

ชื่อเรื่องการค้นคว้าแบบอิสระ

การจัดกลุ่มข้อมูลด้วยวิธีการใช้ต้นไม้แบบทอดข้ามต่ำสุด

ผู้เขียน

นายชุมพล บุญคุ้มพรภัทร

ปริญญา

วิทยาศาสตรมหาบัณฑิต (วิทยาการคอมพิวเตอร์)

อาจารย์ที่ปรึกษาการค้นคว้าแบบอิสระ

อาจารย์ ดร.สรพรพรรณ กันตะบุตร

บทคัดย่อ

อัลกอริทึมเค-มินส์ เป็นวิธีที่นิยมใช้ในการจัดกลุ่มข้อมูลในปัจจุบัน อย่างไรก็ตามเมื่อข้อมูลเข้ามีขนาดใหญ่อัลกอริทึมจะทำงานช้ามาก เพราะอัลกอริทึมเค-มินส์มีความซับซ้อนของเวลาเป็น $O(kn)$ เมื่อ k เป็นจำนวนของกลุ่มที่ต้องการ r เป็นจำนวนของการดำเนินงาน และ n เป็นจำนวนของจุด เป็นที่รู้กันว่า เค-มินส์มีลักษณะเป็นเชิงเส้น (linear rate) เมื่อเข้าใกล้การเบนเข้าหากัน (convergence) ถ้า r เป็นสัดส่วนกับ n เวลาที่ใช้ในการทำงานของ เค-มินส์จะเท่ากับ $O(n^2)$ ซึ่ง n ที่มีขนาดใหญ่มาก จะไม่สามารถทำงานได้จริงในขอบเขตของเวลาที่ต้องการ ในผลการค้นคว้านี้ได้นำเสนออัลกอริทึมการจัดกลุ่มข้อมูลด้วยวิธีการใช้ต้นไม้แบบทอดข้ามต่ำสุด มีความซับซ้อนของเวลาเป็น $O(n \lg n)$ ซึ่งมีเวลาทำงานเร็วกว่าอัลกอริทึมเค-มินส์ นอกจากนี้อัลกอริทึมการจัดกลุ่มข้อมูลด้วยวิธีการใช้ต้นไม้แบบทอดข้ามต่ำสุดยังมีผลลัพธ์ที่ได้ดีกว่าอัลกอริทึมเค-มินส์

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่

Copyright © by Chiang Mai University

All rights reserved

Independent Study Title Minimum Spanning Tree-based Clustering

Author Mr. Chumphol Bunkhumpornpat

Degree Master of Science (Computer Science)

Independent Study Adviser Lecturer Dr. Sanpawat Kantabutra

ABSTRACT

K-means algorithm is one of the most popular clustering algorithms in use today. However, when the input size is large, the algorithm runs very slow. This is because it has time complexity of $O(krn)$ where k is the number of desired clusters, r is the number of iterations, and n is the number of points. It is known that K-means approaches a linear rate when it is close to convergence. If r is proportional to n , the running time of K-means is $O(n^2)$ which is impractical for a large n . In this papers, we offer the minimum spanning tree-based clustering (MST-based clustering). It has time complexity of $O(n \lg n)$ and its running time is faster than that of K-means algorithm. Besides, the clustering result of our algorithm is better than that of K-means algorithm.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved