فصلنامه علمی – پژوهشی مطالعات مدیریت صنعتی – سال سیزدهم، شماره۳۹، زمستان ۹۴
صفحات ۳۴-۱

یک روش تکاملی برای طبقه‌بندی اعتباری مبتنی بر رویکرد تجمیع زدایی ترجیحات

امیر دانشور،[۱] مصطفی زندیه،[۲] جمشید ناظمی[۳]

تاریخ دریافت(۹۳/۷/۲۹)تاریخ پذیرش(۹۴/۶/۷)

چکیده:

مدل‌های مبتنی بر روابط برتری یک شاخه مهم از روش‌های تصمیم چندمعیاره هستند که نیاز به تعریف مقدار قابل توجهی اطلاعات ترجیحی در قالب پارامترها توسط تصمیم گیرنده دارند. تعدد پارامترها، معنای اغلب گیج کننده آنها در فضای مسئله و طبیعت غیردقیق داده‌ها، این فرآیند را خصوصاً در مسائل طبقه بندی اعتباری با ابعاد بزرگ که نیاز به تصمیم‌گیری بلادرنگ است بسیار زمان‌بر و پیچیده می‌سازد. بدین منظور رویکرد تجمیع زدایی ترجیحات این اطلاعات را از طریق قضاوت‌های جامعی که توسط تصمیم گیرنده فراهم می‌شوند استنتاج می‌کند. این رویکرد در تصمیم گیری چندمعیاره معادل یادگیری ماشینی در حوزه هوش مصنوعی است.

تحت این رویکرد، این مقاله یک روش جدید پیشنهاد می‌کند که در آن الگوریتم ژنتیک طی فرآیند یادگیری، به طور همزمان تمامی پارامترهای مدل ELECTRE TRI را از داده های آموزشی استنتاج و در خاتمه فرآیند، پارامترهای استنتاج شده برای طبقه بندی داده های آزمایشی بکار گرفته می‌شوند. تحلیل آزمایشات روی دیتاست های اعتباری نشان از کیفیت بالا و قابل رقابت روش پیشنهادی در مقایسه با مدل‌های استاندارد طبقه‌بندی دارد.

واژگان کلیدی:

مدل ELECTRE TRI، تجمیع زدایی ترجیحات، طبقه بندی اعتباری، یادگیری ماشینی، الگوریتم ژنتیک

---

۱ – دانشجوی دکتری گروه مدیریت صنعتی دانشگاه آزاد اسلامی واحد علوم و تحقیقات (نویسنده مسئول)
Daneshvar.amir@gmail.com
۲ دانشیار دانشکده مدیریت و حسابداری دانشگاه شهید بهشتی
۳ – استادیار دانشکده مدیریت و اقتصاد دانشگاه آزاد اسلامی واحد علوم و تحقیقات

مقدمه

صنعت اعتباری[1] در طول چند سال اخیر به سرعت در حال گسترش است. به دلیل رقابت شدید در این صنعت، افراد بیشتر و بیشتری می‌توانند به کارت‌های اعتباری دسترسی داشته باشند بدون اینکه وضعیت اعتباری آنان به دقت توسط بانک‌ها بررسی شود[9]. در این راستا طبقه‌بندی اعتباری[2] به عنوان ابزار مدیریت ریسک اعتباری به طور گسترده توسط بانک‌ها و موسسات مالی استفاده شده و به آنها کمک می‌کند مشتریان اعتباری خود را بر اساس معیارهایی چون درآمد، سن، گردش حساب بانکی، شغل، تحصیلات به مشتریان خوش حساب و بد حساب طبقه‌بندی کنند. کاهش امکان ریسک ناشی از تصمیمات اعتباری نادرست، صرفه جویی در زمان و هزینه و نیز بهبود تصمیمات اعتباری از مزایای استفاده از تکنیک‌های طبقه بندی اعتباری است[25].

به طور سنتی بانک‌ها و مؤسسات مالی از تکنیک‌های آماری مانند تحلیل تمایزی[3]، تحلیل پروبیت[4]، رگرسیون لجستیک[5] برای مسئله طبقه بندی اعتباری استفاده می‌کرده‌اند. آلتمن [4] و توماس [32] این تکنیک‌ها را به طور گسترده مرور کرده‌اند. مطالعات اخیر روش‌های جدید مانند درخت تصمیم[6][33,34]، شبکه عصبی‌مصنوعی[7][1,24]، ماشین بردار پشتیبان[8] [7,20]، برنامه‌ریزی ریاضی[9][8,30] و مدل‌های چند معیاره پشتیبان تصمیم[10] [5,16,35] را بدین منظور به کار گرفته‌اند.

1 -Credit Industry
2 -Credit Scoring
3 -Discriminate Analysis
4 -Probit and Logit Analysis
5 -Lofistics Regression
6 -Decision Trees
7 -Artificial Neural Networks(ANN)
8 -Support Vector Machines(SVM)
9- Mathematical Programming
10- Multi Criteria Decision Aid(MCDA)

بیان **مسئله**

مدل‌های پشتیبان تصمیم چندمعیاره تاکنون کاربردهای فراوانی در زمینه حل مسائل مالی مانند طبقه‌بندی اعتباری، پیش بینی ورشکستگی، مدیریت پورتفولیو، سرمایه‌گذاری مخاطره‌پذیر[1] و ریسک کشورها[2] داشته‌اند[37].

غالب مدل های طبقه بندی MCDA مبتنی بر روابط برتری[3] هستند که مدل ELECTRE TRI شناخته شده ترین و پرکاربردترین مدل این گروه است. پیاده سازی این مدل علیرغم قابل فهم بودن و توسعه یافتگی مستلزم تعیین تعداد زیادی پارامتر است و این فرآیند بسیار مشکل و وقتگیر است بنابراین عملاً در موقعیت‌های واقعی طبقه بندی اعتباری که تصمیمات در آن باید به صورت بلادرنگ گرفته شوند کاربرد نخواهند داشت.

مرور ادبیات

تاکنون چندین مقاله به استفاده از مدل‌های پشتیبان تصمیم چندمعیاره در طبقه بندی اعتباری پرداخته‌اند. داومپوس و زوپونیدیس [16] سه مدل پشتیبان تصمیم چندمعیاره الکتر تری[4]، مطلوبیت افزایشی تمایزی[5] و مجموعه های راف[6] را برای مسائل طبقه‌بندی مالی شامل طبقه‌بندی اعتباری، پیش بینی ورشکستگی و ارزیابی سهام استفاده‌کردند. مقایسه نتایج نشان داد که مدل‌های پشتیبان تصمیم چندمعیاره یک آلترناتیو قابل رقابت با مدل‌های آماری در حوزه مسائل مالی و بانکی هستند. خلیل، مارتل و جوتراس [23] یک سیستم مبتنی بر الکتر تری جهت ارزیابی کیفیت اعتباری شرکت‌ها و طبقه‌بندی آنها به کلاس‌های همگن ارائه دادند. دامپوس، کازدیمودو، بوراکیس و زوپونیدیس [13]، کاربرد روش تمایز سلسله مراتبی چند گروهی[7] در طبقه بندی شرکتهای متقاضی وام به گروه های همگن را بررسی و نتایج را با

1- Venture capital investment
2 - Country risk
3 -Outranking relations
4- ELECTRE TRI
5-Utility Additivr Discrimination(UTADIS)
6 -Rough Sets
7 -Multi Group Hierarchical Discrimination(M.H.DIS)

چند تکنیک استاندارد طبقه بندی مقایسه کردند. بوراکیس و دیگران [5] قابلیت روش تصمیم گیری چندمعیاره مطلوبیت افزایشی تمایزی در توسعه سیستمهای ارزیابی ریسک اعتباری مؤسسات مالی را مورد بررسی قرار دادند. داومپوس و پاسیوراس [15] مطلوبیت افزایشی تمایزی را با استفاده از داده‌هایی که به صورت عمومی در دسترس هستند پیاده‌سازی نمودند، هدف این بود که نتایج آنها رتبه بندی انجام شده آژانس‌های رتبه بندی محلی را تکرار کند.

مدل‌های استفاده شده درمقالات فوق الذکر را می‌توان در سه گروه دسته‌بندی کرد؛ اول مدل‌های مبتنی بر قانون[1] مانند مجموعه های راف[18]، دوم مدل‌های مبتنی بر توسعه توابع مطلوبیت مانند مدل مطلوبیت افزایشی تمایزی و سوم مدل‌های مبتنی بر روابط برتری[2] که شناخته‌شده‌ترین آنها ELECTRE TRI است.

در مقایسه این سه دسته، مدل‌های دسته اول آزاد از پارامتر[3] هستند، ولی پیاده‌سازی مدل‌های مربوط به دسته‌های دوم و سوم مستلزم تعیین تعداد زیادی پارامتر توسط تصمیم‌گیرنده است. این پارامترها با یکی از دو روش زیر بدست می‌آیند:

- **روش مستقیم[4]:**

در این روش پارامترها به طور مستقیم از تصمیم‌گیرنده پرسیده می‌شوند. در بسیاری موقعیت‌های تصمیم‌گیری مثل‌طبقه‌بندی اعتباری تعیین ارزش این پارامترها مستلزم تلاش طاقت‌فرسا و هزینه فراوان است[2,3,6,11,14,26,29] لذا انجام چنین فرآیندی در بسیاری از مسائل مالی که تصمیمات باید به سرعت گرفته شوند عملاً غیرقابل توجیه است[36].

