



دانشگاه آزاد اسلامی

واحد تهران جنوب

دانشکده فنی و مهندسی

پایان نامه برای دریافت درجه کارشناسی ارشد “M.Sc”

مهندسی کامپیوتر - نرم افزار

عنوان :

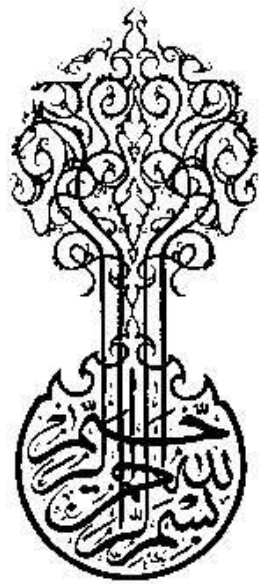
کشف قوانین انجمنی عددی چند هدفه با استفاده از الگوریتم بهینه سازی کلونی

مورچه‌ها برای دامنه پیوسته

استاد راهنما :

استاد مشاور :

نگارش :



تقدیم به

پدر و مادر عزیزم که همواره در زندگی حامی و پشتیبان من بودند. همچنین به تمام کسانی که در راه کسب دانش و معرفت گام بر می‌دارند.

سپاسگزاری

منت خدای را عز و جل که طاعتش موجب قربت است و به شکر اندرش مزید نعمت.
سپاس از پدر و مادر مهربانم که همواره با یاری و کمک بسیار مشوق اصلی من در راه زندگی و کسب دانش بوده‌اند.

سپاس از استاد راهنما و مشاور ارجمندم و از کلیه اساتید محترم که در دوران تحصیل در حل همه مشکلات با صبر و حوصله، اینجانب را راهنما بودند.

و هم چنین سپاس از کلیه کسانی که در انجام این رساله مرا یاری نموده‌اند.

فهرست مطالب

شماره صفحه	عنوان مطالب
1.....	چکیده:
2.....	مقدمه:
7.....	فصل اول.....
7.....	کلیات
8.....	فصل اول: کلیات
8.....	1-1- هدف
9.....	2-1- پیشینه تحقیق
14.....	3-1- روش کار و تحقیق
15.....	فصل دوم.....
15.....	مروری بر ادبیات موضوع
16.....	فصل دوم: مروری بر ادبیات موضوع
16.....	2- مقدمه‌ای بر داده کاوی
16.....	1-2- کشف دانش در پایگاه داده
17.....	2-2- داده کاوی
18.....	3-2- کاوش قوانین انجمنی
19.....	1-3-2- مجموعه اقلام و عدد پشتیبان
19.....	2-3-2- مجموعه اقلام مکرر

- 19..... 3-3-2 قانون انجمنی
- 20..... 4-3-2 قاعده حل مسائل کاوش قوانین انجمنی
- 21..... 5-3-2 اصل *Apriori*
- 21..... 6-3-2 معیارهای جذابیت قوانین انجمنی
- 22..... 7-3-2 الگوریتم *Apriori*
- 25..... 4-2 قوانین انجمنی عددی
- 28..... 5-2 بهینه سازی
- 30..... 1-5-2 الگوریتم های کلونی مورچه ها برای بهینه سازی ترکیبی
- 35..... 1-1-5-2 الگوریتم سیستم اجتماع مورچگان
- 37..... 2-5-2 الگوریتم های کلونی مورچه ها برای بهینه سازی پیوسته
- 40..... 1-2-5-2 الگوریتم بهینه سازی کلونی مورچه ها برای دامنه پیوسته ($ACOR$)
- 42..... 1-1-2-5-2 ساختار آرشیو، مقدار دهی اولیه، بروز رسانی
- 43..... 2-1-2-5-2 ساخت راه حل احتمالی
- 46..... 3-1-2-5-2 ستبری در مقابل سرعت همگرایی
- 47..... 6-2 الگوریتم های بهینه سازی چند هدفه
- 49..... فصل سوم
- 49..... مدل پیشنهادی
- 50..... فصل سوم: مدل پیشنهادی
- 51..... 1-3 الگوریتم کلی سیستم
- 53..... 2-3 معیار جذابیت در سیستم پیشنهادی
- 53..... 3-3 توصیف الگوریتم پیشنهادی کاوش قوانین انجمنی چند هدفه با استفاده از $ACOR$

