Archive of SID

Journal of Information Technology Management
ISSN: 2008-5893
EISSN: 2423-5059
Vol. 9, No. 1; PP. 123 - 142
Spring 2017

University of Tehran
Faculty of Management

Attribute Reduction in Incomplete Information System based on Rough Set Theory Using Fuzzy Imperialist Competitive Algorithm

Mohammad Ghanei Ostad [1], Hosein Khosravi Mahmoee [2], Majid Abdolrazzagh Nezhad [3]

Abstract: In recent years, rough set theory has been considered as a strong solution to solve artificial intelligence problem such as data mining. But, the classic rough set theory is not effective in the case of attribute reduction in incomplete information systems. Since there are null values for some of attributes in a data set, an incomplete information system is created. In this paper, a novel method proposed to solve attribute reduction in incomplete information system based on rough set theory by combining and modifying imperialist competitive algorithm with fuzzy logic. Utilizing the fuzzy logic to control the parameters of the algorithm was useful and generated better solutions compared to its classic draft. In this research, no changes imposed on incomplete data, and it was just considered as a complete systems. The fuzzy imperialist competitive algorithm acted intelligently to reduce the number of attribute in incomplete information system, providing appropriate results that is worthy of attention.

Key words: *Attribute reduction, Fuzzy logic, Imperialist competitive algorithm, Incomplete information system, Rough set theory.*

1. MSc. Student in IT Engineering, University of Birjand, Iran
2. MSc. Student in IT Engineering, University of Birjand, Iran
3. Assistant Prof., Faculty of Engineering, Dep. of Computer, Bozorgmehr University of Qaenat, Qaen, Iran

Submitted: 13/ February / 2016
Accepted: 24 / January / 2017
Corresponding Author: Mohammad Ghanei Ostad
Email: m.ghanei1992@gmail.com

www.SID.ir

Archive of SID

مدیریت فناوری اطلاعات
Journal of Information Technology Management
دانشکدهٔ مدیریت دانشگاه تهران
دورهٔ ۹، شمارهٔ ۱
بهار ۱۳۹۶
صص. ۱۴۲- ۱۲۳

کاهش ویژگی سیستم‌های اطلاعاتی ناقص بر مبنای تئوری مجموعهٔ راف با استفاده از الگوریتم رقابت استعماری فازی

محمد قانعی استاد[۱]، حسین خسروی مهموئی[۲]، مجید عبدالرزاق نژاد[۳]

چکیده: در سال‌های اخیر، تئوری مجموعهٔ راف به یکی از راه‌حل‌های قدرتمند در حل مسائل هوش مصنوعی همچون داده‌کاوی تبدیل شده است. اما نسخهٔ کلاسیک تئوری مجموعهٔ راف برای بحث کاهش ویژگی در سیستم‌های اطلاعاتی ناقص، چندان مناسب نیست. یک سیستم اطلاعاتی ناقص به جدول‌هایی از داده‌ها اطلاق می‌شود که برخی درایه‌های صفات آن مقداری ندارند. در این مقاله، راه‌حل نوینی که ترکیبی از الگوریتم رقابت استعماری و منطق فازی است، برای حل مسئلهٔ کاهش ویژگی سیستم‌های اطلاعاتی ناقص مبتنی بر تئوری مجموعهٔ راف ارائه شده است. نتایج نشان داد استفاده از منطق فازی در کنترل پارامترهای الگوریتم رقابت استعماری مفید است و در مقایسه با نسخهٔ کلاسیک الگوریتم، جواب‌های بهینه‌تری به‌دست می‌آورد. در روند اجرای این پروژه، تغییری روی داده‌های ناقص اعمال نشد و به آن همچون سیستم اطلاعاتی کامل نگاه شد. الگوریتم رقابت استعماری فازی به‌صورت هوشمند عمل کرده و برای کاهش ویژگی در سیستم‌های اطلاعاتی ناقص، نتایج مناسبی ارائه داد که درخور تأمل است.

واژه‌های کلیدی: الگوریتم رقابت استعماری، تئوری مجموعهٔ راف، سیستم اطلاعاتی ناقص، کاهش ویژگی، منطق فازی.

۱. دانشجوی کارشناسی ارشد مهندسی فناوری اطلاعات، دانشکدهٔ برق و کامپیوتر، دانشگاه بیرجند، بیرجند، ایران
۲. دانشجوی کارشناسی ارشد مهندسی فناوری اطلاعات، دانشکدهٔ برق و کامپیوتر، دانشگاه بیرجند، بیرجند، ایران
۳. استادیار دانشکدهٔ فنی مهندسی، گروه مهندسی کامپیوتر، دانشگاه بزرگمهر قائنات، قائن، ایران

تاریخ دریافت مقاله: ۱۳۹٤/۱۱/۲٤
تاریخ پذیرش نهایی مقاله: ۱۳۹۵/۱۱/۰۵
نویسندهٔ مسئول مقاله: محمد قانعی استاد
E-mail: m.ghanei1992@gmail.com

www.SID.ir

Archive of SID

مقدمه

امروزه در بیشتر سازمان‌ها، داده‌ها به‌سرعت جمع‌آوری و ذخیره می‌شوند، اما می‌توان ادعا کرد علی‌رغم این حجم انبوه داده، سازمان‌ها با فقر دانش در تصمیم‌گیری روبه‌رو هستند (هان و کمبر، ۲۰۰۶). همچنین با توجه به ذخیرهٔ ده‌ها، صدها یا حتی هزاران ویژگی در پایگاه دادهٔ برنامه‌های کاربردی، بحث کاهش ویژگی در جهت تحلیل این داده‌ها، امر ضروری و انکارناپذیر است (گایون و الیزف، ۲۰۰۳؛ یو، ۲۰۰۵). در بررسی پایگاه‌های داده، گاهی با سیستم‌های اطلاعاتی ناقص[1] مواجه‌ایم. سیستم اطلاعاتی ناقص به جدول‌هایی از داده‌ها اطلاق می‌شود که برخی درایه‌های صفات آن بی‌مقدارند (اورلاسکا و پاولاک، ۱۹۸۴). یکی از رویکردهای تحلیل این سیستم‌ها، این است که سطر مربوط به مقادیر ناقص را حذف کنیم (کمیلسکی، گریزمالا، پترسن و تان، ۱۹۹۳)؛ اما این رویکرد مناسب نیست، زیرا اندازهٔ داده‌ها را کاهش می‌دهد و امکان حذف اطلاعات مفید وجود دارد. روش دیگر برای تحلیل این نوع سیستم‌های اطلاعاتی این است که داده‌های ناموجود را به‌صورت‌های مختلف مقداردهی کنیم. برای مثال از تحلیل‌های آماری استفاده‌کرده و داده‌های ناموجود را حدس بزنیم (گونلان، ۱۹۸۶؛ کمیلسکی و همکاران، ۱۹۹۳). هر دو رویکرد برای کامل‌شدن سیستم‌های اطلاعاتی کاربرد دارند، اما به‌کارگیری روش مناسب برای کامل‌کردن سیستم‌های اطلاعاتی ناقص، خود یک موضوع تحقیق پیچیده است (منگ و ژی، ۲۰۰۹).

