บทความวิจัย

Computer Science Department
Faculty of Informatics, Mahasarakham University

การศึกษาโมเดลแบบทรานสฟอร์มเมอร์สำหรับการจำแนกความรู้สึก

A Study of Transformer-based Model for Sentiment Classification

เกียรตินำชัย อินทร์นอก, จันทิมา พลพินิจ

Kiatnumchai Innork and Jantima Polpinij

บทคัดย่อ

การวิเคราะห์ความรู้สึกเป็นประโยชน์สำหรับการดำเนินการทางธุรกิจเป็นอย่างมาก เพราะการประเมินความรู้สึกของลูกค้าที่มีต่อสินค้าและบริการนั้น สามารถนำมาใช้ประโยชน์อย่างมาก จากการศึกษาที่ผ่านมาพบว่ากระบวนการในการจำแนกความรู้สึกสามารถแบ่งออกเป็น 3 ประเภท ได้แก่ กระบวนการที่ใช้คำศัพท์ (Lexicon-based Approaches) กระบวนการที่ใช้อัลกอริทึมการเรียนรู้ของเครื่อง (Machine Learning-based Approaches) และกระบวนการที่ใช้อัลกอริทึมการเรียนรู้เชิงลึก (Deep Learning-based Approaches) แต่เทคนิคเหล่านี้มองข้ามข้อมูลของอารมณ์จากการพิจารณาบริบทในบทวิจารณ์ ดังนั้นเวกเตอร์ที่สร้างขึ้นมาอาจจะเป็นเพียงเวกเตอร์ของคำ ซึ่งเป็นสาเหตุของจำแนกข้อมูลที่ผิดพลาด (Misclassification) ไม่กี่ปีที่ผ่านมามีงานวิจัยด้านการจำแนกความรู้สึก ได้ประยุกต์นะเอากระบวนการแบบทรานสฟอร์มเมอร์ (Transformer-based Approach) มาใช้ ซึ่งเป็นกระบวนการที่มีการพิจารณาโครงสร้างและความหมายของคำควบคู่กับการวิเคราะห์อารมณ์ในบริบทของข้อมูล เนื่องจากโมเดลแบบทรานสฟอร์มเมอร์ (Transformer Model) ที่พัฒนาด้วยโครงข่ายประสาทเทียมด้วยโครงสร้างแบบ Encoder-Decoder ที่เรียนรู้บริบทและความหมายแบบ Self-Attention ที่จะวิเคราะห์และเปรียบเทียบข้อมูลอินพุตทั้งหมดในลักษณะการวิเคราะห์แบบ Sequence-to-Sequence และโมเดลแบบทรานสฟอร์มเมอร์ ยังสามารถนำมาใช้กับงานทางด้านการประมวลผลภาษาธรรมชาติ (Natural Language Processing : NLP) ได้หลากหลาย ดังนั้นโครงงานฉบับนี้จึงได้นำเสนอโมเดลการวิเคราะห์ความรู้สึกที่ผ่านการ Fine-tuning ด้วยข้อมูลบทวิจารณ์โรงแรมภาษาอังกฤษ กับ Pre-trained Model อย่างโมเดลทรานสฟอร์มเมอร์แบบ BERT-base ภายใต้แนวคิดของการวิเคราะห์ความรู้สึก (Sentiment Classification) ด้วยการประมวลผลภาษาธรรมชาติ และการจำแนกเอกสารข้อความ (Text Classification)

คำสำคัญ การวิเคราะห์ความรู้สึก, โมเดลแบบทรานสฟอร์มเมอร์, โมเดลทรานสฟอร์มเมอร์แบบ BERT-base, การประมวลผลภาษาธรรมชาติ, การจำแนกเอกสารข้อความ

Computer Science Department
Faculty of Informatics, Mahasarakham University

บทนำ

การวิเคราะห์ความรู้สึก (Sentiment Analysis) [1-3] เป็นสาขาวิจัยทางด้านการประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) ที่ใช้ในการประเมินความรู้สึกของลูกค้าหรือผู้บริโภคจากบทวิจารณ์ของลูกค้า (Customer Review) ที่มีต่อสินค้าหรือบริการว่ามีความรู้สึกเป็นบวก (Positive) ลบ (Negative) หรือเป็นกลาง (Neutral) โดยทั่วไปเทคนิคที่นิยมใช้ในการวิเคราะห์ความรู้สึกคือ เทคนิคด้านการจำแนกเอกสาร (Text Classification) ดังนั้นการวิเคราะห์ความรู้สึกด้วยเทคนิคด้านการจำแนกเอกสาร จึงเรียกว่าการจำแนกความรู้สึก (Sentiment Classification) [2]

การวิเคราะห์ความรู้สึกเป็นประโยชน์สำหรับการดำเนินการทางธุรกิจเป็นอย่างมาก เพราะการประเมินความรู้สึกของลูกค้าที่มีต่อสินค้าและบริการนั้น สามารถนำมาใช้ประโยชน์อย่างมากเช่น เจ้าของสินค้าและบริการสามารถใช้ข้อมูลด้านการวิเคราะห์ความรู้สึกในการปรับปรุงสินค้าและบริการของตนเอง เป็นข้อมูลที่ใช้ในการรักษาแบรนด์หรือฐานลูกค้า รับรู้แนวโน้มทางการตลาดของสินค้าและบริการ หรือการตัดสินใจเกี่ยวกับสินค้าและบริการของตน ในขณะที่ในส่วนของลูกค้าหรือผู้บริโภค ข้อมูลดังกล่าวจะช่วยในการตัดสินใจว่าควรจะซื้อสินค้าหรือบริการเหล่านั้นหรือไม่

จากการศึกษาที่ผ่านมาพบว่า กระบวนการในการจำแนกความรู้สึกสามารถแบ่งออกเป็น 3 ประเภท [1] ได้แก่ กระบวนการที่ใช้คำศัพท์ (Lexicon-based Approaches) กระบวนการที่ใช้อัลกอริทึมการเรียนรู้ของเครื่อง (Machine Learning-based Approaches) และกระบวนการที่ใช้อัลกอริทึมการเรียนรู้เชิงลึก (Deep Learning-based Approaches) กระบวนการเหล่านี้มีการประยุกต์ใช้งานอย่างกว้างขวาง อย่างไรก็ตามพบว่า เทคนิคเหล่านี้มองข้ามข้อมูลของอารมณ์จากการพิจารณาบริบทในบทวิจารณ์ [4] ดังนั้นเวกเตอร์ที่สร้างขึ้นมาอาจจะเป็นเพียงเวกเตอร์ของคำ และบ่อยครั้งยังพบปัญหาที่เรียกว่า Out-of-Vocabulary (OOV) [5] นั่นคือคำศัพท์ที่ไม่ได้เป็นส่วนหนึ่งของพจนานุกรมทั่วไปที่พบในสภาพแวดล้อมการประมวลผลภาษาธรรมชาติ สาเหตุนี้นำไปสู่การสูญหายของข้อมูล [4] นอกจากนั้นความท้าทายอีกประการหนึ่งสำหรับการจำแนกความรู้สึกก็คือการขาดข้อมูลที่มีคำอธิบายประกอบ (Annotated Data) ซึ่งจะส่งผลให้เกิดการเรียนรู้โมเดลที่ให้ค่า false negative สูง [6] และบางครั้งยังพบว่า มีความขัดแย้งระหว่างบริบทของการรีวิวและคลาสลาเบลซึ่งเป็นสาเหตุของจำแนกข้อมูลที่ผิดพลาด (Misclassification) [4]