- **روش غیرمستقیم[5]:**

در این روش تصمیم‌گیرنده به جای تعیین مقدار پارامترها، مجموعه‌ای از مثال‌های تخصیص[6] را ارائه می‌کند که هر یک از این مثالها، نمایشی از خط مشی قضاوتی و سیستم ترجیحات[1]

1 -Rule-based
2 -Outranking Relations
3 -Parameter free
4 -Direct method
5 -Indirect method
6 -Assignment Examples

وی بوده و از درون این مثال‌ها می‌توان پارامترها را استنتاج کرد به نحوی که نتایج مدل با مثال‌های تخصیص حداکثر سازگاری را داشته باشد[21]. روش غیرمستقیم به نام تجمیع زدایی ترجیحات[۲] خوانده می‌شود. تجمیع زدایی ترجیحات در حوزه پشتیبانی تصمیم چند معیاره معادل یادگیری ماشینی در حوزه هوش مصنوعی و مثال‌های تخصیص نیز معادل مجموعه آموزشی[۳] در یادگیری ماشینی است[22].

با مطالعه ادبیات تحقیق با دو رویکرد در روش غیرمستقیم مواجه هستیم:

-رویکرد اول:

استفاده از مدل‌های کلاسیک برنامه‌ریزی ریاضی است[11,12,25,26,27,29]. این روش‌ها قابلیت تخمین کلیه پارامترهای مدل‌های برتری را ندارند یا آنکه برای مسائل با اندازه واقعی کاربردی نیستند [14]، ضمن اینکه آنها صرفاً رویه تخصیص بدبینانه مدل ELECTRE TRI را در نظر گرفته‌اند.

-رویکرد دوم:

استفاده از الگوریتم های فراابتکاری است. جابر و گیتونی[22] از الگوریتم ژنتیک برای استنتاج پارامترهای مدل PROAFTN استفاده و مدل خود را در یک مسئله نظامی بکار گرفتند آنها پروفایل ها و وزن معیارها را در فرآیند استنتاج پارامترها منظور نکردند.گولتسیس[17] روابط برتری را با الگوریتم ژنتیک ترکیب و در یک مسئله پزشکی با دو کلاس استفاده نمود او صرفاً از روابط برتری استفاده کرد و پارامترهای وتو، پروفایل و درجه اعتبار را در نظر نگرفت. دامپوس و دیگران[14] از الگوریتم فراابتکاری تفاضل تکاملی برای بدست آوردن پارامترهای یک مدل برتری استفاده کردند. تمرکز اصلی آنها ارزیابی روش پیشنهادی با استفاده از تولید داده‌های مصنوعی و تصادفی بود و از داده‌های واقعی برای مسائل با اندازه بزرگ استفاده نشد ضمن آنکه عملکرد روش خود را نیز با سایر الگوریتم‌های طبقه بندی مقایسه نکردند. ال اُبیدات و دیگران[2,3] الگوریتم‌های فراابتکاری

1 -Preferential system
2 -Preferences disaggregation
3 -Training set

بهینه‌سازی ازدحام ذرات[1] و تفاضل تکاملی[2] را برای بدست آوردن پارامترهای مدل تصمیم‌گیری PROAFTN بکار گرفتند و روش طبقه بندی خود را برای چندین دیتاست پزشکی با تعداد کلاس‌های مختلف، پیاده سازی کرده و نتایج رضایت‌بخشی گرفتند. در هر دو مقاله مزبور، آنها پارامتر وتو را به علت پیچیدگی‌های آن نادیده گرفتند.

بیان مسئله

مدل‌های پشتیبان تصمیم چندمعیاره تاکنون کاربردهای فراوانی در زمینه حل مسائل مالی مانند طبقه بندی اعتباری، پیش بینی ورشکستگی، مدیریت پورتفولیو، سرمایه‌گذاری مخاطره-پذیر[3] و ریسک کشورها[4] داشته‌اند[38].

غالب مدل‌های طبقه بندی MCDA مبتنی بر روابط برتری[5] هستند که مدل ELECTRE TRI شناخته شده‌ترین و پرکاربردترین مدل این گروه است. پیاده سازی این مدل علیرغم قابل فهم بودن و توسعه یافتگی مستلزم تعیین تعداد زیادی پارامتر است و این فرآیند بسیار مشکل و وقتگیر است بنابراین عملاً در موقعیت‌های واقعی طبقه بندی اعتباری که تصمیمات در آن باید به صورت بلادرنگ گرفته شوند کاربرد نخواهند داشت.

همانگونه که در جدول ۱ آمده است روشهای برنامه‌ریزی ریاضی برای استخراج پارامترهای مدل‌های مبتنی بر روابط برتری، اغلب برای استنتاج یک یا چند پارامتر سایر پارامترها را ثابت نگه داشته یا نادیده می‌گرفتند هر چند تلاشهایی برای استخراج همزمان کلیه پارامترها نیز منجر به توسعه مدل‌های غیرخطی و غیرمحدب پیچیده و ناکارآمد بود و مدل‌های حاصل از رویکرد برنامه‌ریزی ریاضی برای کاربردهای دنیای واقعی از جمله طبقه‌بندی اعتباری کارایی نداشت ضمن آنکه همه این روشها رویه تخصیص بدبینانه را در نظر گرفتند. روش‌های فرابتکاری نیز به منظور استخراج پارامترهای مدل های برتری به کار

1- Particle Swarm Optimisation(PSO)
2 -Differential Evolution(DE)
3 -Venture capital investment
4 -Country risk
5 -Outranking relations

گرفته شدند هر چند هیچیک قابلیت استخراج همزمان کلیه پارامترها در دو رویه خوشبینانه و بدبینانه را ندارند و تاکنون قابلیت آنها در طبقه بندی اعتباری آزمایش نشده است.
تحقیق حاضر با در نظر داشتن موارد فوق در نظر دارد با ارائه یک روش بر این محدودیت‌ها فائق آید لذا در روش پیشنهادی این تحقیق با استفاده از الگوریتم ژنتیک کلیه پارامترهای مدل ELECTRE TRI در یک فرآیند تکاملی برای هر دو رویه خوش بینانه و بدبینانه، استنتاج و برای مسئله طبقه بندی مشتریان اعتباری با داده های واقعی در ابعاد بزرگ بکار گرفته می‌شوند. نتایج با سایر الگوریتم‌های شناخته شده طبقه‌بندی مقایسه و نتیجه‌گیری می‌شود.

سوالات تحقیق

۱. چگونه پارامترهای بهینه مدل ELECTRE TRI را از درون مثال‌های تصمیم استنتاج کنیم؟
۲. جزئیات توسعه روش فراابتکاری الگوریتم ژنتیک برای طبقه بندی اعتباری مشتریان مبتنی بر ELECTRE TRI در دو نسخه خوشبینانه و بدبینانه به چه صورت است؟
۳. چه روشی برای اعتبارسنجی روشهای پیشنهادی مناسب است؟

جدول(۱): خلا تحقیقاتی موجود در پیشینه تحقیق

مقالات مرتبط با استخراج پارامترهای مدل های مبتنی بر (*Outranking relations*)		[27]	[26]	[12]	[29]	[17]	[6]	[22]	[3]	[2]	تحقیق حاضر
رویکرد	برنامه ریزی ریاضی	√	√	√	√						
	روشهای فراابتکاری					√	√	√	√	√	√
پارامترهای قابل استخراج	*p*	√				√	√	√	√	√	√
	q	√				√	√	√	√	√	√
	v			√				√			√
	w	√	√			√			√	√	√
	profiles	√			√		√		√	√	√
رویه تخصیص	بدبینانه	√		√	√	√	√	√	√	√	√
	خوشبینانه										√
تعداد اهداف	تک هدفه	√	√	√	√	√	√	√	√	√	√
	چندهدفه										√
کاربرد	طبقه بندی عتباری										√
	سایر کاربردها	√	√	√	√	√	√	√	√	√	

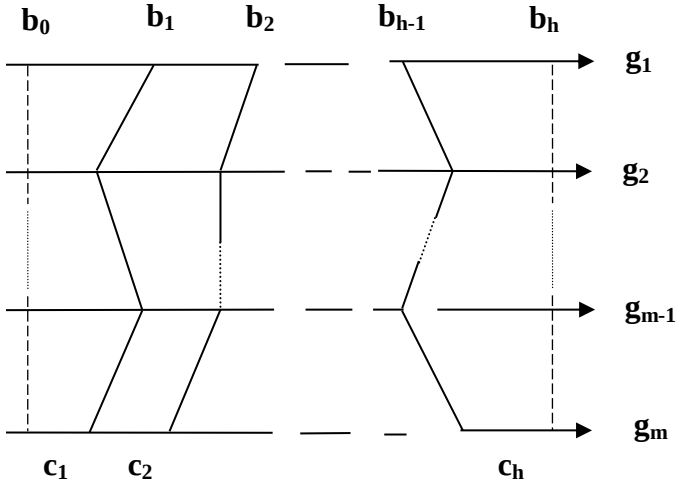
متدولوژی تحقیق

در این بخش ابتدا مدل الکتر تری معرفی، سپس یک روش تکاملی جدید مبتنی بر الگوریتم ژنتیک پیشنهاد می‌شود.