در سال‌های اخیر تئوری مجموعهٔ راف[2] به یکی از راه‌حل‌های قدرتمند در حل مسائل هوش مصنوعی همچون داده‌کاوی تبدیل شده است (پاولاک و اسکورن، ۲۰۰۷). یکی از راهکارهای اساسی برای استخراج دانش و کشف الگوهای پنهان از سیستم‌های اطلاعاتی[3] بزرگ، استفاده از داده‌کاوی[4] است (هان و کمبر، ۲۰۰۶). کاهش ویژگی[5] که همان انتخاب ویژگی[6] نامیده می‌شود، یکی از موضوعات مهم در تئوری مجموعهٔ راف است، اما نسخهٔ کلاسیک تئوری مجموعهٔ راف برای بحث کاهش ویژگی در سیستم‌های اطلاعاتی ناقص چندان مناسب نیست (کیان، لیانگ، پدریک و دانگ، ۲۰۱۱). برای کاهش ویژگی، دو استراتژی کلی به نام بسته‌بندی[7] (کوهاوی و جان، ۱۹۹۷) و فیلتر[8] مطرح شده است. در گذشته، یک الگوریتم یادگیری برای ارزیابی مجموعهٔ

---

1. Incomplete Information System
2. Rough Set Theory
3. Information System
4. Data Mining
5. Attribute Reduction
6. Feature Selection
7. Wrapper
8. Filter

www.SID.ir

Archive of SID

ویژگی‌های انتخاب‌شده به‌کار می‌رفت، اما به مرور زمان مجموعهٔ ویژگی‌ها به‌وسیلهٔ معیارهای با معنی‌ای همچون افزایش اطلاعات[1] (لی و لی، ۲۰۰۶؛ کونلان، ۱۹۸۶)، سازگاری (داش و لیو، ۲۰۰۳؛ کوین، لیانگ و دانگ، ۲۰۰۸)، مسافت (کایرا و رندل، ۱۹۹۲)، وابستگی (مورزجسکی، ۱۹۹۲) و غیره بررسی می‌شوند. این معیارها به دو دستهٔ کلی مبتنی بر مسافت و مبتنی بر سازگاری، طبقه‌بندی شده‌اند (هو، ژیو و یو، ۲۰۰۷).

اغلب روش‌های بهینه‌سازی مبتنی بر هوش مصنوعی همانند الگوریتم ژنتیک، شبیه‌سازی فرایندهای طبیعی هستند. یکی از دلایل این امر ملموس‌بودن، سادگی فرموله‌کردن و درک تکامل این الگوریتم‌هاست. الگوریتم رقابت استعماری[2]، با الهام‌گیری از نوعی فرایند اجتماعی ـ سیاسی نسبت به سایر روش‌های مطرح‌شده در این زمینه توانایی زیادی دارد و از سرعت مناسبی نیز برخوردار است (آتش‌پز و لوکاس، ۲۰۰۷؛ صفوی، پورجعفریان و صفوی، ۱۳۹۳: ۱۰۴).

راه‌حل نوین ارائه‌شده در این مقاله، ترکیبی از الگوریتم رقابت استعماری و منطق فازی است که در حل مسئلهٔ کاهش ویژگی سیستم‌های اطلاعاتی ناقص مبتنی بر تئوری مجموعهٔ راف کاربرد دارد. استفاده از منطق فازی در کنترل پارامترهای الگوریتم رقابت استعماری بسیار مفید است و در مقایسه با نسخهٔ بدون فازی الگوریتم، جواب‌های بهینه‌ای می‌دهد. منطق فازی از اجزای محاسبات نرم[3] است که روش محاسباتی جدیدی محسوب می‌شود و توانایی‌های شاخص ذهن انسان را برای استدلال و یادگیری در محیط نامعین و نادقیق گرد هم می‌آورد (لطفی‌زاده، ۱۹۹۲). الگوریتم رقابت استعماری فازی هوشمندانه عمل می‌کند و روی سیستم‌های اطلاعاتی ناقص نتایج مناسبی ارائه می‌دهد.

بخش‌های مختلف این مقاله بدین شرح است؛ ابتدا پیشینهٔ پژوهش بررسی‌شده و پژوهش‌های موفق در زمینهٔ کاهش ویژگی مبتنی بر تئوری مجموعهٔ راف مرور می‌شوند. بخش روش‌شناسی پژوهش، به توضیح روند اجرای پروژه اختصاص دارد. در قسمت یافته‌های پژوهش، نتایج به‌دست‌آمده از اجرای روش، به‌صورت مقایسه‌ای مطرح می‌شود و در نهایت نتیجه‌گیری و پیشنهادها ارائه خواهد شد.

پیشینهٔ پژوهش

در دو دههٔ گذشته، تعداد زیادی از روش‌های کاهش ویژگی مبتنی بر تئوری مجموعهٔ راف ارائه شده است (لی، ژانگ و لیونگ، ۲۰۰۴؛ می، وا و ژانگ، ۲۰۰۳؛ وا، ژانگ، و لی، ۲۰۰۵؛ شاو و

1. Information gain
2. Imperialist Competitive Algorithm (ICA)
3. Soft Computing

www.SID.ir

Archive of SID

ژانگ، ۲۰۰۵؛ کیان، لیانگ و دانگ، ۲۰۰۸). یکی از مشکلات اساسی این روش‌ها، زمان زیادی است که صرف پردازش مجموعهٔ داده‌های بزرگ می‌شود. در ادامه تعدادی از تحقیق‌های انجام‌شده در این حوزهٔ پژوهشی بررسی می‌شود.

پیشینهٔ نظری

در بررسی پیشینهٔ نظری کاهش ویژگی بر مبنای تئوری مجموعهٔ راف، دو رویکرد اساسی را در مقابل داریم. ابتدا رویکردهایی که کلاسیک هستند و از همان نسخهٔ اصلی تئوری مجموعهٔ راف استفاده می‌کنند. رویکرد دوم، توسعه‌یافتهٔ رویکرد قبلی است که الگوریتم‌های اکتشافی کاهش ویژگی را ارائه می‌کند. نکتهٔ مهم در رویکردهای مختلف کاهش ویژگی بر مبنای تئوری مجموعهٔ راف، نوع سیستم‌های اطلاعاتی است. در اغلب پژوهش‌های اجراشده بر سیستم‌های اطلاعاتی کامل تمرکز شده است؛ اما در دنیای واقعی، مجموعهٔ داده‌های جمع‌آوری‌شده همیشه کامل نیستند. پس چالش اساسی کاهش ویژگی مبتنی بر تئوری مجموعهٔ راف، کار روی سیستم‌های اطلاعاتی ناقص است.

پیشینهٔ تجربی

در بررسی پیشینهٔ تجربی، تعدادی از پروژه‌های موفق در زمینهٔ کاهش ویژگی مبتنی بر تئوری مجموعهٔ راف مرور می‌شوند. این پژوهش‌ها در همان قالب توضیح داده‌شده در بخش قبل ارائه شده‌اند.

یکی از تحقیقات ابتدایی انجام‌شده، روی چند روش مبتنی بر تئوری مجموعهٔ راف تمرکز دارد که برای استخراج قوانین از جدول‌های تصمیم به‌کار می‌روند. معمولاً این روش‌ها با استفاده از ماتریس‌های عینی عمل می‌کنند (اسکورن، ۱۹۹۵). در این پژوهش، کارایی روش‌های مطرح‌شده و دقت استخراج قوانین پایین بود.

برای افزایش کارایی عملکرد کاهش ویژگی در تئوری مجموعهٔ راف، چندین الگوریتم اکتشافی توسعه داده شده است (ها و همکاران، ۲۰۰۷؛ ها، یو و ژیو، ۲۰۰۶؛ ها و سرکون، ۱۹۹۵؛ لیانگ، چین، دانگ و ریچد، ۲۰۰۲؛ کیان و لیانگ، ۲۰۰۸؛ اسلزاک، ۲۰۰۲؛ وانگ، ژو و یانگ، ۲۰۰۲؛ وانگ، ژائو و آن، ۲۰۰۵؛ وا، لی، هانگ و لیو، ۲۰۰۴). هر یک از این الگوریتم‌ها در حفظ ویژگی‌های خاص سیستم اطلاعاتی داده شده کمک می‌کنند.

برای کاهش ویژگی در سیستم‌های اطلاعاتی ناقص، مشابه تحقیقات ابتدایی، یک ماتریس عینی به‌صورت کلی مطرح‌شده که بحث کاهش ویژگی را روی جدول‌های تصمیم ناقص انجام

www.SID.ir

Archive of SID

می‌دهد (کریزکیوسز، ۱۹۹۸). برای افزایش کارایی در بحث کاهش ویژگی روی داده‌های ناقص، چندین الگوریتم اکتشافی ارائه‌شده که در ادامه تعدادی از آنها بررسی می‌شود.

