

ความรู้สึก Emotions

พชรพล แดงมณี , ภูริภัทร สุนสุข

สาขาวิทยาการข้อมูลประยุกต์ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม

Emails: 64011212040@msu.ac.th, 64011212033@msu.ac.th,

บทคัดย่อ

"Emotions" คือประสบการณ์ทางจิตใจที่เกิดขึ้นในตัวบุคคลตามสถานการณ์และประสบการณ์ที่ต่างๆ ในชีวิตประจำวัน. ความรู้สึกมีลักษณะหลายแบบ เช่น ความสุข, เศร้า, โกรธ, หรือเครียด เป็นต้น. มีผลกระทบต่อทั้งร่างกายและจิตใจของบุคคล. ความรู้สึกสามารถแสดงออกผ่านทางภาษา, ท่าทาง, และพฤติกรรม. การเข้าใจและจัดการความรู้สึกมีความสำคัญในการสร้างความสัมพันธ์ที่ดีและความเข้าใจในสังคม. ความรู้สึกมีผลต่อความสุขและคุณภาพชีวิตของบุคคล.

"Emotions" ซึ่งเป็นคอลเลกชันข้อความ Twitter ภาษาอังกฤษที่อธิบายอย่างพิถีพิถันด้วยอารมณ์พื้นฐาน 6 อารมณ์ ได้แก่ ความโกรธ ความกลัว ความสุข ความรัก ความเศร้า และความประหลาดใจ ชุดข้อมูลนี้ทำหน้าที่เป็นแหล่งข้อมูลอันมีค่าสำหรับการทำความเข้าใจและวิเคราะห์อารมณ์ที่หลากหลายที่แสดงออกมาในรูปแบบข้อความสั้นบนโซเชียลมีเดีย

คำสำคัญ -- ความโกรธ(anger), ความกลัว(fear), ความสุข(joy), ความรัก(love), ความเศร้า(sadness), ประหลาดใจ(surprise)

1. บทนำ

ความรู้สึกเป็นส่วนสำคัญของประสบการณ์มนุษย์ที่เติบโตและมีอารมณ์ส่วนตัวที่หลากหลายต่อหลายสถานการณ์ การที่เราสามารถเข้าใจและจัดการกับความรู้สึกของเราเองเป็นพื้นฐานสำคัญในการพัฒนาทักษะสังคมและความสัมพันธ์กับผู้อื่น ความรู้สึกมีความหลากหลายตามสถานการณ์ที่เกิดขึ้นและ

ส่วนตัวของแต่ละบุคคล มันมีความผูกพันกับทั้งการรับรู้สิ่งต่างๆ ที่เกิดขึ้นภายนอกและกระบวนการในจิตใจของเราเอง

ในทางปฏิบัติ, ความรู้สึกมีบทบาทสำคัญในการสื่อสารและสร้างความเข้าใจในความหมายของสิ่งต่างๆ ที่เกิดขึ้นในชีวิตประจำวัน การเข้าใจความรู้สึกของตนเองและผู้อื่นช่วยให้สามารถจัดการกับสถานการณ์ที่ซับซ้อนได้อย่างมีประสิทธิภาพ การแสดงออกถึงความรู้สึกอาจเป็นทางภาษา, ท่าทาง, และพฤติกรรมต่างๆ ซึ่งมีผลต่อทั้งระดับสังคมและระดับบุคคล

"Emotions" ซึ่งเป็นคอลเลกชันข้อความ Twitter ภาษาอังกฤษที่อธิบายอย่างพิถีพิถันด้วยอารมณ์พื้นฐาน 6 อารมณ์ ได้แก่ ความโกรธ ความกลัว ความสุข ความรัก ความเศร้า และความประหลาดใจ ชุดข้อมูลนี้ทำหน้าที่เป็นแหล่งข้อมูลอันมีค่าสำหรับการทำความเข้าใจและวิเคราะห์อารมณ์ที่หลากหลายที่แสดงออกมาในรูปแบบข้อความสั้นบนโซเชียลมีเดีย

