สารบัญ

หน้า

บทคัดย่อ ..... ก
กิตติกรรมประกาศ ..... ข
สารบัญ ..... ค
สารบัญตาราง ..... ฉ
สารบัญภาพประกอบ ..... ช
บทที่ 1 บทนำ 1
1.1 หลักการและเหตุผล 1
1.2 วัตถุประสงค์ของโครงงาน 2
1.3 ขอบเขตของโครงงาน 2
1.4 ประโยชน์ที่คาดว่าจะได้รับ 3
1.5 อุปกรณ์และเครื่องมือที่ใช้ในการดำเนินงาน 3
1.6 แผนการดำเนินงาน 3
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 5
2.1 ข้อมูลที่ไม่สมดุล (Imbalanced Data) 5
2.2 การจำแนกความรู้สึก (Sentiment Classification) 6
2.3 เทคนิคและอัลกอริทึมที่เกี่ยวข้อง 6
2.3.1 การจำแนกหมวดหมู่เอกสาร (Text Classification) 6
2.3.2 ขั้นตอนการเตรียมเอกสาร (Document Pre-processing) 7
2.3.3 การสร้างตัวแทนเอกสาร (Document Representation) 9
2.3.4 การเลือกคุณลักษณะ (Feature Selection) 9
2.3.5 การให้น้ำหนักคำ (Term Weighting) 10
2.3.6 นาอีฟเบย์ (Naïve Bayes) 14
2.3.7 วิธีการค้นหาเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor: KNN) 15
2.3.8 โครงข่ายประสาทแบบคอนโวลูชัน (Convolutional Neural Network: CNN) 15
2.3.9 การประเมิน (Evaluation) 17
2.4 งานวิจัยที่เกี่ยวข้อง (Related work) 18

สารบัญ (ต่อ)

หน้า

บทที่ 3 วิธีดำเนินงานวิจัย....21

3.1 กรอบการดำเนินงาน....21
3.2 ชุดข้อมูล (Data set)....21
3.3 การสร้างโมเดลเพื่อการจำแนกความรู้สึกของบทวิจารณ์ (Classifier Modeling)....23
3.3.1 การเตรียมข้อมูลก่อนการประมวลผล....23
3.3.2 การสร้างโมเดลการจำแนกความรู้สึกของบทวิจารณ์....34
3.4 การวัดประสิทธิภาพของตัวจัดกลุ่มเอกสาร (Evaluation)....45
3.4.1 การนำโมเดลเพื่อการจำแนกกลุ่มของบทวิจารณ์ไปใช้....45
3.4.2 การวัดประสิทธิภาพของตัวจัดกลุ่มเอกสาร (Evaluation)....56
3.5 การปรับปรุงประสิทธิภาพโมเดลเพื่อการจำแนก....57
3.5.1 ปัญหาจากการทำ Lemmatization....57
3.5.2 ปัญหาด้านการใช้ภาษา....58
3.6 ตัวอย่างหน้าจอโปรแกรม....58

บทที่ 4 ผลการทดลอง....59

4.1 ข้อมูลที่ใช้ในการทดสอบ....59
4.2 Algorithm Setup....59
4.2.1 KNN Setup....59
4.2.2 Naïve Bayes....60
4.2.3 CNN Setup....60
4.3 ผลการทดลอง (Results)....61
4.3.1 การทดสอบโมเดลในการจำแนกบทวิจารณ์โดยอัลกอริทึม KNN....61
4.3.2 การทดสอบโมเดลในการจำแนกบทวิจารณ์โดยอัลกอริทึม Naïve Bayes....64
4.3.3 การทดสอบโมเดลในการจำแนกบทวิจารณ์โดยอัลกอริทึม CNN....67
4.3.4 ภาพรวมผลการทดลอง....70
4.4 การทดสอบการจำแนกบทวิจารณ์ที่มีข้อมูลที่ต่างกัน 3 ชุดข้อมูลในทุกสัดส่วน....70
4.4.1 ทดสอบโมเดลกับ 3 สัดส่วนด้วยข้อมูล 3 ชุดที่ต่างกันกับอัลกอริทึม KNN....71

สารบัญ (ต่อ)

