Archiv

Industrial Management Journal

Print ISSN: 2008-5885
Online ISSN: 2423-5369

A Multi-objective Approach to the Problem of Subset Feature Selection Using Meta-heuristic Methods

Amir Daneshvar
*Corresponding Author, Assistant Prof. Department of Information Technology Management, Faculty of Management, Electronic Branch, Islamic Azad University, Tehran, Iran. E-mail: a_daneshvar@iauec.ac.ir

Mahdi Homayounfar
Assistant Prof., Department of Industrial Management, Faculty of Management and Accounting, Rasht Branch, Islamic Azad University, Rasht, Iran. E-mail: homayounfar@iaurasht.ac.ir

Bijan Nahavandi
Assistant Prof., Department of Industrial Management, Faculty of Management and Economics, Science and Research Branch, Islamic Azad University, Tehran, Iran. E-mail: bijan.nahavandi@srbiau.ac.ir

Fariba Salahi
Assistant Prof., Department of Industrial Management, Faculty of Management, Electronic Branch, Islamic Azad University, Tehran, Iran. E-mail: f_salahi@iauec.ac.ir

Abstract

Objective: Finding a subset of features is an issue that has been widely used in a variety of fields such as machine learning and statistical pattern recognition. Since increasing the number of features increases the computational cost of a system, it seems necessary to develop and implement systems with minimum features and acceptable efficiency.

Methods: Considering objective, it's developmental research and in terms of two Meta-heuristic algorithms, namely genetic algorithm (GA) and multi-objective non-dominated sorting genetic algorithm (NSGA II). The multi-objective method compared to the single-objective method has reduced the number of features to 50% in all instances; it doesn't make much difference in classification accuracy. The proposed method is applied on six datasets of credit data, and the results were analyzed using two common classifiers namely, support vector machine (SVM) and K-nearest neighbors (KNN). Comparing two classifiers applied on datasets, K- nearest neighbors (KNN) compared to the support vector machine (SVM) has shown relatively better performance in increasing the classification accuracy and reducing the number of attributes.

www.SID.ir

Archive of SID

Results: Genetic algorithm and multi objective non-dominated sorting genetic algorithm have a good performance in increasing the accuracy of classification and reducing the number of attributes in feature selection problem of multi-class data. The results also indicate an increase in classification accuracy, simultaneously with a significant decrease in the number of features in both KNN and SVM methods.

Conclusion: According to the results, the proposed approach has a high efficiency in features selection problem.

Keywords: Multi-objective programming, Feature subset selection, Meta-heuristic algorithm, Genetic algorithm, NSGA II algorithm

Citation: Daneshvar, Amir; Homayounfar, Mahdi; Nahavandi, Bijan and Salahi, Fariba (2021). A Multi-objective Approach to the Problem of Subset Feature Selection Using Meta-heuristic Methods. *Industrial Management Journal*, 13(2), 278- 299. (*in Persian*)

Industrial Management Journal, 2021, Vol. 13, No.2, pp. 278-299
https://doi.org/10.22059/IMJ.2021.315625.1007809
© Amir Daneshvar, Mahdi Homayounfar, Bijan Nahavandi, Fariba Salahi
Published by University of Tehran, Faculty of Management

Article Type: Research Paper
Received: December 19, 2020
Accepted: May 01, 2021

www.SID.ir

Archive of SID

مدیریت صنعتی
شاپا چاپی: ۲۰۰۸-۵۸۸۵
شاپا الکترونیکی: ۲۴۲۳-۵۳۶۹

رویکرد چندهدفه مبتنی بر روش‌های فرا ابتکاری برای مسئله انتخاب زیرمجموعه ویژگی‌ها

امیر دانشور
* نویسنده مسئول، استادیار، گروه مدیریت فناوری اطلاعات، دانشکده مدیریت، واحد الکترونیکی، دانشگاه آزاد اسلامی، تهران، ایران.
رایانامه: a_daneshvar@iauec.ac.ir

مهدی همایون‌فر
استادیار، گروه مدیریت صنعتی، دانشکده مدیریت و حسابداری، واحد رشت، دانشگاه آزاد اسلامی، رشت، ایران. رایانامه: homayounfar@iaurasht.ac.ir

بیژن نهاوندی
استادیار، گروه مدیریت صنعتی، دانشکده مدیریت و اقتصاد، واحد علوم و تحقیقات تهران، دانشگاه آزاد اسلامی، تهران، ایران. رایانامه: bijan.nahavandi@srbiau.ac.ir

فریبا صلاحی
استادیار، گروه مدیریت صنعتی، دانشکده مدیریت، واحد الکترونیکی، دانشگاه آزاد اسلامی، تهران، ایران. رایانامه: f_salahi@iauec.ac.ir

چکیده

هدف: پیدا کردن زیرمجموعه‌ای از مجموعه ویژگی‌ها، مسئله‌ای است که در زمینه‌های مختلفی مانند یادگیری ماشین و شناسایی آماری الگوها، کاربرد گسترده‌ای دارد. با توجه به اینکه افزایش تعداد ویژگی‌ها، هزینه محاسباتی سیستم را به‌طور تصاعدی افزایش می‌دهد، این پژوهش به‌دنبال طراحی و پیاده‌سازی سیستم‌هایی با کمترین تعداد ویژگی و کارایی قابل قبول است.

روش: با توجه به لزوم جست‌وجوی کارآمد در فضای جواب، در این پژوهش برای انتخاب ویژگی در داده‌های چندکلاسه، از الگوریتم ژنتیک (GA) و الگوریتم ژنتیک با مرتب‌سازی نامغلوب (NSGA II) چندهدفه با هدف افزایش دقت طبقه‌بندی و کاهش تعداد ویژگی‌ها استفاده شده است. روش ارائه شده، بر مبنای دو روش طبقه‌بندی ماشین بردار پشتیبان (SVM) و K نزدیک‌ترین همسایه (KNN) روی ۶ مجموعه داده اعتباری به اجرا درآمد و نتایج آن تجزیه و تحلیل شد.

یافته‌ها: الگوریتم ژنتیک و الگوریتم ژنتیک با مرتب‌سازی نامغلوب چندهدفه برای افزایش دقت طبقه‌بندی و کاهش تعداد ویژگی‌ها در مسئله انتخاب ویژگی در داده‌های چندکلاسه کارکرد مناسبی دارند. نتایج به‌دست‌آمده، نشان‌دهنده بهبود در دقت طبقه‌بندی، هم‌زمان با کاهش چشمگیر در تعداد ویژگی‌ها در هر دو روش ماشین بردار پشتیبان و نزدیک‌ترین همسایه است.

نتیجه‌گیری: با توجه به نتایج، رویکرد پیشنهادشده در این پژوهش، برای مسئله انتخاب ویژگی‌ها کارایی بسیار خوبی دارد.

کلیدواژه‌ها: برنامه‌ریزی چندهدفه، انتخاب زیرمجموعه ویژگی‌ها، الگوریتم‌های فرا ابتکاری، الگوریتم ژنتیک، الگوریتم NSGA II

استناد: دانشور، امیر؛ همایون‌فر، مهدی؛ نهاوندی، بیژن و صلاحی، فریبا (۱۴۰۰). رویکرد چندهدفه مبتنی بر روش‌های فرا ابتکاری برای مسئله انتخاب زیرمجموعه ویژگی‌ها. *مدیریت صنعتی*، ۱۳(۲)، ۲۷۸- ۲۹۹.

مدیریت صنعتی، ۱۴۰۰، دوره ۱۳، شماره ۲، صص. ۲۷۸- ۲۹۹
https://doi.org/10.22059/IMJ.2021.315625.1007809
دریافت: ۱۳۹۹/۰۹/۲۹، پذیرش: ۱۴۰۰/۰۲/۱۱
نوع مقاله: علمی پژوهشی
© امیر دانشور، مهدی همایون‌فر، بیژن نهاوندی و فریبا صلاحی
ناشر: دانشکدهٔ مدیریت دانشگاه تهران

www.SID.ir

Archive of SID

مقدمه

استفاده از فناوری‌های نوین در صنایع امروزی، بیش از آنکه مزیت به‌شمار رود، ضرورت محسوب می‌شود (رزمی، حیدریه و شهابی، ۱۳۹۳). این فناوری‌ها در نتیجه ضرورت انطباق محصولات و خدمات با خواسته‌های در حال تغییر مشتریان توسعه یافته‌اند (همایون فر، باقرسلیمی، نهاوندی و ایزدی شیجانی، ۱۳۹۷). رشد سریع صنعت اطلاعات و توسعه فناوری، موجب شکل‌گیری ویژگی‌های بسیاری برای مجموعه داده‌ها شده است که از پیامدهای آن ایجاد اطلاعات مرتبط، زائد و نامرتبط می‌باشد (هنسر، ژو، ژانگ، کارابوگا و آکای[1]، ۲۰۱۸). بسیاری از ویژگی‌های زائد و نامرتبط با مجموعه داده‌های اصلی در حل مسائل طبقه‌بندی غیراثربخش هستند و به دلیل فضای جست‌وجوی گسترده‌ای که ایجاد می‌کنند، به‌طور شایان توجهی موجب کاهش دقت طبقه‌بندی و افزایش پیچیدگی محاسباتی می‌شوند (گائو، هو، ژانگ و هی[2]، ۲۰۱۸). داده‌های نیازمند پردازش در تشخیص الگو و یادگیری ماشین، هم از نظر تعداد ویژگی‌ها و هم تعداد موارد به‌طور فزاینده‌ای رشد یافته‌اند (بولون ـ کان‌دو، سانچز مارنو و آلنسو ـ بتانزوس[3]، ۲۰۱۵). چنین داده‌هایی به‌طور قابل توجهی موجب افزایش زمان محاسبات و حافظه لازم برای تجزیه و تحلیل می‌گردند. علاوه بر این، معمولاً ویژگی‌های غیرمرتبط و زائدی در داده‌ها وجود دارند که علاوه بر افزایش بار محاسباتی، تأثیر منفی بر مدل‌های یادگیری دارند (تنگ، دای و ژیانگ[4]، ۲۰۱۹). انتخاب ویژگی یکی از مهم‌ترین و پرکاربردترین مسائل مطرح در پیش پردازش داده‌ها است که کاربرد گسترده‌ای در حذف ویژگی‌های غیرمرتبط، داده‌های تکراری و از بین بردن نویز در مجموعه داده‌ها دارد.

انتخاب ویژگی یکی از مسائل چالش‌برانگیز در مبحث یادگیری ماشینی و همچنین شناسایی آماری الگوها، به‌خصوص در مجموعه داده‌های بزرگ است که به‌طور گسترده و موفقیت‌آمیزی در سیستم‌های خبره و هوشمند، پردازش تصویر، بیوانفورماتیک، متن‌کاوی و طراحی کنترل‌کننده‌ها (چن، ژو و یوان[5]، ۲۰۱۹)، کنترل کیفیت آماری فرایندهای تولیدی (ژائو، کائو، ژانگ و لی[6]، ۲۰۲۱؛ لی و هی[7]، ۲۰۲۱)، تعمیرات و نگهداری پیشگیرانه (آرموا، کدیک، هیلند- وود و مک‌آری[8]، ۲۰۲۰)، مهندسی خطوط تولید نرم‌افزار (زیو و همکاران[9]، ۲۰۱۶) و پیش‌بینی ریزش مشتری در استفاده از خدمات مخابراتی (هوانگ، باکلی و کچادی[10]، ۲۰۱۰)، مدیریت سیستم‌های انرژی (مولر[11]، ۲۰۲۱)، مدیریت ارتباط با مشتری (تسنگ و هوانگ[12]، ۲۰۰۷) و غیره به‌کار گرفته شده است.

