ตัวแยกประเภทเชิงลึกสำหรับคำถามและการใช้เอไอเชิงอธิบาย สำหรับการตรวจพบปัญหาที่อาจเกิดขึ้นได้ในข้อมูลฝึก

โดย
นายอภินันท์ พีรชัยเดโช

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล แผน ก แบบ ก 2 ระดับปริญญามหาบัณฑิต
ภาควิชาคอมพิวเตอร์
มหาวิทยาลัยศิลปากร
ปีการศึกษา 2566
ลิขสิทธิ์ของมหาวิทยาลัยศิลปากร

ตัวแยกประเภทเชิงลึกสำหรับคำถามและการใช้เอไอเชิงอธิบาย สำหรับการตรวจพบปัญหาที่อาจเกิดขึ้นได้ในข้อมูลฝึก

โดย
นายอภินันท์ พีรชัยเดโช

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล แผน ก แบบ ก 2 ระดับปริญญามหาบัณฑิต
ภาควิชาคอมพิวเตอร์
มหาวิทยาลัยศิลปากร
ปีการศึกษา 2566
ลิขสิทธิ์ของมหาวิทยาลัยศิลปากร

DEEP QUESTION CLASSIFIER AND EXPLAINABLE AI FOR DETECTION OF POTENTIAL PROBLEMS IN TRAINING DATA

By

MR. Aphinan PEERACHAIDACHO

A Thesis Submitted in Partial Fulfillment of the Requirements

for Master of Science INFORMATION TECHNOLOGY AND DIGITAL INNOVATION

Department of COMPUTER SCIENCE

Silpakorn University

Academic Year 2023

Copyright of Silpakorn University

หัวข้อ	ตัวแยกประเภทเชิงลึกสำหรับคำถามและการใช้เอไอเชิงอธิบาย สำหรับการตรวจพบปัญหาที่อาจเกิดขึ้นได้ในข้อมูลฝึก
โดย	นายอภินันท์ พีรชัยเดโช
สาขาวิชา	เทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล แผน ก แบบ ก 2 ระดับปริญญามหาบัณฑิต
อาจารย์ที่ปรึกษาหลัก	ผู้ช่วยศาสตราจารย์ ดร. ทัศนวรรณ ศูนย์กลาง
อาจารย์ที่ปรึกษาร่วม	ผู้ช่วยศาสตราจารย์ ดร. ภิญโญ แท้ประสาทสิทธิ์

คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร ได้รับพิจารณาอนุมัติให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต

.....คณบดีคณะวิทยาศาสตร์

( )

พิจารณาเห็นชอบโดย

.....ประธานกรรมการ

(รองศาสตราจารย์ ดร. ศุภวงศ์ ทั่วรอบ)

.....อาจารย์ที่ปรึกษาหลัก

(ผู้ช่วยศาสตราจารย์ ดร. ทัศนวรรณ ศูนย์กลาง)

.....อาจารย์ที่ปรึกษาร่วม

(ผู้ช่วยศาสตราจารย์ ดร. ภิญโญ แท้ประสาทสิทธิ์)

.....ผู้ทรงคุณวุฒิภายใน

(ผู้ช่วยศาสตราจารย์ ดร. วีณาวดี ม่วงอ้น)

61318305 : เทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล แผน ก แบบ ก 2 ระดับปริญญามหาบัณฑิต
คำสำคัญ : การเรียนรู้เชิงลึก, การประมวลภาษาธรรมชาติ, ความสามารถในการเข้าใจ

นาย อภินันท์ พีรชัยเดโช: ตัวแยกประเภทเชิงลึกสำหรับคำถามและการใช้เอไอเชิงอธิบายสำหรับการตรวจพบปัญหาที่อาจเกิดขึ้นได้ในข้อมูลฝึก อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก : ผู้ช่วยศาสตราจารย์ ดร. ทัศนวรรณ ศูนย์กลาง

