

Learning Membership Functions for Mining Fuzzy Association Rules Using Learning Automata

Zohreh Anari¹, Abdolreza Hatamlou^{2*} and Mohammad Masdari³

1- Department of Computer Engineering, Urmia Branch, Islamic Azad University, Urmia, Iran.

2*- Department of Computer Engineering, Khoy Branch, Islamic Azad University, Khoy, Iran.

3- Department of Computer Engineering, Urmia Branch, Islamic Azad University, Urmia, Iran.

¹ zanari323@yahoo.com, ^{2*} hatamlou@iaukhoy.ac.ir, and ³ M. Masdari @ iaurmia.ac.ir

Corresponding author's address: Abdolreza Hatamlou, West Azarbaijan Province, Khoy, Islamic Azad University, Khoy Branch, Faculty of Computer Engineering.

Abstract- Transactions in web data are often quantitative data and fuzzy set theory can be used to represent such data. The time spent by users on web pages is one of the types of data stored in web log files, which can be used as an important factor in analyzing users' browsing behavior. In all existing researches for mining fuzzy association rules in web usage data, the number and parameters of membership functions for the time parameter are assumed to be constant across all web pages. However, the number and parameters of the membership functions used for each web page are different from other web pages, which can influence the behavioral analysis of users. Therefore, to address this challenge, in this paper, a reinforcement-based optimization approach using learning automata (LA) called LA-OMF is proposed. The proposed method is able to simultaneously find the number and parameters of trapezoidal membership functions to derive fuzzy association rules in web data. A new heuristic has also been proposed to increase the convergence speed of the proposed method and eliminate inappropriate membership functions. The performance of the proposed algorithm was compared to FWMA, MAOMF and FTARM + PFGM algorithms on the real dataset. The results of experiments on datasets of different sizes show that the proposed LA-OMF by extracting the optimized membership functions increased the average efficiency of the cost function and fuzzy support.

Keywords- Web usage mining, Learning automata, Fuzzy set, Fuzzy association rule, Trapezoidal membership function.

یادگیری توابع عضویت برای کاوش قوانین انجمنی فازی با استفاده از آتوماتای یادگیر

زهره اناری^۱، عبدالرضا حاتم لو^{۲*}، محمد مصدری^۳

۱- دانشکده مهندسی کامپیوتر، واحد ارومیه، دانشگاه آزاد اسلامی، ارومیه، ایران.

۲- دانشکده مهندسی کامپیوتر، واحد خوی، دانشگاه آزاد اسلامی، خوی، ایران.

۳- دانشکده مهندسی کامپیوتر، واحد ارومیه، دانشگاه آزاد اسلامی، ارومیه، ایران.

¹zanari323@yahoo.com, ²hatamlou@iaukhoy.ac.ir, and ³m.masdari@iaurmia.ac.ir

* نشانی نویسنده مسئول: عبدالرضا حاتم لو، آذربایجان غربی، خوی، دانشگاه آزاد اسلامی واحد خوی، دانشکده مهندسی کامپیوتر.

چکیده- تراکنش‌ها در داده‌های وب، اغلب داده‌های کمی بوده و تئوری مجموعه‌های فازی می‌تواند برای نشان دادن چنین داده‌هایی مورد استفاده قرار گیرد. زمان سپری شده توسط کاربران در صفحات وب، یکی از انواع داده‌های ذخیره شده در فایل‌های لاگ وب بوده که می‌توان از آن به عنوان یک عامل مهم برای تحلیل رفتار حرکتی کاربران استفاده کرد. در تمامی کارهای گزارش شده قبلی در زمینه‌ی کاوش قوانین انجمنی فازی در داده‌ی استفاده از وب، تعداد و پارامترهای توابع عضویت برای پارامتر زمان، در تمام صفحات وب ثابت فرض شده است. این در حالی است که تعداد و پارامترهای توابع عضویت مورد استفاده برای هر صفحه وب با سایر صفحات وب متفاوت بوده که این موضوع می‌تواند بر تحلیل رفتار حرکتی کاربران تأثیرگذار باشد. بنابراین برای حل این چالش، در این مقاله، یک روش مبتنی بر یادگیری تقویتی با استفاده از آتوماتای یادگیر، بنام LA-OMF پیشنهاد شده است. روش پیشنهادی قادر است همزمان، تعداد و پارامترهای توابع عضویت دوزنقه‌ای را برای استخراج قوانین انجمنی فازی، پیدا نماید. همچنین برای افزایش سرعت همگرایی روش پیشنهادی و حذف توابع عضویت نامناسب، هیوریستیک جدیدی ارائه شده است. کارایی الگوریتم پیشنهادی در مقایسه با الگوریتم‌های MAOMF، FWMA و FTARM+PFGM بر روی مجموعه داده‌ی واقعی مورد ارزیابی قرار گرفت. نتایج آزمایش‌ها بر روی مجموعه داده با اندازه‌های مختلف نشان می‌دهد که روش پیشنهادی LA-OMF با استخراج توابع عضویت بهینه، میانگین کارایی تابع هزینه و پشتیبان فازی را افزایش می‌دهد.

واژه‌های کلیدی: کاوش استفاده از وب، آتوماتای یادگیر، مجموعه فازی، قوانین انجمنی فازی، توابع عضویت دوزنقه‌ای

۱- مقدمه

برای یافتن و استخراج الگوهای مفید از داده‌های وب در نظر گرفت [۳]. کاوش وب را می‌توان به سه دسته به نام‌های کاوش استفاده از وب، کاوش محتوای وب و کاوش ساختار وب تقسیم کرد. کاوش وب از تکنیک‌های داده‌کاوی برای کشف الگوهای معنی‌دار از لاگ‌های وب به عنوان تعامل کاربر با وب سایت‌ها استفاده می‌کند. کاوش محتوای وب، فرایند کشف اطلاعات مفید از محتوای صفحات وب، اسناد و سرویس‌ها است. کاوش ساختار وب برای کشف دانش از سایت‌های وب و لینک‌ها در وب استفاده

شبکه جهانی وب (WWW) شامل داده‌های عظیمی بوده و به سرعت در حال رشد است. برای یافتن اطلاعات مناسب از میان تعداد زیادی از داده به زمان زیادی نیاز است [۱]. بنابراین، تکنیک‌های مؤثری برای تجزیه و تحلیل داده‌ها مورد نیاز است. کاوش وب نقش مهمی در کشف چنین دانشی ایفا می‌کند [۲]. کاوش وب را می‌توان به عنوان روش‌ها و تکنیک‌های داده‌کاوی

متعدد، یک روش کاوش فازی وب ارائه نموده تا بتواند الگوهای حرکتی وب را از اصطلاحات زبانی^۲ با سطوح و دیدگاه‌های متنوع کشف کند.

با این حال، مشکل اصلی در تمام مطالعات قبلی انجام شده در مورد الگوریتم‌های کاوش فازی وب این است که تعداد توابع عضویت و پارامترهای آنها برای هر متغیر زبانی توسط ایده‌های کارشناسان خبره مشخص شده است. به عبارت دیگر، آنها فرض می‌کنند که تعداد و پارامترهای توابع عضویت، ثابت و از پیش تعریف شده که در این صورت، ممکن است برخی از نتایج کاوش، معتبر برای هر صفحه‌ی وب نباشند. علاوه بر این، توابع عضویت مختلف ممکن است، دانش متفاوتی را استنتاج کند. بنابراین، یافتن توابع عضویت مناسب برای استخراج قوانین فازی یک کار مهم بوده که دلیل آن دو مورد است: (الف) داشتن دانش قبلی در به دست آوردن توابع عضویت مناسب، که تمام متغیرهای دامنه را در برگیرد، آسان نیست و (ب) داده‌ای که برای یک صفحه‌ی وب مناسب است، ممکن است برای صفحه‌ی وب دیگر مناسب نباشد.

در این مقاله، به حل مسأله‌ی فوق پرداخته و یک روش جدید مبتنی بر اتوماتای یادگیر، برای بهینه‌سازی خودکار تعداد و پارامترهای توابع عضویت دوزنقه‌ای، برای کاوش قوانین انجمنی فازی از داده‌های وب پیشنهاد شده است. اتوماتای یادگیر به عنوان ابزار بهینه‌سازی در کاربردهای مختلفی از جمله شبکه‌های کامپیوتری [۴۲]، پردازش تصویر [۴۳]، بهینه‌سازی توابع عضویت فازی [۴۴]، تشخیص گفتار [۴۵]، تجزیه و تحلیل سیگنال [۴۶]، بازیابی اطلاعات [۴۷]، محاسبات ابری [۴۸]، بهینه‌سازی پویا [۴۹] و خوشه‌بندی داده‌ها مورد استفاده قرار گرفته است [۵۰].

برای انجام این کار، زمان صرف شده توسط کاربران در هر صفحه‌ی وب به عنوان یک متغیر فازی با توابع عضویت دوزنقه‌ای در نظر گرفته شده است. همچنین از مراکز و پهنای هر تابع عضویت به عنوان پارامترهای مسأله‌ی بهینه‌سازی استفاده شده است. برای یافتن مقدار بهینه‌ی این پارامترها، یک روش جدید با استفاده از اتوماتای یادگیر پیشنهاد شده است. این چارچوب تعداد توابع عضویت (اصطلاحات زبانی) را به عنوان ورودی گرفته و تیمی از اتوماتای یادگیر را می‌سازد. هر یک از این پارامترها به چارچوب پیشنهادی، ما یک متد برای یافتن خودکار تعداد و پارامترهای توابع عضویت توسعه دادیم.

به طور خلاصه اهداف این مقاله به صورت زیر است:

- با توجه به مطالعات صورت گرفته شده، این اولین کاری

می‌کند. دانش استخراج شده از داده‌های وب را می‌توان در بسیاری از زمینه‌ها مانند شخصی‌سازی وب [۷-۴]، سیستم‌های توصیه‌کننده [۸-۱۰]، خوشه‌بندی [۱۱، ۱۲]، کاوش الگوهای ترتیبی [۱۳-۱۶] و کاوش قوانین انجمنی استفاده [۱۷، ۱۸] کرد. استخراج قوانین انجمنی یکی از تکنیک‌های مهم داده‌کاوی بوده که از رکوردهای لاگ وب برای این منظور استفاده می‌کند [۱۵، ۱۸-۲۱]. سادگی استفاده از مجموعه‌های فازی و شباهت آنها به استنتاج انسانی، استفاده از آنها را در زمینه‌های مختلف، از جمله کاوش الگوهای ترتیبی، استخراج قوانین، داده‌کاوی، استخراج دانش بیولوژیکی، تشخیص نفوذ، مهندسی و تشخیص، آسان می‌کند [۲۲-۳۲]. داده‌های وب اغلب از داده‌های کمی، مبهم و زبانی تشکیل شده که در نتیجه می‌توان از منطق فازی در این زمینه استفاده کرد [۳۳-۳۷].

قوانین انجمنی با استفاده از داده‌ی استفاده از وب، برای دستیابی به کشف روابط بین صفحات وب استفاده می‌شود. به عنوان مثال، اگر یک تراکنش شامل مجموعه‌ای از صفحات وب باشد، در حالی که هر صفحه‌ی وب یک آدرس مستقلی دارد، الگوی حرکتی در لاگ فایل ثبت شده و دانش استخراج شده می‌تواند به شرح زیر باشد: ۴۵٪ کاربرانی که صفحه وب با آدرس `/main/food-beverage/milk.html` را ملاقات کرده‌اند، همچنین صفحه وب `/main/food-beverage/butter.html` را نیز ملاقات خواهند کرد. الگوریتم‌های کاوش فازی وب زیادی برای استخراج قوانین انجمنی یا الگوهای حرکتی پیشنهاد شده است. هونگ و همکاران [۳۸] الگوریتم استخراج فازی وب را با استفاده از مفاهیم فازی و شی گرا ارائه دادند. در روش آنها، هر صفحه وب به عنوان یک کلاس و هر صفحه وب ملاقات شده توسط کاربر به عنوان یک نمونه در نظر گرفته شد. روش آنها شامل دو مرحله بود. در مرحله اول، قوانین انجمنی از صفحات وب تولید و در مرحله دوم، روابط بین تمام صفحات وب مشخص می‌شدند. هونگ و همکاران [۳۹]، با توجه به زمان سپری شده توسط کاربران در هر صفحه‌ی وب، اهمیت هر صفحه‌ی وب و حداقل پشتیبان، یک روش کاوش فازی وب را توسعه دادند. علاوه بر این، متیو و همکاران [۴۰]، یک الگوریتم کاوش فازی مبتنی بر الگوریتم ژنتیک برای مسأله‌ی از دست دادن قوانین انجمنی فازی موقتی از لاگ های وب معرفی کردند. آنها با تبدیل مجموعه‌ی داده به یک گراف و تغییر تابع شایستگی^۱، کار قبلی خود را بهبود دادند. نتایج شبیه‌سازی در مجموعه داده‌ی واقعی نشان داد که روش آنها می‌تواند قوانین معتبر بیشتری را استخراج کند. ونگ و همکاران [۴۱] با استفاده از طبقه‌بندی صفحات وب به سطوح

که الگوریتم آنها می‌تواند همزمان فضای جستجو و زمان اجراء را کاهش دهد.

