

تشخیص تقلب در کارت‌های اعتباری با استفاده از الگوریتم گرگ خاکستری و شبکه عصبی مصنوعی

فرهاد کریم خانی¹، وحید حاجی زاده²

¹ کارشناسی فناوری اطلاعات، دانشگاه آزاد اسلامی واحد بین المللی ماکو، بانک رفاه کارگران، ارومیه، ایران

refah14544@gmail.com

² دکترای حسابداری، دانشگاه آتاتورک ترکیه، بانک رفاه کارگران، ارومیه، ایران

Vahid_7796@yahoo.com

چکیده - در دنیای امروزی با توجه به اینکه گسترش روزافزون تجارت الکترونیک فرصت‌های بسیار خوبی برای تجارت جهانی ایجاد نموده، تهدیداتی نیز برای بانک‌ها و موسسات مالی و اعتباری فراهم داشته است که در این میان، تقلب در کارت‌های اعتباری، تقلب در بانکداری الکترونیک و پول شویی توجه و نگرانی‌های زیادی را به سمت خود جلب کرده است. در معاملات بانکی، اطمینان و اعتبار برای صاحبان کارت اعتباری از اهمیت بالایی برخوردار است طوریکه افراد بتوانند با خیال راحت پرداخت‌های الکترونیکی خود را انجام دهند. یک سیستم تشخیص کلاهبرداری باید بین معامله مشروع، کلاهبرداری مشکوک و معامله نامشروع تمایز قائل شود. در این مقاله یک روش تشخیص تقلب در کارت اعتباری با کمک الگوریتم بهینه سازی شده گرگ خاکستری و بر مبنای شبکه عصبی با هدف حل مشکلات همگرایی، بهینگی محلی، نقص شبکه و پایداری سیستم ارائه شده است. ما در روش پیشنهادی خود برای بهینه‌سازی وزن شبکه عصبی، ابتدا از الگوریتم گرگ خاکستری برای بدست آوردن مقدار اولیه بهینه استفاده میکنیم و سپس از الگوریتم شبکه عصبی برای تصحیح مقدار خطا استفاده میکنیم تا از این طریق مقدار بهینه را بدست آوریم. نتایج آزمایشگاهی با مدل‌های تصادفی نشان می‌دهد که روش پیشنهادی به شناسایی فعالیت‌های کلاهبرداری به طور موثر کمک می‌کند و تولید هشدارهای دروغین را به حداقل می‌رساند که باعث صرفه جویی در زمان و بهینه شدن نتیجه می‌گردد.

کلیدواژه- بانکداری هوشمند، هوش مصنوعی، شبکه عصبی مصنوعی، گرگ خاکستری

مقدمه

تجارت در سراسر دنیای امروزی هستیم. تجارت الکترونیک بر مبنای پرداخت‌های الکترونیک برنامه ریزی شده است. به طور کلی پرداخت کارت بانکی و اعتباری به دو شکل آنالیت و آفلاین اتفاق می‌افتد. بنابراین خطرات کلاهبرداری نیز در هر دو حال وجود خواهد داشت. کلاهبرداری یک مشکل اساسی برای بازرگانان، به ویژه در بخش آنالاین است [1،2]. کاربران نیز با رعایت اصول امنیتی می‌توانند خطر کلاهبرداری را تا حد بالایی کاهش دهند، یکی از مهم‌ترین موارد پیشگیری استفاده از تایید دو مرحله‌ای و رعایت اصول حفاظت فیزیکی (عدم تحویل رمز و کارت به فروشندگان) می‌باشد. سالانه میلیون‌ها دلار ضرر و زیان بر اثر تقلب در پرداخت‌های الکترونیک به کاربران، صاحبان تجارت و بانک‌ها تحمیل می‌شود. برای پردازش بهتر نیاز به درک تکنیک‌های سایبری کلاهبرداری از کارت اعتباری داریم. کلاهبرداری از کارهای بانکی به دو شکل آنالاین و آفلاین اتفاق می‌افتد. در تقلب آنالاین از اطلاعات کارت اعتباری برای

با افزایش روزافزون گردش‌های مالی و گسترش تجارت بین‌المللی، هزینه تولید، نگهداری و تبدیلات ارزی برای کشورها بار گزافی را در پی دارد. برای همین در کنار روند رو به رشد، فناوری تشکیل مفاهیم در برگیرنده پول مجازی با استفاده از ابزار الکترونیکی، باعث پیدایش تحولی در صنعت بانکداری و گرایشی با عنوان تجارت الکترونیکی گشته است. با توجه به گسترش تبادلات بانکی از طریق کارت‌های بانکی و کلاه برداری‌های فراوان از بانک‌ها و مردم، تشخیص تقلب و شناسایی راه‌های پیش‌گیرانه حائز اهمیت است که بوسیله این الگوریتم درصد قابل ملاحظه‌ای از آن قابل تشخیص می‌باشد. یکی از مهم‌ترین مسایل معرفی ایده برای کاهش زمان و سرعت بخشیدن به شناسایی کلاه بردار و پیشگیری از تخلفات بیشتر می‌باشد. با شکوفا شدن تجارت الکترونیک و فعالیت‌های حرفه‌ای سایت‌های فروش به ویژه در شبکه‌های اجتماعی شاهد رضایت روزافزون کاربران از این