ไม่กี่ปีที่ผ่านมามีงานวิจัยด้านการจำแนกความรู้สึก ได้ประยุกต์นะเอากระบวนการแบบทรานสฟอร์มเมอร์ (Transformer-based Approach) มาใช้ ซึ่งเป็นกระบวนการที่มีการพิจารณาโครงสร้างและความหมายของคำควบคู่กับการวิเคราะห์อารมณ์ในบริบทของข้อมูล [4] เนื่องจากโมเดลแบบทรานสฟอร์มเมอร์ (Transformer Model) พัฒนาด้วยโครงข่ายประสาทเทียมด้วยโครงสร้างแบบ Encoder-Decoder (Neural Network with an Encoder-Decoder Structure) ที่เรียนรู้บริบทและความหมายแบบ Self-Attention [7] ที่จะวิเคราะห์และเปรียบเทียบข้อมูลอินพุตทั้งหมดในลักษณะการวิเคราะห์แบบ Sequence-to-Sequence เข้าด้วยกันเพื่อคำนวณหาความสัมพันธ์และลำดับของคำในภาษาหรือประโยค ดังนั้นจึงมักจะมีการกล่าวว่าโมเดลแบบทรานสฟอร์มเมอร์ก็คือโมเดลของภาษา (Language Model)

จากเหตุผลข้างต้น งานวิจัยนี้จึงได้นำเสนอโมเดลการจำแนกความรู้สึกบทวิจารณ์โรงแรมภาษาอังกฤษจากโมเดลทรานสฟอร์มเมอร์แบบ

Computer Science Department Faculty of Informatics, Mahasarakham University

BERT-base ที่มีการให้ผลการวอเคราะห์เป็น 3 ระดับ คือ

ระดับที่ 5 ความรู้สึกแบบบวก (positive)

ระดับที่ 3 ความรู้สึกแบบกลาง (Neural)

ระดับที่ 1 ความรู้สึกแบบลบ (Negative)

ภายใต้แนวคิดของการวิเคราะห์ความรู้สึก (Sentiment Classification) ด้วยการประมวลผลภาษาธรรมชาติ และการจำแนกเอกสารข้อความ (Text Classification)

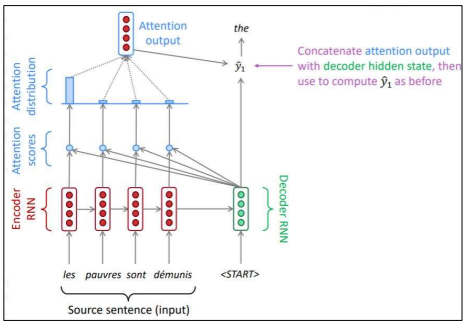
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

1.ความเป็นมาของโมเดลแบบทรานฟอร์มเมอร์ (History of Transformer-based Model)

ประเภทของงานที่ประยุกต์ใช้อัลกอริทึมการเรียนรู้ของเครื่อง (Machine Learning: ML) อาจจะแบ่งตามลักษณะของข้อมูลอินพุตที่เข้ามาและเอาต์พุตที่ได้ ซึ่งโดยทั่วไปก็อาจจะเป็นงานด้านการจำแนกข้อมูล (Classification) หรือรีเกรสชัน (Regression) แต่กรณีที่ข้อมูลอินพุตเป็น Sequence แล้วให้เอาต์พุตเป็นอีก Sequence ในลักษณะนี้เรียกว่า Sequence-to-Sequence Learning ซึ่งในการประมวลผลในงานด้าน Machine Translation ก็จัดเป็นงานในลักษณะ Sequence-to-Sequence Learning และนับว่าเป็นสิ่งที่ผู้คนให้ความสำคัญมากตั้งแต่ยุคสมัยแรกๆ ของ AI มาจนถึงปัจจุบัน

ปัจจุบันวิธีที่เป็นมาตรฐานสำหรับทำงานแบบ Sequence-to-Sequence Learning ก็คือ Sequence-to-Sequence Model (seq2seq) หรือเรียกอีกชื่อว่าโมเดล RNN Encoder–Decoder ซึ่งโมเดลนี้จะมีองค์ประกอบหลักอยู่ 2 ส่วน เรียกว่า Encoder กับ Decoder โดยส่วนของ Encoder จะรับอินพุตเข้ามาทีละหน่วยผ่านทาง RNN และเก็บสะสม Information ที่จำเป็นไว้ จากนั้นจะผ่าน Information นี้ไปยังส่วนของ Decoder ซึ่งก็จะเป็น RNN อีกตัวนึงที่ใช้ผลิตเอาต์พุตออกมาทีละหน่วย สำหรับการผลิตเอาต์พุตก็พิจารณาจาก Information ที่ได้รับมาและเอาต์พุตตัวก่อนหน้า ซึ่งแผนภาพ ของโมเดล seq2seq

อย่างไรก็ดี โมเดล seq2seq จะมีปัญหาคอขวดเกิดขึ้น นั่นคือการส่ง Information เป็นทอดๆ ที่เป็นสายยาวๆ เช่นนี้ ก็อาจจะทำให้มี Information ที่จำเป็นบางอย่างสูญหายไประหว่างทางได้ ยกตัวอย่างเช่น จากรูปข้างบน สมมติว่าเป็นงาน Machine Translation ที่แปลจากภาษาไทยไปเป็นภาษาอังกฤษ โดย “A B C” คือคำว่า “ฉัน เลี้ยง แมว” ตามลำดับ และ “W X Y Z” เป็นคำว่า “I have a cat” ตามลำดับ จะเห็นว่าเอาต์พุตของคำว่า “cat” จะขึ้นอยู่กับอินพุตคำว่า “แมว” โดยตรง แต่ข้อมูลจาก C กว่าจะส่งมาถึง Z ต้องผ่านตัวกลางหลายทอด ทำให้อาจจะสูญหายได้ โดยเฉพาะอย่างยิ่งถ้าเป็นประโยคยาวๆ ดังนั้นจึงเกิดแนวคิดที่ว่า น่าจะดีกว่าถ้าเราให้กระบวนการของการสร้างเอาต์พุต สามารถโฟกัสไปที่อินพุตส่วนใดส่วนหนึ่งของประโยคได้โดยตรง และนี่คือที่มาของ “Attention” นั่นเอง ซึ่งภาพที่ 1 แสดงการทำ Attention สำหรับโมเดล seq2seq



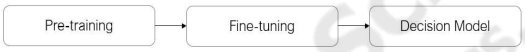
ภาพที่ 1 การทำ Attention สำหรับโมเดล seq2seq

Computer Science Department
Faculty of Science, Mahasarakham University

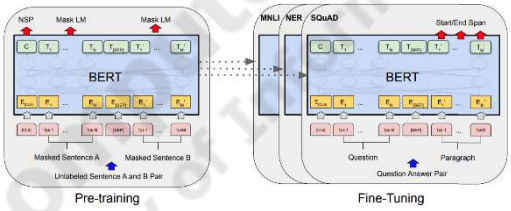
2. BERT (Bidirectional Encoder Representations from Transformer)