หน้า

4.4.2 ทดสอบโมเดลกับ 3 สัดส่วนด้วยข้อมูล 3 ชุดที่ต่างกันกับอัลกอริทึมนาอีฟเบย์ 74
4.4.3 ทดสอบโมเดลกับ 3 สัดส่วนด้วยข้อมูล 3 ชุดที่ต่างกันกับ CNN 77
4.5 การวิเคราะห์ผล 80
บทที่ 5 สรุปและอภิปรายผลการทดลอง 84
5.1 สรุปผลและอภิปรายผล 84
5.2 ปัญหาและอุปสรรคในการดำเนินงาน 85
5.2.1 ปัญหาเกี่ยวกับอัลกอริทึมในการสร้างโมเดล 85
5.2.2 ปัญหาเกี่ยวกับชุดข้อมูลที่ใช้ในการสร้างโมเดล 85
5.3 ข้อเสนอแนะ 86
เอกสารอ้างอิง 87
ภาคผนวก 90
ภาคผนวก ก คู่มือการติดตั้ง 91
ภาคผนวก ข คู่มือการใช้งาน 101
บทความวิจัย 105
โปสเตอร์โครงงาน 121
ประวัติผู้จัดทำโครงงาน 123

สารบัญตาราง

หน้า

ตารางที่ 1.1 แผนการดำเนินงาน 4
ตารางที่ 2.1 แสดงการตัดคำหยุด 7
ตารางที่ 2.2 สัญลักษณ์สำหรับ Supervised Term Weighting (STW)....11
ตารางที่ 2.3 ตัวอย่างการแจกแจงเอกสารสองเทอม 13
ตารางที่ 2.4 ผลลัพธ์การคำนวณการกระจายน้ำหนัก 14
ตารางที่ 3.1 แสดงการนำเสนอความสัมพันธ์ระหว่างคำและเอกสาร 25
ตารางที่ 3.2 *BOW* แสดงค่าและน้ำหนักค่าในแต่ละเอกสารด้วยการให้น้ำหนักแบบ *tf-idf* 29
ตารางที่ 3.3 *BOW* แสดงค่าและน้ำหนักค่าในแต่ละเอกสารด้วยการให้น้ำหนักแบบ *Delta TF-IDF* 30
ตารางที่ 3.4 *BOW* แสดงค่าและน้ำหนักค่าในแต่ละเอกสารด้วยการให้น้ำหนักแบบ *TF-IDF-ICF* 32
ตารางที่ 3.5 *BOW* แสดงค่าและน้ำหนักค่าในแต่ละเอกสารด้วยการให้น้ำหนักแบบ *TF-RF* 33
ตารางที่ 3.6 *BOW* แสดงค่าและน้ำหนักค่าในแต่ละเอกสารด้วยการให้น้ำหนักแบบ *TF-IGM* 34
ตารางที่ 3.7 โมเดลการจำแนกความรู้สึกของบทวิจารณ์สินค้าอิเล็กทรอนิกส์ด้วย Naïve Bayes โดยใช้การให้น้ำหนักคำแบบ *tf-idf* 37
ตารางที่ 3.8 โมเดลการจำแนกความรู้สึกของบทวิจารณ์สินค้าอิเล็กทรอนิกส์ด้วย Naïve Bayes โดยใช้การให้น้ำหนักคำแบบ *Delta TF-IDF* 38
ตารางที่ 3.9 โมเดลการจำแนกความรู้สึกของบทวิจารณ์สินค้าอิเล็กทรอนิกส์ด้วย Naïve Bayes โดยใช้การให้น้ำหนักคำแบบ *TF-IDF-ICF* 39
ตารางที่ 3.10 โมเดลการจำแนกความรู้สึกของบทวิจารณ์สินค้าอิเล็กทรอนิกส์ด้วย Naïve Bayes โดยใช้การให้น้ำหนักคำแบบ *TF-RF* 40
ตารางที่ 3.11 โมเดลการจำแนกความรู้สึกของบทวิจารณ์สินค้าอิเล็กทรอนิกส์ด้วย Naïve Bayes โดยใช้การให้น้ำหนักคำแบบ *TF-IGM* 41
ตารางที่ 3.12 โมเดลวิเคราะห์ระดับคะแนนบทวิจารณ์ด้วย *KNN* โดยการให้น้ำหนักคำด้วย *tf-idf* 42
ตารางที่ 3.13 โมเดลวิเคราะห์ระดับคะแนนบทวิจารณ์ด้วย *KNN* การให้น้ำหนักคำด้วย *Delta TF-IDF* 42
ตารางที่ 3.14 โมเดลวิเคราะห์ระดับคะแนนบทวิจารณ์ด้วย *KNN* โดยการให้น้ำหนักคำด้วย *TF-IDF-ICF* 43
ตารางที่ 3.15 โมเดลวิเคราะห์ระดับคะแนนบทวิจารณ์ด้วย *KNN* โดยการให้น้ำหนักคำด้วย *TF-RF* 43
ตารางที่ 3.16 โมเดลวิเคราะห์ระดับคะแนนบทวิจารณ์ด้วย *KNN* โดยการให้น้ำหนักคำด้วย *TF-IGM* ..43
ตารางที่ 3.17 แสดงคำสำคัญที่ได้หลังจากผ่านกระบวนการ pre-processing ในการทดสอบ NV....46