54	1-3-3- ساختار آرشیو راه حل در الگوریتم پیشنهادی
56	2-3-3- تابع هدف الگوریتم پیشنهادی
58	3-3-3- استخراج قوانین
59	4-3-3- آغاز دهی و کنترل پارامترهای الگوریتم
61	4-3- شرایط همگرایی و پارامترهای توقف سیستم پیشنهادی
62	5-3- نمودار گردش کار الگوریتم پیشنهادی
64	6-3- پیچیدگی محاسباتی الگوریتم پیشنهادی
66	فصل چهارم
66	نتایج و تفسیر آنها
67	فصل چهارم: نتایج و تفسیر آنها
67	1-4- آزمایش الگوریتم پیشنهادی در پایگاه داده مصنوعی
72	2-4- استخراج و مقایسه نتایج اجرا در پایگاه داده‌های مختلف
80	3-4- نتیجه گیری
81	فصل پنجم
81	جمع‌بندی و پیشنهادها
82	فصل پنجم: جمع‌بندی و پیشنهادها
82	1-5- جمع‌بندی
84	2-5- پیشنهادها
86	منابع و مآخذ
86	فهرست منابع فارسی
87	فهرست منابع لاتین

فهرست جداول و نمودارها

شماره صفحه

عنوان

61	جدول (1-4) : مجموعه‌های ساختگی
63	جدول (2-4) : مقادیر پارامترهای تشکیل بازه‌ها در پایگاه داده مصنوعی
64	جدول (3-4) : مقادیر پارامترهای آرشیو راه حل پایگاه داده مصنوعی
66	جدول (4-4) : قوانین انجمنی یافت شده در پایگاه داده‌های مصنوعی
67	جدول (5-4) : نتایج آزمایش الگوریتم همراه با درجات مختلفی از اختلال
69	جدول (6-4) : مشخصات پایگاه داده‌ها
70	جدول (7-4) : مقادیر پارامترهای تشکیل بازه‌ها
72	جدول (8-4) : مقادیر پارامترهای آرشیو در اجرا روی پایگاه داده‌ها
73	جدول (9-4) : مقایسه نتایج در قالب تعداد قوانین و مقادیر اطمینان آن‌ها
74	جدول (10-4) : مقایسه نتایج در قالب معیار پشتیبان و اندازه قوانین

فهرست شکل‌ها

شماره صفحه

عنوان

21 شکل (1-2) : فرآیند کشف دانش در پایگاه داده
22 شکل (2-2) : مثالی از الگوریتم Apriori
24 شکل (3-2) : شبه کد الگوریتم Apriori
25 شکل (4-2) : روش‌های کاوش قوانین انجمنی عددی
32 شکل (5-2) : رفتار واقعی مورچه‌ها
41 شکل (6-2) : ساختار آرشیو راه حل
44 شکل (7-2) : پارامترهای سیستم
49 شکل (1-3) : الگوریتم کلی سیستم
50 شکل (2-3) : توزیع نرمال گوسی با متوسط μ و انحراف معیار σ
51 شکل (3-3) : ساختار آرشیو راه حل الگوریتم پیشنهادی
57 شکل (4-3) : تأثیر β بر شکل گیری بازه‌های راه حل

چکیده:

در حال حاضر اکثر الگوریتم‌های جستجوی به کار گرفته شده در ابزارهای کاوش قوانین انجمنی عددی از روش‌های گسسته سازی متغیرهای پیوسته استفاده کرده و فرض می‌کنند که تمامی متغیرها گسسته هستند. این مسئله، منجر به تولید قوانین انجمنی می‌شود که از دقت و صحت قابل قبولی برخوردار نیستند. از آن جایی که مسئله کاوش قوانین انجمنی عددی یک مسئله بهینه سازی سخت به شمار می‌رود، تا یک مسئله گسسته سازی ساده، الگوریتم‌های جستجوگری که می‌توانند با متغیرهای پیوسته سروکار داشته باشند و جواب‌هایی دقیق‌تر از جواب‌های روش‌های مرسوم فراهم آورند، می‌توانند جالب توجه باشند. در این پایان نامه، الگوریتم جدیدی برای کاوش قوانین انجمنی عددی چند هدفه ارائه شده که قادر است که بدون نیاز به مشخص کردن آستانه حداقل پشتیبان و حداقل اطمینان و در یک مرحله به کشف بازه‌هایی از صفات عددی که قوانین انجمنی جذاب، با پشتیبان و اطمینان بالا بپردازد. برای این کار، الگوریتم کلونی مورچه‌ها در حوزه پیوسته (ACOR) به گونه‌ای به کار گرفته شده است که به تولید قوانین انجمنی عددی بهینه یک مجموعه داده، که شامل متغیرهای عددی پیوسته است، بینجامد.

