

# SID



سرویس های ویژه



سرویس ترجمه تخصصی



کارگاه های آموزشی



بلاگ مرکز اطلاعات علمی



سامانه ویراستاری STES



فیلم های آموزشی

## کارگاه های آموزشی مرکز اطلاعات علمی جهاد دانشگاهی

کارگاه آنلاین  
بررسی مقابله ای متون (مقدماتی)

کارگاه آنلاین  
پروپوزال نویسی و پایان نامه نویسی

کارگاه آنلاین آشنایی با پایگاه های اطلاعات علمی بین المللی و ترند های جستجو

# استخراج توابع عضویت در داده کاوی فازی با استفاده از سیستم کلونی مورچه بدون تعیین کمینهی تکیه گاه

قمرناز تدین تبریزی	مسعود نیازی ترشیز	فاطمه سعادت‌مند	احسان وجدانی محمودی
گروه کامپیوتر - دانشگاه آزاد اسلامی واحد مشهد	گروه کامپیوتر - دانشگاه آزاد اسلامی واحد مشهد	گروه کامپیوتر - دانشگاه آزاد اسلامی واحد مشهد	باشگاه پژوهشگران جوان - دانشگاه آزاد اسلامی واحد مشهد
gtadayon@gmail.com	niazi@mshdiau.ac.ir	fsaadatmand@gmail.com	e.vejdani@mshdiau.ac.ir

در حال حاضر به کاربردن سیستم های کلونی مورچه برای چندین مسئله با مرتبه ی NP مانند مسئله فروشنده دوره گرد (TSP)، مسئله زمان بندی کار (JSP)، مسئله های مسیریابی کامیون (VRP) و غیره موفقیت آمیز بوده است [۳]، [۴] و [۵]. مدل‌های ذکر شده، مدل‌های اکتشافی برگرفته شده از جامعه ی مورچه ها هستند. مورچه ها ردپای شیمیایی شان را که "فرمون" نام دارد جهت ارتباط با دیگران روی زمین به عنوان ردپا می گذارند. با استفاده از "فرمون" مورچه ها می توانند کوتاه ترین مسیر بین مبدأ و مقصد را بیابند. داده کاوی بر اساس سیستم کلونی مورچه موضوع جدیدی در حوزه داده کاوی محسوب می شود. کارهای قبلی روی اکتشاف قانون بر اساس ACS به وسیله Parpinelli [۶]، Cordon و همکارانش ارائه شده بود [۷]، که در این کارها کاوش قوانین دسته بندی برای سیستم های کنترل فازی ارائه شده است. تحقیقات کمتری در زمینه کشف قوانین وابستگی انجام شده است. Han، Wang، Lu و Tzvetov [۸] اشاره کرده اند که تعیین کمینه ی تکیه گاه کار ماهرانه و ظریفی است که می تواند کاربردهای گسترده ی این الگوریتم ها را تحت تاثیر قرار دهد. تجربیات جدید ما از کاوش دیتابیس های تراکنشی ثابت می کند که فرایند تعیین کردن کمینه ی تکیه گاه به هیچ عنوان کار ساده ای نیست.

در این مقاله یک الگوریتم بر اساس سیستم کلونی مورچه را برای استخراج توابع عضویت از داده های کمی جهت استفاده در کاوش فازی، بدون تعیین کردن کمینه ی تکیه گاه ارائه شده است. نکته ی منحصر به فرد روش ما عدم نیاز به تعیین آستانه ی کمینه ی تکیه گاه توسط کاربر است. این روش باعث (۱) کارایی جهت جستجو ی کلی و (۲) اتوماسیون سیستم می شود، چرا که مدل ما به آستانه ی کمینه ی تکیه گاه تعیین شده توسط کاربر احتیاج ندارد. آزمایشات عددی روی الگوریتم ارائه شده نیز جهت ارائه کارایی انجام شده اند. بخشهای بعدی این مقاله به صورت زیر سازماندهی شده است، چارچوب کاوش فازی براساس سیستم کلونی مورچه در بخش ۲ معرفی

