Archive of SID

فصلنامهٔ مطالعات مدیریت کسب و کار هوشمند سال ششم، شماره ۲۲، زمستان ۹۶
صفحات ۱۲۹ تا ۱۵۴

انتخاب ویژگی‌های بهینه به‌منظور تعیین ریسک اعتباری مشتریان بانکی

مجتبی صالحی *
علیرضا کرد کتولی **

چکیده

ریسک اعتباری که به معنی احتمال عدم بازپرداخت تعهدات توسط مشتریان در سررسید تعبیر می‌شود به‌عنوان یکی از عوامل ورشکستگی مؤسسات مالی قلمداد می‌شود. بدین منظور از تکنیک‌های داده‌کاوی نظیر شبکه عصبی، درخت تصمیم، شبکه بیز، k نزدیک‌ترین همسایه و ماشین بردار پشتیبان برای دسته‌بندی مشتریان به مشتریان پر ریسک و کم ریسک استفاده شده است. در این مقاله یک روش ترکیبی از الگوریتم بهینه‌سازی رقابت استعماری و شبکه عصبی برای افزایش دقت دسته‌بندی در ارزیابی و سنجش ریسک اعتباری مشتریان بانکی ارائه می‌دهد. این روش با شناسایی زیرمجموعه‌ی ویژگی‌های بهینه و حذف ویژگی‌های غیرضروری از تمامی ویژگی‌های موجود در داده‌ها به کاهش ابعاد مسئله و افزایش دقت طبقه‌بندی می‌پردازد. رویکرد پیشنهادشده بر روی مجموعه داده‌های واقعی پایگاه داده UCI و همچنین داده‌های واقعی یک بانک خصوصی ایرانی به‌منظور اعتبارسنجی اعمال شد. نتایج تجربی به‌دست‌آمده نشان داد میزان خطای شبکه عصبی برای مجموعه آزمون با انتخاب ویژگی‌های مؤثر و حذف ویژگی‌های کم اثر توسط الگوریتم بهینه‌سازی صفر و یک

---

* استادیار، گروه مهندسی صنایع، دانشکدهٔ فنی و مهندسی، دانشگاه پیام نور، تهران، ایران. (نویسنده مسئول)؛ m.salehi61@chmail.ir
** کارشناسی ارشد، مهندسی صنایع، دانشکدهٔ فنی و مهندسی، دانشگاه پیام نور، تهران، ایران.
تاریخ دریافت:۱۳۹۵/۰۲/۲۳ تاریخ پذیرش:۱۳۹۶/۰۴/۲۷

www.SID.ir

Archive of SID

رقابت استعماری کاهش می‌یابد. بعلاوه، برای سایر روش‌ها طبقه‌بندی استفاده شده، میزان خطای داده آزمون در حد قابل قبولی باقی می‌ماند. برای اولین بار در این مقاله از الگوریتم رقابت استعماری برای ارزیابی ریسک اعتباری مشتریان بانکی استفاده شده است.

کلیدواژگان: ویژگی‌های بهینه، ریسک اعتباری، مشتریان بانکی، طبقه‌بندی، داده‌کاوی.

www.SID.ir

Archive of SID

مقدمه

ریسک اعتباری به‌طور ساده عبارت است از این امکان بالقوه که قرض گیرنده از بانک در اجرای تعهدات خود در مقابل بانک در مدت مشخصی ناتوان شود (مهرآرا و همکاران، ۱۳۸۸). آمارها نشان می‌دهد با توجه به ضمانت‌هایی که بانک‌ها از مشتریان اعتباری دریافت می‌کنند همچنان مطالبات عقب‌افتاده آن‌ها حجم قابل‌توجهی از اعتبارات را دربر می‌گیرد. امروزه مسئله اعتبارسنجی به یکی از مهم‌ترین مسائل پیش روی مدیران بانکی تبدیل شده است. بانک‌ها برای شناسایی مشتریان خود نیازمند پردازش و کشف دانش نهفته‌ی پایگاه‌های اطلاعاتی با استفاده از روش‌های نوین هستند. با بهره‌گیری از تحلیل اطلاعات مربوط به مشتریان بانک با استفاده از فرایند داده‌کاوی[1] می‌توان به رتبه‌بندی اعتباری متقاضیان وام و طبقه‌بندی آن‌ها بدون قضاوت شخصی و بر اساس سامانه‌های هوشمند پرداخت. این امر موجب کاهش ریسک‌های بانکی از جمله ریسک اعتباری می‌شود.

با توجه به افزایش سریع حجم داده‌ها در پایگاه‌های داده، نیاز به روش‌هایی با سرعت و دقت بیشتر برای کشف دانش درون پایگاه‌های داده در اولویت قرار می‌گیرد. داده‌کاوی اصلی‌ترین قدم در کشف دانش درون داده‌ها است. جستجو در حجم عظیمی از داده‌ها کاری بس دشوار، زمان‌بر و پرهزینه است زیرا همه‌ی داده‌های موجود نمی‌تواند داده‌های مفیدی برای پردازش و کشف دانش باشد. داده‌کاوی هم‌زمان با ایجاد و استفاده از پایگاه داده‌ها در اوایل دهه ۸۰ برای جستجوی دانش در داده‌ها شکل گرفت. شاید بتوان لوول[2] (۱۹۸۳) را اولین شخصی دانست که مقاله‌ای در مورد داده‌کاوی تحت عنوان "شبیه‌سازی فعالیت داده‌کاوی" ارائه نمود. هم‌زمان با او پژوهشگران و متخصصان علوم رایانه، آمار، هوش مصنوعی، یادگیری ماشین و ... نیز به پژوهش در این زمینه و زمینه‌های مرتبط با آن پرداخته‌اند. البته پژوهش جدی روی موضوع داده‌کاوی از اوایل دهه ۱۹۹۰ شروع شد و تا آن زمان این واژه به شکل و معنای امروزی بکار برده نمی‌شد.

از جمله روش‌های مختلف داده‌کاوی که در زمینه ریسک اعتباری کاربرد داشته است

---

1. Data Mining
2. Lovel

www.SID.ir

Archive of SID

می‌توان به درخت تصمیم، شبکه عصبی، شبکه بیز، نزدیک‌ترین همسایگی، ماشین‌های بردار پشتیبان، رگرسیون و K میانگین اشاره کرد. شبکه‌های عصبی مصنوعی به‌عنوان یکی از ابزارهای داده‌کاوی، شبکه‌هایی از نرون‌ها متصل‌به‌هم هستند که سعی می‌کنند کارکرد مغز انسان را در تصمیم‌گیری شبیه‌سازی کنند. نرون‌ها کوچک‌ترین واحدهای تصمیم‌گیری در این شبکه‌ها هستند. سیستم شبکه عصبی با تقلید از سیستم عصبی مغز انسان می‌کوشد ارتباط بین داده‌ها و ستاده‌ها را از راه تکرار نمونه‌برداری از مجموعه داده‌ها یاد بگیرد. شبکه‌های عصبی بر مبنای سه ویژگی داده‌های ورودی، وزن‌ها و لایه‌های پنهان تعریف می‌شوند.

فرایند انتخاب ویژگی‌ها، از مسائل بهینه‌سازی ترکیبی کلی در یادگیری ماشین شناخته می‌شود که تعداد ویژگی‌ها را کاهش داده و داده‌های غیر مرتبط و زائد را حذف می‌کند. فرایند انتخاب ویژگی‌ها در حوزه‌هایی که تعداد ویژگی‌ها زیاد است بسیار مورد توجه قرار گرفته است. هدف اصلی انتخاب ویژگی، شناسایی زیرمجموعه‌ای از ویژگی‌های بهینه است که بیشترین تأثیر را بر روی یک متغیر پاسخ معلوم دارند (زعفرانی و سیف-برقی، ۱۳۹۲). کشف زیرمجموعه بهینه‌ای از ویژگی معمولاً مشکل بوده و نشان داده شده است که بسیاری از مسائل مرتبط NP-hard شناخته می‌شوند (کوهاوی و جوهن[1]، ۱۳۹۷). بسیاری از پایگاه‌های داده شامل چند صد میلیارد رکورد ثبت شده می‌باشند که تحلیل و استخراج و پردازش اطلاعات آن‌ها با روش‌های معمول آماری مستلزم صرف زمان و هزینه بسیار بالایی است که اغلب کمترین توجیه اقتصادی برای آن یافت نمی‌شود. بی‌تردید بسیاری از این داده‌ها بی‌کیفیت، پر خطا و یا متناقض می‌باشند و بدیهی است که استفاده از داده‌های بی‌کیفیت همیشه منجر به نتایج بی‌کیفیت‌تر می‌شود. بدین منظور برای کاهش حجم داده و کشف ویژگی‌های بهینه از روش‌های انتخاب ویژگی استفاده می‌شود.