در بعضی از زمینه‌ها، هزاران ویژگی برای اندازه‌گیری وجود دارند که یا همه آنها به خروجی مسئله مربوط نیستند، یا دارای افزونگی می‌باشند. به علاوه، بررسی تعداد زیادی ویژگی غالباً هزینه‌بر است و دقت را کاهش می‌دهد. بنابراین،

1. Hancer, Xue, Zhang, Karaboga, and Akay
2. Gao, Hu, Zhang, and He
3. Bolon-Canedo, Sanchez-Marono, and Alonso-etanzos
4. Tang, Dai, and Xiang
5. Chen, Zhou, and Yuan
6. Zhao, Cao, Zhang, and Li
7. Li, and He
8. Aremua, Codyc, Hyland-Woodb, and McAree
9. Xue, Zhong, Tan, Liu, Cai, Chen, and Sun
10. Huang, Buckley, and Kechadi
11. Muller
12. Tseng, and Huang

www.SID.ir

Archive of SID

شناسایی ویژگی‌هایی که وابستگی زیادی با خروجی دارند از اهمیت بالایی برخوردار است. هدف از انتخاب ویژگی، دستیابی به کوچکترین زیرمجموعه از ویژگی‌های ورودی با بیشترین قدرت پیش‌بینی کنندگی است (زو، ژانگ و براون[1]، ۲۰۱۳). در انتخاب ویژگی، ویژگی‌هایی از مجموعه داده‌ها که برای پیش‌بینی خروجی مؤثرترند، انتخاب می‌شوند و مفهوم موجود در ویژگی‌ها را بعد از انتخاب ویژگی حفظ می‌کنند. در اصل، فرایند انتخاب ویژگی، انتخاب زیرمجموعه‌های بهینه ویژگی‌ها از مجموعه داده‌های اصلی است. بنابراین، رویکردهای بسیار کارآمد انتخاب ویژگی‌ها آنهایی هستند که می‌توانند به شیوه مؤثری داده‌های اصلی را به فضایی با ابعاد کم تقلبل دهند (چن و همکاران، ۲۰۱۹).

در مبانی نظری پژوهش، روش‌های انتخاب ویژگی به سه دسته اصلی بسته‌بندی، تعبیه‌سازی و فیلترسازی تقیسم می‌شوند (گائو و همکاران، ۲۰۱۸). بسته‌بندی زیرمجموعه‌ای از ویژگی را به‌کمک یک طبقه‌بند مشخص، مانند ماشین بردار پشتیبان انتخاب می‌کند. با وجود عملکرد مناسب در طبقه‌بندی، این دسته از مدل‌ها از ریسک بیش برازش برخوردارند. به‌طور مشابه، مدل‌های تعبیه‌سازی، بهترین زیرمجموعه ویژگی را در طی فرایند یادگیری یک الگوریتم یادگیری خاص، انتخاب می‌کنند. مدل‌های فیلترسازی هر ویژگی بر اساس مشخصه‌های احتمالی یا آماری خود و بدون استفاده از هیچ روش یادگیری یا طبقه‌بندی، رتبه‌بندی می‌شوند، بنابراین مستقل از یک طبقه‌بندی از پیش تعریف شده هستند (چن و همکاران، ۲۰۱۹). در این پژوهش، مسئله انتخاب ویژگی بر اساس رویکرد تعبیه‌سازی با استفاده از الگوریتم ژنتیک و الگوریتم ژنتیک با مرتب‌سازی نامغلوب (NSGA II) صورت می‌گیرد. به علاوه، برای کاهش حجم محاسباتی، کاهش تعداد ویژگی‌ها مسئله ضروری است، ولی تعیین این هدف نباید منجر به کاهش بازدهی یا افت عملکرد سیستم یادگیری ماشین شود.

اغلب پژوهش‌های صورت گرفته در زمینه انتخاب ویژگی، این مسئله را به‌صورت تک‌هدفه مورد بررسی قرار داده‌اند. با این وجود، مطالعات نشان داده است که مسئله انتخاب ویژگی یک مسئله چند هدفه با اهداف عمدتاً متضاد است که از آن میان می‌توان به حداکثرسازی عملکرد طبقه‌بندی و حداقل‌سازی تعداد ویژگی‌های انتخاب شده، اشاره نمود. برای ایجاد توازن میان این دو، انتخاب ویژگی را می‌توان به‌عنوان یک مسئله چندهدفه مورد بررسی قرار داد. پیشینه پژوهش در زمینه انتخاب ویژگی چندهدفه در مقایسه با پژوهش‌های انجام شده روی تکنیک‌های تک‌هدفه بسیار محدود است. با وجود این، پژوهش‌های انگشت‌شماری به ارائه چارچوب بهینه‌سازی چندهدفه پرداخته‌اند. یک رویکرد در انتخاب ویژگی چندهدفه، تبدیل آن به مسئله تک‌هدفه از طریق وزن‌دهی به اهداف و سپس تجمیع آنها است (بولون کان دو و همکاران، ۲۰۱۵). نقطه ضعف این روش آن است که وزن اهداف باید به‌صورت خارجی توسط تصمیم‌گیرنده تعیین شود که با توجه به عدم قطعیت و مقیاس‌های اندازه‌گیری متفاوت اهداف گوناگون، مسئله چالش برانگیزی است. در رویکرد دیگر، پژوهشگران بدنبال حداقل‌سازی تعداد ویژگی‌ها، با تعریف سطح مشخصی از عملکرد هستند. رویکرد دیگر به‌کار رفته در بررسی مسائل چندهدفه، در نظر گرفتن مسئله بهینه‌سازی به‌صورت تک‌هدفه با لحاظ کردن محدودیت بودجه پیشنهادی است (پنتزپی، بلانکوئرو، کاریزوسا و رامیرزکوبو[2]، ۲۰۱۸). استراتژی بهتر، در نظر گرفتن

1. Xue, Zhang, and Browne
2. Bentez-Pea, Blanquero, Carrizosa, & Ramırez-Cobo

www.SID.ir

Archive of SID

اهداف جداگانه و جست‌وجو برای مجموعه‌ای نامغلوب بر حسب چندین هدف، به جای تمرکز صرف روی یک هدف است. مجموعه جواب‌های نامغلوب که به‌عنوان مرز کارای پارتو شناخته می‌شوند و نمایانگر نقاطی هستند که در آنها، نمی‌توان یک هدف را جز با کاهش هدف دیگر بهبود بخشید. نوآوری این تحقیق در نظر گرفتن هم‌زمان دو هدف کاهش تعداد ویژگی و افزایش دقت عملکرد مدل با استفاده از الگوریتم فرابتکاری NSGAII جهت استخراج مرز کارای پارتو و پیاده‌سازی آن روی داده‌های اعتباری مشتریان بانکی و مقایسه عملکرد آن با الگوریتم چندهدفه NRGA است. سؤال اصلی تحقیق آن است که چگونه می‌توان یک مدل هوشمند چند هدفه به‌منظور انتخاب بهترین زیرمجموعه ویژگی‌ها جهت اعتبارسنجی مشتریان بانک ارائه داد؟

مرور ادبیات

پژوهش در رابطه با انتخاب ویژگی، برای داده‌کاوی و تشخیص الگو دارای اهمیت بسیار بوده و از سال‌های دهه ۱۹۷۰ آغاز شده است (چن و همکاران، ۲۰۱۹). در سال‌های اخیر روش‌های متنوعی برای انتخاب ویژگی ارائه شده‌اند که شامل روش‌های کاهش ابعاد و استخراج داده‌های مؤثر از ویژگی‌های اولیه می‌باشند. کاربرد الگوریتم‌های تکاملی چندهدفه واقعی به دهه ۱۹۸۰ باز می‌گردد. نخستین بار شافر[1] (۱۹۸۵) یک الگوریتم ژنتیک با سه‌عملگر ساده (انتخاب، آمیزش و جهش) را توسط چرخه‌های انتخاب مستقل مطابق با هر هدف، اصلاح کرد. الگوریتم پیشنهاد شده برای برخی نسل‌ها به خوبی عمل می‌کرد، اما در برخی موارد از مسئله جهت‌گیری اولیه به سمت یک تک عضو از جمعیت یا نواحی خاص (به‌خصوص اعضای قهرمان در یک هدف) مستثنی نبود. فونسکا و فلمینگ[2] (۱۹۹۳) یک مدل الگوریتم ژنتیک چندهدفه (MOGA) را پیشنهاد دادند که در آن اعضای نامغلوب جمعیت به‌صورت ۱ رتبه‌بندی می‌شدند. رتبه‌بندی دیگر اعضا با محاسبه‌ی تعداد جواب‌هایی که بر یک جواب مشخص غالب بودند، محاسبه می‌شد. مدل پیشنهادی آنها بر جواب‌هایی که شایستگی یکسان دارند، تکنیک نیچینگ (همسایگی) را اعمال می‌کرد تا پراکندگی جمعیت حفظ شود. در پژوهش آنها نیچینگ بر مقادیر پارامترها و نه بر مقادیر توابع هدف اعمال شده است. هورن، نفپلویتیز وگلدبرگ[3] (۱۹۹۴) از الگوریتم ژنتیک مبتنی بر نیچینگ پارتو برای تعیین مجموعه بهینه پارتو، استفاده نمودند. آنها برای نمایش قابلیت الگوریتم پیشنهادی در کشف و حفظ یک «جمعیت بهینه پارتو»، از یک مثال عددی و یک مسئله واقعی در سیستم‌های آبی استفاده کردند. دب، آکروال، پرتپ و میاریوان[4] (۲۰۰۰) الگوریتم ژنتیک با مرتب‌سازی نامغلوب (NSGA) را توسعه دادند که از نظر تعیین میزان شایستگی و روش اجرای نیچینگ با الگوریتم ژنتیک چند هدفه (MOGA) تفاوت دارد.

در سال‌های اخیر از روش‌های بهینه‌سازی چندهدفه تکاملی، شبکه عصبی، ID3 و درخت تصمیم در انتخاب ویژگی در سیستم‌های یادگیری ماشین استفاده‌های گوناگونی شده است. نصرتی ناهوک و افتخاری (۱۳۹۲) یک معیار انتخاب ویژگی اصلاح شده با استفاده از منطق فازی برای انتخاب تعداد ویژگی‌های مورد نیاز ارائه نمودند. عملکرد روش پیشنهادی روی مجموعه داده‌های منتشر شده از UCI ارزیابی شد و نتایج حاصل نشان‌دهنده کارایی روش به‌کار رفته در

1. Schaffer
2. Fonseca and Fleming
3. Horn, Nafpliotis, and Goldberg
4. Deb, Agrawal, Pratap, and Meyarivan

www.SID.ir

Archive of SID

مقایسه با نسخه غیرفازی آن است. خان و بایگ[1] (۲۰۱۵) از یک روش مبتنی بر الگوریتم تکاملی برای حل مسئله انتخاب زیرمجموعه ویژگی‌های چند هدفه استفاده کردند. در مطالعه آنها مقدار برازش یک زیرمجموعه ویژگی مشخص با استفاده از ID3 محاسبه گردیده است. هوانگ و همکاران (۲۰۱۰) به‌منظور پیش‌بینی وقفه در خدمات مخابراتی، یک رویکرد جدید چند هدفه مبتنی بر NSGA-II برای انتخاب ویژگی‌های ارائه کردند. آنها از روش FBSM برای تعیین آستانه برازش و استفاده از آن به‌منظور انتخاب بهترین راه‌حل‌ها استفاده کردند. کای، ونگ، تنگ، امریچ و وربک[2] (۲۰۱۶) از یک معیار فازی در انتخاب ویژگی چند هدفه نظارت نشده با استفاده از رویکرد ترکیبی فیلتر شده (FC-MOFS) استفاده نمودند. به علاوه، آنها از الگوریتم NSGA-II برای حل مسئله بهینه‌سازی استفاده کردند. زنگ، ونگ، شن و شی[3] (۲۰۱۷) از الگوریتم ژنتیک برای انتخاب ویژگی و بهینه‌سازی پارامترها در ماشین‌های تاکر پشتیبان استفاده کردند. الگوریتم پیشنهاد شده قادر بود اطلاعات نامناسب در داده‌های تنسور را از بین برده و دقت بهتری را به‌دست آورد.