ปัจจุบันการประมวลภาษาธรรมชาติมีขอบเขตการใช้งานเฉพาะทางที่หลากหลาย ซึ่งปัญหาในงานเฉพาะทางที่พบได้บ่อยคือปริมาณข้อมูลมีจำกัด ทำให้มีความจำเป็นในการหาแบบจำลองและชุดข้อมูลที่เหมาะสมกับข้อจำกัดดังกล่าว สำหรับผู้ให้บริการแพล็ตฟอร์มด้านงานอีเว้นท์ ปัญหาที่พบคือมีคำถามอยู่จำนวนมากในฐานข้อมูล มีความซ้ำซ้อน และไม่ได้รับการจัดประเภทคำถาม งานวิจัยนี้จึงนำเสนอการพัฒนาแบบจำลองเชิงลึกสำหรับการจำแนกประเภทคำถามของงานอีเว้นท์ด้วยโครงข่ายประสาทเทียมแบบผสม CNN-BiLSTM ผลลัพธ์การทดลองแสดงให้เห็นว่าแบบจำลองที่เรานำเสนอมีประสิทธิภาพเหนือกว่าแบบจำลองอื่นๆอย่างเห็นได้ชัด นอกจากนี้เราได้นำเสนอวิธีการตรวจหาปัญหาที่อาจเกิดขึ้นกับชุดข้อมูลฝึกด้วยการใช้เอไอเชิงอธิบาย วิธีการของเราสามารถอธิบายผลลัพธ์การทำนายของแบบจำลองเพื่อช่วยให้ผู้วิจัยสามารถเข้าใจพฤติกรรมการทำงานของแบบจำลองได้มากยิ่งขึ้น และนำปัญหาที่พบไปพิจารณาเพื่อปรับปรุงคุณภาพของชุดข้อมูลฝึก เมื่อข้อมูลมีคุณภาพจะส่งผลให้แบบจำลองมีประสิทธิภาพการทำนายผลลัพธ์ได้ดียิ่งขึ้น

มหาวิทยาลัยศิลปากร

61318305 : Major INFORMATION TECHNOLOGY AND DIGITAL INNOVATION
Keyword : deep learning, natural language processing, interpretability

MR. Aphinan PEERACHAIDACHO : Deep Question Classifier and Explainable AI for Detection of Potential Problems in Training Data Thesis advisor : Assistant Professor Doctor Tasanawan Soonklang

Currently, the processing of natural language holds diverse applications with distinct limitations. A common issue in specialized tasks is the limited availability of data, necessitating the search for appropriate models and datasets that align with these constraints. For event platform service providers, a prevalent challenge is the abundance of unorganized questions in the database. These questions often exhibit repetition and lack proper categorization. This research presents the development of a deep learning model for question categorization within event-related content using a CNN-BiLSTM hybrid neural network. Experimental results demonstrate that the presented model consistently outperforms other existing models, exhibiting significant improvements in performance. Furthermore, a method is proposed to identify potential issues within the training dataset by utilizing interpretability through artificial intelligence. This approach facilitates the explanation of the model's prediction outcomes, aiding researchers in better understanding the model's behavior. This, in turn, enables the researchers to analyze and address the model's performance more effectively. As the dataset quality improves, it enhances the model's predictive capabilities, resulting in better prediction outcomes.

กิตติกรรมประกาศ

วิทยานิพนธ์เล่มนี้สำเร็จสมบูรณ์ได้ด้วยดีเพราะได้รับความกรุณา แนะนำ ให้คำปรึกษา และแก้ไขข้อบกพร่องต่างๆ ด้วยความเอาใจใส่อย่างดียิ่งจาก อ.ดร.ภิญโญ แท้ประสาทสิทธิ์ ที่ได้ให้ความช่วยเหลือตั้งแต่เริ่มต้นจนสำเร็จเสร็จสิ้น ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูง

