کاربرد تئوری مجموعه های راف در نظریه تصمیم گیری

آصف زارع
استادیار گروه مهندسی برق
دانشگاه آزاد اسلامی واحد گناباد

غلامرضا نساجیان
دانشجوی کارشناسی ارشدمهندسی کنترل
دانشگاه آزاد اسلامی واحد گناباد

چکیده: در این مقاله تئوری مجموعه راف که تعمیمی از نظریه های مجموعه کلاسیک بر پایه منطق سه مقداری ،جهت کار باداده های ناقص و ناسازگار و کاهش داده های مازاد بر نیاز پایگاه داده ها است معرفی می شود . پس از ارائه تاریخچه مختصری از نحوه ارائه شدن این نظریه ،مفهوم مجموعه راف ومزیت های استفاده از آن ذ کر خواهد شد.آنگاه به طور مختصر تئوری مجموعه راف در آنالیز منطقی داده و استخراج حداقل مجموعه قواعد " اگر ... آنگاه ..."ارائه شده ودر پایان کاربرد تئوری مجموعه راف در آنالیز داده که شامل کشف وابستگی ها در داده و حذف قسمتهای اضافی از داده است با مثال واقعی ارائه می شود.

واژه های کلیدی:مجموعه های راف،سیستم های اطلاعاتی ،کاهش داده ها ،ویژگیهای تصمیم ، ویژگیهای موقعیت

مقدمه:

یافتن یک واژه معادل برای عبارت"ROUGH SETS"کار مشکلی است .در فرهنگ لغت برای کلمه"ROUGH"معادلهایی مانند زبر،درشت،تقریبی،بی ادب،متلاطم و ناصاف در نظر گرفته شده است[1] که در میان این واژه هاکلمه تقریبی شباهت بیشتری با مفهوم مورد نظر بنیانگذار این نظریه دارد.اما هیچ یک از این کلمات بار معنایی خود واژه لاتین را ندارند به همین علت دراین مقاله عبارت "مجموعه های راف" به عنوان معادل آن به کار گرفته خواهدشد. تئوری مجموعه های راف دراوایل سال۱۹۸۰میلادی توسط پروفسورزدیسلاوپاولاک پایه گذاری شد.این تئوری باتحلیل جدولهای داده سروکار دارد.دراین تئوری جدولهای داده می توانند توسط اندازه گیری یاافراد متخصص وآگاه (خبره)بدست آمده باشد.هدف اصلی از تحلیل مجموعه راف به دست آوردن مفاهیم تقریبی از داده های اکتسابی می باشد. این تئوری،یک ابزار قدرتمند ریاضی برای استدلال در موارد ابهام و نایقینی است که روشهایی را برای زدودن و کاستن اطلاعات دانش نامربوط یامازاد بر نیاز ازپایگاههای داده ها مهیا می سازد.این فرایند حذف داده های زائد ،بر مبنای آموزش وظیفه اصلی سیستم،وبدون از دست دادن داده های اساسی پایگاه داده ها صورت می پذیرد. در نتیجه تقلیل اطلاعات ،مجموعه ای از قواعد تلخیص شده و پرمعنا حاصل می گرددکه کار تصمیم گیرنده را بسیار ساده تر می کند.در حقیقت میتوان گفت که مجموعه راف با کاهش فضای داده ها وبر گزیدن ترمهای مهم،یک نگاشت از فضای داده های خام و ترمها به فضای سمانتیک (مفاهیم)انجام می دهد.لذا ،با توجه به رشد انفجاری حجم اطلاعات ،مجموعه راف می توانندنقش بسیار موثری در سیستمهای پشتیبان تصمیم گیری

داشته باشند[2].تئوری مجموعه راف نقاط اشتراک زیادی با تئوری مجموعه های فازی ،تئوری شهود،روش های استدلال بولی و تحلیل تفکیکی دارد.اما تئوری مجموعه راف به عنوان یک تئوری مستقل در نظر گرفته می شود.