داده‌کاوی سودمندترین ابزار کشف دانش از میان تراکنش‌ها است [1] [2] [3]. یکی از کاربردهای مهم داده‌کاوی، کشف قوانین انجمنی می‌باشد که یکی از مهم‌ترین روش‌های بازشناسی الگو در سیستم‌های بدون نظارت است. کشف قوانین انجمنی مانند جستجوی طلا در یک پایگاه داده بسیار بزرگ است، که در اینجا منظور از طلا یک قانون جذاب که هنوز کشف نشده است، می‌باشد. از این طریق می‌توان تمام قوانین ممکن را در یک پایگاه داده پیدا کرد، اما مسئله اینجاست که تحلیل تعداد زیادی قانون کار سختی است. به همین جهت، معیارهایی همچون ضریب پشتیبان¹ و ضریب اطمینان² برای مشخص کردن قوانین با کیفیت بالاتر به کار می‌روند.

اکثر الگوریتم‌های کاوش قوانین انجمنی، مانند الگوریتم Apriori و AIS، مبتنی بر روش‌هایی هستند که توسط Agrawal در [4] و [5] پیشنهاد شده‌اند. با استفاده از این متدها نمی‌توان قوانینی را که شامل صفات خاصه عددی هستند کشف نمود. این الگوریتم‌ها مسئله کاوش قوانین انجمنی را به دو قسمت تقسیم می‌کنند [6]: (1) تولید مجموعه اقلام مکرر، که در آن همه اقلامی که معیار حداقل پشتیبان را ارضا می‌کنند پیدا می‌شوند. (2) تولید قوانین انجمنی، که در آن قوانین انجمنی که حد اقل اطمینان را ارضا می‌کنند از مجموعه اقلام مکرر تولید شده در مرحله قبل استخراج می‌شوند. از میان این دو مرحله، تولید قوانین انجمنی، پیچیدگی محاسباتی بالاتری دارد، لذا روش‌هایی که به طور کارا مجموعه اقلام مکرر را تولید می‌کنند می‌توانند جالب توجه واقع شوند. به محض پیدا شدن مجموعه اقلام مکرر، قوانین انجمنی از این مجموعه اقلام استخراج می‌شوند.

در بسیاری از مسائل، ممکن است با اقلام داده‌ای سر و کار پیدا کنیم. که صریح³ یا عددی⁴ هستند. قوانین حاصل از این مجموعه اقلام را قوانین انجمنی عددی می‌نامند. به همین دلیل، الگوریتم‌هایی برای

¹Support

²Confidence

³Categorical

⁴Numeric

کاوش قوانین انجمنی عددی ارائه شدند. در یک قانون انجمنی عددی صفات خاصه داده‌ها محدود به نوع بولین نیستند، بلکه می‌توانند دو نوع عددی (مثل سن، حقوق و گرما) و یا صریح (مثل جنسیت و علامت تجاری) باشند [7]. از آنجایی که کاوش قوانین انجمنی عددی یک مسئله بهینه سازی سخت است تا یک مسئله گسسته سازی ساده، دسته مهمی از مسائل کاوش قوانین انجمنی وجود دارند که تنها زمانی می‌توانند با یک الگوریتم بهینه سازی ترکیبی حل شوند که بازه پیوسته مقادیر مجاز صفات خاصه عددی به مجموعه محدودی تبدیل شوند. مسئله تبدیل بازه پیوسته به مجموعه متناهی همیشه کار راحتی نیست، به خصوص زمانی که بازه اولیه، بازه وسیعی باشد و شفافیت و دقت بالایی مد نظر باشد. بنابراین، در این گونه موارد، معمولاً الگوریتم‌هایی که به طور طبیعی و ساده قادر به کار کردن با متغیرهای پیوسته عددی هستند، بهتر عمل می‌کنند [8].

هم چنین، کاوش قوانین انجمنی عددی باید به عنوان یک مسئله چند هدفه تلقی شود، چرا که یک قانون انجمنی باید قانونی جذاب¹ و با مقدار پشتیبان و اطمینان بالایی باشد؛ لذا الگوریتم‌های کاوش قوانین انجمنی بایستی چند هدفه بوده و به طور هم‌زمان همه معیارها را برای کشف قوانین انجمنی مفید بررسی نمایند.

