



سجش از دور

GIS ایران



سنجش از دور و GIS ایران
Iranian Remote Sensing & GIS

سال نهم، شماره دوم، تابستان ۱۳۹۶
Vol.9, No. 2, Summer 2017

۶۱-۷۸

ارزیابی قابلیت‌های ترکیب الگوریتم‌های مجموعه‌های راف و ژنتیک برای داده‌کاوی و استخراج قوانین مرتبط با آب مصرفی در شهر تهران

وحید احمدی^۱، عباس علیمحمدی^{۲*}، جلال کرمی^۳

۱. کارشناس ارشد GIS و سنجش از دور، دانشگاه تربیت مدرس
۲. دانشیار گروه مهندسی GIS، دانشگاه صنعتی خواجه نصیرالدین طوسی
۳. استادیار گروه سنجش از دور و GIS، دانشگاه تربیت مدرس

تاریخ پذیرش مقاله: ۱۳۹۶/۶/۲۹

تاریخ دریافت مقاله: ۱۳۹۵/۵/۱

چکیده

مدیریت و برنامه‌ریزی آب شهری، به‌ویژه در کلان‌شهرها، اهمیت بسیار زیادی دارد. توسعه مناطق شهری، تبدیل شهرها به کلان‌شهر و افزایش پیچیدگی عوامل تأثیرگذار در مصرف آب در شهرها سبب دشواری مدیریت مصرف، تأمین و توزیع آب شده است. بنابراین، استخراج قوانین نقش مهمی در کشف الگوهای حاکم بر مجموعه داده و کاهش پیچیدگی‌ها دارد. اصل نظریه مجموعه‌های راف، که پائولاک در دهه ۸۰ مطرح کرد، روشی توانمند و انعطاف‌پذیر در پردازش داده‌های دارای عدم قطعیت شمرده می‌شود و در این تحقیق، به‌منظور استخراج قوانین حاکم بر مصرف آب، به‌کار رفته است. در این تحقیق، از روش ترکیب الگوریتم‌های مجموعه‌های راف و ژنتیک از روش‌های داده‌کاوی، برای بهبود استخراج قوانین و طبقه‌بندی داده‌های آب مصرفی، با کاربری مسکونی در شهر تهران به‌منزله منطقه مورد مطالعه استفاده شده است. داده‌های مورد استفاده در این تحقیق شامل داده‌های اجتماعی-اقتصادی، محیطی، اقلیمی و فنی-مدیریتی می‌شوند. این داده‌ها به دو زیرمجموعه، شامل ۶۰٪ با هدف آموزش و ۴۰٪ به‌فصد ارزیابی نتایج، تقسیم شدند. نتایج نشان می‌دهند که تلفیق الگوریتم‌های ژنتیک و مجموعه‌های راف کارایی بالاتری برای استخراج مؤثر قوانین از داده‌های مصرف آب شهر تهران را دارند. دقت طبقه‌بندی مجموعه داده آزمون، از طریق قوانین استخراج‌شده از مجموعه‌های راف، ۷۷٪ بود. پس از بهینه‌سازی قوانین با استفاده از الگوریتم ژنتیک در مجموعه‌های راف، دقت طبقه‌بندی در نسل ششم، با سرعت همگرایی متوسط، به ۸۸٪ و در نسل دهم، به ۹۲٪ افزایش یافت. براساس قوانین استخراج‌شده، عوامل مؤثر در مصرف سالیانه آب به‌ترتیب میزان تأثیرگذاری، جمعیت ساکن، آب‌بها، تراکم جمعیت در واحد سطح، بعد خانوار، موقعیت مکانی (عرض جغرافیایی)، تحصیلات ساکنان و سرانه فضای سبز به‌شمار می‌روند.

کلیدواژه‌ها: استخراج قوانین، مجموعه راف، الگوریتم ژنتیک، داده‌کاوی مکانی.

* نویسنده‌دار مکاتبات: تهران، خیابان ولیعصر، دانشگاه صنعتی خواجه نصیرالدین طوسی، دانشکده مهندسی ژئودزی و ژئوماتیک، کدپستی ۱۹۹۶۷۱۵۴۳۳، تلفن:

۰۹۱۲۱۱۴۸۲۱۷۴

۱- مقدمه

برنامه‌ریزی و مدیریت سیستم آب شهری موضوعی پیچیده و دشوار است. این سیستم متأثر از عوامل ناشناخته متعدد و متفاوت اجتماعی-اقتصادی، محیطی، مدیریتی-فنی و اقلیمی است (Mortazavi, 2009). توزیع ناهمگن عوامل تأثیرگذار در مصرف آب در سطح شهر تهران، وسعت مناطق شهری، تراکم بالای جمعیت، جابه‌جایی‌های سالیانه منازل مسکونی در مناطق این کلان‌شهر و همچنین بحران آب آشامیدنی، که شهر تهران با آن مواجه است (بهران، ۱۳۸۷)، پیچیدگی موضوع را دوچندان کرده است. این پیچیدگی ضرورت توسعه زیرساخت‌ها و سیستم‌های پردازش قدرتمند را، به‌منظور ساده‌سازی سیستم و تجزیه آن به اجزای گوناگون و قابل بررسی، بیش از پیش نشان می‌دهد.

روش‌های کشف دانش یا داده‌کاوی دارای قابلیت‌های پردازش، تحلیل داده‌ها و استخراج قواعد مفید از داده‌های بسیار حجیم و گسترده است (Cabena et al., 1998). تحلیل چنین داده‌هایی به‌سبب ابعاد، پیچیدگی، مقیاس، الگو و روابط پنهان بین داده‌ها دارای دشواری است؛ بنابراین، نیاز فزاینده برای سازماندهی و استخراج اطلاعات و کشف دانش و روابط بین داده‌های حجیم ضروری به‌نظر می‌رسد (Larose, 2005).

ویژگی‌های ذکرشده درمورد آب شهر تهران، به‌ویژه نقل و انتقال‌های سالیانه به منازل مسکونی، به‌دلیل استیجاری‌بودنشان در این کلان‌شهر، منجر به عدم قطعیت معتناهایی در داده‌های موجود و افزایش پیچیدگی تأثیر عوامل هیدرولوژیک، اقلیمی و مدیریتی (Mortazavi, 2009) می‌شود. بنابراین، لازم است از روش‌های مناسب کار، با داده‌های دارای عدم قطعیت، استفاده شود. با ظهور تئوری مجموعه‌های راف^۱ در دهه ۸۰ از سوی پائولاک^۲، روشی مؤثر و امیدبخش برای کار با این داده‌ها ایجاد شد (YinZhai, 2006). استفاده از روش‌های بهینه‌سازی، مانند الگوریتم ژنتیک، در مجموعه‌های راف در افزایش دقت قوانین استخراج‌شده از این مجموعه‌ها مؤثر است؛ در نتیجه، کشف عوامل و نیز میزان تأثیرگذاری آن‌ها در مصرف آب شهری،

به‌کمک قوانین استخراجی، نقش مهمی در برنامه‌ریزی و مدیریت آب شهری در نواحی گوناگون ایفا می‌کند. تا کنون تحقیقات بسیاری درباره کاربرد الگوریتم‌های ژنتیک و مجموعه‌های راف در حل مسائل متفاوت انجام شده اما تحقیقات محدودی با رویکرد تلفیقی صورت گرفته است. لوئیس سانتانا^۳ (۲۰۱۰) به ترکیب الگوریتمی تکاملی و روش‌های جست‌وجوی محلی مبتنی بر تئوری مجموعه‌های راف اقدام کردند تا به الگوریتمی قوی برای حل مسائل بهینه‌سازی متفاوت، همچون بهبود اقتصاد ایالات مکزیک، دست یابند. نتایج حاصل نشان‌دهنده کارایی بالای این الگوریتم در حل مسائل است (Santana, 2010). چینگ هو چنگ^۴ (۲۰۱۰) و فنگ خو و همکاران^۵ (۲۰۰۲)، به‌منظور رویکرد استخراج یکپارچه ویژگی‌ها، از ترکیب الگوریتم ژنتیک و مجموعه‌های راف استفاده کرده‌اند. مدل اولیه سیستم ادغام قابلیت‌های مجموعه‌های راف در مدیریت عدم قطعیت با یک موتور جست‌وجوی قوی براساس الگوریتم ژنتیک را نشان می‌دهد (Pheng-Khoo, 2002). از سیستم جدید پیشنهادی (ترکیب مجموعه‌های راف و الگوریتم ژنتیک)، به‌منظور کشف قوانین از داده‌های ناسازگار، استفاده کرده‌اند. نتایج نشان می‌دهد رویکرد پیشنهادی جدید برای کار با داده‌های دارای عدم قطعیت موفقیت‌آمیز بوده است (Pheng-Khoo & Yin, 2001). همچنین به‌منظور پیش‌بینی هزینه‌های بالای اجتماعی، رویکردی ترکیبی از مجموعه‌های راف و ژنتیک پیشنهاد داده‌اند که از الگوریتم ژنتیک برای ساخت یک مدل پیش‌بینی، با متغیرهای دریافت‌شده از مجموعه‌های راف، استفاده می‌کند. یافته‌ها نشان می‌دهد الگوریتم ژنتیک همراه با تئوری مجموعه‌های راف می‌تواند رویکرد ترکیبی کارآمد و مؤثری ایجاد کند (McKee & Lensberg, 2002).

1. Rough Set
3. Santana et al.
5. Pheng Khoo et al.

2. Zdzislaw Pawlak
4. Ching Hsue Cheng

درجه و ۵۰ دقیقه شمالی به عرض تقریبی ۳۰ کیلومتر گسترده شده است (طالقانی، ۱۳۸۵). میانگین مصرف روزانه آب تهران، طی سال ۱۳۸۹، بیش از ۱۴۰۰۰۰۰ مترمکعب بوده که نشان‌دهنده اهمیت موضوع مورد بررسی است.



