

# ارایه یک الگوریتم ترکیبی برای خوشه بندی با استفاده از رویکرد فرا ابتکاری جستجوی ممنوعه

مسعود یقینی<sup>۱</sup>؛ ناهید غضنفری<sup>۲</sup>

## چکیده

الگوریتم K-Means یکی از روشهای رایج خوشه بندی می باشد که علیرغم مزایای بسیار از جمله سرعت بالا و سهولت پیاده سازی، در دام بهینه محلی قرار گرفته و همیشه جواب بهینه مساله را تولید نمی نماید. از سوی دیگر الگوریتمهای فراابتکاری مانند خوشه بندی کلونی مورچه ها و یا الگوریتم ژنتیک نیز که با منطق تصادفی و با تکرار اجرای الگوریتم، قادر به خوشه بندی داده ها با شکلهای پیچیده می باشد، لزوماً تولید جواب بهینه خوشه بندی را تضمین نمی نماید.

در این مقاله جهت الگوریتمهایی که مراکز اولیه خوشه ها در آن به صورت تصادفی انتخاب می گردند، الگوریتم ترکیبی جدیدی با استفاده از روش فراابتکاری جستجوی ممنوعه طراحی و پیاده سازی شده است که می تواند سبب خروج از مینیمم محلی شده و با درصد بالایی جواب بهینه مساله را تولید نماید. الگوریتم جدید "Ant+Tabu" و "K-Means+Tabu" با داده های مختلف تست و نتایج آن با سایر الگوریتمهای فراابتکاری نیز مقایسه شده است. بررسی نتایج نشان می دهد که الگوریتم جدید قادر است با دقت بالایی جواب بهینه را در مقایسه با نتایج حاصل از الگوریتمهای K-Means و Ant Clustering تولید نماید.

**کلمات کلیدی:** داده کاوی، خوشه بندی، جستجوی ممنوعه، بهینه سازی کلونی مورچه ها، الگوریتم K-Means، مینیمم محلی

---

<sup>۱</sup> استادیار و عضو هیات علمی دانشگاه علم و صنعت ایران

<sup>۲</sup> دانشجوی کارشناسی ارشد مدیریت اجرایی، دانشگاه علم و صنعت ایران

# **A Hybrid Clustering Algorithm Based on Tabu Search Approach**

Masoud Yaghini; Nahid Ghazanfari

## **Abstract**

The K-Means algorithm is one of the most widely used techniques for clustering. However, K-Means has two shortcomings: dependency on the initial state and convergences to local optima and global solutions of large problems can not found with reasonable amount of computation effort. On the other hand, evolutionary metaheuristic algorithms including ant colony optimization and genetic algorithms does not necessarily ensure that it converges to a solution which is global optimum. In order to overcome local optima problem lots of studies done in clustering.

This paper proposed an efficient hybrid algorithm based on combining K-Means and Tabu search, called K-means+Tabu, Ant clustering and Tabu search called “Ant+Tabu” for optimally clustering N object into K clusters. The new algorithm is tested on several data sets and its performance is compared with those of K-Means, Tabu search, ACO and other metaheuristic algorithms. The simulation results show that the proposed evolutionary optimization algorithm is robust and suitable for handling data clustering.

**Keywords:** Data Mining, Clustering, Tabu Search, Ant Clustering, K-Means, Local Optima

## ۱- مقدمه

خوشه بندی فرایند دسته بندی مجموعه ای از اشیاء به خوشه هایی می باشد که اعضا درونی هر خوشه بیشترین شباهت را به یکدیگر و کمترین شباهت را نسبت به اعضا سایر خوشه ها داشته باشند. از خوشه بندی در علوم مختلفی مانند مهندسی، پزشکی، علوم اجتماعی و بازاریابی استفاده می شود. در واقع خوشه بندی از جمله تکنیکهای توصیفی داده کاوی می باشد که بدون هدف از قبل مشخص شده، الگوهایی را از داده ها استخراج می نماید.

جهت تجزیه و تحلیل خوشه ها، اشیاء با نقاطی در فضای  $n$  بعدی نمایش داده می شوند که بردارها نشان دهنده خصیصه اشیا بوده و هدف مساله نیز دسته بندی  $N$  شی به  $K$  خوشه می باشد که اشیا هر خوشه بیشترین شباهت را به یکدیگر داشته باشند. در این مقاله از معیار فاصله اقلیدسی استفاده شده است که هدف حداقل نمودن انحراف نقاط درونی خوشه ها از مراکز خوشه می باشد.

جهت خوشه بندی داده ها، تکنیکها و الگوریتمهای مختلفی به کار گرفته می شود. بیشتر مسایل خوشه بندی غیرخطی و محدب می باشند که ممکن است جوابهای مینیمم محلی داشته باشند و در نتیجه جواب بهینه مساله را تولید ننمایند. در توسعه الگوریتمهای خوشه بندی از الگوریتمهای فراابتکاری بسیاری مانند GA، TS، ACO و SA استفاده شده است که همگی جواب بهتری را تولید می نمایند که لزوماً جواب بهینه مساله نمی باشد. در واقع در الگوریتمهایی که جواب اولیه الگوریتم به صورت تصادفی انتخاب می شود، احتمال تولید جواب بهینه محلی وجود دارد که مانع از رسیدن به جواب بهینه مساله می گردد. در این مقاله با ترکیب الگوریتم فراابتکاری جستجوی ممنوعه با الگوریتمهای K-Means و بهینه سازی کلونی مورچه ها، الگوریتم موثر جدیدی طراحی و ارایه شده است که سبب خروج از دام بهینه محلی گردیده و جواب بهینه مساله را با درصد بالایی تولید می نماید.

این مقاله به صورت ذیل سازماندهی شده است: بخش ۲ توضیحی از مسایل خوشه بندی و برخی از الگوریتمهای آن نشان داده شده است. بخش ۳ مراحل ساخت مدل جدید شامل جزییات الگوریتم طراحی شده و نمودار چرخشی آن را بیان می کند. در بخش ۴ نتایج محاسبات انجام شده با داده های تصادفی و استاندارد، جهت تست و ارزیابی الگوریتم جدید ارایه شده است. در پایان نتایج پژوهش نیز در بخش ۵ آورده شده است.