مختلف با استفاده از معیارهای مختلف بر روی مجموعه داده‌های نامتعادل انجام شده است. این مطالعه همچنین هدف دارد تا برخی الگوریتم‌های تشخیص کلاهبرداری کارآمد را پیشنهاد کند که دقت را بهبود می‌بخشند و مثبت‌های کاذب را کاهش می‌دهند. با رشد تجارت الکترونیک و روش‌های پرداخت آنلاین، تعداد تراکنش‌های تقلبی نیز افزایش یافته است. موسسات مالی با سیستم‌های پرداخت آنلاین باید از سیستم‌های تشخیص تقلب خودکار برای کاهش خسارات ناشی از فعالیت‌های تقلبی استفاده کنند. مشکل تشخیص تقلب معمولاً به عنوان یک مدل طبقه‌بندی دوتایی فرموله می‌شود که می‌تواند تراکنش‌های تقلبی را تشخیص دهد. جاسازی داده‌های ورودی مجموعه داده تقلب در یک نمایش کم‌بعدتر برای ساخت سیستم‌های تشخیص تقلب قوی و دقیق بسیار مهم است. مقاله [5] یک چارچوب دو مرحله‌ای را برای تشخیص تراکنش‌های تقلبی پیشنهاد می‌دهد که شامل یک خودرمنگار عمیق به عنوان یک روش یادگیری نمایش و تکنیک‌های یادگیری عمیق نظارت‌شده است. ارزیابی‌های آزمایشی نشان داد که روش پیشنهادی عملکرد طبقه‌بندی‌کننده‌های مبتنی بر یادگیری عمیق مورد استفاده را بهبود می‌بخشد.

به طور خاص، طبقه‌بندی‌کننده‌های یادگیری عمیق مورد استفاده که روی مجموعه داده تبدیل شده توسط خودرمنگار عمیق آموزش دیده‌اند، از نظر تمام معیارهای عملکرد، به طور قابل توجهی نسبت به طبقه‌بندی‌کننده‌های پایه خود که روی داده‌های اصلی آموزش دیده‌اند، عملکرد بهتری دارند. علاوه بر این، مدل‌های ایجاد شده با استفاده از خودرمنگار عمیق نیز نسبت به مدل‌های ایجاد شده با استفاده از مجموعه داده به دست آمده از تحلیل مؤلفه‌های اصلی (PCA) و مدل‌های موجود، عملکرد بهتری دارند. رویکرد پیشنهادی مقاله [6] بر اساس دو مرحله ساخته شده است. مرحله اول با هدف معرفی بهترین سه الگوریتم یادگیری ماشین از بین نه الگوریتم است. مرحله دوم با هدف ادغام بهترین سه الگوریتم با نوزده تکنیک نمونه‌برداری مجدد است. هر مدل در هر دو مرحله بر اساس مساحت زیر منحنی عملکرد گیرنده (AUC)، دقت، فراخوانی، دقت و امتیاز F1 ارزیابی می‌شود. در مرحله اول، نه الگوریتم رگرسیون لجستیک (LR)، همسایگی‌های نزدیکترین K (KNN)، درخت تصمیم (DT)، بیز ساده‌لوح (NB)، جنگل

خرید الکترونیکی کالاها و خدمات را بر روی اینترنت استفاده می‌شود. در کلاهبرداری آنلاین از کارت‌های سرقت شده یا تقلبی (کپی اطلاعات اصلی بر روی یک کارت جعلی) در یک فروشگاه محلی و یا یک مکان فیزیکی برای خرید برخی از کالا و یا خدمات استفاده می‌گردد. در جامعه‌ی امروزی با توجه به گسترش معاملات و تراکنش‌های بانکی و با توجه به وجود انسان‌های سودجو و خلافکار در همه‌ی زمینه‌ها و وجود ضعف‌های شدید در سیستم‌های انتقال وجه و اطلاعات ناکافی کاربران در حفاظت از رمزها و اطلاعات خود در پرداخت‌ها به ویژه در زمینه‌ی کارت‌های بانکی و نقل و انتقالات، اینترنتی بانک‌ها و مردم متحمل ضرر و زیان‌های هنگفتی در این نوع تراکنش‌ها می‌شوند. در برخی از کشورها با استفاده از الگوریتم‌های داده‌کاوی یادگیری ماشین و یا شناسایی تراکنش‌های ناهنجاری این تخلفات را به حداقل می‌رسانند. با این حال میلیاردها دلار ضرر و زیان هر ساله در اثر تراکنش‌های متقلبانه کارت اعتباری ایجاد می‌شود [3، 5]. طراحی الگوریتم‌های کارآمد برای تشخیص کلاه برداری برای کاهش این تلفات مهم است. 2- ادبیات تحقیق و کارهای گذشته در زمینه تشخیص تقلب در کارت اعتباری تاکنون مقالات متعددی ارائه شده است که در این قسمت به جدیدترین و مرتبط‌ترین آن‌ها پرداخته خواهد شد. استفاده از کارت اعتباری روز به روز در حال افزایش است و نقش مهمی در تجارت الکترونیک ایفا می‌کند. کلاهبرداری با کارت اعتباری (CCFs) به دلیل افزایش استفاده از کارت‌های اعتباری، اکنون رایج‌تر شده است. موسسات مالی و دارندگان کارت با خسارات مالی مواجه هستند. چالش اصلی در تشخیص کلاهبرداری با کارت اعتباری، ماهیت نامتعادل مجموعه داده‌های موجود است. مجموعه داده‌های موجود اطلاعات بسیار کمتری در مورد تراکنش‌های کلاهبرداری نسبت به تراکنش‌های مشروع دارند. روش‌های مختلف مبتنی بر یادگیری ماشین مانند جنگل تصادفی، SVM، KNN، ANN و برخی روش‌های ترکیبی در ادبیات موجود است. با این حال، به دلیل دقت پایین و نرخ بالای مثبت کاذب تکنیک‌های یادگیری ماشین، الگوریتم‌های یادگیری عمیق همراه با روش‌های نمونه‌برداری مختلف باید برای کاهش خسارات مالی اعمال شود. در مقاله [4]، تجزیه و تحلیل گسترده روش‌های ML و DL برای ارزیابی در مورد دقت طبقه‌بندی‌های