สารบัญตาราง (ต่อ)

หน้า

ตารางที่ 3.18 คำสำคัญที่ได้หลังจากผ่านกระบวนการ pre-processing ในการทดสอบ TF-IDF.....49
ตารางที่ 3.19 คำสำคัญที่ได้หลังจากผ่านกระบวนการ pre-processing ในการทดสอบ Delta TF-IDF 50
ตารางที่ 3.20 คำสำคัญที่ได้หลังจากผ่านกระบวนการ pre-processing ในการทดสอบ TF-IDF-ICF ..52
ตารางที่ 3.21 คำสำคัญที่ได้หลังจากผ่านกระบวนการ pre-processing ในการทดสอบ TF-RF 53
ตารางที่ 3.22 คำสำคัญที่ได้หลังจากผ่านกระบวนการ pre-processing ในการทดสอบ TF-IGM.....55
ตารางที่ 3.23 ตัวอย่าง Confusion Matrix.....56
ตารางที่ 4.1 ตารางการทดสอบประสิทธิภาพของค่า *k* 60
ตารางที่ 4.2 ค่าเฉลี่ยในการทดลองค่า input ในการทดสอบกับอัลกอริทึม *CNN* 60
ตารางที่ 4.3 ผลการทดสอบด้วยอัลกอริทึม *KNN* 62
ตารางที่ 4.4 ผลการทดสอบด้วยอัลกอริทึม *Naïve Bayes*.....65
ตารางที่ 4.5 ผลการทดสอบด้วยอัลกอริทึม *CNN* 68
ตารางที่ 4.6 ตารางค่าเฉลี่ย *F-measure* การให้น้ำหนักร่วมกับอัลกอริทึม 70
ตารางที่ 4.7 ทดสอบโมเดลที่มีสัดส่วน 100 : 10 กับข้อมูล 3 ชุดที่ต่างกันกับทุกอัลกอริทึม *KNN* 72
ตารางที่ 4.8 ทดสอบโมเดลที่มีสัดส่วน 100 : 20 กับข้อมูล 3 ชุดที่ต่างกันกับทุกอัลกอริทึม *KNN* 72
ตารางที่ 4.9 ทดสอบโมเดลที่มีสัดส่วน 100 : 30 กับข้อมูล 3 ชุดที่ต่างกันกับทุกอัลกอริทึม *KNN* 73
ตารางที่ 4.10 ทดสอบโมเดลที่มีสัดส่วน 100 : 10 กับข้อมูล 3 ชุดที่ต่างกันกับทุกอัลกอริทึมนาอีฟเบย์ 75
ตารางที่ 4.11 ทดสอบโมเดลที่มีสัดส่วน 100 : 20 กับข้อมูล 3 ชุดที่ต่างกันกับทุกอัลกอริทึมนาอีฟเบย์ 75
ตารางที่ 4.12 ทดสอบโมเดลที่มีสัดส่วน 100 : 30 กับข้อมูล 3 ชุดที่ต่างกันกับทุกอัลกอริทึมนาอีฟเบย์ 76
ตารางที่ 4.13 ทดสอบโมเดลที่มีสัดส่วน 100 : 10 กับข้อมูล 3 ชุดที่ต่างกันกับทุกอัลกอริทึม *CNN*.....78
ตารางที่ 4.14 ทดสอบโมเดลที่มีสัดส่วน 100 : 20 กับข้อมูล 3 ชุดที่ต่างกันกับทุกอัลกอริทึม *CNN*.....78
ตารางที่ 4.15 ทดสอบโมเดลที่มีสัดส่วน 100 : 30 กับข้อมูล 3 ชุดที่ต่างกันกับทุกอัลกอริทึม *CNN*.....79

