

دانشکده مهندسی کامپیوتر و فناوری اطلاعات
دانشگاه صنعتی امیرکبیر
تهران ایران
hossein.morshedlou@yahoo.com, mmeybodi@aut.ac.ir

: امروزه با توجه به حجم بسیار بالا و نیز روند رو به رشد میزان داده‌هایی که بشر با آن سرو کار دارد، نیاز به روش‌هایی همچون داده‌کاوی برای تحلیل این حجم بزرگ از اطلاعات احساس می‌گردد. از طرفی با توجه به نرخ بسیار بالای تولید داده‌ها و اطلاعات و نیاز به تحلیل سریع آنها احتیاج به روش‌های داده‌کاوی داریم که بتوانند با سرعت بالایی داده‌ها را تحلیل نمایند. یکی از معمول‌ترین روش‌های افزایش سرعت، موازی سازی اعمال می‌باشد. با توجه کاربرد روزافزون آtomاتاتی سلولی در کاربردهای مختلف و نیز توان موازی‌سازی بسیار بالای آن، بنظر می‌رسد استفاده از آtomاتاتی سلولی برای داده‌کاوی زمینه‌ای بسیار جالب و قابل توجه باشد. در این مقاله روشی برای خوشبندی، که یکی از مهم‌ترین شاخه‌های داده‌کاوی می‌باشد، ارائه شده است که در آن از مفاهیم آtomاتاتی سلولی و کلونی مورچه‌ها استفاده گردیده است. در روش ارائه شده در این مقاله هر داده متعلق به مجموعه داده‌ای که باید خوشبندی گردد، به یک مورچه مناسب می‌گردد و این مورچه‌ها هستند که بطور همزمان در فضای آtomاتاتی سلولی جابجا شده و سعی می‌کنند تا با مورچه‌هایی که دارای داده‌های مشابه می‌باشند، اجتماع تشکیل دهند و بدین طریق عمل خوشبندی را انجام دهند. آزمایشات انجام شده برای تست روش پیشنهاد شده، توانایی و قابلیت بالای این روش را برای خوشبندی داده‌ها نشان می‌دهد.

:

A Cellular Automata based Data Clustering Method

H. Morshedlou M. R. Meybodi

Soft Computing Lab.

Computer Engineering and Information Technology Department

Amirkabir University of Technology, Tehran, Iran

hossein.morshedlou@yahoo.com, mmeybodi@aut.ac.ir

Abstract: Nowadays, regarding huge data volumes and increasing rate of data generation, the need for data analysis methods like data mining is increased. Furthermore, with the real time requirement and size of data that should be processed, these methods must be able to act very fast. One of the well-known techniques to increase speed is parallelism. With respect to increasing use of cellular automata in various domains and its high power in parallelism, it seems that using cellular automata for data mining is an appealing approach. This paper presents a clustering method named CAC, which is based on the cellular automata and ant-colony concepts. In this method, each data point from the dataset that should be clustered is assigned to an ant. The ants move in grid space simultaneously and find other ants with similar data to form a community. Clustering operation is done by constitution of these communities. Experiments conducted on standard datasets using this method show the capability and applicability of our approach for data clustering.

Keywords: Cellular Automata, Clustering, Ants Colony

[]

[]

[]

[]

(Cellular Automata based Clustering) CAC

[]

(
[])

[]

[]

[]

CAC

(CAC)

CAC

[]

CAC

CAC

CAC

[]

CAC

[]

[]

CAC

CAC

[][]

•

•

•

•

()

Q d S_d (S_d, Q, N, δ)
 δ N

(CAC)

CAC

S_d

()

n

$2 \times \lfloor \sqrt{n} \rfloor$

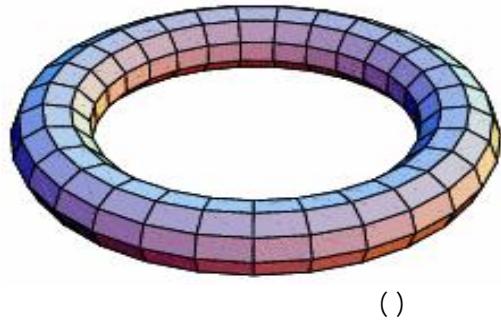
$2 \times \lfloor \sqrt{n} \rfloor$

C_i

i

C_i i

(i, C_i)



()