ایدهٔ کاهش نواحی مثبت بدین صورت مطرح شد که یک الگوریتم کاهش ویژگی اکتشافی، روی جدول‌های تصمیم ناقص اعمال می‌شود. روش آن به این شکل است که ناحیهٔ مثبت هدف تصمیم را بدون تغییر نگه می‌دارد (یانگ و شو، ۲۰۰۶). همچنین تعریف جدید از درگاشت اطلاعات[1]، برای اندازه‌گیری اجمالی از سیستم‌های اطلاعاتی ناقص عنوان شد (لیانگ، شی، لی و وایرمن، ۲۰۰۶). علاوه‌بر این، اِعمال متناظر درگاشت شرطی[2] در کاهش ویژگی‌های دارای افزونه، شامل این تعریف بود (لیانگ و ژو، ۲۰۰۲). دو دانشمند دیگر ترکیب درگاشت برای اندازه‌گیری اجمالی از یک سیستم اطلاعاتی ناقص و همچنین استفاده از درگاشت شرطی برای دستیابی به زیرمجموعه‌ای از ویژگی‌ها را پیشنهاد دادند (کیان، لیانگ و وانگ، ۲۰۰۹). بعد از درگاشت اطلاعات که برای کاهش ویژگی در تئوری مجموعهٔ راف به‌صورت کلاسیک مطرح شد، نمونهٔ توسعه‌یافته‌ای از آن برای تکمیل روش قبلی ارائه گردید (اسلزاک، ۲۰۰۲).

در یکی از پژوهش‌های انجام‌شده به بحث بهینه‌سازی و اثربخشی[3] الگوریتم‌های اکتشافی کاهش ویژگی در سیستم‌های اطلاعاتی ناقص پرداخته شده است. چارچوب جدیدی که در این پژوهش برای تئوری مجموعهٔ راف ارائه شد، «تقریب مثبت»[4] در سیستم‌های اطلاعاتی ناقص نامیده می‌شود. یکی از نتایج این چارچوب، این است که می‌تواند ساختار ناقص تئوری مجموعهٔ راف را شناسایی کند (کیان، لیانگ، پدریک و دانگ، ۲۰۱۱).

اما یکی از چالش‌های اساسی پژوهش‌های انجام‌شده در این حوزه، صرف زمان زیاد برای پردازش داده‌های ناقص با حجم بالاست. نکتهٔ دیگری که در تحقیقات پیشین همواره به‌عنوان یک چالش مطرح است، کارایی کم روش‌های ارائه شده است.

روش‌شناسی پژوهش

در این بخش، پس از تعریف مفاهیم اساسی همچون سیستم‌های اطلاعاتی ناقص، تئوری مجموعهٔ راف، الگوریتم رقابت استعماری و قواعد فازی، مجموعهٔ دادهٔ این پژوهش، تشریح می‌شود.

---

1. Information Entropy
2. Conditional Entropy
3. Efficiency
4. Positive Approximation

www.SID.ir

Archive of SID

سیستم‌های اطلاعاتی ناقص

یک سیستم اطلاعاتی معمولاً به فرم زیر ارائه می‌شود.

رابطهٔ ۱) $$IS = (U, A, \{V_a\}, f_a)_{a \in A}$$

که در آن، U تعداد متناهی از اشیای دارای مقدار و A تعداد محدودی از ویژگی‌های دارای مقدار است. V_a مجموعهٔ مقادیر ویژگی‌ها و f_a، یک تابع اطلاعاتی از سیستم اطلاعات است که V_a را به U نگاشت می‌کند. اگر در سیستم اطلاعاتی یک شیء x و یک ویژگی به نام a وجود داشته باشد که در تابع f_a (x) مقداری نداشته باشد، به آن سیستم اطلاعاتی ناقص گفته می‌شود. معمولاً مقادیر ناموجود را با * و به فرم ذیل نمایش می‌دهند:

رابطهٔ ۲) $$IIS = (\mathrm{U}, \mathrm{A}, \{V_a\}, f_a)_{a \in A} *\in \bigcup_{a \in A} V_a$$

مثال: جدول شمارهٔ ۱ نمونه‌ای از سیستم اطلاعاتی ناقص را نشان می‌دهد. این سیستم شامل اطلاعات مربوط به چند فرد با ویژگی‌های قد، وزن و سن است. هر یک از مقادیر مربوط به این ویژگی‌ها با مقادیر کیفی مشخص شده‌اند.

جدول ۱. نمونه‌ای از سیستم اطلاعاتی ناقص

افراد	قد	وزن	سن
شمارهٔ ۱	بلند	زیاد	خیلی بالا
شمارهٔ ۲	کوتاه	زیاد	بالا
شمارهٔ ۳	*	کم	پایین
شمارهٔ ۴	بلند	زیاد	خیلی بالا
شمارهٔ ۵	خیلی بلند	*	خیلی بالا
شمارهٔ ۶	بلند	متوسط	*

تئوری مجموعهٔ راف

تئوری مجموعهٔ راف در سال ۱۹۸۰ ارائه شد. این ابزار ریاضی برای بیان و بررسی مسائلی است که در آنها عدم قطعیت و ابهام وجود دارد و معمولاً برای پیدا کردن ناهمگونی در مجموعهٔ داده‌ها به‌کار می‌رود. همان‌طور که در جدول ۱ مشخص است، فرد ۱ و ۴ مقادیر ویژگی مشابهی دارند و با این مقادیر از هم تفکیک نمی‌شوند (پاولاک، ۱۹۹۸؛ کیانا، ژانگ‌ب، سانگب و لیانگ، ۲۰۱۴).

www.SID.ir

Archive of SID

رابطهٔ زیر را رابطهٔ تفکیک‌ناپذیر می‌گویند:

$$I_B = \{(x, y) \in U \times U | \forall Attr \in A, Attr\ (x) = Attr\ (y)\}$$ رابطهٔ ۳)

اگر x و y دو شیء باشند و برای هر ویژگی در مجموعهٔ B، مقدار آن ویژگی در دو شیء یکسان باشد، I_B را رابطهٔ تفکیک ناپذیر B می‌نامند. همچنین برچسب کلاس، ویژگی تصمیم‌گیری تعریف می‌شود و هر به سیستم اطلاعاتی که حاوی این ویژگی باشد، سیستم تصمیم‌گیری می‌گویند.

جدول ۲. نمونه‌ای از سیستم اطلاعاتی کامل

افراد	قد	وزن	سن	درآمد
شمارهٔ ۱	بلند	زیاد	خیلی بالا	کم
شمارهٔ ۲	کوتاه	متوسط	بالا	متوسط
شمارهٔ ۳	بلند	زیاد	خیلی بالا	زیاد
شمارهٔ ۴	خیلی بلند	خیلی کم	خیلی بالا	زیاد
شمارهٔ ۵	بلند	زیاد	پایین	کم
شمارهٔ ۶	متوسط	کم	بالا	خیلی کم

در جدول ۲ ستون درآمد، ویژگی تصمیم‌گیری است. در این جدول افراد ۱ و ۳ ویژگی‌های یکسانی دارند، اما از نظر درآمد وضعیت مشابهی ندارند. به بیان دیگر، ویژگی تصمیم‌گیری آنها متفاوت است. مجموعهٔ توانی ویژگی‌ها عبارت‌اند از {(قد)، (وزن)، (سن)، (قد و وزن)، (قد و سن)، (وزن و سن)، (قد، وزن، سن)}. حال اگر مجموعهٔ (قد و وزن) را در نظر بگیریم، رابطهٔ تفکیک‌ناپذیری آن به‌صورت رابطهٔ ۴ است.

$$U|\ I_{\{\text{وزن و قد}\}}=\{\{1, 3, 5\}, \{2\}, \{4\}, \{6\}\}$$ رابطهٔ ۴)

با توجه به رابطهٔ ۴، مجموعهٔ {۱، ۳، ۵} در رابطهٔ ۴، توسط قد و وزن تفکیک‌ناپذیرند و به یک کلاس هم‌ارزی تعلق دارند. در یک رابطهٔ هم‌ارزی به نام x، کلاس هم‌ارزی مربوط به عضو a شامل اعضایی است که با a ارتباط دارند. نمایش کلاس هم ارزی به‌صورت $[x]_a$ است.