الگوریتم‌های بهینه‌سازی معرفی‌شده، به‌طور عمده الهام گرفته از فرایندهای طبیعی می‌باشند و در ارائه این الگوریتم‌ها به سایر نمودهای تکامل انسانی توجهی نشده است. در این نوشتار الگوریتم جدیدی برای بهینه‌سازی مطرح می‌شود که نه از یک پدیده طبیعی، بلکه از یک پدیده اجتماعی ـ انسانی الهام گرفته است. به‌طور ویژه این

1. Kohavi & John

www.SID.ir

Archive of SID

الگوریتم به فرایند استعمار، به‌عنوان مرحله‌ای از تکامل اجتماعی ـ سیاسی بشر نگریسته و با مدل‌سازی ریاضی این پدیده تاریخی، از آن به‌عنوان منشأ الهام یک الگوریتم قدرتمند در زمینه بهینه‌سازی بهره می‌گیرد. در مدت کوتاهی که از معرفی این الگوریتم می‌گذرد، از آن برای حل مسائل بسیاری در حوزه بهینه‌سازی استفاده شده است.

در ادبیات مسئله ریسک اعتباری از الگوریتم‌هایی مانند الگوریتم ژنتیک[1] و الگوریتم جفت‌گیری زنبورعسل[2] به‌منظور انتخاب ویژگی‌های بهینه استفاده شده است (اورسکی و ارسکی[3]، ۲۰۱٤، مارناکی و همکاران[4]، ۲۰۱۲) برای اولین بار در این مقاله با استفاده از الگوریتم بهینه‌سازی رقابت استعماری به انتخاب ویژگی‌های بهینه مشتریان بانکی پرداخته شده است. مسئله اصلی، انتخاب ویژگی‌های بهینه مشتریان بانکی برای افزایش دقت و سرعت عملیات داده‌کاوی با استفاده از الگوریتم بهینه‌سازی رقابت استعماری است.

در بخش ۲ این مقاله به مروری بر کارهای انجام‌شده در این حوزه پرداخته شده است. بخش ۳ مفاهیم شبکه عصبی را تبیین خواهد کرد. بخش ٤ مروری بر الگوریتم بهینه‌سازی رقابت استعماری خواهد بود. در بخش ۵ روش پیشنهادی برای انتخاب ویژگی‌های بهینه مشتریان بانکی و ٤ روش طبقه‌بندی دیگر معرفی خواهد شد. بخش ٦ نتایج آزمایش‌ها بر روی پایگاه‌های داده بررسی خواهد شد و درنهایت بخش ۷ نتیجه‌گیری پژوهش و کارهای آتی خواهد بود.

مروری بر مطالعات پیشین

امروزه اکثر تحقیقات و کاربردها در اعتبارسنجی مبتنی بر دو روش هستند: ۱ـ روش‌های سنتی (آماری) ۲ـ روش‌های داده‌کاوی (هوش مصنوعی). می‌توان مدل‌های اعتبار سنجی را به مدل‌های پارامتریک و نا پارامتریک تقسیم کرد. روش‌های پارامتریک مثل پروبیت[5] و لوجیت[6]، تحلیل تمایزی خطی[1] و رگرسیون خطی[2] که از

1. Genetic Algorithm Optimization
2. Honey Bees Mating Optimization
3. Oreski & Oreski
4. Marinaki et al.
5. Probit
6. Logit

www.SID.ir

Archive of SID

ابتدای ظهور اعتبارسنجی مورد استفاده قرار گرفتند و سپس روش نا پارامتریک و داده‌کاوی مثل درختان تصمیم‌گیری، شبکه‌های عصبی و ماشین بردار پشتیبان که امروزه بیشتر به کار گرفته می‌شود.

هسیه[3] (۲۰۰٤) یک مدل امتیازدهی رفتاری و داده‌کاوی یکپارچه برای مدیریت مشتریان کارت‌های اعتباری موجود در یک بانک مطرح می‌کند. این مطالعه نشان داد که شناسایی مشتریان توسط یک مدل امتیازدهی رفتاری برای توسعه استراتژی تسهیلات بازاریابی و ویژگی‌های مشتری مفید است. با وجود داده‌های حجیم در پایگاه‌های داده استفاده از روش مطرح شده در این مقاله نمی‌تواند کارایی لازم را داشته باشد. برای برطرف کردن این نقص ما از روش انتخاب ویژگی با استفاده از الگوریتم رقابت استعماری استفاده کرده‌ایم.

پیراموسو[4] (۲۰۰٦) از پیش‌پردازش داده ارزیابی ریسک اعتباری که به‌عنوان ورودی روش‌های یادگیری ماشین استفاده می‌شوند برای بهبود عملکرد ابزارهای پشتیبانی تصمیم استفاده کرد.

کارلینگ[5] و همکاران (۲۰۰۷) یک مدل زمانی برای مدت بقاء تا نکول برای مشتری در پرتفوی وام تجاری از یک بانک بزرگ سوئدی در سال‌های ۱۹۹٤ـ۲۰۰۰. مقایسه آن‌ها نشان داد که مدل می‌تواند یک رتبه‌بندی دقیق منطقی از شرکت‌ها مطابق با ریسک نکول ارائه دهد.

ماریناکی و همکاران[6] (۲۰۱۰) با استفاده از الگوریتم بهینه‌سازی زنبورعسل و روش نزدیک‌ترین همسایگی به انتخاب ویژگی‌های بهینه و طبقه‌بندی مشتریان در ارزیابی ریسک اعتباری پرداختند. نتایج حاصل با الگوریتم ازدحام ذرات، کلونی مورچگان، الگوریتم ژنتیک و الگوریتم جستجوی ممنوع مقایسه شد. در پژوهش ذکرشده از ماریناکی و همکاران، تنها از روش طبقه‌بندی نزدیک‌ترین همسایگی استفاده شده است و ویژگی‌های انتخاب‌شده با سایر مدل‌های طبقه‌بندی مورد آزمایش قرار نگرفته است. ما در پژوهش خود علاوه بر شناسایی ویژگی‌های بهینه و مقایسه روند همگرایی

1. Linear Discriminant Analysis
2. Linear regression
3. Hsieh
4. Piramuthu
5. Carling et al.
6. Marinaki et al.

www.SID.ir

Archive of SID

الگوریتم بهینه‌سازی استفاده شده با دیگر الگوریتم‌ها، به مقایسه عملکرد دیگر روش‌های طبقه‌بندی نظیر درخت تصمیم، شبکه‌های بیزی، نزدیک‌ترین همسایگی و ماشین بردار پشتیبان بر روی ویژگی‌های انتخابی پرداختیم.

فین لای[1] (۲۰۱۱) کارایی چند سیستم طبقه‌بندی چندگانه را از نظر توانایی آن‌ها برای طبقه‌بندی درست مشتریان با ریسک‌های اعتباری خوب یا بد بررسی کرد. نتایج تجربی او نشان داد که برخی از سامانه‌های طبقه‌بندی چندگانه به‌طور قابل‌توجهی کارایی بهتری از بهترین طبقه‌بندی تکی ارائه می‌دهند اما بسیاری نیز چنین نیستند.