## ۲- خوشه بندی

مسایل خوشه بندی جهت گروه بندی اشیاء به دسته هایی به کار می رود به نحویکه مجموع مربعات فاصله اقلیدسی بین اشیاء و مراکز خوشه ای که به آن تعلق دارند، حداقل گردد. پارامترها، تابع هدف و محدودیتهای مسایل خوشه بندی در رابطه شماره ۱ نمایش داده شده است.

$$\text{CP:} \quad \text{Min } F(w, m) = \sum_{j=1}^K \sum_{i=1}^N \sum_{v=1}^n w_{ij} \|x_{iv} - m_{jv}\|^2 \quad (1)$$

$$\text{Subject to :} \quad m_{jv} = \frac{\sum_{i=1}^N w_{ij} x_{iv}}{\sum_{i=1}^N w_{ij}} \quad j = 1, \dots, K, v = 1, \dots, n$$

$$\sum_{j=1}^K w_{ij} = 1, \quad i = 1, \dots, N$$

$$\sum_{i=1}^N w_{ij} \geq 1, \quad j = 1, \dots, K$$

$$w_{ij} = \begin{cases} 1 & \text{if data } i \text{ is contained in Cluster } j \\ 0 & \text{Otherwise} \end{cases}$$

که در آن

ماتریس داده  $X=\{x_1, \dots, x_n\}$

تعداد اشیاء:  $N$

تعداد خوشه ها:  $K$

تعداد خصیصه ها:  $n$

نمایشگر اختصاص شی به هر خوشه:  $W_{ij}$

فاصله اقلیدسی بین هر شی و مرکز خوشه:  $\|x_{iv} - m_{jv}\|^2$

تابع هدف در اکثر مسایل خوشه بندی محدب و غیرخطی می باشد که ممکن است در دام بهینه محلی قرار گرفته و لزوماً جواب بهینه مساله را تولید ننماید.

## ۲-۱ خوشه بندی مرکز - محور

الگوریتم K-Means [۱] یکی از الگوریتمهای محبوب خوشه بندی مرکز- محور در تکنیک خوشه بندی می باشد. K-Means با مقداری اولیه مراکز خوشه آغاز می شود. سایر اشیاء با توجه به معیار فاصله اقلیدسی به خوشه ای اختصاص داده می شوند که کمترین فاصله را از مرکز خوشه داشته باشند. در هر تکرار الگوریتم، دو مرحله اصلی اجرا می گردد. ابتدا هر شی در مجموعه داده، به خوشه ای با کمترین فاصله از مرکز آن اختصاص داده می شود. در ادامه پس از گروه بندی نقاط به  $k$  خوشه، مراکز خوشه های جدید محاسبه و الگوریتم تکرار می گردد. الگوریتم زمانی خاتمه می یابد که در محاسبه مراکز خوشه تغییری ایجاد نگردد.

K-Means [۲،۳] دو اشکال عمده دارد. اول اینکه الگوریتم به مقادیر اولیه تعداد خوشه و مرکز خوشه ها وابسته بوده و دوم اینکه به آسانی در دام بهینه محلی قرار می گیرد و در نتیجه جواب بهینه مساله را تولید نمی نماید. تاکنون تحقیقات بسیاری جهت رفع اشکالات الگوریتم K-Means با انتخاب صحیح مراکز اولیه خوشه جهت اجتناب از تولید جواب بهینه محلی انجام یا سعی در یافتن جواب نهایی بهینه خوشه بندی شده است.

## ۲-۲ الگوریتمهای فراابتکاری

الگوریتمهای فراابتکاری [۴] الگوریتمهایی هستند که با الهام از طبیعت، فیزیک و انسان طراحی شده اند و در حل بسیاری از مسایل بهینه سازی از جمله خوشه بندی استفاده می شوند. معمولاً از الگوریتمهای فراابتکاری در ترکیب با سایر الگوریتمها، جهت رسیدن به جواب بهینه یا خروج از وضعیت جواب بهینه محلی استفاده می گردد. دو گروه الگوریتمهای فراابتکاری خوشه بندی [۳] عبارتند از:

**الگوریتمهای جواب محور:** الگوریتمهای جستجوی ممنوعه و شبیه سازی سرد شدن تدریجی از جمله الگوریتمهای فراابتکاری جواب محور می باشند که از یک جواب اولیه خوشه بندی شروع نموده و در تکرار و با جابجایی اشیاء بهبود خوشه ها در هر خوشه بندی، به جواب قابل قبول می رسند. در شکل ساده آن احتمال جابجایی اشیاء خوشه ها به صورت تصادفی و در نوع تکاملی آن با منطقی جهت بهبود تابع هدف صورت می پذیرد.

**الگوریتمهای جمعیت محور:** در این گروه از الگوریتمها، در مجموعه ای از جوابهای خوشه بندی اولیه، جستجوی محلی انجام و در یک فرایند تکرارشونده و بهبود جوابها، به جوابی نزدیک به جواب بهینه می رسند. با جستجوی جوابهای همسایگی جواب فعلی، در صورت یافتن جوابی بهتر از جواب فعلی، عمل جایگزینی جواب انجام و جستجو ادامه پیدا می کند. الگوریتم خوشه بندی کلونی مورچه ها و الگوریتم ژنتیک از جمله الگوریتمهای جواب محور می باشند.

## ۲-۲-۱- الگوریتم جستجوی ممنوعه

جستجوی ممنوعه [۵] یکی از روشهای فراابتکاری می باشد که با استفاده از فرایندهای جستجوی محلی، قادر به حل مسایل بهینه سازی می باشد. جستجوی محلی یک روال جستجو تکرارشونده است که از یک جواب مساله شروع نموده و سپس با انجام حرکاتی در همسایگی جواب فعلی، جوابهای بهتری را در مساله پیدا می نماید. نکته کلیدی الگوریتم جستجوی ممنوعه، پذیرش جوابهای همسایگی بدون بهبود در جواب مساله می باشد که سبب خروج از بهینه محلی می گردد. جهت پیشگیری از ایجاد حلقه تکرار، از حافظه کوتاه مدتی به نام لیست ممنوعه استفاده می شود که با نگهداری جوابهای بررسی شده در همسایگی، مانع از بررسی مجدد آنها می گردد.

## ۲-۲-۲- الگوریتم خوشه بندی کلونی مورچه ها

الگوریتم خوشه بندی کلونی مورچه ها از جمله الگوریتمهای فراابتکاری جمعیت محور می باشد که در سالهای اخیر جهت حل مسایل بهینه سازی از جمله خوشه بندی استفاده می شوند. این الگوریتم قادر است جواب بهینه را با سرعت بالاتری در خوشه هایی با شکلهای پیچیده نسبت به سایر الگوریتمهای فراابتکاری تولید نماید. روش خوشه بندی کلونی مورچه ها با از رفتار مورچه ها در طبیعت طراحی شده است. مورچه ها حشراتی هستند که به صورت گروهی زندگی می کنند. آنها کور هستند و جهت یافتن کوتاهترین مسیر رسیدن به غذا و بازگشت به آشیانه از رد فرومون استفاده می کنند. فرومون ماده شیمیایی است که مورچه ها از خود ترشح می نمایند که اثر آن در مسیری که عبور می نمایند باقی می ماند. وقتی تعداد زیادی مورچه از یک مسیر عبور نماید، میزان فرومون مسیر افزایش می یابد. سایر مورچه ها نیز با دنبال نمودن رد فرومون، کوتاهترین مسیر را پیدا می کنند.