فرض کنید یک سیستم اطلاعاتی S= (U, A) داشته باشیم که در آن U مجموعهٔ اشیا و A ویژگی‌هاست. همچنین X⊆U و B⊆A را داریم و دو مجموعه به شرح زیر تعریف می‌شوند:

$$\underline{B}X = \{x | [x]_B \subseteq X\}$$ رابطهٔ ۵)

www.SID.ir

Archive of SID

رابطهٔ ۶) $\overline{B}X = \{x|[x]_B \cap X \neq \phi\}$

با توجه به تعریف B، مجموعه‌های $\overline{B}X$ و $\underline{B}X$ احتمالاً دارای اشیایی متعلق بـه X هسـتند. محدودهٔ مرزی روی X به‌صورت زیر تعریف می‌شود و حاوی اشیایی است که قطعاً در X نیستند.

رابطهٔ ۷) $BN_B(X) = \overline{B}X - \underline{B}X$

یک مجموعه در صورتی راف است که با توجه به B، مجموعهٔ مرزی آن تهی نباشـد. بـرای مثال، X برابر است با افرادی که درآمد زیادی دارنـد، یعنـی $X = \{۳ , ۴\}$ و مجموعـهٔ B برابـر {قد، وزن و سن} فرض شود، آنگاه داریم:

$\underline{B}X = \{4\}, \overline{B}X = \{1, 3, 4\}, BN_B(X) = \{1, 3\}, U\text{-}BN_B(X) = \{2, 5, 6\}$

همان‌طور که مشخص است، مجموعهٔ مـرزی تهـی نیسـت. افـراد ۱ و ۳ دارای درآمـد بـالا هستند، افراد ۴، ۳ و ۱ احتمالاً درآمد بالایی دارند. افراد ۳ و ۱ را نمی توان بـا B تعیـین تکلیـف کرد. همچنین افراد ۲، ۵ و ۶ قطعاً درآمد بالایی ندارند. میزان دقت مجموعهٔ راف را مـی‌تـوان از رابطهٔ ۸ تعیین کرد.

رابطهٔ ۸) $\alpha_B(X) = \frac{|\underline{B}X|}{|\overline{B}X|}$

درجهٔ تعلق x به X را می توان با تابع تعلق راف بیان کرد. این تابع تعلق، درجهٔ هـم‌پوشـانی مجموعه X و کلاس هم‌ارزی $[x]_B$ را تعیین می‌کند. در مثال بالا کلاس‌های هم‌ارزی به شـرح زیر هستند:

$[1]_B = [3]_B = \{1, 3\}, [2]_B = \{2\}, [4]_B = \{4\}, [5]_B = \{5\}, [6]_B = \{6\}$

رابطهٔ زیر تابع تعلق را مشخص می‌کند:

رابطهٔ ۹) $\mu_X^B: U \to [0, 1] \text{ and } \mu_X^B(x) = \frac{|[x]_B \cap X|}{|[x]_B|}$

در بررسی داده‌ها، یافتن رابطهٔ بین ویژگی‌های شرطی و تصمیم اهمیـت دارد. بـا اسـتفاده از همین وابستگی بین ویژگی‌ها می‌توان آنهایی که اهمیت ندارند را حذف کرد. اگـر Td مجموعـهٔ ویژگی‌های تصمیم‌گیری و Tc مجموعهٔ ویژگی‌های شرطی باشد، وابسـتگی بـین آنهـا بـه ایـن شکل بیان می‌شود $Tc \leq Td$؛ به این معنا که تمام مقادیر تصمیم‌گیری از مقادیر شرطی به‌دسـت می‌آیند و البته حالت جزئی هم می‌تواند داشته باشد.

www.SID.ir

Archive of SID

تعریف رسمی برای این خاصیت بدین صورت است که اگر C و D زیر مجموعه‌های A باشند، به‌طوری که اشتراک C و D تهی و اجتماع آنها A باشد، می‌گوییم D وابسته به C است. و به شکل $C \Rightarrow_k D$ نمایش داده می‌شود، اگر رابطهٔ ۱۰ برقرار باشد.

$$K = \gamma(C, D) = \sum_{X \in U|D} \frac{|\underline{C}X|}{|U|} \qquad \text{رابطهٔ ۱۰)}$$

اگر k برابر ۱ باشد، یعنی D کاملاً به C وابسته است. در مثال قبلی ویژگی تصمیم‌گیری درآمد با درجهٔ $\frac{۲}{۳}$ به مجموعهٔ ویژگی‌های {قد، وزن و سن} وابسته است.

با توجه به تشریح تئوری مجموعهٔ راف، کاربردهای مختلف آن قابل درک خواهد بود که بحث کاهش ویژگی یکی از کاربردهای مهم آن است.

الگوریتم رقابت استعماری

الگوریتم رقابت استعماری در سال ۲۰۰۷، به‌عنوان روش جست‌وجوی فرامکاشفه‌ای مطرح شد که از پدیدهٔ اجتماعی ـ انسانی الهام می‌گیرد (آتشپز و لوکاس، ۲۰۰۷). این الگوریتم با تعدادی جمعیت اولیه آغاز می‌شود که هر عنصر جمعیت یک کشور نام دارد و کشورها به دو گروه استعمارگر و مستعمره دسته‌بندی می‌شوند. هر استعمارگر بسته به قدرت خود تعدادی از کشورهای مستعمره را به سلطه درمی‌آورد و آنها را کنترل می‌کند. سیاست جذب، رقابت استعماری و انقلاب، هستهٔ اصلی این الگوریتم را تشکیل می‌دهند (صنیعی آباده و جبل، ۱۳۹۲). رویهٔ اجرای این الگوریتم به‌صورت زیر است (حاجی حسنی، ارمغانی و مارتو، ۲۰۱۵):

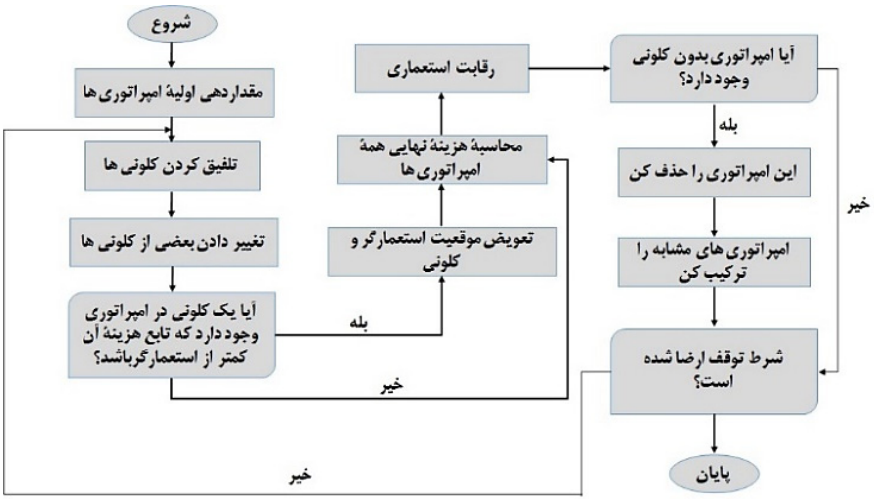
- شکل‌دهی امپراتوری‌های اولیه: تعدادی از بهترین عناصر جمعیت به‌عنوان استعمارگر انتخاب می‌شوند و باقی‌ماندهٔ جمعیت نیز مستعمره در نظر گرفته می‌شوند. استعمارگران بسته به قدرتشان، این مستعمرات را با روند خاصی به سمت خود می‌کشند. در بهینه‌سازی، هدف، یافتن یک جواب بهینه بر حسب متغیرهای مسئله است. ما آرایه‌ای از متغیرهای مسئله را که باید بهینه شوند، ایجاد می‌کنیم و آن را یک کشور می‌نامیم. در مسئلهٔ بهینه‌سازی N_{var} بعدی، یک کشور، آرایه‌ای به طول ۱ * N_{var} است. این آرایه به این صورت تعریف می‌شود: کشور = $[p_۱, p_۲, \ldots, p_{Nvar}]$. مقادیر متغیرها در یک کشور، به‌صورت اعداد اعشاری نمایش داده می‌شوند. از دیدگاه تاریخی و فرهنگی، اجزای تشکیل‌دهندهٔ یک کشور را می‌توان ویژگی‌های اجتماعی ـ سیاسی آن کشور، همچون فرهنگ، زبان، ساختار اقتصادی و سایر ویژگی‌ها در نظر گرفت.

