی که مجموعه داده با یک آستانه بسیار کم از سرعت کشف و شناسایی حداقل اعمال می شود. بنابراین، در حال حاضر استخراج الگوی کنتراست می تواند به خوبی در کشف تقلب در بانکداری آنلاین کمک کند. همچنین این مورد توسط آزمایش های ما ثابت شده است.



شکل 2. طرح پیش بینی الگوی کانترست مبتنی بر پشتیبان.

دلایل زیر محدودیت های موجود در کشف الگوها در داده های بسیار نامتوازن را نشان می دهد :

(1) MDB-LL$_{border}$ از الگوریتم ماکس ماینر (7) برای محاسبه مستقیم مرز ها در هر دوی D_f وD_g صرف نظر از کوچک بودن پشتیبانی حداقل محاسبه می کند. ارایه نتایج برای پشتیبانی بسیار کوچک در زمان قابل قبول و مناسب، خارج از توان و ظرفیت ماکس ماینر است.

(2) MDB-LL$_{border}$ فضای کل از مجموعه آیتم در طول عملیات تمایز مرز خود است ، به ویژه هنگامی که مرزها حاوی الگوهای طولانی است.

(3) الگوریتم استخراج الگوی در حال ظهور ، خروجی تعداد زیادی از الگوها با یک نرخ رشد مناسب است. از این رو یک فیلتر موثرتر برای از بین بردن الگوهای ناچیز در داده ها بسیار نامتوازن مهم است.

برای حل مشکلات ذکر شده در بالا، ما الگوریتم کنتراست ماینبر با بهبود الگوریتم $MDB\text{-}LL_{border}$ از طریق استراتژی های خاص، که جزئیات در بخش 5 توضیح داده خواهد شد ارائه داده ایم.

2.4. شبکه عصبی حساس به هزینه

شبکه عصبی حساس به هزینه (CNN) یک روش اصلاح شده عصبی مبتنی بر شبکه است ، که برای بانکداری آنلاین طراحی شده است.

مشکل طبقه بندی بسیار نامتوازن در تشخیص تقلب در بانکداری آنلاین باعث می شود روش های طبقه بندی کلاسیک (مانند روش آمار، هزینه SVM، درخت تصمیم، طبقه بندی متضاد بر اساس الگو، شبکه های بیزی، شبکه عصبی) به خوبی انجام نشوند. با این حال، مشخص شده است که بعضی از شبکه های عصبی دیگر هم در دقت و بهره وری عملکرد بهتری نسبت به آن ها دارند، و یادگیری به هزینه اثبات شده به یک راه حل خوب برای این مشکل عدم تعادل کلاس حساس است [62]. بنابراین، ما این ایده از شبکه عصبی توسعه داده و طراحی یک مدل شبکه عصبی حساس به هزینه را انجام می دهیم. به عملکرد پیش بینی بسیار بهتری نسبت به روش های دیگر ذکر شده در بالا می رسیم.

شبکه های عصبی مصنوعی شبکه های الکترونیکی نسبتا خام "نورون" بر اساس ساختار عصبی مغز هستند. آن ها با پردازش سوابق در یک زمان، و با مقایسه طبقه بندی آن ها از رکورد (که، در آغاز، تا حد زیادی دلخواه می باشد) با طبقه بندی واقعی از رکورد تفکیک می شوند. و برای تغییر الگوریتم شبکه در دور دوم، و غیره مورد استفاده است. میزان خطای پارامتری کلیدی برای تصمیم گیری زمانی است.

یک خطا مناسب می تواند آموزش و قابلیت پیش بینی بالا را تضمین کند. دو نوع خطا در مرحله آموزش وجود دارد :

(1) خطای مثبت (PE) : خطا PE تعیین کننده چگونگی خروجی مختلف یک شبکه عصبی از خروجی ایده آل برای انجام معاملات واقعی است ، با تمرکز بر معاملات واقعی به اشتباه به عنوان تقلبی.

(2) خطای منفی (NE) : خطا NE تعیین کننده چگونگی خروجی مختلف یک شبکه عصبی از خروجی ایده آل برای انجام معاملات تقلبی است ، با تمرکز بر معاملات تقلبی به اشتباه به عنوان واقعی است.

برای تشخیص تقلب، از دست دادن NE به مراتب پر هزینه تر از آن است که PE از دست رود. با این حال، شبکه عصبی حساس به هزینه سنتی PE و NE به همان شکل رفتار می کنند . بنابراین، خطای کلی کم حتی با NE بسیار بالا در طبقه بندی داده ها نامتوازن به دست آمده است، و در نتیجه دقت پیش بینی کم است. از این رو، ما مدل شبکه عصبی را با توجه به تاثیر بالاتر از NE در آموزش استفاده می کنیم، که این ایده یک شبکه عصبی حساس به هزینه به شکل تغییر یافته است.