ขอขอบคุณบริษัท ซิปอีเว้นท์ จำกัด สำหรับการสนับสนุนชุดข้อมูลคำถามของงานอีเว้นท์ ซึ่งเป็นสิ่งสำคัญเป็นอย่างยิ่งสำหรับการทำงานวิจัยนี้ให้สำเร็จไปได้ด้วยดี

ขอขอบคุณคณาจารย์ประจำภาควิชาคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากรที่ให้ความรู้ ความเมตตา ในการศึกษาระดับปริญญาโท ทำให้ผู้วิจัยสามารถทำงานวิจัยนี้สำเร็จไปได้ด้วยดี

ขอขอบคุณภาควิชาคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากรที่ให้โอกาสในการศึกษาต่อในระดับปริญญาโท

คุณค่าทั้งหลายที่ได้รับจากการทำวิทยานิพนธ์เล่มนี้ ผู้เขียนขอมอบเป็นกตัญญูกตเวทีแด่บิดามารดา และบูรพาจารย์ที่เคยอบรมสั่งสอน รวมทั้งผู้มีพระคุณทุกท่าน

อภินันท์ พีรชัยเดโช

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

บทที่ 1 1

บทนำ 1

ความเป็นมาและความสำคัญของปัญหา 1

วัตถุประสงค์ของงานวิจัย 2

สมมติฐานของงานวิจัย 2

ขอบเขตการวิจัย 2

ประโยชน์ที่คาดว่าจะได้รับ 3

บทที่ 2 4

งานวิจัยที่เกี่ยวข้อง 4

Deep Learning Model Used in Text Classification 4

News Text Classification Based on Improved Bi-LSTM-CNN 7

Text classification based on hybrid CNN-LSTM hybrid model 10

Question Classification Based on MAC-LSTM 13

บทที่ 3 19

ทฤษฎีที่เกี่ยวข้อง 19

การแปลงคำเป็นตัวเลข (Word Embedding) 19

การเรียนรู้เชิงลึก (Deep Learning) 19

โครงข่ายประสาทแบบคอนโวลูชัน (Convolutional Neural Network)..... 19
โครงสร้างเครือข่ายประสาทเทียมหน่วยความจำระยะสั้นแบบยาวแบบสองทิศทาง (Bidirectional Long Short Term Memory)..... 19
เทคนิคสำหรับการปรับปรุงความเร็วสมรรถนะและความเสถียรของเครือข่ายประสาทเทียม..... 20
(Batch Normalization)..... 20
การอธิบายคำทำนายของแบบจำลองด้วย SHAP (SHapley Additive exPlanations)..... 20
บทที่ 4..... 21
วิธีการทดลอง..... 21
ลักษณะข้อมูลที่ใช้..... 21
การเตรียมข้อมูลก่อนประมวลผล..... 22
การเรียนรู้เชิงลึกของแบบจำลอง..... 22
การปรับพารามิเตอร์..... 23
การอธิบายคำทำนายของตัวแบบแยกประเภทคำถาม..... 25
การตรวจหาปัญหาที่อาจเกิดขึ้นจากข้อมูลฝึก..... 26
การปรับปรุงคุณภาพของข้อมูลฝึกจากปัญหาที่ตรวจพบ..... 27
การวัดประสิทธิภาพ..... 31
บทที่ 5..... 32
ผลการดำเนินการวิจัย..... 32
ผลการทดสอบแบบจำลอง..... 32
การเปรียบเทียบการใช้ Batch Normalization..... 33
การตามหาตัวอย่างจากข้อมูลทดสอบที่เครื่องทายผิดในลักษณะที่มั่นใจมาก..... 35
การตามหาประโยคที่มีความคล้ายกันสูงระหว่างตัวอย่างที่ทายผิด..... 44
การเปรียบเทียบประโยคที่มีความคล้ายกันสูงระหว่างตัวอย่างที่ทายผิด..... 49
การปรับปรุงข้อมูลชุดฝึกจากการหาประโยคที่มีความคล้ายกันสูงระหว่างตัวอย่างที่ทายผิด..... 51