چکیده: یکی از مهم ترین کاربردهای داده کاوی فازی تبدیل داده های کمی به عبارات زبانی است. بنابراین استخراج توابع عضویت تقشموثریدر فرایند کاوش فازی دارد. سیستم کلونی مورچه اگرچه کاربرد وسیع در مسائل بهینه سازی در سالهای اخیر داشته است، روش مناسبی برای دست آوردن توابع عضویت از داده های گسسته است. یکبار مشکلات تعدد هیاستفاده از داده کاوی فازی جهت استخراج توابع عضویت تعیین کمینه ی تکیه گاه (minimum support) است که می تواند کاربرد وسیع داده کاوی فازی را تحت تاثیر قرار دهد. در این مقاله الگوریتمی ارائه شده است که تابع عضویت هر آیتم را به وسیله ی سیستم کلونی مورچه و با استفاده از مکانیزمی اتوماتیک جهت تولید کمینه ی تکیه گاه هر آیتم با توجه به خصوصیات همان آیتم به دست می آورد و مناسب ترین توابع را جهت استفاده در کاوش فازی انتخاب می کند. پیاده سازی الگوریتم با استفاده از ۱۰۰۰۰ تراکنش فروشگاه، جهت ارائه ی کارایی الگوریتم ارائه شده، انجام گرفته است.

**واژه های کلیدی:** سیستم کلونی مورچه، داده کاوی، مجموعه فازی، تابع عضویت، قوانین وابستگی.

## ۱- مقدمه

اخیرا تئوری مجموعه ی فازی بارها و بارها در سیستم های هوشمند به دلیل سادگی و شباهتش به استنتاج بشری استفاده شده است [۱]. نقش مجموعه های فازی در داده کاوی به تبدیل مقادیر عددی به اصطلاحات زبانی و در نتیجه کاهش مجموعه آیتم های ممکن در فرایند کاوش کمک می کند. در زمینه داده کاوی فازی، Kuo و Hong یک مدل کاوش که نظریات مجموعه فازی را با الگوریتم کاوش Apriori تلفیق می کند ارائه کردند [۲]. در داده کاوی فازی توابع عضویت داده شده می توانند تاثیر به سزایی بر نتایج پایانی کاوش داشته باشند.

می شود، در بخش ۳ الگوریتم کاوش ارائه شده را توضیح می دهیم. در بخش ۴ نتایج شبیه سازی نشان داده شده است و در نهایت در بخش ۵ نتایج مورد بحث و بررسی قرار می گیرد.

## ۲- چارچوب کاوش فازی بر اساس سیستم کلونی مورچه

چارچوب ارائه شده، چارچوب کاوش براساس الگوریتم ACS در [۹] را اصلاح می کند. این چارچوب به دو قسمت تقسیم می شود. در فاز اول یک جستجوی کلی برای یافتن توابع عضویت مناسب برای آیتمها به وسیله الگوریتم کاوش ACS انجام می شود. سپس در فاز دوم بهترین مجموعه توابع عضویت بدست آمده در فاز اول را برای داده کاوی فازی به کار می بریم.

الگوریتم ACS نقش مهمی در استخراج توابع عضویت در فاز اول بازی می کند. در گذشته Parpinelli و همکارانش AntMiner را برای کشف قوانین وابستگی کشف کردند [۶]. آنها روی صفات دسته ای و مقادیر گسسته کار کردند و خوبی عمل کرد الگوریتم ACS در کنترل مقادیر گسسته در یک فضای راه حل را نشان دادند. در این مقاله، ما پارامترهای توابع عضویت را گسسته فرض کردیم و بنابراین سعی کردیم از الگوریتم ACS برای یافتن آنها استفاده کنیم. ما استخراج توابع عضویت را به یک مسئله ی یافتن مسیر تبدیل کردیم. یک مسیر نمایانگر یک مجموعه ممکن از توابع عضویت است. بنابراین مورچه های مصنوعی می توانند جهت یافتن نزدیک ترین راه حل بهینه بکار گرفته شوند.