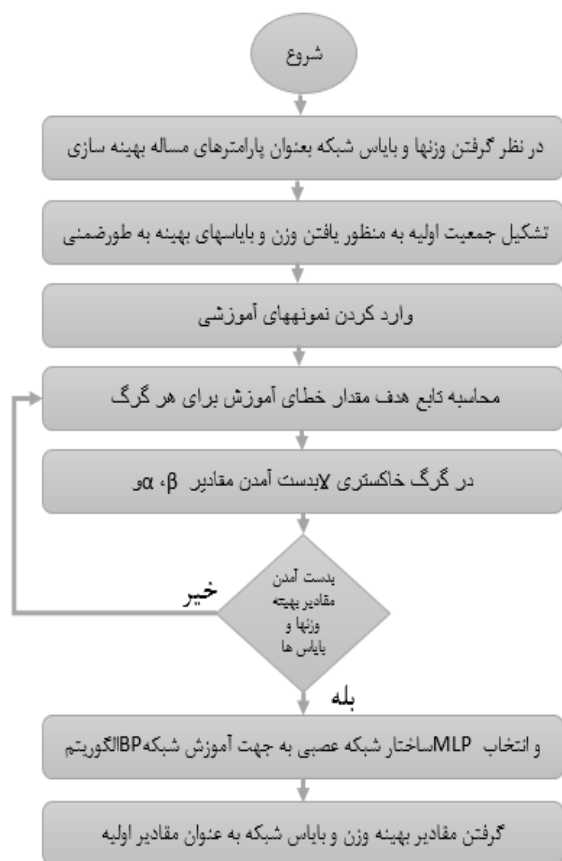
گذشت مدت زمانی الگوهایی را برای خود در نظر می گیرد و استفاده از این سیستم ما را در طبقه بندی بهتر یاری می رساند، در این مقاله از شبکه های عصبی چند لایه پیشخور استفاده شده است الگوریتم یادگیری پس انتشار خطا یکی از رایج ترین الگوریتم ها جهت آموزش شبکه عصبی MLP می باشد [9،11]. حجم زیاد از دادگان و شباهت بین آنها منجر به انباشتگی زمان و دقت کمتر روی نمونه رفتارهای سالم و ناسالم توسط طبقه بندی های سنتی می شود [7].

از این رو ما از الگوریتم های فراابتکاری به منظور بهبود عملکرد در شبکه و کاهش پیچیدگی های محاسباتی در مقایسه با روش حداقل مربعات و گرادیان نزولی استفاده می کنیم. همگرایی الگوریتم BP به انتخاب مقادیر اولیه وزن های شبکه بردارهای بایاس و پارامترهای موجود در الگوریتم مانند نرخ یادگیری وابسته است. این الگوریتم ممکن است به نقاط مینیمم محلی در فضای پارامتر، همگرا شود. بنابراین زمانیکه الگوریتم BP همگرا شود، نمیتوان مطمئن شد که به یک جواب بهینه رسیده باشیم. پس با توجه به اهمیت موضوع امروزه ضرورت وجود الگوریتم بهینه سازی که بتواند بر اساس چندین پارامتر و قرار دادن در تابع، شایستگی مقادیر اولیه مناسب وزنی و بایاس را برای اجرای بهتر شبکه، انتخاب کند به شدت احساس میشود. در مرور مقالات در بخش قبل دیدیم که الگوریتم های بهینه سازی برای اینکار وارد عمل شدند ولی هنوز روشی که به تنهایی و بدون ترکیب با الگوریتم های بهینه سازی محلی با پارامترهای کمتر برای سرعت بخشیدن به الگوریتم در مقالات پیشین ارائه نشده است. همچنین آن الگوریتم هایی که به تنهایی برای مسئله بهینه سازی استفاده شده است یا دچار کندی سرعت در اواسط اجرا می شوند و یا تنظیم پارامترهایشان به شبکه عصبی دلیل زیاد بودن سخت می باشد و این معایب در مسائلی همچون تشخیص کلاهبرداری در کارت های اعتباری که سرعت و دقت برایش حائز اهمیت می باشد تاثیر گذار می باشد. دیگر روش های موجود، محاسبات زیادی به جهت حل مسئله ذکر شده دارند و زمانیکه حجم اطلاعات مسئله بالا می رود قادر به طبقه بندی دقیق و با سرعت بهینه نمی باشند لذا ما در این مقاله روشی مبتنی بر هوش مصنوعی بر پایه الگوریتم بهینه سازی گرگ خاکستری که از جمله روش های ساده و دارای تنظیم پارامترهای بسیار کم ولی با دقت بالا می