سیستمهای اطلاعاتی و مجموعه های راف:

در بیشترموارد،اطلاعات بصورت جدول های داده ها که به نام سیستمهای اطلاعاتی ،جدولهای ویژگی-مقدار ویا جدولهای اطلاعات در اختیار است.ستون های جدول اطلاعاتی باویژگیها،وسطرهای آن با شئ ها نام گذاری می شوند و ورودی های جدول با مقادیر آن ویژگیها پر می شوند.شئ های دارای مقادیر ویژگیهای یکسان –ازنظرآن ویژگی – مشابه تلقی می شوند و به بلوک یکسانی از تقسیم بندی (کلاسه سازی)-که توسط آن مجموعه از ویژگی ها تعیین می شوند- تعلق می یابند. میتوان از مجموعه راف در حل مسائل اساسی در زمینه تحلیل داده ها استفاده نمود،از جمله[3,4]:

* مشخص کردن مجموعه ای از اشیاءبرحسب مقادیرویژگیها
*یافتن وابستگیها بین ویژگیها
*زدودن (کاهش یا تقلیل)ویژگیهای مازاد(داده ها)
*یافتن مهمترین ویژگیها
*تولید قواعد تصمیم گیری

تئوری مجموعه راف یک روش ریاضی جدید برای تحلیل داده های هوشمند و داده کاوی می باشد.بعداز تقریبا بیست و اندی سال از پایه گذاری تئوری مجموعه راف روشهای کاربردی از ان به یک درجه مشخص از کمال رسیده است و در چند سال اخیریک رشد سریع از توجه به تئوری مجموعه راف و کاربردهایش در سرتاسر جهان به وضوح دیدمیشود.بسیاری از کارگاههای آموزشی، کنفرانسها و سمینارها در برنامه هایشان مجموعه های راف را در نظر می گیرند. نزدیک به دو هزار مقاله و چندین کتاب تا هم اکنون روی جنبه های مختلف از مجموعه های راف منتشر شده است.فلسفه مجموعه های راف بر این فرض که هر شئ از جهان را می توان به عنوان اطلاعات(داده،معرفت)در نظر گرفت بنا شده است. اشیاء توصیف شده به وسیله اطلاعات یکسان از نقطه نظر اطلاعات در دسترس درباره آنها غیر قابل تشخیص هستند.رابطه غیر قابل تشخیص بودن (رابطه علی- معلولی) به دست آمده در این روش اساس ریاضیات تئوری مجموعه های راف می باشد. هر مجموعه ای از اشیاء غیر قابل تشخیص را یک مجموعه بنیادی می نامند و شکل یک جرء اصلی (اتم) از دانش درباره جهان است. به هر اجتماعی از مجموعه های بنیادی عنوان مجموعه های کریسپ(دقیق) نسبت می دهندو در غیر اینصورت مجموعه مبهم و غیر صریح است که عنوان مجموعه راف برای آن در نظر گرفته می شود.تئوری مجموعه های راف مبتنی بر مفهوم کلاسه سازی (دسته بندی)است. برای روشنتر شدن موضوع،به عنوان مثال ،گروهی از بیماران که از یک بیماری معین رنج می برند رادرنظر بگیرید.با هر بیمار،یک فایل داده ها شامل اطلاعاتی از قبیل نام،آدرس،سن، جنسیت، دمای بدن، فشار خون و مانند آن- همراه است.تمام بیمارانی که علائم مشابهی را نشان می دهند – از نظر اطلاعات دردسترس –مانند یکدیگر هستند،ومی توان آنان را در دسته هایی- به عنوان اجزای بنیادی دانش و معرفت موجودنسبت به بیماران –کلاسه بندی نمود.این اجزاءبه نام مجموعه های بنیادی یا مفاهیم بنیادی شناخته میشوند و میتوانند به عنوان بلوک های سازنده دانش در مورد بیماران در نظر گرفته شوند. مفاهیم بنیادی می توانند با مفاهیم مختلط – یعنی مفاهیمی که به شکل یکتایی برحسب مفاهیم بنیادی تعریف می شوند- ترکیب گردند. هراجتماعی از مفاهیم بنیادی، مجموعه کریسپ- به معنای مجموعه ای با تعریف و مرز های دقیق –نامیده می شود،وهر مجموعه دیگری که کریسپ نباشد،مجموعه راف –به معنای مجموعه مبهم و نادقیق شناخته می شود.با هرمجموعه X دومجموعه کریسپ به نامهای تقریب بالایی وتقریب پایینی مجموعه X مرتبط و همراه می شود.تقریب پایینی X اجتماع تمام مجموعه های بنیادینی است که مجموعهX شامل آنها می شود،درحالیکه تقریب بالایی Xاجتماع تمام مجموعه های بنیادینی است که اشتراک ناتهی باX دارند.به عبارت روشن تر ،تقریب پایینی یک مجموعه X ،مجموعه تمام عناصری است که مطمئنا به Xتعلق دارند،درحالیکه تقریب بالایی مجموعه X، مجموعه تمام عناصری است که احتمالا به X تعلق دارند.تفاوت تقریب های بالایی و پایینی مجموعه Xدر نواحی مرزی آنان می باشد.یک مجموعه راف است اگر ناحیه مرزی ناتهی داشته باشد،ودرغیر اینصورت ،کریسپ است.عناصر ناحیه مرزی با اطلاعات دردسترس از خود مجموعه یا مکمل

آن قابل کلاسه سازی نیستند.تقریب مجموعه ها عملیات اصلی تئوری راف ست است که برای تعریف کردن وابستگی ها (کلی یاجزئی) بین ویژگیها، تقلیل(زدودن)ویژگیها تولید قواعد تصمیم و غیره به کار گرفته می شود[3,4,5].