داس و داس[4] (۲۰۱۷) انتخاب ویژگی را در قالب یک مسئله بهینه‌سازی دو هدفه که در آن هر ویژگی دارای وزن مشخصی است، مدلسازی کردند. بنابراین، زیرمجموعه‌ای از ویژگی‌های وزنی به‌عنوان بهترین زیرمجموعه برای طبقه‌بندی بعدی داده‌ها انتخاب شده است. در الگوریتم بهینه‌سازی تکاملی چند هدفه ارائه شده، مرتبط بودن و افزونگی در طراحی توابع هدف مورد استفاده قرار گرفتند. سهرابی و تاجیک[5] (۲۰۱۷) یک رویکرد انتخاب ویژگی برای تصمیم‌گیری در مورد دُز وارفارین ارائه دادند. در روش ارائه شده آنها ترکیبی از الگوریتم‌های بهینه‌سازی چند هدفه مانند: NSGA II، بهینه‌سازی ازدحام ذرات چندهدفه و شبکه عصبی مصنوعی استفاده شده است. بر اساس نتایج، عملکرد روش‌های چندهدفه نسبت به روش‌های کلاسیک مناسب‌تر است. هانسر، ژو، ژانگ، کارابوگا و آکای[6] (۲۰۱۸) یک رویکرد چندهدفه متشکل از دو الگوریتم زنبور عسل و NSGA II را برای انتخاب ویژگی ارائه نمودند. روش آنها در قالب دو رویکرد ABC با نمایش دوتایی و ABC با نمایش پیوسته ارائه شده است. نتایج به‌دست آمده نشان می‌دهد که رویکرد با نمایش دوتایی نسبت به سایر روش‌ها، هم از نظر کاهش ابعاد و هم دقت طبقه‌بندی، از عملکرد بهتری برخوردار است. تانگ، وای و ژیانگ[7] (۲۰۱۹) برای انتخاب ویژگی از تعاملات پنج سویه استفاده کردند. آنها برای جمع‌آوری اطلاعات پنج بعدی متقابل، از یک معیار کارآمد محاسباتی برای تعریف شرایط تعاملی استفاده نمودند. به علاوه، برای اجتناب از برآورد بیش از اندازه، از رویکرد غیر خطی «حداکثرِ حداقل‌ها» استفاده کردند. نتایج تجربی نشان‌دهنده کارایی مناسب روش ارائه شده در انتخاب ویژگی‌هاست.

ذاکری و حکم آبادی[8] (۲۰۱۹) یک روش جدید انتخاب ویژگی بر مبنای مدل ریاضی تعامل بین ملخ‌ها در یافتن منابع غذایی پیشنهاد نمودند. در الگوریتم ارائه شده، اصلاحاتی در الگوریتم بهینه‌سازی ملخ (GOA) صورت گرفت تا با

1. Khan and Baig
2. Cai, Wang, Tang, Emmerich, and Verbeek
3. Zeng, Wang, Shen and Shi
4. Das and Das
5. Sohrabi, and Tajik
6. Hancer, Xue, Zhang, Karaboga, and Akay
7. Tang, Dai, and Xiang
8. Zakeri, and Hokmabadi

www.SID.ir

Archive of SID

مسئله انتخاب ویژگی سازگار گردد. مقایسه نتایج الگوریتم ارائه شده و دوازده روش شناخته انتخاب ویژگی، نشان‌دهنده اهمیت روش پیشنهاد شده در مقایسه با سایر روش‌های انتخاب ویژگی است. چن، ژو و یوان[1] (۲۰۱۹) از یک رویکرد ترکیبی بهینه‌سازی ذرات با مکانیزم مارپیچی شکل (HPSO-SSM) برای انتخاب زیرمجموعه بهینه ویژگی‌ها در طبقه‌بندی بر مبنای رویکرد بسته‌بندی استفاده کردند. گائو و همکاران (۲۰۱۸) یک روش انتخاب ویژگی مبتنی بر تئوری اطلاعات ارائه نمودند. در روش آنها ترکیبی از مربوط بودن ویژگی مورد توجه است و از قابلیت حداکثرسازی اطلاعات جدید و حداقل‌سازی افزونگی اطلاعات به‌طور هم‌زمان برخوردار است. نعمت‌زاده، عنایتی فرد، محمود و اکبری[2] (۲۰۱۹) یک روش جدید انتخاب ویژگی بر اساس الگوریتم تکاملی ارائه نمودند. در روش آنها از روش تراکم متقابل به‌عنوان روش جدید نام‌گذاری طبقات و از الگوریتم وال برای افزایش دقت کلی تراکم متقابل استفاده شده است. کیورا[3] (۲۰۱۹) یک الگوریتم جست‌وجوی محلی موازی برای انتخاب ویژگی‌ها و طبقه‌بندی داده‌ها ارائه نمودند. علاوه بر این، روش پیشنهادی در مدت زمان مناسب، قادر است راه‌حل مسائلی را بیابد که تعداد ویژگی‌های بسیار زیادی دارند.

محمد و همکاران[4] (۲۰۲۰) یک رویکرد جدید بهینه‌سازی فرا ابتکاری به نام الگوریتم PPA را برای مسئله انتخاب ویژگی ارائه دادند که از تعامل بین شکارچی (گربه)، انگل (فاخته) و میزبان (کلاغ) در قالب سیستم گربه–فاخته–کلاغ تبعیت می کند. این مدل برای غلبه بر مشکلات همگرایی کم و ابعاد داده‌های بزرگ ارائه شده است. مقایسه نتایج الگوریتم PPA و چهار الگوریتم جست‌وجوی اکتشافی استاندارد نشان داد که الگوریتم پیشنهادی از نظر طبقه بندی و کاهش ابعاد، هم مؤثر و هم رقابتی است. چندهاری و ساهو[5] (۲۰۲۰) یک روش ترکیبی از فیلتر– بسته‌بندی برای انتخاب ویژگی ارائه دادند. در این مطالعه تاپسیس به‌عنوان فیلتر استخراج مناسبترین ویژگی‌ها و الگوریتم باینری جاوا با تابع انتقال متغیر در طول زمان برای یافتن زیرمجموعه بهینه ویژگی‌ها استفاده شده است. نتایج تجربی در ۱۰ مجموعه داده نشان می‌دهد که روش پیشنهادی از نظر دقت طبقه‌بندی، عملکرد بهتری دارد و ۱۰ برابر سریعتر از روش‌های پیشرفته موجود است. آگراوال، کائورا و شارما[6] (۲۰۲۰) یک الگوریتم بهینه‌سازی نهنگ کوانتومی (QWOA) را برای انتخاب ویژگی ارائه دادند که تلفیقی از مفاهیم کوانتومی و الگوریتم بهینه‌سازی نهنگ است. روش پیشنهادی با استفاده از نمایش بیت کوانتوم از افراد جمعیت و اپراتور دروازه چرخش کوانتومی، به‌عنوان یک اپراتور تغییر، قدرت اکتشاف و بهره برداری الگوریتم کلاسیک بهینه‌سازی نهنگ را افزایش می دهد. کارآیی روش پیشنهادی با الگوریتم معمولی و با الگوریتم‌های تکاملی ازدحام ذرات و کوانتوم بر اساس ۱۴ مجموعه داده از حوزه‌های متنوع مقایسه گردید که نتایج تجربی، عملکرد برتر روش پیشنهادی QWOA را نشان می‌دهد.

عبدالعزیز، اویس، ابراهیم و لو[7] (۲۰۲۰) یک روش جایگزین برای ایجاد یک زیرمجموعه بهینه از ویژگی‌هایی ارائه دادند که نمایانگر کل ویژگی‌ها از طریق بهبود کارایی الگوریتم بهینه‌سازی شعله پروانه (MFO) در جست‌وجوی

1. Chen, Zhou and Yua
2. Nematzadeh, Enayatifar, Mahmud, and Akbari
3. Cura
4. Mohamed, Hassan, Hemeida, Alkhalaf, Mahmoud, and Baha-Eldin
5. Chaudhuri, and Sahu
6. Agrawal, Kaura, and Sharma
7. Abd-Elazizb, Ewees, Ibrahim, and Lu

www.SID.ir

Archive of SID

زیرمجموعه‌های بهینه است. روش پیشنهادی ذکر شده به‌عنوان OMFODE توانایی جلوگیری از توقف در یک مقدار بهینه محلی را دارد. نتایج تجربی نشان داد که الگوریتم پیشنهادی از نظر اندازه‌گیری عملکرد نسبت به الگوریتم‌های فراابتکاری پیشرفته، عملکرد بهتری دارد. گارسیا-پدراجاس، دِ کاستیلو و سروئلا-گارسیا[1] (۲۰۲۰) یک الگوریتم جدید مبتنی بر جست‌وجوی ابتکاری و چندین مکانیزم مقیاس‌بندی را برای بررسی هم‌زمان نمونه‌ها و ویژگی‌های تعداد بالا ارائه دادند که می‌تواند با موفقیت در مجموعه داده‌هایی با میلیون‌ها ویژگی و نمونه اعمال شود. روش ارائه شده ضمن دستیابی به کاهش قابل توجه داده، قادر است نتایج عملکرد طبقه‌بندی بهتری را نسبت به رویکردهای پیشرفته فراهم آورد. الاعظم، شریعه و صبری[2] (۲۰۲۰) یک الگوریتم انتخاب ویژگی بسته‌بندی را برای سیستم تشخیص نفوذ ارائه دادند. این الگوریتم از بهینه‌ساز الهام گرفته از کبوتر برای استفاده از فرایند انتخاب استفاده می‌کند. الگوریتم پیشنهادی بر اساس سه مجموعه داده شناخته شده با چندین الگوریتم انتخاب ویژگی پیشرفته مقایسه شده که نشان‌دهنده همگرایی سریع‌تر این روش نسبت به سایر روش‌ها است.

لی، ژانگ، یون و وون[3] (۲۰۲۰) با توسعه الگوریتم ماشین بردار پشتیبان، هزینه‌های ویژگی را (همراه با حفظ دقت طبقه بندی) با استفاده از محدودیت بودجه کاهش دادند. مطالعه تجربی مدل روی انواع داده‌های معیار و مصنوعی نشان داد که مدل برنامه‌ریزی خطی عدد صحیح (MILP) ارائه شده می‌تواند از نظر عملکرد پیش‌بینی و اقتصادی، رقابتی عمل کنند. رودریگوئز، دِ آلبوکیورکو و پاپا[4] (۲۰۲۰) در مطالعه خود به ارائه نسخه‌های تک هدفه، چند هدفه و پرشمار هدفه از الگوریتم بهینه‌سازی پروانه مصنوعی (ABO) در زمینه انتخاب ویژگی پرداختند. مقایسه نتایج حاصل از پیاده‌سازی الگوریتم پیشنهادی با الگوریتم‌های ازدحام ذرات، کرم شب‌تاب، گرده افشانی گل، سیاه چاله و بهینه‌سازی طوفان مغزی بر اساس ۸ مجموعه داده نشان داد که عملکرد دودویی تک هدفه ABO با توجه به انتخاب ویژگی کمتر و همچنین بار محاسباتی کمتر، بهتر از سایر تکنیک‌های فراابتکاری است. امینی و هو[5] (۲۰۲۰) یک رویکرد جدید دو لایه برای مسئله انتخاب ویژگی ارائه نموده‌اند که ترکیبی از یک روکش و یک روش تعبیه شده در ساخت زیرمجموعه مناسب از پیش‌بینی‌ها است. در اولین لایه روش پیشنهادی، الگوریتم ژنتیک (GA) به‌عنوان یک بسته‌بندی برای جست‌وجوی زیرمجموعه بهینه پیش‌بینی‌کننده‌ها استفاده شده است که هدف آن کاهش تعداد پیش‌بینی‌ها و خطای پیش‌بینی است. شبکه الاستیک (EN) به‌دلیل انعطاف‌پذیری در تنظیم شرایط مجازات در فرایند تنظیم و بهره‌وری از زمان، به‌عنوان روش تعبیه شده در لایه دوم انتخاب شده است. نتایج عددی برتری مدل پیشنهادی را تأیید می‌کند.