شکل ۱. مناطق شش‌گانه آبفای شهر تهران

داده‌های به‌کاررفته در این تحقیق شامل این موارد است: داده‌های سرشماری سال ۱۳۸۵ در سطح واحدهای بلوک شهری (مرکز آمار ایران)، حاوی ۴۸ واحد اطلاعاتی؛ داده‌های اقلیمی شامل کمترین و بیشترین میزان دما و کمترین میزان رطوبت، به‌صورت فصلی، در بازه زمانی سال ۱۳۸۲ تا ۱۳۹۰ هجری شمسی (سازمان هواشناسی ایران)؛ داده‌های محیطی شامل طول و عرض جغرافیایی در سیستم متریک؛ ارتفاع منطقه مورد مطالعه، با دقت مکانی ۳۰ متر؛ سطح فضای سبز و پوشش گیاهی در منطقه و همچنین داده‌های فنی-مدیریتی، شامل آب‌بها و فشار آب موجود در شبکه‌های آب‌رسانی. در مجموع، پایگاه داده این تحقیق ۵۷ واحد اطلاعاتی و ۳۰۲۱۷ رکورد در سطح بلوک و نواحی آبفای شهر تهران را دربر می‌گیرد. داده‌های مصرف آب شهری با کاربری مسکونی در سطح نواحی آبفای شهر تهران، در بازه

رویکرد ترکیبی الگوریتم ژنتیک و مجموعه‌های راف را به‌کار برده‌اند و نتایج آن کارآیی بالای رویکرد ترکیبی پیشنهادی را نشان می‌دهد (Yan-Liang & Che Huang, 2009).

پروژه‌های تحقیقاتی بسیاری در مورد آب شرب مصرفی شهر تهران انجام شده ولی اهمیت مکان و همچنین، تحلیل‌های مکانی در این موضوع کمتر از آنچه باید مورد توجه قرار گرفته است. بنابراین، هدف از طرح این تحقیق بررسی و ارزیابی کارآیی و دقت خروجی از تلفیق الگوریتم‌های مجموعه‌های راف و ژنتیک، در استخراج قوانین مرتبط با آب مصرفی نواحی، با کاربری مسکونی در شهر تهران است. نوآوری این تحقیق به‌کارگیری تکنیک‌های داده‌کاوی با رویکرد مکانی، به‌ویژه تلفیق الگوریتم ژنتیک و مجموعه‌های راف با موضوع آب کلان‌شهر تهران، با توجه به بحران کمبود آب این کلان‌شهر است که در سطح شهر تهران انجام نشده است. با توجه به قوانین استخراج‌شده و اعتبارسنجی آن‌ها، اثر هر یک از عوامل مورد بررسی در آب مصرفی تهران تحلیل می‌شود. نتایج به‌دست‌آمده از این تحقیق را می‌توان در سیاست‌گذاری و مدیریت آب مصرفی شهر تهران به‌کار برد.

۲- مواد و روش‌ها

۲-۱- موقعیت جغرافیایی منطقه مورد مطالعه

با توجه به موضوع تحقیق و پیچیدگی‌های مسئله آب شهری و بحران آب حاکم بر شهر تهران و نیز ویژگی‌های منحصربه‌فرد این شهر، که موجب پیچیدگی بیشتر مسئله می‌شود، تهران به‌منزله منطقه مورد مطالعه انتخاب شده است. از نظر آب‌وفاضلاب، شهر تهران به شش منطقه آبفا و نوزده ناحیه زیرمجموعه مناطق تقسیم‌بندی می‌شود (شکل ۱). شهر تهران در شمال ایران و جنوب دامنه رشته‌کوه البرز، در حدفاصل طول جغرافیایی ۵۱ درجه و ۲ دقیقه شرقی تا ۵۱ درجه و ۳۶ دقیقه شرقی، به طول تقریبی ۵۰ کیلومتر و عرض جغرافیایی ۳۵ درجه و ۳۴ دقیقه شمالی تا ۳۵

جدول تصمیم با اشیا و ستون‌ها با ویژگی‌ها مرتبط باشند. کلاس‌های طبقه‌بندی شده «ویژگی‌های تصمیم» نامیده می‌شوند و دیگر ویژگی‌ها «ویژگی‌های شرطی» به‌شمار می‌روند. مجموعه C برای تعیین ویژگی‌ها و مجموعه D برای تشخیص ویژگی تصمیم استفاده می‌شود (Thangavel, 2009); به‌گونه‌ای که:

$$C \cap D = \emptyset$$

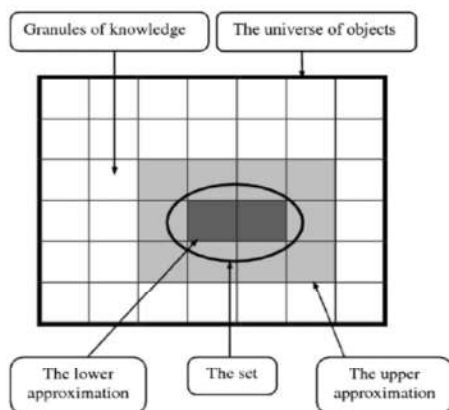
در نظریه مجموعه‌های راف، مجموعه براساس کلاس‌های هم‌ارز شامل lower approximation, upper approximation و boundary region به سه بخش تقسیم می‌شود. بر این اساس، سه عملگر برای تعریف سه بخش یادشده، به شرح زیر، تعریف می‌شوند:

$$B_*(x) = \{x \in U: B(x) \subseteq X\}$$

$$B^*(x) = \{x \in U: B(x) \cap X \neq \emptyset\}$$

$$BN_B(x) = B^*(x) - B_*(x)$$

در معادلات بالا $B_*(x)$ ، $B^*(x)$ و $BN_B(x)$ به ترتیب lower approximation، upper approximation و boundary region شمرده می‌شوند. کران پایین شامل همه اشیا است که، به‌طور قطع، براساس داده‌های گردآوری شده طبقه‌بندی می‌شوند. کران بالا همه اشیا را شامل می‌شود که به‌صورت احتمالی، براساس داده‌های گردآوری شده طبقه‌بندی می‌شوند. ناحیه مرزی اختلاف بین کران بالا و کران پایین است. این سه بخش در شکل ۲ نشان داده شده‌اند.



شکل ۲. نمایش یک مجموعه راف با سه ناحیه برآورد بالا، برآورد پایین و ناحیه مرزی

زمانی فصلی، از بهار ۱۳۸۲ تا زمستان ۱۳۸۹ است که این داده‌ها از مناطق شش‌گانه آبفای این شهر گردآوری شده‌اند. با توجه به اینکه داده‌های مصرف آب شهری در واحد نواحی نوزده‌گانه آبفای این شهرند و دسترسی به واحدهای کوچک‌تر امکان‌پذیر نیست، داده‌های اجتماعی-اقتصادی که در واحد بلوک‌های شهری در دسترس بودند نیز براساس نواحی آبفای شهر تهران گردآوری شدند. همچنین، با توجه به نبود اطلاعات سرشماری در تمامی سال‌های مورد تحلیل در تحقیق، با استفاده از سرشماری سال‌های ۱۳۷۵ و ۱۳۸۵، اقدام به درون‌یابی و محاسبه جمعیت در سال‌های مورد نظر شد. شایان ذکر است که فقدان داده‌های سرشماری دقیق در هر سال و جابه‌جایی‌های متعدد شهروندان در منازل مسکونی در سطح شهر، به‌ویژه به‌علت استیجاری بودن آن‌ها، موجب افزایش عدم قطعیت در داده‌های سرشماری می‌شود.

۲-۲- مجموعه‌های راف

تئوری مجموعه‌های راف یکی از توانمندترین و انعطاف‌پذیرترین ابزار ریاضی برای استدلال در زمینه عدم قطعیت به‌شمار می‌رود (Triantaphyllou & Felici, 2006). این تئوری ابزاری برای تولید و تحلیل مجموعه قواعد تصمیم از جدول داده است و قابلیت‌های بسیاری برای پردازش و استخراج قوانین و روابط بین عوامل ناشناخته تأثیرگذار و نیز کاهش پیچیدگی و افزونگی داده‌ها در پایگاه داده‌ها دارد. بنابراین در این تحقیق، قابلیت‌های الگوریتم مجموعه‌های راف، با هدف حل مسئله آب و استخراج قوانین و الگوهای حاکم بر مصرف آب شهری، بررسی شده است.

نقطه آغاز نظریه مجموعه‌های راف ایجاد رابطه تشابه است که این رابطه با کمک اطلاعات درباره اشیا ایجاد می‌شود. رابطه مشابهت این واقعیت را بیان می‌کند که، به‌علت کمبود داده و دانش، قادر نخواهیم بود برخی از اشیا را تفکیک کنیم. در نظریه مجموعه‌های راف، داده‌ها در یک جدول، به نام جدول تصمیم، سازماندهی می‌شوند؛ به‌صورتی که سطرهای

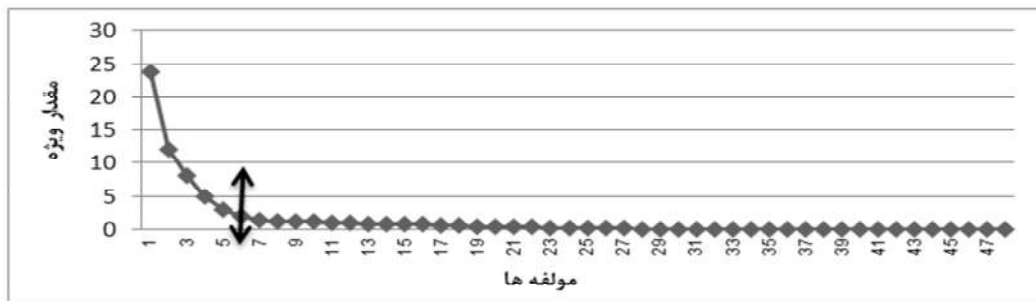
داده‌های بزرگ و استخراج قوانین و عوامل تأثیرگذار در موضوع آب شهر تهران، و نیز بهینه‌سازی این قوانین مؤثر می‌تواند به برنامه‌ریزی و مدیریت آب شهری کمک شایانی کند. بنابراین با توجه به پیچیدگی ماهیت مسئله آب و نیز قوانین استخراج‌شده از اجرای مجموعه‌های راف، برای بهینه‌سازی قوانین استخراجی از مجموعه‌های راف از الگوریتم ژنتیک استفاده می‌شود. مشخصه‌های مهم در نظر گرفته‌شده در ارزیابی سرعت همگرایی و دقت پاسخ‌اند.

۲-۴- تبدیل و آماده‌سازی داده‌ها

مطابق فلوجارت روش تحقیق، داده‌ها در ابتدا ویرایش و هم‌خوان‌سازی شدند و با استفاده از رابطه (۱)، بین مقادیر صفر تا ۱۰ درجه‌بندی و نرمال شدند.

$$y = \left(\frac{x - x_{\min}}{x_{\max} - x_{\min}} \right) * 10 \quad (1)$$

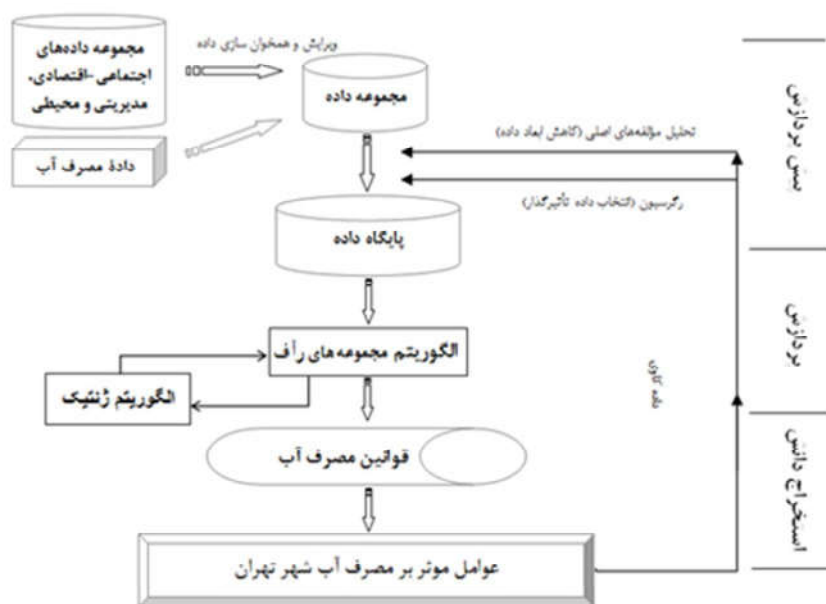
سپس با اجرای تحلیل مؤلفه‌های اصلی^۳ روی داده‌ها، ابعاد و همبستگی بین داده‌ها کاهش یافت. همان‌طور که در نمودار مقادیر ویژه^۴ استخراج‌شده روی داده‌های اجتماعی-اقتصادی شهر تهران (شکل ۳) مشاهده می‌شود، از مؤلف، ششم (PC6) به بعد، مقادیر ویژه به‌شدت کاهش یافته که وجود اطلاعات ناچیز و قابل حذف در این قسمت از داده‌ها را نشان می‌دهد. روند نمای تحقیق در شکل ۴ نشان داده شده است.



شکل ۳. نمودار مقادیر ویژه موجود در مؤلفه‌های گوناگون داده‌های اجتماعی-اقتصادی مناطق شهر تهران

- | | |
|---------------------------------|-----------------|
| 1. genetic algorithm | 2. John Holland |
| 3. principal component analysis | 4. eigen values |

ارزیابی قابلیت‌های ترکیب الگوریتم‌های مجموعه‌های راف و ژنتیک ...

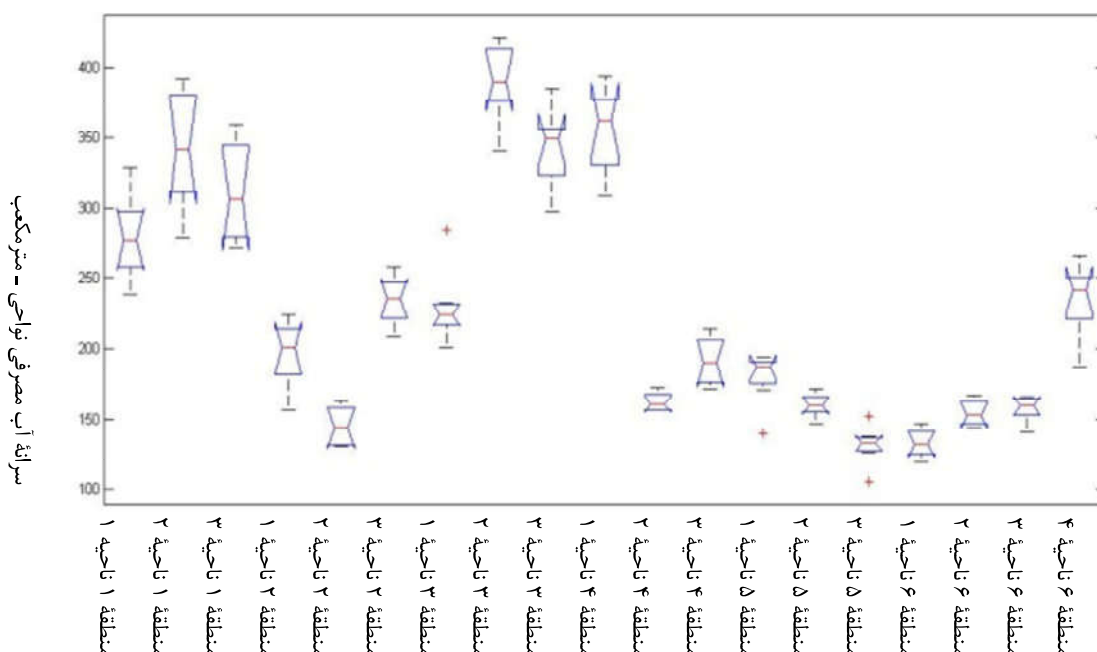


شکل ۴. روند نمای تحقیق

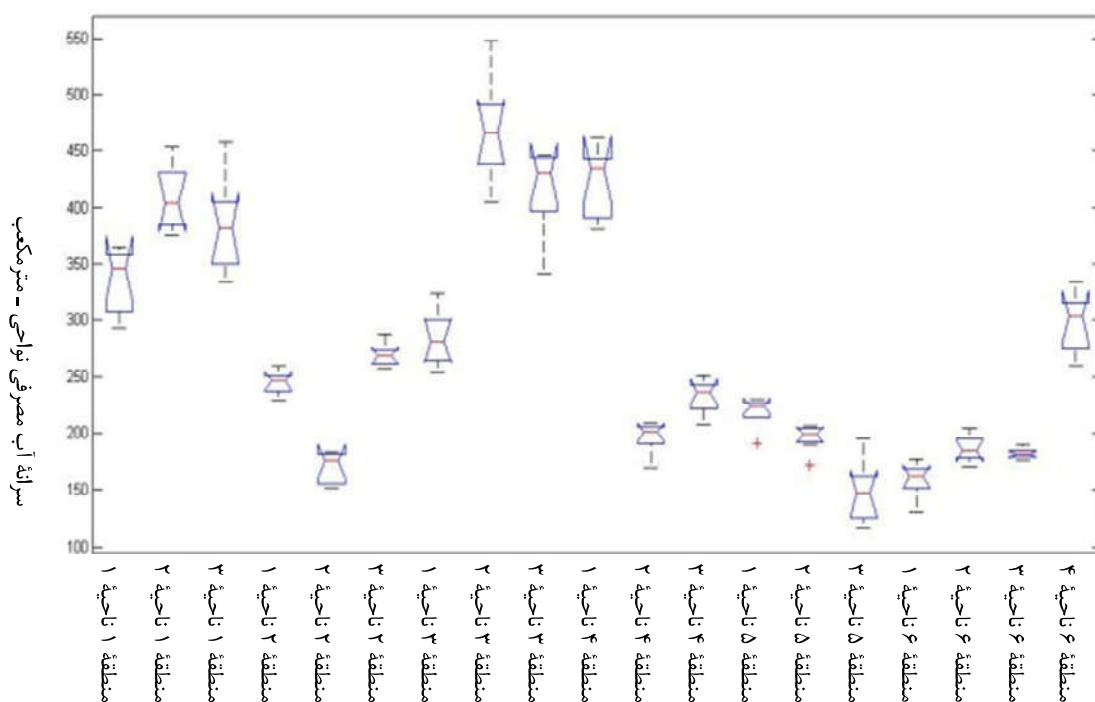
۳- مصارف آب نواحی آبیای شهر تهران

در بررسی و تحلیل داده‌های مصرف آب نواحی مسکونی تهران، برای نرمال‌سازی داده‌های مصرفی روزانه، بر مبنای متغیر جمعیت، اقدام شد و میانگین مصرف آب روزانه، برای یک نفر در هر ناحیه، از طریق