www.SID.ir

Archive of SID

- سیاست جذب (همگون‌سازی): این سیاست با هدف تحلیل فرهنگ و ساختار اجتماعی مستعمرات در فرهنگ حکومت مرکزی اجرا می‌شد. کشورهای استعمارگر برای افزایش نفوذ خود، شروع به ایجاد زیرساخت‌های حمل و نقل، تأسیس دانشگاه و غیره می‌کردند. در واقع، حکومت مرکزی با اعمال سیاست جذب تلاش می‌کرد کشور مستعمره را در راستای ابعاد مختلف اجتماعی ـ سیاسی به خود نزدیک کند.
- انقلاب: روند انقلاب تغییرات ناگهانی را در ویژگی‌های اجتماعی ـ سیاسی یک کشور ایجاد می‌کند. در الگوریتم رقابت استعماری، انقلاب با جابه‌جایی تصادفی یک کشور مستعمره به موقعیت تصادفی جدید، مدل‌سازی می‌شود.
- جابه‌جایی موقعیت مستعمره و استعمارگر: در حین حرکت مستعمرات به سمت کشور استعمارگر، ممکن است بعضی از این مستعمرات به موقعیتی بهتر از استعمارگر برسند. در این حالت، کشور استعمارگر و کشور مستعمره، جای خود را با یکدیگر عوض می‌کنند.
- رقابت استعماری: قدرت امپراتوری، به‌صورت قدرت کشور استعمارگر به اضافهٔ درصدی از قدرت کل مستعمرات آن تعریف می‌شود. هر امپراتوری‌که نتواند بر قدرت خود بیفزاید و قدرت رقابت خود را از دست بدهد، در جریان رقابت‌های استعماری، به‌تدریج سقوط می‌کند و مستعمراتش به‌دست امپراتوری‌های قوی‌تر می‌افتد.

در شکل ۱ نمودار الگوریتم رقابت استعماری مشاهده می‌شود.



شکل ۱. فلوچارت الگوریتم رقابت استعماری

www.SID.ir

Archive of SID

در ادامه به معرفی پارامترهای الگوریتم رقابت استعماری پرداخته می‌شود.

جدول ۳. معرفی پارامترهای الگوریتم رقابت استعماری

نام متغیرها	توضیح
Nvar	تعداد متغیرها
Npop	تعداد اعضای جمعیت
Nimp	تعداد استعمارگران
Maxdecades	تعداد تکرار
Beta	ضریب جابه‌جایی
Prevolve	ضریب انقلاب (شبیه جهش در ژنتیک)
Zeta	ضریب میانگین قدرت کلونی‌های یک استعمارگر

در هر بار تکرار الگوریتم، مقادیر اولیهٔ مختلفی را برای پارامترهای آن انتخاب کردیم که در نهایت، بهترین مقادیر آنها به‌صورت تجربی به این ترتیب مشخص شد: تعداد تکرار برابر با ۱۰۰ تا ۲۰۰، تعداد اعضای جمعیت ۵۰ تا ۱۰۰، تعداد استعمارگران بسته به تعداد اعضای جمعیت بین ۵ تا ۱۰، با توجه به پیشنهاد ارائه‌دهندگان الگوریتم ضریب جابه‌جایی مقدار ۰/۲، ضریب انقلاب ۰/۱ و ضریب میانگین قدرت کلونی‌ها بین ۰/۰۵ تا ۰/۰۰۵.

قوانین فازی تعریف‌شده

استفاده از قواعد فازی در اجرای الگوریتم باعث شد بتوانیم پارامترهای کلیدی الگوریتم رقابت استعماری را بهتر کنترل کنیم. مجموعهٔ قواعد فازی برای این برنامه به‌صورت زیر تعریف شد:

1. If (UN is not L) and (I is not L) then (Prevolve is M) (zeta is L) (Beta is H)
2. If (nimp is M) and (I is M) then (Prevolve is L) (zeta is M) (Beta is M)
3. If (I is H) then (Prevolve is M) (Beta is L)
4. If (I is L) then (Prevolve is L) (Beta is H)
5. If (I is M) then (Prevolve is M) (Beta is M)
6. If (nimp is not H) and (I is H) then (Prevolve is L) (zeta is M) (Beta is L)

ورودی‌ها و خروجی‌های مجموعهٔ قواعد در جدول ۴ مشخص شده‌اند.

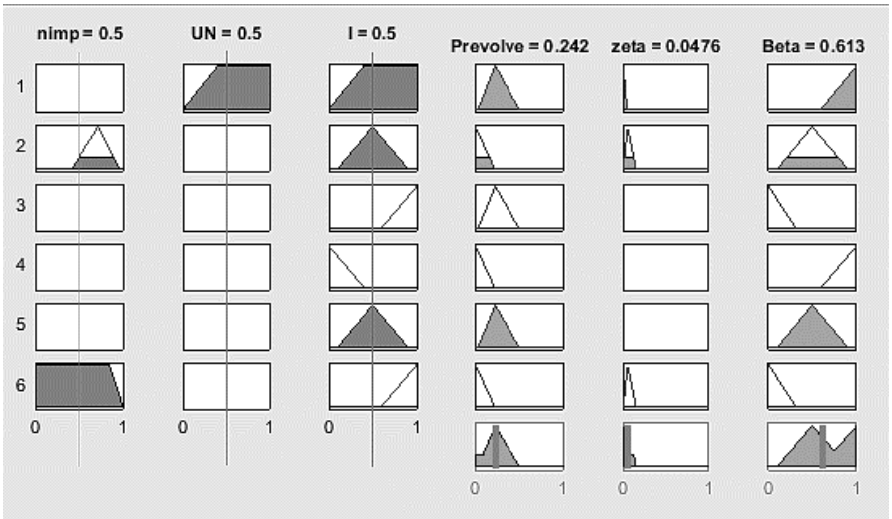
www.SID.ir

Archive of SID

جدول ٤. توضیح ورودی‌ها و خروجی‌های قوانین فازی تعریف‌شده

ورودی‌ها	توضیح ورودی‌ها	خروجی‌ها	توضیح خروجی‌ها
I	تعداد تکرارها	Prevolve	ضریب انقلاب (جهش)
UN	تعداد اجراهایی که جواب تغییر نمی‌کند	zeta	ضریب میانگین قدرت کلونی‌های یک استعمارگر
Nimp	تعداد استعمارگران	Beta	ضریب جابه‌جایی

همچنین در مشخص‌کردن مقدار ورودی‌ها و خروجی‌ها، L به معنای کم[1]، M به معنای متوسط[2] و H به معنای زیاد[3] است. در شکل ۲، نمودارهای مربوط به قوانین نشان داده شده است.