لین و همکاران [۵۲]، یک ساختار لیست فازی با طراحی مجموعه اقلام مکرر فازی برای کاوش الگوهای مکرر فازی پیشنهاد دادند. الگوریتم طراحی شده (MFFI-Miner) به طور کارا چندین مجموعه از اقلام مکرر فازی را بدون تولید کاندیداها بر اساس ساختار لیست فازی طراحی شده، استخراج می‌کرد. همچنین استراتژی‌هایی برای هرس کارآمد ساختار درخت برای کاهش فضای جستجوی و کاهش محاسبات مکانیسم تولید و تست مرحله‌ی کاوش با استفاده از ساختار لیست فازی ارائه نموده‌اند. چن و همکاران [۵۳] یک الگوریتم داده کاوی فازی را با استفاده از الگوریتم ژنتیک، برای به دست آوردن توابع مناسب عضویت و کاوش الگوهای مفهومی معرفی کرده‌اند. آنها از یک طرح نمایش دوتایی برای متغیرهای زبانی به منظور رمزگذاری توابع عضویت از اقلام داده به کروموزومها استفاده کرده‌اند. آنها ادعا کردند که روش آنها در مقایسه با سایر روش‌هایی که از توابع عضویت از پیش تعریف شده استفاده می‌کنند، الگوهای مفهومی بیشتری را تولید می‌کند.

علاوه بر این، رودزیکنی [۵۴] یک روش بهینه‌سازی چند هدفه برای سیستم‌های طبقه‌بندی بر اساس قانون فازی با استفاده از الگوریتم‌های ژنتیک ارائه کرد. او از دو معیار برای ارزیابی صحت و تفسیرپذیری استفاده کرد.

چن و همکاران [۵۵] یک روش کاوش فازی را توسعه داده و یک نمایش کروموزومی جدید برای یافتن تعداد مناسب توابع عضویت و پارامترهای بهینه‌ی آنها ارائه نمودند. در این روش، هر کروموزوم مجموعه‌ای از توابع عضویت را نشان داده و شامل دو بخش بود. بخش اول فعال‌سازی هر تابع عضویت را تعیین کرده و با رشته‌های باینری نمایش داده می‌شد. بخش دوم پارامترهای بهینه‌سازی شده مربوط به توابع عضویت را مشخص می‌کرد. آنها با استفاده از عملگرهای ژنتیکی و بررسی وضعیت‌های مختلف توابع عضویت غیر فعال و فعال، بهترین مجموعه از توابع عضویت را بدست آوردند.

وو و همکاران [۵۶] برای یافتن توابع مناسب از اقلام عددی، از یک نمایش رمزگذاری استفاده و با معرفی برخی از عملگرهای معین، راه حل بهینه‌ی سرتاسری واقعی را در فضای پیوسته جستجو کردند. روش پیشنهادی آنها یال‌ها و نودهای ثابتی در فرآیند جستجو نداشت. آنها با استفاده از توزیع فرمون‌ها در فضای راه حل، یال‌های جستجو را به صورت پویا در گراف ایجاد کردند.

است که تلاش می‌کند به طور همزمان تعداد و پارامترهای توابع عضویت دوزنقه‌ای را به صورت پویا برای کاوش قانون انجمنی فازی پیدا کند.

- یک روش جدید به نام LA-OMF برای یافتن تعداد بهینه‌ی توابع عضویت دوزنقه‌ای و همچنین پارامترهای آنها در الگوریتم‌های کاوش فازی وب پیشنهاد کردیم.
- در روش پیشنهادی، برای بهبود سرعت همگرایی و کاهش فضای جستجو، دو هیوریستیک برای حذف توابع عضویت نامناسب در طی فرایند یادگیری پیشنهاد شده است. علاوه بر این، ما یک روش جدید برای پیاده سازی هیوریستیک‌ها توسعه داده‌ایم.
- تابع هزینه با در نظر گرفتن توابع عضویت دوزنقه‌ای تغییر داده شده است.
- کارایی بالای چارچوب و روش پیشنهادی در مورد کاوش قوانین انجمنی فازی برای کاوش استفاده از وب بر اساس آزمایش‌های انجام شده بر روی مجموعه‌ی داده با اندازه‌های مختلف، تأیید شده است.

ادامه مقاله به صورت زیر سازماندهی شده است: در بخش ۲ مروری بر پژوهش‌های انجام شده در زمینه کاوش قوانین انجمنی فازی و در بخش ۳ پیشینه‌ی لازم از جمله اتوماتای یادگیر آورده شده است. در بخش ۴، الگوریتم پیشنهادی LA-OMF شرح داده شده است. در بخش ۵ مجموعه‌ی داده‌ی استفاده شده و نتایج آزمایش نشان داده شده است. در انتها بخش ۶ بیان نتیجه‌گیری و پیشنهاد برای کارهای آتی است.

۲- مروری بر کارهای پیشین

بهینه‌سازی توابع عضویت، تأثیر مهمی در کشف قوانین انجمنی فازی دارد. چندین الگوریتم کاوش فازی برای استخراج بهینه‌ی توابع عضویت مثلثی پیشنهاد شده است. این الگوریتم‌ها می‌توانند به طور پویا توابع عضویت را با استفاده از روش‌های فرا اکتشافی تنظیم و از آنها برای تبدیل تراکنش‌های عددی به مقادیر فازی استفاده کنند.

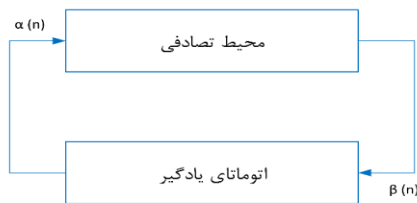
پالاسیوس و همکاران [۵۱] یک روش دو مرحله‌ای برای یافتن خودکار توابع عضویت مناسب و قوانین انجمنی فازی از داده‌ی مبهم (غیر دقیق) ارائه داده‌اند. در مرحله‌ی اول، آنها با استفاده از یک الگوریتم ژنتیک، بهترین مجموعه از توابع عضویت را استخراج و در مرحله دوم، با استفاده از الگوریتم کاوش الگوی رشد مکرر فازی، قوانین انجمنی فازی را استخراج کرده‌اند. آنها نشان دادند

استفاده از وب به همراه توابع عضویت از پیش تعریف شده استفاده می‌کرد.

۳- پیشینه

۳-۱- آتوماتای یادگیر

آتوماتای یادگیر^۵ یک مدل ریاضی بوده که می‌تواند تعداد محدودی عمل^۶ را انجام دهد. در این مدل، هر آتوماتای یادگیر سعی می‌کند با استفاده از تعامل با یک محیط تصادفی، عمل بهینه‌ی خود را از بین مجموعه‌ی اعمال پیدا کند [۶۴، ۶۵]. هرگاه این ماشین عملی را انتخاب کند، عمل انتخاب شده توسط محیط تصادفی ارزیابی و نتیجه‌ی آن به صورت یک سیگنال بازخوردی مثبت (در صورت مناسب عمل بودن) یا منفی (در صورت نامناسب بودن عمل) به آتوماتا بازگردانده می‌شود. مقدار این سیگنال در انتخاب عمل بعدی تأثیر می‌گذارد. هدف این فرایند این است که آتوماتا بعد از گذشت مدتی به سمت مناسب‌ترین عمل خود در محیط میل کرده، یا به عبارت دیگر یاد می‌گیرد که کدام عمل، عمل بهینه است [۶۶، ۶۷]. نحوه تعامل آتوماتای یادگیر و محیط در شکل ۱ نشان داده شده است.



شکل ۱: تعامل آتوماتای یادگیر با محیط

مدل ریاضی محیط به سه صورت سه‌تایی $\langle \alpha, \beta, c \rangle$ تعریف شده که در آن $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_r\}$ مجموعه‌ی ورودی‌ها، $\beta = \{\beta_1, \beta_2, \dots, \beta_m\}$ مجموعه‌ی مقادیر سیگنال تقویتی که توسط محیط تولید شده و $c = \{c_1, c_2, \dots, c_r\}$ مجموعه‌ای از احتمال‌های جریمه می‌باشد. هر c_i با α_i رابطه دارد. در واقع c_i ‌ها مشخصات و رفتار محیط را تعریف می‌کنند. اگر احتمال‌های جریمه ثابت باشند، محیط را ایستا و اگر در طول زمان تغییر کند، غیر ایستا می‌نامند. آتوماتای یادگیر به دو خانواده‌ی آتوماتای یادگیر با ساختار ثابت^۷ و آتوماتای یادگیر با ساختار متغیر^۸ تقسیم می‌گردد. در ادامه، آتوماتای یادگیر با ساختار متغیر که در این مقاله از آن استفاده شده، شرح داده شده است. آتوماتای یادگیر با ساختار متغیر، توسط چهارتایی $\langle \alpha, \beta, p, T \rangle$ نشان داده شده که در آن $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_r\}$ مجموعه‌ی اعمال آتوماتای یادگیر، $\beta = \{\beta_1, \beta_2, \dots, \beta_m\}$

تینگ و همکاران [۵۷] برای یافتن توابع عضویت مناسب برای کاوش قوانین انجمنی فازی، یک الگوریتم ممتیک^۲ ارائه دادند. الگوریتم آنها برای نمایش کروموزوم‌ها، نوع ساختار توابع عضویت را در نظر می‌گرفت. آنها موفق شدند با در نظر گرفتن نوع ساختار در روش خود، توابع عضویت نامناسب را حذف و در نتیجه فضای جستجو را کاهش دهند. علاوه بر این، آنها یک اپراتور جستجوی محلی برای کاهش فضای جستجو گسترش دادند.

روش دیگری توسط تینگ و همکاران [۵۸] با استفاده از الگوریتم ژنتیک ارائه شده است. در روش آنها، هر تابع عضویت مثلثی توسط سه پارامتر نشان داده می‌شد. تمام پارامترها به عنوان کروموزوم‌ها، کدگذاری و با استفاده از عملگرهای ژنتیکی بهترین مجموعه از توابع عضویت استخراج می‌شدند. آنها همچنین دو هیوریستیک^۴ را در شکل توابع عضویت توسعه دادند. هیوریستیک‌ها اشکال نامناسب توابع عضویت را حذف و همچنین فضای جستجو برای یافتن توابع عضویت مناسب را کاهش می‌داد. سونگ و همکاران [۵۹] یک الگوریتم خفاش فازی را برای کاوش قوانین انجمنی پیشنهاد داده تا بتوانند به طور پویا توابع عضویت را از مقادیر کمی (عددی) پیدا کنند.

چمازی و همکاران [۶۰] با ترکیب الگوریتم زنبور عسل و یک روش داده‌کاوی موقتی فازی، الگوریتم استخراج موقتی فازی را معرفی کردند. هدف از این کار استخراج توابع عضویت مناسب و استخراج قوانین انجمنی موقتی فازی بود. علی‌خادمی و همکاران در [۶۱] با استفاده از مجموعه‌های فازی و بهینه‌سازی ذرات، چارچوبی را برای استخراج توابع عضویت مناسب برای کشف قوانین انجمنی فازی ارائه دادند.

هونگ و همکاران [۶۲] یک الگوریتم کاوش فازی مبتنی بر الگوریتم ژنتیک ارائه داده تا توابع عضویت مناسب و قوانین انجمنی فازی را پیدا کنند. پردازنده اصلی، مانند یک الگوریتم ژنتیک، مقدار تابع شایستگی را بین پردازنده‌های سرویس گیرنده تقسیم می‌کرد. هر پردازنده تابع شایستگی را ارزیابی و نتیجه خود را به پردازنده اصلی ارسال می‌کرد. سپس، پردازنده اصلی نتایج شایستگی بدست آمده از کلیه پردازنده‌ها را ارزیابی و بهترین مقدار شایستگی را به عنوان یک تابع عضویت مناسب انتخاب می‌کرد. سرانجام، بهترین مجموعه از توابع عضویت به دست آمده توسط پردازنده اصلی در الگوریتم‌های داده‌کاوی فازی برای استخراج قوانین مفید مورد استفاده قرار می‌گرفت.