جدول ۱. مروری بر ادبیات مسئله ریسک اعتباری و انتخاب ویژگی با تأکید بر الگوریتم‌های فرا ابتکاری و روش‌های داده‌کاوی

عنوان	نویسندگان	روش مورد استفاده	نوع شبکه عصبی	روش انتخاب ویژگی	حوزه کاربرد
الگوریتم ابتکاری ژنتیک برای انتخاب ویژگی در ارزیابی ریسک اعتباری	(اورسکی و ارسکی[2]، ۲۰۱٤)	شبکه عصبی	پرسپترون چندلایه	الگوریتم ژنتیک	ارزیابی ریسک اعتباری مشتریان بانکی
مدل ارزیابی ریسک اعتباری برای بانک‌های تجاری اردن: رویکرد امتیازدهی اعتباری	بخت و التر[3] (۲۰۱٤)	شبکه عصبی	شعاعی	-	امتیازدهی اعتباری بانک‌های تجاری اردن
سیستم ترکیبی با الگوریتم ژنتیک و شبکه عصبی مصنوعی برای ارزیابی ریسک اعتباری خرده‌فروشی	(اورسکی و همکاران[4]، ۲۰۱۲)	شبکه عصبی	پرسپترون چندلایه	الگوریتم ژنتیک	ارزیابی ریسک اعتباری مشتریان بانکی
مدل‌های انتخاب ویژگی احتمالی و متمایزکننده گروه دانا برای تحلیل ریسک اعتباری	گونن[5] و همکاران (۲۰۱۱)	رگرسیون پروبیت		گروه دانا	ارزیابی ریسک اعتباری

1. Finlay
2. Oreski & Oreski
3. Bekhet & Eletter
4. Oreski et al.
5. Gönen et al.

www.SID.ir

Archive of SID

عنوان	نویسندگان	روش مورد استفاده	نوع شبکه عصبی	روش انتخاب ویژگی	حوزه کاربرد
کاربرد داده‌کاوی برای ارزیابی ریسک اعتباری از طریق مدل‌های امتیازدهی اعتباری	یاپ و همکاران[1] (۲۰۱۱)	رگرسیون لجستیک درخت تصمیم کارت امتیازی اعتباری	–	–	امتیازدهی اعتباری

جدول ۲. مروری بر ادبیات مسئله ریسک اعتباری و انتخاب ویژگی با تأکید بر الگوریتم‌های فرا ابتکاری و روش‌های داده‌کاوی

عنوان	نویسندگان	روش مورد استفاده	نوع شبکه عصبی	روش انتخاب ویژگی	حوزه کاربرد
توسعه مدل‌های ارزیابی ریسک اعتباری با استفاده از طبقه‌بندی بر اساس ماشین‌های بردار پشتیبان	داناس و همکاران[2] (۲۰۱۱)	ماشین بردار پشتیبان	–	الگوریتم جستجوی ممنوع	ارزیابی ورشکستگی مالی
روش بهینه‌سازی ازدحام ذرات گسسته برای انتخاب ویژگی در مسائل طبقه‌بندی باینری	اونلر و مورات[3] (۲۰۱۰)	رگرسیون لجستیک	–	الگوریتم ازدحام ذرات	مسائل طبقه‌بندی
الگوریتم جفت‌گیری زنبورعسل برای مسائل طبقه‌بندی مالی	مارینا‌کی و همکاران[4] (۲۰۱۰)	نزدیک‌ترین همسایگی	–	الگوریتم جفت‌گیری زنبورعسل	ارزیابی ریسک اعتباری مشتریان حقوقی
الگوریتم کلونی مورچگان و ازدحام ذرات برای مسئله طبقه‌بندی مالی	مارینا‌کی و همکاران[5] (۲۰۰۹)	نزدیک‌ترین همسایگی	–	الگوریتم کلونی مورچگان و ازدحام ذرات	طبقه‌بندی مالی
امتیازدهی اعتباری با رویکرد داده‌کاوی مبنی بر ماشین‌های بردار پشتیبان	هیوانگ و همکاران[6] (۲۰۰۷)	ماشین بردار پشتیبان	-	الگوریتم ژنتیک	امتیازدهی اعتباری مشتریان کارت‌های اعتباری
خوشه‌بندی اعتباری مشتریان برای ارائه تسهیلات متناسب	افسر و همکاران (۱۳۹۲)	–	کوهن	–	امتیازدهی اعتباری
مقایسه کارایی مدل‌های کلاسیک و شبکه‌های عصبی در برآورد ریسک و ظرفیت اعتباری مشتریان شواهدی از	عیسی زاده و منصوری (۱۳۹۲)	شبکه عصبی	پرسپترون چندلایه	–	اعتبارسنجی مشتریان

1. Yap et al.
2. Danenas et al.
3. Unler & Murat
4. Marinaki et al.
5. Marinaki et al.
6. Huang et al.

www.SID.ir

Archive of SID

عنوان	نویسندگان	روش مورد استفاده	نوع شبکه عصبی	روش انتخاب ویژگی	حوزه کاربرد
بانک تجارت					
به‌کارگیری تکنیک‌های خوشه‌بندی و الگوریتم ژنتیک در بهینه‌سازی درختان تصمیم‌گیری برای اعتبارسنجی مشتریان بانکی	البرزی و همکاران (۱۳۹۱)	درخت تصمیم	-	الگوریتم ژنتیک	اعتبارسنجی مشتریان بانکی

یاپ و همکاران[1] (۲۰۱۱) کاربرد داده‌کاوی برای بهبود ارزیابی ارزش اعتباری با استفاده از مدل امتیازدهی اعتباری بیان کردند. آن‌ها عملکرد طبقه‌بندی مدل امتیازدهی اعتباری، رگرسیون لجستیک و درخت تصمیم را مقایسه کردند.

اورسکی و همکاران[2] (۲۰۱۲) بررسی کردند که مجموعه داده‌های متعلق به یک بانک تا چه اندازه می‌تواند برای پیش‌بینی توانایی وام‌گیرنده در بازپرداخت به‌موقع وام مؤثر باشد. آن‌ها یک تکنیک انتخاب ویژگی با الگوریتم ژنتیک و شبکه عصبی برای یافتن زیرمجموعه ویژگی بهینه که دقت طبقه‌بندی را افزایش می‌دهد، ارائه دادند. آزمایش‌ها برای ارزیابی دقت روش آن‌ها بر روی داده‌های یک بانک کرواتی انجام شد. در پژوهش ذکرشده از اورسکی و همکاران، با استفاده از الگوریتم ژنتیک و شبکه عصبی به انتخاب ویژگی‌های بهینه پرداخته و با سایر روش‌های انتخاب ویژگی مقایسه شده است؛ اما کاربردی از روش‌های طبقه‌بندی دیگر برای تعیین تائید صحت طبقه‌بندی دیده نمی‌شود. ما در پژوهش خود روشی مبتنی بر الگوریتم رقابت استعماری و شبکه عصبی ارائه داده‌ایم که عملکرد بهتری نسبت به مدل اورسکی و همکاران داشته است و برای تصدیق ویژگی‌های انتخابی از چهار روش طبقه‌بندی دیگر نیز استفاده کرده‌ایم.

ماندالا و همکاران[3] (۲۰۱۲) مقاله‌ای با هدف تشخیص فاکتورهایی که برای شناسایی برنامه‌های اعتباری یک بانک روستایی هستند، برای کاهش تعداد وام‌های غیر اجرایی و معیارهای تصمیم‌گیری متداول در ارزیابی ریسک اعتباری مورد بررسی قرار دادند. پس‌ازآن یک مدل درخت تصمیم با استفاده از متدولوژی داده‌کاوی ارائه دادند.

1. Yap et al.
2. Oreski et al.
3. Mandala et al.

www.SID.ir

Archive of SID

چن و همکاران[1] (۲۰۱۲) یک تکنیک ترکیبی داده‌کاوی را پیشنهاد کردند که حاوی دو مرحله پردازش است. تفاوت روش آن‌ها با سایر روش‌های امتیازدهی اعتباری آن است که نمونه‌ها بجای دو دسته خوب و بد به سه یا چهار دسته تقسیم می‌شوند. نتایج آن‌ها بر روی داده‌های یک بانک محلی در چین نشان می‌دهد که با انتخاب یک نقطه مناسب، صحت دسته‌بندی بیشتر از اعتبار خوب و بد حاصل شد.

محمد و تاسیر[2] (۲۰۱۳) به بررسی روش‌های داده‌کاوی محققان قبلی و آخرین روند داده‌کاوی آموزشی پرداختند و محدودیت‌های مختلف تحقیقات موجود مطرح و برخی جهت‌گیری‌ها برای تحقیقات آتی پیشنهاد شد.

بخت و التر[3] (۲۰۱٤) دو مدل امتیازدهی اعتباری با استفاده از تکنیک‌های داده‌کاوی برای حمایت از تصمیم‌گیری وام برای بانک‌های تجاری اردن پیشنهاد کردند. نتایج آن‌ها نشان داد که مدل رگرسیون لجستیک قدری بهتر از تابع شعاع محور عمل می‌کند.

اورسکی و ارسکی[4] (۲۰۱٤) یک الگوریتم ابتکاری جدید پیشرفته ارائه دادند. آن‌ها ترکیبی از الگوریتم ژنتیک و شبکه‌های عصبی به‌منظور شناسایی زیرمجموعه ویژگی بهینه در ارزیابی ریسک اعتباری ارائه کردند. اجرای مدل پیشنهادی با استفاده پایگاه داده‌های امتیازی دنیای واقعی که در یک بانک کرواتی جمع‌آوری شده ارزیابی شد؛ و داده‌ها با یک پایگاه داده دنیای واقعی اعتبارسنجی شد.