คำสำคัญ -- ความโกรธ(anger), ความกลัว(fear), ความสุข(joy), ความรัก(love), ความเศร้า(sadness), ประหลาดใจ(surprise)

2. ทฤษฎีที่เกี่ยวข้อง

2.1 การประมวลผลภาษาธรรมชาติ (Natural Language Processing)

การประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) คือสาขาหนึ่งของปัญญาประดิษฐ์ ที่เกี่ยวข้องกับการพัฒนาอัลกอริทึมเพื่อทำให้คอมพิวเตอร์สามารถเข้าใจและประมวลผลภาษาธรรมชาติของมนุษย์ได้ ภาษาธรรมชาติเป็นภาษาที่มนุษย์ใช้สื่อสารกัน ซึ่งมีลักษณะซับซ้อนและหลากหลาย คอมพิวเตอร์จำเป็นต้องมีความสามารถในการเข้าใจภาษาธรรมชาติเพื่อที่จะสื่อสารกับมนุษย์ได้อย่างมีประสิทธิภาพ

2.2 Long Short-Term Memory (LSTM)

LSTM ย่อจาก Long Short-Term Memory ถือเป็นประเภทหนึ่งของสถาปัตยกรรมแบบ Recurrent Neural Network (RNN) อยู่ในกลุ่มของ Deep Learning ถูกออกแบบให้จดจำ Patterns ในช่วงเวลานานๆ มีประสิทธิภาพสำหรับปัญหาการทำนายที่เป็น Sequential เนื่องจากสามารถเก็บข้อมูลก่อนหน้าและนำมาร่วมใช้ในการประมวลผลได้สามารถแก้ปัญหา Long-term Dependency ได้ โดย RNN แบบดั้งเดิมจะเผชิญกับความท้าทายในเรื่อง Long-range Dependency และมีปัญหา Vanishing Gradient LSTM ถูกออกแบบมาให้จดจำ Long-term Information โดยใช้ Gating Mechanisms ที่ออกแบบมาเฉพาะมีความยืดหยุ่นที่ดี LSTM สามารถใช้ Model ได้ทั้ง Long-term และ Short-term Temporal Sequences

2.3 Word Cloud

Word Cloud (คลาวด์คำ) เป็นวิธีที่นิยมในการแสดงข้อมูลข้อความในรูปแบบกราฟิก โดยการนำเอาคำที่พบบ่อยในข้อความมาแสดงในขนาดที่ใหญ่กว่า ซึ่งทำให้สามารถดูได้ชัดเจนถึงคำที่มีน้ำหนักมากที่สุดข้อความนั้น ๆ

กระบวนการสร้าง Word Cloud มักจะประกอบด้วยขั้นตอนต่อไปนี้:

Tokenization (ตัดคำ): แยกข้อความเป็นคำๆ ที่อยู่ในประโยคหรือเอกสาร.

Counting (นับคำ): นับจำนวนครั้งที่แต่ละคำปรากฏในข้อความ.

Filtering (กรองคำ): กรองคำที่ไม่สนใจหรือไม่เกี่ยวข้อง เช่น คำที่ใช้บ่อยมากที่จะไม่ให้ปรากฏใน Word Cloud.

Visualization (การแสดงผล): นำคำที่ได้จากขั้นตอนก่อนหน้ามาสร้าง Word Cloud และแสดงผลกราฟิกที่ใหญ่ขึ้นตามความถี่ของคำ.

2.4 Visualization

"Visualization" หมายถึงกระบวนการแสดงผลข้อมูลในรูปแบบกราฟิกหรือภาพเพื่อให้ข้อมูลที่ซับซ้อนเป็นมิตรและเข้าใจได้ง่ายขึ้น โดยใช้สื่อต่าง ๆ เช่น กราฟ, แผนภูมิ, แผนที่, หรือภาพวิดีโอ เพื่อเน้นและสื่อความหมายของข้อมูล.

การทำ Visualization มีหลายวัตถุประสงค์ เช่น:

การสื่อสารข้อมูล: ทำให้ข้อมูลที่ซับซ้อนกลายเป็นเรื่องราวที่เข้าใจได้ง่าย.