پاتیل و همکاران [۶۳] یک الگوریتم کاوش فازی وب تعمیم یافته برای استخراج قوانین انجمنی جالب معرفی کردند. الگوریتم آنها از یک تابع عضویت از پیش تعیین شده برای فازی‌سازی داده‌های

به طوری که L_1 مجموعه‌ای از دنباله‌های بزرگ به طول یک بوده که از توابع عضویت بدست می‌آید. $fuzzy\ support(Y_{j,k})$ ، یعنی پشتیبان فازی از دنباله‌های بزرگ به طول یک از رکوردهای لاگ وب^{۱۲} تشکیل می‌شود. مناسب بودن توابع عضویت^{۱۳} به عنوان مجموع عوامل همپوشانی^{۱۴} و پوشش^{۱۵} توصیف می‌شود. مناسب بودن در تابع هزینه، وقوع دو نوع از توابع عضویت نامناسب را همان طوری که در شکل ۲ مشخص شده، کاهش می‌دهد. به طوری که در مورد اولی (۲-الف)، توابع عضویت خیلی از هم جدا و در دومی (۲-ب) همپوشانی خیلی زیاد است. مناسب بودن توابع عضویت برای هر صفحه وب به صورت زیر توصیف می‌شود [۶۸].

$$Suitability = \text{Overlap factor} + \text{Coverage factor} \quad (۵)$$

که در آن فاکتور همپوشانی، نشان دهنده‌ی ناحیه‌ی پوشش داده شده توسط دو تابع عضویت و فاکتور پوشش، نشان دهنده‌ی ناحیه‌ی پوشش داده شده توسط تمام توابع عضویت می‌باشد. فاکتور همپوشانی در مساله‌ی ما به صورت زیر تعریف می‌شود:

$$\text{Overlap factor} = \sum_{i < j} (\max(\text{overlap ratio}(R_i, R_j), 1) - 1) \quad (۶)$$

با

$$\text{overlap ratio}(R_i, R_j) = \frac{\text{The area covered by both } R_i \text{ and } R_j}{\min(R_{i,4} - R_{i,3}, R_{j,2} - R_{j,1})} \quad (۷)$$

نسبت همپوشانی دو تابع عضویت R_1 و R_2 به عنوان ناحیه‌ی تحت پوشش توسط دو تابع عضویت، تقسیم بر حداقل پهنا‌ی نیمه‌ی سمت راست از تابع عضویت در سمت چپ و نیمه‌ی چپ از تابع عضویت در سمت راست تعریف می‌شود. فاکتور همپوشانی برای جلوگیری از همپوشانی بیش از حد توابع عضویت استفاده می‌شود. بنابراین، فاکتور پوشش توابع عضویت برای هر صفحه‌ی وب مانند p_i به صورت زیر تعریف می‌شود:

$$\text{Coverage factor} = \frac{\max(t_i)}{\text{range}(R_1, R_2, \dots, R_m)} \quad (۸)$$

به طوری که $\text{range}(R_1, R_2, \dots, R_m)$ طول محور افقی پوشش داده شده برای توابع عضویت صفحه‌ی وب p_i ، $\max(t_i)$ حداکثر مدت زمان صفحه‌ی وب p_i در لاگ داده و m تعداد نواحی فازی (تعداد توابع عضویت) است. بهترین مقدار فاکتور پوشش برای پوشش کامل مدت زمان صفحه وب برابر ۱ است. در قسمت پیوست مثالی برای بهینه‌سازی توابع عضویت آورده شده است.

$p = \{p_1, p_2, \dots, p_r\}$ بردار احتمال انتخاب هر یک از عمل‌ها و T الگوریتم یادگیری بوده که براساس پاسخ دریافتی از محیط بردار احتمال عمل اتوماتا را مطابق با پاسخ دریافتی از محیط بروزرسانی می‌کند. همچنین، نشان دهنده‌ی تعداد اعمالی بوده که توسط اتوماتا قابل انتخاب است. فرض کنید $\alpha_i(n) \in \alpha$ و $p(n)$ به ترتیب نشان دهنده‌ی عمل انتخاب شده توسط اتوماتای یادگیر در لحظه‌ی t و بردار احتمال اعمال بر روی مجموعه‌ی اعمال باشد. در لحظه‌ی t ، زمانی که عمل انتخاب شده $\alpha_i(n)$ توسط محیط پاداش داده شود، بردار احتمال $p(n)$ توسط الگوریتم یادگیر خطی، مطابق رابطه‌ی ۱ بروزرسانی می‌شود. در حالت انتخاب جریمه نیز این بروزرسانی توسط معادله‌ی ۲ انجام می‌شود. در معادلات ۱ و ۲ $a \geq 0$ و $b \geq 0$ پارامترهای پاداش و جریمه بوده که به ترتیب، مقدار افزایش کاهش احتمال اعمال را مشخص می‌کنند.

$$p_j(n+1) = \begin{cases} p_j(n) + a[1 - p_j(n)] & j = i \\ (1-a)p_j(n) & \forall j \neq i \end{cases} \quad (۱)$$

$$p_j(n+1) = \begin{cases} (1-b)p_j(n) & j = i \\ \left(\frac{b}{r-1}\right) + (1-b)p_j(n) & \forall j \neq i \end{cases} \quad (۲)$$

برای $a=b$ ، معادلات (۱) و (۲) الگوریتم L_{R-P} نامیده می‌شود؛ زمانی که $a \gg b$ باشد، معادلات داده شده به عنوان $L_{R-\epsilon p}$ و اگر $b=0$ باشد الگوریتم را L_{R-I} می‌نامند.

۴- الگوریتم پیشنهادی (LA-OMF)

۴-۱- تابع هزینه

برای کاوش قوانین انجمنی فازی و به منظور جستجوی توابع عضویت بهینه، از یک تابع ارزیابی به نام تابع هزینه^{۱۶} استفاده می‌شود. در این مقاله، ما تابع ارزیابی پیشنهادی توسط هونگ و همکاران [۶۸] که در توابع مثلثی کاربرد دارد را تغییر داده تا بتوانیم از آن در ارزیابی توابع عضویت دوزنقه‌ای مناسب استفاده کنیم. تابع هزینه از پشتیبان فازی^{۱۷} و همچنین مناسب بودن^{۱۱} تشکیل شده است. برای هر صفحه‌ی وب، تابع هزینه به صورت زیر توصیف می‌شود:

$$\text{Cost function} = \frac{1}{\text{objective function}} \quad (۳)$$

$$\text{Objective function} = \frac{\sum_{X \in L_1} \text{fuzzy support}(Y_{j,k})}{\text{Suitability}} \quad (۴)$$

برای R_i فرض می‌کنیم که $wl_{i,j}$ نشان دهنده پهنای سمت چپ ناحیه‌ی فازی R_i ، $cl_{i,j}$ نشان دهنده‌ی مرکز سمت چپ ناحیه‌ی فازی R_i ، $cr_{i,j}$ مرکز سمت راست ناحیه‌ی فازی R_i و $wr_{i,j}$ نشان دهنده‌ی پهنای سمت راست ناحیه‌ی فازی R_i است. در روش پیشنهادی، یک آتوماتای یادگیر به هر پارامتر از چهارتایی $(wl_{i,1}, cl_{i,2}, cr_{i,3}, wr_{i,4})$ اختصاص داده می‌شود. این آتوماتای یادگیر به ترتیب با برچسب‌های شماره ۱- $A_{wl_{i,1}}^i$ ، ۲- $A_{cl_{i,2}}^i$ ، ۳- $A_{cr_{i,3}}^i$ ، ۴- $A_{wr_{i,4}}^i$ شماره گذاری می‌شوند. بنابراین، هر تابع عضویت R_i برای بهینه‌سازی پارامترهای خود به چهار آتوماتا نیاز دارد. هر آتوماتا به طول r بوده، که r تعداد اعمال هر آتوماتا را نشان می‌دهد. برای اولین و آخرین تابع عضویت که با دو پارامتر مشخص می‌شوند، فقط دو آتوماتا نیاز است.

فرض کنید K_{max} تعداد توابع عضویت در نظر گرفته شده از کاربر به عنوان ورودی باشد. برای بهینه‌سازی این پارامترها، تیمی از آتوماتای یادگیر به تعداد $4 - 4 \times K_{max}$ ساخته می‌شود. این مقاله، تیم آتوماتای یادگیر به صورت زیر است:

$$LA = \left\{ \begin{array}{l} A_{cr_{1,3}}^1, A_{wr_{1,4}}^1, \dots, A_{wl_{i,1}}^i, A_{cl_{i,2}}^i, A_{cr_{i,3}}^i, A_{wr_{i,4}}^i, \dots, \\ A_{wlm-1,1}^{m-1}, \\ A_{clm-1,2}^{m-1}, A_{crm-1,3}^{m-1}, A_{wrm-1,4}^{m-1}, A_{wlm,1}^m, A_{clm,2}^m \end{array} \right\}$$

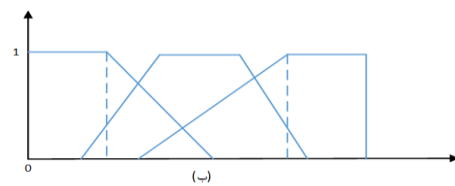
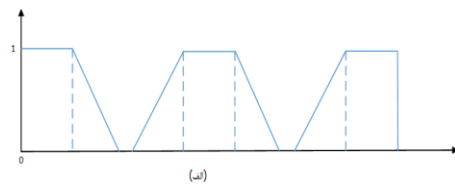
۴-۴- تولید توابع عضویت مناسب

ناحیه‌ی فازی توابع عضویت دوزنقه‌ای با چهار پارامتر مشخص می‌شوند. توابع عضویت دوزنقه‌ای باید دو قانون را برآورده کنند. این قوانین پارامترها را مطابق با شکل توابع عضویت تنظیم کرده که به ترتیب، در معادلات ۹ و ۱۰ توصیف شده‌اند.

$$cr_{i,3} \leq cl_{i+1,2} \leq cr_{i+1,3} \leq \dots \leq cl_{m-1,2} \quad (9)$$

$$\begin{aligned} &\leq cr_{m-1,3} \leq cl_{m,2} \\ &wl_{i,1} \leq cl_{i,2} \leq cr_{i,3} \leq wr_{i,4} \end{aligned} \quad (10)$$

قانون اول ترتیب مراکز توابع عضویت را تنظیم و قانون دوم شکل دوزنقه‌ای را حفظ می‌کند. شکل ۵ یک تابع عضویت دوزنقه‌ای با سه ناحیه را نشان داده که قوانین فوق را برآورده می‌کنند. مناسب بودن توابع عضویت شامل دو بخش است: همپوشانی و پوشش. همپوشانی ناحیه تحت پوشش، توسط دو تابع عضویت را نشان می‌دهد. پوشش، ناحیه تحت پوشش توسط همه توابع عضویت را نشان می‌دهد. با در نظر گرفتن این دو محدودیت، فضای جستجو به مقدار قابل توجهی کاهش می‌یابد. معادلات ۱۱ و ۱۲ پوشش کامل را برآورده کرده، همچنین معادله ۱۳ نیز همپوشانی مناسب را برآورده می‌کند.