غلامی، پورپناه و وانگ[6] (۲۰۲۰) یک الگوریتم جست‌وجوی‌هارمونی جهانی دودویی بهبود یافته (IBGHS) را برای مسئله انتخاب ویژگی مورد استفاده قرار دادند. یک مرحله بداهه‌سازی اصلاح شده برای افزایش توانایی جست‌وجوی جهانی و افزایش سرعت همگرایی الگوریتم معرفی شده است. علاوه بر این، از الگوریتم K نزدیک‌ترین همسایه به‌عنوان

1. García-Pedrajas, de Castillo, and Cerruela-García
2. Alazzam, Sharieh, and Sabri
3. Lee, Zhang, Yoon, and Won
4. Rodrigues, de Albuquerque, and Papa
5. Amini, and Hu
6. Gholami, Pourpanah, and Wang

www.SID.ir

Archive of SID

یک مدل یادگیری اساسی برای ارزیابی اثربخشی زیرمجموعه ویژگی انتخاب شده استفاده می‌شود. بر اساس ۱۸ مسئله الگو، الگوریتم پیشنهادی IBGHS قادر است در مقایسه با سایر روش‌های پیشرفته مانند: الگوریتم ژنتیک، بهینه‌سازی ازدحام ذرات، شیر- مورچه، جست‌وجوی‌هارمونی جهانی و الگوریتم بهینه‌سازی نهنگ در حل مسئله انتخاب ویژگی نتایج قابل مقایسه‌ای ایجاد کند. وی، ژائو، فنگ، هه و یو[1] (۲۰۲۰) یک الگوریتم انتخاب ویژگی جدیدی با نام الگوریتم انتخاب ویژگی بر مبنای اهمیت پویای ویژگی (DFIFS) و شاخص جدیدی متناظر با آن (DFI) ارائه نمودند. به منظور دستیابی به دقت طبقه‌بندی بالاتر با تعداد کمتر از ویژگی‌ها، یک الگوریتم اصلاح‌شده DFIFS با نام M-DFIFS با ترکیب DFIFS با فیلترهای کلاسیک ایجاد شده است. براساس آزمایش‌های انجام شده با ۱۴ مجموعه داده عمومی با ابعاد بالا، الگوریتم M-DFIFS از لحاظ میانگین دقت آنها با زمان محاسبه قابل قبول، عملکرد بهتری نسبت به پنج الگوریتم فیلتر متداول نشان می‌دهد.

افروسیندیس و آرامپازیس[2] (۲۰۲۱) یک مطالعه تجربی جامع از ۱۲ روش فردی و ۶ روش گروهی برای انتخاب ویژگی‌ها در طبقه‌بندی داده‌های زیست‌محیطی ارائه دادند. آنها به بررسی بهترین عملکرد در روش‌های فردی و گروهی و اندازه‌گیری پایداری آنها پرداختند. ژنگ، چن، نی و هوانگ[3] (۲۰۲۱) یک روش تحلیل تشخیصی انطباقی جدید را برای انتخاب ویژگی نیمه نظارت شده (SADA) ارائه دادند که بجای محاسبه ماتریس شباهت، یک ماتریس تشابه تطبیقی و ماتریس فرافکنی را با یک فرایند تکراری فرا می‌گیرد. نتایج تجربی برای مجموعه داده‌های مصنوعی، برتری SADA را در مقایسه با ۶ روش انتخاب ویژگی نیمه نظارت شده نشان می‌دهد. تیرومورثی و مونس واران[4] (۲۰۲۱) یک روش انتخاب ویژگی ترکیبی جدید را بر اساس الگوریتم بهینه‌سازی باینری ضعیف و قوی (HBPRO) ارائه نمودند. روش پیشنهادی را با پنج الگوریتم شناخته شده انتخاب ویژگی مقایسه گردید که نتایج تجربی بیانگر دقت بیشتری و تولید تعداد کمتر ویژگی‌ها توسط الگوریتم ارائه شده بود. ژانگ، فان و یانگ[5] (۲۰۲۱) یک الگوریتم انتخاب ویژگی اکتشافی مبتنی بر آنتروپی تصمیم‌گیری نسبی همسایه فازی با عنوان AFNRDE ارائه دادند. بر اساس نتایج، AFNRDE الگوریتم اولیه FSMRDE را بر اساس آنتروپی تصمیم‌گیری نسبی در پردازش داده‌های عددی ارتقا می‌دهد و همچنین از دو روش معمول FNRS وFNGRS برای طبقه‌بندی پیشی می‌گیرد.

ابراهیم، عبدالعزیز، اویس العبد و لیو[6] (۲۰۲۱) در مطالعه خود به توسعه یک روش انتخاب ویژگی فوق ابتکاری پرداختند. در گام اول الگوریتم ارزیابی تکاملی برای کنترل بهترین ترکیب عناصر گام بعد به‌کار رفته است. در گام دوم ترکیب به‌دست آمده با استفاده از یک مجموعه آزمایشی ارزیابی می شود. ارزیابی تجربی روش پیشنهادی بر اساس ۱۸ مجموعه داده نشان دهنده عملکرد بهتر روش ارائه شده در مقایسه با ۸ الگوریتم فراابتکاری معروف است. مولر (۲۰۲۱) یک رویکرد انتخاب ویژگی جدید مبتنی بر خوشه بندی، مدل‌سازی لانه پرندگان و رگرسیون (CNR) را برای مواردی

1. Wei, Zhao, Feng, He, and Yu
2. Effrosynidis, and Arampatzis
3. Zhong, Chen, Ni, and Huang
4. Thirumoorthy, and Muneeswaran
5. Zhang, Fan, and Yang
6. Ibrahim, Abd Elaziz, Ewees, El-Abd, and Lu

www.SID.ir

Archive of SID

که نیاز به انتخاب‌پذیری بالایی دارند و از مجموعه داده‌های مختلف استفاده می‌کنند، ارائه نمودند. نتایج نشان می‌دهد که CNRوقتی مستقیماً با تک تک الگوریتم‌های تشکیل دهنده خود مقایسه شود، به طور متوسط تا ۱۰۱درصد میانگین قدر مطلق خطا (MAE) کمتری دارد. ژو، کانگ، ونگ و ژانگ[1] (۲۰۲۱) با در نظر گرفتن تعارضات بین اهداف مختلف، یک چارچوب بهینه‌سازی چند هدفه تکاملی را برای مسائل انتخاب ویژگی مبتنی بر داده‌های گسسته ارائه نمودند. چارچوب پیشنهادی با استفاده از سه الگوریتم تکاملی چند هدفه اجرا و با برخی از روش‌های پیشرفته مقایسه گردید. بر اساس نتایج، روش پیشنهادی از نظر دقت طبقه‌بندی، به طور قابل توجهی از سایر روش‌ها بهتر است. بایا، هان، رن و کین[2] (۲۰۲۱) برای تشخیص عیب در ماشین‌های چرخان، یک الگوریتم ترکیبی انتخاب ویژگی مبتنی بر بهینه‌سازی وال و تحلیل مولفه‌ای آنتروپی کرنل با نام WOSKECA ارائه نمودند. نتایج نشان می‌دهد که روش پیشنهادی قادر است با موفقیت و دقت بیشتری نسبت به یادگیری منیفولد سنتی، عیب‌های ماشین‌های چرخان را شناسایی و طبقه‌بندی کند. برخی از مطالعات اخیر در رابطه با انتخاب ویژگی در جدول ۱ ذکر شده است.

جدول ۱. مطالعات انجام شده در زمینه انتخاب ویژگی

روش به‌کار رفته	نویسندگان
NSGA II، درخت تصمیم	هوانگ، بوکلی وکچادی (۲۰۱۰)
الگوریتم ژنتیک، منطق فازی	نصرتی ناهوک و افتخاری (۱۳۹۲)
NSGA – II , ID3	خان و بایگ (۲۰۱۵)
ترکیب الگوریتم ژنتیک دوهدفه و معیار Gain	داس، میشرا و شاو (۲۰۱۶)
الگوریتم ازدحام ذرات	گیو، چنگ و جین (۲۰۱۶)
NSGA II و منطق فازی	کای، ونگ، تنگ، امریچ و وربک (۲۰۱۶)
الگوریتم ژنتیک	زنگ، ونگ، شن و شی (۲۰۱۷)
NSGA II، ماشین بردار پشتیبان، K نزدیک‌ترین همسایه	داس و داس (۲۰۱۷)
بهینه‌سازی ازدحام ذرات چندهدفه، NSGA II، شبکه عصبی مصنوعی	سهرابی و تاجیک (۲۰۱۷)
الگوریتم زنبور عسل	هانسر، ژو، ژانگ، کارابوکا و آکای (۲۰۱۸)
ماشین بردار پشتیبان، بیز ساده، تئوری اطلاعات	گائو، هو، ژانگ وهی (۲۰۱۸)
بهینه‌سازی ذرات	چن و همکاران (۲۰۱۹)
تعاملات پنج سویه، برنامه‌ریزی غیرخطی	تانگ، وای، ژیانگ (۲۰۱۹)
الگوریتم تکاملی، الگوریتم وال	نعمت‌زاده، عنایتی فرد، محمود و اکبری (۲۰۱۹)
الگوریتم ارزیابی تکاملی	کیورا (۲۰۱۹)
الگوریتم جست‌وجوی محلی موازی	محمد و همکاران (۲۰۲۰)
الگوریتم PPA (الگوریتم ازدحام گربه، الگوریتم جست‌وجوی فاخته و الگوریتم جست‌وجوی کلاغ)	چتدهاری و پراسادساهو (۲۰۲۰)

1. Zhou, Kang, Kwong, Wang, and Zhang
2. Baia, Han, Ren, and Qin

www.SID.ir

Archive of SID

ادامهٔ جدول ۱

روش به‌کار رفته	نویسندگان
TOPSIS و الگوریتم باینری جاوا	آگراوال، کائورا و شارما (۲۰۲۰)
الگوریتم بهینه‌سازی نهنگ کوانتومی	عبدالعزیز، اویس، ابراهیم و لو (۲۰۲۰)
الگوریتم بهینه‌سازی شعله- پروانه	گارسیا پدراجاس، دِکاستیلو و سروئلا گارسیا (۲۰۲۰)
الگوریتم جدید مبتنی بر یک جست‌وجوی ابتکاری	الاعظم، شریعه و صبری (۲۰۲۰)
الگوریتم از بهینه ساز الهام گرفته از کبوتر	لی، ژانگ، یون و وون (۲۰۲۰)
الگوریتم ماشین بردار پشتیبان	رودریگوئز، دِ آلبوکیورکو و پاپا (۲۰۲۰)
الگوریتم بهینه‌سازی پروانه مصنوعی	امینی و هو (۲۰۲۰)
الگوریتم ژنتیک و شبکه الاستیک	غلامی، پورپناه و وانگ (۲۰۲۰)
الگوریتم جست‌وجوی هارمونی جهانی دودویی بهبود یافته	وی، ژائو، فنگ، هه و یو (۲۰۲۰)
الگوریتم انتخاب ویژگی بر مبنای اهمیت پویای ویژگی	هاشمی و باقر (۲۰۲۱)
۱۲ الگوریتم کلاسیک موجود	ژنگ، چن، نی و هوانگ (۲۰۲۱)
الگوریتم بهینه‌سازی باینری ضعیف و قوی	تیرومورثی و مونس واران (۲۰۲۱)
روش تحلیل تشخیصی انطباقی نیمه نظارت شده	ژانگ، فان و یانگ (۲۰۲۱)
الگوریتم ارزیابی تکاملی	ابراهیم، عبدالعزیز، اویس العبد و لیو (۲۰۲۱)
الگوریتم ترکیبی (خوشه‌بندی، لانه پرندگان و رگرسیون)	مولر (۲۰۲۱)
آنتروپی تصمیم‌گیری نسبی همسایه- فازی	ژو، کانگ، ونگ و ژانگ (۲۰۲۱)
الگوریتم ترکیبی بهینه‌سازی وال و تحلیل مولفه‌ای آنتروپی کرنل	بایا،هان، رن و کین (۲۰۲۱)

همان گونه که بیان شد، نوآوری تحقیق حاضر در نظر گرفتن هم‌زمان دو هدف کاهش تعداد ویژگی و افزایش دقت عملکرد مدل با استفاده از الگوریتم فراابتکاری NSGAII جهت استخراج مرز کارای پارتو و پیاده‌سازی آن روی داده‌های اعتباری مشتریان بانک است. ضمن اینکه مقایسه عملکرد الگوریتم ارائه شده با الگوریتم چندهدفه NRGA صورت گرفته است.