จากปัญหาของการแสดงเอกสารข้อความ (Text Representation) โดยเฉพาะในเรื่องของการพิจารณาตำแหน่งของคำ และการให้ความหมายของคำ จึงมีความพยายามที่จะพัฒนาโมเดลอื่นๆ เพื่อปรับปรุงประสิทธิภาพของงานด้าน NLP ซึ่งโมเดลที่ได้มีการนำเสนอขึ้นมาคือโมเดลแบบทรานสฟอร์มเมอร์ ซึ่งโมเดลที่เด่นๆ เช่น BERT ซึ่ง BERT คือ Google (AI) Algorithm ที่ในเวอร์ชันล่าสุดนี้มันได้พัฒนาบนพื้นฐานของเทคโนโลยี AI Neural Network เพื่อให้ระบบอัลกอริทึมเข้าใจภาษามนุษย์มากขึ้น

BERT เป็น Transfer Learning ที่ประกอบด้วย 3 ส่วนหลัก คือ Pre-training, Fine-tuning และ Decision Model โดย Pre-training สถาปัตยกรรมของ BERT จะเป็นโมเดลที่มีการใช้ Encoder อย่างเดียว แสดงได้ดังภาพที่ 2 และภาพที่ 3



ภาพที่ 2 กรอบการดำเนินงานใน BERT



ภาพที่ 3 Pre-training และ Fine-tuning ของ BERT

2.1 Pre-training of BERT

ในการเรียนรู้โมเดลของภาษา (Language Model Learning) สมมติว่า การทำนาย “คำ”ในประโยค เช่น “The child came home from ___.” ในส่วนที่ละไว้ หลายๆ โมเดลจะทำนาย “คำ” ถัดไปในลำดับของประโยค (Sequence) ซึ่งวิธีการที่พิจารณาลำดับของประโยคแบบ Directional Model จึงเป็นวิธีการที่มักจะมีข้อจำกัดของการเรียนรู้จากบริบท (Context Learning) ซึ่ง BERT ได้ใช้กลยุทธ์ 2 อย่างในการแก้ปัญหานี้

กลยุทธ์ที่ 1 Masked Language Model (MLM)

BERT จะมีการแทน “คำ” ในแต่ละ sequence ด้วยโทเค็น [MASK] เรียกว่า “Masked Word” จำนวน 15% หลังจากนั้นโมเดลจะทำการทำนายค่าดั้งเดิม (Original Value) ของ Masked Word ด้วยคำที่ไม่มี MASK ที่เรียกว่า “Non-masked Word” ซึ่งเป็นบริบทที่อยู่รอบ “Masked Word” ในword sequence นั้นๆ สำหรับการทำนายคำที่เป็นเอาต์พุต

กลยุทธ์ที่ 2 Next Sentence Prediction (NSP)

BERT จะรับประโยคเข้ามาเป็นคู่ และจะเรียนรู้เพื่อทำนาย ถ้าประโยคที่สองในคู่ประโยคที่รับเข้ามา เป็น subsequence ในเอกสารต้นฉบับ ระหว่างการเรียนรู้นั้น 50% ของข้อมูลอินพุตจะเป็นคู่ประโยค (Pairs of Sentences) โดยประโยคที่สองที่ตามหลังประโยคแรกต้องเป็น subsequent sentence ในเอกสารต้นฉบับ ในขณะที่ อีก 50% ของข้อมูลอินพุตที่เหลือ จะเป็น random sentence จากคลังเอกสารที่ถูกเลือกมาเป็นประโยคที่สอง

2.2 Fine-tuning of BERT

สำหรับงานด้านการจำแนกเอกสารข้อความ (Text Classification) เช่น การวิเคราะห์ความรู้สึก (Sentiment Analysis) ทำได้คล้ายกับ Next Sentence Classification ดังนั้นจะมีการเพิ่มเลเยอร์ของการจำแนกเอกสารที่ด้านบนของเอาต์พุตของทรานสฟอร์มเมอร์สำหรับโทเค็น [CLS]

Computer Science Department
Faculty of Informatics, Mahasarakham University

2.3 งานวิจัยที่เกี่ยวข้อง (Related Works)

Fine-Tuning BERT for Sentiment Analysis of Vietnamese Reviews [8]

ในปี 2020 Nguyen และคณะ ได้ทำการวิจัยโดยใช้ BERT ในการทำ sentiment Analysis กับภาษาเวียดนาม โดยทดลองแบ่งการ Fine-tuning เป็นสองวิธี วิธีที่ 1 ใช้โทเค็น [CLS] เพิ่มไปที่ตำแหน่งเริ่มต้นของประโยค จากนั้นเวกเตอร์ output ของโทเค็นนี้จะถูกส่งผ่าน feed-forward neural network เพื่อจำแนกข้อความ วิธีที่ 2 ใช้ผลลัพธ์ที่ได้จาก BERT รวมไปถึงโทเค็น [CLS] มาสร้างเป็นเมทริกซ์ SEQ_LENxh โดยที่ SEQ_LEN คือความยาวสูงสุดของลำดับอินพุต และ h คือความยาวของ hidden vectors จากนั้นนำ output matrix ที่ได้จากกระบวนการข้างต้นไปใช้เป็น input ไปยังโมเดลการจำแนกประเภท 3 โมเดล ได้แก่ LSTM, TextCNN, RCNN โดยพวกเขาใช้ชุดข้อมูลจากสองแหล่ง ชุดข้อมูลแรกคือ Ntc-sv2 เป็นชุดข้อมูลรีวิวร้านอาหารและอาหารบน Foody ประกอบด้วยชุดข้อมูลตัวอย่าง 50,000 ชุดเอกสาร label จะถูกกำหนดจากคะแนนเฉลี่ย (avg_score) โดย คะแนนที่สูงกว่า 8.5 เป็น positive และ คะแนนที่น้อยกว่า 5 เป็น Negative ชุดข้อมูลที่ 2 คือ Vreview เป็นชุดข้อมูลบทวิจารณ์ผลิตภัณฑ์บนเว็บไซต์อีคอมเมิร์ซต่างๆ และยังมีบทวิจารณ์บางส่วนจากที่เกี่ยวกับอาหารและร้านอาหารบน Foody4 label จะถูกกำหนดจากคะแนนเฉลี่ย (avg_score) โดย คะแนนที่สูงกว่า 7.5 เป็น positive และ คะแนนที่น้อยกว่า 5 เป็น Negative จากการทดลองพบว่าในการทำ Sentiment Analysis กับภาษาเวียดนาม พบว่าโมเดล BERT-RCNN ได้ประสิทธิภาพที่ดีที่สุด โดยค่า F1 88.22% แม้โมเดลนี้จะได้ไม่ได้มีประสิทธิภาพสูงกว่าโมเดลอื่นมากนัก แต่สำหรับงานในอนาคต พวกเขาตั้งเป้าที่จะขยายวิธีการที่เสนอสำหรับการวิเคราะห์ความรู้สึกตามแง่มุมต่างๆ และอาจมีการทดลองที่กว้างขวางมากขึ้นในชุดข้อมูลต่างๆ

Comparative study of Twitter Sentiment on COVID - 19 Tweets [9]