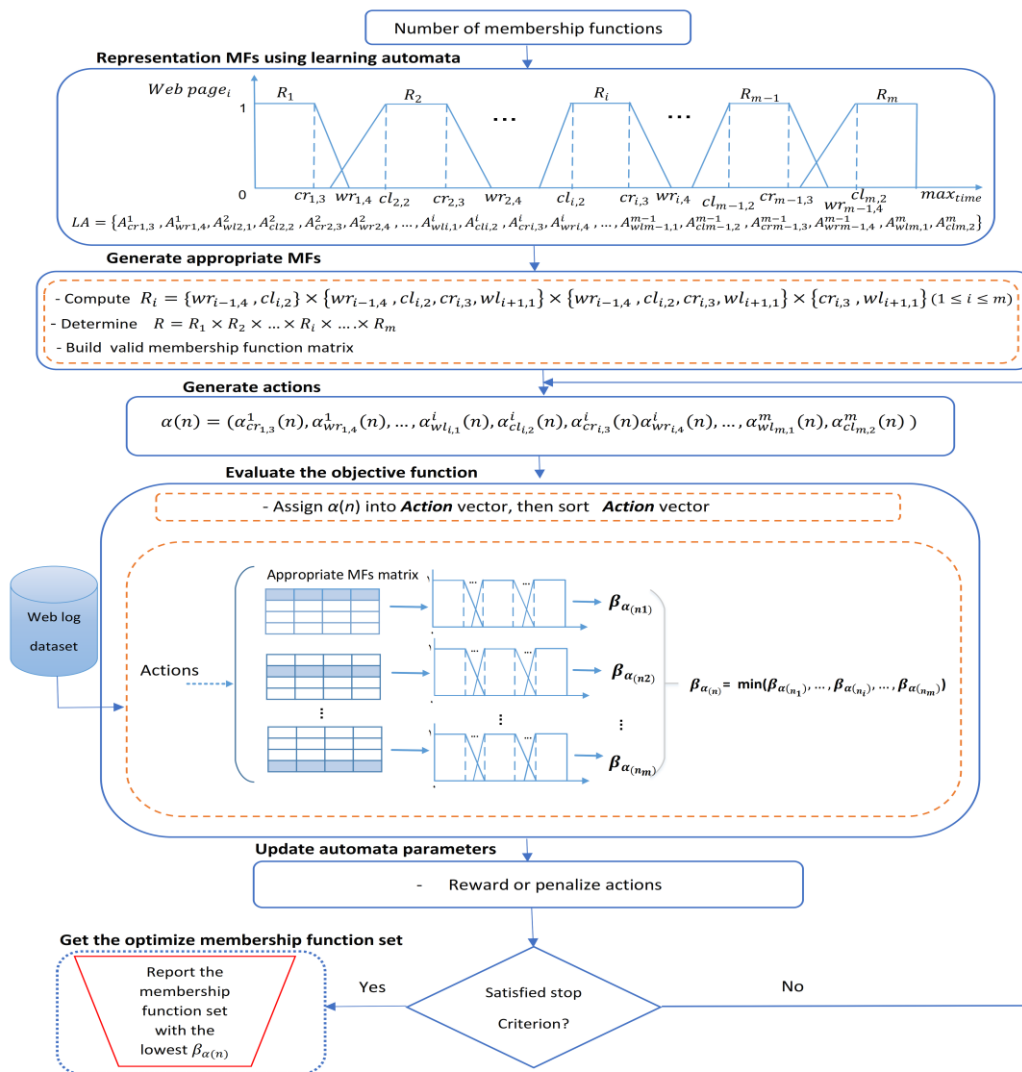
شکل ۲: دو نوع از توابع عضویت نامناسب

۴-۲- چارچوب پیشنهادی

مراکز و پهنای توابع عضویت دوزنقه‌ای پارامترهای فضای جستجو بوده و یافتن توابع عضویت دوزنقه‌ای مناسب به عنوان یک مسأله‌ی بهینه‌سازی در نظر گرفته می‌شود. در چارچوب پیشنهادی، تعداد توابع عضویت به عنوان ورودی گرفته شده و تیمی از آتوماتای یادگیر ساخته می‌شود. پس از تعیین تیمی از آتوماتای یادگیر، روند یادگیری تکرار شده تا زمانی که حداقل مقدار تابع هزینه پیدا شود. علاوه بر این، در طی فرآیند یادگیری، رهیافت‌های پیشنهادی به حذف توابع عضویت نامناسب کمک می‌کنند. سرانجام، پارامترهای بهینه برای تابع عضویت داده شده بدست می‌آیند. چارچوب پیشنهادی در شکل ۳ نشان داده شده و شامل مراحل زیر است: (۱) نمایش توابع عضویت با استفاده از آتوماتای یادگیر (۲) تولید توابع عضویت مناسب (۳) تولید اعمال آتوماتا (۴) ارزیابی تابع هزینه (۵) بروزرسانی پارامترهای آتوماتای یادگیر و (۶) استخراج تعداد بهینه توابع عضویت. هر مرحله در زیر با جزئیات توضیح داده شده است.

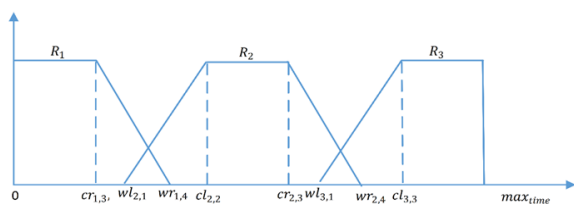
۴-۳- نمایش توابع عضویت با استفاده از آتوماتای یادگیر

زمان صرف شده توسط کاربران در هر صفحه‌ی وب یکی از پارامترهایی است که می‌تواند برای تحلیل رفتار حرکتی کاربران استفاده شود. این پارامتر به صورت توابع عضویت دوزنقه‌ای نشان داده می‌شود. فرض کنید $R = \{R_1, R_2, \dots, R_m\}$ نشان دهنده‌ی مجموعه‌ی توابع عضویت دوزنقه‌ای و m تعداد توابع عضویت باشد. R_i ، $(1 \leq i \leq m)$ تابع عضویت i امین ناحیه فازی و R_m تابع عضویت آخرین ناحیه فازی را نشان می‌دهد. هر R_i توسط چهارتایی $(wl_{i,j}, cl_{i,j}, cr_{i,j}, wr_{i,j})$ توصیف شده، به طوری که اندیس‌های i و j نشان دهنده‌ی j امین پارامتر از i امین تابع عضویت است. به عنوان مثال، در شکل ۴ ناحیه‌ی R_2 با چهارتایی $(wl_{2,1}, cl_{2,2}, cr_{2,3}, wr_{2,4})$ نشان داده شده است. علاوه بر این،



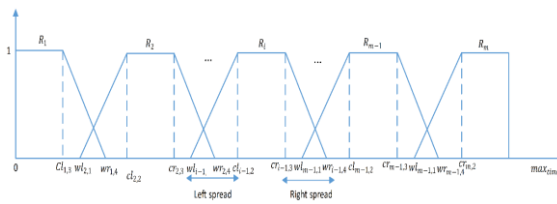
شکل ۳: چارچوب روش پیشنهادی

استفاده از محدودیت‌های ذکر شده در بالا ارائه شده است.



شکل ۵: یک تابع عضویت دوزنقه‌ای با سه ناحیه‌ی فازی

این رویکرد، توابع عضویت نامناسب را از حذف و مقادیر پارامتر بهینه را در بین توابع عضویت مناسب جستجو می‌کند. روش پیشنهادی به صورت زیر توصیف می‌شود:
هر R_i ($1 \leq i \leq m$) توسط چهارتایی $(wl_{i,j}, cl_{i,j}, cr_{i,j}, wr_{i,j})$ نشان داده می‌شود. برای اولین تابع عضویت (R_1) و آخرین تابع عضویت (R_m) ، فقط دو پارامتر $(cr_{1,3}, wr_{1,4})$ و $(wl_{m,1}, cl_{m,2})$ مورد نیاز است. فرض کنید:



شکل ۴: مدت زمان هر صفحه‌ی وب به عنوان توابع عضویت دوزنقه‌ای

$$wl_{i-1,1} \leq wl_{i,1} \leq wr_{i-1,4} \quad (11)$$

$$wl_{i+1,1} \leq wr_{i,4} \leq wr_{i+1,4} \quad (12)$$

$$wr_{i,4} \leq wl_{i+2,1} \quad (13)$$

در این تحقیق، رویکرد جدیدی برای کاهش فضای جستجو و

جدول ۱: تعداد توابع عضویت برای سه تابع عضویت دوزنقه‌ای

1	$cr_{1,3}$	$wl_{2,1}$	$wr_{1,4}$	$cl_{2,2}$	$cr_{2,3}$	$wl_{3,1}$	$wr_{2,4}$	$cl_{3,3}$
2	$cr_{1,3}$	$wl_{2,1}$	$wr_{1,4}$	$cl_{2,2}$	$cr_{2,3}$	$wl_{3,1}$	$cl_{3,3}$	$wr_{2,4}$
3	$cr_{1,3}$	$wl_{2,1}$	$wr_{1,4}$	$cl_{2,2}$	$wl_{3,1}$	$cr_{2,3}$	$wr_{1,4}$	$cl_{3,3}$
4	$cr_{1,3}$	$wl_{2,1}$	$wr_{1,4}$	$cl_{2,2}$	$wl_{3,1}$	$cr_{2,3}$	$cl_{3,3}$	$wr_{2,4}$
5	$cr_{1,3}$	$wl_{2,1}$	$wr_{1,4}$	$wl_{3,1}$	$cl_{2,2}$	$cr_{2,3}$	$wr_{2,4}$	$cl_{3,3}$
6	$cr_{1,3}$	$wl_{2,1}$	$wr_{1,4}$	$wl_{3,1}$	$cl_{2,2}$	$cr_{2,3}$	$cl_{3,3}$	$wr_{2,4}$
7	$cr_{1,3}$	$wl_{2,1}$	$cl_{2,2}$	$wr_{1,4}$	$cr_{2,3}$	$wl_{3,1}$	$wr_{2,4}$	$cl_{3,3}$
8	$cr_{1,3}$	$wl_{2,1}$	$cl_{2,2}$	$wr_{1,4}$	$cr_{2,3}$	$wl_{3,1}$	$cl_{3,3}$	$wr_{2,4}$
9	$cr_{1,3}$	$wl_{2,1}$	$cl_{2,2}$	$wr_{1,4}$	$wl_{3,1}$	$cr_{2,3}$	$wr_{2,4}$	$cl_{3,3}$
10	$cr_{1,3}$	$wl_{2,1}$	$cl_{2,2}$	$wr_{1,4}$	$wl_{3,1}$	$cr_{2,3}$	$cl_{3,3}$	$wr_{2,4}$
11	$cr_{1,3}$	$wl_{2,1}$	$cl_{2,2}$	$cr_{2,3}$	$wr_{1,4}$	$wl_{3,1}$	$wr_{2,4}$	$cl_{3,3}$
12	$cr_{1,3}$	$wl_{2,1}$	$cl_{2,2}$	$cr_{2,3}$	$wr_{1,4}$	$wl_{3,1}$	$cl_{3,3}$	$wr_{2,4}$
13	$wl_{2,1}$	$cr_{1,3}$	$wr_{1,4}$	$cl_{2,2}$	$cr_{2,3}$	$wl_{3,1}$	$wr_{2,4}$	$cl_{3,3}$
14	$wl_{2,1}$	$cr_{1,3}$	$wr_{1,4}$	$cl_{2,2}$	$cr_{2,3}$	$wl_{3,1}$	$cl_{3,3}$	$wr_{2,4}$
15	$wl_{2,1}$	$cr_{1,3}$	$wr_{1,4}$	$cl_{2,2}$	$wl_{3,1}$	$cr_{2,3}$	$wr_{2,4}$	$cl_{3,3}$
16	$wl_{2,1}$	$cr_{1,3}$	$wr_{1,4}$	$cl_{2,2}$	$wl_{3,1}$	$cr_{2,3}$	$cl_{3,3}$	$wr_{2,4}$
17	$wl_{2,1}$	$cr_{1,3}$	$wr_{1,4}$	$wl_{3,1}$	$cl_{2,2}$	$cr_{2,3}$	$wr_{2,4}$	$cl_{3,3}$
18	$wl_{2,1}$	$cr_{1,3}$	$wr_{1,4}$	$wl_{3,1}$	$cl_{2,2}$	$cr_{2,3}$	$cl_{3,3}$	$wr_{2,4}$
19	$wl_{2,1}$	$cr_{1,3}$	$cl_{2,2}$	$wr_{1,4}$	$cr_{2,3}$	$wl_{3,1}$	$wr_{2,4}$	$cl_{3,3}$
20	$wl_{2,1}$	$cr_{1,3}$	$cl_{2,2}$	$wr_{1,4}$	$cr_{2,3}$	$wl_{3,1}$	$cl_{3,3}$	$wr_{2,4}$
21	$wl_{2,1}$	$cr_{1,3}$	$cl_{2,2}$	$wr_{1,4}$	$wl_{3,1}$	$cr_{2,3}$	$wr_{2,4}$	$cl_{3,3}$
22	$wl_{2,1}$	$cr_{1,3}$	$cl_{2,2}$	$wr_{1,4}$	$wl_{3,1}$	$cr_{2,3}$	$cl_{3,3}$	$wr_{2,4}$
23	$wl_{2,1}$	$cr_{1,3}$	$cl_{2,2}$	$cr_{2,3}$	$wr_{1,4}$	$wl_{3,1}$	$wr_{2,4}$	$cl_{3,3}$
24	$wl_{2,1}$	$cr_{1,3}$	$cl_{2,2}$	$cr_{2,3}$	$wr_{1,4}$	$wl_{3,1}$	$cl_{3,3}$	$wr_{2,4}$

$$= \begin{pmatrix} \alpha(n) \\ \alpha_{cr_{1,3,k_1}^1}(n), \alpha_{wr_{1,4,k_2}^1}(n), \\ \alpha_{wl_{2,1,k_3}^2}(n), \alpha_{cl_{2,2,k_4}^2}(n), \\ \alpha_{cr_{2,3,k_5}^2}(n), \alpha_{wr_{2,4,k_6}^2}(n), \dots, \\ \alpha_{wl_{m,1,k}^m}(n), \alpha_{cl_{m,2,k}^m}(n) \end{pmatrix} \quad (14)$$

$$1 \leq k_1 \leq k_2 \leq k_3 \leq \dots \leq L$$

فرض کنید $LA_1 = (\alpha_1, \alpha_2, \dots, \alpha_r)$ نشان دهنده‌ی یک آتوماتون به طول r بوده که $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_r)$ نشان دهنده‌ی مجموعه‌ی عمل و $p = \{p_1, p_2, \dots, p_r\}$ توزیع احتمال اعمال آتوماتون LA_1 را نشان دهد. روش انتخاب یک عمل در آتوماتون LA_1 به شرح زیر است:

(۱) با استفاده از بردار احتمال، طول بازه برای هر عمل LA_1

به صورت معادله‌ی ۱۵ تعیین می‌شود.

$$\begin{aligned} \alpha_1 &\in [0, 0 + p_1) \\ \alpha_2 &\in [0 + p_1, 0 + p_1 + p_2) \\ \alpha_3 &\in [0 + p_1 + p_2, 0 + p_1 + p_2 + p_3) \\ &\vdots \\ &\vdots \\ \alpha_r &\in [0 + p_1 + p_2 + \dots + p_{L-1}, 0 + p_1 \\ &\quad + p_2 + \dots + p_r) \end{aligned} \quad (15)$$

(۲) یک عدد تصادفی بین صفر و یک انتخاب می‌شود.