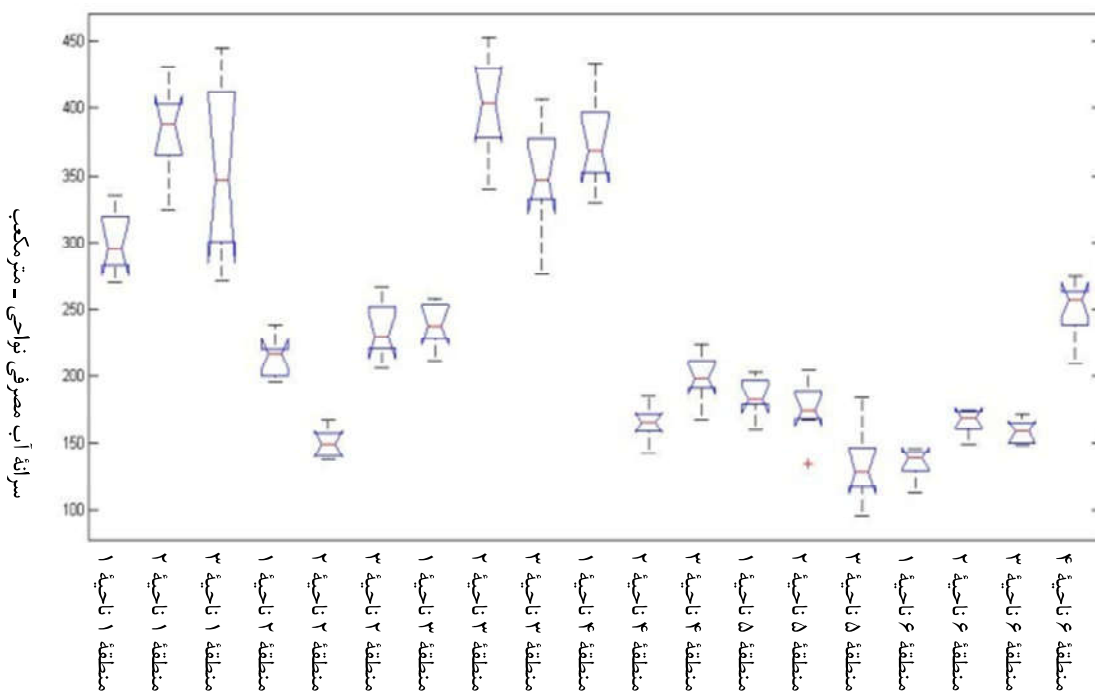
تقسیم کل مصرف آب روزانه هر ناحیه بر جمعیت محاسبه شد. نمودار مصرف سرانه آب مشترکان شهر تهران، به تفکیک نواحی آبیای شهر تهران در چهار فصل سال، در بازه زمانی هشت‌ساله جداگانه محاسبه شد (شکل‌های ۸-۵).



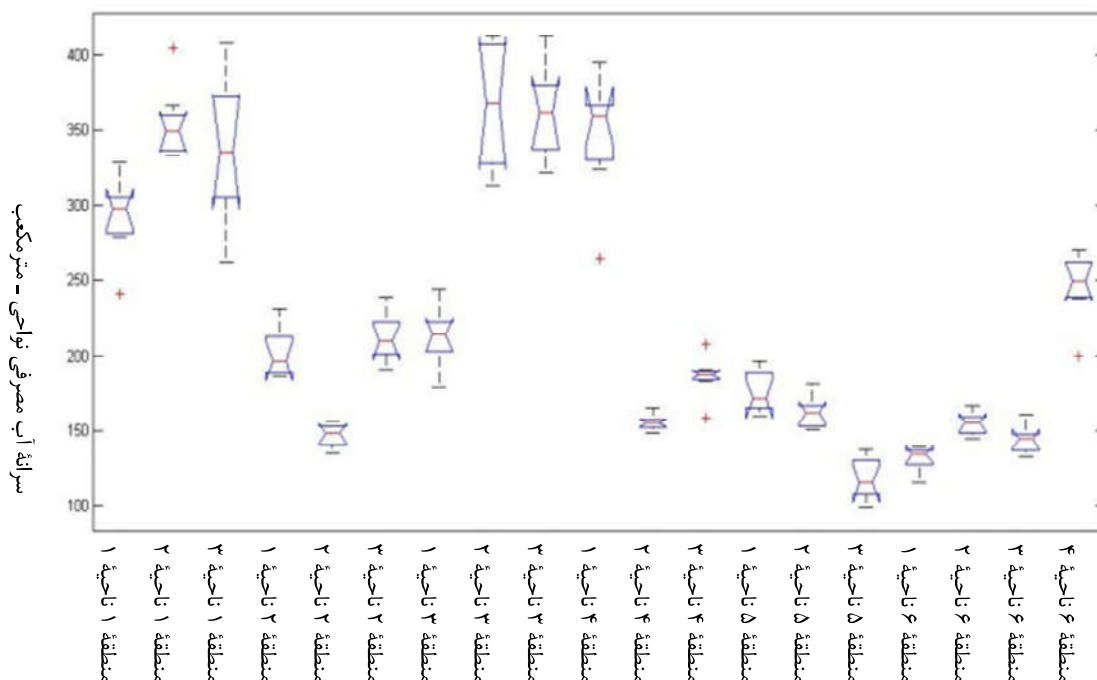
شکل ۵. نمودار سرانه آب مصرفی روزانه مناطق و نواحی آبیای شهر تهران در فصل بهار



شکل ۶. نمودار سرانه آب مصرفی روزانه مناطق و نواحی آبفای شهر تهران در فصل تابستان



شکل ۷. نمودار سرانه آب مصرفی روزانه مناطق و نواحی آبفای شهر تهران در فصل پاییز



شکل ۸. نمودار سرانه آب مصرفی روزانه مناطق و نواحی آبفای شهر تهران در فصل زمستان

۴- استخراج الگوها و قوانین حاکم بر مجموعه داده

در اجرای الگوریتم مجموعه‌های راف، ابتدا طبقه‌بندی داده‌های مصرف سرانه آب به منزله متغیر تصمیم (وابسته) در چهار دسته در نظر گرفته می‌شوند: کلاس‌های مصرف کم، نرمال، زیاد و بسیار زیاد. سپس پایگاه داده به دو مجموعه آموزشی (۶۰٪) و آزمایشی (۴۰٪) تقسیم شد. افزایشی یا کاهش‌ی بودن هر یک از متغیرهای مستقل تعیین شد و سپس جدول تصمیم و استخراج قوانین حاکم بر مجموعه داده‌ها شکل گرفت. قوانین استخراج‌شده، با در نظر گرفتن هر یک از کلاس‌های آب مصرفی (کلاس‌های تصمیم) و کران‌های بالا و پایین متغیرها به صورت «اگر (شرط) آن‌گاه (نتیجه)» بیان می‌شوند.

برای اجرای الگوریتم ژنتیک کروموزوم‌ها شامل تمامی متغیرهای مستقل تأثیرگذار تعریف شدند و به منزله جمعیت اولیه^۱، برای ورود به الگوریتم ژنتیک،

با توجه به مقایسه نمودارهای یادشده، اگرچه در نواحی شهری پراکندگی مصرف و میانگین مصرف آب در کاربری مسکونی، در فصل‌های گوناگون، متفاوت است؛ همچنان روند یکسان و مشابهی در تمامی نمودارها مشاهده می‌شود که بیانگر وابستگی‌های مکانی- زمانی موجود در داده‌های مصرف آب شهری است؛ به گونه‌ای که، با تغییر در دو بعد زمان و مکان، مصرف آب شهری نیز تغییر بارزی می‌یابد. شایان ذکر است که عوامل متأثر از بعد زمان تأثیری تقریباً مشابه در مصرف آب نواحی دارند. در حالی که بعد مکان نشان‌دهنده اختلافات بیشتر در مصرف آب نواحی شهر تهران است. با تقسیم شهر تهران به دو نیمه شمالی و جنوبی، مناطق نیمه جنوبی مصرف سرانه‌ای حدود ۱۰۰ تا ۲۰۰ مترمکعب آب، با تغییرپذیری اندک، دارند. در حالی که مناطق نیمه شمالی دارای مصرف سرانه بالاتر از ۲۵۰ مترمکعب، با تغییرپذیری بیشتری هستند. این اختلاف‌ها نشان‌دهنده تغییرپذیری زمانی- مکانی مصرف سرانه آب ساکنان شهر است.

1. initial population preparation

تولیدشده از اجرای الگوریتم مجموعه‌های راف آمده است. این کیفیت با این سه معیار متفاوت اندازه‌گیری می‌شود (حسین‌زاده، ۱۳۸۶):

پشتیبانی^۱: درصدی از داده‌های آموزشی که سمت چپ قانون برای آن‌ها صدق می‌کند. اگر در مشاهده‌ای قانون LHS^۳ صدق کند، گفته می‌شود که قانون برای آن مشاهده «به‌کار می‌رود». این شاخص بیان‌کننده قابلیت کاربرد قانون موردنظر است.

اطمینان^۴: بر مبنای رکوردهای آموزشی محاسبه می‌شود که قانون LHS را می‌توان در آن‌ها به‌کار برد و درصدی از این رکوردها در نظر گرفته می‌شود که RHS^۵ آن‌ها صحیح باشد. به عبارت دیگر، این شاخص درستی قوانین را بررسی می‌کند.

دریافت^۶: چند درصد از رکوردهای مورد نظر به‌درستی از طریق یک قانون دریافت می‌شوند. در واقع، نوعی انعکاس از ساختار مسئله است. اگر قانونی با دریافت نزدیک به ۱۰۰٪ وجود داشته باشد، به این معناست که در فضای پیش‌بینی‌کننده، تمامی مشاهدات با این طبقه نزدیک یکدیگر قرار گرفته‌اند و قانون توانسته است آن بخش از فضای پیش‌بینی‌کننده را به‌خوبی دریافت کند.

در اعتبارسنجی صورت‌گرفته از قوانین استخراجی با دو روش مجموعه‌های راف و الگوریتم ژنتیک، دقت طبقه‌بندی مجموعه‌آزمون، با استفاده از قوانین استخراج‌شده با کمک ترکیب الگوریتم ژنتیک، و در مقایسه با دقت طبقه‌بندی حاصل از مجموعه‌های راف، بدون الگوریتم ژنتیک، بیشتر است. به‌صورتی که تعداد قوانین در ترکیب مجموعه‌های راف و الگوریتم ژنتیک ۳۵ قانون با دقت ۹۲/۲٪ است. اما مجموعه‌های راف بدون الگوریتم ژنتیک دارای ۲۲ قانون و دقت ۷۷٪ است که توانایی تعمیم قوانین را،

استفاده شدند؛ بدین‌معنا که جدول تصمیم تشکیل‌شده دارای ویژگی‌های شرطی و تصمیم است و هریک از ویژگی‌ها معرف ژن‌های گوناگون تشکیل‌دهنده کروموزوم محسوب می‌شود. تابع برازش^۱، در واقع، کیفیت مجموعه قوانین را نشان می‌دهد و به‌صورت درصدی از نمونه‌هایی که به‌درستی با استفاده از قوانین تصمیم طبقه‌بندی شده‌اند، با استفاده از رابطه (۲)، محاسبه می‌شود (Sheng, 2008).

$$F = \frac{C}{N} \quad \text{رابطه (۲)}$$

در این رابطه، F و C و N به ترتیب نشان‌دهنده تابع برازش‌اند؛ شامل درصد رکوردهای به‌درستی طبقه‌بندی‌شده، تعداد نمونه‌های به‌درستی طبقه‌بندی‌شده و تعداد کل نمونه‌ها.

برای تولید نسل‌های جدید در الگوریتم ژنتیک، ژن‌ها متشکل از ویژگی‌های شرط و تصمیم‌اند و با استفاده از کران‌های بالا و پایین در نظر گرفته‌شده برای هر ویژگی در الگوریتم مجموعه‌های راف تشکیل می‌شوند. با توجه به تابع برازش (F) تعریف‌شده که باید بیشترین مقدار را داشته باشد، فرزندان از نسل جدید انتخاب می‌شوند که بالاترین مقدار از تابع برازش تعریف‌شده را داشته باشند.

توقف الگوریتم ژنتیک براساس یکی از این شرایط اعمال می‌شود:

- حد از پیش تعیین‌شده یک نسل
- هنگامی که بهترین تناسب کروموزومی به حد آستانه‌ای از پیش تعیین‌شده می‌رسد.
- هنگامی که بهترین تناسب کروموزومی هیچ پیشرفتی به نسبت تعداد مشخصی از نسل‌ها نداشته باشد.

۵- نتایج و بحث

اعتبارسنجی قوانین تولیدی با استفاده از الگوریتم‌های مجموعه‌های راف و ژنتیک در جدول‌های ۱ و ۲ ارائه شده است. اعتبار هریک از قوانین در خلاصه قوانین

1. fitness function
2. support
3. left hand side
4. confidence
5. right hand side
6. capture

ارزیابی قابلیت‌های ترکیب الگوریتم‌های مجموعه‌های راف و ژنتیک ...

می‌کنند؛ قوانین استخراج‌شده از طریق مجموعه‌های راف با الگوریتم ژنتیک، به نسبت قوانین استخراج‌شده از مجموعه‌های راف، بدون استفاده از الگوریتم ژنتیک (۳۸٪)، میانگین اطمینان بیشتری (۷۲٪) دارند. همچنین، میانگین دریافت قوانین استخراج‌شده از طریق مجموعه‌های راف با الگوریتم ژنتیک و بدون آن، به ترتیب، ۵۱٪ و ۴۱٪ است (جدول‌های ۱ و ۲).

به علت جست‌وجوی مؤثر در فضای پایگاه داده در الگوریتم ژنتیک و همگرانشدن به نقاط بهینه محلی، نشان می‌دهد. افزون‌بر آن، اگرچه هریک از قوانین استخراج‌شده با مجموعه‌های راف، بدون الگوریتم ژنتیک، در مواردی رکورد بیشتری (بالاتر از ۱۷۰ رکورد) به نسبت مجموعه‌های راف با الگوریتم ژنتیک (بالاتر از ۶۰ رکورد) داشتند که از قوانین پشتیبانی

جدول ۱. کیفیت قوانین تولیدشده از اجرای الگوریتم مجموعه‌های راف

شماره قانون	کلاس اختصاص یافته به قانون	طول قانون	پشتیبانی		اطمینان	دریافت	مناطق مورد پشتیبانی قانون
			تعداد	درصد			
۳	Low	۲	۴۹	۲۰.۰	۳۰.۶	۲۱.۱	۶۳، ۵۲، ۴۱، ۳۱
۴	Low	۲	۱۳	۵.۳	۸۴.۶	۱۵.۵	۲۲
۵	Normal	۲	۱۰۳	۴۲.۲	۵۵.۳	۸۰.۳	۶۳، ۶۱، ۵۳، ۵۱، ۴۳، ۴۲، ۲۲، ۲۱
۷	Normal	۱	۱۰۶	۴۳.۴	۵۶.۶	۷۷.۹	۳۱، ۲۳، ۵۳، ۵۲، ۵۱، ۴۳، ۴۲، ۲۱، ۱۱
۹	High	۱	۱۷۳	۷۰.۹	۱۴.۵	۵۲.۱	۴۲، ۳۳، ۳۱، ۲۳، ۲۲، ۲۱، ۱۳، ۱۱
۱۰	High	۲	۱۶۳	۶۶.۸	۲۵.۲	۸۵.۴	۴۳، ۲۱، ۶۴، ۴۱، ۳۱، ۲۳، ۱۲، ۱۳، ۱۱
۱۱	High	۲	۶۱	۲۵.۰	۱۸.۰	۲۲.۹	۲۳، ۱۲
۱۲	High	۲	۱۴۶	۵۹.۸	۹.۶	۲۹.۲	۴۳، ۱۳، ۴۱، ۶۴، ۲۱
۱۳	High	۳	۱۳	۵.۳	۸۴.۶	۲۲.۹	۱۱
۱۷	High	۲	۵۵	۲۲.۵	۵۴.۵	۶۲.۵	۱۱، ۳۳، ۲۳، ۱۳، ۱۲
۱۸	High	۳	۴۷	۱۹.۳	۳۸.۳	۳۷.۵	۶۴، ۳۳، ۱۳
۲۱	Normal	۳	۲۶	۱۰.۷	۷.۷	۲.۶	۳۳، ۳۲
۲۲	Normal	۲	۱۰۳	۴۲.۲	۱۷.۵	۲۳.۴	۲۳، ۶۲، ۶۴، ۱۱

جدول ۲. کیفیت قوانین تولیدشده از ترکیب الگوریتم‌های مجموعه‌های راف و ژنتیک

شماره قانون	کلاس اختصاص یافته به قانون	طول قانون	پشتیبانی		اطمینان	دریافت	مناطق مورد پشتیبانی قانون
			تعداد	درصد			
۱	Normal	۲	۴۲	۱۷.۲	۸۳.۳	۴۵.۵	۵۱-۴۳-۲۱
۲	Low	۳	۳۷	۱۵.۲	۷۰.۲	۳۶.۶	۶۲-۵۳-۵۱
۳	Low	۳	۴۴	۱۸.۰	۷۲.۷	۴۵.۱	۵۳-۵۲-۴۲
۴	Low	۳	۴۳	۱۷.۶	۴۶.۵	۲۸.۲	۵۳-۶۲-۵۲-۵۱
۵	Low	۳	۵۶	۲۲.۹	۸۳.۹	۶۶.۲	۶۳-۶۲-۶۱-۵۳
۶	Very High	۲	۳۸	۱۵.۶	۵۰.۰	۳۹.۶	۳۳-۳۲

ادامه جدول ۲.