ในปี 2021 Nair และคณะ กล่าวว่า เมื่อเร็วๆ นี้ จำนวนทวีตเกี่ยวกับ COVID-19เพิ่มขึ้นในอัตราที่ไม่เคยมีมาก่อน โดยรวมถึงทวีตเชิงบวก เชิงลบ และเป็นกลางลักษณะทวีตที่หลากหลายนี้ดึงดูดให้นักวิจัยทำการวิเคราะห์ความรู้สึกและวิเคราะห์อารมณ์ที่หลากหลายของประชาชนจำนวนมากที่มีต่อโควิด-19 งานวิจัยนี้จึงได้นำเสนอการวิเคราะห์ความรู้สึกจากความคิดเห็นใน Twitter ที่มีต่อโควิด-19 โดยใช้เดลที่ 3 ตัวคือ Logistic Regression sentiment analysis, VADER sentiment analysis, BERT sentiment analysis ชุดข้อมูลที่ผู้วิจัยนำมาใช้คือ ความคิดเห็นที่มีต่อโควิด-19 จำนวน 20,000 ชุดเอกสาร โดยแบ่งออกเป็น 3 label ได้แก่ positive, neutral, negative แบ่งข้อมูลเป็นข้อมูลชุดสอน(train) ต่อชุดข้อมูลทดสอบ(Test) เป็น 75 : 25 จากการทดลองเปรียบเทียบผลลัพธ์จากทั้ง 3 โมเดลพบว่า BERT เป็นโมเดลที่มีค่า Accuracy สูงที่สุดคือ 92%

BERT-Based Stock Market Sentiment Analysis [1จ]

ในปี 2020 Lee และคณะ งานวัยจัยของพวกเขาคือการวิเคราะห์ความรู้สึกจากความคิดเห็นของผู้คนเกี่ยวของหุ้นในหุ้นสหรัฐจากเว็บไซต์ Stocktwits โดยใช้ BERT-based เป็น pre-trained และ Fine-tuned ด้วยชุดข้อมูลภาษาอังกฤษโดยใช้ 2 ชุดคือ Bullish และ bearish เป็นชุดข้อมูลที่มี label ที่บอกความรู้สึกของข้อความเป็น positive และ Negative แต่ชุดข้อมูลจาก Bullish มีขนาดใหญ่กว่า bearish อยู่มาก พวกเขาจึงสุ่มเลือกข้อมูลจาก Bullish มา 1.3 เท่า ของชุดข้อมูล bearish เพื่อ

Computer Science Department Faculty of Science, Mahasarakham University

หลีกเลี่ยงปัญหาการกระจายข้อมูลที่ไม่สมดุล จากนั้นแบ่งข้อมูลออกเป็นชุดข้อมูลสอน(Training) และ ชุดข้อมูลตรวจสอบ (Validation) อัตราส่วนอยู่ที่ 80 : 20 ตามลำดับ โดยผลลัพธ์ที่พวกเขาได้จากการทดลอง ได้ค่า accuracy : 87.3% สุดท้ายพวกเขากล่าวว่าแบบจำลอง BERT ได้รับผลลัพธ์ที่น่าทึ่งในการจำแนกความคิดเห็น สิ่งนี้จะช่วยในการศึกษาเกี่ยวกับเกี่ยวกับความสัมพันธ์ระหว่างโซเชียลมีเดียของหุ้นกับราคาหุ้นจริง

Sentiment Analysis of Reviews in Kazakh With Transfer Learning Techniques [11]

ในปี 2022 Nugumanova และคณะ ได้ศึกษาเกี่ยวกับการวิเคราะห์ความรู้สึกกับภาษา Kazakh โดยการทดลองของพวกเขาเลือกใช้การ transfer learning สองแบบคือ zero-shot learning และ fine-tuning สำหรับโมเดล Pre-training จากเขาเลือกใช้โมเดล BERT-base จาก HuggingFace's Transformers Library [12] มา 3 แบบ ได้แก่ NLPtown Savacy และ BERTurk ชุดข้อมูลที่ใช้ในการทดลองเป็นชุดข้อมูลที่รวบรวมจาก Facebook เว็บไซต์ร้องเรียนของผู้บริโภค zhalobi.kz เว็บไซต์แอปพลิเคชัน 2GIS ข้อมูลมี label แบ่งเป็นสองแบบคือ positive และ Negative แบ่งชุดข้อมูลออกเป็น 3 ส่วน ชุดข้อมูลทดสอบ (Training) ชุดข้อมูลทดสอบ (Testing) ชุดข้อมูลตรวจสอบ (Validation) แบ่งเป็น 30 : 30 : 100 ชุดข้อมูล ตามลำดับ จากการทดลองพวกเขาพบว่า ค่า Accuracy สูงที่สุดคือ NLPtown ที่ทำการ transfer learning แบบ fine-tuning ได้ค่า Accuracy 0.73

Comparison of BERT Models and Machine Learning Methods for Sentiment Analysis on Turkish Tweets [13]

ในปี 2021 Guven ได้ศึกษาการวิเคราะห์ความคิดเห็นบนทวีตโดยใช้โมเดล BERT และนอกจากนี้ ยังมีการเปรียบเทียบโมเดล BERT ที่ผ่านการฝึกอบรมและวิธีการเรียนรู้ของเครื่อง ในบรรดาวิธีการเรียนรู้ของเครื่อง Random Forest, Naive Bayes และ Logistic Regression นั้น Logistic Regression เป็นวิธีที่ประสบความสำเร็จมากที่สุดถึง 98.4% ในอีกด้านหนึ่ง โมเดล BERT มีความแม่นยำถึง 98.75% และแซงหน้าวิธีการเรียนรู้ของเครื่อง ผลเชิงบวกของแบบจำลอง BERT ในการวิเคราะห์ความรู้สึกได้แสดงให้เห็นในการศึกษานี้

วิธีดำเนินงานวิจัย

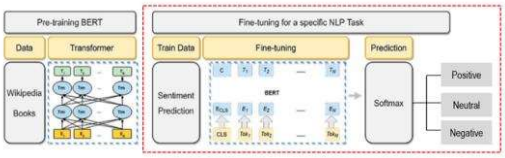
1.ชุดข้อมูล (Dataset)

ในงานวิจัยฉบับนี้ใช้ชุดข้อมูลบทวิจารณ์โรงแรมจาก TripAdvisor ที่สามารถดาวน์โหลดได้จาก www.tripadvisor.com โดยมีลักษณะของข้อมูลดังภาพ 4

ภาพที่ 4 ชุดข้อมูลบทวิจารณ์โรงแรมจาก TripAdvisor

2.กรอบการดำเนินงาน

กรอบการดำเนินงานในการประยุกต์โมเดลแบบทรานสฟอร์มเมอร์สำหรับการจำแนกความรู้สึกแบ่งเป็น 2 ส่วนหลักๆ คือ



ภาพที่ 5 กรอบการดำเนินงานในการประยุกต์โมเดลแบบทรานสฟอร์มเมอร์สำหรับการจำแนกความรู้สึก

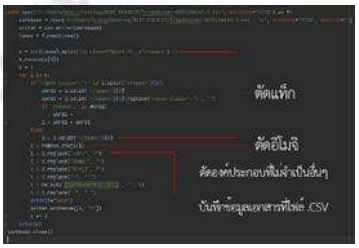
ส่วนที่ 1 คือ Pre-Training BERT คือการสร้างโมเดลด้วยข้อมูลเอกสารจำนวนมากๆ ซึ่งเป็นเอกสารที่ไม่จำเป็นต้องมีคลาสลาเบล เพื่อให้โมเดลเกิดความเข้าใจ Language Model นั่นคือ ทำให้โมเดลเข้าใจในลักษณะของภาษาหรือการใช้ภาษา