(۳) بررسی این که عدد تصادفی متعلق به کدام بازه است،

عمل انتخابی (اکشن انتخابی) را نشان می‌دهد.

$$P = \left\{ \begin{array}{l} cr_{1,3}, wl_{2,1}, wr_{1,4}, \dots, wl_{i,1}, wr_{i-1,4}, cl_{i,2}, cr_{i,3}, \\ wr_{i,4}, \dots, wl_{m-1,1}, \\ cl_{m-1,2}, cr_{m-1,3}, wl_{m,1}, wr_{m-1,4}, cl_{m,2} \end{array} \right\}$$

نشان دهنده‌ی مجموعه‌ی همه‌ی پارامترهای توابع عضویت دوزنقه‌ای باشد. هر جایگشت از این مجموعه می‌تواند نشان دهنده‌ی یک تابع عضویت دوزنقه‌ای باشد. با این حال، بسیاری از این جایگشت‌ها نامعتبر هستند. برای هر تابع عضویت دوزنقه‌ای R_i ، هر دنباله‌ی غیر تکراری از حاصل ضرب دکارتی:

$$\{wr_{i-1,4}, cl_{i,2}\} \times \{wr_{i-1,4}, cl_{i,2}, cr_{i,3}, wl_{i+1,1}\} \times \{wr_{i-1,4}, cl_{i,2}, cr_{i,3}, wl_{i+1,1}\} \times \{cr_{i,3}, wl_{i+1,1}\}$$

به عنوان یک توالی معتبر (توابع عضویت مناسب) در نظر گرفته می‌شود. پس از محاسبه‌ی هر R_i ، هر توالی غیر تکراری از حاصل ضرب دکارتی $R_1 \times R_2 \times \dots \times R_m$ به عنوان تابع عضویت نهایی مناسب در نظر گرفته می‌شود. به عنوان مثال، برای سه تابع عضویت، اگر از روش پیشنهادی استفاده نشود، تعداد $12! = 479,001,600$ دنباله خواهیم داشت. از میان این دنباله‌ها، رویکرد پیشنهادی فقط ۲۴ تابع عضویت را به عنوان تابع عضویت مناسب در نظر می‌گیرد. یعنی این توابع عضویت، پوشش کامل و همپوشانی مناسب دارند. جدول ۱ تعداد توابع عضویت مناسب برای سه تابع عضویت دوزنقه‌ای را نشان می‌دهد.

۴-۵- تولید اکشن (اعمال اتوماتا)

فرض کنید i نشان دهنده i امین تابع عضویت بوده به طوری که $z \in \{wl_{i,j}, cl_{i,j}, cr_{i,j}, wr_{i,j}\}$ و $(1 \leq i \leq m)$ نشان دهنده‌ی i امین پارامتر از i امین تابع عضویت دوزنقه‌ای و k نشان دهنده‌ی تعداد اعمال انتخاب شده توسط آتوماتای یادگیر باشد. k امین عمل انتخاب شده توسط آتومای یادگیر برای i امین پارامتر از i امین تابع عضویت در لحظه‌ی n با $\alpha_{j,k}^i(n)$ نمایش داده می‌شود. برای مثال، $\alpha_{cr_{1,3,2}}^1(n)$ نشان می‌دهد که در اولین تابع عضویت، آتوماتای یادگیر برای پارامتر $cr_{1,3}$ در لحظه n دومین عمل خود را انتخاب کرده است. فرض کنید t_{max} حداکثر مدت زمان در بین تمام صفحات وب باشد. عمل هر آتوماتا در لحظه‌ی n با در نظر گرفتن t_{max} و نرمال‌سازی مدت زمان هر صفحه‌ی وب بین مقادیر صفر و یک، به صورت $\alpha_{j,k}^i(n)$ در نظر گرفته می‌شود. در لحظه‌ی n ، هر آتوماتون یادگیر، عملی را به طور تصادفی و مستقل از دیگر آتوماتا انتخاب می‌کند. اعمال انتخاب شده توسط تیم آتوماتا در لحظه‌ی n در معادله‌ی ۱۴ آورده شده است.

جدول ۲: شبه کد الگوریتم LA-OMF

Proposed LA-OMF Algorithm
<p>Input:</p> <p>n: number of Web log records; m: number of Web pages ; p: Web page number ; K_{max} : maximum number of membership functions; Maxiter: maximum number of iterations; ltr: number of iterations ; α: minimum support .</p> <p>Output: optimal number of membership functions $OptMF$ and its associated set of membership functions $BestMF$</p>
<p>Step 1: Define appropriate membership functions matrix v [] [] and action vector Action [] .</p> <p>Step 2: Define $Global \beta_{\alpha}(n)$ and initialize it with a maximum value.</p> <p>Step 3: initialize $BestMF = \emptyset$.</p> <p>Step 4: initialize $F \leftarrow 0$.</p> <p>Step 5: For each $MF_i, 2 \leq i \leq K_{max}$ do</p> <p>Step 5.1: Create a team of LA with $(K_{max} \times 4 - 4)$ automata such that each automaton has the length of r .</p> <p>Step 5.2: Initialize parameters of each automaton, such as a, b, r (see Section 3.2).</p> <p>Step 5.3: Define the action-set of each automaton (see Section 4.4).</p> <p>Step 5.4: Repeat Steps (5-4-1) to (5-4-3) until the maximum number of iterations is satisfied.</p> <p>Step 5.4.1: Each automaton $A_j^i (1 \leq i \leq m, j \in \{wl_{i,1}, cl_{i,2}, cr_{i,3}, wr_{i,4}\})$ at instant n selects an action $\alpha_j^{i,k}(n)$ (see Section 4.3).</p> <p>Step 5.4.2: Evaluate the objective function and obtain $\beta_{\alpha}(n)$. (see Section 4.5).</p> <p>Step 5.4.3: Update the parameters of each automaton (see Section 4.6).</p> <p>Step 5.5: If $\beta_{\alpha}(n) < Global \beta_{\alpha}(n)$, then</p> <p>Step 5.5.1: $Global \beta_{\alpha}(n) \leftarrow \beta_{\alpha}(n)$ and save the associated membership functions in $BestMF$.</p> <p>Step 5.5.2: $OptMF \leftarrow i$.</p> <p>End if.</p> <p>End for.</p> <p>Step 6: Report $OptMF, Global \beta_{\alpha}(n)$, and $BestMF$.</p>

این صورت، $\beta_{\alpha}(n)$ در $Global \beta_{\alpha}(n)$ ذخیره شده و به همهی اعمال انتخاب شده در هر اتوماتا طبق معادله‌ی ۱ پاداش داده می‌شود. در غیر این صورت، به تمام اعمال انتخاب شده در هر اتوماتا طبق معادله‌ی ۲ جریمه داده می‌شود.

۴-۸- استخراج تعداد بهینه‌ی توابع عضویت

پس از تعیین چارچوب پیشنهادی، الگوریتمی برای یافتن تعداد بهینه‌ی توابع عضویت و پارامترهای مربوطه توسعه داده شد. فرض کنید تعداد بهینه‌ی توابع عضویت را با پارامتر $OptMF$ نشان دهیم، به طوی که: $2 \leq OptMF \leq K_{max}$ و نشان دهنده‌ی حداکثر تعداد توابع عضویت مشخص شده توسط کاربر باشند. برای تعیین مقدار $OptMF$ ، الگوریتم پیشنهادی به شرح زیر عمل می‌کند: برای هر مقدار $OptMF$ با استفاده از چارچوب پیشنهادی، مقدار $\beta_{\alpha}(n)$ تعیین می‌شود. سپس، $OptMF$ با کمترین $\beta_{\alpha}(n)$ به عنوان تعداد بهینه‌ی توابع عضویت در نظر گرفته می‌شود. شکل ۶ نحوه یافتن تعداد بهینه‌ی توابع عضویت را نشان می‌دهد. همچنین شبه کد الگوریتم LA-OMF در جدول ۲ نشان داده شده است.

برای تعیین احتمال عمل انتخابی i با توجه به اینکه r نشان دهنده‌ی تعداد اکشن‌ها باشد، از فرمول زیر استفاده می‌کنیم:

$$Optimize Value = \frac{i-1}{r-1} \alpha_i, 1 \leq i \leq r \quad (16)$$

۴-۶- ارزیابی تابع هزینه

بعد از تعیین $\alpha(n)$ مرحله بعدی، ارزیابی مقدار تابع هزینه یعنی یافتن مقدار $\beta_{\alpha}(n)$ است. برای این منظور، هر بردار $\alpha(n)$ به بردار دیگری که بردار عمل (Action vector) نامیده می‌شود، اختصاص داده می‌شود. پس از آن، بردار عمل، به صورت صعودی مرتب و برای یافتن مقدار $\beta_{\alpha}(n)$ مورد استفاده قرار می‌گیرد. مقادیر مرتب شده‌ی این بردار در معادله ۱۷ نشان داده شده است.

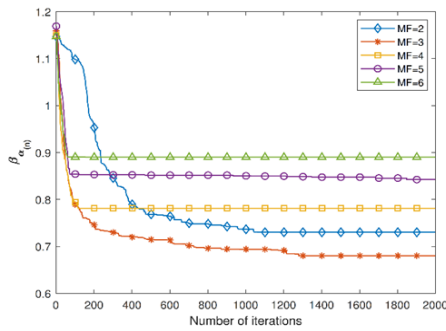
$$Action = (\alpha_j^{i1}(n), \alpha_j^{i2}(n), \alpha_j^{i3}(n), \dots, \alpha_j^{im}(n)), \quad (17)$$

$$where \alpha_j^{i1}(n) \leq \alpha_j^{i2}(n) \leq \alpha_j^{i3}(n) \leq \dots, \\ \leq \alpha_j^{im}(n), \\ (i1 \neq i2 \neq i3, \dots, \neq im)$$

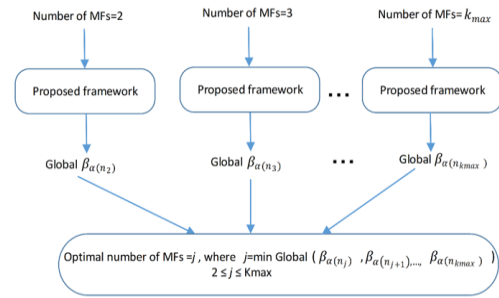
توابع عضویت مناسب را می‌توان با یک بردار دو بعدی، به نام ماتریس تابع عضویت معتبر نشان داد. هر سطر از این ماتریس یک دنباله‌ی معتبر را مشخص می‌کند. تعداد سطرهای موجود در این ماتریس، تعداد توابع عضویت معتبر را نشان می‌دهد. برای محاسبه‌ی $\beta_{\alpha}(n)$ به صورت زیر عمل می‌کنیم: برای هر سطر از این ماتریس، ابتدا تابع عضویت را می‌سازیم. سپس تابع هزینه با توجه به معادله‌ی ۴ ارزیابی می‌شود. با در نظر گرفتن حداقل مقدار تابع هزینه از بین همه‌ی توابع عضویت معتبر، مقدار $\beta_{\alpha}(n)$ مشخص می‌شود. پس از تعیین حداقل مقدار $\beta_{\alpha}(n)$ توسط تیم اتوماتا، مجموعه‌ی تابع عضویت مرتبط به عنوان بهترین تابع عضویت انتخاب می‌شود.