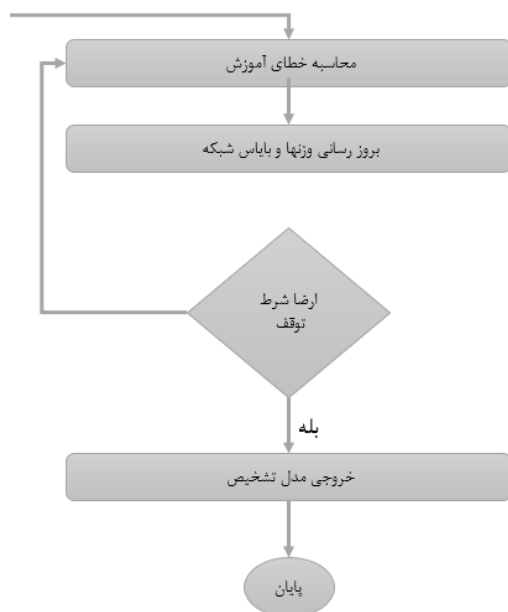
تصادفی (RF)، ماشین های تقویت گرادیان (GBM)، ماشین تقویت گرادیان سبک (LightGBM)، تقویت گرادیان شدید (XGBoost)، و تقویت دسته (CatBoost) هستند. بهترین مدل از بین همه این ها AllKNN همراه با CatBoost (AllKNN-CatBoost) است. در نهایت، AllKNN-CatBoost با کارهای قبلی با همان مجموعه داده و رویکردهای مشابه مقایسه می شود. در واقع، AllKNN-CatBoost در زمینه AUC (97.94٪)، فراخوانی (95.91٪) و امتیاز (87.40٪ F1) از مدل های قبلی بهتر عمل می کند. در مطالعه [13]، یک مدل به نام IF-GWO ارائه شده است که از طریق تحلیل تراکنش های گذشته، الگوهای تقلب را یاد می گیرد. این روش از یک روش جدید یادگیری آنسامبل با استفاده از جنگل انزوا (IF) و بهینه سازی گرگ خاکستری (GWO) استفاده می کند. نتایج تجربی نشان دهنده اولویت سیستم تشخیص تقلب ارائه شده بر اساس تعداد قابل توجهی از تراکنش های حساب کارت اعتباری است. در مقایسه با مدل های متداول مورد استفاده برای تشخیص ناهنجاری، مدل پیشنهادی این مقاله می تواند با خطای مثبت کاذب کمتر، حساب های تقلبی بیشتری را تشخیص دهد. بر اساس مقایسه با سایر مدل ها با استفاده از مجموعه داده ای شامل 284,807 تراکنش انجام شده توسط دارندگان کارت اروپایی، مدل پیشنهادی این مقاله عملکرد بهتری نسبت به سایر رویکردها داشته و بالاترین عملکرد را از نظر F-Measure (93.52٪)، مساحت زیر منحنی عملکرد گیرنده (94.17٪ AUC) و میانگین (94.10٪ G) به دست آورده است.

3- روش پیشنهادی

همان طور که در ابتدا اشاره کردیم با مطالعه در مرور مراجع دیدیم که با وجود تنوع کارهای انجام شده در زمینه مسئله تشخیص کلاهبرداری در کارت های اعتباری همچنان نیاز به پیدا شدن راه حل بهینه که با سرعت بالاتر و دقت بیشتر روند تشخیص را بهبود دهد، احساس میشود [7،8]. به دنبال حل این مسئله، استفاده از شبکه های عصبی برای این منظور نسبت به دیگر روش های طبقه بندی نتایج بهتری از خود نشان داده است. ما به دنبال پیدا کردن الگوهایی برای بهبود رفتارهای دارندگان کارت های اعتباری هستیم و چون در سیستم های شبکه عصبی روند یادگیری وجود دارد و بعد از



شکل 1: فلوچارت روش پیشنهادی



شکل 1: فلوچارت روش پیشنهادی

باشد، استفاده کرده ایم که برای اولین بار در حل این مساله استفاده خواهد شد. این الگوریتم از دقت و مقاومت خوبی و همچنین سرعت مناسبی در برابر مواجه شدن با مسائل با ابعاد بالا مانند تراکنش های بانکی برخوردار می باشد. از این رو تفاوت اصلی روش پیشنهادی با روش پایه [13] در زیر آورده شده است:

- ✓ کاهش احتمال همگرایی به مینیمم های محلی: الگوریتم های فراابتکاری با توانایی جستجوی گسترده در فضای پارامترها، احتمال همگرایی به مینیمم های محلی را کاهش می دهند.
- ✓ بهبود عملکرد شبکه: با پیدا کردن مقادیر اولیه بهینه برای وزن ها و بایاس ها، می توان دقت و سرعت شبکه را بهبود بخشید.
- ✓ کاهش پیچیدگی محاسباتی: برخی از الگوریتم های فراابتکاری نسبت به الگوریتم BP ساده تر هستند و می توانند به کاهش پیچیدگی محاسباتی کمک کنند.

3-1- جزئیات مراحل الگوریتم پیشنهادی

ما در این مقاله به مسئله تشخیص کلاهبرداری از کارت اعتباری پرداخته ایم. الگوریتم پیشنهادی با نام تشخیص کلاهبرداری کارت اعتباری بهینه شده توسط الگوریتم بهینه سازی گرگ خاکستری می باشد. روال کار بدین صورت می باشد که در ابتدا با الگوریتم بهینه سازی گرگ خاکستری به دنبال بهینه ترین مقادیر برای وزن ها و بایاس های روش BP باشیم [10، 12]، چون تاثیر زیادی در روند یادگیری شبکه دارند، سپس عملکرد شبکه MLP بهینه شده را بر روی مجموعه داده ای که در نظر گرفته شده اجرا شود. در روش پیشنهادی به دلیل سادگی و کم بودن پارامترهای الگوریتم به جهت تنظیم باعث شده که نسبت به دیگر روش های قبلی از پایداری بیشتر و سرعت بالاتر در رسیدن به تشخیص طبقه بندی بهینه و سربار کمتری برخوردار باشد.

در این بخش ما به جزئیات گام های روش پیشنهادی پرداخته ایم. فلوچارت کلی آن در شکل 1 نشان داده شده است.

جزئیات روش پیشنهادی

روش کار بدین صورت است که ابتدا جمعیت اولیه ای از وزن و بایاس ها به صورت تصادفی توسط الگوریتم گرگ خاکستری ایجاد شده و خروجی آن باعث بدست آمدن وزن و بایاس های بهینه بعنوان مقادیر اولیه برای شروع آموزش شبکه می باشند. این روند تا پایان رسیدن به جواب بهینه تکرار خواهد شد. در ادامه به جزئیات مراحل الگوریتم پیشنهادی پرداخته ایم.

3-2- گام های روش پیشنهادی

گام های روش پیشنهادی به شرح زیر می باشد:

گام اول:

در این گام مقداردهی پارامترهای مسئله و الگوریتم پیشنهادی به صورت تصادفی صورت می گیرد.

✓ تعیین شمارنده تکرار الگوریتم، تعیین کردن تعداد افراد در جمعیت و تعیین تعداد وزن و بایاس ها و تعداد لایه های ورودی و پنهان و خروجی.

✓ تعیین کردن مقدار برای پارامتر $a, r1, r2$ که $r1$ و $r2$ مقادیر تصادفی بین ۰ و ۱ هستند و a یک مقدار خطی از ۲ به ۰ در نظر گرفته می شود. این پارامترها به عنوان مقادیری در ضرایب ثابت بردارها A و C که برای مدل ریاضی حلقه زدن دور شکار در نظر گرفته می شود. همچنین این پارامترها نقش تعادل بین اکتشاف و استخراج را ایفا میکنند.

✓ تعیین کردن حداکثر تعداد تکرار اجرای الگوریتم پیشنهادی

گام دوم:

در این گام پخش افراد جمعیت که هر کدام از آنها نشان دهنده موقعیت ضمنی وزن و بایاس ها که به صورت تصادفی در فضای مسئله میباشد.

گام سوم:

در این گام نمونه آموزشی را به شبکه وارد میکنیم.

گام چهارم

در این گام محاسبه تابع هدف (مقدار خطا شبکه) هر نفر در جمعیت انجام می شود و بهترین فرد از لحاظ داشتن کمترین مقدار تابع هدف در متغیری ریخته می شود.

گام پنجم :

بدست آوردن مقادیر α, γ, β در هر تکرار الگوریتم گرگ خاکستری

گام ششم:

در این گام چک می شود آیا مقادیر وزن و بایاس ها به اندازه کافی بهینه شده اند، اگر جواب مثبت است رفتن به گام هشتم در غیر این صورت رفتن به گام چهارم.

گام هفتم

تشکیل معماری شبکه عصبی MLP و انتخاب الگوریتم BP به جهت فرایند آموزش شبکه.

ساختار شبکه در این مقاله به این صورت است که دو لایه ورودی، ۲۰ لایه پنهان و دو لایه خروجی داریم. تابع انتقال در لایه مخفی، تابع sigmoid است. تعداد دفعات آموزش شبکه ۱۰۰ بار و نرخ یادگیری ۰.۱ در نظر گرفته ایم.