شکل ۲. نمودار ورودی‌ها و خروجی‌های قوانین فازی

با توجه به بررسی شکل ۲، هنگامی که ورودی‌های قواعد فازی را متوسط (M) فرض کنیم، خروجی‌های آن به‌صورت Prevolve برابر ۰/۲۴۲، zeta برابر ۰/۰۴۷۷۶ و Beta برابر ۰/۶۱۳ است.

1. Low
2. Medium
3. High

www.SID.ir

Archive of SID

مجموعه دادهٔ در دست بررسی

همان‌طور که در قسمت مقدمهٔ مقاله اشاره کردیم، کار روی داده‌های ناقص انجام می‌شود. داده‌های ناقص به جدول‌هایی از داده‌ها گفته می‌شود که برخی درایه‌های صفات آن مقداری ندارند. در روند انجام این پروژه، هیچ تغییری روی داده‌های ناقص اعمال نکردیم و آن را سیستم اطلاعاتی کامل فرض کردیم. مجموعهٔ دادهٔ در دست بررسی در جدول ۵ معرفی شده است. (کیان، لیانگ، پدریک و دانگ، ۲۰۱۱؛ یو. سی. آی، ۲۰۱۶).

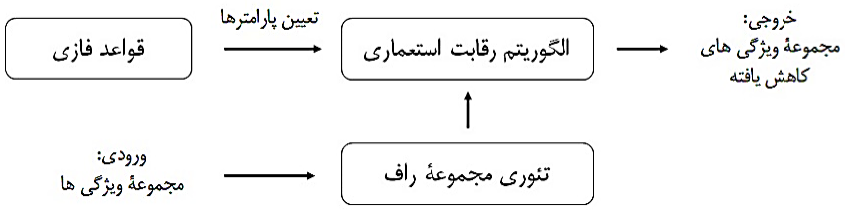
جدول ۵. معرفی مجموع دادهٔ مورد بررسی

ردیف	نام مجموعه داده‌ها	تعداد سطرها	تعداد ستون‌ها	شمارهٔ ستون تصمیم
۱	Lungcancer	۳۲	۵۷	۵۷
۲	Dermatology	۳۶۶	۳۵	۳۵
۳	Wisconsin_Breast_Cancer	۶۹۹	۱۱	۱۱
۴	Hepatitis	۱۵۵	۲۰	۲۰
۵	Flag	۱۹۴	۳۰	۳

در مجموعهٔ دادهٔ معرفی‌شده، دادهٔ شمارهٔ ۱ مربوط به سرطان ریه، دادهٔ شمارهٔ ۲ مربوط به مشخصات پوست، دادهٔ شمارهٔ ۳ مربوط به سرطان سینه، دادهٔ شمارهٔ ۴ مربوط به هپاتیت و دادهٔ شمارهٔ ۵ مربوط به مشخصات پرچم‌هاست.

یافته‌های پژوهش

شکل ۳ چارچوب ساده‌ای از اجرای این پژوهش را نشان می‌دهد که نقش قواعد فازی در تعیین پارامترهای الگوریتم و نحوهٔ ارتباط آن با تئوری مجموعهٔ راف را تعیین می‌کند. همچنین در این چارچوب، ورودی و خروجی مشخص شده است.



شکل ۳. چارچوب پژوهش

www.SID.ir

Archive of SID

دو نسخهٔ الگوریتم رقابت استعماری و رقابت استعماری فازی طراحی‌شده، به دفعات زیاد روی مجموعهٔ داده‌ها اجرا شد. بهترین جواب‌های به‌دست‌آمده از اجرای دو نسخهٔ الگوریتم رقابت استعماری با بهترین جواب‌های به‌دست‌آمده تاکنون در جدول ۶ آمده است (کیان، لیانگ، پدریک و دانگ، ۲۰۱۱؛ کی، ژو و گیانگ، ۲۰۰۸). همان‌طور که در جدول مشاهده می‌شود، بهترین جواب‌های به‌دست‌آمده از هر دو نسخهٔ الگوریتم برای مجموعهٔ دادهٔ ۲ تا ۵، بهتر یا برابر با بهترین جواب‌های به‌دست‌آمده تاکنون است. از سویی، نسخهٔ فازی الگوریتم نتایج بهینه‌تری را ارائه کرده است. نتیجهٔ اجرای نسخهٔ فازی الگوریتم روی مجموعه دادهٔ ۱ بهتر از نسخهٔ کلاسیک آن است، اما بهینه‌تر از بهترین جواب به‌دست‌آمده تاکنون نبود.

جدول ٦. بهترین جواب‌های به‌دست‌آمده از دو نسخهٔ الگوریتم با بهترین جواب‌های به‌دست‌آمده تاکنون

مجموعهٔ داده‌ها	الگوریتم رقابت استعماری فازی		الگوریتم رقابت استعماری		بهترین جواب به‌دست‌آمده تاکنون
	بهترین جواب	شمارهٔ ستون‌ها	بهترین جواب	شمارهٔ ستون‌ها	
۱	۵	۳، ۱۱، ۲۸، ۳۶، ۴۱	۶	۳، ۷، ۱۱، ۲۸، ۳۵، ۳۹	۴
۲	۸	۲، ۳، ۴، ۹، ۱۶، ۱۷، ۱۹، ۲۸	۱۰	۲، ۳، ۴، ۵، ۸، ۹، ۱۶، ۱۸، ۲۲، ۲۸	۱۰
۳	۲	۱، ۶	۳	۱، ۳، ۶	۴
۴	۳	۲، ۶، ۱۷	۳	۲، ۸، ۱۶	۵
۵	۱	۴	۳	۴، ۹، ۱۱	۲

با مقایسهٔ دو نسخهٔ الگوریتم رقابت استعماری روی مجموعهٔ دادهٔ مشخص‌شده، درمی‌یابیم که نسخهٔ فازی، راه‌حل‌های بهینه‌تری را ارائه کرده است. بهترین جواب‌های به‌دست‌آمده از الگوریتم رقابت استعماری فازی برای پنج دادهٔ تعیین‌شده، به‌ترتیب {۵، ۸، ۲، ۳، ۱} و بدترین پاسخ‌های به‌دست‌آمده شامل {۱۰، ۱۱، ۴، ۵، ۶} است، اما بهترین پاسخ‌های به‌دست‌آمده از اجرای الگوریتم رقابت استعماری برای مجموعهٔ داده‌ها به‌ترتیب {۶، ۱۰، ۳، ۳، ۳} و بدترین جواب‌ها به‌صورت {۹، ۱۴، ۴، ۵، ۵} هستند. در جدول ۷ مقایسهٔ ریز نتایج دو نسخهٔ الگوریتم درج شده است.

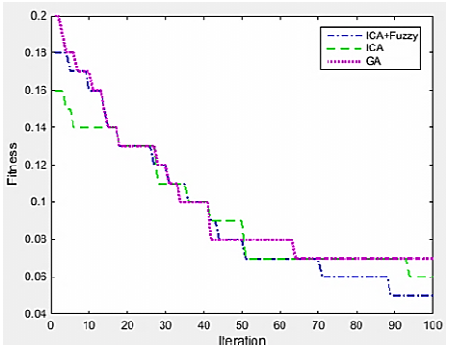
www.SID.ir

Archive of SID

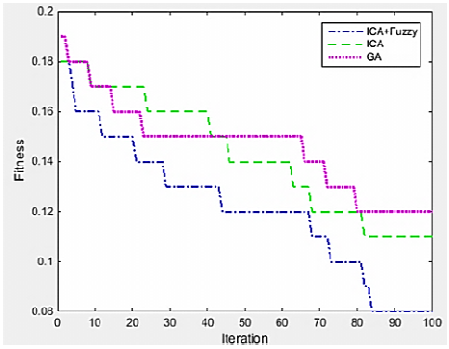
جدول ۷. مقایسهٔ بهترین و بدترین جواب‌های به‌دست آمده از دو نسخهٔ الگوریتم