مناطق مورد پشتیبانی قانون	دریافت	اطمینان	پشتیبانی		طول قانون	کلاس اختصاص یافته به قانون	شماره قانون
			تعداد	درصد			
۵۳-۲۲	۴۷.۹	۸۹.۴	۳۸	۱۵.۶	۲	Low	۷
۳۲-۱۳-۱۲	۵۲.۱	۸۰.۶	۳۱	۱۲.۷	۲	Very High	۸
۵۱-۴۳-۲۱	۴۰.۳	۸۱.۶	۳۸	۱۵.۶	۲	Normal	۹
۵۳-۴۳-۴۲	۲۸.۲	۶۰.۶	۳۳	۱۳.۵	۳	Low	۱۰
۴۳-۵۳-۶۲-۶۱-۴۲	۳۸.۰	۶۲.۸	۴۳	۱۷.۶	۳	Low	۱۱
۵۳-۵۲-۴۲	۳۳.۸	۷۰.۶	۳۴	۱۳.۹	۳	Low	۱۲
۵۳-۴۲-۲۲	۵۷.۷	۸۵.۴	۴۸	۱۹.۷	۲	Low	۱۳
۵۱-۴۳-۲۱	۴۴.۲	۸۵.۰	۴۰	۱۶.۴	۳	Normal	۱۴
۳۳-۳۲-۱۳-۱۲	۶۴.۶	۷۹.۵	۳۹	۱۶.۰	۳	Very High	۱۵
۶۲-۶۱-۵۳-۴۲	۵۳.۵	۶۷.۸	۵۶	۲۳.۰	۲	Low	۱۶
۳۳-۳۲-۱۲	۵۰.۰	۷۰.۶	۳۴	۱۳.۹	۳	Very High	۱۷
۶۳-۵۲-۶۱-۵۳	۴۶.۵	۷۵.۰	۴۴	۱۸.۰	۲	Low	۱۸
۵۳-۶۲-۶۱-۵۲-۳۲	۴۰.۸	۵۳.۷	۵۴	۲۲.۱	۲	Low	۱۹
۵۳-۶۲-۶۱-۵۱-۴۲	۴۳.۷	۶۰.۸	۵۱	۲۰.۹	۲	Low	۲۰
۶۱-۵۳-۶۲-۴۲	۳۸.۰	۵۲.۹	۵۵	۲۲.۵	۲	Low	۲۱
۴۳-۶۳-۶۱-۵۳-۴۲	۵۰.۷	۷۳.۵	۴۹	۲۰.۱	۲	Low	۲۲
۳۳-۳۲	۲۰.۸	۳۲.۲	۳۱	۱۲.۷	۲	Very High	۲۳
۳۳-۱۲-۳۲	۶۰.۴	۷۶.۳	۳۸	۱۵.۶	۳	Very High	۲۴
۶۱-۵۳-۶۲-۵۱-۴۳-۴۲	۵۰.۷	۷۸.۲	۴۶	۱۸.۹	۳	Low	۲۵
۵۳-۶۲-۶۱-۴۲	۴۶.۵	۶۶.۰	۵۰	۲۰.۵	۳	Low	۲۶
۶۱-۵۳	۶۳.۴	۹۳.۸	۴۸	۱۹.۷	۲	Low	۲۷
-۶۲-۵۳-۵۲-۵۱-۴۳-۴۲-۲۱-۱۱-۶۴	۶۴.۹	۸۴.۷	۵۹	۲۴.۲	۲	Normal	۲۸
۴۱-۳۳-۱۲-۱۱-۳۲-۱۳	۸۳.۳	۷۵.۵	۵۳	۲۱.۷	۲	Very High	۲۹
-۵۲-۵۱-۴۳-۴۲-۳۱-۲۳-۲۱-۱۱-۶۴-۶۲-۵۳	۵۸.۴	۸۱.۸	۵۵	۲۲.۵	۲	Normal	۳۰
۳۱-۱۱-۳۳-۱۳-۱۲	۶۸.۸	۷۸.۶	۴۲	۱۷.۲	۲	High	۳۱
۱۱-۶۴-۳۳-۲۳-۱۳	۴۵.۸	۶۸.۸	۳۲	۱۳.۱	۲	High	۳۲
۳۳-۴۱-۱۲-۱۱-۳۲-۱۳	۹۳.۸	۸۳.۳	۵۴	۲۲.۱	۲	Very High	۳۳
۳۱-۱۱-۴۱-۳۳-۳۲-۱۳	۸۳.۳	۶۹.۰	۵۸	۲۳.۸	۲	High	۳۴
۳۱-۱۳-۱۱	۴۵.۸	۶۶.۷	۳۳	۱۳.۵	۲	High	۳۵

می‌رسد. پس از آن، با افزایش سرعت و تعداد نسل، تأثیر چندانی در دقت طبقه‌بندی ندارد (شکل ۹ و جدول ۳). بنابراین، با افزایش سرعت همگرایی و افزایش تعداد نسل شکل گرفته، تا مرحله‌ای، دقت قوانین خروجی و نیز دقت طبقه‌بندی مجموعه‌آزمون طبق قوانین تصمیم افزایش می‌یابد. اما پس از آن، با افزایش پارامترهای تعیین شده، دقت طبقه‌بندی تغییر چندانی نمی‌کند و تغییرات آن جزئی و قابل چشم‌پوشی است.

تعداد نسل پدیدآمده و سرعت همگرایی الگوریتم پارامترهای در نظر گرفته شده برای اجرای الگوریتم ژنتیک به‌شمار می‌روند. نتایج نشان می‌دهد، با افزایش تعداد نسل و ثابت ماندن سرعت همگرایی (سرعت همگرایی پایین)، دقت طبقه‌بندی افزایش چشمگیری نشان نمی‌دهد. اما با افزایش سرعت همگرایی به سرعت معمول و تعداد شش نسل پدیدآمده، دقت طبقه‌بندی به ۰.۸۸٪ و پس از تولید ۱۰ نسل، دقت به ۰.۹۲/۲٪



شکل ۹. رابطه دقت طبقه‌بندی با تعداد نسل و سرعت همگرایی

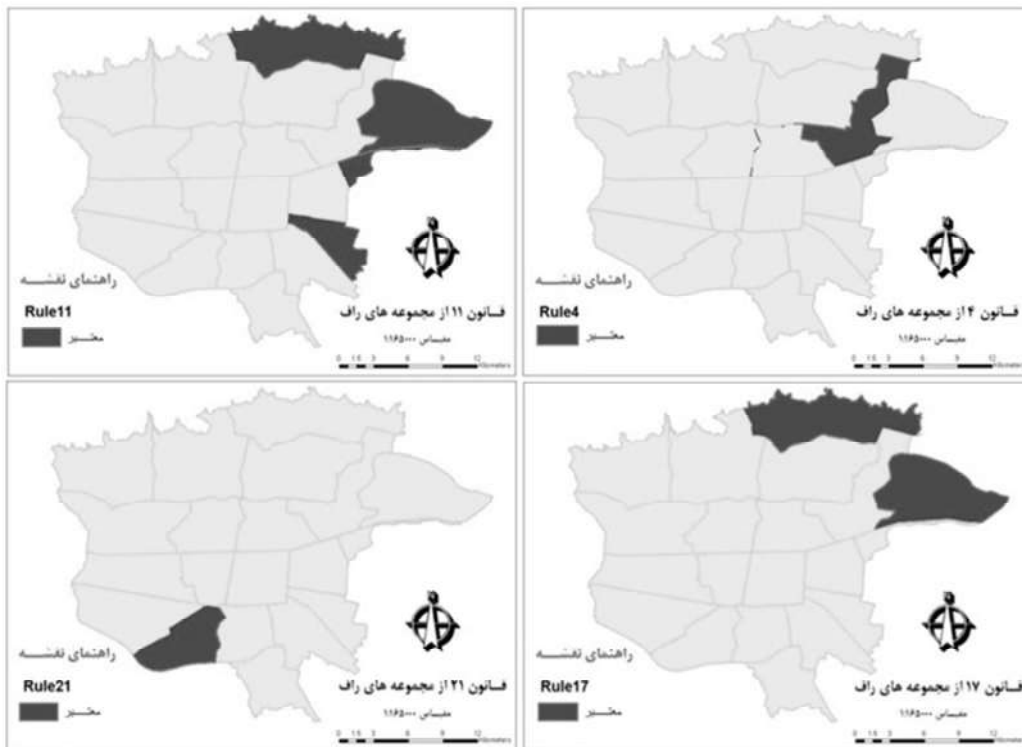
جدول ۳. دقت طبقه‌بندی مجموعه‌آزمون در مجموعه‌های راف و ترکیب آن با الگوریتم ژنتیک

قوانین	تعداد نسل	سرعت همگرایی	دقت طبقه‌بندی (درصد)
مجموعه‌های راف	-	-	۰.۶۲۸
برگزیده‌ای از مجموعه‌های راف	-	-	۰.۷۷۲
ترکیب مجموعه‌های راف و الگوریتم ژنتیک	۵ نسل	پایین	۰.۷۶۶
	۱۰ نسل		۰.۷۶۶
	۱۵ نسل		۰.۷۶۶
	۵ نسل	نرمال	۰.۷۸۱
	۶ نسل		۰.۸۸۸
	۱۰ نسل		۰.۹۲۲
	۱۵ نسل		۰.۹۲
	۵ نسل		۰.۹۲۴
	۱۰ نسل		۰.۹۲
	۱۵ نسل	بالا	۰.۹۲۴
	۱۵ نسل		۰.۹۲۴

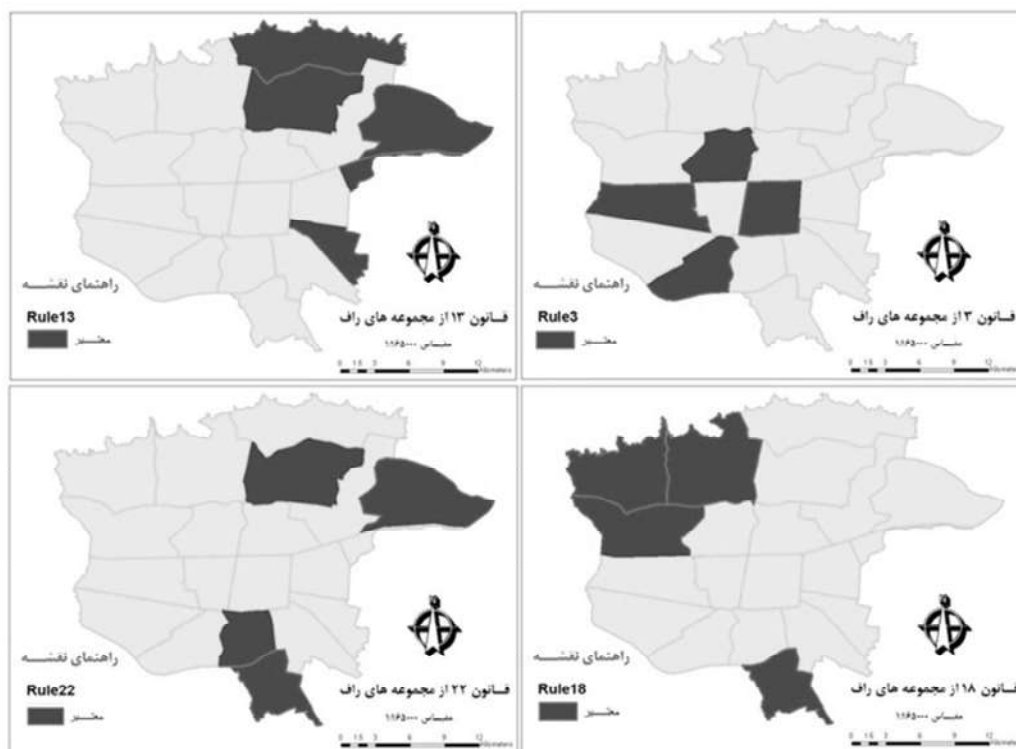
بین مصرف سرانه آب و تراکم جمعیت در واحد سطح رابطه معکوس وجود دارد. همان‌طور که مشاهده می‌شود، عامل تراکم جمعیت در واحد سطح به همراه موقعیت مکانی نواحی (طول و عرض جغرافیایی) به استخراج قوانین ویژه یادشده منجر شده‌اند.