3.4. جنگل تصمیم

درخت تصمیم گیری به طور گسترده ای در تجزیه و تحلیل استفاده می شود، و قوانین تولید شده توسط درخت تصمیم گیری به آسانی قابل درک هستند. با این حال، برخی از معایب جهت استفاده از درخت تصمیم گیری در تشخیص تقلب وجود دارد. قوانین عمومی به سمت ویژگی های خاص مغرضانه عمل می کنند. در یک سناریوی حساس به هزینه، شاخه های بسیار کوچک نیز بیش از حد در درخت وجود دارد و بیش از مشکل اتصالات جدی است. همچنین، از آنجایی درخت تصمیم گیری تنها یک ریشه دلرد، قوانین عمومی به صورت محلی مهم است ، بلکه در سطح جهان بی اثر هستند.

یک جنگل تصمیم، برای کشف قوانین موثر تر است، که متشکل از چند درخت تصمیم گیری قوی می باشد. این کار بهتر از روش درخت تصمیم گیری کلاسیک بر روی داده های نامتوازن در ایجاد یک مدل می باشد.

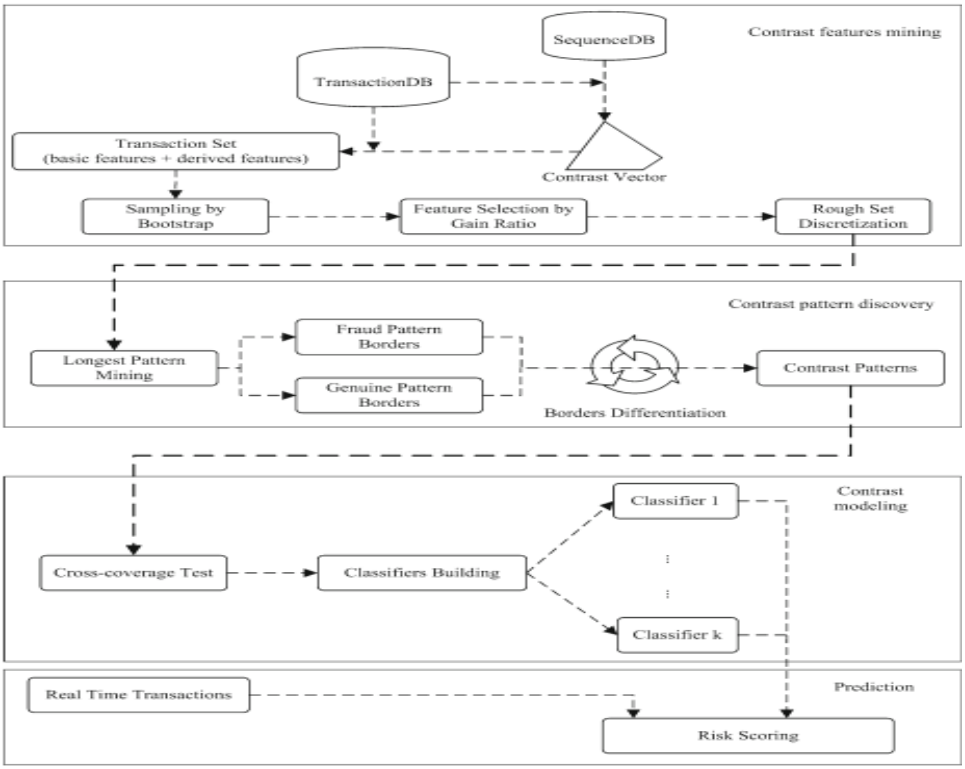
5. رفتار بانکداری ماینینگ کانترست یا کانترست کاوی

این بخش مدلی برای شناسایی الگوهای مقابل در رفتار بانکداری آنلاین معرفی می کند.

1.5. چارچوب

چارچوب ماینینگ کانترست در شکل 3 نشان داده شده . علاوه بر مجموعه داده دو منبع نشان داده شده است: پایگاه داده معامله رابطه ای و پایگاه داده پی در پی، سیستم متشکل از چهار مرحله می باشد: در مقابل ویژگی ماینینگ کنتراست ، کشف زمینه، مبتنی بر روش مدل سازی کنتراست های متعدد، و خطر به ثمر رساند. مرحله اول به طور عمده در پردازش اولیه متمرکز است. در مرحله اول، دنباله ای از هر یک از مشتریان معاملات بانکی آنلاین مشتری

استفاده شده است ، که برای تولید بردار مقابل برای هر معامله ساخته شده است. پس از آن ترکیب ویژگی های اساسی و بردار کنتراست (ویژگی های مشتق شده) به شکل مجموعه ای از ویژگی های خام است. نمونه بوت استرپ در اجرای نمونه گیری برای به دست آوردن مجموعه آموزش به تصویب رسیده است. پس از آن ما با استفاده از نسبت وزن به دست آمده اطلاعات برای انتخاب ویژگی های مهم که وزن بالاتری از حد آستانه خاص دارند. در نهایت، مجموعه ای برای ویژگی های گسسته ما معرفی شده است.



شکل 3 چارچوب ماینینگ کانترست برای رفتار بانکداری آنلاین.