การแสดงสถิติการปรับปรุงคุณภาพของชุดข้อมูลฝึกหลังจากตรวจพบปัญหา..... 53
การเปรียบเทียบประสิทธิภาพของแบบจำลองก่อนและหลังปรับปรุงคุณภาพของชุดข้อมูลฝึก 53
บทที่ 6..... 55
สรุปผลการดำเนินงานวิจัยและข้อเสนอแนะ 55
รายการอ้างอิง 56
ประวัติผู้เขียน..... 59

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันการประมวลประโยคในภาษาธรรมชาติเพื่อทำความเข้าใจเนื้อหาด้วยคอมพิวเตอร์เป็นสิ่งที่มีการใช้งานที่หลากหลาย แต่การประมวลประโยคก็มีขอบเขตการใช้งานเฉพาะทางที่หลากหลาย เช่น การทำความเข้าใจความต้องการของนักศึกษาจากข้อความที่ส่งถึงผู้สอน การจำแนกประเภทเนื้อหาซึ่งมีที่มาและชนิดที่แตกต่างจากที่พบในชีวิตประจำวัน ซึ่งปัญหาในงานเฉพาะทางที่พบได้บ่อยคือปริมาณข้อมูลจะมีจำกัด ทำให้มีความจำเป็นที่ในการหาตัวแบบและชุดข้อมูลที่เหมาะสมกับข้อจำกัดดังกล่าว

หากกล่าวถึงการประมวลประโยคเพื่อการใช้งานเฉพาะทางอีกอย่างหนึ่ง ในช่วงหลายปีที่ผ่านมา ในสังคมเรามีการจัดงานอีเว้นท์ การสัมมนาหรือการพบปะ (Meeting) ทั้งแบบออนไลน์และแบบออนไซต์จำนวนมาก แต่ละงานนั้นมักมีแบบสอบถามเพื่อเก็บข้อมูลและความเห็นของผู้เข้าร่วมงาน ซึ่งผู้จัดงานมักจะมีคำถามที่แตกต่างกันไป สำหรับผู้ทำแพล็ตฟอร์มด้านข้อมูลข่าวสารการจัดงานจะเป็นผู้รับคำถามเหล่านั้นมาจากผู้จัดงาน หรือในบางครั้งก็จะเป็นผู้ร่วมสร้างแบบสอบถามด้วย ปัญหาที่พบในเรื่องของการจัดทำแบบสอบถามก็คือปริมาณคำถามอาจจะมีมากเกินไปรวมถึงมีความซ้ำซ้อนและทุกครั้งที่มีการจัดงาน คำถามในแบบสอบถามก็ต้องถูกจัดทำขึ้นมาใหม่ ทางผู้จัดทำแพล็ตฟอร์มจึงมีความคิดว่าจะเป็นการดีหากสามารถใช้คำถามเดิมได้มากที่สุด ไม่ซ้ำซ้อน รวมถึงอาจจะทำการตั้งคำถามตามบุคคล (Personalized Question) เพื่อให้ได้ข้อมูลตรงตามวัตถุประสงค์และประหยัดเวลาในการทำและตอบแบบสอบถาม

อย่างไรก็ตาม การจะทำเช่นนั้นได้ ผู้ทำระบบจะต้องแบ่งแยกประเภทคำถามให้ได้เสียก่อน ซึ่งรวมไปถึงความพร้อมในการเปลี่ยนแปลงเรื่องจำนวนและประเภทของคำถามหากมีการจัดกลุ่มธุรกิจหรือประเภทงานต่างไปจากเดิม การสร้างและฝึกตัวแบบด้วยข้อมูลที่จำกัดจึงเป็นเรื่องที่ท้าทายและการปรับปรุงคุณภาพของข้อมูลฝึก รวมถึงการทำความเข้าใจความผิดพลาดในการทายผลของตัวแบบจึงเป็นสิ่งที่สำคัญ