توزیع مکانی قوانین ۳، ۱۳، ۱۸ و ۲۲ در شکل ۱۱ نشان داده شده است. مشاهده می‌شود که این قوانین با تعداد نواحی بیشتری تطابق دارند؛ بنابراین، اعتبار مکانی بالاتری دارند. قوانین یادشده در بالا بیشتر از عوامل جمعیت، تراکم جمعیت در واحد سطح و نیز بعد خانوار متأثرند. با افزایش بعد خانوار، ملاحظه می‌شود که مصرف سرانه آب شهری نیز کاهش می‌یابد؛ بنابراین، بین مصرف سرانه آب شهری و بعد خانوار ساکنان رابطه معکوس وجود دارد.

یکی دیگر از معیارهای اعتبار قوانین استخراج‌شده از مجموعه‌های راف پراکندگی و توزیع مکانی قوانین در سطح شهر تهران است؛ به گونه‌ای که هرچه میزان شمول قانون (از نظر زمانی-مکانی) بیشتر باشد، آن قانون اعتبار مکانی بیشتری خواهد داشت. شکل ۱۰ توزیع مکانی قوانین در نواحی شهری را نشان می‌دهد. قوانین ۴، ۱۱، ۱۷ و ۲۱ دارای توزیع مکانی محدود به یک یا دو ناحیه شهری‌اند که بیانگر ویژه‌بودن این قوانین است. افزایش تراکم جمعیت در واحد سطح در قوانین ۴، ۱۱، ۱۷ و ۲۱، به ترتیب، منجر به کاهش (قانون ۴)، افزایش (قوانین ۱۱، ۱۷) و نیز مصرف متعادل آب (قانون ۲۱) می‌شود. با کاهش تراکم جمعیت، به ویژه در مناطق شمالی شهر (قانون ۱۷)، مصرف سرانه آب نیز افزایش بارزی می‌یابد. بنابراین،



شکل ۱۰. توزیع مکانی اعتبار قوانین ۴، ۱۱، ۱۷ و ۲۱



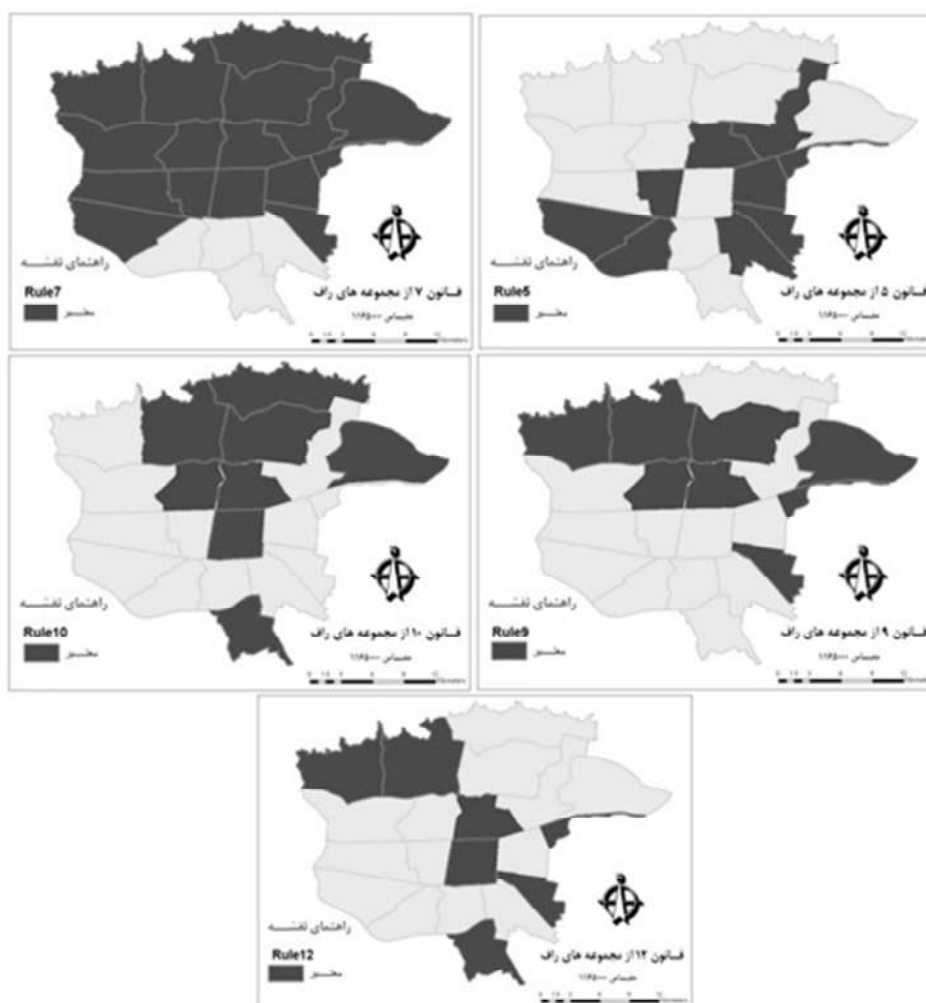
شکل ۱۱. توزیع مکانی قوانین ۳، ۱۳، ۱۸ و ۲۲

تراکم جمعیت در واحد سطح برگرفته از قوانین ۴، ۵ و ۱۷ مجموعه‌های راف و نیز قوانین ۱، ۷، ۸، ۹، ۱۳، ۱۴ و ۲۷ الگوریتم ژنتیک تأثیری کاهنده در مصرف سرانه آب دارد. عامل تراکم جمعیت در واحد سطح با درآمد خانوار (عامل اقتصادی) نیز مرتبط است. به همین دلیل، افزایش تراکم جمعیت در واحد سطح، در نواحی جنوبی شهر (با تراکم جمعیت بالا)، موجب کاهش سرانه آب مصرفی و در نواحی شمالی شهر (با تراکم جمعیت پایین)، موجب افزایش سرانه مصرف آب شهری می‌شود. باید گفته شود تعداد قوانین بیانگر اهمیت و تأثیرگذاری عامل تراکم جمعیت، در واحد سطح، در مصرف سرانه آب است.

قانون ۵ از مجموعه‌های راف و قوانین ۵، ۹، ۳۰ و ۳۳ از الگوریتم ژنتیک نیز تأثیر کاهنده عامل بعد خانوار در مصرف سرانه آب شهری را نشان می‌دهد. بدین ترتیب، با افزایش بعد خانوار، مصرف سرانه آب روندی کاهشی دارد. به علت اندک بودن بعد خانوار در مناطق شمالی و غربی شهر تهران، سرانه مصرف آب در این نواحی کم است.

با توجه به شکل ۱۲، با مشاهده توزیع و پراکندگی مناسب مناطق مربوط به قوانین ۵، ۷، ۹، ۱۰ و ۱۲ می‌توان اظهار کرد که این قوانین با نواحی شهری بیشتری مطابق‌اند و اعتبار مکانی بهتری دارند. با توجه به همین مطابقت ویژه، قوانین عام قلمداد می‌شوند. این قوانین بیشتر از عواملی همچون آب‌بها، فضای سبز و تراکم جمعیت در واحد سطح تأثیر می‌پذیرند؛ به‌صورتی که افزایش عامل آب‌بها در کل نواحی شهر تهران، به‌ویژه در مناطق شمالی آن، موجب کاهش در مصرف سرانه آب می‌شود. همان‌طور که ملاحظه می‌کنید، در بیشتر قوانین، عامل تراکم جمعیت در واحد سطح در مصرف سرانه آب مؤثر است که اهمیت این عامل در کنترل مصرف سرانه آب شهری را می‌رساند.

برای بررسی عوامل تأثیرگذار در سرانه مصرف آب شهری، از قوانین با میزان اطمینان بالا در الگوریتم ژنتیک (۸۰٪ و بالاتر) و مجموعه‌های راف (۵۰٪ و بالاتر) استفاده شد. همان‌طور که بیان کردیم، عامل



شکل ۱۲. توزیع مکانی اعتبار قوانین ۵، ۷، ۹، ۱۰ و ۱۲

عامل سرانه فضای سبز نیز در قانون ۷ مجموعه‌های راف و قوانین ۷ و ۱۳ الگوریتم ژنتیک انتخاب شده است. این عامل نیز بیشتر در قوانین عام وجود دارد و در مصرف سرانه آب تأثیری کاهنده دارد. به عبارتی، عامل سرانه فضای سبز با مصرف سرانه آب رابطه معکوس نشان می‌دهد.