مرحله دو مرحله اصلی الگوهای مقابل است. در ابتدا، حداکثر ماینینگ مورد استفاده برای ایجاد مرزهای الگوی برای تقلب و مجموعه واقعی. مرز-تمایز و فرایند اعتبار سنجی الگوهای کنتراست به طور مداوم در خروجی اجرا شده اند. در مرحله سه ، مدل گروه توسط خروجی الگوهای مرحله قبل طراحی می شود. تست متقابل پوشش بهترین تعیین الگو می باشد، که به طور موثر نشان دهنده قدرت طبقه بندی در میان تمام الگوهای است. مجموعه نمونه مورد استفاده

در آزمون مقطع پوشش به چندین قسمت تقسیم خواهد شد و هر قسمت یک مدل تولید می کند. در نتیجه، ما چند مدل برای به ثمر رساند ریسک به دست می آوریم.

آخرین مرحله پیش بینی زمان واقعی است. نمرات با مدل های متعدد با در نظر گرفتن وزن هر مدل برای یک معامله ارائه می شوند ، که توسط میزان پوشش در آزمون مقطع پوشش جمع شده اند. نمره نهایی داده شده توسط یک معامله. ارزش خود را در برابر سطح خطر نشان می دهد.

5.2. مدل سازی رفتار پیچیده

تغییر و بهبود کارایی الگوریتم در حال ظهور $\text{MDB-LL}_{\text{border}}$ الگوی ماینر [24] توسط استراتژی های زیر:

(1) به جای الگوی محاسبه مرزی به طور مستقیم با δ_1 کوچک در D_g پشتیبانی می شود(شکل 2)، ما δ_1 پشتیبانی حداقل به δ_2 ، که تضمین می کند ماینر حداکثر می توانید D_g را بالا ببرد. سپس، استفاده از مرزهای الگوی در مستطیل ACDE و تفریق مرزها در مثلث BCD اتفاق می افتد .

(2) در تابع مرزی متفاوت است [24]، تکرار فضای کل زیر مجموعه آیتم طولانی وقت گیر است، بنابراین همه مجموعه طول بزرگتر از 5و در یک جدول هش تحت فشار قرار دادند و پردازش در طول اعتبار در حال چک کردن عبارت است .

(3) ما در آزمون مقطع پوشش، که هرس تعداد بسیار بالایی از الگوهای کار برکنار شده، قبل از ساخت طبقه ما را برگزیند. کنتراست ماینستر برای الگوهای رفتاری مقابل ماینر در بانکداری آنلاین: الگوریتم 1 ایده اصلی الگوریتم را معرفی می کند.

Algorithm 1 ContrastMiner

Data: D_f,D_g,β,θ,len.
Result: Contrast patterns.

```
S' = {X|S_{D_f}(X) ≥ θ}            /* Retrieve pattern sets in fraud set */
S'' = {X|S_{D_g}(X) ≥ β}           /* Retrieve patterns sets in genuine set */
for each E in S' do
    U = {u|u ∈ S'', |E − u| ≤ len} /* Execute Border-Differ only for short borders */
    store BORDER-DIFF([ø, E], [ø, U]) in CandCPs
    add {u|u ∈ S'', |E − u| > len} into H  /* Store long borders for later checking */
for each [L, R] in CandCPs do
    Check the support of item set in[L, R] against H and remove the redundancy;
```

مراحل 1 و 2 محاسبه مرزهای الگوی مکرر برای داده تقلبی و واقعی مجموعه به ترتیب از طریق ماکس ماینستر. به دنبال یک بررسی صلاحیت هر یک از زیر مرز و خروجی الگوهای کنتراست در یافتیم که مرز متفاوت است.

3.5. گزیده ای از الگوهای رفتاری

هر چند که ما می توانیم پارامترها را برای استخراج الگوی ماینستر کنترل کنیم، بسیاری از الگوهای ارائه شده ساختار مشابه دارند. علاوه بر این، قابلیت تمایز میان الگوهایی با توزیع آن ها متفاوت است، بنابراین ما روش های زیر را برای فیلتر افزونگی اتخاذ می کنیم :

(1) مرتب کردن بر اساس الگوها به صورت نزولی، و استفاده از الگوهای قوی برای حذف آن هایی که ضعیف هستند. برای مثال، اگر الگوی P_1 یک الگو زیر P_2 و $F(p_1) \geq F(p_2)$ باشد، سپس P_2 حذف خواهد شد. هدف از حذف آن است که موارد عمومی تر را بدون کاهش قدرت تمایز نگه داریم.

(2) انجام آزمون پوشش متقابل در انتخاب الگوی برجسته. الگوریتم برای آزمون مقطع پوشش در الگوریتم 2 ارائه شده است.

Algorithm 2 Cross-coverage test

Data: Pattern set $\mathbb{P}$, coverage threshold η, D_f, D_g
Result: Positive set P^+ and negative set P^-
Initialize matchCount of samples in D_f and D_g to 0
Sort $\forall X \in \mathbb{P}$ by F(X) descending order and then by $|X|$ ascending order
for *each* $X \in \mathbb{P}$ **do**
    if $\exists T \in D_f$ *and* $X \subseteq T$ **then**
        $P^+ = P^+ \cup X$
        $Count(T) = Count(T) + 1$
        if $Count(T) \geq \eta$ **then**
            remove T from D_f
    if $\exists T \in D_g$ *and* $X \subseteq T$ **then**
        $P^- = P^- \cup X$
        $Count(T) = Count(T) + 1$
        if $Count(T) \geq \eta$ **then**
            remove $T from D_g$
return P^+ and P^-

