

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

แบบสอบถาม (Questionnaire) เป็นเครื่องมือวิจัยชนิดหนึ่งที่นิยมใช้กันมากในหมู่นักวิจัย ทั้งนี้เพราะการเก็บรวบรวมข้อมูลด้วยแบบสอบถามเป็นวิธีที่สะดวกและสามารถใช้วัดได้อย่างกว้างขวาง แบบสอบถามส่วนใหญ่จะอยู่ในรูปของคำถามเป็นชุด ๆ ที่ได้ถูกรวบรวมไว้อย่างมีหลักเกณฑ์และเป็นระบบ เพื่อใช้วัดสิ่งที่ผู้วิจัยต้องการจะวัดจากกลุ่มตัวอย่างหรือประชากร เป้าหมายให้ได้มาซึ่งข้อเท็จจริงทั้งในอดีต ปัจจุบันและการคาดคะเนเหตุการณ์ในอนาคต การเก็บข้อมูลด้วยแบบสอบถามสามารถทำได้ด้วยการสัมภาษณ์หรือให้ผู้ตอบตอบด้วยตนเอง (มารยาท โยทองยศ, 2554) ข้อคำถามในแบบสอบถามอาจแบ่งได้เป็น 2 ประเภท คือ คำถามปลายเปิด (Open Ended Question) และคำถามปลายปิด (Closed Question) ซึ่งคำถามปลายเปิดเป็นคำถามที่เปิดโอกาสให้ผู้ตอบสามารถตอบได้อย่างไม่จำกัดวงคำตอบ ซึ่งคาดว่าจะได้คำตอบที่เป็นอิสระ ไม่แน่นอน ผู้ตอบสามารถตอบได้ตามสภาพความเป็นจริงมากกว่าคำตอบที่จำกัดวงให้ตอบ คำถามปลายเปิดจะนิยมใช้กันมากในกรณีที่ผู้วิจัยไม่สามารถคาดเดาได้ล่วงหน้าว่าคำตอบจะเป็นอย่างไร หรือใช้คำถามปลายเปิดในกรณีที่ต้องการได้คำตอบเพื่อนำมาเป็นแนวทางในการสร้างคำถามปลายปิด แบบสอบถามแบบนี้มีข้อเสีย คือ การรวบรวมความคิดเห็น และการแปลผลมีความยุ่งยาก สำหรับคำถามปลายปิดเป็นคำถามที่ผู้วิจัยมีแนวคำตอบไว้ให้ผู้ตอบเลือกตอบจากคำตอบที่กำหนดไว้เท่านั้น คำตอบที่ผู้วิจัยกำหนดไว้ล่วงหน้ามักได้มาจากการทดลองใช้คำถามในลักษณะที่เป็นคำถามปลายเปิด หรือการศึกษากรอบแนวความคิด สมมติฐานการวิจัย และนิยามเชิงปฏิบัติการ คำถามปลายปิดมีวิธีการเขียนได้หลาย ๆ แบบ เช่น แบบให้เลือกตอบอย่างใดอย่างหนึ่ง แบบให้เลือกคำตอบที่ถูกต้องเพียงคำตอบเดียว แบบผู้ตอบจัดลำดับความสำคัญหรือแบบให้เลือกคำตอบหลายคำตอบ (พิชิต ฤทธิ์จรูญ, 2544) การวิเคราะห์ผลทางสถิติในส่วนข้อคิดเห็นหรือข้อเสนอแนะของผู้ตอบแบบสอบถามแบบปลายเปิด เป็นเรื่องยากและต้องใช้เวลามาก โดยเฉพาะในกรณีที่มีการเก็บข้อมูลแบบสำรวจเป็นจำนวนมาก ทำให้ผู้วิเคราะห์ผลไม่นำข้อมูลดังกล่าวมาประมวลผล ซึ่งข้อมูลดังกล่าวมีความสำคัญและเป็นประโยชน์ต่อการวิจัย จึงได้ประยุกต์ใช้การจำแนกหมวดหมู่เอกสารแบบอัตโนมัติ (Automatic Text Classification) กับข้อความภาษาไทยที่เป็นข้อคิดเห็นในแบบสอบถามปลายเปิด