:

$$N(ant_i) = \{ x \bmod w(n), y \bmod h(n) \mid |x - x_i| \leq 1, |y - y_i| \leq 1 \} \quad (1)$$

$$\delta \times \left\lfloor \sqrt{n} \right\rfloor \quad h(n) - w(n)$$

$$Ant_i \quad data_j = (z_{j1}, z_{j2}, \dots, z_{jk}) \quad data_i = (z_{i1}, z_{i2}, \dots, z_{ik}) \\ dist_{ij} \quad Ant_j$$

$$dist_{ij} = \sqrt{\sum_{x=1}^k ((1 - \frac{|data_i(x) - data_j(x)|^2}{d_{ij}^2}) \times |data_i(x) - data_j(x)|^2)} \quad (2)$$

$$d_{ij}$$

$$d_{ij} = d(ant_i, ant_j) = d(data_i, data_j) = \|data_i - data_j\| \quad (3)$$

$$|data_i(x) - data_j(x)|^2 \quad (4) \quad 1 - \frac{|data_i(x) - data_j(x)|^2}{d_{ij}^2} \\ |data_i(x) - data_j(x)|^2$$

$$() \quad \text{(Information Gain)}$$

fitness : **fitness**

$$t+1 \quad i \quad f_{t+1}(ant_i) \quad () \\ t \quad f^{\frac{1}{5}}_t(ant_j)$$

$$f_{t+1}(ant_i) = \min\{ -1, \max\{ 0, \frac{1}{8} \sum_{ant_j \in N(ant_i)} (m \times f^{\frac{1}{5}}(ant_j))^x \times (1 - \frac{dist_{ij}}{\alpha_i}) \} \}$$

if ($dist_{ij} > \gamma$) $x = 0$

(4)

if ($dist_{ij} < \gamma$ and $f_t(ant_j) \leq \beta$) $x = 1, m = 1$

if ($dist_{\text{..}} \leq \gamma$ and $f_{\cdot}(\text{ant}_{\cdot}) \geq \beta$) $x \equiv 1$, $m \equiv 1$

2

7

$$f_t$$

β

()

$$\alpha_i$$

$$\alpha_i = \frac{1}{n-1} \sum_{j=1}^n dist_{ij}$$

(ω)

$$p_a(\text{ant}_i)$$

10

$$p_a(ant_i) = \min\{p, (\frac{\beta}{\beta + f_t} + \frac{f'_t}{f'_t + m \times f_t})\} \quad (8)$$

$$\frac{\beta}{\beta + f_t}$$

$$\frac{\beta}{\beta + f_t}$$

β

$$f_t$$

$$) \qquad \beta \qquad f_t$$

$$\beta f_t$$

$$\frac{\beta}{\beta + f_t}$$

(

$$f_t'$$

()

f_t'

$$\frac{f_t'}{f_t' + m \times f_t}$$

$$) \quad f_t' \quad f_t \quad . \quad (f_t') \quad ($$

$$f'(ant_i) = \max\left\{0, \frac{1}{8} \sum_{ant_j \in Pheromone} \frac{1}{\sqrt{t_{current} - t_{ph_j}}} \times \left(1 - \frac{dist_{ij}}{\alpha_i}\right)\right\} \quad ()$$

$$i \quad C_i$$

$$(i, C_i)$$

C_i

i

C_i

$$\frac{1}{\sqrt{t_{Current} - t_{ph_j}}} \quad ()$$

()

$$v = (t_{current} - t_{ph_j})^{\frac{1}{3}} \times \sqrt{\sum_{x=1}^k (1 - \frac{|data_i(x) - data_j(x)|^2}{d_{ij}^2}) \times |data_i(x) - data_j(x)|^2} \quad ()$$

$$\alpha_i$$

()

()

α_i

($p \geq 0.90$) p

$$(1-p)$$

Glass Iris

[]

Iris

:Iris

class

Glass

:Glass

(*id*

Iris

()

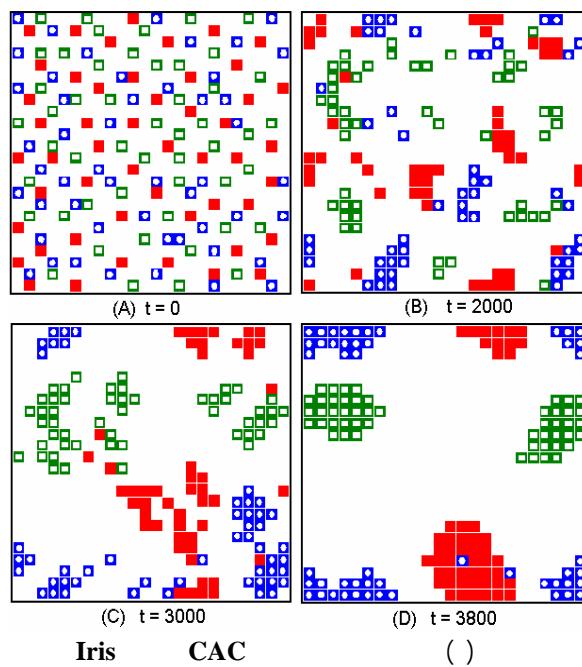
[] SA4C A4C

Glass

CAC

()