گام هشتم

گرفتن مقادیر بهینه وزن و بایاس بعنوان مقادیر اولیه به جهت آموزش.

گام نهم

در این گام خطای آموزش BP محاسبه میشود.

گام دهم

در این گام بروز رسانی وزن و بایاس های شبکه صورت می گیرد.

گام یازدهم

آموزش زمانی متوقف می شود که اختلاف خطای میانگین بدست آمده در دو دوره (EPOCH) متوالی به اندازه کافی کوچک باشد در غیر این صورت رفتن به گام نهم انجام می شود.

گام دوازدهم

در این گام خروجی مدل آموزش داده شده تولید می شود. در بخش قبلی به روش پیشنهادی با جزئیات کامل پرداخته شد. در این بخش شبیه سازی های روش پیشنهادی بر روی مجموعه داده Kaggle بازگو می شود. همچنین روش پیشنهادی ذکر شده را با روشهای Whale و PSO و GA در این مسئله مقایسه خواهیم کرد.

4- شبیه سازی ها

برای یک شبکه MLP برابر با تعداد ویژگی های مجموعه داده مورد نظر می باشد و تعداد نورون های خروجی برابر با تعداد کلاس های در آن مجموعه داده است. توقف آموزش زمانی صورت میگیرد که میانگین مربعات خطا به صفر یا تعداد EPOCH به حداکثر مقدار از قبل تعیین شده، رسیده باشد. از تابع انتقال sigmoid برای لایه های مخفی استفاده کردیم. همچنین داده را به دو دسته آموزش و آزمایش تقسیم بندی کردیم. در ابتدا شبکه با دادگان آموزشی آموزش می بیند. سپس صحت ارزیابی دادگان آزمایشی ارزیابی می شود. تعداد بار آموزش شبکه را ۱۰۰ بار و نرخ یادگیری را 0.1 در نظر گرفتیم. در جدول 1 مجموعه داده Kaggle آورده شده است. در جدول 2 توزیع داده ها آورده شده است. مشخصات لازم به ذکر است که نتایج بدست آمده در جداول با میانگین ۲۵ بار اجرای مستقل روش پیشنهادی و دیگر روش های ذکر شده می باشد. در ادامه نتایج و ارزیابی ها آورده شده است. ما در روش پیشنهادی توانستیم تعداد تکرارها و طبقه بندی شدن معاملات نرمال نسبت به معاملاتی که در آنها کلاهبرداری صورت گرفته را با دقت بالاتر انجام دهیم.

جدول 1: مجموعه داده Kaggle

تعداد ویژگی	30
تعداد کلاس	2
تعداد نمونه ها	284807

جدول 2: توزیع داده ها به داده آموزشی و داده آزمایشی

تعداد معاملات کلاهبرداری شده	تعداد معاملات نرمال	تعداد	تعداد داده
450	1550	2000	تعداد داده آزمایشی
40	460	500	تعداد داده آموزشی

جدول 3: نتایج بدست آمده ارزیابی ها برای داده های آموزشی

F	فراخوان	دقت	صحت	الگوریتم پیشنهادی
95.81%	97.63%	96.42%	95.07%	الگوریتم پیشنهادی
95.44%	97.81%	93.18%	92.75%	Whale
94.42%	97.16%	91.83%	91.10%	PSO
95.35%	97.23%	93.54%	92.65%	GA

جدول 4: نتایج بدست آمده ارزیابی داده های آزمایشی

برای بررسی میزان صحت و دقت طبقه بندی در تشخیص کلاهبرداری روش پیشنهادی ما، از چهار متریک با نام های Precision، Recall، Accuracy و F استفاده شده است [8،9]. همچنین دیتاست استفاده شده در این پژوهش مجموعه داده Kaggle می باشد که برای تشخیص تقلب در کارت های اعتباری، یکی از منابع ارزشمند برای پژوهشگران و دانشمندان داده است که به دنبال توسعه مدل های پیش بینی برای تشخیص تراکنش های تقلب هستند. این مجموعه داده، با ارائه اطلاعات دقیق و گسترده در مورد تراکنش های مختلف، امکان ارزیابی و مقایسه عملکرد الگوریتم های مختلف را فراهم می کند. ویژگی های کلیدی این مجموعه داده عبارت است از:

- ✓ حجم داده: این مجموعه داده شامل صدها هزار تراکنش است که حجم قابل توجهی از داده را برای آموزش و ارزیابی مدل ها فراهم می کند.
- ✓ تنوع تراکنش ها: تراکنش ها در این مجموعه داده از نظر زمان، مبلغ و سایر ویژگی ها متنوع هستند که این امر باعث می شود مدل های توسعه یافته، در برابر انواع مختلف تقلب مقاوم تر باشند.

- ✓ عدم تعادل داده ها: یکی از چالش های اصلی در تشخیص تقلب، عدم تعادل داده ها است.