ساختار کلی الگوریتم خوشه بندی کلونی مورچه ها به شرح زیر می باشد:

- تمام اشیا به صورت تصادفی بر روی شبکه پراکنده می شوند.
- هر مورچه تصادفی یکی از اشیا را برمی دارد.
- هر مورچه به صورت تصادفی در موقعیتی روی شبکه قرار می گیرد.
- پس از مراحل فوق الذکر مرحله جستجوی محلی آغاز می گردد:
- یک مورچه تصادفی انتخاب می شود.
- مورچه در یک جهت بر روی شبکه شروع به حرکت می کند.
- مورچه به صورت احتمالی شی موردنظر را در یکی از سلولهای خالی قرار می دهد.

احتمال برداشتن یا انداختن اشیاء توسط مورچه ها بر مبنای محاسبه چگالی محلی اشیاء مشابه روی شبکه اندازه گیری می شود. الگوریتم خوشه بندی کلونی مورچه ها در هر تکرار، شامل سه مرحله زیر است:

- ۱) ساخت خوشه های جدید با استفاده از ماتریس تعدیل شده فرومون توسط مورچه ها در تکرارهای قبلی
- مطابق با فرمول ۲ هر شیء  $i$  ام با احتمال  $P_{ij}$  به خوشه  $j$  ام اختصاص داده می شود.  $\tau_{ij}$  فرومون مسیر و  $\eta_{ij}$  طول مسیر پیموده شده را نمایش می دهد.

$$P_{ij} = \frac{\tau_{ij}\eta_{ij}}{\sum_{k=1}^K \tau_{ik}\eta_{ik}}, \quad j = 1, \dots, K \quad (2)$$

- ۲) انجام جستجوی محلی بر روی خوشه های ایجاد شده

در این مرحله از جوابهای تولید شده، تعدادی را انتخاب و با اجرای فرایند جستجوی محلی، جواب بدست آمده بهبود می یابد.

- ۳) بروزرسانی ماتریس فرومون

در پایان هر تکرار ماتریس فرومون بر اساس جوابهای بهتر بدست آمده بروزرسانی می گردد. همچنین جهت جلوگیری از همگرایی زودرس و قرار گرفتن در دام بهینه محلی از تبخیر فرومون با ضریب  $(1 - \rho)$  استفاده می شود که روش محاسبه آن در فرمولهای ۳ و ۴ به شرح ذیل آمده است:

$$\tau_{ij}(t+1) = (1 - \rho)\tau_{ij} + \sum_{l=1}^L \Delta\tau_{ij}^l \quad (3)$$

$$\tau_{ij}^l \Delta = \begin{cases} \frac{1}{F_1} & \text{اگر شی } i \text{ توسط مورچه } l \text{ به خوشه } j \text{ اختصاص داده شود} \\ 0 & \text{در غیر این صورت} \end{cases} \quad i = 1, \dots, N, \quad j = 1, \dots, K$$

نکته قابل توجه در این الگوریتم این است که با توجه انتخاب تصادفی اشیاء توسط مورچه ها و تعداد تکرار، ممکن است مساله در دام بهینه محلی قرار گیرد و جواب بهینه را تولید ننماید.

### ۳- الگوریتم جدید خوشه بندی مبتنی بر رویکرد جستجوی ممنوعه

در سالهای اخیر جهت رفع اشکال الگوریتمهای خوشه بندی در تولید جوابهای مینیمم محلی، از الگوریتمهای فراابتکاری استفاده شده و در بسیاری از موارد با ترکیب الگوریتمهای فراابتکاری با سایر الگوریتمهای خوشه بندی مطابق با جدول شماره ۱، الگوریتم جدیدی طراحی شده است که کارایی و اثربخشی الگوریتمهای خوشه بندی را بهبود داده اند. بهبود الگوریتمهای خوشه بندی به روشهای ذیل انجام می گردد:

- ۱- تعیین روش انتخاب پارامترهای اولیه
- ۲- تغییر در الگوریتم پایه
- ۳- ترکیب الگوریتم خوشه بندی با سایر الگوریتمهای فراابتکاری

جدول ۱- فهرستی از پژوهشهای انجام شده خوشه بندی با استفاده از الگوریتمهای فراابتکاری

نویسنده	مرجع	سال	K-Means	GA	AC	SA	TS	نوع الگوریتم
K.S. Al-Sultan	۵	۱۹۹۵					*	الگوریتم پایه
P.S Shelokar, V.K. Jayaraman and B.D. Kulkarni	۹	۲۰۰۴			*			الگوریتم پایه
L. CHEN'X.XU, Y. CHEN	۱۰	۲۰۰۴			*			الگوریتم پایه
A.L. Vizine, L.N. de Castro, E.R. Hruschka1,	۱۵	۲۰۰۵			*			الگوریتم پایه
Chen. Wang and K.Y. Liu, L	۱۶	۲۰۰۵		*			*	الگوریتم ترکیبی
Chen. Wang and k. Liu, L. Liu, Y.Y	۱۳	۲۰۰۵	*				*	الگوریتم ترکیبی
Hui and Y. Sim. C. He, S.Y	۱۷	۲۰۰۶			*		*	الگوریتم ترکیبی
S.M. Pan and K. Sheng	۱۴	۲۰۰۷	*	*			*	الگوریتم ترکیبی
Chuang and K. Liu. Lee, S. Su, C.Z	۱۸	۲۰۰۸		*	*			الگوریتم ترکیبی
Z. Gungor and A. Unler	۱۱	۲۰۰۸	*				*	الگوریتم ترکیبی
T. Niknam, J. Olamaei and B. Amiri	۱۹	۲۰۰۸			*	*		الگوریتم ترکیبی

در این مقاله نیز از قابلیت‌های الگوریتم جستجوی ممنوعه استفاده و الگوریتم ترکیبی "K-Means+Tabu" و "Ant + Tabu" طراحی و پیاده سازی شده است که سبب خروج از دام بهینه محلی و تولید جواب بهینه مساله با دقت بالا می گردد.

### ۳-۱- ساختار الگوریتم جدید

برای پیاده سازی مدل به پنج عامل اساسی ذیل نیازمندیم :

#### ۱- نقطه شروع

مدل پیشنهادی با جواب بدست آمده از الگوریتم K-Means یا خوشه بندی کلونی مورچه ها آغاز می گردد. ابتدا مراکز نهایی خوشه ها تعیین می گردد. سپس به ترتیب خوشه ها را انتخاب نموده، منطق الگوریتم مدل جدید را بر روی خوشه منتخب اجرا می نمایم. مرکز خوشه منتخب، به عنوان نقطه آغاز الگوریتم بوده و سایر خوشه ها با همان مراکز خوشه قبلی بررسی می شوند.

#### ۲- مولد حرکت

این مولد وظیفه دارد مرکز خوشه جدید را از اشیاء در همسایگی مرکز خوشه فعلی، با این شرط که شی موردنظر در لیست ممنوعه نباشد، انتخاب کند و اشیاء را به خوشه هایی اختصاص دهد که کمترین فاصله را از مراکز خوشه ها داشته باشند. همواره برای انتخاب جهت حرکت در خوشه ها، سه استراتژی وجود دارد:

انتخاب نزدیکترین شی به مرکز خوشه اولیه منتخب: در این روش مرکز خوشه منتخب برای خوشه بندی، نزدیکترین شی به مرکز خوشه اولیه ورودی مدل می باشد. در واقع فضای جواب به صورت کروی شکل حول مرکز خوشه ورودی مدل بررسی می گردد.