یک مثال ساده برای سیستم های اطلاعاتی در جدول زیر آمده است[6]:

store	E	Q	L	P
1	High	Good	no	profit
2	Med	good	no	less
3	Med	good	no	profit
4	No	ave	no	less
5	Med	ave	yes	less
6	High	ave	yes	profit

در جدول زیر شش فروشگاه بر حسب چهار ویژگی E،Q،LوPتوصیف شده است:

E:اختیارات پرسنل فروشگاه Q:کیفیت کالاها و اجناس

L:موقعیت رفت و آمد زیاد P:سود یا ضرر فروشگاه

هر فروشگاه بر حسب چهار ویژگی E،Q،LوPدارای توصیف مختلفی می باشد.بنابراین فروشگاهها را می توان با بکارگیری اطلاعات فراهم شده توسط ویژگیها ازیکدیگر متمایز نمود. فروشگاه ۲و۳ برحسب ویژگیهای E،QوL غیر قابل تشخیص هستند زیرا دارای ارزش مشابه در این ویژگیها می باشند.به همین نحوفروشگاههای۱و۲و۳ بادر نظر گرفتن ویژگیهایLو Q غیر قابل تشخیص هستند زیرا دارای ارزش مشابه در این ویژگیها می باشند.می توان فروشگاههارا برحسب ویژگیهای مشابهی که دارند دریک کلاس جمع کردو به عنوان یک زیر مجموعه در نظر گرفت.به عنوان مثال می توان با توجه به ویژگیهای Lو Qهمه فروشگاهها را در کلاسهای(زیر مجموعه های) زیرطبقه بندی نمود:

{1,2.3} Q:good , L:no

{4} Q:ave , L:no

{5,6} Q:ave L: yes

سوال:آیا می توان مجموعه {1,3,6}را بر حسب ویژگیهای E،QوL توصیف کرد؟ بدیهی است که نمی توان به صورت منحصر به فرد به این سوال جواب داد زیرا فروشگاههای ۲و۳ حالتهای شبیه بر حسب ویژگی های E، Qو L دارند اما فروشگاه ۲سود داشته است درحالیکه فروشگاه ۳ ضرر داشته است.بنابراین اطلاعات درجدول برای پاسخ دادن به این سوال کافی نیست درنتیجه میتوان یک پاسخ ناقص به این سوال داد. می توان در نظر گرفت که اگر یک فروشگاه مشخص دارای ارزش high برای ویژگی E باشد آن فروشگاه دارای سود خواهدبود و اگر ارزش ویژگی E کم باشد(low) آن فروشگاه ضرر خواهد داد.بنابراین به طور مطمئن می توانیم بگوییم که فروشگاه ۱و۶ دارای سود هستند.فروشگاههای ۴و ۵ ضرر می دهند ودر مورد فروشگاههای ۲و ۳ نمی توان گفت که سود دارند یا ضرر می دهند . با بکارگیری ویژگیهای E، Q وL می توان گفت که فروشگاههای ۱و۶ مطمئنا دارای سود هستند واژه مطمئنا متعلق به مجموعه {1,3,6}است.در حالیکه فروشگاههای 1,2,3,6ممکن است دارای سود باشند واژه شاید(ممکن است) متعلق به مجموعه {1,3,6}باشند.در نتیجه می توان گفت که مجموعه {1,6}تقریب پایینی ازمجموعه {1,3,6}ومجموعه {1,2,3,6} تقریب بالایی از مجموعه {1,3,6}

می باشد. مجموعه {2,3} تفاوت بین تقریب بالایی و پایینی است و به عنوان ناحیه مرزی از مجموعه {1,3,6} در نظر گرفته می شود.

