T165522

บทคัดย่อ

ลักษณะการจัดกลุ่มแบบไม่มีการชี้นำนั้นจะทำงานได้ดีเมื่อข้อมูลแยกออกจากกันอย่างชัดเจน แต่ข้อมูลโดยทั่วไปนั้นมีลักษณะแยกออกจากกันไม่ชัดเจนนัก จึงทำให้การจัดกลุ่มแบบไม่มีการชี้นำแบบเดิมนั้นมีประสิทธิภาพไม่เพียงพอเท่าที่ควร การจัดกลุ่มเอกสารโดยใช้โคโฮเนนนิวรอลเน็ตเวิร์คร่วมกับกฎระหว่างสองข้อมูลเป็นการพัฒนาปรับปรุงเท็กโปรเซสซิงโคโฮเนนนิวรอลเน็ตเวิร์ค ซึ่งเป็นเทคนิคการจัดกลุ่มแบบไม่มีการชี้นำให้มีประสิทธิภาพมากขึ้น โดยนำแนวคิดของโคโฮเนนเซ็ลฟออร์แกไนซิงแม็ป มาปรับปรุงความสามารถเพื่อทำการจัดกลุ่มข้อมูลประเภทข้อความได้โดยตรงโดยทำการประยุกต์แนวคิดเรื่องการเปรียบเทียบความแตกต่างของข้อมูลเชิงสัญลักษณ์ รวมทั้งนำแนวคิดการจัดกลุ่มข้อมูลโดยใช้กฎระหว่างสองข้อมูลมาใช้ในการจัดกลุ่มข้อมูลประเภทข้อความด้วย แนวคิดการจัดกลุ่มเอกสารที่นำเสนอนี้สามารถจัดกลุ่มข้อมูลประเภทข้อความได้อย่างถูกต้อง และมีประสิทธิภาพ

ABSTRACT

TE 165522

Unsupervised clustering obtains a good performance when data are completely separated. However, most data are incompletely separated. As a result, unsupervised clustering gives poor performance, when they are applied to such data. This paper proposes document clustering using Kohonen neural network with pairwise constraints to improve the text processing Kohonen neural network. This algorithm works directly on textual information without mapping document into some representation which has quantitative features. The input level of the proposed neural network can directly receive a qualitative value without mapping the qualitative value into numerical value. The proposed neural network is based on the architecture of text processing Kohonen neural network, the concepts of dissimilarity measure of symbolic objects and pairwise constrained concepts. As a result, the model can successfully assign cluster label to the objects.