لوپز و جرونیمو[5] (۲۰۱۵) یک رویکرد گروهی مبتنی بر درخت‌های تصمیم ادغام شده بنام جنگل‌های تصمیم تعدیل‌شده همبسته[6] (CADF)، برای تولید مدل‌های دقیق و قابل‌فهم ارائه دادند. نتایج تجربی آن‌ها نشان داد CADF یک راه حل امیدوارکننده برای مسئله ریسک اعتباری است که قابل رقابت با دقت طرح‌های بسیار پیچیده است.

ماشین‌های بردار پشتیبان خطی[7] همراه با ارزیابی خارجی و تکنیک اسلاید ویندو با تمرکز بر نرم‌افزار در دیتاست‌های بزرگ ارائه کردند. روش آن‌ها یک تکنیک برای انتخاب طبقه‌بندی کننده ماشین‌های بردار پشتیبان خطی بهینه بر اساس روش

---

1. Chen et al.
2. Mohamad & Tasir
3. Bekhet & Eletter
4. Oreski & Oreski
5. Lopez & Jeronimo
6. Correlated-Adjusted Decision Forest
7. Linear Support Vector Machines classifier

www.SID.ir

Archive of SID

بهینه‌سازی ازدحام ذرات با تمرکز روی مسئله یادگیری نامتوازن. نتایج به‌دست‌آمده با استفاده از مجموعه داده‌های مالی دنیای واقعی SEC EDGARD نشان داد که این روش قابل رقابت با مدل رگرسیون لجستیک و شبکه عصبی خودسازمانده است.

هریس[1] (۲۰۱۵) در مقاله خود عمل امتیازدهی اعتباری را بررسی می‌کند و کاربردهای ماشین‌های بردار پشتیبان خوشه‌ای[2] (CSVM) برای بهبود امتیازدهی اعتباری را معرفی می‌کند. این الگوریتم که به‌تازگی طراحی شده برخی از محدودیت‌های موجود در ادبیات مرتبط با طبقه‌بندی را با روش‌هایی بر پایه ماشین‌های بردار پشتیبان غیرخطی مرسوم نشان می‌دهد. این نکته بر اساس اطلاعات گذشته مشخص است که دیتاست‌های ریسک اعتباری بزرگ‌تر می‌شوند. این رویکرد غیرخطی درحالی‌که دقت بسیار بالای داشته باشد بسیار پرهزینه است. این مطالعه نشان داد CSVM قادر است سطح دقت طبقه‌بندی قابل‌مقایسه‌ای درحالی‌که هنوز کم‌هزینه است به دست بیاورد.

در این مقاله یک روش ترکیبی از الگوریتم رقابت استعماری و شبکه عصبی برای پیش‌بینی ریسک اعتباری مشتریان بانکی با دقت بالاتر ارائه می‌دهد. این روش با کاهش حجم داده و شناسایی زیرمجموعه‌ی ویژگی‌های بهینه به کاهش ابعاد مسئله می‌پردازد و افزایش دقت طبقه‌بندی را در پی دارد. رویکرد پیشنهادشده بر روی دو مجموعه داده‌های واقعی جرمن اعمال شد. برای اولین بار در این مقاله از الگوریتم رقابت استعماری برای ارزیابی ریسک اعتباری مشتریان بانکی استفاده شده است.

طبقه‌بندی

طبقه‌بندی یکی از عملیات رایج و مورد استفاده در داده‌کاوی است. طبقه‌بندی عملیاتی است که سازمان‌ها را قادر می‌سازد که در حل مسائل خاص در مجموعه‌های بزرگ و پیچیده به کشف الگوها دست یابند. طبقه‌بندی فرآیندی است که مجموعه داده‌ها را به قسمت‌های مشخص تقسیم می‌کند. طبقه‌بندی داده‌ها یک فرآیند دومرحله‌ای است. در مرحله اول که مرحله یادگیری[3] نامیده می‌شود، الگوریتم طبقه‌بندی توسط تحلیل یا همان (یادگیری) از مجموعه داده‌های آموزشی متشکل از سطرهای پایگاه داده و

1. Harris
2. Clustered Support Vector Machine
3. Training Step

www.SID.ir

Archive of SID

برچسب‌های دسته مرتبط با آن‌ها، طبقه‌بند[1] (طبقه‌بندی کننده) را می‌سازد. در مرحله دوم، از مدل برای طبقه‌بندی داده‌ها استفاده می‌شود. در ابتدا صحت پیش‌بینی دسته‌بند تخمین زده می‌شود. برای مثال مشتریان یک بانک بر اساس خصوصیاتشان به دو گروه با ریسک بالا و ریسک پایین تقسیم می‌شوند، با این کار در واقع مشتریان این بانک طبقه‌بندی شده‌اند. برای سنجش یک طبقه‌بندی کننده می‌توان قسمتی از داده‌هایی که در هنگام آموزش استفاده نشده را به‌عنوان داده آزمایش استفاده کرد. بدین‌صورت می‌توان میزان خطای طبقه‌بندی کننده‌های مختلف را با هم مقایسه کرد. شبکه عصبی، درخت تصمیم، ماشین بردار پشتیبان، k نزدیک‌ترین همسایه و شبکه بیزی از جمله روش‌های طبقه‌بندی پرکاربرد هستند.

شبکه‌های عصبی

شبکه عصبی مصنوعی یک سیستم پردازش داده است که از مغز انسان ایده گرفته و پردازش داده‌ها را به عهده پردازنده‌های کوچک و بسیار زیادی سپرده که به‌صورت شبکه‌ای به‌هم‌پیوسته و موازی رفتار می‌کنند تا یک مسئله را حل نمایند. در این شبکه‌ها به کمک دانش برنامه‌نویسی، ساختار داده‌ای طراحی می‌شود که می‌توانند همانند نورون‌های مغز انسان عمل کنند. شبکه‌های عصبی انواع مختلفی دارند و از معروف‌ترین این شبکه‌ها می‌توان از شبکه‌های عصبی پرسپترون ساده و چندلایه نام برد. از شبکه‌های پرسپترون تک لایه برای مسائل خطی ساده و برای مسائل پیچیده‌تر از شبکه‌های عصبی چندلایه استفاده می‌شود. به‌طورکلی یک شبکه عصبی چندلایه شامل سه پخش لایه ورودی، لایه پنهان و لایه خروجی است. تنظیم وزن‌های بین نرون‌های لایه‌های مختلف شبکه عصبی به‌عنوان آموزش شبکه عصبی شناخته می‌شود. برای یادگیری شبکه عصبی مجموعه‌ای به نام مجموعه یادگیری به شبکه عصبی داده می‌شود تا به وسیله آن خروجی مطلوب بدست آید. زمانی که از یک شبکه عصبی با n لایه صحبت می‌شود تنها لایه‌های میانی و لایه خروجی مد نظر است و لایه ورودی شمارش نمی‌شود زیرا این نرون‌ها محاسبه‌ایی انجام نمی‌دهند.

در این پژوهش از یک شبکه عصبی پرسپترون دو لایه (با یک لایه مخفی) وتعداد ۲۰

1. Classifier

www.SID.ir

Archive of SID

نرون استفاده شده است. ورودی این شبکه عصبی زیرمجموعه‌ی ویژگی‌های انتخابی مشتریان بانکی توسط الگوریتم رقابت استعماری است. بدین صورت که شبکه عصبی با زیرمجموعه‌های مختلف از ویژگی‌ها به‌عنوان داده ورودی، آموزش داده می‌شود. هر زیرمجموعه ویژگی که کمترین میزان خطای داده تست را به همراه داشته باشد، زیرمجموعه‌ی ویژگی‌های بهینه خواهد بود. در این شبکه عصبی ۷۰ درصد از داده‌های ورودی به‌عنوان داده آموزش، ۱۵ درصد به‌عنوان داده ارزیابی و ۱۵ درصد به‌عنوان داده تست در نظر گرفته شده است.

www.SID.ir

Archive of SID

درخت تصمیم

درخت تصمیم تکنیکی است که به سادگی به نمایش دانش در حجم زیادی از داده‌ها پرداخته و مشتریان اعتباری را به طبقات مشتری خوب و مشتری بد تقسیم‌بندی می‌کند. استنتاج درختان تصمیم‌گیری، به معنای یادگیری آن‌ها از سطرهای آموزشی، با برچسب‌های طبقه‌ی مشخص است. ساختار این درختان شبیه به ساختار فلوچارت است که در آن هر گره میانی نشان‌دهنده آزمایش بر روی یک یا چند خصوصیت بوده، هر شاخه نشان‌دهنده‌ی خروجی آزمایش و هر گره برگ نیز دربرگیرنده یک برچسب طبقه می‌باشد.