در هرسیستم اطلاعاتی ویژگیها رامی توان به دو دسته تقسیم کرد:

۱)ویژگیهای تصمیم

۲)ویژگیهای موقعیت

در جدول تصمیم مربوط به فروشگاهها ویژگی Pبه عنوان یک ویژگی تصمیم است در حالیکه ویژگیهای EوQوLبه عنوان ویژگیهای موقعیت شناخته می شوند.ویژگیهای تصمیم معین می کنند که چه تصمیم هایی باید اجرا شود تا ویژگیهای موقعیت بصورت متقاعد درآیند.در هر سطر از یک جدول تصمیم می توان قواعد تصمیم رابه فرم **اگر.....آنگاه.....** بدست آورد. به عنوان مثال در جدول تصمیم ارائه شده داریم:

if (E=high)and (Q=good)and(L=no),then(P=profit)

if (E=no)and (Q=ave)and(L=no),then(P=loss)

قواعد متناقض(مخالف):قواعدی هستند که دارای ویژگیهای موقعیت یکسان اما ویژگیهای تصمیم متفاوت می باشند.

قواعد سازگار(موافق):قواعدی هستند که دارای ویژگیهای موقعیت یکسان و ویژگیهای تصمیم یکسان می باشند.

درمثال ارائه شده (فروشگاهها) قواعد ۲ و ۳ جزء قواعد متناقض و بقیه قواعد ازقواعد سازگار می باشند. قواعد سازگار تحلیلگررا قادر می سازد تایک تصمیم منحصر به فرد بگیرد ولی در مورد قواعد متناقض نمیتواند تصمیم یکتایی رادر نظر بگیرد.

حد پایداری (اندازه استحکام)جدول تصمیم:

نسبت قواعد سازگاربه همه قواعد رادر یک جدول تصمیم حد پایداری گویند. اگر حد پایداری یک جدول تصمیم یک باشد(یعنی تمام قواعد جدول تصمیم سازگار باشند)درنتیجه می توان به صورت منحصر به فرد تصمیم گیری کرد ولی اگر حد پایداری یک سیستم اطلاعاتی کمتراز یک باشد می توان یک تصمیم نسبتا کافی رادر نظر گرفت. تئوری راف ست ابزاری را برای تولید،تحلیل و بهینه سازی مجموعه قواعد تصمیم از جدول داده ها را فراهم می کند.جدول تصمیم می تواندبه عنوان یک نتیجه از مشاهدات (اندازه گیری)،شبیه سازی کامپیوتری ومعرفت(دانش)افرادخبره بدست آید.چندین سیستم نرم افزاری مبتنی بر تئوری راف به شکل زیرمی باشد:

LERS Rough Das

Rough Class

DATALOGIC

برای مشخص کردن یک فرایند کنترلی جدول تصمیم می تواندمفید باشد. بنابراین قواعد تصمیم می تواند به عنوان قواعد کنترلی مورد استفاده قرار گیرد. در نتیجه تئوری مجموعه راف می تواند برای نتیجه گیری یک مجموعه بهینه از قواعد کنترلی از جدول سیستم به کار گرفته شود.

تئوری مجموعه راف در تحلیل قواعد اگر..... آنگاه .....[7] :

فرض کنید که Uیک مجموعه جهانی از اشیاء باشد:

$U=\{x,y,\ldots\}$

$T=\{A1,A2,\ldots,An\}$ Tمجموعه ویژگیهاست:

Dom(Ai) مجموعه ای از ارزشهای ویژگی Ai:

$Dom=dom(A1)Udom(A2)U\ldots Udom(An)$

هر عضو در مجموعه Uرامی توان به صورت منحصر به فرد به صورت یک نگاشت نمایش داد:

$t:T\rightarrow Dom\ \ t(A)\in dom$ برای هر $A\in T$

هر جدول اطلاعاتی شامل موارد زیر می باشد:

1) U={u,v,…}
2) T={A1,A2,…,An}
3) Dom(Ai)
Dom=dom(A1)Udom(A2)U…Udom(An)
4)ρ:UxT →Domتابع توصیف

تابع توصیف یک نگاشت است که:

Ai ϵ T و u ϵ U برای هر ρ(U,Ai) ϵ dom(Ai)

باید در نظر داشت که تابع توصیف ρ باعث یک مجموعه از نگاشتها می شودو هر نگاشت را یک چند تایی می نامند:
t=(ρ(U,A1), ρ(U,A2),…, ρ(U,Ai),…, ρ(U,An))
توجه کنید که چندتاییtدر یک جدول اطلاعات الزاما وابسته به شئ(موضوع) منحصربه فردی نیست و دو شئ متفاوت می توانند دارای نمایش چندتایی یکسان باشند که این در مورد پایگاه داده های وابسته غیر مجاز می باشد. جدول تصمیم(DECISION TABLE) را با مشخصه (U,T,V, ρ)نمایش می دهند که در آن Uمجموعه جهانی وTمجموعه ویژگیهاست که از اجتماع دو مجموعه غیر تهی CوDتشکیل شده است. T=CUD
C مجموعه ویژگیهای موقعیت وDمجموعه ویژگیهای تصمیم می باشد.