การจำแนกหมวดหมู่เอกสารแบบอัตโนมัติเป็นการนำวิธีการเรียนรู้ด้วยคอมพิวเตอร์ (Machine Learning) ประยุกต์ร่วมกับการประมวลผลภาษาธรรมชาติ การจัดแบ่งกลุ่มเอกสารแบบอัตโนมัติ สามารถแบ่งได้ 2 ลักษณะ คือ การจัดกลุ่ม (Clustering) และการจำแนกหมวดหมู่ (Classification หรือ Categorization) การจัดกลุ่มเอกสาร คือ การแบ่งกลุ่มตามเนื้อหาของเอกสาร โดยไม่มีการกำหนดกลุ่มหรือหมวดหมู่ของเอกสารไว้ก่อน ซึ่งจะเป็นการแบ่งกลุ่มตามลักษณะของเอกสาร โดยเอกสารที่มีลักษณะเหมือนกันจะอยู่ด้วยกัน ส่วนการจำแนกหมวดหมู่เอกสาร คือ การแบ่งกลุ่มตามเนื้อหาของเอกสาร โดยที่มีการกำหนดกลุ่มหรือหมวดหมู่ของเอกสารไว้ก่อน โดยจะเปรียบเทียบเอกสารกับต้นแบบในแต่ละหมวดหมู่เอกสารจะถูกจัดอยู่ในหมวดหมู่ที่ต้นแบบมีลักษณะคล้ายกับตัวมันเองมากที่สุด โดยผลลัพธ์ที่ได้จากวิธีการเรียนรู้ด้วยคอมพิวเตอร์นั้น ความถูกต้องใกล้เคียงกับผลการจำแนกหมวดหมู่ของเอกสารที่ทำโดยมนุษย์ ทำให้ประหยัดแรงงานมนุษย์เป็นอย่างมากเพราะไม่ต้องอาศัยผู้เชี่ยวชาญในการจำแนกประเภทเอกสารหรือปรับเปลี่ยนหมวดหมู่ของเอกสาร (Sebastiani, 2002: 1-47) สำหรับการจำแนกหมวดหมู่เอกสารภาษาไทยแบบอัตโนมัติ มีงานวิจัยหลายงานที่ได้ศึกษาทดลองจำแนกหมวดหมู่เอกสารภาษาไทยโดยใช้อัลกอริทึมในการเรียนรู้ที่แตกต่างกัน โดยอัลกอริทึมหรือวิธีการที่มีการนำมาใช้กับเอกสารภาษาไทยที่มีประสิทธิภาพดี ได้แก่ เนอ์ฟเบย์ (Naïve-Bayes) (Kruengkrai and Jaruskulchai, 2001) และซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) (วัลลภ อินทร์น้ำ, 2548)

จากปัญหาดังกล่าว จึงได้ทำการจำแนกหมวดหมู่ข้อความแสดงข้อคิดเห็นในแบบสอบถามปลายเปิดโดยอัตโนมัติ เพื่อใช้สำหรับการวิเคราะห์ผลทางสถิติ โดยใช้วิธีเนอ์ฟเบย์เปรียบเทียบประสิทธิภาพกับซัพพอร์ตเวกเตอร์แมชชีน

1.2 วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อจำแนกหมวดหมู่ข้อความของข้อคิดเห็นภาษาไทยในแบบสอบถามปลายเปิด

1.2.2 เพื่อทดสอบประสิทธิภาพการจำแนกหมวดหมู่ของข้อคิดเห็นภาษาไทยในแบบสอบถามปลายเปิด ด้วยวิธีเนอ์ฟเบย์และซัพพอร์ตเวกเตอร์แมชชีน

1.3 สมมติฐานการวิจัย

การจำแนกหมวดหมู่ของข้อคิดเห็นในแบบสอบถามปลายเปิด โดยใช้วิธีซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพดีกว่าเนอ์ฟเบย์

1.4 ขอบเขตการวิจัย

1.4.1 วิธีที่ใช้ในการจำแนกหมวดหมู่ของข้อคิดเห็นในแบบสอบถามปลายเปิด คือ เนอ็ฟเบย์ และซัพพอร์ทเวกเตอร์แมชชีน

1.4.2 ข้อมูลที่ใช้ในงานวิจัยเป็นข้อมูลที่ได้จากแบบสอบถาม เรื่องความคิดเห็นที่มีต่อสถาบันการศึกษาไทยระหว่างภาครัฐและเอกชนของประชาชนในเขตกรุงเทพมหานคร