در سال‌های اخیر، الگوی هوش جمعی² و به خصوص از این میان بهینه سازی کلونی مورچه‌ها³ و بهینه سازی گروه ذرات⁴، توجه زیادی را در تحقیقات به خود جلب کرده است. همین‌طور، این الگوریتم‌ها رایج‌ترین متاهوریستیک‌های⁵ هوش جمعی برای داده کاوی هستند.

هوش جمعی، شاخه نسبتاً جدیدی از هوش مصنوعی است که به بررسی هوش جمعی حاصل از گروهی از عامل‌های ساده می‌پردازد. الگوریتم‌های هوش جمعی مبتنی بر رفتار اجتماعی هستند که در

¹ Interesting

² Swarm intelligence

³ Ant Colony Optimization (ACO)

⁴ Particle Swarm Optimization (PSO)

⁵ Metaheuristic

طبیعت قابل مشاهده است، مانند کلونی مورچه‌ها، دسته پرندگان، گروه ماهی‌ها و کندوی زنبور عسل که در آن‌ها تعدادی از اعضاء که قابلیت‌های محدودی دارند با همکاری هم قادرند به راه‌حلهایی برای مسائل پیچیده، دست پیدا کنند.

اولین الگوریتم ACO، توسط Dorigo و همکارانش در سال 1992 در [8] جهت حل مسئله فروشنده دوره‌گرد ارائه شد. این الگوریتم از رفتار جستجوی غذا در مورچه‌های حقیقی الهام گرفته شده است. اساس این رفتار، ارتباط غیر مستقیم مورچه‌ها از طریق دنباله‌های شیمیایی فرومون است که آن‌ها را قادر به کشف کوتاه‌ترین مسیر از لانه تا منبع غذا می‌سازد.

تا کنون الگوریتم‌های ACO متعددی برای حل مسائل بهینه‌سازی گسسته¹ و پیوسته²، ارائه شده است. اخیراً، نسخه پیوسته متاهوریستیک بهینه‌سازی کلونی مورچه‌ها، با هدف حل مسائل بهینه‌سازی پیوسته توسط Socha و Dorigo، به نام ACO_R، در [8]، [10] و [11] ارائه شده است. نسخه ACO_R، به جای استفاده از تابع احتمالی که، یک مقدار برای هر رأس (که نشان دهنده یک مقدار منفرد برای متغیر گسسته است) در نظر می‌گیرد، از یک توزیع احتمال که شامل توابع چگالی احتمال گوسی متعددی است و به راحتی قابل نمونه‌گیری می‌باشد، استفاده می‌کند. این الگوریتم در داده‌کاوی برای آموزش وزن‌های یک شبکه عصبی در [12] به کار رفته است. بر اساس آزمایشات انجام گرفته روی سه مجموعه داده تشخیص پزشکی، آن‌ها دریافته‌اند، نسخه هیبرید، که الگوریتم Levenberg-Marquart [13] را با ACO_R ترکیب می‌کند، بهتر از الگوریتم‌های Levenberg-Marquart و انتشار به عقب³ عمل می‌کند. نتایج نشان می‌دهد، با وجود این که محیط آزمایشگاهی نسبتاً محدود است، پتانسیل کاربرد ACO_R در داده‌کاوی امید بخش است.

تا کنون هیچ مطالعه‌ای صورت نگرفته است که از ACO_R برای کاوش قوانین انجمنی عددی استفاده کرده باشد. در این پایان نامه، به توصیف چگونگی پیدا کردن بازه‌های پر تکرار دادگان عددی و

¹ Combinatorial optimization problem

² Continuous optimization problem

³ Back Propagation

قوانین انجمنی حاصل از آن‌ها توسط ACO_R، در یک مرحله و بدون نیاز به مشخص کردن حد اقل پشتیبان و اطمینان قوانین پرداخته شده است. در الگوریتم پیشنهادی، تابع هدفی که قرار است بهینه شود، تابعی وزن دار بوده، که سه معیار پشتیبان، اطمینان و جذابیت را به طور هم‌زمان بهینه می‌کند و عملکرد چند هدفه دارد. نتایج نشان می‌دهد که قوانین انجمنی دقیق و صحیحی از این طریق، قابل تولید خواهند بود.