روش‌شناسی پژوهش

در مسائل انتخاب ویژگی‌ها، مزیت روش‌های فرا ابتکاری نسبت به مدل‌های برنامه‌ریزی ریاضی مبتنی بر روابط برتری، در آن است که تخمین هم‌زمان پارامترها، به مدل‌های پیچیده غیرخطی و غیرمحدب نیاز دارد و می‌توان از فرضیه‌هایی که به سادگی و کاهش دقت مدل منجر می‌شوند، صرف نظر کرد. در پژوهش حاضر، الگوریتم ژنتیک، الگوریتم ژنتیک با مرتب‌سازی نامغلوب، ماشین بردار پشتیبان و K نزدیک‌ترین همسایه برای مسائل انتخاب ویژگی‌ها پیشنهاد شده است. بر این اساس، پژوهش حاضر از نظر هدف، توسعه‌ای و از نظر روش تحلیل داده‌ها کمی و از نوع مدل‌سازی ریاضی است. به‌منظور پیاده‌سازی روش پیشنهادی از داده‌های آرشیوی موجود در پایگاه‌های داده معتبر و برنامه‌نویسی روش ارائه شده در نرم‌افزار متلب استفاده خواهد شد. شبه کد الگوریتم به‌کار رفته در این پژوهش در شکل ۱ نشان داده شده است:

www.SID.ir

Archive of SID

گام ۱: پارامترهای اولیه الگوریتم را تعریف کنید.
گام ۲: جمعیت اولیه (P) را با سایز N ایجاد کنید.
گام ۳: داده‌ها را طبقه‌بندی کنید.
گام ٤: توابع هدف (تعداد ویژگی‌ها و دقت) را ارزیابی کنید.
گام ٥: بر اساس جواب‌های نامغلوب، رتبه (جبهه) را تخصیص دهید.
گام ٦: فاصله ازدحامی را برای جمعیت محاسبه کنید.
گام ۷: تا برآورده نشدن معیار پایان، گام‌های زیر را دنبال کنید:
گام ۸: تقاطع را انجام دهید و فرزند (زاد و ولد) Q_1 را ایجاد کنید.
گام ۹: جهش را انجام دهید و فرزند (زاد و ولد) Q_2 را ایجاد کنید.
گام ۱۰: طبقه بندی را انجام دهید و توابع هدف را برای Q_1 و Q_2 ارزیابی کنید.
گام ۱۱: یک جمعیت واحد تشکیل دهید: $P \cup Q_1 \cup Q_2 = \hat{P}$
گام ۱۲: تا زمانیکه جمعیت کوچکتر از N است، گامهای زیر را دنبال کنید:
گام ۱۳: جواب‌های حاصل از جبهه کنونی را بر اساس فاصله ازدحامی‌شان مرتب کنید.
گام ١٤: برای هر جواب موجود در جبهه مرتب شده، موارد زیر را انجام دهید:
گام ۱۵: اگر جمعیت کوچکتر از N باشد، آنگاه:
گام ١٦: جواب را در جمعیت قرار بدهید.
گام ۱۷. پایان شرط **اگر**.
گام ۱۸: پایان شرط **برای**.
گام ۱۹: به جبهه پارتو بعدی بروید.
گام ۲۰: پایان شرط **تا**.
گام ۲۱: پایان شرط **تا**.
گام ۲۲: به P بازگردید.

شکل ۱. شبه کد الگوریتم پیشنهادی

با توجه به اینکه تفاوت الگوریتم NRGA با NSGAII در استراتژی انتخاب والدین است (الگوریتم NRGA از استراتژی باینری تورتمنت استفاده می‌کند)، شبه کد آن نیز جز در این قسمت، مشابه شبه کد شکل ۱ است.

در راستای پیاده‌سازی الگوریتم پیشنهادی از پایگاه داده‌های اعتباری ۶ کشور استرالیا، آلمان، ژاپن، ایران، لهستان و آمریکا استفاده شده است که مشخصات آنها در جدول ۲ آورده شده است.

جدول ۲. پایگاه‌های داده مورد استفاده در پژوهش

پایگاه داده	تعداد ویژگی	تعداد کلاس	تعداد نمونه
استرالیا	۱۴	۲	۶۹۰
آلمان	۲۴	۲	۱۰۰۰
ژاپن	۱۵	۲	۶۵۳
ایران	۲۷	۲	۱۰۰۰
لهستان	۳۰	۲	۲۴۰
آمریکا	۳۸	۲	۲۴۳۵

www.SID.ir

Archive of SID

در نهایت، برای ارزیابی سیستم طبقه‌بندی‌کننده از پارامتر صحت (AC) استفاده می‌شود که مهم‌ترین معیار برای ارزیابی میزان نزدیکی نتایج به‌دست آمده به نتایج حقیقی است. هرچه جواب آزمایش به مقدار حقیقی نزدیک‌تر باشد صحت آن بیشتر است. برای محاسبه صحت از رابطه ۱ استفاده می‌شود:

$$AC = \frac{TP + TN}{TN + FP + FN + TP} \times 100$$ رابطهٔ ۱)

در این رابطه TP نشان‌دهنده موارد مثبت حقیقی (طبقه‌بندی مشتری خوش‌حساب را تحت عنوان مشتری خوش‌حساب)، TN موارد منفی حقیقی (طبقه‌بندی مشتری بدحساب، تحت عنوان مشتری بدحساب)، FN موارد منفی کاذب (طبقه‌بندی مشتری بدحساب، تحت عنوان مشتری خوش‌حساب) و FP موارد مثبت کاذب (طبقه‌بندی مشتری خوش‌حساب، تحت عنوان مشتری بدحساب) است.

مدل پیشنهادی

همان طور که گفته شد، در این پژوهش از الگوریتم ژنتیک و الگوریتم ژنتیک با مرتب‌سازی نامغلوب برای انتخاب زیرمجموعه ویژگی استفاده شده است. همچنین، از طبقه‌بندهای ماشین بردار پشتیبان با کرنل‌های خطی، پرسپترون چندلایه‌ای، مدل تابع پایه شعاعی و شبکه‌های عصبی چندجمله‌ای و کوادراتیک و طبقه‌بند KNN با مقادیر K=۱,۳,۵,۷,۹ استفاده شده است. ابتدا کدهای مربوط به مباحث پیرامون الگوریتم ژنتیک و الگوریتم چندهدفه NSGAII، شامل توابع و عملگرهای تقاطع، جهش، انتخاب مسابقه‌ای دوتایی، انتخاب چرخ رولت، جمعیت مرتب و همچنین تسلط، فاصله ازدحام، مرتب‌سازی نامغلوب در محیط متلب نوشته می‌شود. سپس کدهای مربوط به قسمت پیش پردازش نوشته شده و نوع فرایند یادگیری و آزمایش مشخص می‌گردد. در این پژوهش برای انجام فرایند یادگیری و آزمون از روش اعتبارسنجی متقاطع K فولد بهره گرفته شده است. به این ترتیب که هر دسته داده به ۱۰ قسمت تقسیم شده است و در هر اجرا ۹ قسمت به‌عنوان دسته آموزش و یک قسمت به‌عنوان دسته آزمون مورد استفاده قرار گرفته است و این رویه تا مورد آموزش قرار گرفتن برای کلیه داده‌ها، تکرار شده است. درصد آموزش نیز برابر با ۰/۶ است.

در ادامه، مسئله انتخاب ویژگی از طریق الگوریتم ژنتیک (با هدف افزایش میزان دقت طبقه‌بندی) کدنویسی شده و از طریق طبقه‌بندهای KNN و SVM اجرا می‌گردند. در این پژوهش، الگوریتم ژنتیک با پارامترهای؛ تعداد تکرارها ۵۰ تکرار، تعداد جمعیت اولیه ۳۰ نمونه، عملگر تقاطع ۵۰ درصد، عملگر جهش ۴۰ درصد و ۱۰ درصد مربوط به نمونه‌های با بهترین نتیجه در تکرار قبلی، درصد تغییر ژن در هر کروموزوم ۰/۰۵ درصد، اجرا گردیده است. با اجرای کدهای الگوریتم ژنتیک در موضوع انتخاب ویژگی، نمودار بهترین، میانگین و بدترین دقت طبقه‌بندی در تکرارهای مختلف ترسیم شده است که از طریق این نمودار، همگرایی، سرعت همگرایی و میزان دقت طبقه‌بندی را می‌توان مشاهده نمود (شکل‌های ۲ و ۳). برای تعیین پارامترهای الگوریتم NSGAII از روش آزمایش و خطا استفاده شده است. پارامترهای الگوریتم NSGA II شامل؛ تعداد تکرارها ۵۰ تکرار، تعداد جمعیت اولیه ۲۰ نمونه، عملگر تقاطع ۸۰ درصد، عملگر جهش ۳۰ درصد و درصد تغییر ژن در هر کروموزوم ۱۰ درصد، اجرا گردیده است. پس از اجرای کدهای الگوریتم NSGA II در موضوع انتخاب ویژگی، کمترین میزان خطا در طبقه‌بندی، تعداد و شماره ویژگی‌های انتخاب شده، استخراج می‌گردد.

www.SID.ir

Archive of SID

پس از نوشتن کدهای الگوریتم‌های ژنتیک و NSGA II، شاخص دقت در سه حالت مقایسه شد: ۱. تمام ویژگی‌ها؛ ۲. انتخاب زیرمجموعه ویژگی‌ها با الگوریتم ژنتیک؛ ۳. انتخاب زیرمجموعه ویژگی‌ها با الگوریتم NSGA II. خلاصه پارامترهای قابل تنظیم در الگوریتم‌های پژوهش در جدول ۳ و نتایج هر پایگاه داده در جدول‌های ۴ و ۵ درج شده است.