۴-۷- بروزرسانی پارامترهای اتوماتا

فرض می‌کنیم $Global \beta_{\alpha}(n)$ متغیری باشد که برای ذخیره‌ی حداقل مقدار $\beta_{\alpha}(n)$ که تاکنون به دست آمده استفاده شود. پس از یافتن مقدار $\beta_{\alpha}(n)$ به عنوان یک سیگنال تقویت کننده‌ی مشترک، هر اتوماتون یادگیر برای عمل انتخاب شده‌ی خود پاداش یا جریمه را به شرح زیر انجام می‌دهد: در هر تکرار، با بدست آوردن مقدار $\beta_{\alpha}(n)$ ، بررسی می‌کنیم آیا مقدار $\beta_{\alpha}(n)$ کمتر از مقدار $Global \beta_{\alpha}(n)$ است یا نه. اگر این مورد باشد، در



شکل ۷: مقدار $\beta_\alpha(n)$ بدست آمده برای توابع عضویت ۲، ۳، ۴، ۵ و ۶



شکل ۶: یافتن تعداد توابع عضویت دوزنقه‌ای

۵- آزمایش‌ها و تحلیل نتایج

در این بخش، چندین آزمایش برای نشان دادن کارایی الگوریتم پیشنهادی LA-OMF با استفاده از پیاده‌سازی الگوریتم‌ها در زبان جاوا انجام دادیم. در این بخش نتایج به دست آمده از اجرای الگوریتم پیشنهادی LA-OMF در مقایسه با روش کاوش فازی و [۵۷] MAOMF، [۶۰] FTARM+PFGM و [۶۹] FWMA مورد مقایسه قرار گرفت. برای ارزیابی و تحلیل LA-OMF پیشنهادی، مجموعه داده‌ی DePaul CTI را در نظر گرفتیم [۷۰]. این داده‌ها در طی یک دوره‌ی دو هفته‌ای، در آوریل ۲۰۰۲ از نمونه کاربرانی که این وب سایت را مشاهده کرده‌اند، جمع آوری شده و شامل ۱۳۷۴۵ جلسه کاربری^{۱۶} و ۶۸۳ صفحه‌ی وب می‌باشند [۷۰]. در آزمایش‌ها، صفحه‌ی وب با شماره ۳۸۷ (news/default.asp)، به عنوان صفحه‌ی مورد نظر برای انجام آزمایش‌ها، انتخاب شده است. پارامترهای مورد استفاده توسط الگوریتم LA-OMF در جدول ۳ آورده شده است. پارامترهای سایر الگوریتم‌های مورد مقایسه، مطابق مرجع اصلی در نظر گرفته شده‌اند.

جدول ۳: پارامترهای استفاده شده توسط LA-OMF

Parameter	Value
Number of iterations	2000
Reward (a)	0.3
Penalty(b)	0.003
Number of actions(r)	20
Minimum support	[0.002 - 0.01]

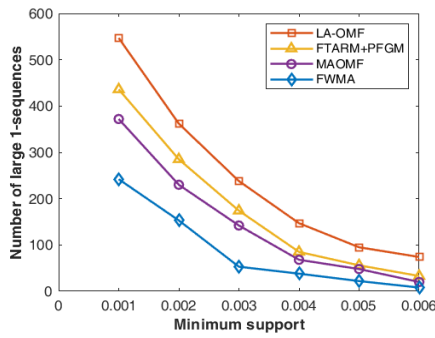
۵-۱- ارزیابی نتایج

آزمایش ۱: این آزمایش با هدف یافتن تعداد توابع عضویت مناسب و پارامترهای بهینه‌ی آنها انجام شده است. در این آزمایش از مقادیر ۲ تا ۶ برای یافتن تعداد توابع عضویت استفاده شده است. شکل ۷ مقدار بهینه‌ی $\beta_\alpha(n)$ برای مقادیر ۲ تا ۶ را برای الگوریتم پیشنهادی نشان می‌دهد. در جدول ۴ تعداد توابع عضویت بهینه با در نظر گرفتن کمترین مقدار تابع هزینه به ترتیب برای

الگوریتم پیشنهادی برابر ۳، الگوریتم FTARM+PFGM برابر ۴ و الگوریتم MAOMF برابر ۲ می‌باشد. از آنجایی که تعداد توابع عضویت در الگوریتم FWMA ثابت در نظر گرفته شده، با آزمایش تعداد توابع عضویت از مقادیر ۲ تا ۶ و با در نظر گرفتن حداقل مقدار تابع هزینه، تعداد بهینه‌ی توابع عضویت ۳ در نظر گرفته شده است. نتایج جدول ۴ نشان می‌دهد که برای تعداد توابع عضویت ۲ تا ۶ میانگین مقدار تابع هزینه به ترتیب ۲۰٪، ۲۳٪، ۱۸٪، ۱۷٪ و ۱۵٪ افزایش یافته است. همچنین، میانگین کارایی تابع هزینه در مقایسه با سایر الگوریتم‌ها ۱۹٪ افزایش یافته است. به طور مشابه، LA-OMF میانگین کارایی پشتیبان فازی را ۱۷٪ افزایش داده است.

با توجه به نتایج جدول ۴، LA-OMF حداقل مقدار تابع هزینه را برای هر تعداد از توابع عضویت به دست آورده است. علاوه بر این، برای یافتن بهترین تعداد توابع عضویت LA-OMF برای تعداد توابع عضویت مختلف مورد آزمایش قرار گرفت. نتایج نشان داد که پس از بهینه‌سازی توابع عضویت، شکل نامناسب توابع عضویت ایجاد نشده و کارایی الگوریتم پیشنهادی افزایش یافته است. شکل ۸ الف-ه توابع عضویت اولیه را به ترتیب برای توابع عضویت ۲ تا ۶ نشان می‌دهد. از شکل ۸ مشخص است که توابع عضویت تولید شده در اولین تکرار الگوریتم پیشنهادی دارای همپوشانی زیاد و یا خیلی از هم جدا هستند. پس از اجرای الگوریتم پیشنهادی، توابع عضویت شکل مناسبی دارند. توابع عضویت بهینه به ترتیب در شکل ۹ الف-ه نشان داده شده است.

آزمایش ۲: این آزمایش با هدف نشان دادن رابطه‌ی بین تعداد دنباله‌های بزرگ به طول ۱ و حداقل مقدار پشتیبان‌های مختلف برای LA-OMF و سایر الگوریتم‌ها انجام شده است. شکل ۱۰ تعداد دنباله‌های بزرگ به طول ۱ به دست آمده توسط LA-OMF و سایر الگوریتم‌ها برای حداقل پشتیبان‌های مختلف پس از ۲۰۰۰ تکرار را نشان می‌دهد. همان‌طوری که از شکل ۱۰

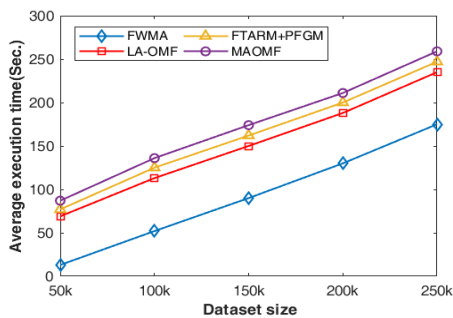


شکل ۱۰: مقایسه تعداد دنباله‌های بزرگ به طول ۱ با حداقل مقادیر مختلف پشتیبان

آزمایش ۳: هدف در این آزمایش، محاسبه‌ی زمان اجرای الگوریتم پیشنهادی و مقایسه آن با سایر روش‌های مورد مقایسه بود. محاسبه‌ی زمان در الگوریتم پیشنهادی شامل دو مرحله است: مرحله ۱، شامل زمان اجرا برای یافتن توابع عضویت بهینه بوده و در مرحله ۲، زمان اجرا برای کاوش الگوهای بزرگ به طول یک با استفاده از توابع عضویت بهینه‌ی به دست آمده محاسبه می‌شود. بنابراین، کل زمان اجرا به صورت رابطه‌ی ۱۸ محاسبه می‌شود:

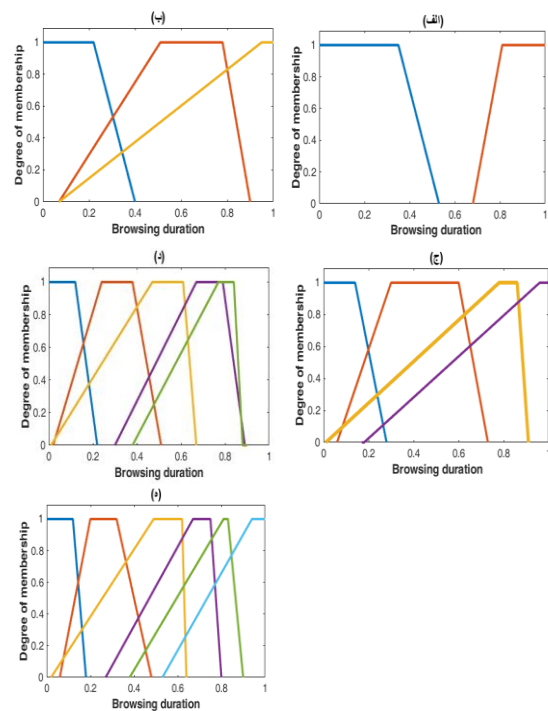
$$Total\ running\ time = \left(\sum_{i=2}^{k_{max}} running\ time\ for\ MF_i \right) + (running\ time\ for\ mining\ large\ 1 - sequences) \quad (18)$$

شکل ۱۱ میانگین زمان اجرا را با افزایش اندازه‌ی مجموعه‌ی داده نشان می‌دهد. نتایج نشان می‌دهد، با افزایش اندازه‌ی مجموعه‌ی داده، زمان اجرای الگوریتم‌ها، افزایش پیدا می‌کند. پرواضح است که زمان اجرای الگوریتم FWMA که از بهینه‌سازی توابع عضویت استفاده نمی‌کند، از همه‌ی الگوریتم‌ها کمتر است. همچنین با افزایش اندازه‌ی مجموعه‌ی داده، میانگین زمان اجرا در الگوریتم پیشنهادی و سایر الگوریتم‌ها به صورت خطی افزایش می‌یابد.

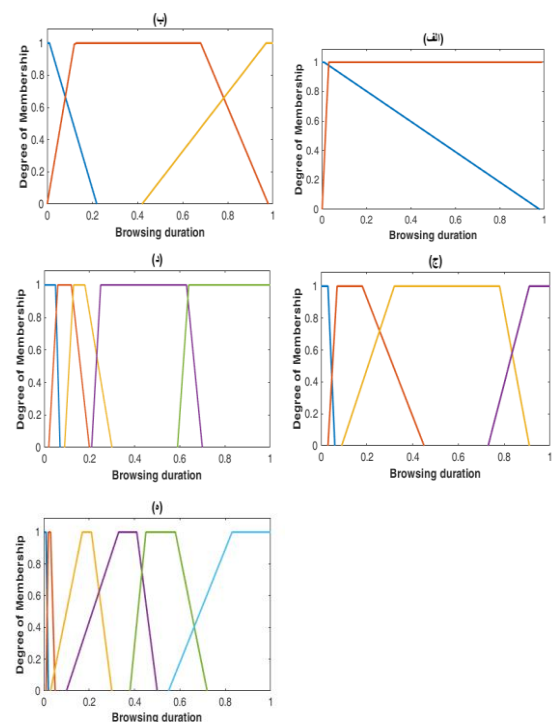


شکل ۱۱: میانگین زمان اجرا برای کاوش دنباله‌های بزرگ به طول ۱ با افزایش اندازه‌ی مجموعه‌ی داده

مشخص است، LA-OMF می‌تواند تعداد دنباله‌های بیشتری نسبت به سایر الگوریتم‌ها تولید کند.



شکل ۸: توابع عضویت بدست آمده قبل از اجرای LA-OMF در مجموعه‌ی داده CTI



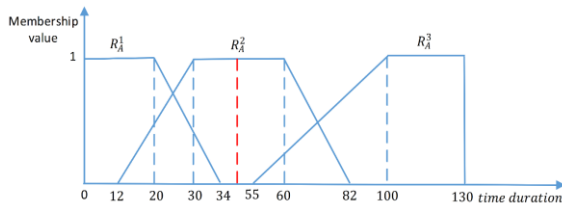
شکل ۹: توابع عضویت بهینه شده بعد از اجرای LA-OMF در مجموعه‌ی داده CTI

همچنین این آزمایش نشان می‌دهد، با افزایش حداقل پشتیبان مختلف، تعداد دنباله‌های بزرگ به طول ۱ کاهش می‌یابد.

فازی.

پیوست:

مثال: فرض کنید توابع عضویت برای مدت زمان صفحه‌ی وب A به صورت شکل ۱۲ باشد. براساس این شکل، صفحه‌ی وب A دارای سه ناحیه‌ی فازی R_A^1 ، R_A^2 و R_A^3 خواهد بود.