مجموعهٔ داده	شمارهٔ ستون‌ها	بهترین جواب	بدترین جواب	میانگین بهترین جواب‌ها	واریانس بهترین جواب‌ها
Fuzzy-ICA					
۱	۳، ۱۱، ۲۸، ۳۶، ۴۱	۵	۱۰	۷	۲/۶۶۶۷
۲	۲، ۳، ۴، ۹، ۱۶، ۱۷، ۱۹، ۲۸	۸	۱۱	۹/۹	۰/۷۶۶۷
۳	۱، ۶	۲	۴	۲/۶	۰/۴۸۸۹
۴	۲، ۸، ۱۷	۳	۵	۳/۶	۴/۴۸۸۹
۵	۴	۱	۶	۲/۷	۲/۸۴۴۴
ICA					
۱	۳، ۷، ۱۱، ۲۸، ۳۵، ۳۹	۶	۹	۶/۹	۰/۹۸۸۹
۲	۲، ۳، ۴، ۵، ۸، ۹، ۱۶، ۱۸، ۲۲، ۲۸	۱۰	۱۴	۱۱/۲	۱/۹۵۵۶
۳	۱، ۳، ۶	۳	۴	۳/۲	۰/۲۳۳۳
۴	۲، ۸، ۱۶	۳	۵	۳/۹	۰/۵۴۴۴
۵	۴، ۹، ۱۱	۳	۵	۴	۰/۴۴۴۴

در بررسی نتایج این مقاله، دو نسخهٔ الگوریتم رقابت استعماری طراحی‌شده با الگوریتم ژنتیک مقایسه شدند. با توجه به بهترین نتایج به‌دست‌آمده، مشاهده می‌شود که نسخهٔ فازی الگوریتم رقابت استعماری از الگوریتم ژنتیک جواب‌های بهینه‌تری را ارائه می‌کند. در شکل‌های ۴ تا ۸ نمودارهای همگرایی این نتایج مشاهده می‌شود.



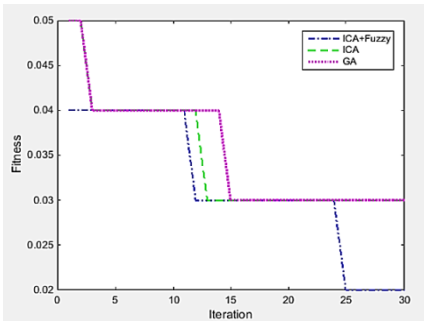
شکل ٤. نمودار همگرایی برای دادهٔ ۱



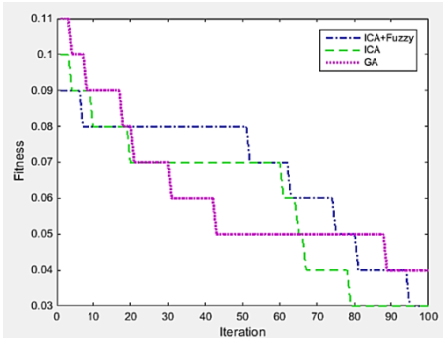
شکل ۵. نمودار همگرایی برای دادهٔ ۲

www.SID.ir

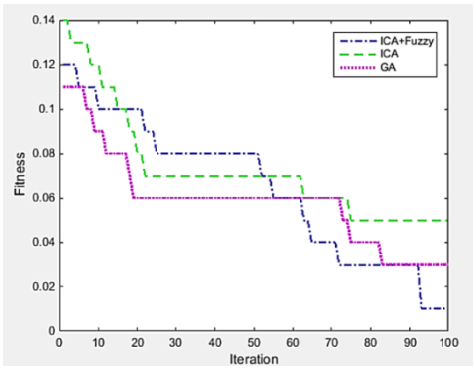
Archive of SID



شکل ٦. نمودار همگرایی برای دادهٔ ۳



شکل ۷. نمودار همگرایی برای دادهٔ ٤



شکل ۸. نمودار همگرایی برای دادهٔ ٥

نتیجه‌گیری و پیشنهادها

امروزه در دنیای واقعی، مواجهه با حجم شایان توجهی از پایگاه‌های داده که برخی درایه‌های صفت آنها مقداری ندارند، اجتناب‌ناپذیر است. یکی از دلایل ناقص‌بودن داده‌ها، خطا در جمع‌آوری آنهاست. به بیان دیگر، وجود سیستم‌های اطلاعاتی ناقص حجیم در شرکت‌ها و مؤسسه‌های گوناگون، امری طبیعی است و استخراج دانش از دل این داده‌های ناقص را از دو جهت دچار چالش می‌کند. از یک سو، وجود داده‌های ناقص در پایگاه‌های داده دسترسی به بخشی از اطلاعات این پایگاه‌ها را امکان‌ناپذیر می‌کند و از سوی دیگر، حجم گستردهٔ داده‌های موجود سبب افزایش زمان شناسایی و استخراج دانش از درون این پایگاه‌های داده می‌شود. از این رو، در پژوهش حاضر به‌منظور مدیریت چالش اول، تئوری توسعه‌یافتهٔ مجموعهٔ راف برای محاسبهٔ کیفیت زیرمجموعهٔ انتخاب‌شده از ویژگی‌های سیستم اطلاعاتی ناقص درنظر گرفته شد.

www.SID.ir

Archive of SID

همچنین برای حل چالش دوم، الگوریتم اصلاح‌شدهٔ رقابت استعماری توسط منطق فازی طراحی شد. طراحی شش قانون فازی براساس نتایج به‌دست‌آمده از پیاده‌سازی الگوریتم کلاسیک رقابت استعماری روی داده‌های نمونهٔ کوچک نیز سبب شد عملکرد این الگوریتم در کنترل پارامترهای خود پویا و هوشمندانه باشد. تجمیع نوآوری‌های خلاقانهٔ اشاره‌شده به‌منظور رفع این چالش‌ها موجب شد که جواب‌های به‌دست‌آمده از پیاده‌سازی این ایده‌ها، کیفیت مناسب و درخور تأملی نسبت به سایر روش‌های موجود داشته باشد.

با بررسی نتایج به‌دست‌آمده (جدول‌های ۶ و ۷)، بهبود عملکرد الگوریتم اصلاح‌شدهٔ رقابت استعماری فازی با نسخهٔ کلاسیک آن، به‌طور آشکاری شایان توجه است. این میزان بهبود برای داده‌های نمونهٔ شمارهٔ ۲ تا ۵، حتی از بهترین جواب‌های به‌دست‌آمده تاکنون برای این داده‌ها نیز پیشی گرفته و کران جدیدی برای کاهش ویژگی این داده‌ها معرفی می‌کند. همچنین جزئیات عملکرد سه الگوریتم ژنتیک، رقابت استعماری فازی و رقابت استعماری کلاسیک در نمودارهای همگرایی این الگوریتم‌ها (شکل‌های ۴ تا ۸) برای هر یک از داده‌های نمونه به‌صورت مجزا با یکدیگر مقایسه شدند. توجه به این نمودارهای همگرایی، تأیید دیگری برای اثبات عملکرد مناسب و درخور تأمل ایده‌های خلاقانهٔ طرح‌شده نسبت به سایر روش‌های موجود است.

نتایج مناسب به‌دست‌آمده از پیاده‌سازی ایده‌های طرح‌شده، راه را برای توسعهٔ این الگوریتم‌ها در مواجهه با پایگاه داده‌های بسیار بزرگ، شامل داده‌های ناقص که ساختار معماری متفاوتی دارند، باز می‌کند. از این رو در فاز بعدی این پژوهش، ابتدا می‌توان پیاده‌سازی و افزایش سطح هوشمندسازی و انعطاف‌پذیری سایر الگوریتم‌های فوق‌ابتکاری جدید را در کانون توجه قرار داد؛ سپس با در نظر گرفتن معماری متفاوت داده‌های کلان، ایده‌های پیشنهادی و آزمایش‌شده را برای کاهش ابعاد و حجم داده‌های کلان با اطلاعات ناقص به‌کار گرفت.