در استفاده و تشخیص ناهنجاری، تعداد نمونه مثبت که معمولا کاملا نسبت به آن هایی که منفی هستند محدود است. آن دشوار و برای به دست آوردن یک مجموعه الگوی که می تواند تمام خواص نمونه ها منفی را بگیرد اغلب بسیار پر هزینه هستند. از سوی دیگر، خواص نشان می دهد رفتار مثبت اغلب دستکاری و در معرض تغییر قرار دارند، که آن را برای تشخیص استثنا در داده های نا متوازن سخت می کند. در واقع، برای انتخاب یک مجموعه الگوی بسیار چالش

برانگیز است. در روش ما، P^-است که از طریق تست پوشش در برابر D_g بدست آمده، که ارزیابی تا چه حد یک معامله C_f (برچسب تقلب) انتخاب شده، در حالی که P^+ بر D_fبرای انتخاب و نشانه یک معامله متعلق به C_f را فراهم می کند.

6. خطر رفتار بانکی اینترنتی بر اساس مدل های ترکیبی

در این بخش، ما مدل به ثمر رساند خطر ترکیبی از سه مدل داده کاوی برای تولید نرخ خطر هر معامله ، مورد بحث است.

1.6. منطق

به ثمر رساند خطر خطر رفتار مشتری فرد ارزیابی می کند. پس از تجزیه و تحلیل، به هر معامله نمره اختصاص داده، نمره بالاتر نشان دهنده بالاتر بودن احتمال تقلب است. آزمایش بر روی داده های بلند مدت معامله بانکداری آنلاین را نشان می دهد که هیچ خطری نداشته و به دلیل تنوع و پویایی تقلب به تنهایی عمل می کند. بنابراین ما با استفاده از روش های متعدد به تجزیه و تحلیل خطر از جنبه های مختلف و ترکیب نتایج می پردازیم. در سیستم ما، نمرات از سه مدل: الگوی ماینستر، شبکه های عصبی حساس به هزینه و جنگل تصمیم ، با استفاده از روش رای گیری ترکیب شده اند.

2.6. مدل منحصر به فرد به ثمر رساند خطر

1.2.6. امتیاز دهی در برابر الگوی ماینستر

نمرات پایه هر الگوی X_iدر P^+ و P^- پایه (X_i, P^+) و پایه (X_i, P^-) به ترتیب، می تواند توسط (6.1) و (6.2) محاسبه شود :

$$Base\ (X_i, P^+) = (S_{Df}(X_i) * F(X_i)/(1 + F(X_i))) \quad (6.1)$$

$$Base\ (X_i, P^-) = (S_{Dg}(X_i) * F(X_i)/(1 + F(X_i))) \quad (6.2)$$

نمرات هر یک t معامله در D_fو D_g برای P^+ و P^- هستند $\mathbb{S}^+(t)$ و $\mathbb{S}^-(t)$ به ترتیب :

$$\mathbb{S}^{+}(t) = \sum_{X_i \in P^{+}, X_i \subseteq t} Base\,(X_i, P^{+}) \,/\, \sum_{X_i \in P^{+}} Base\,(X_i, P^{+}) \qquad (6.3)$$

$$\mathbb{S}^{-}(t) = \sum_{X_i \in P^{-}, X_i \subseteq t} Base\,(X_i, P^{-}) \,/\, \sum_{X_i \in P^{-}} Base\,(X_i, P^{-}) \qquad (6.4)$$

فرض کنید دو معامله t_1 و t_2 وجود دارد ، اگر $\mathbb{S}^{+}(t_1) > \mathbb{S}^{+}(t_2)$ باشد، آنگاه احتمال برای $t_1 \in C_f$ (تقلب) بزرگتر از t_2 است. از سوی دیگر، اگر $\mathbb{S}^{-}(t_1) > \mathbb{S}^{-}(t_2)$ پس از آن t_1کمتر احتمال دارد به از t_2 را با C_f طبقه بندی شود. از آنجا که مقیاس برای $\mathbb{S}^{+}(t)$ و $\mathbb{S}^{-}(t)$ با توجه به (6.3) و (6.4) ناسازگار است، ترکیب به طور مستقیم آسان است. به منظور ادغام معاملات متقابل شلیک شده توسط $\mathbb{S}^{+}(t)$ و $\mathbb{S}^{-}(t)$ ، ما با محاسبه رتبه در P^{+} و P^{-} به صورت جداگانه و ترکیب رتبه با توجه به (6.5) عمل می کنیم.

رتبه t در D' ، که در معامله پیش بینی می شود، به عنوان R_t^{+} توسط $\mathbb{S}^{+}(t)$ نزولی، و R_t^{-} توسط $\mathbb{S}^{-}(t)$ به ترتیب صعودی اشاره کرد. رتبه t کوچکتر است، احتمال بالاتر بودن در C_f است. رتبه کلی از یک tمعامله $\mathbb{S}(t)_{CP}$ ، که نشان دهنده سطح ریسک کلی تراکنش است.

$$\mathbb{S}(t)_{CP} = \lambda_1 * \mathrm{Max}\left(R_t^{+}, R_t^{-}\right) + \lambda_2 * \mathrm{Min}\left(R_t^{+}, R_t^{-}\right) \qquad (6.5)$$

که در آن λ_1 و λ_2 ($\lambda_1 + \lambda_2 = 1$) ضرایب برای کنترل اولویت تصمیم گیری هستند. در آزمایشات، ما مجموعه ای λ_1 = 0.2 و λ_2 = 0.8 ، ثابت شده است که برای رسیدن به دقت کلی بالا برای بسیاری از مجموعه داده های ما مورد آزمایش قرار داده ایم.