## ۳- الگوریتم داده کاوی فازی بر اساس سیستم کلونی مورچه

### ۱-۳ تعاریف اولیه

به جای ارائه توابع عضویت همه ی آیتمها به صورت یک کد طولانی، تابع عضویت هر آیتم را با یک کد دودویی رمزگذاری می کنیم. ما از روش کدگذاری ارائه شده در [۱۰] استفاده کرده ایم. بعلاوه برخی از قوانین تعریف شده در [۹] مانند؛ قانون انتقال حالت، قانون به روز رسانی فرومون، قانون به روز رسانی محلی و قانون به روز رسانی سراسری را نیز به کار گرفته ایم.

در این مقاله، هر آیتم یک مجموعه ی توابع عضویت به شکل مثلث متساوی الساقین دارد. تابع عضویت به معنای اصطلاحات زبانی مانند Low, Middle, High است. تبدیل این مقادیر عددی به اصطلاحات زبانی احتیاج به یک جمعیت مناسب از پایگاه داده دارد. بنابراین ما نیاز به مقدار دهی اولیه و به روز رسانی یک جمعیت در حین پروسه ی تکامل داریم. در این مقاله، از تابع برازندگی (fitness) ارائه شده توسط Hong و همکارانش [۱۱] برای بدست آوردن مجموعه ای از توابع عضویت استفاده می کنیم.

## ۲-۲ الگوریتم کاوش ارائه شده

بر خلاف الگوریتم ارائه شده در [۹، ۱۱] که یک کمینه ی تکیه گاه ثابت برای تمام آیتم ها در نظر گرفته است، ما در این مقاله برای هر آیتم از یک کمینه ی تکیه گاه مخصوص به خود آن آیتم استفاده کرده ایم به این دلیل که در کاربردهای واقعی آیتم ها دارای فراوانی مختلفی هستند؛ بنابراین استفاده از کمینه ی تکیه گاه متفاوت برای آیتم ها جهت استخراج توابع عضویت به نظر می رسد ایده ای کارا و مناسب باشد. در این روش برخلاف روش قبلی نیازی به تعیین کمینه تکیه گاه توسط کاربر نیست، بلکه کمینه ی تکیه گاه توسط یک پیش پردازش بر روی تمام آیتم ها بدست می آید به طوریکه کمینه ی تکیه گاه هر آیتم متناسب با میانگین تعداد آن آیتم در پایگاه داده ی تراکنشی می باشد. علاوه بر پارامترهای تعریف شده در قسمت های قبل پارامترهای دیگری نیز اضافه خواهند شد که عبارتند از: تعداد مورچه های مصنوعی، مینیمم نسبت فرومون یک مورچه، نسبت تبخیر فرومون، نسبت به روز رسانی داخلی و نسبت به روز رسانی خارجی. الگوریتم بر اساس سیستم کلونی مورچه ارائه شده برای کاوش توابع عضویت و قوانین وابستگی فازی به صورت زیر است.

ورودی: (۱) N تا داده ی تراکنشی

(۲) مجموعه ای از m آیتم که ترمهای زبانی آنها از پیش

تعیین شده است

(۳) G ماکسیمم تعداد تکرارها

خروجی: یک مجموعه مناسب از توابع عضویت برای تمام آیتمها جهت کاربرد در داده کاوی فازی.

گام ۱: p را برابر یک قرار دهید  $p=1$ ، p شناسه ای برای تعداد آیتمهایی است که پردازش شده اند.

گام ۲: یک گراف چند حالتی برای مسئله داده کاوی فازی به صورت  $(N, E)$  که N مجموعه ای از نودها است و E هزینه یالها است در نظر می گیریم. همچنین نود [i] را در i امین مرحله به صورت  $N_{ij}$  مشخص می کنیم، یال  $N_{ij}$  تا  $N_{i(k+1)}$  را به صورت  $E_{ijk}$  و با مقدار ۰.۵ در نظر می گیریم.