ส่วนที่ 2 คือ Fine-tuning for a specific NLP Task ซึ่งก็คือ การนำข้อมูลที่เกี่ยวข้องกับงานที่ต้องการประมวลผลมาทำการปรับค่าในโมเดลที่ได้จาก Pre-training โดยข้อมูลที่นำมา Fine-tuning ควรเป็นข้อมูลที่มีคลาสลาเบล เพื่อใช้ในการปรับ weight เพื่อให้กลายเป็น Decision Model ที่ต่อยอดมาจาก BERT จากนั้นจะนำ Decision Model ที่ได้จาก BERT ไปใช้ร่วมกับฟังชัน SoftMax เพื่อจำแนกความรู้สึกของเอกสารได้

2.1 Pre-training of BERT

ขั้นตอนที่ 1 : การอ่านข้อมูลและทำความสะอาดข้อมูล

นำข้อมูลเข้าสู่กระบวนการทำความสะอาดก่อนที่จะนำไปเข้าสู่ขั้นตอนต่อไป



ภาพที่ 6 ตัวอย่างโปรแกรมสำหรับตัดคำ

ขั้นตอนที่ 2 : การตัดคำ (Tokenization)

การตัดคำเพื่อสร้างโมเดลของ BERT จะใช้ WordPiece Tokenizer ในการแยกข้อความออกเป็น “คำ” ให้อยู่ในรูปแบบที่เรียกว่า “รูปแบบของคำเต็ม (Full Form)” หรือเป็น “ชิ้นส่วนของคำ (Word Pieces)” การใช้ WordPiece Tokenizer จะช่วยลดจำนวนคลังในคลังคำศัพท์ และแก้ปัญหา Out-of-Vocabulary (OOV) [5] อีกด้วย

ขั้นตอนที่ 3 : Token Embeddings

Token embedding เป็นกระบวนกระเปลี่ยนคำให้กลายเป็นตัวเลขรูปแบบเวกเตอร์ โดยใช้หลักการของ WordPiece Embedding จะเปลี่ยนคำแต่ละที่จะได้การตัดคำเป็นตัวเลขจำนวนเต็มที่สามารถบ่งบอกว่าคำเหล่านั้นจะถูกเปลี่ยนเป็นเวกเตอร์ใดใน lookup table ของคลังคำศัพท์ ซึ่งมีจำนวนคำศัพท์อยู่ 30,000 คำศัพท์ ในขณะที่ขนาดของเวกเตอร์ใน lookup table มีขนาด 768 มิติ เพื่อที่จะสามารถแบ่งเวกเตอร์ของแต่ละคำเป็น 12 ส่วนเพื่อให้สอดคล้องกับจำนวน Head Attention ที่ใช้ในขั้นตอนของ Multi-Head Attention และก่อนที่จะทำกระบวนการในกาทำ Token embedding โมเดล BERT จะเพิ่มโทเค็นพิเศษนั้นคือ โทเค็น [CLS] ไว้หน้าสุดของเอกสารข้อความ และเพิ่มโทเค็น [SEP] โดยที่

โทเค็น [CLS] เป็นโทเค็นพิเศษสำหรับการจำแนกประเภท (Classification) ถูกใช้เป็นตัวแทนของข้อมูลเอกสารนั้นๆ สำหรับงานด้านการจำแนกประเภท

โทเค็น [SEP] เป็นโทเค็นพิเศษสำหรับแยกประโยคหรือสิ้นสุดประโยค เพื่อให้โมเดลรับรูปว่าตำใดเป็นของประโยคใด

การทำ Token embedding ด้วย WordPiece Embedding มีขั้นตอนดังนี้

Computer Science Department
Faculty of Informatics, Mahasarakham University

(1) เพิ่มโทเค็น [CLS] ไว้ข้างหน้าสุดและเพิ่มโทเค็น [SEP] ไว้ข้างหน้าสุดของข้อมูลเอกสาร

(2) เปลี่ยนแต่ละคำให้เป็นตัวเลขจำนวนเต็มสำหรับบอกที่อยู่ของเวกเตอร์ใน lookup table

(3) นำตัวเลขของแต่ละคำที่เป็นเหมือนตัวบ่งบอกที่อยู่ของเวกเตอร์ใน lookup table เข้าตรวจสอบกับ lookup table เพื่อดึงเวกเตอร์ขนาด 786 มิติที่ของแต่ละคำในข้อมูลเอกสาร

ขั้นตอนที่ 5 : Segment Embeddings

Segment Embeddings คือการฝังตำแหน่งของประโยคในข้อมูลเอกสารเพื่อให้โมเดลรับรู้ว่า คำไหนอยู่ในประโยคลำดับที่เท่าใดในข้อมูลเอกสาร ซึ่งโมเดลแบบ BERT จะสามารถรับประโยคเข้าสู่โมเดลได้ครั้งละ 2 ประโยค [15] หากกำหนดให้ “i love programming.” เป็นประโยคแรกในชุดข้อมูลเอกสาร ประโยคนี้จะถูกฝัง Segment Embeddings ให้เป็นเวกเตอร์ขนาด 768 มิติโดยที่ทุกมิติมีค่าเท่ากับ 0 (index ที่ 0) และกำหนดให้ “i love my job.” เป็นประโยคที่ 2 ในชุดข้อมูลเอกสารประโยคนี้จะถูกฝัง Segment Embeddings ให้เป็นเวกเตอร์ขนาด 768 มิติโดยที่ทุกมิติมีค่าเท่ากับ 1 (index ที่ 1)

ขั้นตอนที่ 4 : Positional encoding

Positional encoding เนื่องด้วยว่า Transformer ตัดการทำ recurrent (จาก RNNs) และ convolute (จาก CNNs) ออกไป ปัญหาที่เกิดขึ้นคือ ในขณะที่กำลังประมวลผลคำใดคำหนึ่งอยู่ โมเดลไม่มีสิ่งที่จะช่วยในการระบุตำแหน่งของคำปัจจุบัน ว่าตอนนี้กำลังประมวลผลคำที่เท่าไรในประโยค คำไหนอยู่ด้านหน้า คำไหนอยู่ด้านหลัง จึงได้มีการเพิ่มส่วนของ Positional Embedding เข้าไป เพื่อช่วยให้โมเดลสามารถรับรู้ตำแหน่งของคำที่กำลังพิจารณาอยู่ ซึ่งใช้ค่า sin/cos ในการแทน position ต่างๆ ตามสมการด้านต่อไปนี้

$$\mathrm{PE}_{(\mathrm{pos},2\mathrm{i})} = \sin(\mathrm{pos}/10000^{2\mathrm{i}/\mathrm{d}_{\mathrm{model}}}) \quad (1)$$

$$\mathrm{PE}_{(\mathrm{pos},2\mathrm{i}+1)} = \cos(\mathrm{pos}/10000^{2\mathrm{i}/\mathrm{d}_{\mathrm{model}}}) \quad (2)$$

