

บทคัดย่อ

ปัจจุบันคนไทยนิยมติดตอสื่อสารกันผ่านเครือข่ายสังคมออนไลน์เป็นจำนวนมาก โดยเฉพาะเครือข่ายสังคมออนไลน์เฟสบุ๊ก เนื่องจากมีผู้คนนิยมติดตอสื่อสารกันผ่านเฟสบุ๊กจำนวนมาก ทำให้มีการกระจายข้อความประสงค์ร้ายให้กับผู้ใช้ผู้อื่นบนเฟสบุ๊ก บางข้อความนำไปสู่การหลอกลวง ซึ่งสร้างความเสียหายให้กับผู้ที่ถูกหลอกเป็นอย่างมาก ดังนั้นงานวิจัยนี้จึงทำการศึกษาข้อความภาษาไทยที่นำไปสู่การหลอกลวงบนเฟสบุ๊ก โดยทำการศึกษาข้อความหลอกลวงเกี่ยวกับการรับสมัครงานซึ่งเป็นการหลอกลวงที่คนไทยตกเป็นเหยื่อมากที่สุด ทำการศึกษาขั้นตอนวิธีการที่หลากหลายเพื่อใช้ในการตรวจสอบข้อความหลอกลวง และนำเสนอวิธีการที่ใช้ในการค้นหาคำเด่นในข้อความหลอกลวง ชุดข้อมูลที่ใช้ในการทดลองในงานวิจัยนี้เป็นข้อมูลที่อยู่บนเพจเฟสบุ๊ก ทำการดึงข้อมูลในเพจโดยใช้ Graph API ร่วมกับการเขียนโปรแกรม PHP จากนั้นทำการสกัดคุณลักษณะเพื่อใช้ในการจำแนกข้อความหลอกลวง โดยทำการสกัดคุณลักษณะออกมา 2 แบบ คือ คุณลักษณะที่สกัดจากบริบทของข้อความหรือคำที่อยู่ในข้อความ (Context-based features) และคุณลักษณะที่สกัดจากเนื้อหาในข้อความ (Content-based features) จากนั้นทำการประยุกต์ใช้ขั้นตอนวิธี SVM และ KNN เพื่อสร้างตัวจำแนกที่ใช้ในการจำแนกข้อความที่นำไปสู่การหลอกลวง งานวิจัยนี้ได้ทำการทดลองเพื่อศึกษาคุณลักษณะที่สกัดจากบริบท คุณลักษณะที่สกัดจากเนื้อหาในข้อความ และคุณลักษณะที่สกัดจากทั้งบริบทและเนื้อหาในข้อความ เพื่อใช้ในการตรวจสอบข้อความหลอกลวง จากผลการทดลองแสดงให้เห็นว่าคุณลักษณะที่สกัดจากบริบทของข้อความ สามารถตรวจสอบข้อความที่นำไปสู่การหลอกลวงได้อย่างมีประสิทธิภาพ และให้ค่าอัตราการเรียนรู้ (F-measure) ถึง 99% เมื่อใช้ตัวจำแนกที่สร้างขึ้นจากขั้นตอนวิธี SVM นอกจากนี้งานวิจัยนี้ได้ทำการศึกษาคำเด่นในข้อความหลอกลวงซึ่งพบว่า ข้อความหลอกลวงเกี่ยวกับการทำงาน ส่วนใหญ่มักจะกล่าวถึงเกี่ยวกับรายได้ และการทำงานที่บ้าน ซึ่งแตกต่างจากข้อความปกติเกี่ยวกับการสมัครงานที่จะเจาะจงตำแหน่งที่สมัครงานอย่างชัดเจน และจำนวนไลค์ จำนวนแชร์ จำนวนคอมเมนต์ของข้อความหลอกลวงมีจำนวนน้อยมากหรือไม่เลย

คำสำคัญ: เหมืองความคิดเห็น, การวิเคราะห์ความรู้สึก, เหมืองข้อความ, เฟสบุ๊ก, เหมืองข้อมูล