گام ۳: جمعیت اولیه g را برابر با یک قرار دهید.

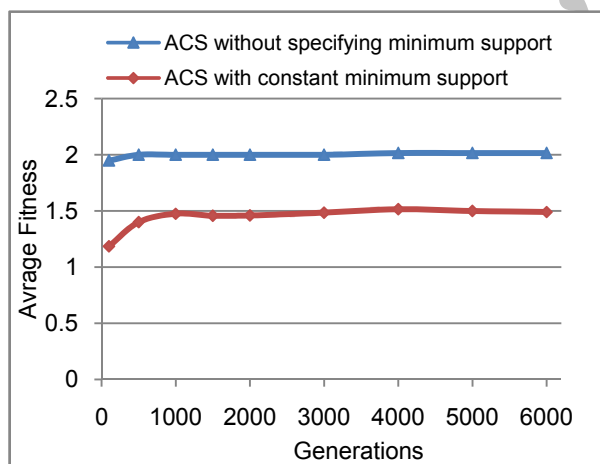
گام ۴: یک مسیر برای هر مورچه  $(Ant_q)$  به وسیله زیر گام های زیر کامل می شود.

گام ۴-۱: بر اساس قانون تغییر حالت، یالها را از شروع تا پایان انتخاب کنید.

#### ۴- آزمایشات عددی

آزمایشاتی جهت نشان دادن کارایی الگوریتم ارائه شده انجام شدند. آزمایشات با زبان C و C++ روی یک کامپیوتر AMD Athlon(tm) 64 با پردازنده اینتل دو هسته ای 1900GHz و رم 2GB پیاده سازی شده اند. مجموعاً ۶۴ آیتم و ۱۰۰۰۰ تراکنش در آزمایشات استفاده شده اند. اندازه اولیه مورچه ها ۱۰ می باشد. پارامترهای الگوریتم ACS به صورت زیر در نظر گرفته شدند: نسبت اولیه فرمون ۰.۰۵، کمترین فرمون مورچه ها ۰.۲، نسبت تبخیر ۰.۹، نسبت به روز رسانی محلی ۰.۱ و نسبت به روز رسانی سراسری ۰.۹. متوسط مقادیر برازندگی مورچه های مصنوعی برای نسلهای متفاوت در شکل ۱ نشان داده شده است.

همان طور که در شکل ۱ مشاهده می کنید مقدار متوسط برازندگی در الگوریتم ارائه شده به مراتب بیشتر از الگوریتم ACS با کمینه ی تکیه گاه ثابت می باشد، و همچنین در روش ما به تعداد تولید نسلهای کمتری برای رسیدن به حالت بهینه نیاز داریم. بعلاوه ما برای توضیح بهتر تفاوت میان الگوریتم ارائه شده و الگوریتم ACS با کمینه ی تکیه گاه ثابت از تعداد نسلهای کوچکتری استفاده کرده ایم که در شکل ۲ نشان داده شده است. واضح است که مدل ما با تعداد نسل ۱۰۰ به بهترین مقدار متوسط برازندگی می رسد، در حالی که در روش ACS با کمینه ی تکیه گاه ثابت این مقدار در تعداد نسل ۴۰۰ حاصل می شود.



شکل ۱: مقادیر متوسط برازندگی برای تعداد نسلهای متفاوت

گام ۴-۲: بر اساس قانون بهینه سازی محلی، فرمون یالهایی را که مورچه ها  $Ant_q$  از آنها عبور کرده اند را به روز رسانی کنید.

گام ۵: با استفاده از زیرگام های زیر مقدار برازندگی یک راه حل (توابع عضویت) که به وسیله هر مورچه بدست می آید را محاسبه کنید.