شکل ۱۲: توابع عضویت برای مدت زمان صفحه‌ی A

در این مثال، همپوشانی‌ها به صورت زیر قابل محاسبه خواهند بود: $\text{overlap}(R_A^1, R_A^2) = 22$ ، $\text{overlap}(R_A^1, R_A^3) = 0$ و $\text{overlap}(R_A^2, R_A^3) = 27$. حداقل پهنای هر ناحیه نیز به صورت زیر قابل محاسبه است:

$$\begin{aligned} \text{minimum spread}(R_A^1, R_A^2) &= \min((34-0), (45-12)) = 33, \\ \text{minimum spread}(R_A^2, R_A^3) &= \min((82-45), (130-55)) = 37 \\ \text{minimum spread}(R_A^1, R_A^3) &= \min((34-0), (130-55)) = 34 \end{aligned}$$

بنابراین، فاکتور همپوشانی صفحه‌ی A به صورت زیر محاسبه می‌شود:

$$\begin{aligned} \text{Overlap - factor} &= \\ &= [\max(\frac{22}{33}, 1) - 1] + [\max(\frac{27}{37}, 1) - 1] + \\ &= [\max(\frac{0}{34}, 1) - 1] = 0 + 0 + 0 = 0 \end{aligned}$$

در این مثال، $\text{range}(R_A^1, R_A^2, R_A^3) = 130$ است. زیرا ناحیه‌ی پوشش داده شده توسط این سه ناحیه‌ی فازی از بازه ۰ تا ۱۳۰ است. فرض کنید حداکثر مدت زمان صفحه‌ی وب A در جلسات ۱۳۰ و برابر با $\max(t_i) = 130$ باشد. فاکتور پوشش توابع عضویت برای صفحه‌ی A به شرح زیر محاسبه می‌شود:

$$\text{coverage factor} = \frac{130}{130} = 1$$

بنابراین در این مثال، مقدار مناسب بودن توابع عضویت برای صفحه‌ی A به صورت زیر محاسبه می‌شود:

$$\text{suitability} = 0 + 1 = 1$$

برای سایر صفحات وب در جلسات، مقدار مناسب بودن توابع عضویت نیز به همین روش محاسبه می‌شود. جدول ۶ مدت زمان

آزمایش ۴: هدف از این آزمایش، بررسی تأثیر اندازه مختلف مجموعه‌ی داده در کارایی الگوریتم LA-OMF است. در این آزمایش، مجموعه‌ی داده با اندازه‌های مختلف از جمله ۵۰K، ۱۰۰K، ۱۵۰K، ۲۰۰K و ۲۵۰K مورد آزمایش قرار گرفت. در جدول ۵ نتایج مربوط به همپوشانی، پوشش، مناسب بودن، پشتیبان فازی و میانگین مقدار تابع هزینه بین LA-OMF و سایر الگوریتم‌ها مورد مقایسه قرار گرفته است. همان‌گونه که از جدول ۵ مشخص است، الگوریتم LA-OMF حداقل مقدار تابع هزینه‌ی بهتر و پشتیبانی فازی بالاتری در مقایسه با سایر الگوریتم‌ها دارد. به علاوه، نتایج جدول ۵ نشان می‌دهد، روش پیشنهادی که از الگوریتم‌ها عملکرد بهتری دارد.

۶- نتیجه‌گیری

در این پژوهش، برای اولین بار یک چارچوب مبتنی بر اتوماتای یادگیر برای استخراج همزمان تعداد و پارامترهای توابع عضویت دوزنقه‌ای، برای استخراج قوانین انجمنی فازی از داده‌های وب، پیشنهاد شد. برای این کار، زمان صرف شده توسط کاربران در هر صفحه‌ی وب، به عنوان یک تابع عضویت دوزنقه‌ای در نظر گرفته شده و به جای استفاده از توابع عضویت از پیش تعریف شده، ما از اتوماتای یادگیر برای یافتن همزمان تعداد و پارامترهای توابع عضویت دوزنقه‌ای استفاده نمودیم. در الگوریتم پیشنهادی LA-OMF، هر تابع عضویت دوزنقه‌ای با چهار پارامتر نشان داده شده بود. برای تنظیم این پارامترها، تیمی از اتوماتای یادگیر ساخته شده و با اختصاص هر اتوماتون به هر پارامتر، مقدار بهینه‌ی آن پارامتر تعیین می‌شد. در الگوریتم پیشنهادی LA-OMF، به منظور کاهش فضای جستجو و حذف توابع عضویت نامناسب، دو هیوربستیک معرفی و یک روش جدیدی برای پیاده‌سازی آنها پیشنهاد گردید. همچنین برای ارزیابی کارایی LA-OMF، آزمایش‌های متعددی با اندازه‌های مختلف بر روی مجموعه‌ی داده انجام شد. نتایج آزمایش نشان داد که الگوریتم پیشنهادی می‌تواند به پشتیبان فازی بالا، همپوشانی و پوشش مناسب دست یابد. نتایج همچنین نشان داد که الگوریتم پیشنهادی نسبت به سایر الگوریتم‌ها، مقدار تابع هزینه‌ی کمتر، پشتیبان فازی بالا و همچنین تعداد قوانین (دنباله‌های بزرگ به طول یک) بیشتری تولید می‌کند. در آینده، ما به دو موضوع خواهیم پرداخت: الف) استفاده از چارچوب روش پیشنهادی در کاوش قوانین انجمنی چند هدفه (ب) استفاده از چارچوب پیشنهادی برای بهینه‌سازی توابع عضویت در سایر سیستم‌های

تمام صفحات وب مرور شده توسط هر کاربر را که شامل ۶ جلسه کاربری است، نشان می‌دهد.

جدول ۴: مقایسه نتایج بین LA-OMF و سایر الگوریتم‌ها

Number of Linguistic terms	Algorithm	Overlap	Coverage	Suitability	Fuzzy support	$\beta_\alpha(n)$
# MFs = 2	LA-FOMF	0.0000	1.0000±0.0000	1.0000	1.3541	0.7368±0.0243
	FWMA	0.0000	1.0000±0.0000	1.0000	0.8216	1.2175±0.0000
	FTARM+PFGM	0.0018	1.0000±0.0000	1.0018	1.2805	0.7823±0.0342
	MAOMF	0.0512	1.0000±0.0000	1.0512	1.3101	0.8024±0.0580
# MFs = 3	LA-FOMF	0.0026	1.0000±0.0000	1.0026	1.4771	0.6842±0.0092
	FWMA	0.0000	1.0000±0.0000	1.0000	0.8285	1.2070±0.0000
	FTARM+PFGM	0.0186	1.0000±0.0000	1.0186	1.1951	0.8523±0.0672
	MAOMF	0.0245	1.0000±0.0000	1.0245	1.2166	0.8421±0.0652
# MFs = 4	LA-FOMF	0.0073	1.0000±0.0000	1.0073	1.2674	0.7895±0.0072
	FWMA	0.0000	1.0000±0.0000	1.0000	0.8261	1.2105±0.0000
	FTARM+PFGM	0.0235	1.0000±0.0000	1.0235	1.3714	0.7463±0.0161
	MAOMF	0.0362	1.0000±0.0000	1.0362	1.1727	0.8836±0.1139
# MFs = 5	LA-FOMF	0.0140	1.0000±0.0000	1.0140	1.2024	0.8421±0.0058
	FWMA	0.0000	1.0000±0.0000	1.0000	0.8173	1.2235±0.0000
	FTARM+PFGM	0.0476	1.0000±0.0000	1.0476	1.1455	0.9145±0.0524
	MAOMF	0.0752	1.0000±0.0000	1.0752	1.1369	0.9457±0.0943
# MFs = 6	LA-FOMF	0.0230	1.0000±0.0000	1.0230	1.1424	0.8947±0.0063
	FWMA	0.0000	1.0000±0.0000	1.0000	0.8055	1.2414±0.0000
	FTARM+PFGM	0.0674	1.0000±0.0000	1.0674	1.1174	0.9552±0.0389
	MAOMF	0.0923	1.0000±0.0000	1.0923	1.0987	0.9941±0.0873

جدول ۵: مقایسه نتایج بدست آمده بر روی داده با اندازه‌های مختلف بر روی دیتاست CTI

Data	Algorithm	Overlap	Coverage	Suitability	Fuzzy support	$\beta_\alpha(n)$
50K	LA-OMF	0.0025	1.0000	1.0025	1.3723	0.7368
	FWMA	0.0000	1.0000	1.0000	0.8462	1.1817
	FTARM+PFGM	0.0165	1.0000	1.0165	1.3284	0.7652
	MAOMF	0.0724	1.0000	1.0724	1.3585	0.7894
100K	LA-OMF	0.0040	1.0000	1.0040	1.4726	0.6842
	FWMA	0.0000	1.0000	1.0000	0.8683	1.1519
	FTARM+PFGM	0.0268	1.0000	1.0268	1.4190	0.7236
	MAOMF	0.0952	1.0000	1.0952	1.4314	0.7651
150K	LA-OMF	0.0009	1.0000	1.0009	1.5640	0.6316
	FWMA	0.0000	1.0000	1.0000	0.8679	1.1522
	FTARM+PFGM	0.0325	1.0000	1.0325	1.4672	0.7037
	MAOMF	0.0863	1.0000	1.0863	1.4231	0.7633
200K	LA-OMF	0.0016	1.0000	1.0016	1.5784	0.6316
	FWMA	0.0000	1.0000	1.0000	0.8662	1.1544
	FTARM+PFGM	0.0409	1.0000	1.0409	1.4910	0.6981
	MAOMF	0.1153	1.0000	1.1153	1.5035	0.7418
250K	LA-OMF	0.0014	1.0000	1.0014	1.2712	0.7895
	FWMA	0.0000	1.0000	1.0000	0.8656	1.1552
	FTARM+PFGM	0.0452	1.0000	1.0452	1.24088	0.8423
	MAOMF	0.1476	1.0000	1.1476	1.2718	0.9023

طول یک برابر $\{\alpha_1, \alpha_3\}$ بوده که در آن α_1 و α_3 مقدار پشتیبانی فازی بزرگتر از ۰.۲۸ دارند. بنابراین، مقدار پشتیبانی کل فازی به صورت زیر محاسبه می‌شود:

$$\text{FuzzySupport}(X) = 0.38 + 0.3 = 0.68.$$

در نتیجه، مقدار تابع هزینه برابر $1/0.68=1.47$ خواهد بود.

فرض کنید که صفحه‌ی وب A در پنج دنباله حرکتی با مدت زمان‌های ۳۰، ۸، ۹۲، ۲۰ و ۱۱۰ وجود داشته باشد. پشتیبانی فازی هر ناحیه R_A^1, R_A^2, R_A^3 به صورت زیر محاسبه می‌شود:

$$\text{FuzzySupport}(\alpha_1) = \frac{1}{6} \times \left(\frac{34-30}{34-20} + 1 + 0 + 1 + 0 \right) = 0.38$$

$$\text{FuzzySupport}(\alpha_2) = \frac{1}{6} \times \left(1 + 0 + 0 + \frac{20-12}{30-12} + 0 \right) = 0.24$$

$$\text{FuzzySupport}(\alpha_3) = \frac{1}{6} \times \left(0 + 0 + \frac{92-55}{100-55} + 0 + 1 \right) = 0.3$$

فرض کنید که حداقل پشتیبانی ۰.۲۸ باشد. دنباله‌های بزرگ به

International Journal of Information Management, 2018. 38(1): p. 201-216.