مدل ELECTRE TRI

مدل ELECTRE TRI از خانواده روش‌های پشتیبان تصمیم چندمعیاره الکتر است که اولین بار توسط یو ارائه شد. این مدل، آلترناتیوها را در کلاس‌های از پیش تعریف شده، طبقه‌بندی

می‌کند. این طبقه‌بندی از مقایسه هر آلترناتیو با پروفایل‌هایی[۱] که مبین مرز[۲] طبقات هستند، حاصل می‌شود[28]. چنانچه مطابق شکل 1 برای معیارهای g_1، g_2، ...، g_m، پروفایل‌های b_1، b_2، ...، b_{h-1} (پروفایل های مجازی b_0 و b_h نشان دهنده آلترناتیوهای ضد ایده‌آل و ایده‌آل هستند)در نظر گرفته شود و b_{h-1} حد بالایی طبقه C_{h-1} و حد پایین طبقه C_h باشد، در این حالت h طبقه وجود خواهد داشت.



شکل(۱): تعریف کلاس‌ها با استفاده از پروفایل‌ها

در این روش، رابطه برتری(S) بین آلترناتیوها و پروفایل‌ها برقرار می‌شود. این روابط که به اختصار با aSb_h و یا b_hSa نشان داده می‌شوند، بدین معنی است که «آلترناتیو a حداقل به خوبی پروفایل b_h است[۳]» و بالعکس.

فرض کنیم رابطه aSb_h برقرار باشد باید میزان هماهنگی[۴] و ناهماهنگی[۵] معیارها با این فرض اندازه‌گیری شود. اگر فاصله b_h از a در معیار jام از q_j تجاوز نکند آنگاه معیار مزبور

1- Prifile
2 -Boundary
3 -"*a* outranks *bh*"
4 -Concordance
5 -Discordance

با فرض فوق هماهنگ است پارامتر q_j آستانه بی تفاوتی[1] نامیده می‌شود. به همین ترتیب اگر فاصله b_h از a در معیار jام از p_j تجاوز کند آنگاه معیار مزبور با (aSb_h) هماهنگ نیست پارامتر p_j آستانه ترجیح[2] نامیده می‌شود.

دو پارامتر دیگر نیز در ساختن روابط برتری نقش دارند؛ مجموعه ضرایب وزن معیارها w_j که در شاخص هماهنگی کل استفاده می‌شوند و آستانه‌های وتو[3] v_j که در شاخص ناهماهنگی محاسبه می‌شوند . v_j نشان‌دهنده حداقل اختلاف $g_j(b_h) - g_j(a)$ است که رابطه aSb_h را رد می کند. به عبارت بهتر اگر فاصله b_h از a در معیار jام از v_j تجاوز کند آنگاه معیار مزبور فرض فوق را وتو می کند.

شاخص هماهنگی بخشی[4]

این شاخص $c_j(a,b_h)$ نشانگر میزان هماهنگی هر یک از معیارها با ادعای «گزینه a حداقل به خوبی پروفایل b_h است» می‌باشد. وقتی g_j ارجحیت افزایشی داشته باشد شاخص هماهنگی بخشی $c_j(a,b_h)$ به صورت زیر محاسبه می‌شود:

رابطه (۱)

$$\begin{cases} if\ \ g_j(a) < g_j(b_h) - p_j & c_j(a,b_h) = 0 \\ if\ \ g_j(b_h) - p_j < g_j(a) < g_j(b_h) - q_j & c_j(a,b_h) = \dfrac{g_j(a) - g_j(b_h) + p_j}{p_j - q_j} \\ if\ \ g_j(b_h) - q_j < g_j(a) & c_j(a,b_h) = 1 \end{cases}$$

1 -Indiffrence Threshold
2 -Preference Threshold
3 -Veto Thereshold
4 -Partial Concordance index

شاخص هماهنگی کل

این شاخص حاصل تجمیع مقادیر شاخص هماهنگی بخشی است و بیان می‌کند معیارها تا چه اندازه با عبارت «گزینه a حداقل به خوبی پروفایل bh است» هماهنگ هستند. طریقه محاسبه این شاخص به شرح زیر است:

رابطه (۲)

$$c(a,b_h) = \frac{\sum_{j\in F} k_j . c_j(a,b_h)}{\sum_{j\in F} k_j}$$

شاخص ناهماهنگی

شاخص ناهماهنگی $d_j(a,b_h)$ نشان می‌دهد تا چه اندازه هر یک از معیارها با عبارت «گزینه a حداقل به خوبی پروفایل b_h است» ناهماهنگ هستند. طریقه محاسبه این شاخص به شرح زیر است:

رابطه (۳)

$$\begin{cases} if\ \ g_j(a) > g_j(b_h) - p_j & d_j(a,b_h) = 0 \\ if\ \ g_j(b_h) - v_j < g_j(a) < g_j(b_h) - p_j & d_j(a,b_h) = \dfrac{g_j(b_h) - g_j(a) - p_j}{v_j - p_j} \\ if\ \ g_j(b_h) - v_j(b_h) > g_j(a) & d_j(a,b_h) = 1 \end{cases}$$

شاخص درجه اعتبار[1]

این شاخص حاصل ترکیب شاخص هماهنگی کل و شاخص‌های ناهماهنگی است و بیانگردرجه اعتبار عبارت«گزینه a حداقل به خوبی پروفایل b_h است» می‌باشد.

رابطه (۴)

$$\sigma(a,b_h) = c(a,b_h).\prod_{j\in\bar{F}} \frac{1-d_j(a,b_h)}{1-c(a,b_h)}$$

$$\bar{F} = \left\{ j \in F \,/\, d_j(a,b_h) > c(a,b_h) \right\}$$

1 -Degree of Credibility index

تبدیل روابط فازی به قطعی

برای تبدیل رابطه برتری فازی به قطعی، شاخص سطح برش[1] معرفی می‌شود. سطح برش (λ) حداقل مقدار شاخص درجه اعتبارست که با عبارت «گزینه a حداقل به خوبی پروفایل b_h است» سازگار می‌باشد یعنی:

$$\sigma(a,b_h)\geq\lambda\Rightarrow aSb_h \qquad \text{رابطه (۵)}$$

انواع روابط زوجی آلترناتیوها

روابط ارجحیت[2] ($\succ$)، بی‌تفاوتی[3] (I) و غیرقابل‌مقایسه‌بودن[4] (R) مطابق جدول۲ بین هر زوج از آلترناتیوها تعریف می‌شود.

جدول(۲):. انواع روابط برتری بین آلترناتیوها

a و b_h هر دو به هم برتری دارند لذا بین آنها بی تفاوت هستیم.	$aIb_h \Leftrightarrow aSb_h \text{ and } b_hSa$
a به b_h برتری دارد ولی b_h نمی‌تواند بر a برتری یابد لذا a بر b_h ارجحیت دارد.	$a\succ b_h \Leftrightarrow aSb_h \text{ and not } b_hSa$
a نمی‌تواند بر b_h برتری یابد ولی b_h به a برتری دارد لذا b_h بر a ارجحیت دارد.	$b_h\succ a \Leftrightarrow b_hSa \text{ and not } aSb_h$
a و b_h هیچیک بر دیگری برتری ندارند لذا این دو آلترناتیو غیرقابل مقایسه هستند.	$aRb_h \Leftrightarrow \text{not } aSb_h \text{ and not } b_hSa$

دو رویه بدبینانه[5] و خوشبینانه[6] برای انجام طبقه‌بندی با مدل ELECTRE TRI وجود دارد. در رویه بدبینانه آلترناتیو a به صورت متوالی از بالاترین طبقه با پروفایل‌های b_i مقایسه می‌شود. اول بار که رابطه aSb_h برقرار شود آلترناتیو a به طبقه C_{h+1} تخصیص می‌یابد.

1 -Cutting level
2 -Preference
3 -Indifference
4 -Incomparability
5- Pessimistic
6 -Optimistic

در رویه خوشبینانه آلترناتیو a به صورت متوالی از پایین‌ترین طبقه با پروفایل‌های b_i مقایسه می‌شود. اولین بار که $b_h \succ a$ برقرار شود، آلترناتیو a به C_h تخصیص می‌یابد.

توسعه روش تکاملی پیشنهادی

برای استفاده از مدل ELECTRE TRI، پارامترهای زیر باید مشخص شوند:

۱. بردار وزن $W = (w_1, \dots w_m)$ به طوری‌که $w_1 + \cdots + w_n = 1$

۲. پروفایل کلاس ها $r_1, r_2 \dots r_{b-1}$.

