

วัตถุประสงค์ของการทำ clustering(การจัดกลุ่ม) คือ การแยกข้อมูลออกเป็นกลุ่มๆ โดยข้อมูลที่อยู่ภายในกลุ่มเดียวกันจะมีค่าความเหมือนกันมากกว่าข้อมูลที่อยู่ต่างกลุ่มกัน เทคนิคในการ clustering สำหรับวัตถุที่มีค่าของคุณสมบัติเป็นค่าเชิงตัวเลข (numerical values) เป็นที่รู้จักกันอย่างแพร่หลาย เช่น K-mean, nearest neighbor, decision tree และ neural networks วิธีเหล่านี้สามารถทำการ clustering ข้อมูลที่มีคุณสมบัติเป็นค่าเชิงตัวเลขได้เป็นอย่างดี แต่ในปัจจุบัน เริ่มมีความสนใจในการทำ clustering ข้อมูลที่เป็นประเภทข้อความหรือเอกสารมากขึ้น แม้ว่าวิธีในการ clustering แบบเดิมจะสามารถใช้ในการ clustering เอกสารหรือข้อความได้ แต่ต้องมีการนำข้อมูลไปผ่านขั้นตอนการแปลงรูปข้อมูลคุณสมบัติของเอกสารเหล่านั้นให้เป็นค่าเชิงตัวเลขเสียก่อน จึงจะนำมาใช้ได้ วิธีนี้แม้จะเป็นที่ยอมรับกัน แต่การแปลงข้อมูลที่เป็นข้อความให้เป็นค่าเชิงตัวเลขนั้น อาจทำให้ข้อมูลสูญเสียความหมายในตัวเองไป นอกจากนี้ยังทำให้เสียเวลาค่อนข้างมากในการแปลงและกำหนดสัญลักษณ์เพื่อแทนข้อมูลเหล่านั้น ในกรณีที่ข้อมูลมีการกระจายตัวกันมากด้วย

Text Processing Kohonen Neural Networks เป็นนิวโรลเน็ตเวิร์คที่ขยายความสามารถของ Self-Organizing Feature Maps ขึ้นเพื่อการทำ clustering ข้อมูลที่มีลักษณะเป็นข้อความได้โดยตรง โดยไม่ต้องผ่านกระบวนการในการแปลงรูปข้อมูล โดยประยุกต์แนวคิดเรื่องการเปรียบเทียบความแตกต่างของข้อมูลแบบ symbolic เข้าไปในส่วนของการทำ competition และ Synaptic Adaptation ของนิวโรลเน็ตเวิร์คแบบ Self-Organizing Feature Maps. นิวโรลเน็ตเวิร์คใหม่สามารถรับข้อมูลที่เป็นข้อความได้โดยตรงและทำการ clustering ข้อมูลที่คุณสมบัติมีค่าเป็นข้อความได้เป็นอย่างดี

Clusterings are methods to group a set of objects into clusters such that objects within the same cluster have a high degree of similarity, while objects belonging to different clusters have a high degree of dissimilarity. Several clustering techniques for objects whose feature values are numerical values are well known, Several technique such as k-mean algorithm, nearest neighbor, decision trees and neural networks are proposed for clustering such objects. Recently, clustering problems are extends for document clustering. To cluster documents by using conventional techniques each document has to be mapped onto some representations that have quantitative features. This mapping use several time in processing and may cause data to loose their real meaning.

The Text Processing Kohonen Neural Network is extened ability of Self-Organizing Feature Maps. The Text Processing Kohonen Neural Network works directly on textual information without mapping documents onto some representation that has quantitative features and can assigns cluster labels to the objects. By modifying competition and Synaptic Adaptation process of the proposed neural network with the concept of dissimilarity measure for symbolic objects, The proposed neural network can directly receive a qualitative value without mapping the qualitative value into numerical value and can cluster the objects.