- [19] Kumar, M. and M. Meenu, *A survey on pattern discovery of web usage mining*. International Journal of Advance Research, Ideas and Innovations in Technology, 2017. 3(1): p. 379-385.
- [20] Sisodia, D.S., V. Khandal, and R. Singhal, *Fast prediction of web user browsing behaviours using most interesting patterns*. Journal of Information Science, 2018. 44(1): p. 74-90.
- [21] Malarvizhi, S. and B. Sathiyabhama, *Frequent pagesets from web log by enhanced weighted association rule mining*. Cluster Computing, 2016. 19(1): p. 269-277.
- [22] Zadeh, L.A., *Fuzzy sets*. Information and control, 1965. 8(3): p. 338-353.
- [23] Lopez, F.J., et al. *Extracting biological knowledge by fuzzy association rule mining*. in *2007 IEEE International Fuzzy Systems Conference*. 2007. IEEE.
- [24] Mamdani, E.H. *Application of fuzzy algorithms for control of simple dynamic plant*. in *Proceedings of the institution of electrical engineers*. 1974. IET.
- [25] Tajbakhsh, A., M. Rahmati, and A. Mirzaei, *Intrusion detection using fuzzy association rules*. Applied Soft Computing, 2009. 9(2): p. 462-469.
- [26] Wang, M., et al. *A cancer classification method based on association rules*. in *2012 9th International Conference on Fuzzy Systems and Knowledge Discovery*. 2012. IEEE.
- [27] Watanabe, T. and R. Fujioka. *Fuzzy association rules mining algorithm based on equivalence redundancy of items*. in *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 2012. IEEE.
- [28] Weber, R. *A class of methods for automatic knowledge acquisition*. in *Proc. Of the 2nd International Conference on Fuzzy Logic and Neural Networks*, 1992. 1992.
- [29] Kudlacik, P., P. Porwik, and T. Wesolowski, *Fuzzy approach for intrusion detection based on user's commands*. Soft Computing, 2016. 20(7): p. 2705-2719.
- [30] Nasser, S.H., A. Ebrahimnejad, and O. Gholami, *Fuzzy stochastic data envelopment analysis with undesirable outputs and its application to banking industry*. International Journal of Fuzzy Systems, 2018. 20(2): p. 534-548.
- [31] Molina, C., M.D. Ruiz, and J.M. Serrano, *Representation by levels: An alternative to fuzzy sets for fuzzy data mining*. Fuzzy Sets and Systems, 2019.
- [32] Sumathi, G. and J. Akilandeswari, *Improved fuzzy weighted-iterative association rule based ontology postprocessing in data mining for query recommendation applications*. Computational Intelligence, 2020.
- [33] Wu, R., W. Tang, and R. Zhao. *Web mining of preferred traversal patterns in fuzzy environments*. in *International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing*. 2005. Springer.
- [34] Lin, C.W. and T.P. Hong, *A survey of fuzzy web mining*. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2013. 3(3): p. 199-199.
- [35] Ansari, Z.A. and A.S. Syed, *Discovery of web usage patterns using fuzzy mountain clustering*. International Journal of Business Intelligence and Data Mining, 2016. 11(1): p. 1-18.
- [36] Ansari, Z.A., S.A. Sattar, and A.V. Babu, *A fuzzy neural network based framework to discover user access patterns from web log data*. Advances in Data Analysis and Classification, 2017. 11(3): p. 519-546.
- [37] Hamidzadeh, J., M. Zabihimayvan, and R. Sadeghi, *Detection of Web site visitors based on fuzzy rough sets*. Soft Computing, 2018. 22(7): p. 2175-2188.
- [38] Hong, T.-P., C.-M. Huang, and S.-J. Horng, *Linguistic object-oriented web-usage mining*. International Journal of Approximate Reasoning, 2008. 48(1): p. 47-61.
- [39] Hong, T.-P., M.-J. Chiang, and S.-L. Wang, *Mining fuzzy weighted browsing patterns from time duration and with linguistic thresholds*. 2008.
- [40] Matthews, S.G., et al., *Web usage mining with evolutionary extraction of temporal fuzzy association rules*. Knowledge-Based Systems, 2013. 54: p. 66-72.
- [41] Wang, S.-L., W.-S. Lo, and T.-P. Hong, *Discovery of fuzzy multiple-level Web browsing patterns*, in *Classification and*

جدول ۶: دنباله‌های حرکتی توسط هر کاربر

User ID	Browsing sequences
1	(A,30) (E,42) (D,98) (C,1)
2	(D,62) (A,8) (D,102)
3	(A,92) (D,89)
4	(A,20) (C,101) (E,118) (B,11) (C,42)
5	(D,64) (A,110) (C,74)
6	(D,80) (C,61) (E,122) (B,17)

مراجع

- [1] Etzioni, O., *The world wide web: Quagmire or gold mine? Communications of the ACM*, 39 (11): 65-68. 1996, November.
- [2] Cooley, R., B. Mobasher, and J. Srivastava. *Web Mining: Information and Pattern Discovery on the World Wide Web*. in *ictai*. 1997.
- [3] Kosala, R. and H. Blockeel, *Web mining research: A survey*. ACM Sigkdd Explorations Newsletter, 2000. 2(1): p. 1-15.
- [4] Kalaianam, K., T. Kushwaha, and K. Rajavi, *How does web personalization create value for online retailers? Lower cash flow volatility or enhanced cash flows*. Journal of Retailing, 2018. 94(3): p. 265-279.
- [5] Wagh, R. and J. Patil, *Enhanced web personalization for improved browsing experience*. Advances in Computational Sciences and Technology, 2017. 10(6): p. 1953-1968.
- [6] Desai, D., *An empirical study of website personalization effect on users intention to revisit E-commerce website through cognitive and hedonic experience*, in *Data Management, Analytics and Innovation*. 2019, Springer. p. 3-19.
- [7] Sharma, S. and V. Rana, *Web Search Personalization Using Semantic Similarity Measure*, in *Proceedings of ICRIC 2019*. 2020, Springer. p. 273-288.
- [8] Kaur, J. and J.S. Bal. *Machine Learning Approach to Recommender System for Web Mining*. in *International Conference on Intelligent Data Communication Technologies and Internet of Things*. 2019. Springer.
- [9] Moreno, M.N., et al., *Web mining based framework for solving usual problems in recommender systems. A case study for movies 'recommendation*. Neurocomputing, 2016. 176: p. 72-80.
- [10] Manikandan, R. and V. Saravanan, *A novel approach on Particle Agent Swarm Optimization (PASO) in semantic mining for web page recommender system of multimedia data: a health care perspective*. Multimedia Tools and Applications, 2020. 79(5): p. 3807-3829.
- [11] Anandhi, D. and M.I. Ahmed, *Prediction of user's type and navigation pattern using clustering and classification algorithms*. Cluster Computing, 2019. 22(5): p. 10481-10490.
- [12] El Aissaoui, O., et al., *A fuzzy classification approach for learning style prediction based on web mining technique in e-learning environments*. Education and Information Technologies, 2019. 24(3): p. 1943-1959.
- [13] Van, T., A. Yoshitaka, and B. Le, *Mining web access patterns with super-pattern constraint*. Applied Intelligence, 2018. 48(11): p. 3914-3922.
- [14] Xiong, L., et al., *Access patterns mining from massive spatio-temporal data in a smart city*. Cluster Computing, 2019. 22(3): p. 6031-6041.
- [15] Liu, Z., et al., *Patterns and sequences: Interactive exploration of clickstreams to understand common visitor paths*. IEEE Transactions on Visualization and Computer Graphics, 2016. 23(1): p. 321-330.
- [16] Sethi, S. and A. Dixit, *A novel page ranking mechanism based on user browsing patterns*, in *Software Engineering*. 2019, Springer. p. 37-49.
- [17] Vojiri, S., et al., *EasyMiner: eu: Web framework for interpretable machine learning based on rules and frequent itemsets*. Knowledge-Based Systems, 2018. 150: p. 111-115.
- [18] Rekik, R., et al., *Assessing web sites quality: A systematic literature review by text and association rules mining*.

- [55] Chen, C.-H., et al., *Finding active membership functions for genetic-fuzzy data mining*. International Journal of Information Technology & Decision Making, 2015. **14**(06): p. 1215-1242.
- [56] Wu, M.-T., T.-P. Hong, and C.-N. Lee, *A continuous ant colony system framework for fuzzy data mining*. Soft Computing, 2012. **16**(12): p. 2071-2082.
- [57] Ting, C.-K., et al., *Mining fuzzy association rules using a memetic algorithm based on structure representation*. Memetic Computing, 2018. **10**(1): p. 15-28.
- [58] Ting, C.-K., et al., *Genetic algorithm with a structure-based representation for genetic-fuzzy data mining*. Soft Computing, 2017. **21**(11): p. 2871-2882.
- [59] Song, A., et al. *Utilizing bat algorithm to optimize membership functions for fuzzy association rules mining*. in *International Conference on Database and Expert Systems Applications*. 2017. Springer.
- [60] Chamazi, M.A. and H. Motameni, *Finding suitable membership functions for fuzzy temporal mining problems using fuzzy temporal bees method*. Soft Computing, 2019. **23**(10): p. 3501-3518.
- [61] Alikhademi, F. and S. Zainudin. *Generating of derivative membership functions for fuzzy association rule mining by Particle Swarm Optimization*. in *2014 International Conference on Computational Science and Technology (ICCSST)*. 2014. IEEE.
- [62] Hong, T.-P., Y.-C. Lee, and M.-T. Wu, *An effective parallel approach for genetic-fuzzy data mining*. Expert Systems with Applications, 2014. **41**(2): p. 655-662.
- [63] Patil, U.M. and J. Patil, *MINING FUZZY ASSOCIATION RULES FROM WEB USAGE QUANTITATIVE DATA*. Computer Science & Information Technology, 2016. **89**.
- [64] TSetlin, M. and M. TSetlin, *Automaton theory and modeling of biological systems*. 1973.
- [65] Thathachar, M.A. and P.S. Sastry, *Networks of learning automata: Techniques for online stochastic optimization*. 2011: Springer Science & Business Media.
- [66] Narendra, K.S. and M.A. Thathachar, *Learning automata: an introduction*. 2012: Courier Corporation.
- [67] Thathachar, M.A. and P.S. Sastry, *Varieties of learning automata: an overview*. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2002. **32**(6): p. 711-722.
- [68] Hong, T.-P., et al., *Genetic-fuzzy data mining with divide-and-conquer strategy*. IEEE Transactions on Evolutionary Computation, 2008. **12**(2): p. 252-265.
- [69] Tao, Y.-H., et al., *A practical extension of web usage mining with intentional browsing data toward usage*. Expert Systems with Applications, 2009. **36**(2): p. 3937-3945.
- [70] <http://www.cs.depaul.edu>.
- [42] Narendra, K.S. and M.A. Thathachar, *On the behavior of a learning automaton in a changing environment with application to telephone traffic routing*. IEEE Transactions on Systems, Man, and Cybernetics, 1980. **10**(5): p. 262-269.
- [43] Anari, B., J.A. Torkestani, and A.M. Rahmani, *Automatic data clustering using continuous action-set learning automata and its application in segmentation of images*. Applied Soft Computing, 2017. **51**: p. 253-265.
- [44] Ghavipour, M. and M.R. Meybodi, *An adaptive fuzzy recommender system based on learning automata*. Electronic Commerce Research and Applications, 2016. **20**: p. 105-115.
- [45] Kumar, N., J.-H. Lee, and J.J. Rodrigues, *Intelligent mobile video surveillance system as a Bayesian coalition game in vehicular sensor networks: Learning automata approach*. IEEE Transactions on Intelligent Transportation Systems, 2014. **16**(3): p. 1148-1161.
- [46] Helmszadeh, A. and S.M. Kouhsari, *Calibration of erroneous branch parameters utilising learning automata theory*. IET Generation, Transmission & Distribution, 2016. **10**(13): p. 3142-3151.
- [47] Torkestani, J.A., *An adaptive learning automata-based ranking function discovery algorithm*. Journal of intelligent information systems, 2012. **39**(2): p. 441-459.
- [48] Ghasemi, S., et al., *A cost-aware mechanism for optimized resource provisioning in cloud computing*. Cluster Computing, 2018. **21**(2): p. 1381-1394.
- [49] Kordestani, J.K., et al., *A novel framework for improving multi-population algorithms for dynamic optimization problems: A scheduling approach*. Swarm and evolutionary computation, 2019. **44**: p. 788-805.
- [50] Anari, B., J. Akbari Torkestani, and A.M. Rahmani, *A learning automata-based clustering algorithm using ant swarm intelligence*. Expert systems, 2018. **35**(6): p. e12310.
- [51] Palacios, A.M., et al., *Genetic learning of the membership functions for mining fuzzy association rules from low quality data*. Information Sciences, 2015. **295**: p. 358-378.
- [52] Lin, J.C.-W., et al., *Efficient mining of multiple fuzzy frequent itemsets*. International Journal of Fuzzy Systems, 2017. **19**(4): p. 1032-1040.
- [53] Chen, C.-H., et al. *A GA-based approach for mining membership functions and concept-drift patterns*. in *2015 IEEE Congress on Evolutionary Computation (CEC)*. 2015. IEEE.
- [54] Rudziński, F., *A multi-objective genetic optimization of interpretability-oriented fuzzy rule-based classifiers*. Applied Soft Computing, 2016. **38**: p. 118-133.

¹⁰ Fuzzy support

¹¹ Suitability

¹² Web log records

¹³ Suitability

¹⁴ Overlap

¹⁵ Coverage

¹⁶ Sessions

¹function Fitness

² Linguistic terms

³ Memetic algorithm

⁴ Heuristic

⁵ Learning automaton

⁶ Action

⁷ Fixed learning automata

⁸ Variable structure learning automata

⁹ Cost function