۳. بردار آستانه بی تفاوتی $q = (q_1, \dots, q_m) \geq 0$

۴. بردار آستانه ارجحیت $P = (p_1, \dots, p_m) \geq q$

۵. بردار آستانه بی تفاوتی $V = (v_1, \dots, v_m) \geq P$

۶. مقدار سطح برش $\lambda \in (0.5,1)$.

این پارامترها به همراه ماتریس تصمیم، ورودی های مدل الکتر تری را تشکیل می‌دهند که در یک رویه تخصیص خوشبینانه یا بدبینانه آلترناتیوها را در کلاسهای از پیش تعریف شده طبقه‌بندی می‌کند. شکل ۲ ورودی و خروجی این مدل را نشان می‌دهد.

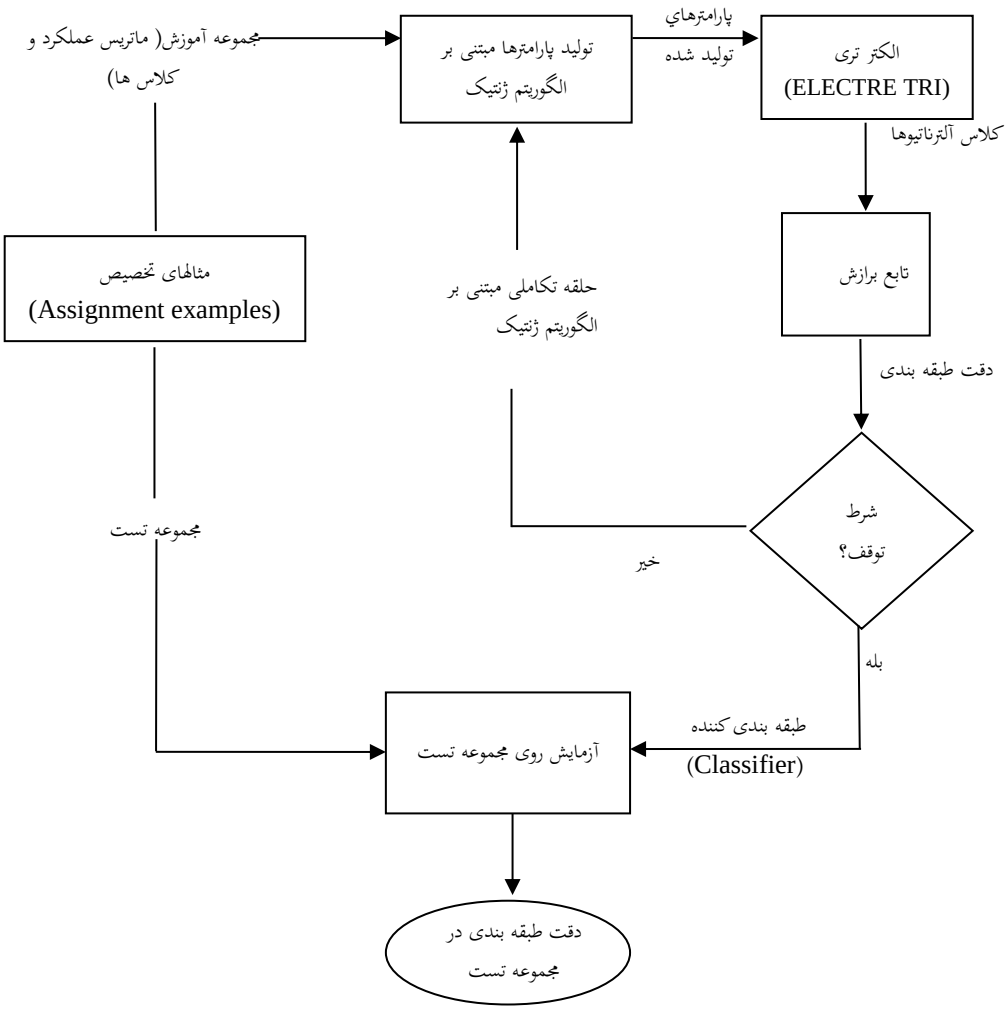
شکل(۲): ورودی و خروجی مدل ELECTRE TRI

اشاره شد که فرآیند تعیین ارزش این پارامترها بسیار طولانی است لذا با مجهول بودن پارامترها عملاً مسئله به صورت یک مسئله یادگیری نظارت شده (طبقه بندی) در می‌آید که در آن با معلوم بودن کلاس آلترناتیوها و ماتریس تصمیم که مجموعاً داده‌های آموزشی را تشکیل می‌دهند، درصدد استنتاج پارامترها هستیم.

- **مدل مفهومی**

در این تحقیق نیز بردار ورودی شامل ماتریس عملکرد است که اطلاعات مربوط به مشتریان اعتباری را در بردارد و بردار خروجی نیز شامل دو کلاس مشتریان خوش حساب و مشتریان بدحساب هستند. درصدد استنتاج طبقه بندی کننده‌ای هستیم که ورودی ها را روی خروجی‌ها نگاشت می کند. این طبقه بندی کننده مدل الکتر تری است که پارامترهای آن مجهول است. مدل مفهومی روش پیشنهادی این تحقیق (که به اختصار GA-ELECTRE نامیده می‌شود) در شکل۳ نشان داده شده است.

الگوریتم ژنتیک در ابتدا جمیعتی از پارامترهای تصادفی را تولید و طبقه‌بندی با استفاده از مدل الکتر تری طبق روابط ۲-۱ الی ۲-۵ انجام می گیرد. نتیجه با طبقه‌بندی واقعی مقایسه و تابع/توابع برازش محاسبه می گردد. طی یک فرایند تکاملی جمعیت‌های جدید از پارامترها تولید و بکار گرفته می‌شوند تا جایی که میزان سازگاری طبقه بندی پیش بینی شده و واقعی به حداکثر مقدار خود برسد یا آنکه سایر شرایط خاتمه الگوریتم محقق شود. مدل الکتر تری با پارامترهای استنتاج شده به عنوان خروجی حاصل از رویه یادگیری برای طبقه بندی داده‌های تست مورد استفاده قرار می گیرند.



شکل(۳): مدل مفهومی روش GA-ELECTRE

تطبیق الگوریتم ژنتیک با مسئله طبقه‌بندی اعتباری مشتریان

در این مقاله، الگوریتم ژنتیک جهت بهینه‌سازی پارامترهای مدل ELECTRE TRI توسعه داده می‌شود. جهت توسعه الگوریتم ژنتیک لازم است پارامترهای آن به شرح زیر تعیین گردد:

-**ماکزیمم تعداد تکرارها:** برابر ۲۰۰ در نظر گرفته می‌شود.

-**اندازه جمعیت:** برابر با عدد ۲۰۰ در نظر گرفته می‌شود.

-**استراتژی انتخاب:** برای انتخاب والدین از استراتژی چرخه رولت[1] استفاده می‌شود.

-**عملگر تقاطع[2]:** در این تحقیق از عملگر تقاطع میانی[3] استفاده می‌شود که تلاش می‌کند میانگین عناصر والدین را محاسبه کند[31]. با معین بودن دو والد p_1 و p_2 عملگر تقاطع میانی، فرزند o را با استفاده از میانگین وزنی حساب می‌کند:

$$O_i = \alpha x_{1i} + (1 - \alpha) x_{2i}$$

عملگر تقاطع میانی یک بار روی کل پارامترها و بار دیگر به طور تصادفی روی برخی پارامترها اعمال می‌شود. دیگر عملگر مورد استفاده، عملگر تقاطع تک نقطه‌ای[4] است. در این مقاله، نرخ تقاطع[5] که نشان دهنده نسبتی از والدین است که روی آنها عملگر تقاطع عمل می‌کند برابر ۱ در در نظر گرفته می‌شود.

-عملگر جهش[6]: برای توسعه الگوریتم مورد نظر از عملگر جهش یکنواخت[7] استفاده می‌شود. این عملگر یک بار روی کل پارامترها و بار دیگر به طور تصادفی روی برخی پارامترها اعمال می‌شود. احتمال جهش[8] یک ژن برابر با احتمال تغییر کردن هر کدام از ژنها می‌باشد که برابر با ۰٫۱ منظور می‌گردد. نرخ جهش[9] که نشان دهنده نسبتی از والدین است که روی آنها عمل جهش انجام می‌شود برابر ۰٫۸ است.

1- Roulette Wheel Strategy
2 -Crossover operator
3- Intermediate crossover
4- Single point crossover
5 -Crossover probability
6 -Mutation operator
7 -Uniform mutation
8 -Mutation Rate
9- Crossover probability

-**اعتبارسنجی**[1]**:** روش استاندارد پیش بینی نرخ خطای یک تکنیک یادگیری با معین بودن یک نمونه از داده های منفرد و ثابت استفاده از روش اعتبارسنجی تقاطعی k بخشی طبقه بندی[2] شده است[25] که در این مقاله مورد استفاده قرار می‌گیرد. تعداد بخشها یا فولدها k=۵ در نظر گرفته می‌شود.

