

วิทยานิพนธ์นี้นำเสนอการแก้ไขปัญหาคำที่มีความหมายกำกวมในภาษาไทย โดยนำเสนอวิธีการเรียนรู้ที่เรียกว่าซัพพอร์ตเวกเตอร์แมชชีน ซึ่งเป็นอัลกอริทึมในการหาระนาบของการแบ่งข้อมูลที่เหมาะสมที่สุด นำมาใช้ในการเรียนรู้เพื่อแก้ปัญหาคำกำกวมของคำ โดยใช้วิธีการเลือกลักษณะ (Feature) 3 แบบ คือคำที่ปรากฏรอบข้างคำกำกวม (Word), คำและหน้าที่ของคำที่ปรากฏรอบข้างคำกำกวม (Word and Part of Speech), คู่ของคำและหน้าที่ของคำที่อยู่ติดกันรอบข้างคำกำกวม (Collocation of Word and Part of Speech) และมีค่าพารามิเตอร์ที่ทำการปรับคือ จำนวนระยะห่างจากคำกำกวมที่แตกต่างกัน (Window size) โดยทดสอบกับคำที่มีความหมายกำกวมในภาษาไทยจำนวน 10 คำ และทำการเปรียบเทียบการเรียนรู้ของเครื่องทั้งหมด 4 แบบคือ SVM(Support Vector Machine), SNOW(Sparse Network of Winnow), Naive Bayes, Neural Network ผลการทดสอบปรากฏว่าการใช้ซัพพอร์ตเวกเตอร์แมชชีนให้ค่าความถูกต้องมากที่สุด

This thesis represents the solution of Word Sense Disambiguation in Thai. The machine learning, called "Support Vector Machine", is the algorithm for finding the best hyperplane to classify data. "Word", "Word and Part of Speech" and "Collocation of Word and Part of Speech" are the data features which have been used to solve the ambiguous words with adjusting "window size" parameter. We choose ten ambiguous words and compare with four types of learning machine algorithms, which are "SVM (Support Vector Machine)", "Snow (Sparse Network of Winnow)", "Naive Bayes" and "Neural Network". Support Vector Machine provides the highest accuracy.