گام ۵-۱: برای هر داده تراکنش  $D_i$  که  $n$  تا  $i = 1$  است، مقدار کمی  $v_p$  آیتم  $I_p$  را بر اساس توابع عضویتی که از طی مسیر مورچه ها بدست آمد را به مقدار فازی  $f_p$  تبدیل کن،  $f_p$  به صورت زیر ارائه شده است:

$$\frac{f_{p1}}{\text{Region}_{p1}} + \frac{f_{p2}}{\text{Region}_{p2}} + \dots + \frac{f_{pi}}{\text{Region}_{pk}} + \dots + \frac{f_{pl}}{\text{Region}_{pl}} \quad (1)$$

که  $\text{Region}_{pk}$ ، کمینه ناحیه فازی برای آیتم  $I_p$  است،  $f_{pk}$  مقدار درجه عضویت فازی  $v_p$  در یک ناحیه است، و  $l$  تعداد توابع عضویت فازی است.

گام ۵-۲: مقدار عدد کاردینالیتی برای هر ناحیه در تراکنش به صورت زیر محاسبه می شود.

$$\text{count}_{pk} = \sum_{i=1}^n f_{pk}^{(i)} \quad (2)$$

که  $f_{pk}^{(i)}$  درجه عضویت فازی ناحیه  $R_{pk}$  برای داده ی  $i$  ام است.

گام ۵-۳:  $\frac{\text{count}_{pk}}{n}$  برای هر ناحیه  $R_{pk}$  چک شود که آیا بزرگتر یا مساوی کمینه ی تکیه گاه است یا نه، اگر بود آن را در مجموعه  $\text{large } 1\text{-itemset}$  (L1) قرار دهید.

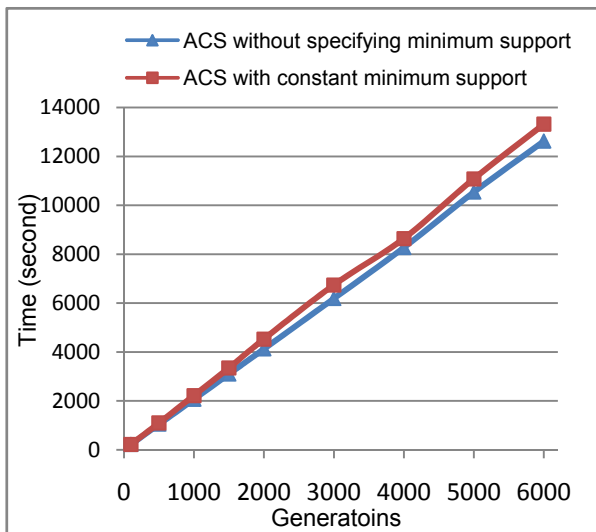
گام ۵-۴: مقدار برازندگی یک راه حل با استفاده از تقسیم  $\text{large } 1\text{-itemset}$  (L1) بر روی شایستگی بدست می آید که در فرمول (۶) معرفی شده است.

گام ۶: یک بار که همه ی مورچه های مصنوعی مسیرها را پیدا کردند. بالاترین درجه برازندگی را برای تغییر در مقدار فرمون بر طبق قانون به روز رسانی سراسری کلی نگه دارید.

گام ۷: اگر  $g$  تولید شده برابر  $G$  باشد، خروجی جاری بهترین تابع عضویت برای آیتم  $I_p$  است، وگرنه  $g = g + 1$  و به گام ۴ برو.

گام ۸: اگر  $p \neq m$  است آنگاه  $p = p + 1$  قرار دهید و به گام ۲ بروید و برای دیگر آیتم ها الگوریتم را تکرار کنید، وگرنه الگوریتم را متوقف کنید.

نهایتاً مجموعه توابع عضویت خروجی گام ۷ و  $1\text{-itemset}$  های بدست آمده را برای کاوش قوانین وابستگی فازی در بانک اطلاعاتی داده شده استفاده می کنیم.