در فصل اول این پایان نامه، مروری بر کلیات تحقیق ارائه می‌شود که شامل هدف تحقیق، پیشینه آن و روش تحقیق می‌باشد. سپس در فصل دوم، مفاهیم داده کاوی و کاوش قوانین انجمنی و هم چنین الگوریتم‌های بهینه سازی کلونی مورچه‌ها تحت عنوان ادبیات موضوع مطرح می‌شود. فصل سوم به توصیف الگوریتم پیشنهادی اختصاص دارد و نتایج و ارزیابی الگوریتم پیشنهادی و مقایسه آن با روش‌های قبلی در فصل چهارم آورده شده است. در نهایت در فصل پنجم، جمع‌بندی صورت گرفته و کارهای آینده پیشنهاد می‌شود.

فصل اول

کلیات

فصل اول : کلیات

1-1- هدف

یکی از چالش‌های عمده، در زمینه کاوش قوانین انجمنی، افزایش دقت و صحت قوانین انجمنی می‌باشد. با استفاده از متاهوریستیک‌هایی از قبیل ACO می‌توان الگوریتمی برای کشف قوانین انجمنی ارائه داد. اما این الگوریتم برای مسائل بهینه سازی ترکیبی مطرح شده و قادر به سر و کار داشتن با دادگان عددی، که حوزه مقادیرشان پیوسته است، نخواهد بود و نیاز به پیش پردازش داده‌های عددی و گسسته سازی آن‌ها قبل از اجرای الگوریتم دارد، که این امر، خود موجب از دست رفتن اطلاعات می‌شود.

در این پایان نامه، نظر به این که کشف قوانین انجمنی عددی یک مسئله بهینه سازی سخت به شمار می‌رود، یک الگوریتم کاوش قوانین انجمنی چند هدفه با استفاده از الگوریتم بهینه سازی کلونی مورچه‌ها برای دامنه مقادیر پیوسته (ACO_R)، ارائه شده است. الگوریتم ACO_R در حوزه بهینه سازی پیوسته کاربرد دارد و بدون نیاز به گسسته سازی، قادر به پیدا کردن بازه بهینه مقادیر یک متغیر پیوسته است.

در این جا هدف، ارائه الگوریتمی چند هدفه است که با استفاده از الگوریتم ACO_R ، قادر باشد در یک مرحله و بدون نیاز به مشخص کردن معیارهای حداقل پشتیبان و حداقل اطمینان قوانین، بازه‌های پر تکرار دادگان عددی و قوانین انجمنی عددی تشکیل شونده از این بازه‌ها را با توجه به معیار پشتیبان و اطمینان قوانین و با در نظر گرفتن میزان جذابیت آن‌ها کشف کند.

نتایج نشان می‌دهد که، الگوریتم پیشنهادی قادر به کشف قوانین انجمنی عددی جذاب با میزان پشتیبان و اطمینان بالا می‌باشد.

1-2- پیشینه تحقیق

همان‌طور که گفته شد، در یک قانون انجمنی عددی، اقلام داده، محدود به نوع بولین نبوده و می‌توانند عددی و یا صریح باشند. لذا، این قوانین گویاتر بوده و قادرند اطلاعات بیشتر و مفیدتری را در اختیار قرار دهند. تا کنون الگوریتم‌های متعددی در زمینه کاوش قوانین انجمنی عددی ارائه شده است.

Agrawal و Sirkan در سال 1996 در [14]، الگوریتمی برای کاوش قوانین انجمنی عددی ارائه دادند که دامنه صفات عددی را به بازه‌های کوچک‌تر افراز نموده و سپس بازه‌هایی که در همسایگی هم قرار دارند را ترکیب می‌کند و آن‌ها را به بازه‌های بزرگ‌تر تبدیل می‌کند، به گونه‌ای که بازه‌های ترکیب شده، معیار حداقل پشتیبان را ارضا کنند. در حقیقت، در این جا یک مسئله عددی با استفاده از گسسته سازی به یک مسئله بولین تبدیل می‌شود.

Yang و Miller در سال 1997 در [15]، با به کار بردن روش خوشه‌بندی Birch در [16]، بازه‌ها را مشخص نموده و یک فرآیند کاوش قوانین انجمنی مبتنی بر فاصله را پیشنهاد دادند، که بازه‌ها را از نظر معنایی بهبود می‌بخشد. Lent و Swami نیز در سال 1997 در [17]، یک الگوریتم با استفاده از روش‌های هندسی، به نام BitOP، برای خوشه‌بندی داده‌های عددی ارائه دادند. آن‌ها نشان دادند که خوشه‌بندی می‌تواند راه حلی برای پیدا کردن نواحی معنادار و پشتیبانی از کشف قوانین انجمنی باشد.