-**تابع برازش**[3]: مطابق ماتریس کانفیوژن که در شکل آمده است از معیار دقت طبقه بندی[4] (CA) به صورت رابطه (۶) به عنوان معیار ارزیابی عملکرد طبقه بندی اعتباری استفاده می‌شودکه نشان دهنده نسبتی از کل مشتریان است(اعم از خوش حساب و بدحساب است) که به درستی طبقه‌بندی شده اند.

	پیش بینی درست	**پیش بینی نادرست**
خوش‌حساب	مشتری خوش حسابی که خوش حساب پیش بینی شده *True Positive(TP)*	مشتری بدحسابی که خوش حساب پیش بینی شده *False Positive(FP)*
بدحساب	مشتری بدحسابی که بدحساب پیش بینی شده *True Negative (TN)*	شتری خوش حسابی که بدحساب پیش بینی شده *False Negative (FN)*

شکل(۴): ماتریس کانفیوژن در طبقه بندی اعتباری

$$CA = \frac{TP + TN}{TP + TN + FP + FN}$$

1-Validation
2-Stratified K-fold cross validation
3-Fitness Function
4- Classification Accuracy

فرم عمومی مسئله بهینه‌سازی به صورت زیر است:

$$\begin{aligned}&\text{Max} \quad CA(w_j, r_k, , q_j, p_j, v_j, \lambda)\\&s.t:\\&\sum W_j = 1\\&v_j \geq p_j \geq q_j \geq 0 \qquad \text{رابطه (۷)}\\&0.5 < \lambda < 1\\&W_j \geq 0\\&j = 1, \dots, m\end{aligned}$$

یکی از مسائل مهم در استفاده از الگوریتم‌های ژنتیک در مسائل بهینه‌سازی محدود شده، چگونگی اداره‌کردن محدودیت‌هاست. از آنجا که عملگرهای ژنتیک کروموزوم‌ها را دستکاری می‌کنند و این ممکن است به تولید فرزندان غیرموجه منجر شود لذا روش‌های مختلفی بدین منظور ارائه شده است که استراتژی جریمه کردن[1] یکی از آنهاست [19]. تخطی[2] از رابطه (7) به صورت زیر تعیین می‌شود:

$$\begin{cases} R_j = Max(q_j - p_j, 0) \\ \bar{R} = \frac{1}{m}\sum(\frac{R_j}{u_j - l_j}) \end{cases} \longrightarrow \quad 0 \leq q_j \leq p_j \text{ جریمه تخطی از محدودیت}$$

$$\begin{cases} S_j = Max(p_j - v_j, 0) \\ \bar{S} = \frac{1}{m}\sum(\frac{S_j}{u_j - l_j}) \end{cases} \longrightarrow \quad 0 \leq p_j \leq v_j \text{ جریمه تخطی از محدودیت}$$

که در آن u_j و l_j حدود بالا و پایین هر معیار، m تعداد کل معیارها می‌معیارهایی هستند که تخطی از هر یک از محدودیت در آنها روی داده و $\bar{R}$ و $\bar{S}$ میانگین جریمه تخطی از هر محدودیت می‌باشند. تابع جریمه به فرم جمع شونده[3] در رابطه (8) آمده است:

$$f: Max \ CA(w_j, r_k, , q_j, p_j, v_j, \lambda) - \frac{1}{2}(\bar{R} + \bar{S}) \qquad \text{رابطه (۸)}$$

1 -Penalizing strategy
2 -Violation
3 -Additive

نمایش جواب‌ها و تولید جمعیت اولیه

با در نظر داشتن متغیرهای تصمیم که همان پارامترهای مدل ELECTRE TRI هستند، ساختار کروموزوم‌ها را به صورت زیر در نظر می‌گیریم:

$$\text{Solution} = (w_1, \dots w_m, r_1, r_2 \dots r_{b-1}, q_1, \dots, q_m, p_1, \dots, p_m, v_1, \dots, v_m, \lambda)$$

در تمامی الگوریتم‌های فراابتکاری، بسیار مهم است که جمعیت اولیه به حد کفایت فضای جواب را پوشش دهد. برای تولید جمعیت اولیه، محدوده بردار وزن W را (1,n) در نظر می‌گیریم و برای وزن هر یک از معیارها یک مقدار با توزیع احتمال یکنواخت به صورت تصادفی در این محدوده، تولید و سپس برای اطمینان از اینکه جمع اوزان برابر با یک می‌شود آنها را نرمال می‌کنیم. برای پارامترهای p، q و v فضای جستجو را محدوده بین می‌نیمم و ماکزیمم مقادیر آلترناتیو در هر معیار در نظر گرفته و مقادیر تصادفی را با توزیع احتمال یکنواخت برای آنها تولید می‌کنیم. پروفایل ها را نیز در محدوده مینیمم و ماکزیمم مقادیر آلترناتیو در هر معیار تعیین کرده و سپس از کوچک به بزرگ مرتب می‌کنیم. برای تولید پارامتر λ نیز همین رویه را در بازه [0.5,1] با تولید اعداد تصادفی از توزیع یکنواخت اعمال می‌کنیم.

رویه کامل که تشریح‌کننده مراحل طبقه‌بندی با مدل پیشنهاد شده GA-ELECTRE است در شکل ۵ به صورت شبه کد ارائه می‌شود. پس از تولید جمعیت اولیه ، بهینه سازی در تکرارهای متوالی اجرا می‌شود. در هر تکرار، یک مقدار برازش(دقت طبقه‌بندی) برای هر کروموزوم مطابق رابطه محاسبه می‌گردد. پس از تکمیل مرحله بهینه سازی، پارامترهای بهینه و داده های تست برای انجام طبقه بندی تحویل ELECTRE TRI می‌شوند و رویه طبقه بندی برپایه روابط (1)- (5) انجام می‌گیرد.

قدم ۱: داده ها را بخوانید

قدم ۲: پارامترهای اولیه الگوریتم ژنتیک را تعریف کنید.

/ N:اندازه جمعیت،MG: ماکزیمم تکرارها،Nc: تعداد فرزندان حاصل از تقاطع،Nm: تعداد فرزندان حاصل از جهش/

قدم۳: داده ها را به ۵ فولد تقسیم کنید

حلقه بیرونی: تا رسیدن به مقدار Nc قدمهای زیر را انجام دهید **(For** i=1 to Nc **Do)**

فولد فعلی را در گروه تست و سایر فولدهارادر گروه آموزش قرار دهید

قدم٤: P_0 را تولید کنید / *جمعیت اولیه*/

t=0;

قدم ۵: حلقه الگوریتم ژنتیک

(While not Termination Criterion (P_t) **Do)**تا محقق نشدن معیار خاتمه قدم هـای زیـر را انجـام دهید

P_tرا ارزیابی کنید

قدم ۵-۱: والدین را انتخاب کنید

حلقه داخلی: تا رسیدن به مقدار Nc قدم‌های زیر را انجام دهید **(For** i=1 to Nc **Do)**

والدین را با استراتژی چرخه رولت انتخاب کنید

پایان حلقه داخلی

والدین را درG^1_t ذخیره کنید

قدم ۵-۲: تقاطع ۱ را انجام دهید

حلقه داخلی: تا رسیدن به مقدار Nc قدم‌های زیر را انجام دهید **(For** i=1 to Nc **Do)**

$G^1_t(j)$و $G^1_t(j+1)$ را با عملگر تقاطع میانی نوع اول ترکیب کنید

پایان حلقه داخلی فرزندان را در P^1_t ذخیره کنید

P^1_t را ارزیابی کنید

قدم ۵-۳: والدین را با استراتژی چرخه رولت انتخاب کنید و آنها را درG^2_t ذخیره کنید

قدم ۵-٤: تقاطع دوم را انجام دهید

حلقه داخلی: تا رسیدن به مقدار Nc قدم‌های زیر را انجام دهید **(For** i=1 to Nc **Do)**

$G^2_t(j)$و $G^2_t(j+1)$ را با عملگر تقاطع میانی نوع دوم ترکیب کنید

پایان حلقه داخلی فرزندان را در P^2_t ذخیره کنید

P^2_t را ارزیابی کنید

قدم ۵-۵: والدین را با استراتژی چرخه رولت انتخاب و آنها را در G^3_t ذخیره کنید

قدم ۵-۶: تقاطع سوم را انجام دهید

حلقه داخلی: تا رسیدن به مقدار Nc قدم‌های زیر را انجام دهید **(For** i=1 to Nc **Do)**

G^3_t (j) و G^3_t (j+1) را با عملگر تقاطع تک نقطه‌ای ترکیب کنید

پایان حلقه داخلی فرزندان را در P^3_t ذخیره کنید

P^3_t را ارزیابی کنید

قدم ۵-۷: P^1_t ,P^2_t,P^3_t را ادغام و آن را P'_t نامگذاری کنید; P'_t =Integrate(P^1_t ,P^2_t,P^3_t)

قدم ۵-۸: مقدار $\alpha\%$ از P'_t را با استراتژی چرخه رولت و تعداد Nm تا از P_t را به صورت تصادفی انتخاب، در M^1_t ذخیره کنید.

