دانشگاه صنعتی اصفهان

دانشکده برق و کامپیوتر

توسعه فازی الگوریتم Apriori برای کاوش قوانین وابستگی

پایان‌نامه کارشناسی ارشد مهندسی کامپیوتر- نرم‌افزار

زهره احمدی دستجردی

استاد راهنما

دکتر محمد حسین سرایی

۱۳۹۰

به نام خدا

دانشگاه صنعتی اصفهان

دانشکده برق و کامپیوتر

توسعه فازی الگوریتم Apriori برای کاوش قوانین وابستگی

پایان‌نامه کارشناسی ارشد مهندسی کامپیوتر- نرم‌افزار

زهره احمدی دستجردی

استاد راهنما

دکتر محمد حسین سرایی

۱۳۹۰

دانشگاه صنعتی اصفهان

دانشکده برق و کامپیوتر

پایان‌نامه‌ی کارشناسی ارشد رشته‌ی مهندسی کامپیوتر- نرم‌افزار خانم زهره احمدی دستجردی

تحت عنوان

توسعه فازی الگوریتم Apriori برای کاوش قوانین وابستگی

در تاریخ ۱۳۹۰/۰۲/۰۵ توسط کمیته‌ی تخصصی زیر مورد بررسی و تصویب نهایی قرار گرفت.

۱- استاد راهنمای پایان‌نامه	دکتر محمد حسین سرایی
۲- استاد مشاور پایان‌نامه	دکتر سید رضا حجازی
۳- استاد داور	دکتر مریم ذکری
۴- استاد داور	دکتر محمدعلی منتظری
سرپرست تحصیلات تکمیلی دانشکده	دکتر سید محمود مدرس هاشمی

از اساتید محترم و همه دوستانی که مـرا در انجـام ایـن تحقیـق یـاری کردنـد تشـکر می‌کنم. همچنین از حمایت‌های پـدر و مـادرم کـه هـر چـه دارم بـه ایشـان مـدیونم سپاسگذارم و از همسر عزیزم که در تمامی مراحل انجام این تحقیق همواره همـراه، همکار و مشوق من بود کمال تشـکر را دارم. از درگـاه خداونـد متعـال، سـلامتی و طول عمر با عزت در پناه حضرت صاحب الزمان (عج) را برایشان مسئلت دارم.

کلیه‌ی حقوق مادی مترتب بر نتایج مطالعات، ابتکارات و نوآوری‌های ناشی از تحقیق موضوع این پایان‌نامه متعلق به دانشگاه صنعتی اصفهان است.

تقدیم به:

مولانا صاحب الامر و الزمان عجل ا... تعالی فرجه

و

مادر عزیزشان حضرت زهرا سلام ا... علیها

فهرست مطالب

عنوان **صفحه**

فهرست مطالب ..... هشت
چکیده ۱

فصل اول : مقدمه

۱-۱ مقدمه ۲
۱-۲ داده‌کاوی ۲
۱-۲-۱ مفاهیم پایه در داده‌کاوی ۳
۱-۲-۲ مدل‌های داده‌کاوی ۴
۱-۳ منطق فازی ۵
۱-۴ اهمیت موضوع ۷
۱-۵ نتیجه‌گیری ۸

فصل دوم : کارهای انجام شده در زمینه کاوش قوانین وابستگی

۲-۱ مقدمه ۹
۲-۲ مروری بر الگوریتم‌های کاوش قوانین وابستگی ۹
۲-۳ کاوش قوانین وابستگی فازی ۱۳
۲-۴ نتیجه‌گیری ۱۷

فصل سوم : الگوریتم یافتن قوانین وابستگی فازی چند سطحی در حالت طبقه‌بندی فازی

۳-۱ مقدمه ۱۸
۳-۲ طبقه‌بندی فازی ۱۸
۳-۳ مفهوم کلی تحلیل عاملی ۱۹
۳-۴ تولید طبقه‌بندی فازی با استفاده از تحلیل عاملی ۲۱
۳-۵ الگوریتم توسعه داده شده ۲۲
۳-۵-۱ پیش‌پردازش داده‌ها ۲۳
۳-۵-۲ تعریف نمادهای مورد نیاز ۲۴
۳-۵-۳ فلوچارت الگوریتم FTARM ۲۵
۳-۵-۴ گام‌های الگوریتم FTARM ۲۵
۳-۵-۵ پیاده سازی گام‌های الگوریتم FTARM روی فلوچارت ۲۸
۳-۵-۶ نکاتی در مورد الگوریتم FTARM ۲۹
۳-۶ نتیجه‌گیری ۳۰