Colla و Vannucci در سال 2004 در [18]، یک شبکه عصبی ارائه دادند که بر محدودیت‌های متدهای گسسته سازی بدون نظارت¹ پیشنهاد شده، که در تلاش است تا توزیع داده‌های نمونه‌گیری شده اصلی را حفظ کند، غلبه می‌کند.

Ke و همکارانش در سال 2006 در [19]، یک روش ارائه دادند که از گسسته سازی، برای کاوش قوانین انجمنی عددی استفاده می‌کند. در این روش، گرافی ساخته می‌شود که ارتباطات آن نشانگر ارتباط اطلاعاتی قوی بین صفات عددی می‌باشد. سپس از گروه‌های مجزایی که در گراف به وجود می‌آید برای

¹Unsupervised

هرس کردن صفات عددی که اطلاعات مفیدی در اختیار نمی‌گذارند، استفاده می‌شود. در نتیجه این عمل، بازه‌های مشترک این صفات نیز هرس می‌شوند.

برخی از محققین، اقدام به افراز داده‌های عددی به کمک مجموعه‌های فازی نمودند. به قوانینی که از این طریق به دست می‌آید قوانین انجمنی فازی گفته می‌شود [20]. قوانین انجمنی فازی در قالب $A=X \rightarrow B=Y$ هستند که در آن A و B شامل مجموعه اقلامی هستند که زیر مجموعه‌ای از صفات عددی می‌باشند. X و Y نیز شامل مجموعه‌های فازی مربوط به مجموعه صفات متناظر در A و B هستند.

Lindell و Aumann در سال 2003 در [21]، از توزیع یک مقدار عددی به عنوان معیاری برای قرار دادن آن متغیر عددی در قانون انجمنی استفاده کردند. آن‌ها این مسئله را مطرح ساختند که یک قانون انجمنی را می‌توان به عنوان یک زیرمجموعه از دادگان در نظر گرفت که رفتار جالب و جذابی را به نمایش می‌گذارد. به عبارت دیگر، تالی یک قانون رفتار جالبی را نشان می‌دهد که همان رخداد مقدم است.

یکی از محدودیت‌های این الگوریتم‌های کاوش قوانین انجمنی، تعداد متغیرهای مجاز در یکی از قسمت‌های قانون انجمنی، یعنی در مقدم و یا در تالی قانون، بود. هم‌چنین، گسسته سازی صفات عددی، ناگزیر، منجر به از دست رفتن اطلاعات می‌شود. زیرا این امر موجب از دست دادن توزیع اصلی صفات‌های عددی شده و ممکن است بازه‌های گسسته سازی شده باعث پنهان شدن قوانین شوند. دلیل این امر این است که، اگر بازه‌ها خیلی بزرگ باشند، ممکن است قوانین مهمی که نیاز به دقت بالاتری دارند از دست بروند؛ همین‌طور اگر بازه‌ها خیلی کوچک باشند، این امکان وجود دارد که به اندازه کافی داده برای کاوش قوانین وجود نداشته باشد.

برخی از محققان برای پیدا کردن بازه‌های مقادیر عددی از ابزار هندسی استفاده کرده‌اند [22]. با این حال، مقدم قوانین به دست آمده، محدود به وجود دقیقاً یک مقدار صریح بود.

مشکل اصلی همه این روش‌ها آماده‌سازی داده‌ها پیش از اعمال الگوریتم است. این آماده‌سازی، چه توسط کاربر صورت گیرد و چه از طریق یک فرآیند خودکار، باعث از دست رفتن اطلاعات می‌شود. زیرا قوانین، جدا از افرازهایی که قبلاً از صفات عددی ایجاد شده‌اند، تولید می‌شوند. ممکن است داده‌های ایجاد

شده برای صفات عددی برای کاربر معنادار نبوده و به اندازه کافی موجز و گویا نباشند تا کاربر بتواند از قوانین کشف شده، اطلاعات با ارزش و کافی به دست آورد. به علاوه، روش‌های پیشنهاد شده، به غیر از مجموعه‌های فازی، ممکن است معایبی داشته باشند. اولین مسئله این است که مرزهای بین بازه‌ها طوری از هم تفکیک شده‌اند که الگوریتم‌ها اعضای نزدیک به مرز بازه‌ها را یا نادیده می‌گیرند و یا بیش از حد به آن‌ها توجه و تأکید دارند. مسئله دیگر این است که، تشخیص درجه عضویت برای روش‌های بازه‌ای بدون داشتن دانش قبلی کار آسانی نیست. به طور مشابه، افراز بازه‌ها با استفاده از مجموعه‌های فازی نیز کار راحتی نیست، زیرا تشخیص مناسب‌ترین مجموعه فازی برای مقادیر داده‌های عددی کار سختی است [23] [24].

به طور کلی، خصوصیات صفات عددی ناشناخته بوده و این امر که همیشه متخصصین سیستم، مناسب‌ترین مجموعه‌های فازی را فراهم می‌کنند امری غیر واقعی است. از این روی، برخی از محققان در [2]، اقدام به طرح الگوریتمی تکاملی برای به دست آوردن مجموعه‌های فازی مقادیر عددی در فاز پیش پردازش داده‌ها، نمودند.

در حقیقت، همان طور که گفته شد، کاوش قوانین انجمنی عددی یک مسئله بهینه سازی سخت است تا یک مسئله گسسته سازی ساده؛ به همین دلیل، برخی از محققان این مسئله را به عنوان یک مسئله بهینه سازی در نظر گرفته و سعی در کاوش قوانین انجمنی با استفاده از الگوریتم‌های بهینه سازی سراسری داشته‌اند.

ایده استفاده از الگوریتم تکاملی¹ برای کاوش مجموعه اقلام مکرر، توسط Mata و همکارانش در سال 2002 در [25]، مطرح شد. آن‌ها با استفاده از الگوریتم تکاملی الگوریتمی ارائه دادند که به کاوش مجموعه اقلام مکرر داده‌های عددی، بدون نیاز به گسسته سازی آن‌ها می‌پردازد. لذا، در این روش فقط معیار ضریب پشتیبان بهینه می‌شود.

Alatas و همکارانش در سال 2006 در [26]، الگوریتم تکاملی مؤثرتری را ارائه دادند. این الگوریتم تکاملی به گونه‌ای دست کاری شده است که، تمام قوانین را با یک بار اجرا، و به طور کارا استخراج

¹ Evolutionary Algorithm (EA)

می‌کند. این الگوریتم، قادر است قوانین انجمنی مثبت و منفی را کاوش کند. منظور از قوانین انجمنی منفی قوانینی است که در آن‌ها عدم عضویت یک صفت عددی در یک بازه نیز مد نظر قرار گرفته می‌شود. به عنوان مثال، یک قانون انجمنی حاصل از این الگوریتم می‌تواند به صورت زیر باشد:

$$B \in \neg[b_1, b_2] \rightarrow C \in [c_1, c_2] \wedge D \in [d_1, d_2]$$

این الگوریتم مستقل از پایگاه داده‌ها بوده و در آن نیازی به مشخص کردن معیار حداقل پشتیبان و حداقل اطمینان قوانین نیست.

Alatas و همکارانش در سال 2007 در [27]، یک الگوریتم تکاملی تفاضلی چند هدفه برای کاوش قوانین انجمنی عددی ارائه دادند. Aourssi, Salleb و همکارانشان در سال 2007 در [28]، الگوریتم QuantMiner را ارائه دادند که از الگوریتم ژنتیک¹ برای کاوش قوانین انجمنی عددی استفاده می‌کند.

بار دیگر، Alatas و همکارانش در سال 2008 در [7]، یک الگوریتم بهینه‌سازی اجتماع ذرات سخت² ارائه کردند که مبتنی بر الگوهای سخت است. الگوهای سخت بر اساس مفهوم مقادیر سخت³ بوده و شامل یک حد بالا و حد پایین می‌باشند، لذا می‌توان از آن‌ها برای نمایش بازه‌ای از مقادیر متغیرها استفاده کرد. این الگوریتم در کاوش قوانین انجمنی عددی چند هدفه به کار گرفته شده است. کاوش قوانین انجمنی عددی در این روش به گونه‌ای است که هم‌زمان با جستجوی بازه‌های صفات عددی، قوانین انجمنی عددی نیز کشف می‌شوند. در این الگوریتم نیز نیازی به مشخص کردن حداقل پشتیبان و حداقل اطمینان قوانین نبوده و عملیات مستقل از پایگاه داده انجام می‌گیرد. این الگوریتم نیز حد بالا و پایین بازه‌ها را ثابت در نظر گرفته و طی عملیاتی پس اتمام جستجوی قوانین انجمنی به کاهش طول بازه‌ها می‌پردازد.

