

ชื่อเรื่องวิทยานิพนธ์	การเปรียบเทียบเทคนิคการลดมิติสำหรับการจำแนก
ผู้เขียน	เอกสารบน โคร่งข่ายประสาท นายศุภชัย มุกดาสนิท
ปริญญา	วิทยาศาสตร์มหาบัณฑิต (วิทยาการคอมพิวเตอร์)
อาจารย์ที่ปรึกษาวิทยานิพนธ์	ผู้ช่วยศาสตราจารย์ ดร.เสมอแหะ สมหอม
	บทคัดย่อ

การจำแนกเอกสารมีความสำคัญเนื่องจากจำนวนเอกสารที่มากมายและ การจัดการเอกสารที่ยุ่งยาก งานวิจัยนี้จึงได้เสนอกระบวนการการจำแนกเอกสาร โดยใช้โครงข่ายประสาท ซึ่งเป็นที่ทราบกันดีว่าโครงข่ายประสาทมีความสามารถสูงในการจำแนกข้อมูลในงานด้านต่างๆรวมถึงการจำแนกเอกสารด้วย แม้ว่าโครงข่ายประสาทมีความสามารถในการเรียนรู้และจัดการความซับซ้อน แต่การลดความซับซ้อนของกรคำนวณของโครงข่ายประสาทโดยการจัดการกับจำนวนของค่าสำคัญในเอกสาร ที่มีปริมาณมากก็ยังคงมีความจำเป็นซึ่งจะส่งผลถึงประสิทธิภาพของการจำแนกเอกสาร งานวิจัยนี้มีการนำเทคนิคการลดมิติสามเทคนิคประกอบด้วยพีแอลเอส ดัชนีความหมายภายใน และรีเลแวนซ์สก็อร์ มาใช้ในการศึกษาเปรียบเทียบและวัดประสิทธิภาพในการใช้การลดมิติแบบต่างๆกับการจำแนกเอกสาร ข้อมูลตัวอย่างที่ใช้ในการสร้างตัวจำแนกและใช้ทดสอบประสิทธิภาพคือเอกสารชีวสารสนเทศ ผลที่ได้คือ การลดมิติด้วยวิธีดัชนีความหมายภายใน และพีแอลเอสส่งผลให้การจำแนกเอกสารมีความถูกต้องที่ใกล้เคียงกัน ซึ่งสูงกว่าวิธีรีเลแวนซ์สก็อร์ ด้านเวลาการเรียนรู้พีแอลเอส สามารถลดเวลาการเรียนรู้ของโครงข่ายประสาทได้ดีที่สุดรองลงมาคือแอลเอสไอ และสุดท้ายคือรีเลแวนซ์สก็อร์

<b>Thesis Title</b>	A Comparison of Dimensionality Reduction Techniques for Document Classification on Neural Network
<b>Author</b>	Mr. Supachai Mukdasanit
<b>Degree</b>	Master of Science (Computer Science)
<b>Thesis Advisor</b>	Assistant Professor Dr. Samerkae Somhom

### ABSTRACT

The document classification is a necessity due to the very large amount of documents and the complexity management of documents. This research presents the document classification process using neural network. The neural network has been recognized as one of the most successful classification methods for many applications including document classification. Even though the learning ability and computational complexity of training in neural network, reducing computational complexity is an essential issue to efficiently handle a large number of terms in practical applications of document classification. This research provides a comparison study of three dimension reduction techniques, namely Partial Least Squares (PLS), Latent Semantic Indexing (LSI) and Relevancy Score (RE), and evaluates the relative performance of classification procedures incorporating those methods. The sample documents are the bio-information documents to create classifier and evaluate performance. The results show that LSI and PLS can improve accuracy higher than RE for classifying by neural network. In the training time, the PLS can improve the training time better of NN than LSI and RE respectively.