وابستگی های دانش:

باتوجه به قوانین ریاضی یک کلاسه سازی یا پارتیشن بندی یک تجزیه از حوزه Uبه زیر مجموعه های گسسته می باشدکه به آن کلاسهای هم ارزی نیز می گویند.هرپارتیشن توسط یک رابطه باینری تعریف می شودکه این رابطه باینری یک رابطه هم ارزی است و سه خاصیت انعکاسی(بازتابی)،متقارن وانتقالی(تراگذر)را دارا می باشد

ویژگیها و روابط هم ارزی:

در جدول اطلاعات (داده ها)هر زیر مجموعه از ویژگیها یک رابطه هم ارزی فراهم می کند. اگر مجموعه Bیک زیر مجموعه غیر تهی ازTودوشئ(موضوع)uوv(غیر قابل تشخیص توسطB)متعلق بهUباشندخواهیم داشت:
برای هر ویژگی Aiدر Bداریم: $u \equiv v(\text{mod } B)$ if $\rho(u,Ai)= \rho(v,Ai)$
علامت ≡ نشان دهنده رابطه هم ارزی است و به آن رابطه تعادل نیز می گویند و آن رابا IND(B)نمایش می دهند.کلاسهای هم ارزی شامل Uرا بوسیله IND(B)[u]یا ساده تر به صورت B[u] نشان می دهندکه Bمی تواند یک مجموعه منفرد باشد.
B={Ai}
دراین حالت به سادگی رابطه هم ارزی به صورت IND(Ai) را تعریف می کنیم وخواهیم داشت:
IND(B)=∩{IND(Ai) : Ai in B}

رابطه بالا بیان می کند که کلاسهای هم ارزی مربوط به IND(B) همه اشتراکهای ممکن از کلاسهای هم ارزی مربوط به IND(Ai) را شامل می شود. با استفاده از مجموعه Bکه زیر مجموعه غیر تهی ازTاست می توان یک مجموعه از رابطه های هم ارزی که آن را باRCol(B) نشان می دهندبدست آوردوآن را به صورت زیر تعریف می کنیم:
Rcol(B)={IND(Ai) : Ai ϵ B}

اگرمجموعه موردنظرمجموعه ای ازهمه ویژگیهاباشددرآن صورت مجموعه متشکل ازرابطه های هم ارزی به صورت زیر خواهد بود:

Rcol(T)={IND(A1),IND(A2),…,IND(A3),…,IND(An)}

پایه های دانش پاولاک(pawlak):
اگر Rcolیک مجموعه از رابطه های هم ارزی باشد،براساس تعریف زوج مرتب زیر را پایه دانش پاولاک نامیده اند.

K=(U,Rcol)

پروفسور زدیسلاو پاولاک مجموعه رابطه های هم ارزی را دانش(knowledge)نامیدزیرا دانش ما درباره یک حوزه اغلب بوسیله کلاس بندیها نمایش داده می شود.

PوQدو رابطه هم ارزی در Kرا در نظر بگیرید:

اگر کلاسهای هم ارزی Q یک اجتماع از کلاسهای هم ارزیP باشند درآن صورت می گوییم Qوابسته به P می باشد یا اصطلاحا Qسخت تراز Pیا Pنرمتر از Qاست. به این ارتباط،وابستگی دانش (Knowledge Dependency)یا KDگویند.

بدیهی است که اشتراک دو رابطه هم ارزی یک رابطه هم ارزی دیگری است که شامل همه کلاسهای هم ارزی مشترک بین دو رابطه هم ارزی است به طور کلی اشتراک همه رابطه های هم ارزی بوسیله IND(Rcol)که به صورت یک رابطه هم ارزی دیگر است مشخص می شود. همچنین IND(Rcol)به عنوان یک رابطه علی_معلولی در RColشناخته می شود.

اگرPColوQColدو زیر مجموعه از RColباشندبنا به تعریف داریم:

1)QColوابسته به PColاست اگر IND(Qcol)سخت تر از IND(Pcol)باشد.

2) PColوQColهم ارز یکدیگر هستنداگر:

if Pcol then Qcol , if Qcol then Pcol

یا به شکل ساده تر PColوQColهم ارز یکدیگر هستنداگرIND(Pcol)=IND(Qcol)

3) PColوQColمستقل هستنداگرنه if Pcol then QColونه if Qcol then Pcol

درنتیجه ما می توانیم یک جدول اطلاعات را به عنوان یک پایه دانش پاولاک ((U,Rcol(T))در نظر بگیریم و مفهوم وابستگی دانش را برای جدول اطلاعات به کار ببریم.