انتخاب نزدیکترین شی به مرکز خوشه جدید تولید شده: در این روش مرکز خوشه منتخب برای خوشه بندی، نزدیکترین شی به مرکز خوشه جدید بدست آمده می باشد. در این حالت فضای جواب به صورت خطی در جهات مختلف بررسی می گردد.

نزدیکترین شی به مرکز بهترین جواب بدست آمده تا کنون: با توجه به اینکه پس از انجام هر خوشه بندی، کیفیت آن متناسب با تابع هدف بررسی می گردد، در این روش مرکز خوشه بعدی، حول مرکز خوشه بهترین جواب بدست آمده خواهد بود. در این حالت امکان جستجو در فضای جواب در دو حالت خطی و کروی امکانپذیر می باشد.

#### ۳- ممنوع نمودن انتخاب مراکز خوشه

نکته کلیدی الگوریتم جستجوی ممنوعه، استفاده از حافظه در پیاده سازی لیست ممنوعه و نگهداری مراکز خوشه انتخاب شده در مراحل قبلی می باشد. این شیوه موجب پرش از دام بهینه محلی و نزدیک شدن به جواب بهینه می گردد. در مدل پیشنهادی نیز پس از هر خوشه بندی، بررسی می گردد آیا نزدیکترین شی به مرکز خوشه جدید در لیست ممنوعه وارد شده است. نتیجه بررسی این شرط، جهت حرکت بعدی و انتخاب مرکز خوشه جدید را مشخص می سازد. در صورتیکه نزدیکترین شی به مرکز خوشه در لیست ممنوعه باشد، به این معناست که ما در وضعیت بهینه محلی قرار داریم و می بایست با ممنوع کردن انتخاب مجدد این شی به عنوان مرکز خوشه، جهت حرکت را تغییر دهیم. در صورتیکه نزدیکترین شی به مرکز خوشه در لیست ممنوعه نباشد، با بررسی تابع هدف و تعیین بهبود یا عدم بهبود جواب، مسیر حرکت بعدی مشخص می گردد.

#### ۴- منطق خوشه بندی در صورت خروج از بهینه محلی

در صورتیکه نزدیکترین جواب به مرکز خوشه در لیست ممنوعه نباشد، در صورت بهبود جواب مطابق با معیار ارزیابی خوشه بندی، دو روش استفاده شده است. در روش اول این اجازه داده می شود تا با منطق الگوریتم K-Means و تا زمانیکه بهبودی در جواب حاصل می گردد، خوشه بندی ادامه یابد. سپس نزدیکترین شی به مرکز خوشه که در لیست ممنوعه نیست، به عنوان مرکز خوشه جدید انتخاب می شود.

در روش دوم یک مرحله خوشه بندی با منطق K-Means اجرا و پس از محاسبه مرکز خوشه و تعیین کیفیت خوشه بندی، مطابق با معیار ارزیابی انحراف کم درون خوشه ها، بررسی می گردد که آیا نزدیکترین شی به مرکز خوشه جدید در لیست ممنوعه می باشد یا خیر. مطابق با شرط تعیین شده، منطق ادامه اجرای الگوریتم تعیین می گردد.

#### ۵- نقطه پایان الگوریتم

الگوریتم زمانی خاتمه می یابد که تمام خوشه ها به صورت ترتیبی بررسی گردند و بهبودی در جواب حاصل نگردد. با توجه به اینکه در بررسی هر خوشه و انتخاب مراکز خوشه جدید، خوشه بندی با منطق K-Means اجرا می گردد، لذا در صورت تغییر در تابع هدف، مراکز سایر خوشه ها نیز تغییر خواهند نمود. بررسی مجدد خوشه ها سبب یافتن جوابهای بهتر و نزدیک شدن به جواب بهینه می گردد.

### ۳-۲- الگوریتم ترکیبی جدید

همانگونه که در بخش قبل در خصوص اجزای الگوریتم توضیح داده شد، سه سناریو مختلف جهت ورود اشیاء به لیست ممنوعه و انتخاب مرکز خوشه بعدی وجود دارد:

- ۱- نزدیکترین شی به مرکز خوشه اولین خوشه بندی
- ۲- نزدیکترین شی به مرکز خوشه آخرین خوشه بندی
- ۳- نزدیکترین شی به مرکز بهترین خوشه بندی تا کنون

پس از طراحی و پیاده سازی معیارهای الگوریتم مطابق با سه سناریو فوق الذکر مشخص گردید که در صورت انتخاب نزدیکترین شی به مرکز بهترین خوشه بندی تا کنون، فضای بیشتری در مجموعه جواب داده ها با شکل‌های متفاوت مورد بررسی قرار گرفته و با دقت بالاتری جواب بهینه مساله خوشه بندی تولید می گردد. لذا الگوریتم جدید ترکیبی "K-Means+Tabu" و "Ant + Tabu" با منطق ورود اشیاء به لیست ممنوعه و انتخاب مرکز خوشه بعدی، از نزدیکترین شی به مرکز بهترین خوشه بندی تا کنون و پارامترهای به کار رفته به شرح ذیل توضیح داده می شود:

تعداد خوشه ها:  $K$

تعداد تکرار الگوریتم:  $Itr$

تعداد اشیاء در لیست ممنوعه:  $NTL$

نمایشگر بهبود در خوشه بندی مطابق با تابع هدف:  $Improve(k)$

بیشترین تعداد اشیاء در لیست ممنوعه:  $NTabuList$



الگوریتم:

۱-  $\hat{C}$  مراکز خوشه ها در جواب نهایی الگوریتم K-Means یا کلونی مورچه ها به عنوان ورودی مدل می باشد.  $\hat{C} = C$  و  $k=1$

۲-  $NTL=0$  و  $Improve(k)=0$

۳- خوشه  $k$ ام با مرکز خوشه  $C(k)$  جهت بررسی لیست ممنوعه انتخاب می شود.

۴- نزدیکترین شی به مرکز خوشه  $k$ ام (بهترین خوشه بدست آمده تا کنون)، که در لیست ممنوعه نیست، در لیست ممنوعه گذاشته می شود.

$$O(k) = \left\{ \min d(C(k), \bar{O}(k)) \mid \begin{array}{l} \forall \bar{O}(k) \in N(C(k)), \\ \bar{O}(k) \notin TL(k) \end{array} \right\}$$

$$TL(k) = TL(k) \cup O(k) \\ NTL = NTL + 1$$

۵- اگر  $NTL > NTabuList$  آنگاه

اگر  $Z(C(k)) < Z(\hat{C}(k))$  آنگاه  $Z(\hat{C}(k)) = Z(C(k))$  و  $Improve(k) = 1$  و  $k = k + 1$  به مرحله ۲ برو.