جدول ۳. پارامترهای الگوریتم‌های در پژوهش

پارامترها	الگوریتم ژنتیک با مرتب‌سازی نامغلوب	الگوریتم ژنتیک
پایگاه داده‌ها	استرالیا، آلمان، ژاپن، ایران، لهستان، آمریکا	
K در K نزدیک‌ترین همسایه	۱، ۳، ۵، ۷، ۹	
کرنل‌های ماشین بردار پشتیبان	خطی، پرسپترون چند لایه‌ای، مدل تابع پایه شعاعی، چند جمله‌ای، کوادراتیک	
طبقه‌بندی	آموزش و آزمون، اعتبارسنجی متقاطع K فولد	
تکرار	۵۰	۵۰
جمعیت	۲۰	۳۰
نرخ ترکیب مجدد	۰/۱	۰/۱
درصد تقاطع	۰/۸	۰/۵
درصد جهش	۰/۳	۰/۴
نرخ جهش	۰/۱	۰/۰۵
قدرت رولت	---	۲

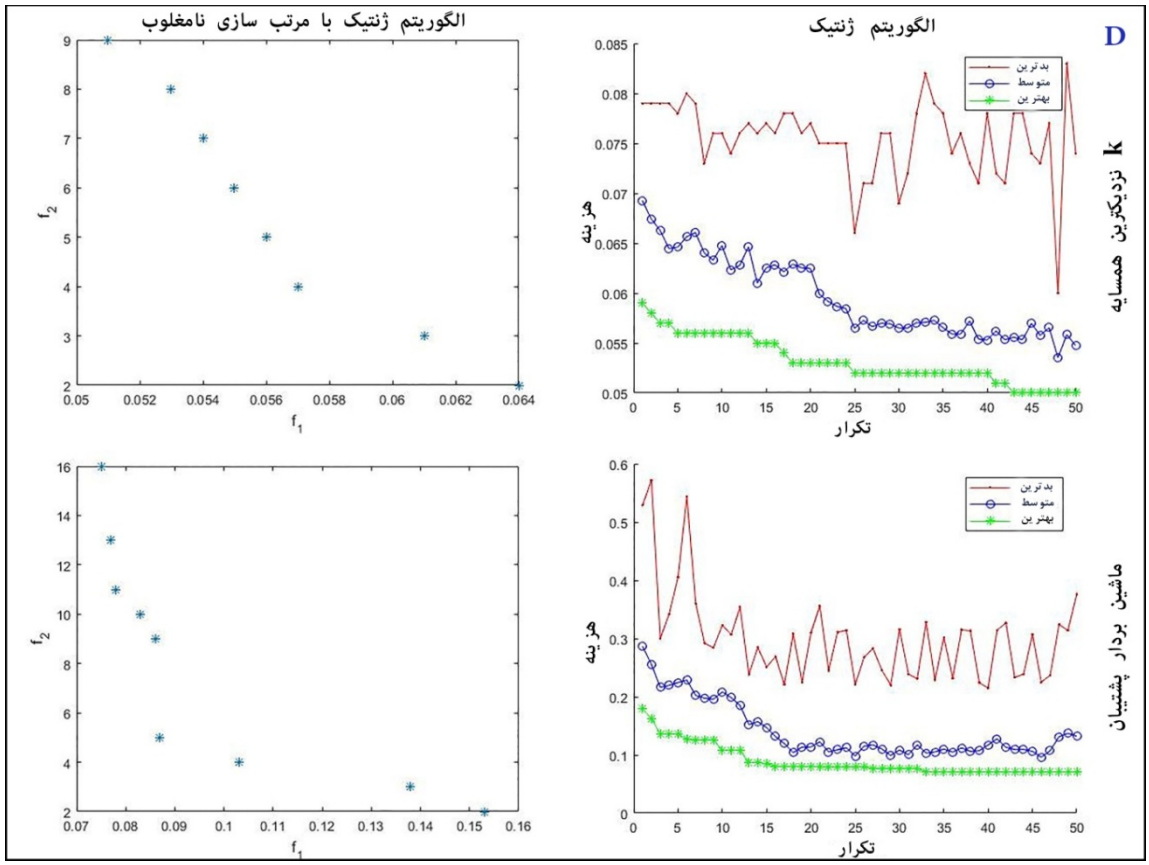
تحلیل داده‌ها

چنانچه پیش از این نیز اشاره شد در این بخش با استفاده از روش بهینه‌سازی چندهدفه الگوریتم ژنتیک NSGA-II به انتخاب بهینه ویژگی‌های برتر در هر مسئله اقدام گردیده است. به این منظور ۶ پایگاه داده را بررسی خواهیم نمود و برای هر یک ابتدا با استفاده از کلیه ویژگی‌ها طبقه‌بندی اتوماتیک را انجام خواهیم داد و نتایج را بر اساس معیارهای ارزیابی طبقه‌بندی، بررسی خواهیم نمود. سپس با استفاده از الگوریتم ژنتیک تک هدفه اقدام به تعیین بهترین زیرمجموعه از ویژگیها خواهیم نمود، به‌شرطی که خطای طبقه‌بندی به کمترین مقدار خود برسد. در مرحله آخر الگوریتم NSGA-II به‌گونه‌ای مورد استفاده قرار خواهد گرفت که هم خطای طبقه‌بندی و هم تعداد ویژگی‌های منتخب به حداقل ممکن برسد. در این راستا از ۶ پایگاه داده مربوط به داده‌های اعتباری ۶ کشور استرالیا، آلمان، ژاپن، ایران، لهستان و آمریکا استفاده شده است که به‌ترتیب دارای ۱۴، ۲۴، ۱۵، ۲۷، ۳۰ و ۳۸ ویژگی هستند. ۳ پایگاه داده اول (استرالیا، ژاپن و آلمان) از سایت معتبر UCI، پایگاه داده ایران از داده‌های پالایش یافته یک بانک خصوصی، پایگاه داده لهستان از داده‌های پالایش یافته ۱۲۰ شرکت ثبت شده در لهستان و پایگاه داده آمریکا از داده‌های پالایش یافته مسابقه داده‌کاوی که توسط دانشگاه کالیفرنیا سان دیگو و شرکت فیر ایساک ترتیب داده شده بود، استفاده نموده است. در هر پایگاه داده، ابتدا مسئله انتخاب ویژگی بر اساس الگوریتم ژنتیک تک هدفه و با هدف افزایش دقت طبقه‌بندی انجام گرفته است. سپس

www.SID.ir

Archive of SID

داده‌های هر پایگاه با الگوریتم ژنتیک چند هدفه NSGA II و با اهداف افزایش میزان دقت طبقه‌بندی و کاهش تعداد ویژگی، اجرا گردیده که در این راستا از دو طبقه‌بند رایج در مقوله داده‌کاوی یعنی؛ SVM و KNN استفاده شده است. شکل ۲، نمودار نتایج اجرای برنامه روی پایگاه دادهٔ ایران و شکل ۳ نتایج اجرای برنامه روی سایر پایگاه‌های داده را نشان می‌دهد.



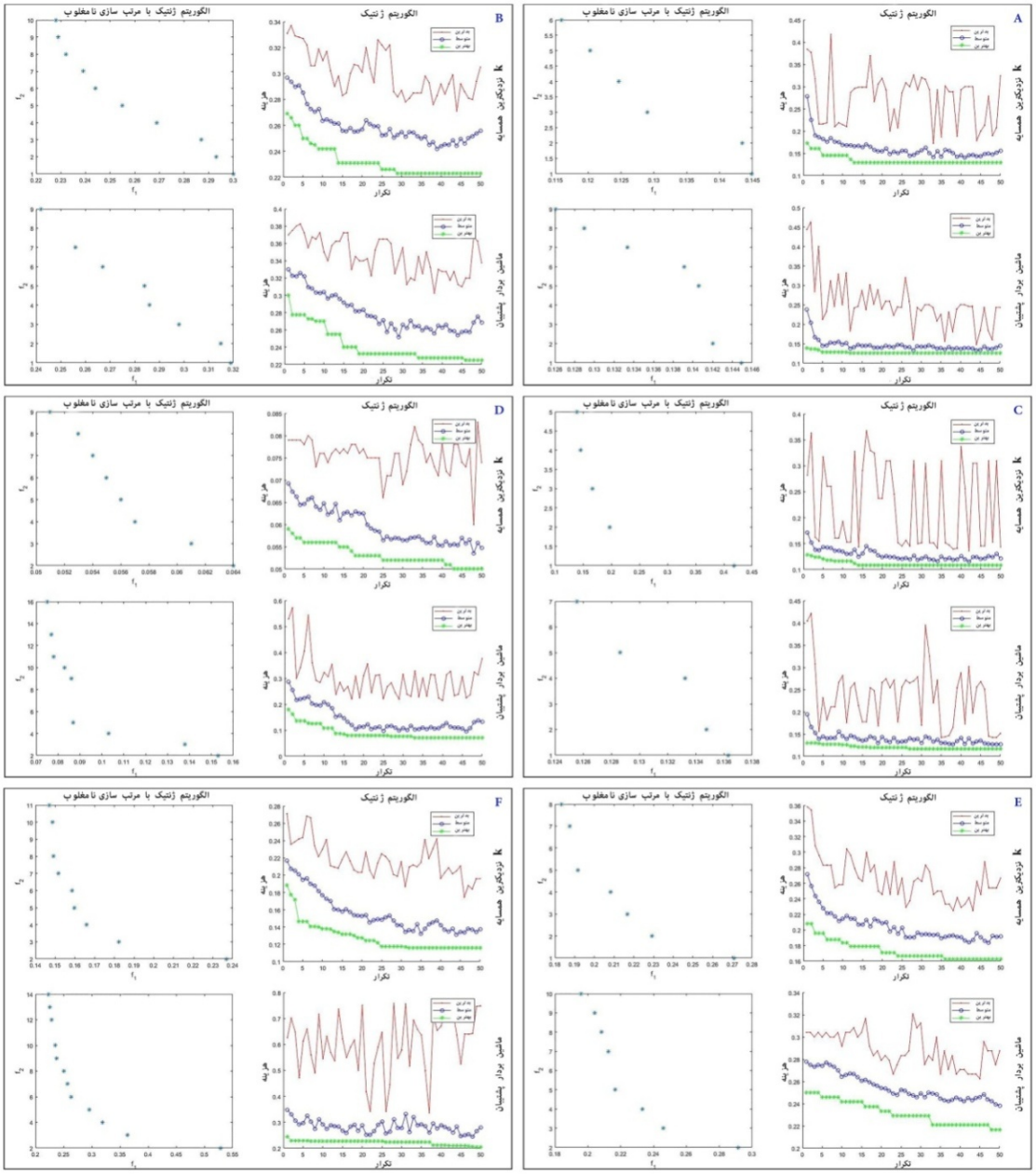
شکل۲. نمودارهای نتایج اجرای برنامه در پایگاه داده ایران (D)

با اجرای الگوریتم ژنتیک با طبقه‌بند KNN، میزان دقت طبقه‌بندی ۹۵ درصد و تعداد ویژگی‌های انتخاب شده ۱۳ مورد است. ضمن اینکه بهترین هزینه ۰/۰۵۰۰ برآورد گردید. با اجرای الگوریتم NSGA II با طبقه‌بند KNN، میزان دقت طبقه‌بندی ۹۴/۹۰ درصد و تعداد ویژگی‌های انتخاب شده برابر با ۹ است. برآورد صورت‌گرفته از بهترین هزینه نیز ۰/۰۵۱۰ می‌باشد. از سوی دیگر، با اجرای الگوریتم ژنتیک تک هدفه با طبقه‌بند SVM، میزان دقت طبقه‌بندی ۹۲/۹۰ درصد و تعداد ویژگی‌ها انتخاب شد ۱۸ مورد است. بهترین هزینه نیز برابر با ۰/۰۷۱۰ می‌باشد. نهایتاً، با اجرای الگوریتم ژنتیک چند هدفه NSGA II با طبقه‌بند SVM، میزان دقت طبقه‌بندی ۹۲/۵۰ درصد و تعداد ویژگی‌های انتخاب شده برابر با ۱۶ است. به علاوه، بهترین هزینه نیز برابر با ۰/۰۷۵۰ تخمین زده شد. با توجه به میانگین اجراها در هر دو طبقه‌بند KNN و SVM، مشاهده می‌شود که با کاهش تعداد ویژگی‌های انتخاب شده، میزان خطای طبقه‌بندی افزایش می‌یابد، با این وجود این میزان افزایش در خطا در مواردی می‌تواند در زمان و هزینه بررسی کلیه ویژگی‌ها، با صرفه‌تر

www.SID.ir

Archive of SID

باشد. نتایج اجرای الگوریتم ژنتیک و NSGA II با طبقه‌بندهای KNN و SVM در پایگاه‌های داده مختلف در شکل ۳ نشان داده شده است.