¹Genetic Algorithm

²Rough Particle Swarm Optimization

³Rough Value

Nasiri و همکارانش در سال 2010 در [29]، یک الگوریتم کاوش قوانین انجمنی عددی چند هدفه پیشنهاد دادند که از الگوریتم شبیه‌سازی تبرید¹ برای کاوش قوانین انجمنی عددی استفاده می‌کند. الگوریتم شبیه‌سازی تبرید از فرآیند سرد کردن تدریجی جامدات الهام گرفته است. این الگوریتم توسط Metropolis و همکارانش در سال 1953 در [30] ارائه شد و توسط Kirkpatrick و همکارانش در [31] مورد استفاده قرار گرفت. الگوریتم کاوش قوانین انجمنی چند هدفه با استفاده از شبیه‌سازی تبرید، از یک رشته عددی برای نمایش مرز بازه‌های عددی و مقدم و تالی قوانین، استفاده می‌کند. در این الگوریتم هنگام کشف قوانین انجمنی، بازه‌های عددی، ثابت در نظر گرفته می‌شوند؛ زیرا الگوریتم شبیه‌سازی تبرید برای حل مسائل بهینه‌سازی ترکیبی مطرح شده است. این امر یکی از ضعف‌های آن به شمار می‌رود و از تنوع قوانین انجمنی عددی حاصل می‌کاهد.

هم‌چنین Qodmanan و همکارانش در سال 2011 در [32]، با به کار بردن یک الگوریتم ژنتیک چند هدفه در کاوش قوانین انجمنی، روشی را برای کاوش قوانین ارائه دادند که نیاز به تعیین معیار حداقل پشتیبان و حداقل اطمینان ندارد.

اولین الگوریتم داده‌کاوی با استفاده از ACO، به نام Ant-Miner توسط Parpinelli و همکارانش در سال 2002 در [33]، جهت رده‌بندی داده‌ها ارائه شد. قوانین استخراج شده توسط این الگوریتم از سادگی و دقت پیش‌بینی قابل توجهی نسبت به سایر الگوریتم‌های رده‌بندی برخوردارند. از آن زمان تا کنون، الگوریتم‌های داده‌کاوی ترتیبی و موازی متعددی با استفاده از ACO در زمینه رده‌بندی و خوشه‌بندی ارائه شده است. با این وجود، از نسخه‌های مختلف ACO در زمینه کاوش قوانین انجمنی استفاده چندانی نشده است.

در سال 2007 میلادی Kuo و همکارانش در [34]، برای نخستین بار، از طریق یکپارچه کردن تحلیل خوشه‌بندی و سیستم کلونی مورچه‌ها² که در زمینه بهینه‌سازی ترکیبی کاربرد دارد، الگوریتمی

¹ Simulated Annealing (SA)

² Ant Colony System (AS)

برای کاوش قوانین انجمنی دادگان بیمه سلامتی تایوان ارائه دادند. هم چنین، اتابکی در [35]، از نسخه سیستم مورچگان¹ برای کاوش قوانین انجمنی در یک محیط توزیعی استفاده کرده است. این الگوریتم‌ها به کاوش قوانین انجمنی در مجموعه دادگانی می‌پردازند که ماهیت گسسته دارند، زیرا ACO یک الگوریتم بهینه سازی ترکیبی² است. در این پایان نامه به استفاده از الگوریتم کلونی مورچه‌ها برای دامنه پیوسته (ACO_R)، که یک الگوریتم بهینه سازی پیوسته³ است، در کاوش قوانین انجمنی عددی چند هدفه پرداخته شده است.

1-3- روش کار و تحقیق

روش تحقیق بر اساس تجزیه و تحلیل، طراحی و پیاده سازی و ارزیابی روش پیشنهادی است. در مرحله تجزیه و تحلیل، ملزومات روش شناسایی می‌شود؛ و پس از طراحی روش آن، در یک محیط، پیاده سازی و ارزیابی خواهد شد. برای به دست آوردن ایده اولیه الگوریتم پیشنهادی در این پایان نامه ابتدا به مطالعه الگوریتم‌های ACO و ACO_R و کارهای مرتبط در زمینه کاوش قوانین انجمنی پرداخته شده است. پس از به دست آوردن ایده، به پیاده سازی الگوریتم پیشنهادی و استخراج نتایج پرداخته شده است. در نهایت، نقاط ضعف و قوت این الگوریتم بررسی شده و نتایج آن با کارهای مرتبط مقایسه شده است.

¹ Ant System

² Combinatorial Optimization Algorithm

³ Continuous Optimization Algorithm