۶- در غیر این صورت، نزدیکترین شی به مرکز خوشه  $C(k)$  (بهترین خوشه بدست آمده تا کنون)، که در لیست ممنوعه نیست، به عنوان مرکز خوشه جدید انتخاب می شود. (سایر خوشه ها با همان مراکز خوشه قبلی)

$$O(k) = \left\{ \min d(C(k), \bar{O}(k)) \mid \begin{array}{l} \forall \bar{O}(k) \in N(C(k)), \\ \bar{O}(k) \notin TL(k) \end{array} \right\}$$

۷- خوشه بندی انجام و جواب مساله مطابق با تابع هدف محاسبه می گردد.

۸- اگر نزدیکترین شی به مرکز خوشه جدید  $k$ ام (بهترین جواب بدست آمده تا کنون) در لیست ممنوعه باشد،

$(O(k) \in TL)$  و  $Z(C(k)) < Z(\hat{C}(k))$  آنگاه به مرحله ۴ برو.

۹- اگر نزدیکترین شی به مرکز خوشه جدید  $k$ ام (بهترین جواب بدست آمده تا کنون) در لیست ممنوعه نباشد  $(O(k) \notin TL)$  و

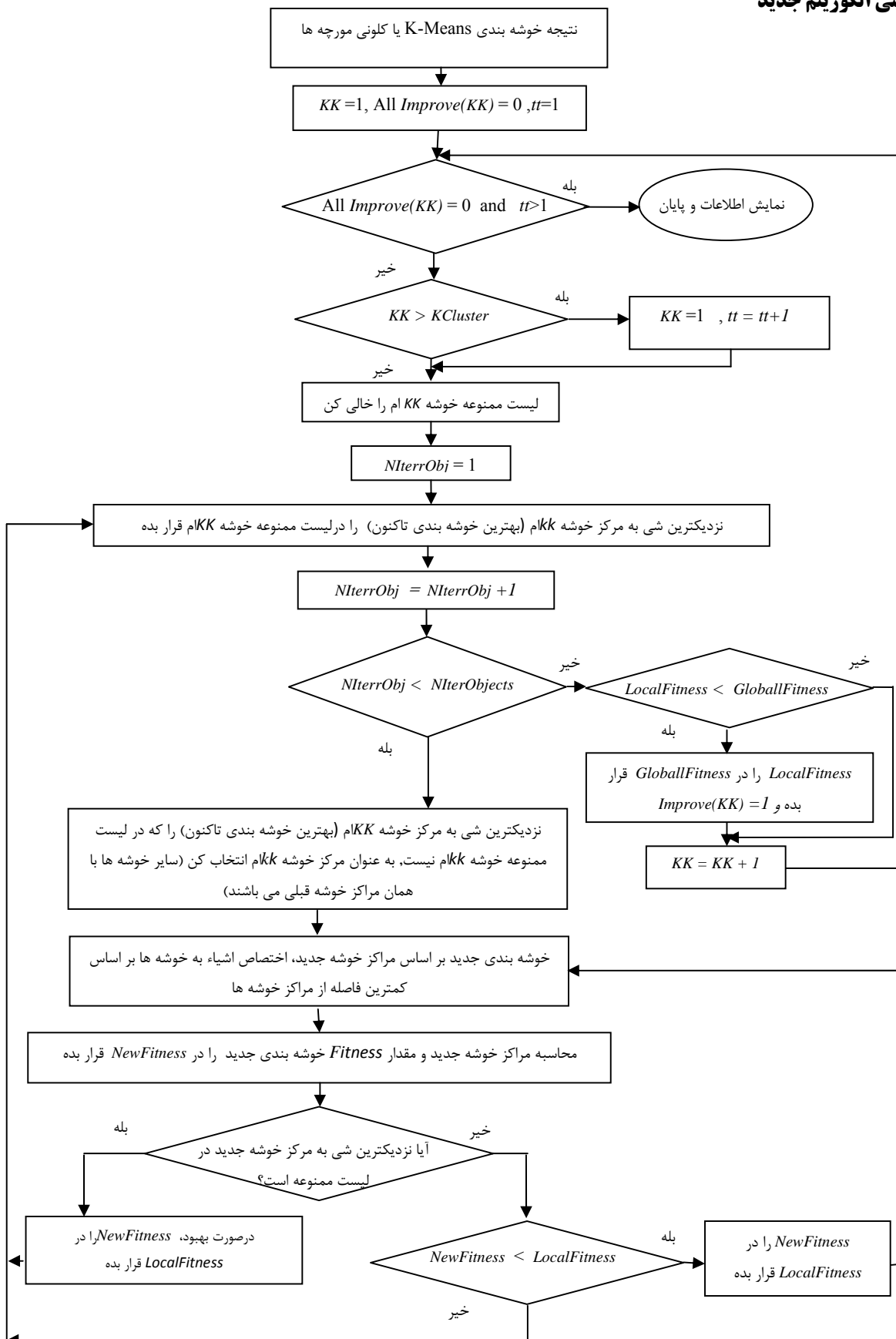
$Z(\hat{C}(k)) < Z(C(k))$  آنگاه  $Z(\hat{C}(k)) = Z(C(k))$  و به مرحله ۴ برو.

۱۰- اگر نزدیکترین شی به مرکز خوشه جدید  $k$ ام (بهترین جواب بدست آمده تا کنون) در لیست ممنوعه نباشد  $(O(k) \notin TL)$  و

$Z(\hat{C}(k)) > Z(C(k))$  آنگاه  $Z(\hat{C}(k)) = Z(C(k))$  به مرحله ۴ برو.

۱۱- در صورت عدم تحقق شرط خاتمه مساله، به مرحله ۲ برو

### ۳-۳- نمودار گردش الگوریتم جدید



## ۴- نتایج تست داده های مختلف

الگوریتم جدید با استفاده از نرم افزار Visual Basic در کامپیوتری با ۲ گیگابایت CPU و ۱ گیگابایت حافظه پیاده سازی و با داده های مختلف مورد تست و ارزیابی قرار گرفت. جهت تست الگوریتم، از دو گروه داده های تستی و استاندارد استفاده شده است. معیار ارزیابی تابع هدف، فاصله اقلیدسی می باشد که در آن مجموع مربعات انحراف درونی اشیاء از مراکز خوشه کمینه می گردد. پارامترهای مورد نیاز جهت پیاده سازی الگوریتمهای "K-Means+Tabu"، "NTL+K" و در الگوریتم "Ant+Tabu"،  $K$ ،  $NTL$ ،  $Itr$  و در الگوریتم  $P_{IS}$ ،  $Iteration$ ،  $Lq$ ،  $\rho$ ،  $\tau_0$ ،  $R$ ،  $Itr$  می باشد. در هر پایگاه داده، نتایج بهترین، میانگین و بدترین جواب در اجرای متفاوت هر الگوریتم مقایسه شده است.

