

การจัดกลุ่มข้อมูลโดยใช้ Self-Organizing Map (SOM) ข้อมูลจะถูกจัดให้อยู่ในรูปแบบของแผนภาพ 2 มิติ จุดเด่นคือกลุ่มของข้อมูลที่มีลักษณะคล้ายกันจะอยู่ในโหนดใกล้เคียงกัน แต่ในกรณีที่แผนภาพมีขนาดใหญ่ข้อมูลที่อยู่ในกลุ่มเดียวกันอาจจะแตกออกเป็นกลุ่มย่อยอยู่ในโหนดที่ห่างออกไป ทำให้ไม่สามารถระบุกลุ่มได้อย่างชัดเจนและทำให้การสำรวจแผนภาพเป็นไปได้ด้วยความยากลำบาก ดังนั้นจำเป็นต้องจัดกลุ่มของโหนดในแผนภาพ SOM หลังจากเสร็จสิ้นกระบวนการเรียนรู้ งานวิจัยนี้นำเสนอการจัดกลุ่มโดยใช้เจเนติกอัลกอริทึมกับแผนภาพ SOM เพื่อแก้ไขปัญหาการกระจายของข้อมูล โดยแบ่งการทำงานออกเป็น 2 ขั้นตอนหลักคือ ขั้นตอนแรกเป็นการจัดกลุ่มข้อมูลโดยใช้แผนภาพ SOM ขั้นตอนที่สองเป็นการจัดกลุ่มโหนดของแผนภาพ SOM โดยใช้เจเนติกอัลกอริทึม ในการทดลองกับข้อมูล KDD cup 1999 ซึ่งเป็นข้อมูลรูปแบบพฤติกรรม การบุกรุกเครือข่าย และเปรียบเทียบกับการจัดกลุ่มโหนดโดยใช้วิธี K-mean ผลปรากฏว่าการจัดกลุ่มโดยใช้เจเนติกอัลกอริทึมที่นำเสนอให้ประสิทธิภาพในการจัดกลุ่มที่ดีกว่า โดยวัดประสิทธิภาพจากค่าเอนโทรปี

ABSTRACT

187702

Data clustering using self-organizing map (SOM) is represented as a two dimensional map. The advantage of this method is that similar feature of data are clustered into the neighbor node. In case of a large map, the SOM may separate those data into sub-groups that are in the other nodes. It is difficult to identify the appropriate groups. Therefore, after finished the training process it is necessary to cluster the node in SOM's map again. In this research we present a clustering method using genetic algorithm in SOM's map. Our proposed algorithm has two processes. First, cluster data using SOM and second, cluster nodes in SOM using genetic algorithm. In our experiments, we applied SOM to KDD cup 1999 dataset. Then genetic algorithm and K-Mean were separately applied to the results from SOM. The final results of genetic algorithm yields a better performance than that of K-mean based on Entropy value.