งานวิจัยนี้จึงถูกจัดทำขึ้นเพื่อศึกษาตัวแบบที่มีความแม่นยำในการจำแนกประเภทคำถามในงานอีเว้นท์ และทำการศึกษาหาข้อจำกัดหรือปัญหาในชุดข้อมูลฝึกจากการใช้เอไอเชิงอธิบาย

(Explainable AI) เพื่อทำให้สามารถตรวจพบข้อจำกัดหรือปัญหาบางประเภทได้อย่างมีประสิทธิภาพ และสามารถนำผลการตรวจพบไปพิจารณาปรับปรุงคุณภาพของข้อมูลฝึกสำหรับฝึกตัวแบบให้ดีขึ้น

วัตถุประสงค์ของงานวิจัย

1. ศึกษาหาตัวแบบการเรียนเชิงลึกที่สามารถจำแนกประเภทคำถามภาษาไทยในงานอีเว้นท์ได้อย่างแม่นยำแม้ข้อมูลฝึกอาจจะไม่มากนัก (โดยเฉลี่ย < 400 ตัวอย่างฝึกต่อประเภท)

2. คิดค้นวิธีการใช้เอไอเชิงอธิบายสำหรับการค้นหาตัวอย่างที่เครื่องมีปัญหาในการแยกประเภทและเชื่อมโยงไปหาสาเหตุจากตัวอย่างฝึกได้

3. นำเสนอแนวคิดของการใช้เอไอเชิงอธิบายในการค้นหาปัญหาในตัวอย่างฝึกที่มีศักยภาพในการนำไปใช้งานได้อย่างเป็นระบบและลดภาระจากฝั่งมนุษย์ในการค้นหาและปรับปรุงตัวอย่างฝึก

สมมติฐานของงานวิจัย

การใช้ตัวแบบชนิดคอนโวลูชันหนึ่งมิติร่วมกับ Bidirectional LSTM (BiLSTM) มีความเหมาะสมในการแยกชนิดคำถามซึ่งเป็นประโยคที่ไม่ยาวนักและเราสามารถปรับปรุงคุณภาพของข้อมูลฝึกได้อย่างเป็นระบบด้วยการใช้เอไอเชิงอธิบายเพื่อวิเคราะห์ความผิดพลาดในการจำแนกประเภทพร้อมเชื่อมโยงกลับไปที่ตัวอย่างฝึก เพื่อหาสาเหตุของความผิดพลาดที่อาจจะเกี่ยวข้องกับตัวอย่างฝึก

ขอบเขตการวิจัย

1. ใช้คำถามภาษาไทยของงานอีเว้นท์ในการฝึกสอนแบบจำลองการเรียนรู้เชิงลึก
2. ชุดข้อมูลฝึกพื้นฐานมาจากบริษัทผู้ให้บริการแพลตฟอร์มด้านการจัดการอีเว้นท์
3. ตัวแบบจะทำการแบ่งคำถามออกเป็น 8 ประเภท ดังนี้
    3.1. ร่างกาย (Anatomy)
    3.2. สังคม (Social)
    3.3. การเงิน (Finance)
    3.4. การงาน (Work)
    3.5. การศึกษา (Education)
    3.6. วิถีการดำเนินชีวิต (Lifestyle)
    3.7. ช่องทางการได้รับข่าวสาร (Channel)

3.8. อื่นๆ (Other)

4. การตรวจหาปัญหาที่อาจเกิดขึ้นได้ในข้อมูลฝึกจะพิจารณาจากสิ่งที่แสดงออกมาทางการอธิบายผลลัพธ์เป็นหลัก และจะเน้นที่กระบวนการที่สามารถทำได้อย่างเป็นระบบ แต่อาจจะยังไม่ได้อยู่ในรูปแบบที่สามารถทำได้อย่างอัตโนมัติโดยปราศจากมนุษย์