ขั้นตอนที่ 5: Multi-Head Attention

Multi-Head Attention คือกระบวนการที่มี Self-Attention ได้พร้อม ๆ กัน โดยเอาหลาย ๆ Self-Attention มาต่อขนานกันไปเพื่อให้โมเดลสามารถเรียนรู้ความสัมพันธ์ได้หลายมุมมองไปพร้อมๆ กัน ตัวอย่างเช่น “Bobby takes his cats to the park because they love climbing trees.” จะเห็นได้ว่า ‘cats’ มีความสัมพันธ์กับ ‘Bobby + takes’ ในฐานะกรรมของประโยค และคำว่า ‘cats’ ยังมีความสัมพันธ์กับ ‘love + climbing’ ในฐานะที่เป็นประธาน (ในที่นี้คือคำว่า ‘they’ ที่อ้างอิงกลับไปที่คำว่า ‘cats’) และในเชิงความหมาย ‘climbing trees’ มีความสัมพันธ์กับ ‘cats’ เพราะว่าเป็นสิ่งที่พวกมันชอบ แต่ก็ยังมีความหมายโดยนัยว่า ‘the park’ ต้องมี ‘trees’ เพื่อให้พวกมันปีนได้อีกด้วย จากตัวอย่างที่กล่าวข้างต้นทำให้เพียง 1 Self-Attention head ซึ่งสามารถเรียนรู้ได้แค่ความสัมพันธ์รูปแบบเดียว ไม่พอในการเรียนรู้ภาพรวมของทั้งประโยค

ขั้นตอนที่ 6 : Add & Norm

Add & Norm ในกระบวนการนี้มีการทำงานอยู่ทั้งหมด 2 ส่วนคือ Residual Connections (หรือที่รู้จักในชื่อ ResNet) คือการนำเวกเตอร์ก่อนหน้าที่จะเข้าสู่ Sublayer ซึ่งในที่นี้หมายถึง Multi-head Attention มาบวกกับเวกเตอร์ที่ถูกประมวลผลด้วย Sublayer มาแล้ว แต่ก่อนที่จะนำเวกเตอร์มาบวกกันได้นั้น จะต้องนำเวกเตอร์ที่ถูกประมวลผลด้วย Sublayer มาผ่านการทำ Dropout เสียก่อน โดย Dropout คือการสุ่มให้เวกเตอร์บางส่วนกลายเป็น 0

Computer Science Department Faculty of Informatics, Mahasarakham University

เพื่อแก้ปัญหา overfitting โมเดล โดยโมเดล Tramsformer นั้นตั้ง dropout rate = 0.1

ส่วนที่ 2 คือ Layer Normalization เมื่อได้ผลลัพธ์จากกระบวนการ Residual Connections ก็จะเข้าสู่กระบวนการ Layer Normalization

ขั้นตอนที่ 7 : Feed Forword

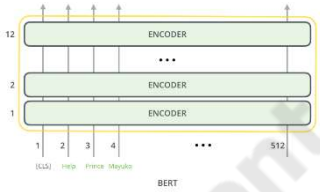
Feed Forword เป็นกระบวนการที่นำ Feed-forward neural networks ที่ถือเป็นโมเดลที่มีโครงสร้างที่เรียบง่ายที่สุด เพราะว่า การดำเนินการของข้อมูลจะเป็นไปในทิศทางเดียว ก็คือ รับข้อมูลจาก input layer แล้วส่งไปต่อไปยัง hidden layer เรื่อยๆ จนกระทั่งถึง output layer ก็จะหยุด โดยการทำ Feed-forward neural networks จำมี Activation function อยู่หนึ่งตัวนั้นคือ ฟังก์ชัน gelu

ภาพที่ 7 การทำงานของ Feed-forward neural networks

ขั้นตอนที่ 8 : Add & Norm

Add & Norm ขั้นตอนนี้ เป็นขั้นตอนเดียวกับขั้นตอนที่ 6 เพียงเปลี่ยนเวกเตอร์ที่ทำการบวกกันเป็น เวกเตอร์ก่อนจะเข้า Feed-forward neural networks กับ เวกเตอร์หลังเข้า Feed-forward neural networks เพียงเท่านั้น เมื่อจบขั้นตอนนี้ก็ถือเป็นการจบการสร้าง pre-training BERT ใน 1 Encoder แล้ว

โดยในการทำ pre-training BERT นั้นจะมี Encoder Block ซึ่งคือชั้นของ Encoder ที่ต้องเรียงกันทั้งหมด 12 ชั้น เพื่อให้โมเดลปรับค่าน้ำหนักของข้อมูลเอกสารเพื่อให้โมเดลเข้าใจโครงสร้างทางภาษา ลำดับของภาษา และการใช้ภาษา นั้นก็คือ Language Model นั้นเอง

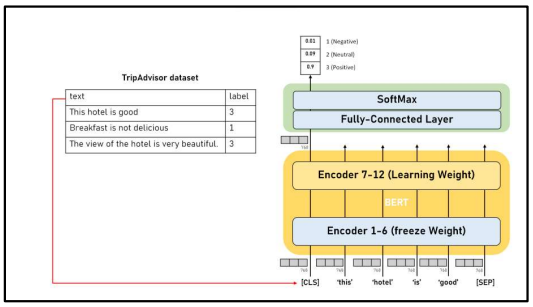


ภาพที่ 8 ชั้นของ Encoder ภายในโมเดลแบบ

BERT2.2 Fine-tuning BERT

Fine-tuning BERT คือการปรับโมเดลแบบ BERT ให้ทำงานเฉพาะ (specific task) ต่างๆได้ เพื่อที่จะให้โมเดลแบบ BERT สามารถทำงานด้านการจำแนกความรู้สึก (Sentiment Classification) ได้นั้น จะต้องนำข้อมูล ที่รวบรวมจาก Tripadvisor ที่เป็นชุดข้อมูลเอกสารแบบมีคลาสลาเบล เข้าสู่กระบวนการนี้ ในขณะเดียวกันโมเดลแบบ BERT จะถูกเชื่อมต่อเข้ากับโมเดลสำหรับการจำแนกความรู้สึก ซึ่งจะประกอบด้วย Fully-connected Layer และฟังก์ชัน SoftMax โดยโมเดลสำหรับการจำแนกความรู้สึกนี้ จะต่อกับเข้ากับโมเดลแบบ BERT ที่ตำแหน่งของโทเค็น [CLS] ซึ่งเป็นโทเค็นสำหรับงานจำแนกเอกสาร (Text Classification) ที่ encoder ชั้นบนสุด นอกจากนั้น โมเดลแบบ BERT ยังมีการแช่การเรียนรู้ของ encoder ชั้นที่ 1 – 6 ไม่ให้ปรับค่าน้ำหนักและพารามิเตอร์ไปในระหว่างการทำกระบวนการ Fine-tuning และเปิดให้ encoder ชั้นที่ 7 – 12 ปรับค่าน้ำหนักและพารามิเตอร์ไปพร้อมๆกับโมเดลสำหรับการจำแนกความรู้สึก เมื่อสิ้นสุดกระบวนการแล้วก็จะได้โมเดลสำหรับการจำแนกความรู้สึก (Sentiment Classification Model) เพื่อใช้ในงานเฉพาะที่ต้องการ