การค้นพบแนวโน้ม: ช่วยในการระบุแนวโน้ม, ความสัมพันธ์, หรือลำดับขั้นของข้อมูล.

การวิเคราะห์: ช่วยในการทำความเข้าใจลึกซึ้งไปในข้อมูล, ค้นพบความสัมพันธ์, หรือลำดับเหตุการณ์.

การตัดสินใจ: สนับสนุนการตัดสินใจที่มีข้อมูลเป็นพื้นฐาน.

การสร้างแรงบันดาลใจ: ทำให้ข้อมูลน่าสนใจและน่าติดตาม.

2.5 Machine Learning

Machine Learning คือ การทำให้คอมพิวเตอร์สามารถเรียนรู้สิ่งต่างๆ และพัฒนาการทำงานให้ดีขึ้นได้ด้วยตัวเองจากข้อมูลและสภาพแวดล้อมที่ได้รับจากการเรียนรู้ของระบบ โดยไม่ต้องมีมนุษย์คอยกำกับหรือเขียนโปรแกรมเพิ่มเติม และไม่ว่าในอนาคตมันจะมีข้อมูลรูปแบบใหม่ๆ ที่เกิดขึ้นมามนุษย์ก็มักจำเป็นที่จะต้องไปนั่งเขียนโปรแกรมใหม่ เพราะคอมพิวเตอร์สามารถตีความและตอบสนองได้ด้วยตัวเอง

3. วิจัยดำเนินการวิจัย

ในส่วนนี้จะอธิบายกระบวนการที่ใช้ในงานวิจัย ซึ่งประกอบไปด้วยขั้นตอนต่อไปนี้

3.1 การรวบรวมข้อมูลและการเตรียมข้อมูล

การเตรียมข้อมูลในงานวิจัยนี้ ข้อมูลประกอบด้วยชุดข้อมูลจำนวน 1 ชุดข้อมูล คือ Emotion จากเว็บไซต์ Kaggle ซึ่งข้อมูลที่รวบรวมประกอบไปด้วยข้อความที่แสดงข้อความ Twitter และป้ายกำกับที่เกี่ยวข้องซึ่งระบุถึงอารมณ์ความรู้สึกจำนวน 416809 ข้อความ ทำการจัดกลุ่มแต่ละข้อความตาม label โดยกำหนด 0 : ความเศร้า(sadness), 1: ความสุข(joy), 2 : ความรัก(love), 3 : ความโกรธ(anger), 4 : ความกลัว(fear) และ 5 : ประหลาดใจ(surprise) ข้อมูลทั้งหมดประกอบไปด้วยข้อความคิดเห็นความเศร้า(sadness) 121,187 ข้อความ ข้อความคิดเห็นความสุข(joy) 141,067 ข้อความ ข้อความคิดเห็นความรัก(love) 34,554 ข้อความ ข้อความคิดเห็นความโกรธ(anger) 57317 ข้อความ ข้อความคิดเห็นความกลัว(fear) 47712 ข้อความ และข้อความคิดเห็นประหลาดใจ(surprise) 14972 ข้อความ จากนั้นทำการปรับเปลี่ยนข้อมูลให้อยู่ในรูปแบบที่เหมาะสมในการวิเคราะห์และประมวลผลข้อมูล ซึ่งข้อความเป็นข้อมูลที่ไม่มีโครงสร้าง ไม่สามารถนำไปประมวลผลได้ทันที จำเป็นต้องผ่านกระบวนการเตรียมข้อมูล เพื่อให้ข้อมูลต่าง ๆ อยู่ในรูปแบบเดียวกันโดยในงานวิจัยนี้มีขั้นตอนการเตรียมข้อมูลด้วยวิธีการประมวลผลข้อความ(preprocess text) คือกระบวนการแปลงข้อความให้อยู่ในรูปแบบตัวพิมพ์เล็ก(lower case) จากนั้นทำการลบอักขระพิเศษ(special characters) ตัวเลข(number) และลิงก์(link)