فصل چهارم : نتایج محاسباتی

۴-۱ مقدمه ۳۱
۴-۲ پیاده‌سازی الگوریتم FTARM روی مجموعه داده‌های هواشناسی ۳۱
۴-۲-۱ پیش‌پردازش داده‌ها ۳۲
۴-۲-۲ پیاده‌سازی گام‌های الگوریتم FTARM روی داده‌های هواشناسی ۳۵
۴-۳ پیاده‌سازی الگوریتم FTARM روی مجموعه داده‌های حمل و نقل جاده‌ای کشور ۴۰
۴-۳-۱ پیش‌پردازش داده‌ها ۴۲
۴-۳-۲ پیاده‌سازی گام‌های الگوریتم FTARM روی داده‌های حمل و نقل ۴۴
۴-۴ نتیجه‌گیری ۵۰

فصل پنجم : نتیجه‌گیری و پیشنهادات

۵-۱ مقدمه ۵۱
۵-۲ جمع‌بندی و نتیجه‌گیری ۵۱
۵-۳ پیشنهاداتی برای مطالعات آتی ۵۳

مراجع ۵۵

چکیده

در دنیای کنونی، اطلاعات به‌عنوان یکی از فاکتورهای تولیدی مهم مطرح است. در نتیجه، تلاش برای استخراج اطلاعات از داده‌ها از جمله چالش‌های اساسی در صنعت اطلاعات و حوزه‌های وابسته به آن است. حجم داده‌ها دائماً در حال رشد است و در همه‌ی محیط‌ها و به صورت متنوع در قالب‌های مختلف وجود دارد، که این نمایانگر پیچیدگی کار تبدیل داده‌ها به اطلاعات است. داده کاوی یکی از پیشرفت‌های اخیر در زمینه فناوری مدیریت داده‌هاست. در داده‌کاوی تئوری‌های پایگاه‌داده‌ها، هوش مصنوعی، یادگیری ماشین و علم آمار را در هم می‌آمیزند تا زمینه‌ای کاربردی فراهم شود. داده‌کاوی روش‌های مختلفی دارد که از جمله مهمترین آنها قوانین وابستگی است. یکی از الگوریتم‌های رایج برای کشف قوانین وابستگی الگوریتم Apriori است. تاکنون مطالعات متعددی جهت توسعه کاوش قوانین وابستگی صورت گرفته است که پایه بسیاری از آنها، الگوریتم Apriori بوده است. آنچه دراین تحقیق بر آن تأکید می‌شود نگاهی جدید بر مبنای منطق فازی جهت کشف قوانین وابستگی است. پس از مطرح شدن منطق فازی، استفاده از آن در سیستم‌های هوشمند به دلیل شباهت با نحوه استدلال انسان به سرعت وارد داده‌کاوی شد. در مقایسه با منطق کلاسیک، منطق فازی جهان‌بینی نوینی است که با نیازهای دنیای پیچیده امروز سازگارتر است. در این نظریه درستی گزاره‌ها عددی بین صفر و یک است در صورتی که در منطق کلاسیک گزاره‌ها یا درست و یا غلط هستند. لذا نتایج حاصل از منطق فازی در مسائل متعدد، کاربردی‌تر و واقعی‌تر خواهد بود. در این تحقیق ضمن مطالعه کارهای انجام شده تاکنون و بررسی نقاط ضعف و قوت آنها، همچنین بررسی مسائل کاربردی و مطرح کنونی، الگوریتمی جدید با نام FTARM برای کشف قوانین وابستگی فازی از روی مجموعه داده‌های عددی مختلف، با در نظر گرفتن یک طبقه‌بندی چندسطحی فازی روی داده‌ها و همچنین قابلیت متفاوت در نظر گرفتن کمترین پشتیبان به ازای اقلام مختلف پیشنهاد شده است. یکی از ایده‌های مهم به کار رفته در این الگوریتم توانایی آن در استفاده روی مجموعه‌داده‌های مختلف با انواع متنوعی از مشخصه‌ها اعم از مشخصه‌های گسسته و پیوسته است. قابلیت الگوریتم ارائه شده در لحاظ نمودن طبقه‌بندی فازی و کشف قوانین وابستگی در تمام سطوح آن با در نظر گرفتن فرض‌های فوق یکی از نقاط قوت پرکاربرد آن است. همچنین در این تحقیق یک روش برای به دست آوردن طبقه‌بندی فازی بر مبنای روش چند متغیره آماری تحلیل عاملی ارائه شده است. استفاده از این روش به منظور استخراج طبقه‌بندی فازی در مواردی توصیه می‌گردد که طبقه‌بندی در دسترس نبوده و یا تولید آن مشکل باشد. از این رو با کمک تحلیل عاملی می‌توان به استخراج طبقه‌بندی فازی پرداخت. این روش با تقسیم‌بندی شاخص‌های سطح پایین‌تر، عوامل پنهان را در سطح بالاتر طبقه‌بندی تولید می‌نماید. روش‌های فوق بر روی دو مثال کاربردی به منظور نشان دادن کارکرد الگوریتم پیاده‌سازی شده است. مثال اول در زمینه هواشناسی و شامل داده‌های میانگین دمای هوای ماهیانه ۲۶ شهر کشور ایران در بین سال‌های ۱۹۹۰ تا ۲۰۰۰ میلادی است. مثال دوم برگرفته از حمل و نقل جاده‌ای کشور بوده و در آن ۱۴ شاخص مهم در زمینه ایمنی حمل و نقل جاده‌ای کشور در سال ۱۳۸۷ مورد استفاده قرار گرفته است.