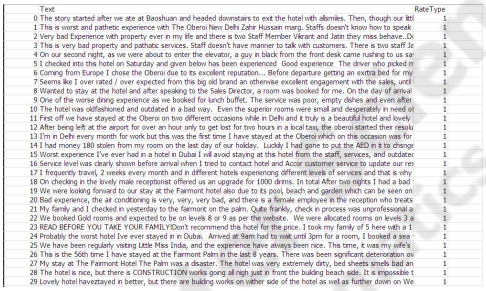
Computer Science Department
Faculty of Informatics, Mahasarakham University



ภาพที่ 9 ขั้นตอนการ Fine-tuning BERT สำหรับงานจำแนกความรู้สึกผลการวิจัย

1.ข้อมูลที่ใช้ในการทดสอบ

ข้อมูลที่ใช้การในทดสอบการสร้างโมเดลสำหรับการจำแนกความรู้สึกบทวิจารณ์โรงแรมนั้น จะเป็นชุดข้อมูลทดสอบ (Test set) ที่มีจำนวนทั้งสิ้น 9024 ข้อความ ที่เก็บอยู่ในรูปแบบของไฟล์ .csv



ภาพที่ 10 ตัวอย่างบทวิจารณ์โรงแรมที่ใช้ในการทดสอบ

2.ผลการทดลองของโมเดลที่ศึกษา

ตารางที่ 1 ผลการทดลองของโมเดลที่ศึกษา

		Actual Results		
		Positive	Neural	Negative
Prediction Results	Positive	1918	66	23
	Neural	63	1716	243
	Negative	11	164	1825

แสดงให้เป็นว่าการสร้างโมเดลเพื่อการจำแนกความรู้สึกจากบทวิจารณ์โรงแรมนั้นให้ผลลัพธ์ในมุมมองของค่าความระลึก ค่าความแม่นยำ และค่าความถูกต้องได้อย่างน่าพอใจ เนื่องจาก

ในมุมมองด้านค่า Recall นั้น BERT-base มีการเรียนรู้แบบต่อเนื่อง (Sequential Learning) ซึ่งช่วยให้โมเดลสามารถเข้าใจความหมายของประโยคหรือข้อความที่ซับซ้อนได้เป็นอย่างดี การเรียนรู้แบบต่อเนื่องช่วยให้ BERT-base สามารถจดจำและระบุลักษณะที่เกี่ยวข้องกับข้อความที่ต้องการจำแนกความรู้สึกได้อย่างแม่นยำ ซึ่งส่งผลให้มีค่า recall สูง เพราะโมเดลสามารถตรวจจับข้อความที่เกี่ยวข้องกับความรู้สึก (Sentiment) ได้หลายกรณี

ในมุมมองด้านค่า Precision นั้น เนื่องจาก BERT-base มีความสามารถในการเข้าใจและระบุลักษณะที่เกี่ยวข้องกับประโยคหรือข้อความที่ต้องการจำแนกความรู้สึก โมเดลสามารถสร้างข้อความที่ถูกต้องเกี่ยวกับความรู้สึกออกมาได้เป็นอย่างดี ซึ่งนำไปสู่การมีค่า precision สูง

ในมุมมองด้านค่า Accuracy นั้น BERT-base เป็นโมเดลที่มีความซับซ้อนและถูกฝึกสอนด้วยข้อมูลจำนวนมาก ซึ่งช่วยให้สามารถจำแนกข้อมูลความรู้สึกได้อย่างถูกต้องแม่นยำ ดังนั้นจึงส่งผลให้มีค่า accuracy สูง

3.การทดสอบเปรียบเทียบกับโมเดลการจำแนกความรู้สึกอื่นๆ

ในหัวข้อนี้จะนำโมเดลการจำแนกความรู้สึกที่สร้างจากอัลกอริทึมอื่นๆ มาเปรียบเทียบกับโมเดลการจำแนกความรู้สึกจากโมเดลแบบทรานสฟอร์มเมอร์แบบ BERT-base โดยโมเดลการจำแนกความรู้สึกที่สร้างจากอัลกอริทึมอื่นๆ นั้นจะสร้างขึ้นจากโมเดลชุดสอน และทดสอบประสิทธิภาพโมเดลด้วยข้อมูลชุดทดสอบเดียวกันกับโมเดลการจำแนกความรู้สึกจากโมเดลแบบทรานสฟอร์มเมอร์แบบ BERT-base และวัดประสิทธิภาพด้วยเกณฑ์เดียวกัน โดยโมเดลที่นำมาทดสอบมีทั้งหมด 4 รูปแบบดังนี้

Computer Science Department
Faculty of Informatics, Mahasarakham University

แบบที่ 1 โมเดลการจำแนกความรู้สึกจากเพื่อนบ้านใกล้เคียงที่สุด (K-Nearest Neighbor: KNN)

แบบที่ 2 โมเดลการจำแนกความรู้สึกจากซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM)

แบบที่ 3 โมเดลการจำแนกความรู้สึกจากป่าสุ่ม (Random Forest: RF)

แบบที่ 4 โมเดลการจำแนกความรู้สึกจากโมเดลแบบทรานสฟอร์มเมอร์แบบ roBERTa-base

ตารางที่ 2 สรุปผลการทดสอบโมเดลและเปรียบเทียบประสิทธิภาพของโมเดลแบบต่างๆ

โมเดล	Accuracy	ค่าความระลึก	ค่าความแม่นยำ	ค่า F1-score
BERT-base	0.9055	0.9125	0.8728	0.8922
		0.8487	0.8818	0.8649
		0.9557	0.9629	0.9593
	ค่าเฉลี่ย	0.9035	0.9041	0.9038
roBERTa-base	0.9382	0.9238	0.9232	0.9235
		0.9125	0.9088	0.9106
		0.9557	0.9629	0.9593
	ค่าเฉลี่ย	0.9307	0.9316	0.9311
KNN	0.5821	0.7066	0.4047	0.5146
		0.6853	0.3886	0.4960
		0.5122	0.9537	0.6665
	ค่าเฉลี่ย	0.6347	0.5823	0.5590
SVM	0.7789	0.6519	0.9519	0.7738
		0.8910	0.4933	0.6350
		0.9004	0.9031	0.9017
	ค่าเฉลี่ย	0.8144	0.7828	0.7702
RF	0.4799	0.7197	0.2606	0.3826
		0.4874	0.2889	0.3628
		0.4345	0.8932	0.5846
	ค่าเฉลี่ย	0.5472	0.4809	0.4433

จากตารางที่ 2 จะเห็นว่า roBERTa-base ให้ประสิทธิภาพดีกว่า BERT base, KNN, SVM และ Random Forest ใน sentiment classification เนื่องจากการฝึกสอนและการปรับปรุงที่ละเอียดและซับซ้อนของโมเดลใน roBERTa-base ช่วยให้โมเดลเรียนรู้และตีความข้อมูลได้ดียิ่งขึ้นและมีความสามารถในการจัดเรียงและจำแนกข้อมูลที่เหมาะสมมากยิ่งขึ้น เนื่องจาก

1.Pre-training อย่างละเอียด: roBERTa-base ได้รับการฝึกสอนในระดับคำศัพท์โดยรวม (word level) และระดับส่วนของคำ (subword level) โดยใช้ข้อมูลจำนวนมาก ซึ่งช่วยให้ระบบเรียนรู้และเข้าใจคำศัพท์และความหมายในประโยคได้ลึกซึ้งมากขึ้น