قضیه: یک تناظر یک به یک جدول اطلاعات پایه های دانش پاولاک وجود دارد.

وابستگی تابعی:

هنگامیکه ارزشهای یک چندتایی روی مجموعه ای از ویژگیهای منحصر به فرد تعیین کننده ارزشهای چندتایی دیگر مجموعه از ویژگیها باشد گوییم یک وابستگی تابعی اتفاق افتاده است.

جدولها و قواعد تصمیم:

تئوری مجموعه راف یک روش مؤثر برای استخراج قواعد از جدولهای اطلاعات ویا صریح تر از جداول تصمیم می باشد.

الف:candidate key :

در پایگاههای اطلاعاتی وابسته یک مجموعه از ویژگیها مانند K(مجموعه مینیمال) یک candidate keyنامیده می شود اگر همه ویژگیها یک وابسته تابعی روی Kباشند.

Extentional Candidate keyیک روش خاص از تقلیل یا زدودن می باشد.

اگر S=(U,T=CUD)یک جدول تصمیم باشد به طوریکه:

C={A1,A2,…,Ai,…,An} D={B1,B2,…,Bi,…,Bn}

کهCویژگیهای موقعیت وDویژگیهای تصمیم می باشد.

می توان گفت که دو پایه دانش روی مجموعه Uبه صورت زیر وجود دارد:

Rcol(C) ={IND(A1),…,IND(A2),…,IND(An)}

Rcol(D)={IND(B1),…,IND(B2),…,IND(Bn)}

S یک جدول تصمیم سازگار است اگرداشته باشیم:

اگر Rcol(C)آنگاه Rcol(D)

B رامی توان یک تقلیل از Sدر نظر گرفت اگرBیک زیر مجموعه مینیمال از Cبا شد وداشته باشیم:

اگر Rcol(B)آنگاه Rcol(D)

بدیهی است که مجموعه Bلازم نیست به صورت منحصر به فرد باشد.

اگر ما مجموعه Dرا مساوی مجموعه Tانتخاب کنیم (D=T)در آن صورت یک تقلیل Extentional Candidate key خواهیم داشت.

ب: تئوری مجموعه راف یک روش مؤثر برای استخراج قواعد از جداول اطلاعاتی (IT) یا با صراحت بیشتر از جداول تصمیم(DT) می باشد. در هر جدول اطلاعاتی می توان چندین جدول تصمیم را مشاهده نمود .هر شئ (موضوع) در Uدر DTمی تواند به عنوان یک قاعده تصمیم گیری تفسیر شود.متخصصین مجموعه راف علاقه مند به ساده کردن قواعد تصمیم هستند.

مثال:سیستم اطلاعاتی(جدول اطلاعاتی)زیررادرنظربگیرید:

U	LOCATION	TEST	NEW	CASE	RESULT
ID-1	Houston	10	92	03	10
ID-2	San Jose	10	92	03	10
ID-3	Palto Alto	10	90	02	10
ID-4	Berkeley	11	91	04	50
ID-5	Newyork	11	91	04	50
ID-6	Atlanta	20	93	70	99
ID-7	Chicago	20	93	70	99
ID-8	Baltimore	20	93	70	99
ID-9	Seattle	20	93	70	99
ID-10	Chicago	51	95	70	94
ID-11	Chicago	51	95	70	95

A: یک جدول تصمیم(DT) انتخاب می کنیم وآن را Sمی نامیم در نتیجه خواهیم داشت: S=(U,T=CUD)

C={TEST,NEW,CASE} D={RESULT}

U:مجموعه جهانی D: مجموعه ویژگیهای تصمیم

C:مجموعه ویژگیهای موقعیت

B:فرمهای ID-10وID-11دو قاعده ناسازگارندو بقیه فرمها قاعده سازگار هستند.در جداول تصمیم قاعده هایی را در نظرمی گیریم که سازگار باشند.

C:با استفاده از رابطه هم ارزی IND(RESULT)می توان اشیاء(ID-1تاID-9)را در سه کلاس هم ارزی که کلاسهای تصمیم نامیده می شونددسته بندی نمود.

DECISION1={ID-1,ID-2,ID-3}=[10] RESULT,
DECISION2={ID-4,ID-5}=[50] RESULT,
DECISION3={ID-6,ID-7,ID-8,ID-9}=[99] RESULT

D : بااستفاده از رابطه هم ارزی ©INDمی توان اشیاء را در جهار کلاس هم ارزی که کلاسهای موقعیت نامیده می شوند تقسیم بندی نمود.