2.2.6. امتیاز دهی توسط شبکه های عصبی حساس به هزینه

هر نورون از لایه خروجی شبکه عصبی ممکن است نشان دهنده یکی از کلاس ها باشد. به عنوان مثال در کلاس که مربوط به نورون با حداکثر خروجی است. با این حال، خروجی شبکه می تواند در معنای احتمالی مشاهده شود.

در نرم افزار تشخیص تقلب بانکداری آنلاین، ما مجموعه ای از خروجی های شبکه به عنوان برچسب تقلب داریم، و در داده های آموزشی مجموعه تنها دو مقدار (1 و 0) وجود دارد. بنابراین، احتمالا خروجی از مدل به عنوان یک نمره خطر برای معامله استفاده می شود؛ نمره بالاتر خطر، نشان دهنده احتمال بیشتر تقلب می باشد. سپس توسط این نمره به ترتیب نزولی قرار می گیرند. و رتبه t معامله به $\mathbb{S}(t)_{CNN}$ اشاره کرده است.

3.2.6. امتیاز دهی به جنگل تصمیم

در الگوریتم 3، ویژگی ها برای اولین بار توسط نسبت بهره گرفته شده اند. ویژگی های K در رأس همه آن ها پس از ساخت درختان K توسط C4.5 انتخاب شده است [51]، در ادامه درختان K به عنوان یک کمیته برای احتمال تقلب در هر معامله خدمت می کنند. مجموع احتمال که خروجی نمره خطر از مدل در الگوریتم 4 خواهد بود. ارائه رتبه معامله t به دستور خروجی نمره خطر از جنگل تصمیم است که $\mathbb{S}(t)_{DF}$ می شود.

Algorithm 3 DecisionForest

```
Data: TransactionDB,k
Result: Decision forest
Sort the features by information gain into hash table H;
for each feature F_i in top k of H do
    Build decision tree T_i with F_i as root node;
    Save T_i into tree table TT;
return TT;
```

Algorithm 4 PredictionByDecisionForest

```
Data: TransDB/* transactions for prediction*/,Decision Forest TT
Result: Risk score
for each transaction t in TransDB do
    ScoreSum=0;
    for each tree T_i in TT do
        ScoreSum=ScoreSum+(score of t by given by T_i);
    Output ScoreSum as risk score for t;
return TT;
```

3.6. ریسک با استفاده از مدل ترکیبی

ترکیب مناسب از مدل های متعدد به طور موثر می تواند به سمت عملکرد تجمعی بهتر [13] اهرم قدرت از هر جزء. در تشخیص تقلب بانکداری آنلاین برود، روش های متعدد ممکن است نمره های مختلف برای معامله داشته باشند، و ما با استفاده از یک بردار وزن جمع بندی می کنیم. نمره نهایی $\mathbb{S}(t)$ را به شرح زیر محاسبه می شود:

$$\mathbb{S}(t) = w_1 * \mathbb{S}(t)_{CP} + w_2 * \mathbb{S}(t)_{CNN} + w_3 * \mathbb{S}(t)_{DF} \qquad (6.6)$$

در اینجا، t پیش بینی معامله است، $w_i \ (i = 1, 2, 3)$ وزن مدل $i - th$ است.

ما یک وزن برای هر مدل با توجه به دقت پیش بینی در داده ها از آزمون قرار می دهیم. برای مثال، مدل های 1 و 2 و وزن مدل اول 0.8 است در حالی که مدل دوم 0.2 است ، که نمره نهایی از معامله نمره جمع خواهد بود.

7. آزمایش و ارزیابی

سیستم i-Alertor در شکل 4 نشان داده شده است. این تکنولوژی از سه مدل (کنتراست ماینستر، CNN و جنگل تصمیم گیری) برای تشخیص تقلب در بانکداری آنلاین ساخته شده است.