شکل ۳. نتایج اجرای برنامه در پایگاه داده استرالیا (A)، آلمان (B)، ژاپن (C)، ایران (D)، لهستان (E) و آمریکا (F)

برای نمونه در پایگاه داده استرالیا، طبقه‌بند KNN باعث کاهش تعداد میانگین ویژگی‌ها از ۱۴ ویژگی در حالت پایه به ۶/۸ ویژگی در الگوریتم ژنتیک و ۴/۴ ویژگی در NSGA II شده است. در مقابل، طبقه‌بند SVM تعداد ویژگی‌ها را از ۱۴ ویژگی به ۷/۶ ویژگی در الگوریتم ژنتیک و ۶/۶ ویژگی در NSGA II کاهش داده است که صرفه‌جویی زیادی در زمان و هزینه را به‌همراه خواهد داشت. در این پایگاه داده، کاهش تعداد ویژگی‌ها در طبقه‌بند KNN نمود بیشتری دارد.

www.SID.ir

Archive of SID

طبقه‌بند KNN افزایش دقت طبقه‌بندی از ۸۳/۶۸ درصد در حالت پایه، به میانگین ۸۸/۰۰ درصد در الگوریتم ژنتیک و ۸۷/۴۲ درصد در NSGA II به‌همراه داشته است. طبقه‌بند SVM نیز موجب افزایش متوسط دقت طبقه‌بندی از ۸۳/۸۳ درصد در حالت پایه به ۸۶/۵۸ درصد در الگوریتم ژنتیک و ۸۶/۴۳ درصد در الگوریتم NSGA II شده است. در مجموع اجرای الگوریتم‌ها روی پایگاه داده استرالیا نشان می‌دهد که با اندکی اغماض در دقت طبقه‌بندی، می‌توان تعداد ویژگی‌های مورد بررسی را کاهش داد و این امر صرفه‌جویی زیادی در زمان و هزینه را در پی خواهد داشت. در این پایگاه داده، طبقه‌بند KNN نسبت به طبقه‌بند SVM، در الگوریتم ژنتیک و الگوریتم NSGA II، تعداد ویژگی کمتر و دقت طبقه‌بندی بالاتری را ارائه می‌دهد.

جدول ٤. نتایج اجرای برنامه در پایگاه‌های داده با طبقه‌بند K نزدیک‌ترین همسایه بر اساس متوسط مقادیر K

پایگاه داده	تعداد ویژگی‌ها			دقت		
	روش پایه	الگوریتم ژنتیک	الگوریتم ژنتیک با مرتب‌سازی نامغلوب	روش پایه	الگوریتم ژنتیک	الگوریتم ژنتیک با مرتب‌سازی نامغلوب
استرالیا	۱۴	۶/۸	۴/۴	۸۳/۶۸	۸۸/۰۰	۸۷/۴۲
آلمان	۲۴	۱۲/۶	۹/۰	۷۱/۰۰	۷۶/۵۴	۷۶/۲۱
ژاپن	۱۵	۷/۰	۵/۸	۸۳/۹۸	۸۸/۲۱	۸۸/۱۲
ایران	۲۷	۱۳/۶	۵/۰	۹۴/۴۲	۹۵/۳۲	۹۵/۲۶
لهستان	۳۰	۱۳/۶	۶/۸	۷۵/۸۳	۸۳/۰۰	۸۱/۷۵
آمریکا	۳۸	۱۵/۶	۹/۴	۸۰/۴۴	۸۶/۰۱	۸۵/۶۷

جدول ۵. نتایج اجرای برنامه در پایگاه‌های داده با طبقه‌بند ماشین بردار پشتیبان بر اساس متوسط مقادیر کرنل‌ها

پایگاه داده	تعداد ویژگی‌ها			دقت		
	روش پایه	الگوریتم ژنتیک	الگوریتم ژنتیک با مرتب‌سازی نامغلوب	روش پایه	الگوریتم ژنتیک	الگوریتم ژنتیک با مرتب‌سازی نامغلوب
استرالیا	۱۴	۷/۶	۶/۶	۸۳/۸۳	۸۶/۵۸	۸۶/۴۳
آلمان	۲۴	۱۴/۰	۹/۰	۷۰/۳۵	۷۴/۵۸	۷۱/۳۸
ژاپن	۱۵	۸/۴	۴/۶	۸۴/۵۹	۸۷/۳۲	۸۶/۸۰
ایران	۲۷	۱۳/۸	۶/۶	۸۳/۲۸	۹۲/۴۴	۹۲/۹۶
لهستان	۳۰	۱۴/۶	۷/۰	۷۳/۰۸	۷۹/۳۳	۷۹/۰۰
آمریکا	۳۸	۲۲/۰	۱۲/۷۵	۶۷/۴۹	۸۱/۱۴	۷۹/۲۱

با توجه به نتایج اجرای الگوریتم‌ها روی پایگاه‌های داده آلمان، ایران، لهستان و آمریکا با استفاده از هر دو طبقه‌بند KNN و SVM، الگوریتم NSGA II نسبت به الگوریتم ژنتیک، تعداد ویژگی کمتری دارد و دلیل این مسئله آن است که کاهش تعداد ویژگی به‌عنوان یکی از توابع هدف استفاده شده است. همچنین در این پایگاه داده، طبقه‌بند KNN نسبت به طبقه‌بند SVM، در الگوریتم ژنتیک و الگوریتم NSGA II، دقت طبقه‌بندی بالاتر و تعداد ویژگی کمتری را

www.SID.ir

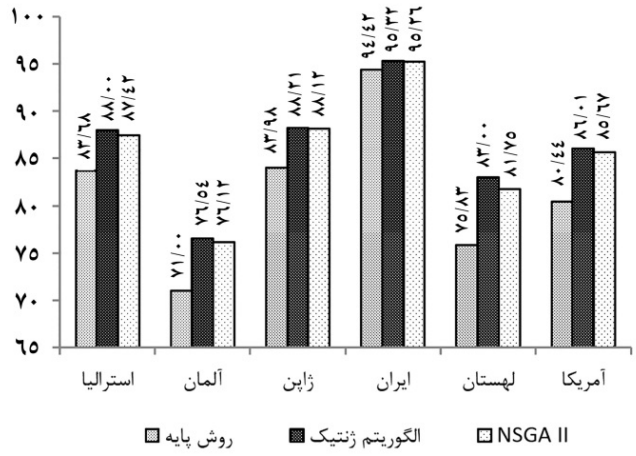
Archive of SID

ارائه می‌دهد. در پایگاه داده ژاپن نیز، طبقه‌بند KNN نسبت به طبقه‌بند SVM، در الگوریتم ژنتیک و الگـوریتم NSGA II، دقت طبقه‌بندی بالاتر و طبقه‌بند SVM نسبت به KNN تعداد ویژگی کمتری را ارائه می‌دهد.

نتیجه‌گیری و پیشنهادها

در این پژوهش از روش‌های بهینه‌سازی تکاملی «ماشین بردار پشتیبان» و «الگوریتم K نزدیک‌ترین همسایه» با قابلیت یادگیری با نظارت، برای انتخاب ویژگی‌های برتر در زمینه‌های مختلف طبقه‌بندی داده‌های اعتباری استفاده شـده اسـت. نتایج به‌دست آمده نشان می‌دهد که استفاده از الگوریتم ژنتیک تک هدفه با هدف بیشینه‌سازی صحت طبقه‌بنـدی قـادر است به‌صورت هم‌زمان تعداد ویژگی‌ها را در حد محدودی کاهش دهد. این در حالی است که تقریباً در تمامی حالـت‌هـا، تفاوت چشم‌گیری از لحاظ صحت نتایج طبقه‌بندی حاصل از دو روش تک هدفه و چندهدفه وجود ندارد. با انتخاب بهینه ویژگی‌ها در هر دو روش تک هدفه و چندهدفه، درصد صحت طبقه‌بندی افزایش و تعـداد ویژگـی‌هـا کـاهش مـی‌یابـد. بدیهی است که در روش چندهدفه که تمرکز خاصی بر کاهش هم‌زمان خطا و تعداد ویژگی‌ها دارد، این کاهش ویژگی‌ها مشهودتر است. بنابراین، با وجود اینکه روش چندهدفه تعداد ویژگی‌ها را نسبت به روش تک هدفه تقریباً در کلیه حالت‌ها به نصف کاهش داده است، تفاوت چندانی در صحت طبقه‌بندی ایجاد نکـرده اسـت. همچنـین در مقایسـه دو طبقـه‌بنـد استفاده شده در این پژوهش در داده‌های مورد اشاره، طبقه‌بند K نزدیک‌ترین همسایه نسبت به ماشین بردار پشـتیبان در افزایش دقت طبقه‌بندی و کاهش تعداد ویژگی‌ها، عملکرد نسبتاً بهتری داشته است.

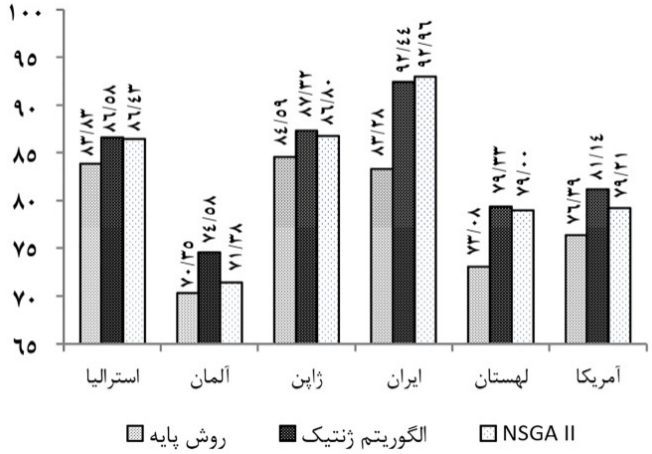
با توجه به نمودارهای ارائه شده می‌توان نتیجه گرفت که کاهش هم‌زمان خطـای طبقـه‌بنـدی و تعـداد ویژگـی‌هـا امکان‌پذیر است و علاوه بر افزایش صحت پیش‌بینی دربارهٔ عملکرد سیستم‌های مختلف داده‌کاوی، امکان کاهش اتلاف وقت و هزینه‌ها را میسر می‌سازد. این نکته با توجه به نتایج به‌دست آمده که نشان می‌دهد که تعداد ویژگی‌های انتخـاب شده در روش NSGA II بسیار قابل اهمیت است. نتایج نشان می‌دهند که قبل و بعد از انتخاب ویژگی با الگوریتم‌هـای ژنتیک و NSGA II، دقت طبقه‌بندی افزایش یافته است. شکل ۴ نتایج ناشی از انتخاب ویژگی با اسـتفاده از طبقـه‌بنـد KNN و شکل ۵ نتایج ناشی از انتخاب ویژگی با استفاده از طبقه‌بند SVM را نمایش می‌دهد.



شکل ٤. نمودار میانگین دقت طبقه‌بندی روش‌های مختلف (قبل و بعد از انتخاب ویژگی) با استفاده از طبقه‌بند K نزدیک‌ترین همسایه

www.SID.ir

Archive of SID



شکل ۵. نمودار میانگین دقت طبقه‌بندی روش‌های مختلف (قبل و بعد از انتخاب ویژگی) با استفاده از طبقه‌بند ماشین بردار پشتیبان

همان طور که مشاهده می‌شود، قبل و بعد از انتخاب ویژگی با الگوریتم‌های ژنتیک و NSGA II، تعداد ویژگی‌های انتخاب شده، کاهش یافته که مقدار آن در الگوریتم NSGA II تقریباً $\frac{1}{3}$ تعداد کل ویژگی‌ها است. مسلماً در روش NSGA II که تمرکز خاصی بر کاهش توأم خطا و تعداد ویژگی‌ها دارد، این کاهش ویژگی‌ها مشهودتر است.