قدم ۵-۹: عملگر جهش یکنواخت نوع اول را روی M^1_t انجام و فرزندان حاصل را در P''_t ذخیره کنید

قدم ۵-۱۰: مقدار $\beta\%$ از P'_t را با استراتژی چرخه رولت و Nm تا از P_t را به صورت تصادفی، انتخاب، در M^2_t ذخیره کنید.

قدم ۵-۱۱: عملگر جهش یکنواخت نوع دوم را روی M^2_t انجام و فرزندان حاصل را در P'''_t ذخیره کنید.

قدم ۵-۱۲: جمعیت(P_t) و فرزندان(P'_t, P''_t, P'''_t) را ادغام کنید.

قدم ۵-۱۳: جمعیت بعدی را در سه سطح انتخاب کنید.

سطح اول: $\rho_1\%$ از جمعیت ادغامی را با استراتژی الیت انتخاب کنید.

سطح دوم: $\rho_2\%$ از جمعیت ادغامی را با استراتژی چرخه رولت انتخاب کنید.

سطح سوم: $1-\rho_1-\rho_2\%$ از جمعیت ادغامی را به صورت تصادفی انتخاب کنید.

P_{t+1} را جایگزین(P_t, P'_t, P''_t, P'''_t) کنید؛ P_{t+1}= Replace(P_t, P'_t, P''_t, P'''_t)

t = t + 1

پایان شرط

قدم ۶: بهترین جواب را ذخیره کنید.

پایان حلقه بیرونی

قدم ۷: طبقه بندی برای داده های دیده نشده(unseen data):

بهترین جواب و مجموعه تست را برای طبقه بندی به electer tri بدهید.

شکل(۵):. شبه کد الگوریتم GA-ELECTRE

داده‌های مورد استفاده در پژوهش

چهار دیتاست واقعی در حوزه طبقه بندی اعتباری در این پژوهش مورد استفاده قرار گرفتندکه عبارتند از: داده‌های اعتباری استرالیا،آلمان، ژاپن و ایران؛ سه دیتاست اول از انباره داده‌های یادگیری ماشینی دانشگاه کالیفرنیا اروایـن (http://archive.ics.uci.edu/ml) گرفته شده‌اند و دیتاست ایران برگرفته شده از یک دیتاست اصلاح شده مربوط به یک بانک خصوصی کوچک در ایران است[25]. جدول۳ خلاصه‌ای از ویژگی‌های اصلی این دیتاست‌ها را گزارش می‌کند.

جدول(۳):. مشخصات دیتاست‌های مورد استفاده

دیتاست	تعداد آلترناتیو	تعداد معیار	تعداد کلاس	تعداد خوش حساب	تعداد بد حساب
استرالیا	690	14	2	307	383
آلمان	1000	24	2	700	300
ژاپن	690	15	2	296	357
ایران	1000	27	2	950	50

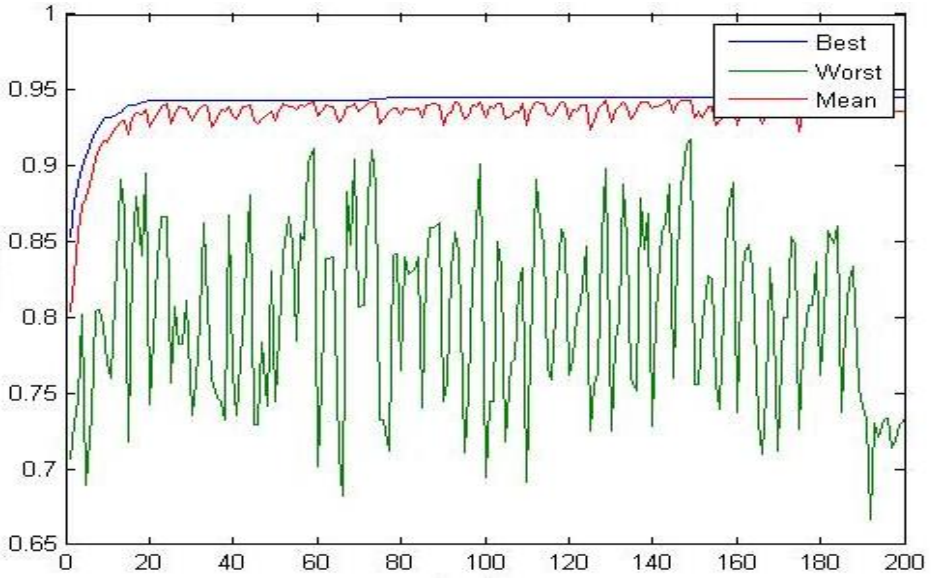
تجزیه و تحلیل داده‌ها

الگوریتم GA-ELECTRE روی داده‌های اعتباری استرالیا، آلمان، ژاپن و ایران اجرا شد. هر یک از این دیتاست‌ها به ۵ فولد طبقه‌بندی شده مجزا تقسیم و هر بار یک فولد برای تست و فولدهای باقیمانده برای آموزش استفاده شدند. این رویه تا جایی که هر یک از ۵ فولد، حداقل یک بار تست شوند تکرار شد. جدول ۴میانگین دقت طبقه بندی در هر اجرای مستقل برای داده‌های اعتباری ایران را نشان می‌دهد.

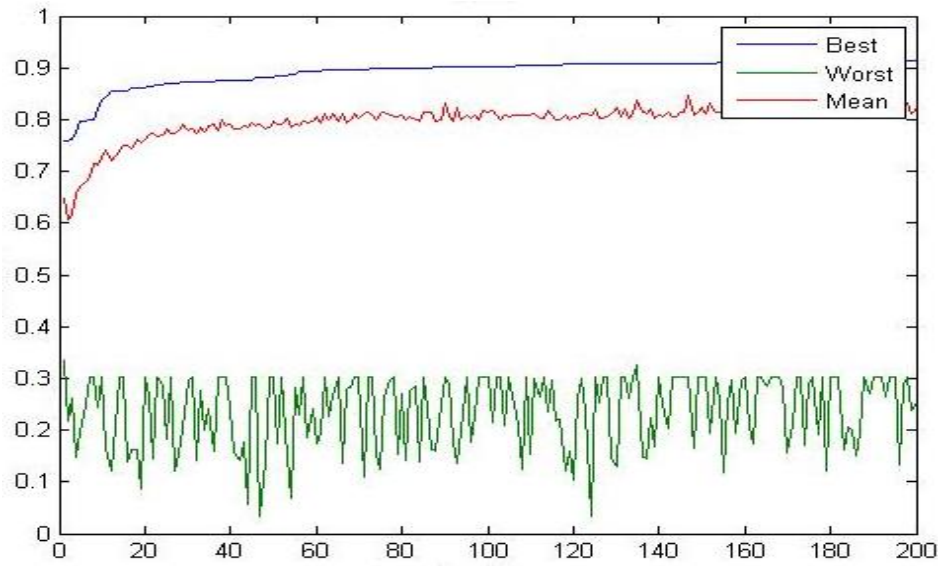
جدول(۴): نتایج پنج اجرای مستقل GA-ELECTRE روی دیتاست ایران

	Train	Test
Fold 1	91.95%	90.05%
Fold 2	92.96%	91.07%
Fold 3	92.47%	90.57%
Fold 4	93.17%	91.28%
Fold 5	94.78%	92.89%
میانگین	93.07%	91.17%

نمودار همگرایی بهترین، میانگین و بدترین جواب‌های مربوط یکی از اجراها روی داده‌های اعتباری ایران و ژاپن به ترتیب در شکل‌های ۶ و۷ نمایش داده شده است.



شکل(۶): نمودار همگرایی دیتاست ایران



شکل(۷): نمودار همگرایی دیتاست ژاپن

در مرحله بعدی، الگوریتم GA-ELECTRE با هفت الگوریتم شناخته شده طبقه‌بندی شامل: نزدیکترین همسایگی[۱]، طبقه‌بندی کننده بیز ساده[۲]، رگرسیون لجستیک، پرسپترون چند لایه-ای[۳]، شبکه عصبی تابع مبتنی بر شعاع[۴]، ماشین بردار پشتیبان با هسته خطی و درخت تصمیم[۵]C4.5، مقایسه و تفاوت آنها از نظر آماری آزمون شد. الگوریتم GA-ELECTRE در متلب[۶] و سایر الگوریتم‌ها در وکا[۷] اجرا شدند.

1 -K-Nearest Neighbourhood(K-NN)
2 -Naïve Bayes Classifier(NBC)
3 -Multi Layer Perceptron(MLP)
4 -Radial Basis Function(RBF)
5 -Decision Tree(DT)
6 -MATLAB
7 -WEKA

آزمون‌های آماری برای رویه تخصیص بدبینانه GA-ELECTRE

در جدول۵، نتایج حاصل از هر یک از الگوریتم‌ها برای دیتاست‌های مختلف ارائه شده است. میانگین جواب هر الگوریتم برای هر دیتاست نشان می‌دهد میانگین جواب حاصل از الگوریتم این پژوهش بهتر از سایر الگوریتم‌ها می‌باشد.