### ۴-۱- داده های تستی

در اولین مرحله پس از پیاده سازی مدل، از ده پایگاه اطلاعاتی با داده های تستی استفاده شده است. این پایگاه داده ها شامل صد رکورد با داده های تصادفی و سه خصیصه می باشد که خصیصه سوم به برچسب خوشه اختصاص داده شده است. در تولید داده های تستی به این مساله توجه شده است که داده هایی انتخاب شوند که در صورت انتخاب تصادفی مراکز خوشه اولیه، در دام بهینه محلی قرار گرفته و ممکن است که جواب بهینه خوشه بندی را تولید ننماید.

با توجه به اینکه ورودی مدلهای پیشنهادی نتایج حاصل از الگوریتم K-Means و خوشه بندی کلونی مورچه ها می باشد، نتایج حاصل از اجرای الگوریتم K-Means در پایگاه داده های نمونه با مدل ترکیبی "K-Means+Tabu" و نتایج حاصل از اجرای الگوریتم کلونی مورچه ها با مدل ترکیبی "Ant + Tabu" مطابق با جداول ۲ و ۳ مقایسه می گردد.

**مشخصات الگوریتم جدید:** نزدیکترین شی به مرکز خوشه بهترین خوشه بندی تا کنون، به عنوان مرکز جدید خوشه بندی و همچنین ورود اشیاء به لیست ممنوعه استفاده می شود. در هر مرحله در صورت بهبود در جواب، یک مرحله خوشه بندی با منطق K-Means اجرا و شرط الگوریتم که آیا نزدیکترین شی به مرکز خوشه جدید در لیست ممنوعه می باشد یا خیر، بررسی می گردد.

جدول ۲- نتایج اجرای الگوریتم جدید با پایگاه داده های تستی و مقایسه با نتایج خوشه بندی K-Means

درصد بهبود	جواب بهینه	K-Means+Tabu			K-Means			پایگاه داده
		بهترین جواب	میانگین جواب	بدترین جواب	بهترین جواب	میانگین جواب	بدترین جواب	
۹۹.۲۷	۵.۶۹	۵.۶۹	۵.۷۴	۶.۰۶	۵.۶۹	۵.۹۴	۶.۵۵	داده نمونه ۱
۹۹.۹۸	۵.۴۸	۵.۴۸	۵.۴۸	۵.۴۸	۵.۴۸	۵.۵۹	۶.۳۹	داده نمونه ۲
۹۷.۳۶	۵.۹۷	۵.۹۷	۶.۱۳	۶.۴۶	۵.۹۷	۶.۱۷	۶.۶۶	داده نمونه ۳
۹۸.۰۷	۶.۱۲	۶.۱۲	۶.۲۴	۶.۷۱	۶.۱۲	۶.۳۸	۶.۷۷	داده نمونه ۴
۹۸.۸۸	۵.۷۸	۵.۷۸	۵.۸۴	۶.۰۷	۵.۷۸	۶.۰۰	۶.۶۸	داده نمونه ۵
۹۹.۵۹	۵.۸۲	۵.۸۲	۵.۸۴	۶.۲۰	۵.۸۲	۶.۰۳	۷.۰۸	داده نمونه ۶
۹۷.۷۰	۶.۰۸	۶.۰۸	۶.۲۲	۷.۱۱	۶.۰۸	۶.۳۳	۷.۵۱	داده نمونه ۷
۱۰۰	۵.۴۸	۵.۴۸	۵.۴۸	۵.۴۸	۵.۴۸	۵.۶۶	۷.۰۴	داده نمونه ۸
۹۹.۶۵	۵.۴۵	۵.۴۵	۵.۴۷	۵.۶۲	۵.۴۵	۵.۶۵	۶.۶۹	داده نمونه ۹
۱۰۰	۵.۴۶	۵.۴۶	۵.۴۶	۵.۴۷	۵.۴۶	۵.۴۷	۵.۵۳	داده نمونه ۱۰
۹۹.۰۵	۵.۷۳	۵.۷۳	۵.۷۹	۶.۰۶	۵.۷۳	۵.۹۲	۶.۶۹	میانگین

جدول ۳- نتایج اجرای الگوریتم جدید با پایگاه داده های نمونه و مقایسه با نتایج خوشه بندی Ant Clustering

پایگاه داده	Ant Clustering			Ant Clustering+Tabu			جواب بهینه	درصد بهبود
	بدترین جواب	میانگین جواب	بهترین جواب	بدترین جواب	میانگین جواب	بهترین جواب		
داده نمونه ۱	۱۰.۱۲	۹.۱۵	۸.۱۹	۵.۷۷	۵.۷۲	۵.۶۹	۵.۶۹	۹۹.۵۶
داده نمونه ۲	۹.۵۳	۸.۶۶	۷.۷۵	۶.۴۴	۵.۶۷	۵.۴۸	۵.۴۸	۹۶.۵۸
داده نمونه ۳	۱۰.۸۴	۱۰.۲۷	۹.۵۲	۶.۴۶	۶.۳۱	۵.۹۷	۵.۹۷	۹۴.۴۷
داده نمونه ۴	۱۱.۶۰	۹.۹۸	۸.۶۷	۶.۷۶	۶.۳۹	۶.۱۲	۶.۱۲	۹۵.۸۳
داده نمونه ۵	۱۰.۹۲	۹.۷۴	۸.۹۶	۶.۶۶	۶.۲۸	۵.۸۶	۵.۷۸	۹۲.۰۰
داده نمونه ۶	۱۰.۹۲	۹.۶۲	۸.۳۴	۶.۲۰	۶.۰۵	۵.۸۲	۵.۸۲	۹۶.۲۱
داده نمونه ۷	۱۱.۶۳	۹.۷۷	۸.۵۳	۷.۱۱	۶.۵۱	۶.۰۸	۶.۰۸	۹۳.۳۶
داده نمونه ۸	۱۰.۱۳	۹.۱۸	۸.۵۰	۵.۴۸	۵.۴۸	۵.۴۸	۵.۴۸	۱۰۰.۰۰
داده نمونه ۹	۱۰.۸۲	۹.۳۷	۸.۰۹	۶.۳۴	۵.۷۳	۵.۴۵	۵.۴۵	۹۵.۱۴
داده نمونه ۱۰	۱۰.۱۱	۹.۴۹	۸.۲۶	۵.۴۹	۵.۴۸	۵.۴۶	۵.۴۶	۹۹.۶۲
میانگین	۱۰.۶۶	۹.۵۲	۸.۴۸	۶.۲۷	۵.۹۶	۵.۷۴	۵.۷۳	۹۶.۲۸

#### ۴-۲- داده های استاندارد

جهت ارزیابی نتایج حاصل از مدل‌های ارایه شده با سایر الگوریتم‌های موجود، از دو پایگاه اطلاعاتی واقعی و استاندارد [۲۰] از سایت مجموعه داده های نمونه یادگیری ماشینی UCI استفاده شده است.

مجموعه داده گل زنبق (Iris): این پایگاه اطلاعاتی شامل ۱۵۰ رکورد تصادفی از نمونه های مختلف گل زنبق در سه خوشه می باشد. این مجموعه داده چهار خصیصه طول گلبرگ، عرض گلبرگ، طول کاسبرگ و عرض کاسبرگ دارد که مقادیر آنها به سانتیمتر داده شده است.