ประโยชน์ที่คาดว่าจะได้รับ

สามารถจำแนกคำถามของงานอีเว้นท์ได้ถูกต้อง โดยการจำแนกคำถามของงานอีเว้นท์เป็นสิ่งที่สำคัญต่อบริษัทผู้ให้บริการแพลตฟอร์มด้านการจัดการอีเว้นท์ โดยทำให้การสร้างชุดคำถามในงานอีเว้นท์เป็นไปอย่างเป็นระบบและมีประสิทธิภาพ ได้ชุดคำถามที่กะทัดรัดแต่ครอบคลุมจุดประสงค์ของการหาข้อมูล นอกจากนี้กระบวนการค้นหาปัญหาในชุดข้อมูลฝึกจะมีประโยชน์ด้านการปรับปรุงข้อมูลฝึก ทำให้ตัวแบบปัญญาประดิษฐ์มีสมรรถนะที่สูงยิ่งขึ้น และยังนำไปประยุกต์ในการฝึกตัวแบบด้านการวิเคราะห์ภาษาได้หลากหลาย ไม่จำกัดเฉพาะการแยกประเภทคำถามเท่านั้น

มหาวิทยาลัยศิลปากร

บทที่ 2

งานวิจัยที่เกี่ยวข้อง

Deep Learning Model Used in Text Classification

ในปี 2018 Jingjing Cai, Jianping Li, Wei Li และ Ji Wang [1] นำเสนอกระบวนการจำแนกข้อความและมุ่งเน้นไปที่แบบจำลองการเรียนรู้ลึกที่ใช้ในการจำแนกข้อความ โดยมีการนำเสนอแบบจำลอง 3 แบบ

แบบที่หนึ่ง TextCNN หรือ CNN (Convolutional Neural Network) ใช้เพื่อดึงข้อมูลสำคัญคล้ายกับ n-gram ในประโยค CNN ถูกเสนอโดย Yoon Kim ในปี 2014 โมเดลประกอบด้วยส่วนต่อไปนี้

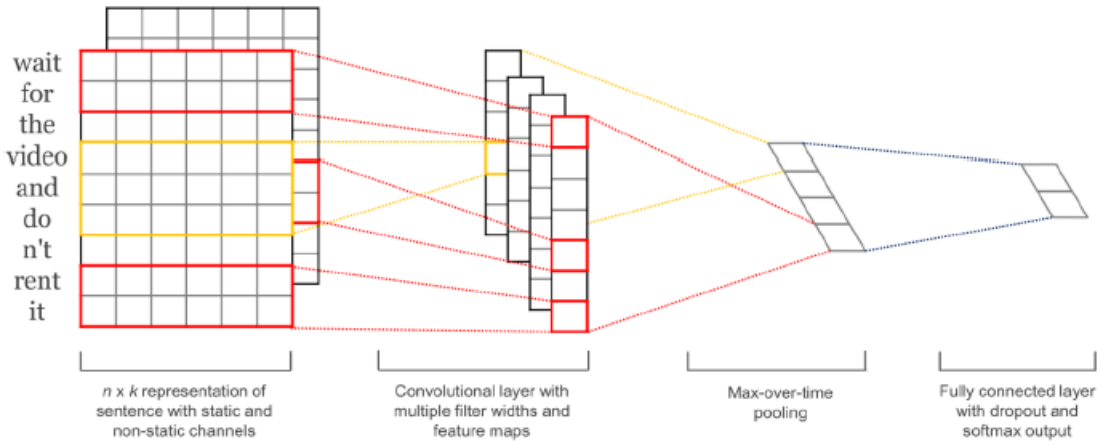
ส่วนที่ 1: เลเยอร์อินพุต ข้อมูลข้อความที่ถูกประมวลผลล่วงหน้าจะถูกป้อนเข้าสู่แบบจำลอง

