

192944

เป็นที่ยอมรับโดยทั่วไปว่าข่ายงานประสาทเทียมเป็นเครื่องมือจำแนกประเภทข้อมูลที่มีความถูกต้องยอดเยี่ยมมาก และสามารถนำไปประยุกต์ใช้ได้ดีในสาขาวิชาต่างๆ แต่ในกรณีที่ข้อมูลฝึกในปริมาณมากเกินไป พบว่าโปรแกรมฝึกข่ายงานประสาทบางตัวไม่สามารถเรียนรู้ข้อมูลได้ เนื่องจากโปรแกรมเกิดการล้มเหลวขึ้นเสียก่อน และถึงแม้ว่าจะเรียนรู้ข้อมูลได้สำเร็จ ก็ยังต้องใช้เวลารเรียนรู้ข้อมูลนานมาก วิทยานิพนธ์ฉบับนี้เสนอวิธีการลดข้อมูลแบบสุ่มตามความหนาแน่น โดยใช้แนวคิดการลดข้อมูลที่ซ้ำซ้อน และวิธีการลดข้อมูลแบบระเบิดลูกโซ่ที่เจาะจงเลือกเก็บเฉพาะข้อมูลตามขอบของข้อมูลแต่ละประเภท ซึ่งผลที่ได้จากการลดข้อมูลนี้ ไม่เพียงแต่จะช่วยลดโอกาสที่โปรแกรมจะเกิดการล้มเหลวในขั้นตอนการฝึกเท่านั้น แต่ยังช่วยลดเวลาในการฝึกให้น้อยลงอีกด้วย นอกจากนี้ผู้วิจัยยังได้นำอัลกอริทึมสำหรับลดจำนวนคุณสมบัติข้อมูลที่มีชื่อว่า อัลกอริทึม Relief มาช่วยในการลดข้อมูล ทำให้การลดข้อมูลแบบระเบิดลูกโซ่สามารถลดข้อมูลได้มากขึ้น ผลการทดลองแสดงให้เห็นว่าข้อมูลที่ลดด้วยวิธีที่เสนอนี้ ให้ความถูกต้องในการจำแนกประเภทโดยเฉลี่ยดีกว่าการลดด้วยวิธีการสุ่มแบบทั่วไป

192944

It has been known that the artificial neural network is a very accurate classifier which is practically used in various fields. Unfortunately, in case of a huge training dataset, we found that some neural network software crashed in the learning process. Even though the software succeeds in learning, it may take a very long time in the learning process. In this paper, we propose a data reduction method, called density-based random data reduction, which uses the concept of competence preservation. We also propose a data reduction method, called Chain-Explosion (CE) data reduction, which keeps only surrounding data of each class. Moreover, by using ReliefF which is an algorithm for dimension reduction, CE data reduction is further enhanced and is able to reduce more data. The experimental results show that the data reduced by the proposed methods give the average classification accuracy better than the data reduced by random reduction.