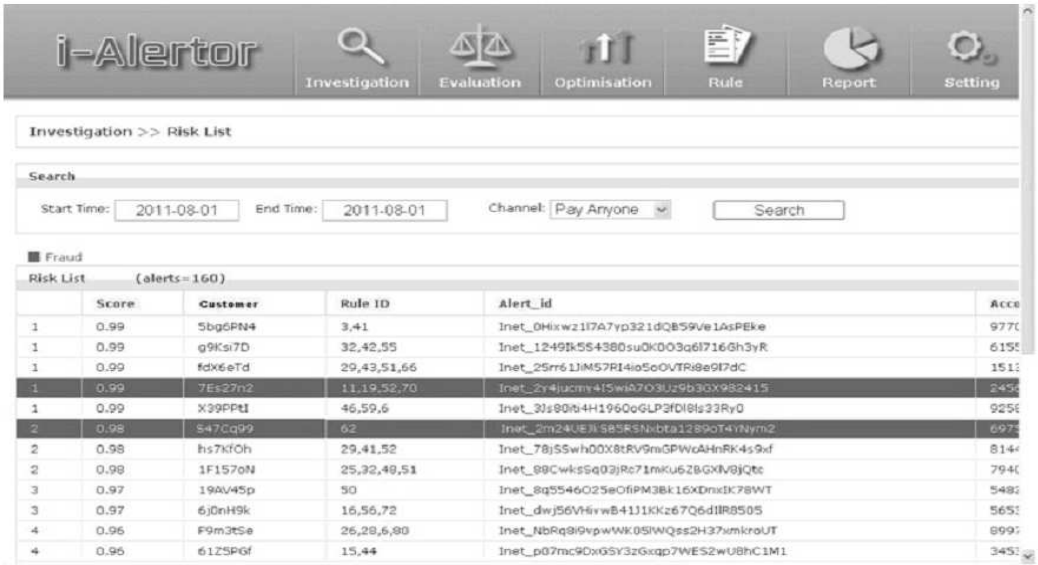
هدف از ارزیابی تجربی و روش پایه مربوطه عبارتند از:

(1) برای مقایسه مدل خطر همراه با سیستم های موجود (سیستم خبره) مورد استفاده در بانک های بزرگ در استرالیا؛

(2) برای مقایسه عملکرد کنتراست ماینستر با الگوریتم های موجود $\text{MDB-LL}_{\text{border}}$ [24] از دیدگاه مقیاس پذیری داده ها، انطباق ابعاد و تحمل عدم تعادل است.

1.7. اطلاعات

مجموعه داده های مورد استفاده در آزمایش ما اطلاعات بانکی کاربردی به یک بانک بزرگ در استرالیا است. که 8،000،000 معاملات واقعی (D_g) و 1500 معامله تقلبی (D_f) را شامل می شود، و تعدادی ویژگی های آن 130 است.



شکل 4. سیستم آنلاین مدیریت ریسک بانکی: i-Alertor.

2.7. تنظیمات تجربی

سیستم I-Alertor علیه سیستم آنلاین تشخیص تقلب بانکی مبتنی بر سیستم خبره ارزیابی شده است. دو معیارهای اصلی برای ارزیابی عملکرد یک سیستم تشخیص تقلب بانکداری آنلاین وجود دارد.

- حجم هشدار ، که تعدادی از هشدار با توجه به روند کسب و کار در بانک هر روز تولید می شود. ، هر هشدار به صورت دستی برای فرآیند بیشتر مورد بررسی قرار می گیرد. تعداد زیادی از هشدارهای تحقیقات کاملا نیاز به کار فشرده و وقت گیر دارند، و در نتیجه هزینه آن ها بالا است.

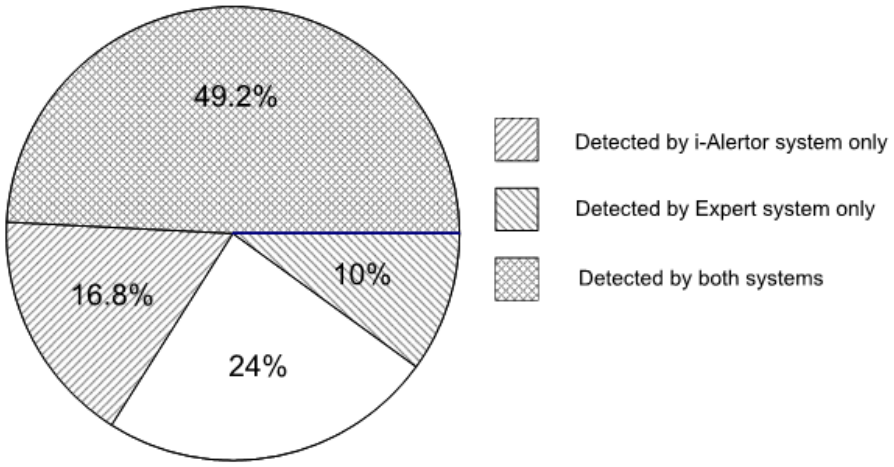
- نرخ تشخیص که درصد تقلب شناسایی شده توسط سیستم است. مورد مناسبی که در آن سیستم می تواند تمام تقلب ها را بگیرد، به این معنی که نرخ تشخیص 100٪ است.

کنتراست ماینستر و MDB-LL_{border} هر دو الگوریتم استخراج الگوی مقابل هستند، بنابراین دقت آنها همان است که به مجموعه داده های مشابه استفاده می شود. به منظور ارزیابی بهره وری کنتراست ماینستر در برابر MDB-LL_{border} ، زمان محاسباتی صرف شده توسط کنتراست ماینستر و MDB-LL_{border} را مقایسه می کنیم. مقایسه عملکرد در هر دو الگوریتم ما را به همان سطح از نرخ متضاد می رسانند.