مجموعه داده بیماری تیروئید (Thyroid): این پایگاه اطلاعاتی شامل ۲۱۵ نمونه از انسانهای بیماری می باشد که نتایج آزمایش آنها نشان دهنده ابتلا به بیماری کم کاری و پرکاری غده تیروئید یا کارکرد صحیح آن می باشد. از ۲۱۵ بیماری که مورد آزمایش غده تیروئید قرار گرفته اند، ۱۵۰ نفر سالم، ۳۰ نفر بیماری پرکاری و ۳۵ بیمار مبتلا به کم کاری غده تیروئید می باشند. جهت تعیین بیماری در آزمایشات از ۵ خصیصه پزشکی استفاده شده است.

الگوریتم "K-Means+Tabu" در هر مجموعه داده ۱۰۰ بار و الگوریتم "Ant +Tabu" در هر مجموعه داده استاندارد ۱۰ بار اجرا شده که بهترین، میانگین و بدترین جوابها را با سایر الگوریتمهای فراابتکاری مانند ژنتیک، شبیه سازی سردشدن تدریجی، کلونی مورچه ها و جستجوی ممنوعه مقایسه و نتایج آن در جداول ۴ و ۵ نمایش داده شده است.

جدول ۴- نتایج اجرای الگوریتم "K-Means+Tabu" با پایگاه داده های تستی گل زنبق و تیروئید

پایگاه داده	K-Means			K-Means+Tabu			جواب بهینه	درصد بهبود
	بدترین جواب	میانگین جواب	بهترین جواب	بدترین جواب	میانگین جواب	بهترین جواب		
گل زنبق	۱۴۵.۲۸	۹۰.۵۱	۷۸.۹۴	۱۴۲.۸۸	۸۶.۶۱	۷۸.۹۴	۷۸.۹۴	۹۱.۱۴
تیروئید	۱۰۴۱۱.۲۸	۱۰۲۵۱.۱۹	۹۵۴۳.۸۳	۱۰۳۴۰.۶۳	۱۰۱۶۲.۱۳	۹۴۰۲.۷۵	۹۴۰۲.۷۵	۹۲.۵۳

جدول ۵- نتایج اجرای الگوریتم "Ant +Tabu" با پایگاه داده های تستی گل زنبق و تیروئید

درصد بهبود	جواب بهینه	Ant+Tabu Search			Ant Clustering			پایگاه داده
		بهترین جواب	میانگین جواب	بدترین جواب	بهترین جواب	میانگین جواب	بدترین جواب	
۹۴.۲۱	۷۸.۹۴	۷۸.۹۵	۱۴۳.۳۳	۱۸۰.۳۸	۱۴۷.۰۰	۱۸۳.۵۹	۲۵۱.۸۰	گل زنبق
۹۴.۲۶	۹۴۰.۲۷۵	۹۴۰.۲۷۵	۹۹۸۰.۱۳	۱۰۱۲۴.۴۸	۲۰۷۲۹.۳۵	۲۲۰۵۴.۳۱	۲۳۷۸۹.۶۶	تیروئید

در ادامه نتایج حاصل از الگوریتمهای جدید با نتایج سایر الگوریتمهای فراابتکاری [۹] از قبیل الگوریتم ژنتیک، شبیه سازی سردشدن تدریجی، خوشه بندی کلونی مورچه ها و جستجوی ممنوعه مقایسه و نتایج آن در جداول ۶ و ۷ نمایش داده شده است.

جدول ۶- مقایسه نتایج اجرای الگوریتم جدید با داده استاندارد گلهای زنبق، با سایر الگوریتمهای فراابتکاری

الگوریتم	بهترین جواب	میانگین جواب	بدترین جواب	زمان اجرا (ثانیه)
ACO	۹۷.۱۰۰۷۷۷	۹۷.۱۷۱۵۴۶	۹۷.۸۰۸۴۶۶	۳۳.۷۲
GA	۱۱۳.۹۸۶۵۰۳	۱۲۵.۱۹۷۰۲۵	۱۳۹.۷۷۸۲۷۲	۱۰۵.۵۳
TS	۹۷.۳۶۵۹۷۷	۹۷.۸۶۸۰۰۸	۹۸.۵۶۹۴۸۵	۷۲.۸۶
SA	۹۷.۱۰۰۷۷۷	۹۷.۱۳۶۴۲۵	۹۷.۲۶۳۸۴۵	۹۵.۹۲
K-Means+Tabu	۷۸.۹۴۰۸۴۱	۷۸.۹۴۰۸۴۱	۹۲.۳۵۸۳۱۴	۰.۹
Ant+Tabu	۷۸.۹۴۵۰۶۶	۱۴۳.۳۳۸۲۲	۱۸۰.۳۸۴۳۲۱	۸۵.۲۵

جدول ۷- مقایسه نتایج اجرای الگوریتم جدید با داده استاندارد تیروئید، با سایر الگوریتمهای فراابتکاری

الگوریتم	بهترین جواب	میانگین جواب	بدترین جواب	زمان اجرا (ثانیه)
ACO	۱۰۱۱۱.۸۲۷	۱۰۱۱۲.۱۲۶	۱۰۱۱۴.۸۱۹	۱۰۲.۱۵
GA	۱۰۱۱۶.۲۹۴	۱۰۱۲۸.۸۲۳	۱۰۱۴۸.۳۸۹	۱۵۳.۲۴
TS	۱۰۲۴۹.۷۲۹	۱۰۳۵۴.۳۱۵	۱۰۴۳۸.۷۸۰	۱۱۴.۰۱
SA	۱۰۱۱۱.۸۲۷	۱۰۱۱۴.۰۴۵۲	۱۰۱۱۸.۹۳۴	۱۰۸.۲۲
K-Means+Tabu	۹۴۰.۲۷۵۲	۱۰۱۶۲.۱۲۶	۱۰۳۴۰.۶۳۰	۱.۱
Ant+Tabu	۹۴۰.۲۷۵۲	۹۹۸۰.۱۳۱	۱۰۱۲۴.۴۷۶	۱۲۳.۷

همانگونه که از مقایسه نتایج حاصل از اجرای مدل‌های جدید خوشه بندی با سایر الگوریتم‌های فراابتکاری در جداول ۲ تا ۷ مشاهده می گردد، الگوریتم ترکیبی جدید، نتایج بهتری را نسبت به سایر الگوریتم‌ها تولید نموده است. در واقع مقایسه بهترین، میانگین و بدترین جواب الگوریتم ترکیبی جدید، با سایر الگوریتم‌های فراابتکاری، نشان می دهد که مدل جدید پیشنهادی توانسته است کارایی و اثربخشی خوشه بندی را افزایش دهد.

در این پژوهش دو الگوریتم ترکیبی جدید "K-Means+Tabu" و "Ant +Tabu" نتایج بهتری را در جواب‌های بهترین، میانگین و بدترین در مقایسه با سایر الگوریتم‌های فراابتکاری تولید نموده است و با درصد بالایی موجب خروج از مینیمم محلی و رسیدن به جواب بهینه مساله شده است.