عامل عرض جغرافیایی مندرج در قانون ۱۷ مجموعه‌های راف، و همچنین قانون ۵ الگوریتم ژنتیک، با مصرف آب رابطه‌ای افزایشده دارد. با توجه به اینکه ارزش ملک در شهر تهران، به صورت تقریبی، از جنوب به شمال شهر افزایش می‌یابد؛ این عامل نیز می‌تواند وابسته به میزان درآمد ساکنان باشد. در این صورت،

عامل آب‌بها از عوامل اقتصادی تأثیرگذار در مصرف آب، در قانون ۱۳ مجموعه‌های راف و قوانین ۲۸، ۳۰ و ۳۳ الگوریتم ژنتیک، شمرده می‌شود. این عامل درزمره قوانین عام (سطح شهر تهران) است و تأثیری کاهنده در مصرف سرانه آب شهری دارد. عامل آب‌بها یکی از عوامل مدیریتی است که بیشتر در مناطق و نواحی شمالی تهران تأثیر می‌گذارد زیرا مناطق جنوبی شهر کمترین میزان مصرف آب را دارند و کمترین آب‌بها را پرداخت می‌کنند. با افزایش مصرف آب، آب‌بها نیز افزایش تصاعدی می‌یابد. بنابراین، ممکن است به ابزاری در دست مدیران تبدیل شود که در جهت ایجاد انگیزه کاهش و صرفه‌جویی در مصرف آب به کار رود.

کاهشی)، عرض جغرافیایی (تأثیر افزایشی)، بعد خانوار (تأثیر کاهشی)، تحصیلات ساکنان (تأثیر افزایشی)، طول جغرافیایی (تأثیر افزایشی)، پوشش گیاهی و سهم فضای سبز موجود در مناطق و نواحی مورد مطالعه (تأثیر کاهشی). در انتها، پیشنهاد می‌شود مدل عرضه‌شده در کلان‌شهرهای دیگر، همچون مشهد و اصفهان و شیراز، نیز اجرا شود و الگوی مصرف آب در کلان‌شهرها، به‌منظور مدیریت بهینه این ماده حیاتی، به‌کار رود. همچنین با توجه به قابلیت بالای مدل پیشنهادی، به‌کارگیری آن در بهینه‌سازی دیگر منابع انرژی، مانند الکتریسیته و سوخت، چه بسا نتایج مفیدی در پی داشته باشد.

۷- منابع

بیران، ص.، ۱۳۸۷، **بحران وضعیت آب در جهان و ایران**، راهبرد، دوره ۱۶، ش ۴۸، صص. ۲۱۲-۱۹۳.

حسین‌زاده، ل.، ۱۳۸۶، **دسته‌بندی مشتریان هدف در صنعت بیمه با استفاده از داده‌کاوی**، پایان‌نامه کارشناسی ارشد مدیریت فناوری اطلاعات، دانشگاه تربیت مدرس.

صالحی، م.، باوی، ا.، ۱۳۸۹، **الگوریتم‌های ژنتیک و بهینه‌سازی سازه‌های مرکب**، تهران، انتشارات عابد.

طالقانی، م.، ۱۳۸۵، **اطلس کامل تهران ۸۵**، مؤسسه جغرافیایی و کارتوگرافی گیتاشناسی.

Cabena, P., Hadjinian, P., Stadler, R., Verhees, J & Zanasi, A., 1998, **Discovering Data Mining: From Concept to Implementation**, Prentice Hall, Upper Saddle River, NJ.

Dempster, A.P., 1967, **Upper and Lower Probabilities Induced by a Multivalued Mapping**, Annals of Mathematical Statistics, 38, PP. 325-339.

متغیر عرض جغرافیایی با مصرف آب شهری رابطه‌ای مستقیم را نشان می‌دهد.

سطح تحصیلات ساکنان، به‌علت رابطه‌اش با میزان درآمد و سطح رفاهی افراد، اثر افزایشی در مصرف سرانه آب دارد. زیرا ارتقای نسبی سطح تحصیلات معمولاً موجب افزایش درآمد خانوار می‌شود و در عوامل دیگر، همچون بعد خانوار و تراکم جمعیت در واحد سطح، تأثیر می‌گذارد. بنابراین و در مجموع، میزان درآمد بیشتر ساکنان دارای سطح تحصیلات بالاتر می‌تواند از علل اصلی افزایش مصرف آب باشد.

فصل‌های سال، که عامل زمانی مهم در مصرف آب در نظر گرفته شده‌اند، تأثیر یکسانی در مصرف آب تمامی مناطق و نواحی دارند. با توجه به شکل‌های ۴ تا ۷، روند افزایش مصرف آب در فصل تابستان، به‌علت افزایش دمای هوا و استفاده از کولرهای آبی، در تهران به‌خوبی تشخیص داده می‌شود و به‌ترتیب، فصل‌های پاییز و زمستان و بهار در رده‌های بعدی مصرف قرار می‌گیرند.

از عوامل فنی دیگر نیز می‌توان فشار آب موجود در شبکه آب‌رسانی را ذکر نمود که این عامل نیز به‌صورت یکسان در تمامی نواحی تأثیرگذار است و کاهش فشار آب موجود در شبکه، موجب کاهش آب مصرفی می‌شود.

۶- نتیجه‌گیری

در این تحقیق، یک مدل ترکیبی از الگوریتم‌های ژنتیک و مجموعه‌های راف، برای افزایش کارایی قوانین استخراج‌شده از داده‌های مصرفی آب در نواحی مسکونی شهر تهران، بررسی شده است. نتایج نشان می‌دهند که مدل پیشنهادی، به‌نسبت مجموعه‌های راف بدون الگوریتم ژنتیک (با دقت ۷۷٪ و میزان اطمینان ۳۸٪)، دارای دقت طبقه‌بندی (۹۲٪) و میزان اطمینان (۷۲٪) بیشتری است. با توجه به قوانین استخراج‌شده، عوامل تأثیرگذار در مصرف آب به‌ترتیب میزان و چگونگی تأثیرگذاری‌اند که شامل این موارد می‌شوند: جمعیت ساکن (تأثیر افزایشی)، تراکم جمعیت در واحد سطح (تأثیر کاهشی)، آب‌بها (تأثیر

- Dubois, D. & Prade, H., 1990, **Rough Fuzzy Sets and Fuzzy Rough Sets**, International Journal of General Systems, 17, PP. 191-209.
- Dubois, D. & Prade, H., 1992, **Putting Rough Sets and Fuzzy Sets Together**, in R. Slowinski, ed. Intelligent Decision Support - Handbook of Applications and Advances of the Rough Sets Theory, Dordrecht: Kluwer Academic Publishers.
- Gorsevski, P.V., Jankowski, P., 2008, **Discerning Landslide Susceptibility Using Rough Sets**, Computers, Environment and Urban Systems, 32, PP. 53-65.
- Hsue Cheng, C., 2010, **A Hybrid Model Based on Rough Sets Theory and Genetic Algorithms for Stock Price Forecasting**, Elsevier\ Information Sciences, 180, P. 20.
- Larose, D., 2005, **Discovering Knowledge (In Data An Introduction to Data Mining)**, Ebook.
- McKee, T. & Lensberg, T., 2002, **Genetic Programming and Rough Sets: A Hybrid Approach to Bankruptcy Classification**, Elsevier/Information Technology, 138, P. 16.
- Mortazavi, M., 2009, **Comparison of Multi-Objective Genetic Algorithm with Ant Colony Optimization: A Case Study for Canberra Water Supply System**, 33rd IAHR Congress: Water Engineering for a Sustainable Environment.
- Pawlak, Z., 1982, **Rough Sets**, International Journal of Computer and Information Science, 11(5), PP. 341-356.
- Pheng Khoo, L., 2002, **Feature Extraction Using Rough Set Theory and Genetic Algorithms an Application for the Simplification of Product Quality Evaluation**, Elsevier/Computers & Industrial Engineering, 43, P. 16.
- Pheng Khoo, L. & Yin Zhai, L., 2001, **A Prototype Genetic Algorithm Enhanced Rough Set Based Rule Induction System**, Elsevier/Computers & Industrial Engineering, 46, P. 12.
- Santana, L., 2010, **DEMORS: A Hybrid Multi-Objective Optimization Algorithm Using Differential Evolution and Rough Set Theory for Constrained Problems**, Elsevier/Computers & Operations Research, 37, P. 11.
- Sheng, 2008, **Modular Feature Selection Using Relative Importance Factors**, Department of Electrical and Computer Engineering National University of Singapore, 13.
- Slowinski, R. & Stefanowski, J., 1989, **Rough Classification in Incomplete Information Systems**, Mathematical & Computer Modeling, 12.
- Thangavel, 2009, **Dimensionality Reduction Based on Rough Set Theory: A Review**, Elsevier/Applied Soft Computing, 9, P. 12.
- Triantaphyllou, E. & Felici, G., 2006, **Data Mining & Knowledge Discovery Approaches Based on Rule Induction Techniques**, Springer.
- Yau Liang, W. & Che Huang, C., 2009, **The Generic Genetic Algorithm Incorporates with Rough Set Theory – An Application of the Web Services Composition**, Elsevier/Expert Systems with Application, 36, P. 8.
- Wygralak, W., 1989, **Rough Sets and Fuzzy Sets - Some Remarks on Interrelations**, Fuzzy Sets and Systems, 29, PP. 241-243.
- Yin Zhai, L., Pheng Khoo, L., Cheong Fok, S., 2006, **Knowledge Acquisition and Uncertainty in Fault Diagnosis: A Rough Sets Perspective**, Data Mining and Knowledge Discovery Approaches Based on Rule Induction Techniques, Springer, Heidelberg, PP. 359-394