

**COMBINING MACHINE-LEARNING AND SEMANTIC-ORIENTATION
APPROACHES FOR SENTIMENT CLASSIFICATION**

TITIMA KASEMSRITANAWAT 5236782 EGTI/M

M.SC. (TECHNOLOGY OF INFORMATION SYSTEM MANAGEMENT)

**THESIS ADVISORY COMMITTEE: TANASANEE PHIENHTRAKUL, Ph.D.,
MINGMANAS SIVARAKSA, Ph.D., WARANYU WONGSEREE, Ph.D.**

ABSTRACT

This study proposed using new extracted features for sentiment classification. These new features were derived from SentiWordNet and they are called “sentiment features”. These features are described by the strength of positivity, negativity, and objectivity for each part-of-speech. Support vector machines (SVM), multinomial Naïve Bayes, and Decision Tree (J48) were used to classify the reviews on the extracted features. The set of proposed features and bag-of-words features were then evaluated on movie reviews. The experimental results showed that using a combination of bag-of-words and sentiment features gave the highest accuracy when SVM were applied. Furthermore, the accuracies of both SVM and multinomial Naïve Bayes on sentiment features were higher than the accuracy of the lexicon analysis in review classification.

**KEY WORDS: SENTIMENT CLASSIFICATION / SENTIWORDNET /
MOVIE REVIEWS / FEATURE EXTRACTION**

126 pages

การรวมวิธีการเรียนรู้ด้วยเครื่องและความหมายทางภาษาเพื่อจำแนกความคิดเห็นตามความรู้สึก
COMBINING MACHINE-LEARNING AND SEMANTIC-ORIENTATION APPROACHES
FOR SENTIMENT CLASSIFICATION

ฐิติมา เกษมศรีธนาวัฒน์ 5236782 EGTI/M

ว.ทม. (เทคโนโลยีการจัดการระบบสารสนเทศ)

คณะกรรมการที่ปรึกษาวิทยานิพนธ์: ธนัสินี เพ็ชรตระกูล, Ph.D., มิ่งมานัส ศิวรักษ์, Ph.D.,
วรัญญู วงษ์เสรี, Ph.D.

บทคัดย่อ

งานวิจัยนี้นำเสนอการสร้างคุณลักษณะใหม่ขึ้นมาเพื่อใช้ในการจำแนกความคิดเห็นตามความรู้สึก คุณลักษณะแบบใหม่นี้ได้มาจากเซนติเวิร์ดเน็ตและถูกเรียกว่า คุณลักษณะตามความรู้สึก คุณลักษณะนี้จะอธิบายถึงความแรงของความเป็นบวก ความเป็นลบ และความเป็นกลางของแต่ละประเภทของคำศัพท์ ซัพพอร์ตเวกเตอร์แมชชีน, เบย์อย่างง่าย (มัลติโนเมียล) และ ต้นไม้ตัดสินใจ (เจ 48) ถูกใช้เพื่อจำแนกความคิดเห็นบนคุณลักษณะตามความรู้สึก ทั้งคุณลักษณะใหม่นี้และคุณลักษณะโดยดุษคำศัพท์ได้ถูกประเมินความถูกต้องบนความคิดเห็นทางภาพยนตร์ ผลการทดลองได้แสดงให้เห็นว่าเมื่อใช้คุณลักษณะที่ได้จากการรวมกันของคุณลักษณะโดยดุษคำศัพท์ กับคุณลักษณะตามความรู้สึกให้ความแม่นยำสูงสุดร่วมกับการใช้ ซัพพอร์ตเวกเตอร์แมชชีน นอกจากนี้ความแม่นยำในการจำแนกความคิดเห็นโดยใช้ ซัพพอร์ตเวกเตอร์แมชชีน และเบย์อย่างง่าย (มัลติโนเมียล) บนคุณลักษณะตามความรู้สึกให้ความแม่นยำในการจำแนกความคิดเห็น มากกว่าการใช้การวิเคราะห์เชิงความหมายของคำศัพท์ด้วยพจนานุกรม

126 หน้า