کلمات کلیدی: ۱- قوانین وابستگی فازی ۲- کمترین پشتیبان ۳- طبقه‌بندی فازی ۴- مجموعه اقلام

فصل اول
مقدمه

۱-۱ مقدمه

در دنیای امروز که آن را عصر فناوری اطلاعات می‌نامند، تلاش برای استخراج اطلاعات از داده‌ها از جمله چالش‌های اساسی در صنعت اطلاعات و حوزه‌های وابسته به آن به شمار می‌رود. به همین منظور روش‌های مختلفی برای به دست آوردن اطلاعات ارائه شده است که یکی از مهمترین و پرکاربردترین آنها داده‌کاوی است. در حال حاضر یکی از رویکردهای مطرح برای توسعه داده‌کاوی، استفاده از منطق فازی است. لذا در این فصل پس از معرفی داده‌کاوی و روش‌های آن و نیز توصیف منطق فازی، اهمیت انجام تحقیق در این زمینه بیان می‌شود.

۱-۲ داده‌کاوی

داده‌کاوی مجموعه‌ای از فنون است که به شخص امکان می‌دهد تا فراتر از داده پردازی معمولی حرکت کند و به استخراج اطلاعاتی که در انبوه داده‌ها مخفی و یا پنهان است کمک می‌کند. انگیزه برای گسترش داده‌کاوی بطور عمده از دنیای تجارت در دهه‌ی ۱۹۹۰ پدید آمد. در داده‌کاوی از بخشی از علم آمار به نام تحلیل اکتشافی داده‌ها[1] استفاده می‌شود که در آن برکشف اطلاعات نهفته و ناشناخته از درون حجم انبوه داده‌ها تأکید

[1] Exploratory Data Analysis

می‌شود. علاوه بر این داده‌کاوی با هوش مصنوعی و یادگیری ماشین نیز ارتباط تنگاتنگی دارد، لذا می‌توان گفت در داده‌کاوی تئوری‌های پایگاه داده‌ها، هوش مصنوعی، یادگیری ماشین و علم آمار را در هم می‌آمیزند تا زمینه‌ای کاربردی فراهم شود.

باید توجه داشت که اصطلاح داده‌کاوی زمانی به کار برده می‌شود که با حجم انبوهی از داده‌ها مواجه باشیم. تقریباً در تمامی منابع داده‌کاوی بر این مطلب تاکید شده که هر چه حجم داده‌ها بیشتر و روابط میان آن‌ها پیچیده‌تر باشد دسترسی به اطلاعات نهفته در میان داده‌ها مشکل‌تر می‌شود و نقش داده‌کاوی به عنوان یکی از روش‌های کشف دانش، روشن‌تر می‌گردد.