สารบัญภาพประกอบ

หน้า

ภาพประกอบที่ 2.1 Bag of words....9
ภาพประกอบที่ 2.2 ตัวอย่างการคำนวณ Convolution....16
ภาพประกอบที่ 2.3 ตัวอย่างการทำ Pooling layer....17
ภาพประกอบที่ 2.4 ตาราง Confusion Matrix....17
ภาพประกอบที่ 3.1 กรอบการดำเนินงานของระบบ....21
ภาพประกอบที่ 3.2 Data Collection....21
ภาพประกอบที่ 3.3 ตัวอย่างเอกสารข้อความแสดงความคิดเห็น....22
ภาพประกอบที่ 3.4 ตัวอย่างเอกสารที่อยู่ในรูปแบบ XML....22
ภาพประกอบที่ 3.5 Classifier Modeling....23
ภาพประกอบที่ 3.6 ตัวอย่าง Unknown word....587
ภาพประกอบที่ 3.7 ตัวอย่างหน้าจอโปรแกรม....588
ภาพประกอบที่ 4.1 ตัวอย่างบทวิจารณ์สินค้าอิเล็กทรอนิกส์ที่ใช้ในการทดสอบ....599
ภาพประกอบที่ 4.2 กราฟค่าเฉลี่ย F-measure การให้น้ำหนักร่วมกับอัลกอริทึม....700
ภาพประกอบที่ 4.3 Curse of Dimensionality....811
ภาพประกอบที่ 4.4 คุณลักษณะที่ไม่ส่งผลต่อการจัดกลุ่ม....822
ภาพประกอบที่ ก-1 ไฟล์ Eclipse สำหรับติดตั้ง....98
ภาพประกอบที่ ก-2 เลือกตัวเลือกการติดตั้งโปรแกรม....98
ภาพประกอบที่ ก-3 ขั้นตอนการติดตั้งไฟล์....99
ภาพประกอบที่ ก-4 ไอคอนโปรแกรม Eclipse....99
ภาพประกอบที่ ก-5 แสดงข้อความ error ของโปรแกรม eclipse....100
ภาพประกอบที่ ก-6 แสดงไฟล์ JDK.exe....100
ภาพประกอบที่ ก-7 แสดงการติดตั้ง JDK ขั้นตอนที่ 1....101
ภาพประกอบที่ ก-8 แสดงการติดตั้ง JDK ขั้นตอนที่ 2....101
ภาพประกอบที่ ก-9 แสดงการติดตั้ง JDK ขั้นตอนที่ 3....102
ภาพประกอบที่ ก-10 แสดงการติดตั้ง JDK ขั้นตอนที่ 4....102
ภาพประกอบที่ ก-11 แสดงการติดตั้ง JDK ขั้นตอนที่ 5....103
ภาพประกอบที่ ก-12 แสดงการติดตั้ง JDK เสร็จสิ้นสมบูรณ์....103
ภาพประกอบที่ ก-13 ไฟล์ Python ที่ดาวน์โหลดมา....104
ภาพประกอบที่ ก-14 แสดงการติดตั้ง Python ขั้นที่ 1....104

สารบัญภาพประกอบ (ต่อ)

หน้า

ภาพประกอบที่ ก-15 แสดงการติดตั้ง Python ขั้นที่ 2 105
ภาพประกอบที่ ก-16 ทำการแตกไฟล์ ClassifierImbalanced 105
ภาพประกอบที่ ก-17 โปรแกรม ClassifierImbalanced 105
ภาพประกอบที่ ก-18 สร้าง shotcut 106
ภาพประกอบที่ ก-19 การติดตั้งโปรแกรมเสร็จสมบูรณ์ 106
ภาพประกอบที่ ข-1 ตัวอย่างโปรแกรมหน้าการสร้างโมเดล 108
ภาพประกอบที่ ข-2 ตัวอย่างโปรแกรมหน้าการสร้างโมเดล 119
ภาพประกอบที่ ข-3 ตัวอย่างโปรแกรมหน้าการนำโมเดลการจำแนกข้อมูลที่มีความไม่สมดุลไปใช้งาน
.....1 10