جدول(۵): نتایج آزمایشات بر حسب دقت طبقه بندی

دیتاست	نزدیکترین همسایگی(k=1)	بیز ساده	رگرسیون لجستیک	پرسپترون چند لایه	تابع-مبتنی‌بر شعاع	ماشین-بردار پشتیبان	درخت تصمیم	GA-Electre
استرالیا	81.45	80.72	84.93	83.04	79.28	85.07	85.51	86.42
آلمان	70.5	74.9	75.7	72.4	74.2	76	73.3	74.98
ژاپن	79.48	82.08	87.29	83.3	83.31	86.37	84.22	91.17
ایران	93	23.2	94.2	94.8	95	95	94.5	94.54
میانگین	81.11	65.23	85.53	83.39	82.95	85.61	84.38	86.77

آزمون فریدمن[1]

برای مقایسه عملکردهای چندین الگوریتم طبقه‌بندی بر چندین دیتاست از این آزمون استفاده می‌شود[10]. ابتدا رتبه هر الگوریتم طبقه‌بندی را برای هر دیتاست مشخص می‌کنیم، به الگوریتم طبقه‌بندی که بهترین عملکرد را داراست رتبه ۱، دومین الگوریتم طبقه‌بندی رتبه ۲ و به همین ترتیب رتبه بندی می‌کنیم. نتایج در جدول ۶ نمایش داده شده است.

1 -Friedman Test

جدول(۶): رتبه بندی و میانگین رتبه های الگوریتم‌های طبقه بندی

دیتاست	نزدیکترین همسایگی(k=1)	بیز ساده	رگرسیون لجستیک	پرسپترون چند لایه	تابع‌مبتنی- بر شعاع	ماشین- بردار پشتیبان	درخت تصمیم	GA-Electre
استرالیا	6	7	4	5	8	3	2	1
آلمان	8	4	2	7	5	1	6	3
ژاپن	8	7	2	6	5	3	4	1
ایران	7	8	6	3	1	1	5	4
میانگین	7.25	6.5	3.5	5.25	4.75	2	4.25	2.25

مقدار میانگین رتبه GA-ELECTRE برابر ۲ بوده که از سایر الگوریتم‌ها بهتر است. اکنون معناداری اختلاف خروجی الگوریتم‌ها را در سطح معنی داری ۰,۰۵، آزمون می‌کنیم. نتایج نشان می‌دهد که مقدار P-value آزمون برابر ۰,۰۱۵ می‌باشد، لذا فرض برابری خروجی الگوریتم ها رد شده و با یکدیگر متفاوت می‌باشند.

آزمون نمنی[1]

برای مقایسه زوجی عملکرد GA-ELECTRE در مقابل هر یک از الگوریتم های طبقه بندی از این آزمون استفاده می‌شود. بر اساس این آزمون، چنانچه میانگین رتبه‌های دو الگوریتم بزرگتر از اختلاف بحرانی[2] باشد، عملکرد دو الگوریتم با یکدیگر اختلاف معناداری خواهد داشت. اختلاف بحرانی برابر است با :

$$CD = q\sqrt{((K(K+1))/6D)}$$

که K و D به ترتیب تعداد الگوریتم‌ها و دیتاست‌ها را نمایش می‌دهد [10].در نهایت مقدار CD=2.94 بدست می‌آید. جدول ۷ اختلاف رتبه GA-ELECTRE با سایر الگوریتم های طبقه بندی را نمایش می‌دهد.

جدول(۷): اختلاف رتبه GA-ELECTRE با سایر الگوریتم‌ها

درخت تصمیم	ماشین-بردار پشتیبان	تابع‌مبتنی بر شعاع	پرسپترون چند لایه	رگرسیون لجستیک	بیز ساده	نزدیکترین همسایگی	
۲	-۰,۲۵	۲,۵	۳	۱,۲۵	٤,۲۵	٥	GA-ELECTRE

مقایسه اعداد جدول ۷ با مقدار CD نشان می‌دهد که الگوریتم GA-ELECTRE به صورت معناداری عملکرد بهتری نسبت به نزدیکترین همسایگی، طبقه‌بندی کننده بیز ساده، پرسپترون چند لایه‌ای دارد. نتایج پیاده سازی رویه تخصیص خوش بینانه GA-ELECTRE را در ضمیمه را ببینید.

1 -Nemenyi's Test

2 -Critical Difference(CD)

نتیجه‌گیری و پیشنهادات

۱. نتایج جدول ۷ نشان می‌دهد که عملکرد GA-ELECTRE نسبت به تعدادی از روشهای شناخته شده طبقه بندی بهتر است. در این روش برای هر معیار و هر کلاس باید ۵ پارامتر بدست آید که این واقعیت باعث انعطاف بیشتر این روش، کسب اطلاعات جزئی تر راجع به معیارها و افزایش شانس طبقه بندی درست آلترناتیوها می‌شود.

۲. هیچ پیش‌پرداز ش یا تبدیلی روی معیارها از قبیل نرمال سازی و... در روش پیشنهادی (GA-ELECTRE) لازم نیست. چرا که رویه طبقه بندی در روشهای مذکور بر مقایسه زوجی آلترناتیوها در هر معیار استوار است و بنابراین مسئله معیارهای با واحدهای اندازه گیری متفاوت عملاً حذف می‌شود.

۳. استفاده از الگوریتم ژنتیک برای بدست آوردن پارامترهای ELECTRE TRI طی فاز یادگیری نشان داد که این الگوریتم رویکرد موفقی برای یادگیری ELECTRE TRI بوده و عملکرد آن را بهبود می‌بخشد. این روش بر خلاف روشهای دقیق قابلیت استنتاج همزمان کلیه پارامترها را در مسائل واقعی با ابعاد بزرگ با دقت قابل قبول طی یک زمان منطقی برای هر دو نسخه بدبینانه و خوشبینانه دارد.

۴. نسخه بدبینانه روش GA-ELECTRE نیز نسبت به نسخه خوشبینانه آن نتایج بهتری تولید می کند. لازم به ذکر است به دلیل پیچیدگی بیشتر تخصیص خوشبینانه نسبت به بدبینانه و به طور کلی توسعه یافتگی بیشتر نسخه بدبینانه چنین نتیجه‌ای نیز انتظار می‌رفت.

۵. تمرکز تحقیقات آتی می‌تواند بر استفاده از روش پیشنهادی در سایر مسائل طبقه‌بندی مالی‌مانند پیش-بینی ورشکستگی، انتخاب پورتفوی سهام و همچنین تعمیم این روش به مسائل طبقه‌بندی چند کلاسه[1] باشد. بکارگیری سایر معیارهای ارزیابی عملکرد طبقه بندی، رویکرد چندهدفه به سایرروش‌های مبتنی بر روابط برتری مانند روش PROMETHEE، توسعه روش‌های پیشنهادی به مسائل بهینه سازی با بیش از چهارهدف(Many-objective optimisation)، استفاده از سایر الگوریتم‌های فراابتکاری برای انتخاب ویژگی[2] و بهینه‌سازی پارامترها و مقایسه با نتایج این تحقیق نیز پیشنهاد می‌شود.

1 -Multi Class

2-Feature selection

ضمیمه:

در جدول۱ نتایج حاصل از هر یک از الگوریتم‌ها برای دیتاست‌های مختلف در حالت تخصیص خوش‌بینانه ارائه شده است. در انتهای جدول نیز میانگین جواب هر الگوریتم برای کلیه دیتاست‌ها ارائه شده که نشان می‌دهد میانگین جواب حاصل از الگوریتم این پژوهش بهتر از سایر الگوریتم‌ها می‌باشد.

جدول(۱): نتایج آزمایشات بر حسب دقت طبقه بندی

دیتاست	نزدیکترین همسایگی(k=1)	بیز ساده	رگرسیون لجستیک	پرسپترون چند لایه	تابع-مبتنی‌بر شعاع	ماشین-بردار پشتیبان	درخت تصمیم	GA-Electre
استرالیا	81.45	80.72	84.93	83.04	79.28	85.07	85.51	86.13
آلمان	70.5	74.9	75.7	72.4	74.2	76	73.3	74.82
ژاپن	79.48	82.08	87.29	83.3	83.31	86.37	84.22	90.93
ایران	93	23.2	94.2	94.8	95	95	94.5	94.7
میانگین	**81.11**	**65.23**	**85.53**	**83.39**	**82.95**	**85.61**	**84.38**	**86.65**

آزمون فریدمن

ابتدا رتبه هر الگوریتم را برای هر دیتاست مشخص می‌کنیم. نتایج حاصل از رتبه‌بندی الگوریتم ها در جدول۲ نمایش داده شده است.