## ۵- نتیجه گیری

الگوریتم K-Means و خوشه بندی کلونی مورچه ها از جمله الگوریتم‌های خوشه بندی جهت استخراج دانش از داده ها می باشند که دارای مشکلاتی بوده و به دلیل انتخاب تصادفی اشیاء در ابتدای الگوریتم، لزوماً جواب بهینه خوشه بندی را تولید نمی نماید. در سال‌های اخیر پژوهش‌های متعددی جهت انتخاب صحیح اولیه اشیاء، تعیین بهینه تعداد خوشه و یافتن جواب بهینه مساله در حل مسائل خوشه بندی انجام شده است.

در این مقاله با استفاده از الگوریتم فراابتکاری جستجوی ممنوعه، الگوریتم‌های خوشه بندی جدیدی طراحی و آرایه شده است که با درصد بالایی جواب بهینه مساله را در خوشه بندی N شی به K خوشه تولید می نماید. الگوریتم پیشنهادی با ترکیب الگوریتم K-Means+Tabu و AntClustering +Tabu Search طراحی و با داده های تستی و استاندارد تست شده است. در طراحی الگوریتم جدید، سه شیوه مختلف جهت ورود اشیاء به لیست ممنوعه و انتخاب مرکز خوشه بعدی مورد توجه قرار گرفته است:

- نزدیکترین شی به مرکز بهترین خوشه بندی تا کنون
- نزدیکترین شی به مرکز خوشه آخرین خوشه بندی
- نزدیکترین شی به مرکز خوشه اولین خوشه بندی

پس از پیاده سازی الگوریتم جدید با روش‌های فوق الذکر، در بررسی کلیه نتایج بدست آمده، استفاده از نزدیکترین شی به مرکز بهترین خوشه بندی تا کنون، جهت ورود اشیاء مرکز خوشه به لیست ممنوعه و انتخاب مرکز خوشه جدید، موجب ایجاد خوشه هایی با کمترین انحراف درونی در اشیاء هر خوشه می گردد و با دقت بالاتری جواب بهینه را تولید می نماید. مقایسه مدل جدید با سایر الگوریتم‌های فراابتکاری ژنتیک، شبیه سازی سردشدن تدریجی، جستجوی ممنوعه و خوشه بندی کلونی مورچه ها نیز نشان می دهد که الگوریتم‌های جدید "K-Means+Tabu" و "Ant +Tabu"، از کارایی و اثربخشی بالاتری جهت خروج از مینیمم محلی و رسیدن به جواب بهینه خوشه بندی برخوردار می باشد.

## ۶- مراجع

- [1]. D. Steinley and M.J.Brusco, "Initializing K-means Clustering: A Critical Evaluation of Several Techniques", Journal of Classification 24. 2007, pp.99-121.
- [2]. Z. Zhang , B.Tian Dai and A.K.H. Tung," On the Lower Bound of Local Optimum in Kmeans Algorithm", Proc. 6<sup>th</sup> Int Conf.Data Mining.(ICDM'06),IEEE. 2006,pp.2701-2709.
- [3]. D. Steinly and L.Hubert, "Order-Constrained Solutions in K-Means Clustering: Even Better Than Being Globally Optimal", The Psychometric Society, Dec. 2008, pp.647-664.
- [4]. A.Misevicius,T.Blazauskas, J.Blonskis and J. Smolinkas, "An Overview of Some Heuristic Algorithms for Combinatorial Optimization Problems", Informacines Technologjos IR Vadymas, 2004.
- [5]. K.S. Al-Sultan. "A tabu search approach to the clustering problem". Pattern Recognition, 28(9):1443-1451, 1995.

- [6]. R.J. Mullen , D.Monekosso ,S.Barman and P.Remagnino, “*A Review of Ant Algorithms*”, Expert Systems with Applications,Elsevier. 2009.
- [7]. Marco. Dorigo and C.Blum, “*Ant Colony Optimization Theory:A Survey*”,Theoretical Computer Science 344,Elsevier. 2005,pp.243-278.
- [8]. U. Boryczka,” *Finding Group in data: Cluster analysis with ants*”, Applied Soft Computing. 2009, pp. 61-70.
- [9]. P.S Shelokar,V.K.Jayaraman and B.D.Kulkarni, “*An Ant Colony approach for clustering*”, Analytica Chimica Acta , Elsevier. 2004, pp.187-195.
- [10]. L. Chen, X.H XU and Y.X CHEN, “*An Adaptive Ant Colony Clustering Algorithm*”, Proc. 3<sup>th</sup> International Conf. Machine Learning and Cybernetics, Shanghai,IEEE. 2004, pp.1387-1392.
- [11]. Z. Gungor and A. Unler, “*K-harmonic means data clustering with Tabu Search Method*”, Applied Mathematical Modelling, Elsevier. 2008, pp.1115-1125.
- [12]. H. Youssef,S.M.Sait and H. Adiche, “*Evolutionary algorithms, simulated annealing and tabu search: a comparative study*”, Engineering Applications of Artificial intelligence, Elsevier. 2001, pp.167-181.
- [13]. Y. Liu, Y. Liu,L. Wang and k. Chen, “*A Hybrid Tabu Search Based Clustering Algorithm*”, KES 2005, LNAI3682, Springer-Verlag Berlin Heidelberg, 2005,pp. 186-192.
- [14]. S.M. Pan and K.S.Senior,” *Evolution-Based Tabu Search Approach to Automatic Clustering*” , IEEE Transaction on Systems,Man, and Cybernetics, Part C-Application and Reviews, Vol 37,No. 5, Sep. 2007, pp.827-838.
- [15].A.L.Vizine,L.N.de Castro, E.R.Hruschka and R.R.Gudwin, “*Towards Improving Clustering Ants: An Adaptive Ant Clustering Algorithm*” , Informatica 29. 2005, pp.143-154.
- [16].Y.Liu, L.Wang and K.Chen, ”*A Tabu Search Based Method for Minimum Sum of Squares Clustering*”, ICAPR 2005, LNCS 3686, Springer-Verlag Berlin Heidelberg, 2005, pp. 248-256.
- [17].Y.He,S.C.Hui and Y.Sim, “*A Novel Ant-Based Clustering Approach for Document Clustering*”, AIRS 2006, LNCS 4182, Springer-Verlag Berlin Heidelberg, 2006, pp. 537-544.
- [18].Z.J.Lee, S.F.Su, C.C.Chuang and K.H.Liu, “*Genetic algorithm with ant colony optimization (GA-ACO) for multiple sequence alignment*”, Applied Soft Computing, Elsevier. 2008, pp.55-78.
- [19].T.Niknam, J.Olamaei, B.Amiri, “*A Hybrid Evolutionary based on ACO and SA for Cluster Analysis*”, Journal of Applied Sciences, Asian Network for Scientific Information, 2008, pp. 2695-2702.
- [20].UCI Repository of Machine Learning Databases retrievevd from the World Wide Web:  
<http://www.ics.uci.edu/~mllearn/MLRepository.html>