ماشین بردار پشتیبان

یک روش برای دسته‌بندی داده‌های خطی و غیرخطی می‌باشد. این روش ابتدا از یک نگاشت غیرخطی برای تبدیل داده اولیه به ابعاد بالا استفاده کرده و سپس در بعد جدید به دنبال بهترین فراصفحه[1] جداساز می‌گردد. این فراصفحه عبارت است از یک مرز تصمیم که رکوردهای یک دسته را از دسته‌های دیگر جدا می‌کند. با یک نگاشت غیرخطی به ابعاد به اندازه کافی بالاتر، داده‌های متعلق به دسته‌های گوناگون به راحتی توسط یک فراصفحه جدا می‌شوند. ماشین بردار پشتیبان این فراصفحه را با استفاده از بردارهای پشتیبان و حاشیه‌ها که توسط بردارهای پشتیبان تعریف می‌شوند، می‌یابد.

نزدیکترین همسایگی هنگام تلاش برای حل مسائل جدید، افراد معمولا به راه‌حل‌های مسائل مشابه که قبلا حل شده‌اند، مراجعه می‌کنند. k نزدیک‌ترین همسایه (KNN) یک تکنیک دسته‌بندی است که از نسخه‌ای از این متد استفاده می‌کند. در این روش تصمیم‌گیری اینکه یک نمونه جدید در کدام دسته قرار گیرد با بررسی تعداد (k) از شبیه‌ترین نمونه‌ها یا همسایه‌ها انجام می‌شود. در بین این k نمونه، تعداد نمونه‌ها برای هر دسته شمرده می‌شوند و نمونه جدید به دسته‌ای که تعداد بیشتری از همسایه‌ها به آن تعلق دارند نسبت داده می‌شود.

---

1. Hyperplane

www.SID.ir

Archive of SID

شبکه بیزی

الگوریتم‌های یادگیری بیزی به طور صریح به روی احتمالات فرض‌های مختلف کار می‌کنند. دسته‌بندی‌کننده‌های بیزی طبقه بندی‌کننده‌های آماری هستند. آن‌ها اعضای طبقه را به‌صورت احتمالی پیشگویی می‌کنند؛ مثلا میزان احتمال اینکه یک نمونه داده شده متعلق به یک دسته خاص باشد. دسته‌بندی‌کننده بیزی بر مبنای تئوری بیز می‌باشد.

الگوریتم رقابت استعماری

الگوریتم رقابت استعماری (ICA) یک الگوریتم جدید در زمینه محاسبات تکاملی است که بر مبنای تکامل اجتماعی ـ سیاسی انسان پایه‌گذاری شده است. همانند دیگر الگوریتم‌های تکاملی، این الگوریتم نیز با تعدادی جمعیت اولیه تصادفی که هر کدام از آن‌ها یک کشور نامیده می‌شوند، شروع می‌شود. تعدادی از بهترین عناصر به‌عنوان استعمارگر انتخاب و باقیمانده جمعیت نیز به‌عنوان مستعمره، در نظر گرفته می‌شوند.
در مسائل بهینه‌سازی با در نظر گرفتن تابع $f(x)$، سعی بر آن است تا آرگومان x به گونه‌ای یافته شود که هزینه متناظر با آن، بهینه باشد (معمولا کمینه). در یک مسئله بهینه‌سازی $1 \times N_{var}$ است. این آرایه به‌صورت زیر تعریف می‌شود.

$$Country = [P_1, P_2, P_3, \dots, P_{N_{var}}] \quad (۱)$$

هزینه یک کشوربا ارزیابی تابع f به ازای متغیرهای $(P_1, P_2, P_3, \dots, P_{N_{var}})$ یافته می‌شود؛ بنابراین داریم:

$$Cost_i = f(country_i) = f(P_1, P_2, P_3, \dots, P_{N_{var}}) \quad (۲)$$

در الگوریتم ICA، برای شروع، تعداد $N_{country}$ کشور اولیه ایجاد و N_{imp} تا از بهترین اعضای این جمعیت (کشورهای دارای کمترین مقدار تابع هزینه) به‌عنوان استعمارگر انتخاب می‌شوند. باقیمانده N_{col} تا از کشورها، مستعمراتی را تشکیل می‌دهند که هر کدام به یک امپراطوری تعلق دارند.
کشورهای استعمارگر با اعمال سیاست جذب (همگون‌سازی) در راستای محورهای مختلف بهینه‌سازی، کشورهای مستعمره را به سمت خود می‌کشند. استعمارگران با

www.SID.ir

Archive of SID

توجه به قدرتشان، این مستعمرات را با رابطه (۳)، به سمت خود می‌کشند. قدرت کل هر امپراطوری، با محاسبه قدرت هر دو بخش تشکیل دهنده آن یعنی قدرت کشور استعمارگر، به اضافه درصدی از میانگین قدرت مستعمرات آن، تعیین می‌شود.

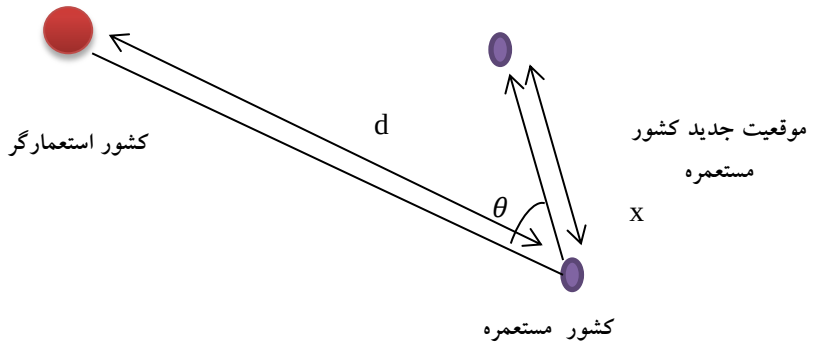
$$T.C._n \, Cost(imperialist_n) + \xi mean\{colonies\ of\ empire_n\} \qquad (۳)$$

کشور مستعمره، به اندازه x واحد در جهت خط واصل مستعمره به استعمارگر، حرکت کرده و به موقعیت جدید، کشانده می‌شود. در شکل ۱، فاصله میان استعمارگر و مستعمره با d نشان داده شده است و x نیز عددی تصادفی با توزیع یکنواخت (و یا هر توزیع مناسب دیگر) می‌باشد؛ یعنی برای x داریم:

$$x \sim U(0, \beta \times d) \qquad (٤)$$

که در آن β عددی بزرگتر از یک و نزدیک به ۲ می‌باشد. یک انتخاب مناسب می‌تواند $\beta = 2$ باشد. همچنین زاویه حرکت را به‌صورت توزیع یکنواخت زیر در نظر گرفته شده است.

$$\theta = U(-\gamma, \gamma) \qquad (۵)$$



شکل ۱. حرکت مستعمرات به سمت استعمارگر (سیاست جذب)

در الگوریتم ICA، با یک انحراف احتمالی، مستعمره در مسیر جذب استعمارگر پیش می‌رود. این انحراف با زاویه θ نشان داده شده است که θ به‌صورت تصادفی و با توزیع یکنواخت انتخاب انتخاب می‌شود. در حین حرکت مستعمرات به سمت کشور استعمارگر، ممکن است بعضی از این مستعمرات به موقعیتی بهتر از استعمارگر برسند در این حالت، کشور استعمارگر و کشور مستعمره، جای خود را با هم عوض می‌کنند.

www.SID.ir

Archive of SID

برای مدل‌سازی این رقابت، ابتدا احتمال تصاحب مستعمرات توسط هر امپراطوری با در نظر گرفتن هزینه کل امپراطوری به‌صورت زیر محاسبه می‌شود.

$$N.T.C._n = \max_i\{T.C._i\} - T.C._n \quad (۶)$$

در این رابطه $T.C._n$، هزینه کل امپراطوری n ام و $N.T.C._n$ نیز، هزینه کل نرمالیزه شده آن امپراطوری می‌باشد و احتمال تصاحب مستعمره، توسط هر امپراطوری به‌صورت زیر محاسبه می‌شود.

$$P_{p_n} = \left|\frac{N.T.C._n}{\sum_{i=1}^{N_{imp}} N.T.C._i}\right| \quad (۷)$$