$$\beta = 0.2 \quad m = 2, p = 0.95$$



CAC

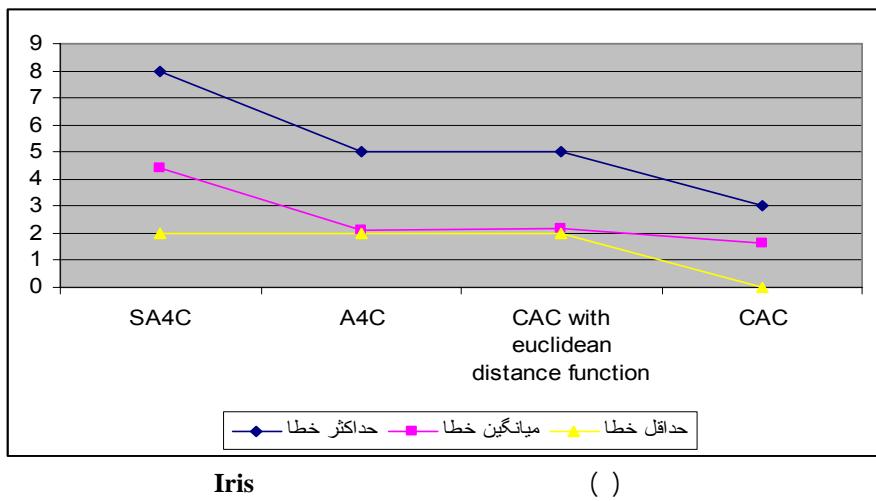
CAC

CAC

CAC

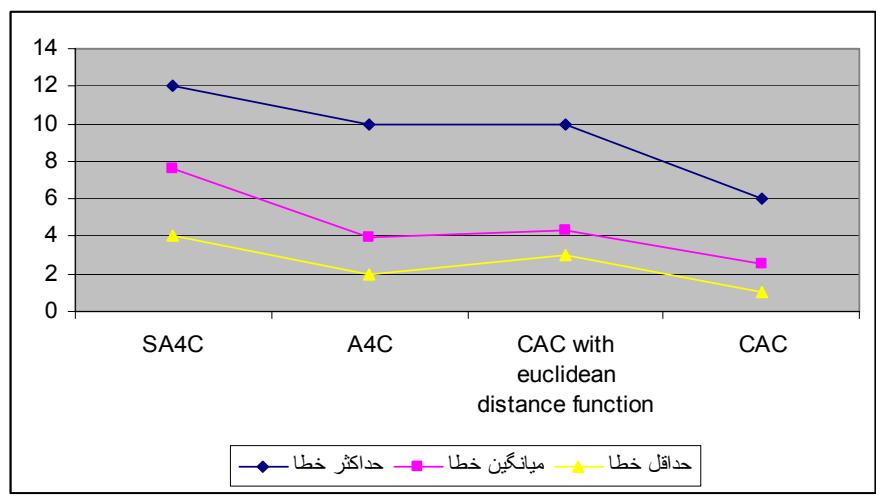
CAC

7



Iris

()



Glass

()

- [1] Chen, L., Xu, X., Chen, Y. and He, P., "A Novel Ant Clustering Algorithm Based on Cellular Automata", Intelligent Agent Technology (IAT 2004), 2004, pp.148-154.
- [2] Vande Moere, A., James Clayden, J., "Cellular Ants: Combining Ant-Based Clustering with Cellular Automata", Proceedings of the 17th IEEE International Conference on Tools with Artificial Intelligence table of contents, Pages: 177 - 184, 2005.
- [3] Ramos, V. and Abraham, A., "Evolving a Stigmergic Self-Organized Data-Mining", Conf. on Intelligent Systems, Design and Applications (ISDA-04), 2004, pp.725-730.
- [4] Schockaert, S., De Cock, M., Cornelis, C. and Kerre, E.E., "Fuzzy Ant Based Clustering", Lecture Notes in Computer Science, 3172. Pp.342-349.
- [5] Handl, J., Knowles, J., and Dorigo, M. (2003). "On the performance of ant-based clustering". In Design and Application of Hybrid Intelligent Systems, Vol. 104 of Frontiers in Artificial Intelligence and Applications (pp. 204-213). Amsterdam, the Netherlands; IOS Press.
- [6] Lumer, E.D. and Faieta, B., "Diversity and Adaptation in Populations of Clustering Ants", From Animals to Animates: Conference on Simulation of Adaptive Behaviour, 1994, pp.501-508.
- [7] Handl, J. and Meyer, B. "Improved ant-based clustering and sorting in a document retrieval interface", PPSN VII, LNCS 2439, 2002.
- [8] Kuntz, A., Layzell, P. and Snyers, D., "A Colony of Ant-Like Agents for Partitioning in VLSI Technology", European Conf.e on Articial Life, 1997, MIT, pp.417-424.
- [9] Wolfram, S., *Universality and complexity in cellular automata*, Physica D, no. 10, pp. 1-35, January 1984.
- [10] Wolfram, S., *Cellular Automata*, Los Alamos Science, vol. 9, pp. 2-21, Fall 1983.
- [11] <http://kdd.ics.uci.edu>