CASE1=[ID-1,ID-2} CASE2={ID-3} CASE3={ID-4,ID-5}
CASE4={ID-6,ID-7,ID-8,ID-9}

E: وابستگی دانش(KD):

بعضی از اشیاء توسط ویژگیهای موقعیت وتصمیم غیر قابل تشخیص می باشند.

CASE1زیر مجموعه ای از DECISION1است.

CASE2زیر مجموعه ای از DECISION1است.

CASE3زیر مجموعه ای از DECISION2است.

CASE4زیر مجموعه ای از DECISION3است.

عبارتهای فوق دلالت دارد بر اینکه رابطه هم ارزی IND(RESULT)وابسته به IND(C)است یا به طور معادل RESULT یکKDرویCمی باشد.

F : استنتاج قواعد:

با استفاده از روابط بالامی توان قوائد زیر را استنتاج کرد:

IF TEST=10,NEW=92,CASE=03 THEN RESULT=10
IF TEST=10,NEW=90,CASE=02 THEN RESULT=10
IF TEST=11,NEW=91,CASE=04 THEN RESULT=50
IF TEST=20,NEW=93,CASE=70 THEN RESULT=99

G:**تقلیل ها:**

واضح است که {TEST,NEW}و{CASE} دو تقلیل هستند.بنابراین موقعیتها روی ویژگی CASE می توانند از قواعد داده شده در بالا حذف شوند.

H :تقلیل کننده های ارزشی:

توجه کنید که قاعده ۱ ازرابطه (CASE1زیر مجموعه DECISION1است)مشتق شده است پس ما می توانیم توصیف ازCASE1را به وسیله قید TEST=10به تنهایی نشان دهیم که آن تقلیل کننده ارزشی نامیده می شود.

CASE1={U:u.TEST=10}

با استفاده از استدلال مشابه می توان چهار قاعده را به صورت زیر خلاصه نمود.

1.IF TEST=10 THEN RESULT=10
2.IF TEST=11 THEN RESULT=50
3.IF TEST=20 THEN RESULT=99

این تقلیل منحصر به فرد نمی باشد و می توان با انتخابهای متفاوت تقلیلهای متفاوت به دست آورد.

مزایای کاربرد تئوری راف:

تئوری راف کاربردهای زیادی در مهندسی،تحلیل داده های پزشکی، پردازش تصویر و.... دارد . برخی از مزایای کاربردی تئوری مجموعه های راف به صورت زیر می باشد[4]:

*یک الگوریتم موثر برای یافتن الگوهای پنهان در داده ها

*یافتن مجموعه های مینیمال داده ها (کاهش یا تقلیل داده ها)

*ارزیابی اهمیت داده ها

*تولید مجموعه های مینیمال از قواعد تصمیم گیری از داده ها

*سادگی فهم و تفسیر آسان نتایج الگوریتم

مثالهای کاربردی تئوری مجموعه راف:

**استخراج قواعد با استفاده از مجموعه راف هنگامیکه ارزشهای تابع تعلق به صورت فواصلی می باشند.

**کاربردمجموعه راف باکشف کننده ها برای انتخاب خصوصیت(طرح)

**روشی برای استخراج تغذیه های پایگاه داده ها با استفاده از مجموعه راف

**کاربرد تئوری راف درکنترل(کنترل راف)[6]

**کاربرد مجموعه راف برای داده کاوی درسیستم های اطلاعاتی بیمارستان

**تحلیل عملی روی مجموعه اطلاعات مراقبت ژنتیک با استفاده از مجموعه راف

**تولید اتوماتیک داستان با روشهایNLP و مجموعه های راف[8]

**کلاسه سازی نقطه بهره برداری سیستم قدرت[9]

**تشخیص خطا در سیستم های قدرت[10]

**استخراج معرفت از پست های توزیع[11]

**استخراج قواعد با تحلیل داده های پزشکی[12]

** کاهش(فشرده سازی)معرفت در جداول اطلاعاتی(اطلاعات خودروها)[13]

**تشخیص خطا درسیستم های حمل و نقل[14]

**کار بر روی هوش پیوندی با استفاده از مجموعه راف

**استخراج معرفت جاسازی شده در شبکه عصبی تعلیم یافته با استفاده از مجموعه راف

**کشف روابط بین ویژگیهاووابستگی قواعد با استفاده از مجموعه راف

**استدلال درباره معرفت با استفاده از مجموعه راف

**مدلسازی تحلیلی سیستم دما با استفاده از مجموعه راف

**مدلسازی مجموعه بیماران قلبی موروثی با استفاده از مجموعه راف

**توسعه چک لیست تشخیصی در یک اتاق اورژانس با استفاده از مجموعه راف

**کلاسه سازی کلمات دورافتاده(پرت)زبان رسمی کشور تایلند با استفاده از مجموعه راف

**بدست آوردن قواعد از اطلاعات ناقص و متناقض با استفاده از مجموعه راف

**روشی برای ارزیابی کیفیت در سیستم حمل و نقل با استفاده از مجموعه راف

**استفاده از مجموعه راف برای انطباق رهنمودهای بالینی برای سازمانهای مراقب سلامتی

نتیجه:

در این مقاله، مفاهیم اولیه و زمینه های تئوری و کاربردی مجموعه های راف در استخراج دانش(معرفت) و داده کاوی و کاربردهای عملی و واقعی آن در زمینه های مختلف مهندسی و علوم دیگر معرفی شد.همچنین به صورت مختصر تئوری مجموعه های راف در تحلیل قواعد اگر.....آنگاه.....همراه با ذکر مثال بیان شد که می توان مباحث تئوری بیشتر را درباره مجموعه های راف در منابع موجود یافت. با توجه به کارهای انجام شده درباره مجموعه های راف می توان نتیجه گرفت که این تئوری یک ابزار قدرتمند در بازیابی اطلاعات ،داده کاوی[15]،تقلیل(کاستن) داده های زائد از پایگاه داده ها می باشد.

مراجع:

[1] Honby, A.S.,"Oxford Advanced Learners Dictionary of Current English",Oxford University Press,UK,1974

[2] Ziarko, W.,"The Discovery ,Analysis and Representation of Data Dependencies in Databases",Knowledge Discovery in Databases,AAAI MIT Press ,Cammridge ,MA,1993,pp.213-228

[3] Pawlak,Z.,"Rough Sets",International Journal of Computer and Information Sciences,Vol.11,1982,pp.341-356

[4] Pawlak,Z.,"Rough Sets and Data Analysis"Proceeding of IEEE Conference ,ISSN:0-7803-3687-9,1996

[5] Pawlak,Z.,"Rough Sets : Theoretical Aspects of Reasoning about Data ",Kluwer Academic Publishers, Dordrecht ,boston ,London, 1991.

[6] Toshinori,M.,"Rough Control Application of Rough Set Theory to Control",Computer and Information Science Department Cleveland State University

[7] T.Y.Lin,"An Overview of Rough Set Theory from the Point of View of Relational Databases", Department of Electrical Engineering and Computer Science,University of California

[8] Beaubouef,T.,Lang,R.,"Rough Set Techniques for Uncertainty Management in Automated Story Generation"Prceeding of the 36th Annual Conference on Southeast Regional Conference,April,1998,pp.326-331

[9] Fereidunian,A.R,Lesani,H.,Lucas,C.,"Distribution Systems Reconfiguration Using Pattern Recognizer Neural Networks",International Journal of Engineering (IJE),Transactions B: Applications,Vol.15,No.2,pp.135-144,July 2002

[10] Zhang ,Q.,Han,Z.,Wen,F.,"A New Approach for Fault Diagnosis in Power Systems Based on Rough Set Theory ", Prceeding of the 4th international Conference on Power Systems Control ,Operation and Management,APSCOM-97,Hong Kong, November 1997,pp.597-602

[11] Hor,C.-L.,Crossley,P.,Dunand,F.,"Knowledge Extraction within Distribution Substation Using Rough Set Approach ",IEEE Power Engineering Society Winter Meeting 2002 Proceeding 0-7803-7322-7/02,January 27-31 ,2002 , New York,NY,USA

[12] Nakayama ,H.,Hattori,Y.,Ishii,R.,"Rule Extraction Based on Rough Set Theory and Its Application to Medical Data Analysis",IEEE Conference Proceeding 0-7803-5731-0/99,1999,pp.v924-v929

[13] Douqian,M.,Jue,W.,"Information-Based Algorithm for Reduction of Knowledge ", Proceeding of IEEE International Conference on Intelligent Processing Systems,Beijing,China,October 28-31,1997.

[14] Xiaolei,L.,Xiaobing,W.,"The Application of Rough Set Theory in Vehicle Transmission System Fault Diagnosis ",Proceeding ISSN:0-7803-5296-3/99,1999,pp.240-242

[15] Grzymala-Busse,J.,Ziarko,W.,"Data Mining and Rough Set Theory",Communications of ACM,April 2000,Vol.43,No.4,pp.108.