روش پیشنهادی (الگوریتم رقابت استعماری صفر و یک برای انتخاب ویژگی‌های بهینه)

پیاده‌سازی

این روش با استفاده از الگوریتم رقابت استعماری و شبکه عصبی به انتخاب ویژگی‌های بهینه می‌پردازد. بدین صورت که الگوریتم رقابت استعماری در فضای N بعدی مسئله، در هر تکرار، شبکه عصبی را با تعدادی از راه‌حل‌ها آموزش می‌دهد. در الگوریتم رقابت استعماری برای شروع، تعداد $N_{country}$ کشور اولیه ایجاد و N_{imp} تا از بهترین اعضای این جمعیت (کشورهای دارای کمترین مقدار تابع هزینه) به‌عنوان استعمارگر انتخاب می‌شوند. هر کشور نشان دهنده یک راه حل است که تعدادی از ویژگی‌ها را در بر دارد. کشورهای استعمارگر با اعمال سیاست جذب (همگون‌سازی) در راستای محورهای مختلف بهینه‌سازی، کشورهای مستعمره را به سمت خود می‌کشند. راه‌حل بهینه راه‌حلی با بهترین مقدار از تابع هدف در همه‌ی تکرارها خواهد بود. این جواب همان ویژگی‌های انتخاب شده از بین ویژگی‌های مشتریان است. این فرایند در نمودار (۱) نشان داده شده است. تابع هدف، خطای تست شبکه عصبی استفاده شده است. در مرحله بعد برای سنجش مدل، به مقایسه میزان خطای اعتبارسنجی ضربدری[1] در ٤ روش طبقه بندی درخت تصمیم، شبکه بیزی، k

1. k-fold Cross Validation

www.SID.ir

Archive of SID

نزدیک‌ترین همسایه و ماشین بردار پشتیبان به‌عنوان تابع هدف می‌پردازد.

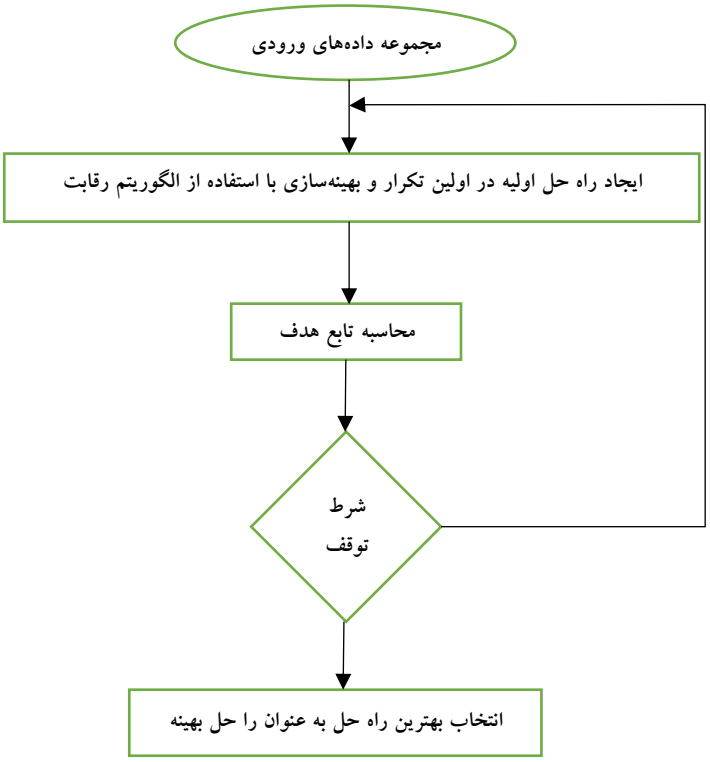
معرفی مجموعه داده

در این مقاله از مجموعه داده اعتباری آلمانی از پایگاه داده UCI[1] که توسط پروفسور هافمن[2] ارائه گردید استفاده شد. این مجموعه داده شامل بیست ویژگی از ۱۰۰۰ مشتری متقاضی وام است. ۷ ویژگی عددی شامل مدت زمان بازپرداخت، مقدار اعتبار، نرخ قسط بر درآمد، مدت اقامت در محل سکونت فعلی، سن، تعداد اعتبارات موجود در بانک، تعدا افراد تحت تکلف و ۱۳ ویژگی کیفی شامل مانده حساب، سابقه اعتبار، هدف از دریافت وام، مقدار موجودی / اوراق قرضه، مدت زمان اشتغال به شغل فعلی، وضعیت تاهل، نوع ضمانت، املاک و مستغلات، برنامه‌های اقساطی دیگر، وضعیت سکونت، وضعیت اشتغال، مالکیت تلفن و کارگر خارجی می‌باشد. ما این مدل را با مجموعه داده یک بانک خصوصی ایرانی شامل اطلاعات ۸٤٥ مشتری با ۱۰ ویژگی مورد ارزیابی قرار دادیم. ویژگی‌های عددی شامل سن، مبلغ تسهیلات، نرخ بهره، مدت زمان بازپرداخت و ویژگی‌های کیفی شامل جنسیت، وضعیت تاهل، عنوان بخش اقتصادی، نوع تسهیلات، روش بازپرداخت و علت دریافت می‌باشد.

1. https://archive.ics.uci.edu/ml/datasets/Statlog+(German+Credit+Data)
2. Hans Hofmann

www.SID.ir

Archive of SID



نمودار ۱. گردش کار روش پیشنهادی

جدول ۳. مشخصات دو مجموعه داده جرمن و بانک ایرانی

نام مجموعه داده	تعداد رکورد	تعداد فیلد	نوع داده	منبع	تاریخ جمع‌آوری
بانک آلمان	۱۰۰۰	۲۱	عددی ـ اسمی	پایگاه داده یو سی آی	
بانک ایرانی	۸٤٥	۱۱	عددی ـ اسمی	–	

انتخاب پارامترها

تنظیم پارامتر مناسب و دقیق الگوریتم‌های فراابتکاری اثرات قابل توجهی بر عملکرد آن‌ها دارد که باید با توجه به مسئله تحت بررسی انجام گیرد. با این وجود بسیاری از محققان این کار را با استفاده از مقادیر مرجع موجود در مقالات مشابه انجام می‌دهند.

www.SID.ir

Archive of SID

در این کار از روش تاگوچی برای تنظیم مقدار پارامترهای الگوریتم رقابت استعماری استفاده شده است. این روش توسط تاگوچی در اوایل ۱۹۶۰ ارائه شده است. با استفاده از روش تاگوچی برای الگوریتم ژنتیک ۹ آزمایش و برای الگوریتم رقابت استعماری ۲۷ آزمایش تعیین شد.

ساختار "کروموزوم" در الگوریتم ژنتیک و "کشور" در الگوریتم رقابت استعماری یک ماتریس صفر و یک با ابعاد 20×1 و مشابه یکدیگر است. این ماتریس صفر و یک همان ویژگی‌های انتخابی از بین بیست ویژگی اصلی است (صفر نشانگر عدم انتخاب ویژگی و یک نشانگر انتخاب ویژگی). تابع هدف استفاده شده در هر دو الگوریتم میزان خطای پیش‌بینی داده‌های تست برای روش طبقه‌بندی شبکه عصبی است.

مقدار پاسخ دریافت شده از الگوریتم رقابت استعماری به پارامترهایی مانند تعداد امپریالیست، حداکثر تعداد کلونی، نرخ انقلاب، نرخ جذب بستگی دارد. پارامترهای بدست آمده از روش تاگوچی بشرح جدول (۳) می‌باشد.

جدول ۳. پارامترهای استفاده شده در الگوریتم بهینه‌سازی رقابت استعماری

تعداد امپریالیست	تعداد کلونی	نرخ انقلاب	zeta	تعداد دهه	نرخ جذب	درجه جذب
٦٠	۷	۰/۲	۰/۱	۱۰۰	۲	تصادفی

مقدار پاسخ دریافت شده از الگوریتم ژنتیک به پارامترهایی مانند تعداد جمعیت اولیه، احتمال جهش، احتمال تقاطع بستگی دارد. پارامترهای بدست آمده از روش تاگوچی بشرح جدول (٤) می‌باشد.