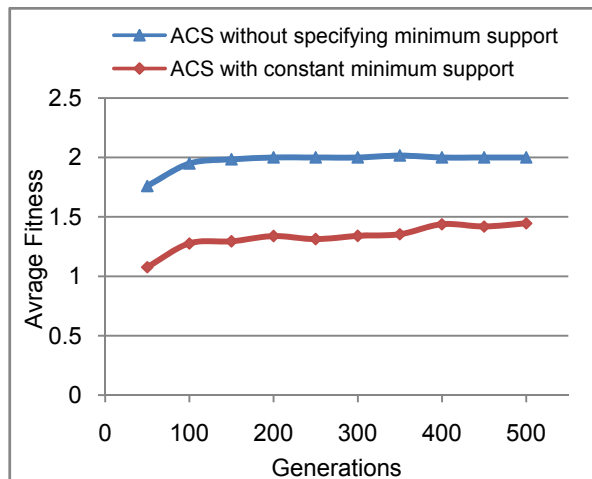
شکل ۴: زمان اجرای الگوریتم کاوش ACS برای تعداد نسلهای متفاوت

#### ۵- نتیجه گیری و کار آینده

در این مقاله به موضوع به کاربردن الگوریتم ACS جهت استخراج توابع عضویت برای داده کاوی فازی پرداختیم و الگوریتمی جهت رسیدن به هدفمان ارائه کردیم. ما با استفاده از نتایج شبیه سازی نشان دادیم که متوسط برازندگی توابع عضویت حاصل از روش ما بیشتر از روش قبلی است. از طرفی زمان اجرا را نیز نسبت به روش کمینه تکیه گاه ثابت اندکی بهبود بخشیدیم. بنابراین استفاده از کمینه ی تکیه گاه متفاوت برای آیتم ها جهت استخراج توابع عضویت ایده ای کارا و مناسب می باشد. الگوریتم ACS بدون مشخص کردن کمینه ی تکیه گاه دو مزیت دارد: (۱) استفاده از کمینه ی تکیه گاه مختص هر آیتم؛ (۲) مکانیزه کردن سیستم. نتایج نشان می دهد که می توان این روش را برای کاوش قوانین وابستگی به نحو موثر به کار برد. با این حال این روش نیازمند کارهای بیشتری در آینده می باشد. به عنوان مثال، استفاده از تکنیک های موازی سازی جهت کاهش زمان اجرا و افزایش کارایی و به کارگیری قوانین فازی در محاسبه ی فاکتورهای پوشش و هم پوشانی جهت واقعی تر شدن مقدار برازندگی تابع. نتایج آزمایشگاهی انگیزه ما را جهت استفاده از این استراتژی در کاربردهای دنیای واقعی بیشتر می کند.

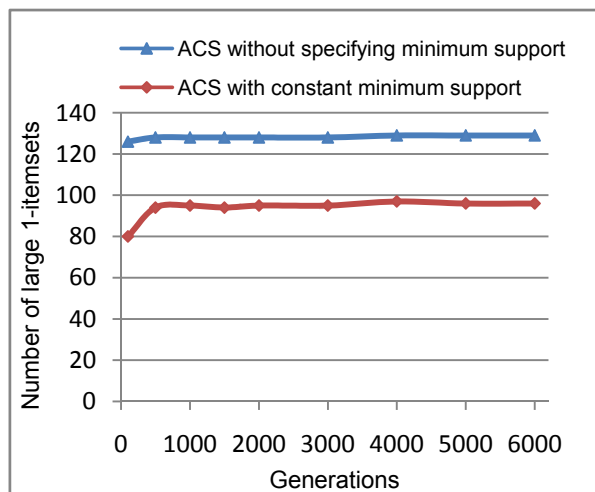
#### مراجع

- [1] Kandel, "Fuzzy expert systems", CRC Press, Boca Raton, pp 8-19, 1992 .
- [2] T .P .Hong, C .S .Kuo, S .C .Chi, "Trade-off between time complexity and number of rules for fuzzy mining from quantitative data", *International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems*, Vol .9, No .5, pp 587-604, 2001 .
- [3] Colorni, M .Dorigo, V .Maniezzo and M .Trubian, "Ant system for job-shop scheduling", *Belgian Journal of*