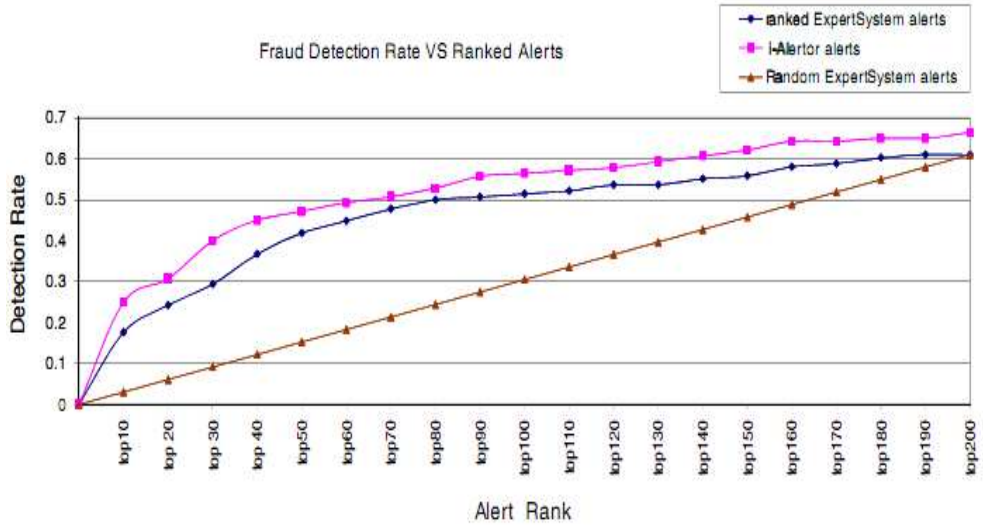
3.7. ارزیابی عملکرد کلی

شکل 5 یک مرور کلی از توزیع تقلب گرفتار شده توسط سیستم خبره و سیستم I-Alertor است. در اینجا I-Alertor بالای 200 پرخطرترین معاملات هر روز است. شکل نشان می دهد که 49.2 درصد از تقلب توسط هر دو سیستم تشخیص داده شده ، در حالی که سیستم ما می تواند 16.8٪ اضافی تر از تقلب ها را تشخیص دهد. در مجموع، در سیستم ما نرخ تشخیص حدود 7 درصد بالاتر است.

شکل 6 سرعت کشف و شناسایی هشدار مختلف تولید شده توسط سیستم های خبره است، سیستم خبره تحت هشئار قرار گرفته است، و به عنوان مثال i-Alertor هیچ رتبه ای در میان هشدار اصلی از سیستم خبره، برای به دست آوردن بالای هشدار N ندارد، هشدار به صورت تصادفی از هشدارهای روزانه سیستم خبره انتخاب شدند. برای هشدار رتبه بندی شده از سیستم خبره، هشدار با توجه به رتبه نمره خطر خود را محاسبه می کند



شکل 5 توزیع تقلب شناسایی شده توسط سیستم های مختلف.



شکل 6 مقایسه نرخ تشخیص هشدار بین i-Alertor و سیستم خبره.

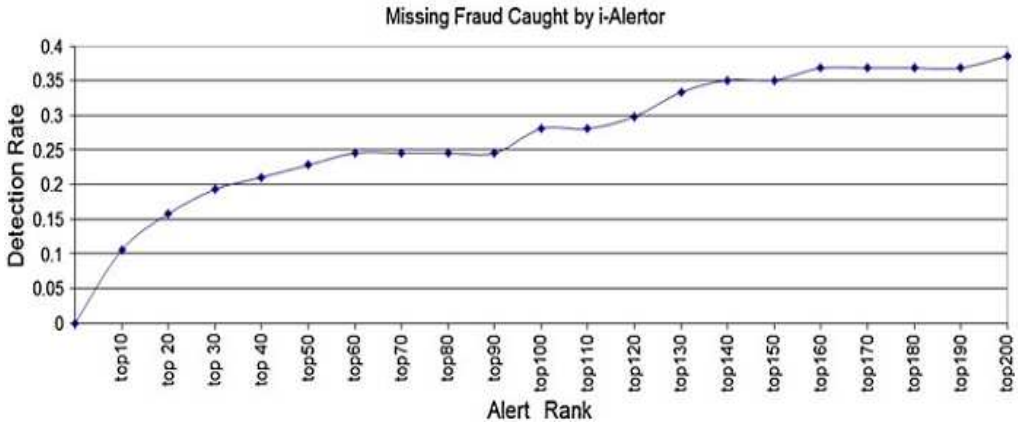
در شکل 6، ما می توانیم ببینیم که سیستم i-Alertor با قرار گرفتن در هشدار خیلی بهتر از هشدار اصلی سیستم خبره عمل می کند. i-Alertor حتی بهتر از سیستم خبره با هشدار کار می کند. در هشدار با حجم زیاد همین، i-Alertor گاهی اوقات حتی یک نرخ تشخیص 10٪ بالاتر دارد. به طور کلی i-Alertor همیشه عملکرد بهتری نسبت به سیستم خبره ارائه می دهد. این است که عمدتا به دلیل پویایی تقلب و i-Alertor می توانید درصد بالایی از تقلب جدید که توسط قوانین در سیستم خبره از دست رفته را بگیرد.