3.2 การสร้างโมเดล

หลังจากที่เตรียมข้อมูลเรียบร้อยแล้ว ขั้นตอนถัดไปจะเป็นการสร้างโมเดลเพื่อวิเคราะห์ความรู้สึก ซึ่งในงานวิจัยนี้ทำการสร้างโมเดลจาก 2 อัลกอริทึม คือ K-Nearest Neighbors, Naive Bayes โดยอัลกอริทึม K-Nearest Neighbors กำหนดค่า $k = 10$ และวัดความใกล้เคียงของข้อมูลที่ต้องการจำแนก (X) กับข้อมูลเรียนรู้ (Y) โดยใช้การวัดระยะทางยูคลิเดียน

(Euclidean distance) ดังสมการที่ 1 ส่วน Naive Bayes ใช้พื้นฐานทฤษฎีของเบส์เพื่อคำนวณความน่าจะเป็นของแต่ละคลาส โดยความน่าจะเป็นของแต่ละคลาสของข้อมูลที่ต้องการจะจำแนก สามารถ คำนวณได้ดังสมการที่ 2 ถ้าความน่าจะเป็นคลาสใดมีค่ามากที่สุดคลาสดังกล่าวคือผลลัพธ์

$$d(X,Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

สมการที่ 1

โดยที่

x_i คือ ค่าคุณลักษณะที่ i ของข้อมูล X (ข้อมูลที่ต้องการจำแนก)

y_i คือ ค่าคุณลักษณะที่ i ของข้อมูล Y (ชุดข้อมูลเรียนรู้)
 n คือ จำนวนคุณลักษณะ

$$p(C|X) = p(C) \prod_{i=1}^n p(x_i|C)$$

สมการที่ 2

โดยที่

$p(C|X)$ คือ ความน่าจะเป็นที่ X จะเป็นคลาส C

$p(C)$ คือ ความน่าจะเป็นคลาส C

$p(x_i|C)$ คือ ความน่าจะเป็นคลาส C ที่มีค่า x_i

2.3 การวัดประสิทธิภาพโมเดล

งานวิจัยนี้ทำการทดสอบประสิทธิภาพของโมเดลโดยแบ่งข้อมูลด้วยเทคนิค Split Test โดยการแบ่งเป็นข้อมูลชุดเรียนรู้ 80% และข้อมูลชุดทดสอบ 20% และใช้ตาราง Confusion matrix เพื่อคำนวณค่าต่างๆ

ค่า Accuracy เป็นการวัดความถูกต้องของโมเดลโดยพิจารณาจากทุกคลาส ค่า Accuracy สามารถคำนวณได้ดังสมการที่ 3

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})}$$

สมการที่ 3

ค่า Precision เป็นการวัดความแม่นยำของโมเดล โดยพิจารณาแยกทีละคลาสดังสมการที่ 4

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})}$$

สมการที่ 4

ค่า Recall เป็นการวัดความถูกต้องของโมเดลเมื่อเปรียบเทียบกับข้อมูลจริงในแต่ละคลาส ค่า recall ของแต่ละคลาสสามารถคำนวณได้ดังสมการที่ 5

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})}$$

สมการที่ 5

ค่า F1-SCORE เป็นการวัดค่า PRECISION และ RECALL พร้อมกันของโมเดลโดยพิจารณาแยกทีละคลาสดังสมการที่ 6

$$\text{F1} = \frac{2 * (\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})}$$

สมการที่ 6

4. ผลการวิจัย

ตาราง 1 ผลการประเมินบน TF Weighting โดยใช้การแบ่งชุดข้อมูล 80:20

อัลกอริทึม		KNN	Bayes
Sadness	Precision	0.31	0.31
	Recall	0.48	0.44
	F1	0.38	0.36
Joy	Precision	0.36	0.37
	Recall	0.49	0.17
	F1	0.41	0.23
Love	Precision	0.11	0.10
	Recall	0.02	0.25
	F1	0.03	0.14
Anger	Precision	0.17	0.17
	Recall	0.06	0.08
	F1	0.09	0.11
Fear	Precision	0.15	0.12
	Recall	0.04	0.10
	F1	0.06	0.11
Surprise	Precision	0.10	0.04
	Recall	0.01	0.07
	F1	0.01	0.05