شکل ۲: مقادیر متوسط برازندگی برای تعداد نسلهای متفاوت در مقیاس کوچکتر

تعداد large 1-itemset ها برای تعداد نسلهای متفاوت در شکل ۳ نشان داده شده است. منحنی الگوریتم ACS با کمینه ی تکیه گاه ثابت پس از حدود ۴۰۰۰ نسل به مقدار ثابتی می رسد، در حالی که در روش ارائه شده منحنی در نسل ۱۰۰۰ به مقدار ثابتی میرسد. بعلاوه، تعداد large 1-itemset ها در روش ارائه ما بیشتر است. زمان اجرای الگوریتم کاوش ACS برای تعداد نسلهای متفاوت در شکل ۴ نشان داده شده است. اگرچه زمان اجرای هر دو الگوریتم با افزایش تعداد نسل افزایش می یابد اما روش ما زمان اجرای کمتری را نسبت به الگوریتم ACS با کمینه ی تکیه گاه ثابت نشان می دهد.



شکل ۳: تعداد large 1-itemset ها برای تعداد نسلهای متفاوت

- Operations Research, Statistics and Computer Science*, Vol . 34, pp .39-53, 1994.
- [4] M .Dorigo, L .M .Gambardella, “Ant colony system :a cooperative learning approach to the traveling salesman problem”, *IEEE Transactions on Evolutionary Computation*, Vol .1, No .1, pp .53-66, 1997.
- [5] A .Wade and S .Shalhi, “An ant system algorithm for the mixed vehicle routing problem with backhauls”, in *Metaheuristics :Computer Decision-Making, Norwell, MA : Kluwer*, pp 699-719, 2004.
- [6] R .S .Parpinelli, H .S .Lopes, A .A .Freitas, “An Ant Colony Based System for Data Mining: Application to Medical Data”, *The Genetic and Evolutionary Computation Conference*, pp.791-798, 2001 .
- [7] J .C .Cordon, F .Herrera, “Learning Fuzzy Rules Using Ant Colony Optimization”, *The Second International Workshop on Ant Algorithms*, pp.13-21, 2002.
- [8] Han, J., Wang, J., Lu, Y., &Tzvetkov, “Mining top-k frequent closed patterns without minimum support”. In *Proceedings of the 2002 IEEE international conference on data mining*, pp 211-218, ICDM 2002.
- [9] T .P .Hong, Y.F .Tung, S.L .Wang, M.T .Wu, Y.L .Wu, “Extracting membership functions in fuzzy data mining by Ant Colony Systems”, *Proceedings of the Seventh International Conference on Machine Learning and Cybernetics, Kunming*, pp 12-15, Jul .2008.
- [10] T .P .Hong, Y.F .Tung, S.L .Wang, M.T .Wu, Y.L .Wu, “An ACS-based Framework for Fuzzy Data Mining”, *Expert Systems with Applications* vol.36, Issue 9, pp 11844-11852, Nov. 2009.
- [11] T .P .Hong, C .H .Chen, and V .S.Tseng, “Cluster-Based Evaluation in Fuzzy-Genetic DataMining” *IEEE Trans . Fuzzy Syst.*, vol .16,no .1, pp .249–262, 2008.

Archive

# SID



سرویس های ویژه



سرویس ترجمه تخصصی



کارگاه های آموزشی



بلاگ مرکز اطلاعات علمی



سامانه ویراستاری STES



فیلم های آموزشی

## کارگاه های آموزشی مرکز اطلاعات علمی جهاد دانشگاهی

توجه: بررسی مقاله ای متون (مقدماتی)

کارگاه آنلاین  
بررسی مقابله ای متون (مقدماتی)

PROPOSAL  
پروپوزال

توجه: پروپوزال نویسی و پایان نامه نویسی

کارگاه آنلاین  
پروپوزال نویسی و پایان نامه نویسی

ISI  
Scopus

توجه: آشنایی با پایگاه های اطلاعات علمی بین المللی و ترند های جستجو

کارگاه آنلاین آشنایی با پایگاه های اطلاعات علمی بین المللی و ترند های جستجو