جدول ٤. پارامترهای استفاده شده در الگوریتم ژنتیک

جمعیت اولیه	احتمال تقاطع	احتمال جهش	تعداد نسل	عملگر تقاطع	عملگر جهش	عملگر انتخاب
۵۰	۰/۷	۰/۲	۱۰۰	یکنواخت	دودویی	چرخ رولت

www.SID.ir

Archive of SID

نتایج عملی

ما روش فوق را بروی مجموعه داده جرمن و مجموعه داده یک بانک خصوصی ایرانی مورد آزمایش قرار دادیم و تابع هدف میزان خطای تست شبکه عصبی را کمینه کردیم. این آزمایشات نشان داد با حذف ویژگی‌های کم اثر و در نتیجه کاهش حجم داده‌ها میزان خطای شبکه عصبی در سطح قابل قبولی باقی می‌ماند و در مواردی بهبود می‌یابد. ویژگی‌های انتخابی مجموعه داده جرمن و مجموعه داده بانک خصوصی با دو الگوریتم رقابت استعماری و الگوریتم ژنتیک در جدول (۵) آمده است. روند همگرایی دو الگوریتم در بهینه‌سازی ویژگی‌های مجموعه داده جرمن مورد ارزیابی قرار گرفت و نتایج آن در نمودار (۲) نشان داده شده است.

در جدول (۵) نتایج حاصل از اجرای الگوریتم ترکیبی رقابت استعماری و شبکه عصبی و همچنین الگوریتم ژنتیک و شبکه عصبی با دو مجموعه داده ورودی (مجموعه داده واقعی پایگاه داده UCI و مجموعه داده بانک خصوصی ایرانی) آمده است. همینطور ویژگی‌های انتخابی هر کدام از دو روش ذکر شده است.

با توجه به جدول ۵ مقدار بهینه الگوریتم رقابت استعماری با ۹ ویژگی انتخابی بهتر از مقدار بهینه الگوریتم ژنتیک با تعداد ویژگی انتخابی بیشتر، است.

نمودار (۲) روند همگرایی الگوریتم رقابت استعماری و الگوریتم ژنتیک را مقایسه کرده است. با توجه به این نمودار نقطه شروع دو الگوریتم ژنتیک و رقابت استعماری متفاوت است. با این وجود الگوریتم ژنتیک از تکرار هفتم تا پانزدهم سرعت همگرایی بیشتری داشته است اما از تکرار پانزدهم به بعد الگوریتم رقابت استعماری بهبود چشمگیری در تابع هدف داشته و سرعت همگرایی آن نسبت به الگوریتم ژنتیک بیشتر بوده است. در نهایت الگوریتم رقابت استعماری با تعداد ویژگی انتخابی کمتر از الگوریتم ژنتیک (طبق جدول (۵)) به جواب بهینه بهتری رسیده است.

www.SID.ir

Archive of SID

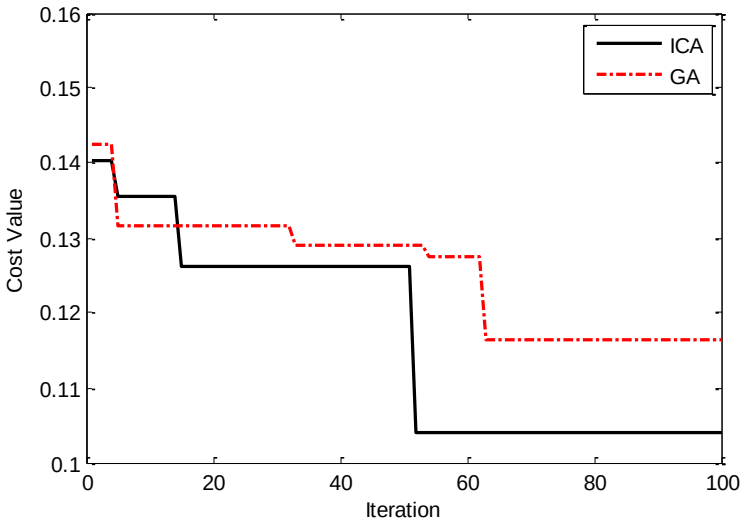
جدول ۳. ویژگی‌های انتخاب شده توسط الگوریتم رقابت استعماری و ژنتیک و مقدار تابع هدف برای مجموعه داده جرمن و بانک خصوصی

مجموعه داده ـ ویژگی	ویژگی‌ها	الگوریتم بهینه‌سازی			
		الگوریتم رقابت استعماری		الگوریتم ژنتیک	
		مجموعه داده جرمن	مجموعه داده بانک خصوصی	مجموعه داده جرمن	مجموعه داده بانک خصوصی
۱-۱	مانده حساب	✓		✓	
۱-۲	مدت زمان پرداخت وام	✓		✓	
۱-۳	سابقه اعتبار	✓		✓	
۱-٤	هدف از دریافت وام			✓	
۱-۵	مقدار اعتبار	✓		✓	
۱-٦	مقدار موجودی اورق قرضه	✓		✓	
۱-۷	مدت زمان اشتغال در شغل فعلی			✓	
۱-۸	نرخ قسط بر درآمد	✓			
۱-۹	وضعیت تاهل	✓		✓	
۱-۱۰	نوع ضمانت			✓	
۱-۱۱	مدت اقامت در محل سکونت فعلی	✓			
۱-۱۲	املاک و مستغلات				
۱-۱۳	سن	✓		✓	
۱-۱٤	برنامه‌های اقساطی دیگر	✓		✓	
۱-۱۵	وضعیت سکونت				
۱-۱٦	تعداد امتیازات موجود در این بانک	✓			
۱-۱۷	وضعیت اشتغال			✓	
۱-۱۸	تعداد افراد تحت تکلف				
۱-۱۹	مالکیت تلفن				
۱-۲۰	کارگر خارجی	✓		✓	
۲-۱	جنسیت				✓
۲-۲	وضعیت تاهل		✓		✓
۲-۳	سن				

www.SID.ir

Archive of SID

مجموعه داده ـ ویژگی	ویژگی‌ها	الگوریتم بهینه‌سازی			
		الگوریتم رقابت استعماری		الگوریتم ژنتیک	
		مجموعه داده جرمن	مجموعه داده بانک خصوصی	مجموعه داده جرمن	مجموعه داده بانک خصوصی
۲-٤	عنوان بخش اقتصادی		✓		✓
۲-٥	نوع تسهیلات		✓		
۲-٦	مبلغ تسهیلات				
۲-۷	نرخ بهره				
۲-۸	مدت زمان وام				
۲-۹	روش بازپرداخت				✓
۲-۱۰	علت دریافت		✓		
مقدار تابع هدف		۰,۱۰٤۱	1.1933×10^{-18}	۰,۱۱٦٤	2.7782×10^{-18}



نمودار ۲. روند همگرایی الگوریتم‌های بهینه سازی رقابت استعماری و ژنتیک

برای تصدیق مدل به مقایسه مقدار خطای شبکه عصبی با مقدار خطای ٤ روش دسته‌بندی دیگر پرداخته شد. بدین منظور به مینیمم‌سازی خطای اعتبارسنجی ضربدری در روش‌های طبقه‌بندی درخت تصمیم، شبکه بیزی، k نزدیک‌ترین همسایه و ماشین

www.SID.ir

Archive of SID

بردار پشتیبان پرداخته شد. نتایج عددی این مقایسه در جدول (٦) آمده است. این نتایج نشان می‌دهد

جدول ٤. مقایسه مجموعه داده اصلی و مجموعه داده بهینه در میزان خطای اعتبارسنجی ضربدری با استفاده از ٥ روش طبقه‌بندی

تابع هدف	میزان خطای اعتبارسنجی ضربدری	
	الگوریتم ژنتیک	الگوریتم رقابت استعماری
درخت تصمیم	۰/۲۳۲۰	۰/۲۲۳۰
شبکه بیزی	۰/۲۸۷۰	۰/۲۸۷۰
K نزدیک‌ترین همسایه	۰/۲۸۷۰	۰/۲۸٦۰
ماشین بردار پشتیبان	۰/۲۲۵۰	۰/۲۲٦۰
شبکه عصبی	۰/۱۲٤۸	۰/۱۲٤۳

با توجه به جدول ٦ مقدار بهینه تابع هدف شبکه عصبی بهترین مقدار نسبت به روش‌های دیگر است. دقت این روش‌ها در مسئله ریسک اعتباری برای داده جرمن به ترتیب نزولی عبارتند از شبکه عصبی، ماشین بردار پشتیبان و درخت تصمیم، k نزدیک‌ترین همسایگی و شبکه بیزی.