1.4.3 การเรียนรู้แบ่งข้อมูลออกเป็น 2 ส่วน คือ ข้อมูลสำหรับการเรียนรู้ 80% และข้อมูลสำหรับการทดสอบ 20%

1.4.4 การวิจัยได้แบ่งกลุ่มของการจำแนกหมวดหมู่ข้อความแสดงความคิดเห็น โดยจะแบ่ง 3 หมวดหมู่ ได้แก่ เชิงบวก (Positive) กลาง (Medium) และลบ (Negative)

1.5 คำจำกัดความที่ใช้ในการวิจัย

1.5.1 การจำแนกหมวดหมู่เอกสาร (Text Classification) หมายถึง เป็นกระบวนการในจัดกลุ่มเอกสารข้อความออกเป็นหมวดหมู่ ซึ่งกระบวนการจัดกลุ่มเอกสารนี้จะมีการกำหนดเป้าหมายชุดของเอกสารเอาไว้ล่วงหน้า

1.5.2 เนอ็ฟเบย์ (Naïve-Bayes) หมายถึง เทคนิคการจำแนกหมวดหมู่ที่ใช้หลักการคำนวณความน่าจะเป็น (กมลลาสน์ อุคมล้ำเลิศ และอานนท์ รุ่งสว่าง, 2553: 312)

1.5.3 ซัพพอร์ทเวกเตอร์แมชชีน (Support Vector Machine: SVM) หมายถึง เทคนิคการจำแนกหมวดหมู่ที่ใช้หลักการสร้างสมการเส้นตรงเพื่อแบ่งกลุ่มข้อมูล 2 กลุ่มออกจากกัน โดย SVM พยายามสร้างเส้นแบ่งตรงกึ่งกลางระหว่างกลุ่มให้มีระยะห่างระหว่างขอบเขตของทั้งสองกลุ่มให้มากที่สุด (พรพล ธรรมรงค์รัตน์, ลัดดา ปรีชาวิรุกุล และวิภาดา เวทย์ประสิทธิ์, 2008)

1.5.4 แบบสอบถามปลายเปิด (Open Ended Questionnaire) หมายถึง เป็นคำถามที่ไม่ได้กำหนดคำตอบไว้ และเปิดโอกาสให้ผู้ตอบมีอิสระในการตอบด้วยความคิดของตนเอง ข้อเสียของแบบสอบถามนี้คือ ตอบยากและเสียเวลาในการตอบมาก และต้องใช้การคิดวิเคราะห์อย่างกว้างขวาง (เขาวเรศ จันทะแสน, 2554)

1.5.5 ข้อมูลสำหรับการเรียนรู้ หมายถึง ข้อมูลที่ต้องใช้สำหรับนำข้อมูลเข้าสู่อัลกอริทึมหรือโมเดล เพื่อให้อัลกอริทึมหรือโมเดลเกิดการเรียนรู้และจดจำข้อมูลก่อนการนำข้อมูลมาทดสอบ

1.5.6 ข้อมูลสำหรับการทดสอบ หมายถึง ข้อมูลที่ใช้ในการทดสอบอัลกอริทึมหรือโมเดล เพื่อทดสอบความถูกต้องของการจำแนกข้อมูล

1.5.7 ข้อคิดเห็น หรือความคิดเห็น หมายถึง อารมณ์ ความรู้สึก ความคิด และข้อสันนิษฐานที่ผู้พูด หรือผู้เขียนมีต่อสิ่งใดสิ่งหนึ่ง

1.6 ประโยชน์ของงานวิจัย

1.6.1 เพื่อประยุกต์การจำแนกเอกสารแบบอัตโนมัติกับข้อความของการแสดงความคิดเห็นของผู้ตอบแบบสอบถาม ด้วยวิธีเอนีฟเบย์และซัพพอร์ตเวกเตอร์แมชชีน

1.6.2 ได้วิธีที่มีประสิทธิภาพในการวิเคราะห์ข้อความแสดงความคิดเห็นของผู้ตอบแบบสอบถามปลายเปิด

1.6.3 แก้ปัญหาความยุ่งยากในการวิเคราะห์ข้อมูลของการแสดงความคิดเห็นในแบบสอบถามปลายเปิด