فهرست منابع

صفوی، ع. ا.؛ پورجعفریان، ن.؛ صفوی، ع. (۱۳۹۳)، *بهینه‌سازی بر پایهٔ الگوریتم‌های فرابتکاری*، تهران: موسسهٔ انتشاراتی پژوهشگران نشر دانشگاهی.

صنیعی آباده، م.؛ جبل عاملیان، ز. (۱۳۹۲). *الگوریتم های تکاملی و محاسبات زیستی*، تهران: نیاز دانش.

Atashpaz-Gargari, E. & Lucas, C. (2007, September). Imperialist competitive algorithm: an algorithm for optimization inspired by imperialistic competition. *Evolutionary Computation* (pp. 4661-4667). Singapore: IEEE.

Chmielewski, M., Grzymala-Busse, J., Peterson, N. & Than, S. (1993). The rule induction system LERS – a version for personal computer. *Foundations of Computing and Decision Science,* 18(3), 181-212.

www.SID.ir

Archive of SID

Dash, M. & Liu, H. (2003). Consistency-based search in feature selection. *Artificial Intelligence,* 151(1), 155-176.

Guyon, I. & Elisseeff, A. (2003). An introduction to variable and feature selection. *The Journal of Machine Learning Research,* 3(1), 1157-1182.

Hajihassani, M., Armaghani, D. J. & Marto, A. (2015). Ground vibration prediction in quarry blasting through an artificial neural network optimized by imperialist competitive algorithm. *Bulletin of Engineering Geology and the Environment,* 74(3), 873-886.

Han, J. & Kamber, M. (2006). *Data Mining: Concepts and Techniques*. San Francisco: Diane Cerra.

Hu, Q., Xie, Z. & Yu, D. (2007). Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation. *Pattern Recognition, 40*(12), 3509–3521.

Hu, Q., Yu, D. & Xie, Z. (2006). Information-preserving hybrid data reduction based on fuzzy-rough techniques. *Pattern Recognition Letters,* 27(5), 414–423.

Hu, X. & Cercone, N. (1995). Learning in relational databases: a rough set approach. *Computational Intelligence,* 11(2), 323–338.

Ke, L., Zuren, F. & Zhigang, R. (2008). An efficient ant colony optimization approach to attribute reduction in rough set theory. *Pattern Recognition Letters,* 29(9), 1351-1357.

Kira, K. & Rendell, L. (1992). The feature selection problem: traditional methods and a new algorithm. *AAAI, 2*(1), 129–134.

Kohavi, R. & John, G. (1997). Wrappers for feature subset selection. *Artificial Intelligence, 97*(2), 273-324.

Kryszkiewicz, M. (1998). Rough setapproach to incomplete information systems. *Information Sciences,* 112(1), 39-49.

Lee, C. & Lee, G. (2006). Information gain and divergence-based feature selection for machine learning-based text categorization. *Information Processing and Management, 42*(1), 155-165.

Li, D., Zhang, B. & Leung, Y. (2004). On knowledge reduction in inconsistent decision information systems. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems,* 12(5), 651-672.

Liang, J. & Xu, Z. (2002). The algorithm on knowledge reduction in incomplete information systems. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems,* 10(1), 95-103.

www.SID.ir

Archive of SID

Liang, J., Chin, K., Dang, C. & YamRichid, C. (2002). A new method for measuring uncertainty and fuzziness in rough set theory. *International Journal of General Systems, 31*(4), 331-342.

Liang, J., Shi, Z., Li, D. & Wierman, M. (2006). Information entropy, rough entropy and knowledge granulation in incomplete information systems. *International Journal of General Systems, 35*(6), 641-654.

Lotfizadeh, L. A. (1992). Fuzzy logic, neural networks and soft computing. *Communications of the ACM, 37*(3), 77-84.

Meng, Z. & Shi, Z. (2009). A fast approach to attribute reduction in incomplete decision systems with tolerance relation-based rough sets. *Information Sciences,* 179(16), 2774–2793.

Mi, J., Wu, W. & Zhang, W. (2003). Comparative studies of knowledge reductions in inconsistent systems. *Fuzzy systems and mathematics,* 17(3), 54-60.

Modrzejewski, M. (1993, April). Feature selection using rough set theory. *Machine Learning* (pp. 213–226). Berlin Heidelberg: Springer.

Orlowska, E. & Pawlak, Z. (1984). Representation of nondeterministic information. *Theoretical Computer Science, 29*(1), 27-39.

Pawlak, Z. (1998). Rough set theory and its applications to data analysis. *Cybernetics and Systems, 29*(7), 662-668.

Pawlak, Z. & Skowron, A. (2007). Rudiments of rough sets. *Information Sciences, 177*(1), 3-27.

Qian, Y. & Liang, J. (2008). Combination entropy and combination granulation in rough set theory. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 16*(2), 179-193.

Qian, Y., Liang, J. & Dang, C. (2008). Consistency measure, inclusion degree and fuzzy measure in decision tables. *Fuzzy Sets and Systems, 159*(18), 2353–2377.

Qian, Y., Liang, J. & Dang, C. (2008). Interval ordered information systems. *Computers & Mathematics with Applications, 56*(8), 1994–2009.

Qian, Y., Liang, J. & Wang, F. (2009). A new method for measuring the uncertainty in incomplete information systems. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 17*(6), 855–880.

Qian, Y., Liang, J., Pedrycz, W. & Dang, C. (2011). An efficient accelerator for attribute reduction from incomplete data in rough set framework. *Pattern Recognition, 44*(8), 1658–1670.

www.SID.ir

Archive of SID

Qiana, Y., Zhangb, H., Sangb, Y. & Lianga, J. (2014). Multigranulation decision-theoretic rough sets. *International Journal of Approximate Reasoning, 55*(1), 225-237.

Quinlan, J. (1986). Induction of decision trees. *Machine Learning, 1*(1), 81-106.

Safavi, A., Pourjafarian, N. & Safavi, A. (2014). *Optimization Based On Meta Hueristic Algorithms*. Shiraz: Researchers academic publishing. (*in Persian*)

Sanei Abadeh, M., & Jebel, Z. (2013). *Evolutionary Algorithms and Biological Computing.* Tehran: Niaz Danesh. (*in Persian*)

Shao, M. & Zhang, W. (2005). Dominance relation and rules in an incomplete ordered information system. *International journal of intelligent systems, 20*(1), 13-27.

Skowron, A. (1995). Extracting laws from decision tables: a rough setapproach. *Computational Intelligence, 11*(2), 371-388.

Slezak, D. (2002). Approximate entropy reducts. *Fundamenta informaticae, 53*(3), 365–390.

UCI. (2016). Retrieved from https://archive.ics.uci.edu/ml/datasets.html

Wang, G., Yu, H. & Yang, D. (2002). Decision table reduction based on conditional information entropy. *Chinese Journal of Computer, 25*(7), 759–766.

Wang, G., Zhao, J. & An, J. (2005). Acomparative study of algebra viewpoint and nattribute reduction. *Foundamenta Informaticae, 68*(3), 289–301.

Wu, S., Li, M., Huang, W., & Liu, S. (2004). An Improved Heuristic Algorithm of Attribute Reduction in Rough Set. *Journal of Systems Science & Information, 2*(3), 557-562.

Wu, W., Zhang, M., Li, H. & Mi, J. (2005). Knowledge reduction in random information systems via Dempster–Shafer theory of evidence. *Information Sciences, 174*(3), 143–164.

Yang, C. & Shu, L. (2006). Attribute reduction algorithm of incomplete decision table based on tolerance relation. *Computer Technology and Development, 16*(9).

Yu, J. (2005). General C-means clustering model. *Pattern Analysis and Machine Intelligence, IEEE Transactions, 27*(8), 1197-1211.

www.SID.ir