در تحقیق حاضر شاخص دقت برای نمایش عملکرد مدل پیش‌بینی در تمامی طبقات استفاده شده است که معیار متداول‌تری برای اندازه‌گیری عملکرد است. در حالی که تشخیص و حساسیت، عملکرد مدل پیش‌بینی را در هر طبقه به‌صورت مجزا محاسبه می‌کنند. در نظر گرفتن هم‌زمان تشخیص و حساسیت به‌عنوان دو شاخص متضاد می‌تواند در پژوهش‌های آتی مورد توجه قرار گیرد. به علاوه، در راستای انجام پژوهش‌های آتی پیشنهاد می‌شود که از سایر الگوریتم‌های بهینه‌سازی فراابتکاری نظیر الگوریتم پرندگان و زنبور عسل برای انتخاب ویژگی استفاده شود و نتایج به‌دست آمده با نتایج این پژوهش مقایسه گردد. به علاوه می‌توان از سایر طبقه‌بندها، مانند: شبکه‌های عصبی و شبکه‌های فازی عصبی و یا ترکیب طبقه‌بندهای مختلف برای بهبود نتایج استفاده کرد

منابع

رزمی، جعفر؛ حیدریه، سید عبدالله؛ شهابی، علی (۱۳۹۳). توسعه مدل پذیرش فناوری در بانکداری ایران (پژوهشی پیرامون بانک رفاه). *مدیریت صنعتی*، ۶ (۳)، ۴۷۱-۴۹۰.

نصرتی ناهوک، حسن؛ افتخاری، مهدی (۱۳۹۲). یک روش جدید برای انتخاب ویژگی مبتنی بر منطق فازی. *هوش محاسباتی در مهندسی برق*، ۴ (۱)، ۷۱-۸۴.

همایون‌فر، مهدی؛ باقرسلیمی، سعید؛ نهاوندی، بیژن؛ ایزدی شیجانی، کاوه (۱۳۹۷). شبیه‌سازی مبتنی بر عامل شبکه تأمین شرکت ملی پخش فراورده‌های نفتی در قالب سیستم انطباقی پیچیده به‌منظور دستیابی به سطح موجودی بهینه. *مدیریت صنعتی*، ۱۰ (۴)، ۶۰۷-۶۳۰.

www.SID.ir

Archive of SID

References

Abd-Elazizb, M., Ewees, A.A., Ibrahim, R.A., and Lu, F. (2020). Opposition-based moth-flame optimization improved by differential evolution for feature selection. *Mathematics and Computers in Simulation*, 168, 48-75.

Agrawal, R.K., Kaura, B., and Sharma, S. (2020). Quantum based Whale Optimization Algorithm for wrapper feature selection. *Applied Soft Computing*, 89. https://doi.org/10.1016/j.asoc.2020.106092.

Alazzam, H., Sharieh, A., Sabri, K.E. (2020). A feature selection algorithm for intrusion detection system based on Pigeon Inspired Optimizer. *Expert Systems with Applications*, 148, https://doi.org/10.1016/j.eswa.2020.113249.

Amini, F., and Hu, G. (2020). A two-layer feature selection method using Genetic Algorithm and Elastic Net. *Expert Systems with Applications*, 166, https://doi.org/10.1016/j.eswa.2020.114072

Baia, L., Han, Z., Ren, J., and Qin, X. (2021). Research on feature selection for rotating machinery based on Supervision Kernel Entropy Component Analysis with Whale Optimization Algorithm. *Applied Soft Computing*, 92, https://doi.org/10.1016/j.asoc.2020.106245.

Bolon- Canedo, V., Sanchez- Marono, N., and Alonso- Betanzos, A. (2015). Recent advances and emerging challenges of feature selection in the context of big data, *Knowledge-Based Systems*, 86, 33-45.

Cai, F., Wang, H., Tang, X., Emmerich, M., and Verbeek, F.J. (2016). Fuzzy Criteria in Multi-objective Feature Selection for Unsupervised Learning, *Procedia Computer Science*, 102, 51- 58.

Chaudhuri, A., and Sahu, T.P. (2020). A hybrid feature selection method based on Binary Jaya algorithm for micro-array data classification. *Computers & Electrical Engineering*, 90(12). https://doi.org/10.1016/j.compeleceng.2020.106963.

Chen, K., Zhou, F.Y., and Yuan, X.F. (2019). Hybrid Particle Swarm Optimization with Spiral-Shaped Mechanism for Feature Selection. Expert *Systems with Applications*, 128, 140-156.

Cura, T. (2019). Use of support vector machines with a parallel local search algorithm for data classification and feature selection. *Expert Systems with Applications*, 145(1), 113-133.

Das, A., and Das, S. (2017). Feature weighting and selection with a Pareto-optimal trade-off between relevancy and redundancy. *Pattern Recognition Letters*, 88, 12-19.

Das, K., Mishra, D., and Shaw, K. (2016). A Meta heuristic optimization framework for informative gene selection. *Informatics in Medicine Unlocked*, 4, 10-20.

Deb, K., Agrawal, S., Pratap, A., and Meyarivan, T. (2000). A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II, *6th International Conference on Parallel Problem Solving from Nature*, 849-858.

www.SID.ir

Archive of SID

Effrosynidis, D., and Arampatzis, A. (2021). An evaluation of feature selection methods for environmental data. *Ecological Informatics*, 61, https://doi.org/10.1016/j.ecoinf.2021.101224.

Fonseca, C.M., and Fleming, P.J. (1993). Genetic Algorithms for Multi-objective Optimization: Formulation, Discussion and Generalization, *5th International Conference on Genetic Algorithms*, 93(3), 416-423.

Gao, W., Hu, L., Zhang, P., and He, J. (2018). Feature selection considering the composition of feature relevancy. *Pattern Recognition Letters*, 112, 70-74.

García-Pedrajas, N., de Castillo, J.A.R., and Cerruela-García, G. (2020). Fast simultaneous instance and feature selection for datasets with many features. *Pattern Recognition*, 111, 107-123.

Gholami, J., Pourpanah, F., and Wang, X. (2020). Feature selection based on improved binary global harmony search for data classification. *Applied Soft Computing*, 93, https://doi.org/10.1016/j.asoc.2020.106402.

Hancer, E., Xue, B., Zhang, M., Karaboga, D., and Akay, B. (2018). Pareto front feature selection based on artificial bee colony optimization, *Information Sciences*, 422, 462-479.

Hashemi, A., and Bagher, M. (2021). A pareto-based ensemble of feature selection algorithms. *Expert Systems with Applications*, 180, 115-130

Homayounfar, M., Baghersalimi, S., Nahavandi, B., and Izadi Sheyjani, K. (2018). Agent-based Simulation of National Oil Products Distribution Company's Supply Network in the Framework of a Complex Adaptive System in Order to Achieve an Optimal Inventory Level. *Industrial Management Journal*, 10(4), 607-630. (*in Persian*)

Horn, J., Nafpliotis, N., and Goldberg, D.E. (1994). A niched Pareto genetic algorithm for multi-objective optimization. Evolutionary Computation, *IEEE World Congress on Computational Intelligence*, 82-87.

Huang, B., Buckley, B., and Kechadi, T.M. (2010). Multi-objective feature selection by using NSGA -II for customer churn prediction in telecommunications. *Expert Systems with Applications,* 37(5), 3638-3646.

Ibrahim, R.A., Abd Elaziz, M., Ewees, A.A., El-Abd, M., and Lu, S. (2021). New Feature Selection Paradigm Based on Hyper-heuristic Technique. *Applied Mathematical Modelling*, 98, 14-37. https://doi.org/10.1016/j.apm.2021.04.018.

Junqueira, N., Marcelo, M., Nagano, S. (2020). Unsupervised feature selection based on bio-inspired approaches. *Swarm and Evolutionary Computation*, 52, https://doi.org/10.1016/j.swevo.2019.100618

Kashef, S., & NezamabadiPour, H. (2015). An advanced ACO algorithm for feature subset selection. *Neurocomputing,* 147, 271–279.

Khan, A., and Baig, A.R. (2015). Multi- Objective Feature Subset Selection using Non-dominated Sorting Genetic Algorithm. *Journal of Applied Research and Technology*, 13(1), 145-159.

www.SID.ir

Archive of SID

Lee, I.G., Zhang, Q., Yoon, S.W., and Won, D. (2020). A mixed integer linear programming support vector machine for cost-effective feature selection. *Knowledge-Based Systems*, 203, https://doi.org/10.1016/j.knosys.2020.106145.

Mohamed, A.A.A., Hassan, S.A., Hemeida, A.M., Alkhalaf, S., Mahmoud, M.M.M., and Baha-Eldin. A.M. (2020). Parasitism – Predation algorithm (PPA): A novel approach for feature selection. *Ain Shams Engineering Journal*, 11(2), 293-308.

Müller, I.M. (2021). Feature selection for energy system modeling: Identification of relevant time series information. *Energy and AI*, 4. https://doi.org/10.1016/j.egyai.2021.100057.

Nematzadeh, H., Enayatifar, R., Mahmud, M., and Akbari, E. (2019). Frequency based feature selection method using whale algorithm, *Genomics*, 111(6), 1946-1955.

Nosrati nahook, H., and Eftekhari, M. (2013). A New Method for Feature Selection Based on Fuzzy Logic, *Computational Intelligence in Electrical Engineering*, 4(1), 71-84. (*in Persian*)

Razmi, J., Heydaeriyeh, S.A., and Shahabi, A. (2014). Development of technology acceptance model in Iranian banking (Case study: Refah Bank of Semnan province). *Industrial Management Journal*, 6(3), 471-490. (*in Persian*)

Rodrigues, D., de Albuquerque, V.H.C., and Papa, J.P. (2020). A multi-objective artificial butterfly optimization approach for feature selection. *Applied Soft Computing*, 94, https://doi.org/10.1016/j.asoc.2020.106442.

Schaffer, J.D. (1985). Some experiments in machine learning using vector evaluated genetic algorithms, *PhD Dissertation*. Vanderbilt University, Nashville, TN, USA.

Sohrabi, M.K., and Tajik, A. (2017). Multi-objective feature selection for warfarin dose prediction, *Computational Biology and Chemistry*, 69, 126-133.

Tang, X., Dai, Y., and Xiang, Y. (2019). Feature selection based on feature interactions with application to text categorization. *Expert Systems with Applications*, 120, 207-216.

Thirumoorthy, K., and Muneeswaran, K. (2021). Feature selection using hybrid poor and rich optimization algorithm for text classification. *Pattern Recognition Letters*, 147, 63-70.

Tseng, T.L., and Huang, C.C. (2007). Rough set-based approach to feature selection in customer relationship management. *Omega*, 35(4), 365-383.

Wei, G., Zhao, J., Feng, Y., He, A., and Yu, J. (2020). A novel hybrid feature selection method based on dynamic feature importance. *Applied Soft Computing*, 93, https://doi.org/10.1016/j.asoc.2020.106337.

Xue, B., Zhang, M., and Browne, W.N. (2013). Particle swarm optimization for feature selection in classification: A multi-objective approach. *IEEE transactions on cybernetics*, 43(6), 1656-1671.

Xue, Y., Zhong, J., Tan, T.H., Liu, Y., Cai, W., Chen, M., and Sun, J. (2016). IBED: Combining IBEA and DE for optimal feature selection in software product line engineering. *Applied Soft Computing*, 49, 1215–1231.

Zakeri, A., and Hokmabadi, A. (2019). Efficient feature selection method using real-valued grasshopper optimization algorithm. *Expert Systems with Applications*, 119, 61-72.

www.SID.ir

Archive of SID

Zeng, D., Wang, S., Shen, Y., and Shi, S. (2017). A GA-based feature selection and parameter optimization for support tucker machine. *Procedia Computer Science*, 111, 17-23.

Zhang, X., Fan, Y., and Yang, J. (2021). Feature selection based on fuzzy-neighborhood relative decision entropy. *Pattern Recognition Letters*, 146, 100-107.

Zhao, X., Cao, Y., Zhang, T., and Li, F. (2021). An improve feature selection algorithm for defect detection of glass Bottles. *Applied Acoustics*, 174, 107794.

Zhong, W., Chen, X., Ni, F., and Huang, J.Z. (2021). Adaptive discriminant analysis for semi-supervised feature selection. *Information Sciences*, 566, 178-194.

Zhou, Y., Kang, J., Kwong, S., Wang, X., and Zhang, Q. (2021). An evolutionary multi-objective optimization framework of discretization-based feature selection for classification. *Swarm and Evolutionary Computation*, 60, https://doi.org/10.1016/j.swevo.2020.100770.

www.SID.ir