

روشی مبتنی بر اتوماتای سلولی برای خوشه‌بندی داده‌ها

حسین مرشدلو محمدرضا میبیدی

آزمایشگاه محاسبات نرم
دانشکده مهندسی کامپیوتر و فناوری اطلاعات
دانشگاه صنعتی امیرکبیر
تهران ایران

چکیده

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 داده-کاوی 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 اتوماتای سلولی در کاربردهای مختلف و نیز توان موازی‌سازی بسیار بالای آن، بنظر می‌رسد استفاده از اتوماتای سلولی برای داده‌کاوی زمینه‌ای بسیار جالب و قابل توجه باشد. در این مقاله روشی برای خوشه‌بندی، که یکی از مهم‌ترین شاخه‌های داده‌کاوی می‌باشد، ارائه شده است که در آن از مفاهیم اتوماتای سلولی و کلونی مورچه‌ها استفاده گردیده است. در روش ارائه شده در این مقاله هر داده متعلق به مجموعه داده‌ای که باید خوشه-بندی گردد، به یک مورچه منتسب می‌گردد و این مورچه‌ها هستند که بطور همزمان در فضای اتوماتای سلولی جابجا شده و سعی می‌کنند تا با مورچه‌هایی که دارای داده‌های مشابه می‌باشند، اجتماع تشکیل دهند و بدین طریق عمل خوشه‌بندی را انجام دهند. آزمایشات انجام شده برای تست روش پیشنهاد شده، توانایی و قابلیت بالای این روش را برای خوشه‌بندی داده‌ها نشان می‌دهد.

کلمات کلیدی: اتوماتای سلولی، خوشه‌بندی، کلنی مورچه‌ها

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

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