۱-۲-۱ مفاهیم پایه در داده‌کاوی

در داده‌کاوی معمولاً به کشف الگوهای مفید از میان داده‌ها اشاره می‌شود. منظور از الگوی مفید، مدلی در داده‌ها است که ارتباط میان یک زیر مجموعه از داده‌ها را توصیف کرده و معتبر، ساده، قابل فهم و جدید باشد. در متون دانشگاهی تعاریف گوناگونی برای داده‌کاوی ارائه شده است. در برخی از این تعاریف داده‌کاوی در حد ابزاری که کاربران را قادر به ارتباط مستقیم با حجم عظیم داده‌ها می‌سازد معرفی گردیده و در برخی دیگر، تعاریف دقیق‌تر که در آن‌ها به کاوش در داده‌ها توجه می‌شود، ارائه شده است. برخی از این تعاریف عبارتند از :

- داده کاوی عبارت است از فرایند استخراج اطلاعات معتبر، از پیش ناشناخته، قابل فهم و قابل اعتماد از پایگاه داده‌های بزرگ و استفاده از آن در تصمیم‌گیری در فعالیت‌های تجاری مهم [۱].
- اصطلاح داده کاوی به فرآیند نیمه خودکار تجزیه و تحلیل پایگاه داده‌های بزرگ به منظور یافتن الگوهای مفید اطلاق می‌شود [۲].
- داده کاوی یعنی جستجو در یک پایگاه داده برای یافتن الگوهایی میان داده‌ها [۳].

همانگونه که در تعاریف گوناگون داده‌کاوی مشاهده می‌شود، تقریباً در تمامی تعاریف به مفاهیمی چون استخراج دانش، تحلیل و یافتن الگوی بین داده‌ها اشاره شده است.

فرآیند کشف دانش از پایگاه داده‌ها شامل پنج مرحله است که عبارتند از [۴] :

- انبار داده‌ها[1]: مجموعه‌ای است موضوعی[2]، مجتمع[3]، متغیر در زمان[1] و پایدار[2] از داده‌ها که به منظور پشتیبانی از فرآیند مدیریت تصمیم‌گیری مورد استفاده قرار می‌گیرد [۱].

---

[1] data warehousing
[2] subject oriented
[3] integrated

- انتخاب داده‌ها
- تبدیل داده‌ها
- کاوش در داده‌ها (داده‌کاوی)
- تفسیر نتیجه

همانگونه که مشاهده می‌شود داده‌کاوی یکی از مراحل فرآیند کشف دانش از پایگاه داده‌ها است که نقش مهمی را در این فرآیند ایفا می‌کند. وجود اطلاعات صحیح و منسجم یکی از ملزوماتی است که در داده‌کاوی به آن نیاز است. اشتباه و عدم وجود اطلاعات صحیح باعث نتیجه‌گیری غلط و اخذ تصمیمات ناصحیح می‌گردد. اکثر سازمان‌ها دچار یک شکاف اطلاعاتی[۳] هستند . در اینگونه سازمان‌ها معمولاً سیستم‌های اطلاعاتی در طول زمان و با معماری و مدیریت‌های گوناگون ساخته شده‌اند، به طوریکه در سازمان اطلاعات یکپارچه و مشخصی مشاهده نمی‌گردد. علاوه براین برای فرآیند داده‌کاوی به اطلاعات خلاصه و مهم در زمینه تصمیم‌گیری‌های حیاتی نیاز است.

۱-۲-۲ مدل‌های داده‌کاوی

در این قسمت هفت روش مشهور داده‌کاوی به اختصار مورد بررسی قرار می‌گیرد.

وابستگی[۴]: هدف این مدل برقراری رابطه بین مواردی است که با هم در یک رکورد معین وجود دارند [۵]. تجزیه و تحلیل سبد بازار و برنامه‌های فروش دو مثالی هستند که این تکنیک به طور معمول برای آن‌ها به کار می‌رود. ابزارهای متداول برای مدل‌سازی وابستگی، آمار، الگوریتم Apriori و توسعه‌های صورت گرفته روی آن و الگوریتم FP-growth [۶] است.