در این مجموعه داده نیز، تعداد تراکنش های عادی به مراتب بیشتر از تراکنش های تقلب است که این امر نیاز به استفاده از تکنیک های خاصی برای مقابله با عدم تعادل داده ها را ایجاد می کند. برخی از ویژگی ها در این مجموعه داده به صورت ناشناس ارائه شده اند که این امر به دلیل حفظ حریم خصوصی انجام شده است. با این حال، این ویژگی های ناشناس نیز می توانند حاوی اطلاعات ارزشمندی برای تشخیص تقلب باشند. داده ها معمولاً در قالب یک فایل CSV ارائه می شوند که به راحتی توسط نرم افزارهای مختلف قابل خواندن است. در کل، مجموعه داده Kaggle در زمینه تشخیص تقلب کارت های اعتباری، یک منبع ارزشمند برای پژوهشگران و دانشمندان داده است که به دنبال توسعه مدل های پیش بینی برای تشخیص تراکنش های تقلب هستند. معادلات این چهار متریک در زیر به ترتیب آورده شده است. ما در روش پیشنهادی ساختار شبکه عصبی MLP را به این صورت در نظر گرفتیم که دو لایه ورودی، ۲۰ لایه مخفی و دو لایه خروجی در نظر گرفته ایم. تعداد نورون های ورودی

دهد حاکی از آن است که در مورد فراخوان ، دقت و صحت بطور متوسط ۱۰۵٪ بهبود نسبت به دیگر روش های مورد مقایسه با روش پیشنهادی داشتیم. همچنین باید خاطر نشان کرد که با استفاده از روش پیشنهادی توانستیم کارایی شبکه را بهبود دهیم و کاهش پیچیدگی های محاسباتی در مقایسه با روش های گرادیان نزولی و حداقل مربعات شاهد باشیم.

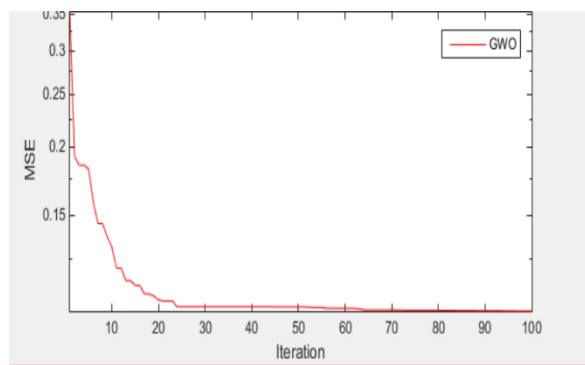
5- نتیجه گیری

تشخیص تقلب معمولاً بعنوان یک مسئله داده کاوری در نظر گرفته شده است. هدف از آن طبقه بندی صحیح معاملات بعنوان مشروع یا کلاه برداری است تشخیص کلاه برداری روی کارت هایی با محدودیت بیشتر، برای بانکداری هوشمند ارزشمندتر از کارت ها با محدودیت کمتر است. هدف اصلی به حداقل رساندن تعداد معاملات اشتباه طبقه بندی شده در بانکداری هوشمند است. در واقعیت طبقه بندی اشتباه در هر معامله تاثیر یکسانی ندارد زیرا اگر کارت در دست کلاه برداران باشد، کل حد مجاز کارت استفاده می شود. بنابراین هزینه طبقه بندی اشتباه باید بعنوان حد مجاز کارت در نظر گرفته شود و باید مینیمم شود. حجم زیاد دادگان و شباهت آنها منجر به انباشتگی زمان و دقت کمتر روی نمونه های سالم و ناسالم توسط طبقه بندی های سنتی می شود. یکی از روش های طبقه بندی استفاده از شبکه های عصبی می باشد که ما در این مقاله از شبکه عصبی MLP با روش BP برای فرایند آموزش MLP انتخاب مقادیر وزن های اولیه نقش بسیار مهمی بر روی فرایند آن استفاده کرده ایم. در شبکه عصبی آموزش شبکه دارند که با استفاده از روش های کلاسیک در شبکه این امر به سادگی میسر نخواهد بود و دارای محاسبات پیچیده بالایی هستند. در نتیجه از الگوریتم های فراابتکاری به جهت بهبود عملکرد شبکه و افزایش سرعت هگرایی استفاده می شود. ما در این تحقیق برای بهبود عملکرد شبکه از الگوریتم گرگ خاکستری به علت دقت بالا و تعداد کم پارامترهای تنظیمی استفاده نموده ایم.

همانطور که در بخش قبل ذکر شد، روش پیشنهادی بر روی مجموعه داده Kaggle که مجموعه معاملات آوری شده در ۲ روز می باشد، استخراج شده اند. این مجموعه داده در ستون آخر کلاسهایش قرار دارد که مقدار 0 نشان دهنده معامله نرمال و مقدار ۱ نشان دهنده کلاهبرداری می باشد. همچنین

	صحت	دقت	فراخوان	F
الگوریتم پیشنهادی	97.53٪	98.74٪	98.21٪	98.56٪
Whale	96.40٪	98.25٪	97.83٪	98.04٪
PSO	96.00٪	97.83٪	97.83٪	97.83٪
GA	96.20٪	98.25٪	97.61٪	97.93٪

همچنین باید خاطر نشان کرد که در زمان آموزش شبکه که توسط الگوریتم گرگ خاکستری بهبود یافته، در تکرار ۵۰ ام برابر با ۰.۱۰۱۹۰ و در تکرار ۱۰۰ ام برابر با ۰.۰۰۰۱۷ می باشد که در شکل 2 می توان مجموع مربع خطا را مشاهده کرد.