شکل 7 ارزیابی i-Alertor در تشخیص تقلب است. نشان می دهد که i-Alertor می تواند 25٪ از تقلب توسط سیستم خبره در عرض 60 هشدار از دست رفته را بگیرد. بنابراین، اگر ما سیستم خبره را با i-Alertor ترکیب کنیم ، i-Alertor می تواند به سیستم خبره جهت تشخیص تقلب از دست رفته از طریق افزایش حجم هشدار کمک کند.

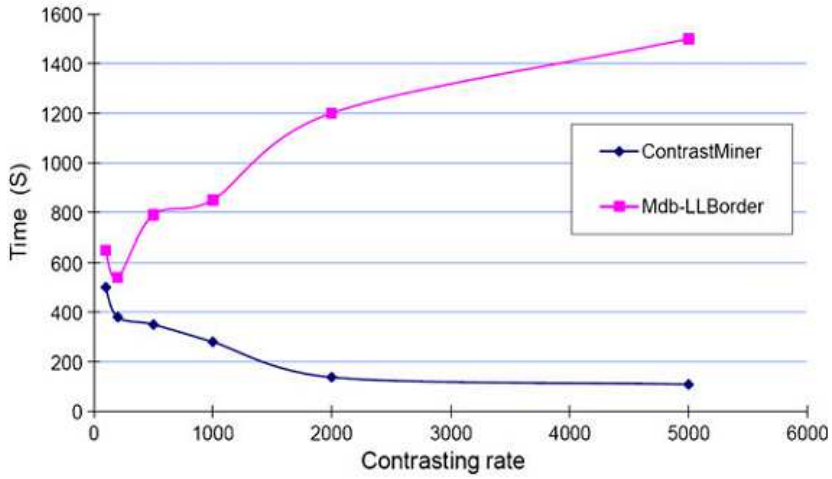
4.7. عملکرد رفتار مدل سازی کنتراست

در نهایت، ما عملکرد کنتراست ماینستر در رفتار مقابل ماینر در داده های نا متوازن را ارزیابی می کنیم.

برای بهره وری $MDB\text{-}LL_{border}$ ، شکل 8 نشان می دهد که، با افزایش نرخ متضاد، کنتراست ماینستر بسیار بهتر از $MDB\text{-}LL_{border}$ عمل می کند. به دلیل اینکه



شکل7. نرخ تشخیص تقلب گمشده از i-Alertor .



شکل8. بهره وری از کنتراست ماینستر در مقابل $MDB\text{-}LL_{border}$.

برای MDB-LL$_{border}$ به الگوی ماینر مرزی وقت گیر است به ویژه هنگامی که آستانه پشتیبانی D_g بسیار نامتوازن و بسیار کوچک است.به عنوان مثال، نرخی که تضاد 2000 را تعیین می کند، به معنی است که MDB-LL$_{border}$ نیاز به کشف تمام الگوهای پشتیبانی <0.0005 دارد، که برای MDB-LL$_{border}$ در موفقیت و در هزینه زمان قابل قبول است. با این حال، کنتراست ماینستر به این دلیل که حمایت نهفته حداقلی دارد هنوز هم می تواند به خوبی کار کند. همانطور که در بخش 5.2 ذکر شد، پشتیبانی کوچک را می توان یک ارزش عملی دانست که در آن ماینر می توانید به راحتی افزایش یابد. بنابراین، با افزایش نرخ مقابل، D_g پردازش می شود و به سرعت در حال کاهش است، و در نتیجه هزینه MDB-LL$_{border}$– مقدار زیادی از زمان پردازش در مرزهای الگوی خروجی را تشکیل می دهد. از سوی دیگر، کنتراست ماینستر دارای زمان کمتر با نرخ کنتراست بالاتر است، چرا که الگوهای کمتری برای انتخاب در منطقه ذوزنقه ای ABDE وجود دارند.

8. نتیجه گیری

تقلب در بانکداری آنلاین شامل منابع متعدد، از جمله خرد انسان، ابزار محاسباتی، تکنولوژی وب و سیستم های تجارت آنلاین است. تشخیص فوری و موثر تقلب روش های تشخیص تقلب های موجود مانند سیستم ها را به چالش می کشد. در این مقاله، ما به مطالعه شیوه ها در دنیای واقعی پرداخته ایم. یک روش تشخیص تقلب بانکداری آنلاین نظام مند معرفی شده است. چارچوب آن استفاده از دامنه دانش ، ویژگی های مخلوط، روش های متعدد داده کاوی، و یک ساختار چند لایه برای یک راه حل نظام مند است. این شامل سه الگوریتم: الگو ماینستر، شبکه عصبی و جنگل تصمیم است، و نتایج با نمره کلی اندازه گیری ریسک در یک معامله آنلاین جعلی و یا واقعی یکپارچه شده است. رویکرد آن ها در کشف تقلب در حجم زیادی از داده های بسیار نامتعادل موثر است. ما روش و سیستم را در یک بانک بزرگ تست کرده ایم. آزمایش های عظیم نشان می دهد که چارچوب ما به طور قابل توجهی بهبود دقت و صحت تشخیص تقلب را به همراه داشته و روش های موجود در تشخیص تقلب ها و سیستم در کارایی و دقت موثر بوده . همچنین این روش می تواند با سیستم تشخیص تقلب بانکی موجود ترکیب شود.