دسته‌بندی[۵]: این مدل یکی از رایج‌ترین مدل‌های یادگیری در داده‌کاوی است [۷]. هدف این روش بر ساختن مدلی به منظور پیش‌بینی رفتار آینده مشتری از طریق دسته‌بندی رکوردهای پایگاه داده به تعدادی دسته از پیش تعیین شده بر اساس معیارهای معین است [۴]. ابزارهای رایج مورد استفاده در این روش شبکه‌های عصبی[۶]، درخت‌های تصمیم[۷] و قوانین اگر- آنگاه[۱] است.

---

[1] time variant
[2] non volatile
[3] information gap
[4] association
[5] classification
[6] neural network
[7] decision tree

خوشه‌بندی[۲]: عمل جداسازی یک جمعیت ناهمگن به چند خوشه همگن را گویند [۴]. این تکنیک با روش دسته‌بندی متفاوت است زیرا در آن خوشه‌ها در ابتدای شروع الگوریتم‌ها ناشناخته‌اند. به عبارت دیگر هیچ خوشه از پیش تعیین شده‌ای در این تکنیک وجود ندارد. رایج‌ترین ابزارها در خوشه‌بندی، شبکه‌های عصبی، تجزیه و تحلیل تمایزات[۳] و k-means است.

پیش‌بینی[۴]: این مدل مقادیر آینده را بر اساس الگوهای یک مشخصه تخمین می‌زند و مرتبط با مدل‌سازی و ارتباطات منطقی مدل در آینده است. پیش‌بینی تقاضا یک مثال برای مدل پیش‌بینی است. ابزارهای متداول این روش شامل شبکه‌های عصبی و تجزیه و تحلیل بقا[۵] است.

رگرسیون[۶]: رگرسیون یک مدل معروف تخمین آماری است. کاربردهای رگرسیون شامل برازش منحنی[۷]، پیش‌بینی، مدلسازی روابط علی[۸] و آزمون فرضیه‌های علمی درباره‌ی روابط بین متغیرها است. یکی از ابزارهای رایج در این زمینه رگرسیون خطی است.

شناسایی توالی[۹]: شناسایی توالی، کشف روابط یا الگوها در طول زمان است [۴]. هدف این روش مدل کردن حالات فرآیند تولید توالی یا استخراج و گزارش انحراف و روندها در طول زمان است. ابزارهای رایج این مدل نظریه مجموعه‌ها و آمار است.

مصورسازی[۱۰]: مصورسازی به نمایش داده‌ها اشاره دارد به نحوی که کاربران بتوانند الگوهای پیچیده را مشاهده کنند. این روش همراه با دیگر مدل‌های داده‌کاوی به منظور درک روشن‌تر الگوها و روابط کشف شده مورد استفاده قرار می‌گیرد. مثال‌هایی برای مصورسازی شامل گراف‌های سه‌بعدی و هایگراف[۱۱] است [۸].

۱-۳ منطق فازی

در سال ۱۹۶۵ پروفسور لطفی عسگرزاده، استاد ایرانی‌الاصل دانشگاه برکلی، اولین مقاله خود را در زمینه فازی تحت عنوان مجموعه‌های فازی منتشر کرد [۹]، در آن زمان هیچ کس باور نداشت که این جرقه‌ای خواهد بود

---

[1] if-then-else rules
[2] clustering
[3] discrimination analysis
[4] forcasting
[5] survival analysis
[6] regression
[7] curve fitting
[8] casual
[9] sequence discovery
[10] visualization
[11] Hygraph

که دنیای ریاضیات را به طور کلی تغییر دهد. گرچه در دهه ۱۹۷۰ و اوایل دهه ۱۹۸۰ مخالفان جدی برای نظریه فازی وجود داشت، اما امروزه هیچ کس نمی‌تواند ارزش‌های منطق فازی را منکر شود.