2. Masked Language Model (MLM): roBERTa-base มีการใช้ MLM ในการฝึกสอนโมเดลโดยการฝึกสอนโมเดลให้เดาคำที่ถูกซ่อนไว้ในประโยค ซึ่งช่วยให้โมเดลเรียนรู้ความสัมพันธ์ระหว่างคำและประโยคได้ดีกว่า

3. Dynamic Masking: roBERTa-base มีการใช้วิธี Dynamic Masking ในการฝึกสอน ซึ่งทำให้คำที่ถูกซ่อนไว้ในประโยคเปลี่ยนไปในแต่ละรอบการฝึกสอน ทำให้โมเดลมีการเรียนรู้ความสัมพันธ์ของคำในทางที่หลากหลายมากขึ้น

4.การทำซ้ำแบบสุ่ม (Random Iteration): roBERTa-base ได้ใช้การทำซ้ำแบบสุ่มในการฝึกสอนโดยการใช้ข้อมูลจากอินเทอร์เน็ตมากขึ้นและใช้กระบวนการสุ่มเพื่อสร้างตัวอย่างการฝึกสอนที่หลากหลายมากขึ้น

5.การปรับปรุงโครงสร้าง: roBERTa-base ได้รับการปรับปรุงในโครงสร้างที่ซับซ้อนขึ้น เช่น การเพิ่มจำนวนเลเยอร์ transformer, การปรับค่าพารามิเตอร์ต่างๆ ซึ่งช่วยให้โมเดลมีความสามารถในการจัดเรียงและจำแนกข้อมูลที่ดีกว่า

สรุปและอภิปรายผลการทดลอง

สรุปเกี่ยวกับการใช้ BERT-base ในการศึกษานี้พบว่าให้ประสิทธิภาพที่น่าพอใจ เพราะ BERT (Bidirectional Encoder Representations from Transformers) เป็นโมเดลการเรียนรู้ของเครื่องที่ใช้การแปลงคำ (word embeddings) และโครงข่าย

Computer Science Department
Faculty of Science, Maha Sarakham University

Transformer เพื่อเข้าใจและเรียนรู้ความหมายของประโยคหรือข้อความที่ให้มากขึ้น ดังนั้น BERT-base มีความสามารถในการจำแนกข้อมูลแบบ sentiment classification ที่ดีเนื่องจากคุณลักษณะต่อไปนี้:

1. ความสามารถในการเข้าใจความหมายของประโยค: BERT ถูกฝึกสอนด้วยการทำนายคำต่อไปในประโยคจากคำที่เหลือ โดยการใช้โครงข่าย Transformer ซึ่งสามารถรับลำดับของคำทั้งหมดในประโยคได้ จึงช่วยให้ BERT สามารถเข้าใจความหมายของประโยคและข้อความอย่างละเอียดได้เป็นอย่างดี ซึ่งเป็นประโยชน์ในการจำแนก sentiment classification เนื่องจากการจำแนกแสดงถึงความรู้สึกหรือทัศนคติที่ซับซ้อนของประโยคหรือข้อความที่ต้องการวิเคราะห์

2. การเรียนรู้ต่อเนื่อง: BERT ถูกฝึกสอนด้วยข้อมูลจำนวนมากและมีโครงสร้างที่ซับซ้อน ทำให้สามารถเรียนรู้ความสัมพันธ์ระหว่างคำและประโยคได้อย่างเหมาะสม โดยการใช้ตัวแทนคำที่ถูกเรียนรู้แล้ว (pre-trained word embeddings) ซึ่งรวมถึงบริบทและความสัมพันธ์ที่เกี่ยวข้อง

ดังนันานวิจัยฉบับนี้ จึงได้นำเสนอโมเดลสำหรับการจำแนกความรู้สึกบทวิจารโรงแรมภาษาอังกฤษจากโมเดลทรานส์ฟอร์มเมอร์แบบ BERT-base โดยเป็นการให้คะแนน 3 ระดับ คือ “1” (Positive) “3” (Neural) “1” (Negative) จากผลการทดสอบมีค่าความถูกต้องอยู่ที่ 90.55%

เอกสารอ้างอิง

[1] A. Ligthart, C. Catal, and B. Tekinerdogan. (2021). Systematic reviews in sentiment analysis: a tertiary study. Artificial Intelligence Review, 54: 4997-5053.

[2] N.N. Yusof, A. Mohamed and S. Abdul-Rahman. (2015). Reviewing Classification Approaches in Sentiment Analysis. International Conference on Soft Computing in Data Science, pp. 43-53.

[3] P. Lin and Z. Luo. (2020). A Survey of Sentiment Analysis Based on Machine Learning. International Conference on Natural Language Processing and Chinese Computing, pp. 372-387.

[4] S.T. Kokab, S. Asghar, and S. Naz. (2022). Transformer-based deep learning models for the sentiment analysis of social media data. Array, 14: 1-12.

[5] S. Kandi. (2018). Handling Out-of-Vocabulary Words in Natural Language Processing based on Context. Available at https://medium.com/analytics-vidhya/handling-out-of-vocabulary-words-in-natural-language-processing-based-on-context-4bbba16214d5, Accessed on 10 October 2022

[6] J. Nelson. (2020). The Difference Between Missing and Null Annotations. Available at https://blog.roboflow.com/missing-and-null-image-annotations/, Accessed on 10 October 2022

[7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, and I. Polosukhin. (2017). 31st Conference on Neural Information Processing Systems (NIPS 2017), pp. 1-11.

[8] Q. T. Nguyen, T. L. Nguyen, N. H. Luong and Q. H. Ngo, "Fine-Tuning BERT for Sentiment Analysis of Vietnamese Reviews," 2020 7th NAFOSTED Conference on Information and

Computer Science Department
Faculty of Informatics, Mahasarakham University

Computer Science (NICS), 2020, pp. 302-307, doi: 10.1109/NICS51282.2020.9335899.

[9] A. J. Nair, V. G and A. Vinayak, "Comparative study of Twitter Sentiment On COVID - 19 Tweets," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), 2021, pp. 1773-1778, doi: 10.1109/ICCMC51019.2021.9418320.

[10] C. Lee, Z. Gao and C. -L. Tsai, "BERT-Based Stock Market Sentiment Analysis," 2020 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan), 2020, pp. 1-2,doi:10.1109/ICCETaiwan49838.2020.9258102
.

[11] A. Nugumanova, Y. Baiburin and Y. Alimzhanov, "Sentiment Analysis of Reviews in Kazakh With Transfer Learning Techniques," 2022 International Conference on Smart Information Systems and Technologies (SIST), 2022,pp.1-6,doi: 10.1109/SIST54437.2022.9945811.

[12] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue et al. “HuggingFace’s Transformers: State-of-the-art Natural Language Processing”. arXiv:1910.03771., 2019.

[13] Z. A. Guven, "Comparison of BERT Models and Machine Learning Methods for Sentiment Analysis on Turkish Tweets," 2021 6th International Conference on Computer Science and Engineering (UBMK), 2021, pp. 98-101, doi: 10.1109/UBMK52708.2021.9559014.

[14] Y. Wu, M. Schuster, Z. Chen, Q. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, et al. 2016. Google’s neural machine translation system: Bridging the gap between human and machine translation. arXiv preprint arXiv:1609.08144.

[15] J. Devlin, M. Chang, K. Lee, K. Toutanova: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805,2018.

Computer Science Department
Faculty of Informatics, Mahasarakham University