ตาราง 2 Confusion Matrix ของ KNN ใช้เทคนิคการให้น้ำหนักค่าแบบ TF Weighting แบ่งชุดข้อมูล 80:20

		Confusion Matrix KNN					
True Labels	Sadness	11627	10475	239	1135	683	42
	Joy	12000	13773	440	1200	702	49
	Love	2749	3637	118	268	151	6
	Anger	5262	5026	125	684	320	24
	Fear	4270	4272	122	531	359	40
	Surprise	1221	1441	37	160	157	17
		Sadness	Joy	Love	Anger	Fear	Surprise
		Predicted Labels					

Applied Data Science : Natural Language Processing.

ตาราง - Confusion Matrix ของ Bayes ใช้เทคนิคการให้น้ำหนักคำแบบ TF Weighting แบ่งชุดข้อมูล 80:20

		Confusion Matrix Bayes					
True Labels	Sadness	10624	3322	4637	1701	2324	1593
	Joy	11185	4715	6223	1806	2271	1964
	Love	2503	1199	1746	365	568	548
	Anger	4652	1668	2350	952	1071	748
	Fear	4081	1361	1861	712	924	655
	Surprise	1210	474	692	171	274	212
		Predicted Labels					
		Sadness	Joy	Love	Anger	Fear	Surprise

เอกสารอ้างอิง

- [1] [Narut Soontranon](https://www.nerd-data.com/deep_learning_lstm/) : LSTM คือ อะไร
https://www.nerd-data.com/deep_learning_lstm/
- [2] Data Base Camp : Long Short-Term Memory Networks (LSTM)- simply explained!
<https://databasecamp.de/en/ml/lstms>
- [3] Wiki Pedia : แท้ก็คลาวด์
<https://th.wikipedia.org/wiki/%E0%B9%81%E0%B8%97%E0%B9%87%E0%B8%81%E0%B8%84%E0%B8%A5%E0%B8%B2%E0%B8%A7%E0%B8%94%E0%B9%8C>
- [4] Wiki Pedia : Visualization
<https://en.wikipedia.org/wiki/Visualization>
- [5] [Peachapong Poolpol](https://peachapong-poolpol.medium.com/na%C3%AFve-bayes-classification-cb6cf905505d) : Naïve Bayes Classification
<https://peachapong-poolpol.medium.com/na%C3%AFve-bayes-classification-cb6cf905505d>

- [6] [KongRuksiam Studio](https://kongruksiam.medium.com/%E0%B8%AA%E0%B8%A3%E0%B8%B8%E0%B8%9B-machine-learning-ep-4-%E0%B9%80%E0%B8%9E%E0%B8%B7%E0%B9%88%E0%B8%AD%E0%B8%99%E0%B8%9A%E0%B9%89%E0%B8%B2%E0%B8%99%E0%B9%83%E0%B8%81%E0%B8%A5%E0%B9%89%E0%B8%97%E0%B8%B5%E0%B9%88%E0%B8%AA%E0%B8%B8%E0%B8%94-k-nearest-neighbors-787665f7c09d) : สรุป Machine Learning(EP.5) —การคำนวณเพื่อนบ้านใกล้สุด (K-nearest Neighbors)
<https://kongruksiam.medium.com/%E0%B8%AA%E0%B8%A3%E0%B8%B8%E0%B8%9B-machine-learning-ep-4-%E0%B9%80%E0%B8%9E%E0%B8%B7%E0%B9%88%E0%B8%AD%E0%B8%99%E0%B8%9A%E0%B9%89%E0%B8%B2%E0%B8%99%E0%B9%83%E0%B8%81%E0%B8%A5%E0%B9%89%E0%B8%97%E0%B8%B5%E0%B9%88%E0%B8%AA%E0%B8%B8%E0%B8%94-k-nearest-neighbors-787665f7c09d>

- [7] Machine Learning เทคโนโลยีประโยชน์ครอบคลุมจักรวาล
<https://www.cyberelite.co.th/blog/machine-learning/>