منطق فازی در مقابل منطق کلاسیک قرار دارد. منطق کلاسیک یا منطق دومقداری بیان می‌دارد که تمام حقیقت، حقیقتی سیاه و سفید، یا حقیقتی کاملاً درست یا کاملاً نادرست است. در حالیکه در منطق فازی به جای درست یا نادرست، سیاه یا سفید، صفر یا یک، سایه‌های نامحدودی از خاکستری بین سیاه و سفید وجود دارد. تمایز عمده منطق فازی با منطق دو ارزشی در این است که در منطق فازی حقیقت هم می‌تواند نادقیق باشد. در منطق فازی مجاز به استفاده از عباراتی مثل " تاحدودی درست است" یا " تقریباً غیر ممکن است" هستیم. در واقع منطق فازی نظام کاملاً منعطفی را در خدمت زبان طبیعی قرار می‌دهد. منطق فازی جهان را دقیق‌تر از منطق کلاسیک به تصویر می‌کشد چراکه دنیایی که ما در آن زندگی می‌کنیم دنیای ابهامات و عدم قطعیت است. مغز انسان نیز عادت کرده است تا در چنین محیطی فکر کند و تصمیم بگیرد و این قابلیت مغز که می‌تواند با استفاده از داده‌های نادقیق و کیفی به یادگیری و نتیجه گیری بپردازد، در مقابل منطق کلاسیک که لازمه آن داده‌های دقیق و کمی است، قابل تأمل است [۱۰].

در کل منطق فازی عبارت از "استدلال با مجموعه‌های فازی" است. همان‌طور که در رابطه با منطق کلاسیک، نظریه مجموعه‌ها مطرح می‌شود، در منطق فازی نیز نظریه مجموعه‌های فازی مطرح می‌گردد. در این نظریه هر عضو در مجموعه‌های فازی می‌تواند میزان عضویتی بین صفر و یک داشته باشد. اگر X مجموعه مرجعی باشد که هر عضو آن با x نمایش داده شود، مجموعه فازی $\widetilde{A}$ در X بوسیله زوج‌های مرتبی به صورت زیر بیان می‌شود که در آن $\mu_{\widetilde{A}}(x)$ بیانگر میزان عضویت x به مجموعه فازی $\widetilde{A}$ است [۱۱].

$$\widetilde{A} = \left\{ \left(x, \mu_{\widetilde{A}}(x) \right) \mid x \in X \right\} \qquad ۱-۱$$

اگر قوانین وابستگی را به فرم کلی $X \Rightarrow Y$ در نظر بگیریم، قوانین وابستگی فازی زمانی که X یا Y مجموعه‌های فازی باشند به همین فرم قابل تعریف است. در واقعیت مسائل مختلفی وجود دارد که منطق کلاسیک قادر به حل آن‌ها نبوده و استفاده از منطق فازی برای حل آن‌ها ضروری است. برای مثال میزان عضویت اقلام به یک مجموعه همیشه قطعی نیست و یا در کاوش قانون وابستگی تعمیم یافته، طبقه بندی[1] قطعی نیست بلکه فازی است [۱۲]. برای مثال در طبقه‌بندی سبزیجات و میوه‌جات "خیار" می‌تواند به یک میزان متعلق به دسته میوه‌ها و به میزان دیگری به متعلق به دسته سبزیجات باشد. همچنین در تصمیم گیری انسانی مدیران اغلب از مقادیر زبانی به

[1] taxonomy

جای مقادیر قطعی برای بیان قواعد تصمیم استفاده می‌کنند [۱۲]. برای مثال ممکن است قوانین وابستگی به صورت زیر باشند:

- اگر قیمت نفت خیلی زیاد شود وابستگی بودجه به قیمت نفت نیز افزایش می‌یابد.
- اگر میزان بارندگی در طول سال کم باشد، خسارت به محصولات کشاورزی افزایش می‌یابد.
- اگر دمای هوا بسیار پایین باشد، احتمال وقوع بارش به صورت برف زیاد است.
- اگر تعداد پایگاه‌های اورژانس جاده‌ای زیاد باشد، احتمال مرگ‌ومیر ناشی از تصادفات کم خواهد بود.