جدول۲. رتبه بندی و میانگین رتبه‌های الگوریتم های طبقه بندی

دیتاست	نزدیکترین همسایگی(k=1)	بیز ساده	رگرسیون لجستیک	پرسپترون چند لایه	تابع-مبتنی‌بر شعاع	ماشین-بردار پشتیبان	درخت تصمیم	GA-Electre
استرالیا	6	7	4	5	8	3	2	1
آلمان	8	4	2	7	5	1	6	4
ژاپن	8	7	2	6	5	3	4	1
ایران	7	8	6	3	1	1	5	4
میانگین	7.25	6.5	3.5	5.25	4.75	2	4.25	2.5

میانگین رتبه الگوریتم این پژوهش دارای مقدار ۱ بوده که نسبت به سایر الگوریتم ها بهتر است.اکنون بررسی می کنیم که آیا خروجی الگوریتم‌ها تفاوت خاصی با هم دارند یا خیر؟ نتایج حاصله نشان می‌دهد که مقدار P-value آزمون برابر ۰,۰۲۶ می‌باشد، بدین معنی که در سطح معنی‌داری ۰,۰۵، فرض برابری خروجی الگوریتم‌ها رد شده به عبارت دیگر خروجی الگوریتم ها با یکدیگر متفاوت می‌باشد.

آزمون نمنی

در نهایت براساس مقادیر حاصله، مقدار CD=2.94 شد. اختلاف رتبه الگوریتم این پژوهش، با سایر الگوریتم ها در جدول ۳ نمایش داده شده است.

جدول(۳): اختلاف رتبه GA-ELECTRE با سایر الگوریتم ها

	نزدیکترین همسایگی	بیز ساده	رگرسیون لجستیک	پرسپترون چند لایه	تابع‌مبتنی بر شعاع	ماشین-بردار پشتیبان	درخت تصمیم
GA-ELECTRE	**4/75**	**4**	**1**	**2**	**2.25**	**-0.5**	**1.75**

با مقایسه اعداد جدول فوق با مقدار CD می‌توانیم نتیجه بگیریم که الگوریتم GA-ELECTRE به صورت معناداری عملکرد بهتری نسبت به الگوریتم بیز ساده، و نزدیکترین همسایگی دارد.

منابع

Abdou, H., Pointon, J. and El-Marsy, A., (2008). *Neural nets versus conventional techniques in credit scoring in Egyptian banking*. Expert Systems with Applications, 35, 1277-1279.

Al-Obeidat, F., Belacel,N., Carretero, J.A., Mahanti,P., (2010). *Differential evolution for learning the classification method PROAFTN*, Knowledge-Based Systems 23 (5) 418‑426.

Al-Obeidat, F., Belacel,N., Carretero, J.A., Mahanti,P., (2011*).* ***An evolutionary framework using particle swarm optimization for classification method PROAFTN***, Applied Soft Computing 11, 4971‑4980

Altman, EI., Saunders, A., (1998*). Credit risk measurement: Developments over the last 20 years*. J Bank Financ 21:1721–1742

Baourakis, G., Conisescu, M., van Dijk, G., Pardalos, P.M., Zopounidis, C., (2009). *A multicriteria approach for rating the credit risk of financial institutions*. Comput. Manage. Sci. 6(3), 347–356

Belacel, N., Bhasker Raval, H., Punnenc, A., (2007). *Learning multicriteria fuzzy classification method PROAFTN from data***.** Computers and Operations Research 34, 1885–1898.

Bellotti, T., Crook, J. (2009). *Support vector machines for credit scoring and discovery of significant features*. Expert Systems with Applications, 36, 3302.

Bugera, V., Konno, H., Uryasev, S., (2002). *Credit cards scoring with quadratic utility function*. Journal of Multi-Criteria Decision Analysis 11, 197–211.

Chuang, C., Lin, R. (2009). *Constructing a reassigning credit scoring model*. Expert Systems with Applications, 36,1685-1687.

Demšar, J. (2006*). Statistical comparisons of classifiers over multiple data sets.* Journal of Machine Learning Research, 7, 1–30

Dias, L., Mousseau, V., (2006). **Inferring Electre's veto-related parameters from outranking examples**. European Journal of Operational Research 170 (1), 172–191.

Dias, L., Mousseau, V., Figueira, J., Climaco, J., (2002). **An aggregation /disaggregation approach to obtain robust conclusions with ELECTRE TRI**. European Journal of Operational Research 138 (2), 332–348.

Doumpos, M., Kosmidou, K., Baourakis,G. Zopounidis, C. (2002). **Credit risk assessment using a multicriteria hierarchical discrimination approach: A comparative analysis**. European Journal of Operational Reseach 138, 392–412.

Doumpos, M., Marinakis, Y., Marinaki, M., Zopounidis, C. (2009). **An evolutionary approach to construction of outranking models for multi criteria classification: The case of the ELECTRE TRI method**. European Journal of Operational Research, 199:496 – 505.

Doumpos, M., Pasiouras, F. (2005). **Developing and testing models for replicating credit ratings: a multicriteria approach**, Computational Economics 25, 327–341.

Doumpos, M., Zopounidis, C. (2002). **Multi-criteria classification methods in financial and banking decisions**, International Transactions in Operational Research, 9 ,567–581.

Goletsis, Y., Papaloukas, C., Fotiadis, D., Likas, A., Michalis, L., (2004). **Automated ischemic beat classification using genetic algorithms and multicriteria decision analysis**. IEEE Transactions on Biomedical Engineering 51 (10), 1717–1725.

Greco, S., Matarazzo, B., Slowinski, R., (2001a). **Rough sets theory for multi criteria decision analysis**. European Journal of Operational Research 129, 1–47.

Homaifar, A., Qi, Charlene X., Lai, Steven H., (1994). **Constrained Optimization Via Genetic Algorithms.** SIMULATION April 62: 242-253

Huang, C., Chen, M., Wang, C., (2007). **Credit scoring with a data**

mining approach based on support vector machines. Expert Systems with Applications, 33(4),847–856.

Jacquet-Lagrèze, E., Siskos, Y., (2001). **Preference disaggregation: Twenty years of MCDA experience**. European Journal of Operational Research 130, 233–245.

Jabeur, K., Guitouni, A., (2007). **Automated learning multi-criteria classifiers for FLIR ship imagery classification**, 10th International Conference on Information ,9 ,12, pp.1-8.

Khalil, J., Martel, J.M., Jutras, P., (2000), **A multicriterion system for credit risk rating**, Gestion: Belgian Management Magazine 15, 125–146.

Lee, T., Chen, I., (2005). **A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines.** Expert Systems with Applications, 28(4), 743‑752.

Marqués, A.I., García, V., Sánchez, J.S. (2012). **Exploring the behaviour of base classifiers in credit scoring ensembles**. Expert Systems with Applications, Vol. 39, No. 11, pp. 10244-10250.

Mousseau, V., Figueira, J., Naux, J.-P., (2001). **Using assignment examples to infer weights for ELECTRE TRI method: Some experimental results**. European Journal of Operational Research 130, 263–275.

Mousseau, V., Slowinski, R., (1998). **Inferring an ELECTRE-TRI model from assignment examples**. Journal of Global Optimization 12 (2), 157–174.

Mousseau, V., Slowinski, R., Zielniewicz, P., (2000). **A user-oriented implementation of the ELECTRE-TRI method integrating preference elicitation support**. Computers & Operations Research 27, 757–777.

Ngo The, A., Mousseau, V., (2002). **Using assignment examples to infer category limits for the ELECTRE TRI method**. Journal of Multi-Criteria Decision Analysis 11, 2943.

Peng, Y., Kou, G., Shi, Y., Chen, Z. (2008). **Multi-criteria convex quadratic programmingmodel for credit data analysis**, Decis. Support Syst. (44) 1016–1030.

Talbi, E. G. (2009). **Metaheuristics: from design to implementation** (Vol. 74). John Wiley & Sons.

Thomas, L. C. (2000). **A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers**. International Journal of Forecasting, 16(2), 149–172.

Zekic-Susac, M., Sarlija, N., Bensic, M., (2004). **Small Business Credit Scoring: A Comparison of Logistic Regression, Neural Networks, and Decision Tree Models.** 26th International Conference on Information Technology Interfaces. Croatia.

Zhang, DF., Leung, S. and Ye, ZM. (2008b). **A decision tree scoring model based on genetic algorithm and K-means algorithm**. In:Proceedings of the 3rd International Conference on Convergence and Hybrid Information Technology, Daejeon, Korea, pp 1043‑1047.

Zhu, X., Li, J., Wu, D., Wang, H. & Liang, C. (2013). **Balancing accuracy, complexity and interpretability in consumer credit decision making: A C-TOPSIS classification approach**.. *Knowl.-Based Syst.*, 52, 258-267.

Zopounidis, C. (2001). **Preference disaggregation in financial modeling: Basic features and some examples**. Operational Research, 1(3), 263-284.

Zopounidis,C. (2013). **Multicriteria decision aid in financial management**, European Journal of Operational Research 119, 404-415.

Zopounidis, C., Galariotis, E., Doumpos, M.,.Sarri, S., Andriosopoulos, K., (2015). **Multicriteria decision aiding for finance**,**An updated bibliographic survey,** European Journal of Operational Research, In Press, Accepted Manuscript.