نتیجه‌گیری و تحقیقات آتی

این مقاله مسئله انتخاب ویژگی‌های بهینه در ارزیابی ریسک اعتباری مشتریان بانکی را مورد بررسی قرار داد. استفاده از الگوریتم‌های فراابتکاری برای انتخاب ویژگی‌های بهینه نتایج قابل قبولی به همراه داشته است در نتیجه مدل ترکیبی الگوریتم بهینه‌سازی رقابت استعماری و شبکه عصبی در این پژوهش مورد آزمایش قرار گرفت. الگوریتم رقابت استعماری و الگوریتم ژنتیک مقایسه شدند ونتایج بدست آمده در جدول (٥) و (٦) و نمودار (٢) نشان می‌دهد مدل ارائه شده می‌تواند ویژگی‌های بهینه مشتریان بانکی را با دقت قابل قبولی شناسایی کند. روش ترکیبی شبکه عصبی و الگوریتم بهیه‌سازی رقابت استعماری برای اولین بار در حوزه ریسک اعتباری مشتریان بانکی در این مقاله ارزیابی شد و نتایج رضایت بخشی را از خود نشان داد. برای ارزیابی مدل به مقایسه مدل شبکه عصبی با روش‌های دیگر طبقه‌بندی همچون درخت تصمیم، شبکه بیزی، k

www.SID.ir

Archive of SID

نزدیک‌ترین همسایه و ماشین بردار پشتیبان به‌عنوان تابع هدف پرداخته شد و نتایج حاصل که در جدول (٦) آمده است نشان داد مدل ارائه شده دقت بالایی داشته و به خوبی قادر به شناسایی ویژگی‌های بهینه از بین کل ویژگی‌ها است.

با توجه به پیچیدگی فرایندهای اقتصادی و مالی نیاز موسسات مالی و بانک‌ها به پیش‌بینی‌های دقیق با قابلیت اطمینان بیشتر روزافزون خواهد بود. بدین ترتیب استفاده از سیستم‌های فازی یا در نظر گرفتن احتمالات در پیش‌بینی ریسک‌های مالی می‌تواند بر دقت روش‌های موجود بیافزاید و روش‌های نوینی خلق کند. همچنین می‌توان با در نظر گرفتن سایر اروش‌های داده‌کاوی به بررسی و توسعه این مدل پرداخت.

www.SID.ir

Archive of SID

منابع

زهرا جلائیان زعفرانی؛ مهدی سیف برقی، طراحی سیستم پشتیبان تصمیم گیری برای ارزیابی عملکرد کارکنان؛ مطالعه ی موردی در بانک توسعه ی صادرات ایران، دوره ۱، شماره ۳، بهار ۱۳۹۲، صفحه ۱-۱۸.

افسر امیر، هوشدارمحجوب رحمت، مینایی بیدگلی بهروز، خوشه بندی اعتباری مشتریان برای ارائه تسهیلات متناسب، **پژوهش‌های مدیریت در ایران**، ۱۷(٤)، زمستان ۱۳۹۲، ۱-۲٤.

البرزی محمود، خان بابایی محمد، محمدپور زرندی محمد ابراهیم، بکارگیری تکنیک های خوشه بندی و الگوریتم ژنتیک در بهینه سازی درختان تصمیم گیری برای اعتبارسنجی مشتریان بانک ها، **آینده‌پژوهی مدیریت**، دوره ۱,۱(۱)، زمستان ۱۳۹۱، ۱۵-۳٤.

عیسی‌زاده سعید، منصوری گرگری حامد، مقایسه کارایی مدل های کلاسیک و شبکه های عصبی در برآورد ریسک و ظرفیت اعتباری مشتریان شواهدی از بانک تجارت، **اقتصاد پولی، مالی (دانش و توسعه)**، ۲۰(۵)، بهار و تابستان ۱۳۹۲، ۸۷ -۱۱٤.

مهرآرا محسن، موسایی میثم، تصوری مهسا، حسن زاده آیت، رتبه بندی اعتباری مشتریان حقوقی بانک پارسیان، **فصلنامه مدلسازی اقتصادی**، سال سوم، ۳(۹)، پاییز ۱۳۸۸، ۱۵۰-۱.

Bekhet H., Eletter S., "Credit risk assessment model for Jordanian commercial banks: Neural scoring approach", **Review of Development Finance**, University of Cairo., 4, 20–28, 2014.

Carling K., Jacobson T., Lindé J., Roszbach K., "Corporate credit risk modeling and the macroeconomy", **Journal of Banking & Finance**, 31, 3, 845–868, 2007.

Chen W., Xiang G., Liu Y., Wang K., "Credit risk Evaluation by hybrid data mining technique", **Systems Engineering Procedia**, 3, 194–200, 2012.

Danenas, P., Garsva, G., & Gudas, S. "Credit risk evaluation model development using support vector based classifiers". **Procedia Computer Science**, 4, 1699-1707, 2011.

Danenas P., Garsva G., "Selection of Support Vector Machines based classifiers for credit risk domain", **Expert Systems with Applications**, Elsevier Ltd, 42, 6, 3194–3204, 2015.

Finlay S., "Multiple classifier architectures and their application to credit risk assessment", **European Journal of Operational Research**, Elsevier B.V., 210, 2, 368–378, 2011.

Florez-Lopez R., Ramon-Jeronimo J. M., "Enhancing accuracy and interpretability of ensemble strategies in credit risk assessment. A correlated-adjusted decision forest proposal", **Expert Systems with Applications**, Elsevier, March, 5737-5753, 2015.

Gönen, G. B., Gönen, M., & Gürgen, F. (2012). Probabilistic and discriminative group-wise feature selection methods for credit risk

www.SID.ir

Archive of SID

analysis. **Expert Systems with Applications**, 39(14), 11709-11717.
Harris T., "Credit scoring using the clustered support vector machine", **Expert Systems with Applications**, 42, 741–750, 2015.
Hsieh N.-C., "An integrated data mining and behavioral scoring model for analyzing bank customers", **Expert Systems with Applications**, 27, 4, 623–633, 2004.
Huang, C. L., Chen, M. C., & Wang, C. J. Credit scoring with a data mining approach based on support vector machines. **Expert systems with applications**, 33(4), 847-856, 2007.
Kohavi R., Kohavi R., "Wrappers for feature subset selection", **Artificial Intelligence**, 97, 1-2, 273–324, 1997.
Mohamad S. K., Tasir Z., "Educational Data Mining: A Review", Procedia - Social and Behavioral Sciences, Elsevier B.V., 97, 320–324, 2013.
Mandala I. G. N. N., Nawangpalupi C. B., Praktikto F. R., "Assessing Credit Risk: An Application of Data Mining in a Rural Bank", **Procedia Economics and Finance**, The Authors, 4, Icsmed, 406–412, 2012.
Marinaki M., Marinakis Y., Zopounidis C., "Honey Bees Mating Optimization algorithm for financial classification problems", **Applied Soft Computing Journal**, Elsevier B.V., 10, 3, 806–812, 2010.
Marinakis, Y., Marinaki, M., Doumpos, M., & Zopounidis, C. Ant colony and particle swarm optimization for financial classification problems. **Expert Systems with Applications**, 36(7), 10604-10611. 2009.
Oreski S., Oreski G., "Genetic algorithm-based heuristic for feature selection in credit risk assessment", **Expert Systems with Applications**, Elsevier Ltd, 41, 4, 2052–2064, 2014.
Oreski S., Oreski G., "Genetic algorithm-based heuristic for feature selection in credit risk assessment", **Expert Systems with Applications**, Elsevier Ltd, 41, 4 PART 2, 2052–2064, 2014.
Oreski S., Oreski D., Oreski G., "Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment", **Expert Systems with Applications**, Elsevier Ltd, 39, 16, 12605–12617, 2012.
Piramuthu S., "On preprocessing data for financial credit risk evaluation", **Expert Systems with Applications,** 30, 3, 489–497, 2006.
RonKohavi, George H.John, "Wrappers for feature subset selection", **Artificial Intelligence**, 97, 1-2, 273–324, 1997.
Yap B. W., Ong S. H., Husain N. H. M., "Using data mining to improve assessment of credit worthiness via credit scoring models", **Expert Systems with Applications**, Elsevier Ltd, 38, 10, 13274–13283, 2011.
Unler, A., & Murat, A. (2010). A discrete particle swarm optimization method for feature selection in binary classification problems. **European Journal of Operational Research**, 206(3), 528-539.
Yap, B. W., Ong, S. H., & Husain, N. H. M. (2011). Using data mining to improve assessment of credit worthiness via credit scoring models. **Expert Systems with Applications**, 38(10), 13274-13283.

www.SID.ir