شکل 2: مجموع خطای شبکه بهبود داده شده با داده های آموزشی

با توجه به مقادیر بدست آمده در جداول بالا برای روش پیشنهادی در مقایسه با روش های قبلی گفته شده برای تشخیص کلاه برداری در کارت اعتباری می توان به این نتیجه رسید که روش پیشنهادی این مقاله از نظر سرعت و دقت تشخیص در اکثر موارد بهتر از دیگر روش های قبل بالاتر بوده است. همانطور که در جدول 5 مشاهده می شود، روش پیشنهادی نتایج بهتری را نسبت به روش پایه [13] کسب کرده است.

جدول 5: نتایج بدست آمده ارزیابی داده های آزمایشی

روش	معیار F
الگوریتم پیشنهادی	98.56٪
مقاله [13]	93.52٪

ما از ۴ ارزیاب استفاده کرده ایم. برای داده های آموزشی میزان صحت حداقل 2٪ بیشتر از دیگر روش های پیشین شده است. در مورد ارزیابی دقت الگوریتم پیشنهادی ما میزان تقریباً ۵.۴٪ بهبود داشته و در مورد F تقریباً 1٪ بهبود را در نتایج مشاهده می کنیم. در جدول 3 نتایجی که نشان می

- [10] Chiu, C. C., & Tsai, C. Y. (2004, March). A web services-based collaborative scheme for credit card fraud detection. In IEEE International Conference on e-Technology, e-Commerce and e-Service, 2004. EEE'04. 2004 (pp. 177-181). IEEE.
- [11] Duman, E., & Elikucuk, I. (2013). Solving credit card fraud detection problem by the new metaheuristics migrating birds optimization. In Advances in Computational Intelligence: 12th International Work-Conference on Artificial Neural Networks, IWANN 2013, Puerto de la Cruz, Tenerife, Spain, June 12-14, 2013, Proceedings, Part II 12 (pp. 62-71). Springer Berlin Heidelberg.
- [12] Dunn, J. C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters.
- [13] Tabrizchi, H., & Razmara, J. (2024). Credit card fraud detection using hybridization of isolation forest with grey wolf optimizer algorithm. *Soft Computing*, 1-19.

روش پیشنهادی ذکر شده را با روشهای PSO و GA و Whale در این مسئله مقایسه شده است. برای بررسی میزان کارایی و صحت طبقه بندی از چهار شاخصه ارزیابی با نام های صحت، فراخوان، دقت و F استفاده شده است. در پژوهش می توان به این نتیجه رسید که روش پیشنهادی ما از نظر سرعت اجرا و دقت طبقه بندی در بیشتر موارد بهتر از دیگر روش های قبل بوده است. همچنین کارایی روش پیشنهادی ما از دیگر روشها در بیشتر موارد بهتر بوده است.

مراجع

- [1] Mienye, I. D., & Jere, N. (2024). Deep learning for credit card fraud detection: A review of algorithms, challenges, and solutions. *IEEE Access*.
- [2] Cherif, A., Badhib, A., Ammar, H., Alshehri, S., Kalkatawi, M., & Imine, A. (2022). Credit card fraud detection in the era of disruptive technologies: A systematic review. *Journal of King Saud University-Computer and Information Sciences*.
- [3] Bin Sulaiman, R., Schetinin, V., & Sant, P. (2022). Review of machine learning approach on credit card fraud detection. *Human-Centric Intelligent Systems*, 2(1-2), 55-68.
- [4] Alarfaj, F. K., Malik, I., Khan, H. U., Almusallam, N., Ramzan, M., & Ahmed, M. (2022). Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms. *IEEE Access*, 10, 39700-39715.
- [5] Fanai, H., & Abbasimehr, H. (2023). A novel combined approach based on deep Autoencoder and deep classifiers for credit card fraud detection. *Expert Systems with Applications*, 217, 119562.
- [6] Alfaiz, N. S., & Fati, S. M. (2022). Enhanced credit card fraud detection model using machine learning. *Electronics*, 11(4), 662.
- [7] Akhilomen, J. (2013). Data mining application for cyber credit-card fraud detection system. In Advances in Data Mining. Applications and Theoretical Aspects: 13th Industrial Conference, ICDM 2013, New York, NY, USA, July 16-21, 2013. Proceedings 13 (pp. 218-228). Springer Berlin Heidelberg.
- [8] Carneiro, N., Figueira, G., & Costa, M. (2017). A data mining based system for credit-card fraud detection in e-tail. *Decision Support Systems*, 95, 91-101.
- [9] Chaudhary, K., & Mallick, B. (2012). Credit Card Fraud: Bang in E-Commerce. Editorial Board, 935.