۱-٤ اهمیت موضوع

اصلی‌ترین دلیلی که باعث شد داده‌کاوی کانون توجهات در صنعت اطلاعات قرار بگیرد، مسئله در دسترس بودن حجم وسیعی از داده‌ها و نیاز شدید به استخراج اطلاعات و دانش سودمند از این حجم عظیم داده بود. داده‌کاوی را می‌توان حاصل سیر تکاملی طبیعی تکنولوژی اطلاعات دانست، که این سیر تکاملی ناشی از یک سیر تکاملی در صنعت پایگاه داده می‌باشد. تکامل تکنولوژی پایگاه داده و استفاده فراوان آن در کاربردهای مختلف سبب جمع‌آوری حجم فراوانی داده شده است. این داده‌های فراوان باعث ایجاد نیاز برای ابزارهای قدرتمند برای تحلیل داده‌ها گشته، زیرا در حال حاضر به لحاظ داده ثروتمند هستیم ولی دچار کمبود اطلاعات می باشیم. تاکنون مطالعات زیادی در حوزه داده‌کاوی و روش‌های متداول آن صورت گرفته که هر کدام به نوعی با افزایش دقت، سرعت و .... سعی در بهبود روند کلی استخراج اطلاعات دارند. یکی از رویکردهای مطرح کنونی برای استخراج اطلاعات مفید و کاربردی استفاده از منطق فازی برای تطبیق بیشتر روش‌ها با دنیای واقعی است.

یکی از متداول‌ترین کاربردهای داده‌کاوی که به آن اشاره شد کشف قوانین وابستگی است. پژوهشگران متعددی در زمینه توسعه روش‌ها و الگوریتم‌های کشف قوانین وابستگی به مطالعه پرداخته‌اند که یکی از این زمینه‌های توسعه، استفاده از منطق فازی است. در حقیقت زمینه توسعه الگوریتم‌های کشف قوانین وابستگی با استفاده از منطق فازی در راستای تطبیق فرضیات مسئله با شرایط واقعی و غلبه بر پیچیدگی‌های آن است؛ لذا برخی از محققین مطالعات خود را در این زمینه متمرکز نموده‌اند. از این رو در این پایان‌نامه نیز یکی از الگوریتم‌های معروف در زمینه کشف قوانین وابستگی با استفاده از منطق و فرضیات فازی مورد مطالعه و تحقیق قرار گرفته است و سعی شده تا با جامعیت بخشیدن به مفاهیم مجموعه‌های فازی و روش‌های داده‌کاوی، قوانین وابستگی فازی از یک پایگاه داده عددی تراکنشی استخراج گردد.

۱-۵ نتیجه‌گیری

در این فصل ابتدا داده‌کاوی و مفاهیم پایه آن مرور شده و سپس مدل‌های پرکاربرد در داده‌کاوی به اجمال معرفی گردید. در ادامه کلیات منطق فازی ارائه و سپس به اهمیت کاربرد آن در روش‌های داده‌کاوی و به خصوص کشف قوانین وابستگی اشاره شد.

فصل دوم
کارهای انجام شده در زمینه کاوش قوانین وابستگی

۲-۱ مقدمه

داده‌کاوی در زمینه‌های مختلف کاربرد داشته و از روش‌های مختلفی استفاده می‌کند. از جمله مهمترین این روش‌ها می‌توان به قوانین وابستگی[1] اشاره نمود. یکی از الگوریتم‌های رایج برای کشف قوانین وابستگی الگوریتم Apriori است [۱۳]. تاکنون مطالعات متعددی جهت توسعه کاوش قوانین وابستگی صورت گرفته است که پایه بسیاری از آنها، الگوریتم Apriori بوده است. در حال حاضر یکی از رویکردهای مطرح و مورد توجه برای کشف بهتر قوانین وابستگی استفاده از منطق فازی است. از آنجا که در این تحقیق تمرکز بر کشف قوانین وابستگی با استفاده از منطق فازی است، در این فصل برخی از مهمترین مطالعات صورت گرفته در زمینه کاوش قوانین وابستگی و همچنین کاوش قوانین وابستگی با استفاده از منطق فازی مورد بررسی قرار می‌گیرد.

۲-۲ مروری بر الگوریتم‌های کاوش قوانین وابستگی

یافتن قوانین وابستگی در پایگاه‌داده‌های دارای تراکنش متداول است ۱۴][. برای محاسبه قوانین وابستگی از دو مفهوم پشتیبانی و اطمینان استفاده می شود که هر کدام از آنها را می توان به صورت زیر بیان کرد.

[1] association rules