ส่วนที่ 2: การฝังเลเยอร์ การแยกคุณสมบัติข้อความ

ส่วนที่ 3: เลเยอร์ Convolution แต่ละชั้นเลเยอร์ convolution ถูกกำหนดโดยตัวกรองที่มีขนาดต่างกันในผลลัพธ์ของการได้รับฟีเจอร์แมปมากมาย

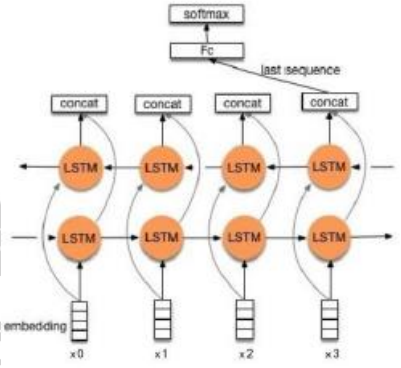
ส่วนที่ 4: เลเยอร์สูงสุดที่รวมกันได้ มิติของชั้นของการบิดลดลง

ส่วนที่ 5: ชั้น SoftMax แสดงผลความน่าจะเป็นของแต่ละหมวดหมู่ในงานหลายหมวดหมู่



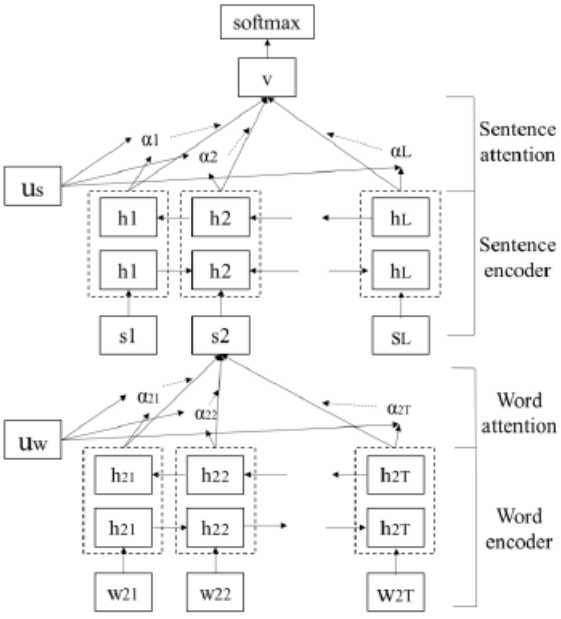
ภาพที่ 1 โครงสร้างแบบจำลอง CNN

แบบที่สอง TextRNN หรือ RNN (Recurrent neural network) โดยหนึ่งในปัญหาที่ใหญ่ที่สุดของ CNN คือขนาดของตัวกรองคือการกำหนดค่าตายตัว ในอีกด้านหนึ่งเป็นไปไม่ได้ที่จะสร้างแบบจำลองข้อมูลลำดับที่ยาวขึ้น ในทางกลับกันการปรับขนาดตัวกรองเกินขนาดเป็นเรื่องที่ไม่ดีนัก อย่างไรก็ตาม TextRNN หรือ RNN แบบสองทิศทาง (LSTM แบบสองทิศทาง) สามารถจับภาพข้อมูล "n-gram" แบบสองทิศทางที่มีความยาวผันแปรได้



ภาพที่ 2 โครงสร้างแบบจำลอง RNN

แบบที่สาม แบบจำลอง TextRNN และกลไกความสนใจ ถูกเพิ่มเข้าไปในแบบจำลองซึ่งสามารถแก้ปัญหาการพึ่งพาข้อความในระยะยาวได้ โดยนำเสนอการมีส่วนร่วมของแต่ละคำกับผลลัพธ์และสร้างกรอบการทำงานของโมเดล Seq2Seq



ภาพที่ 3 โครงสร้างแบบจำลอง RNN + Attention

โดยชุดข้อมูลที่นำมาใช้คือชุดข้อมูลข่าว Sohu ซึ่งมาจากข้อมูลข่าวของ 18 ช่องทางรวมถึงในประเทศต่างประเทศกีฬาสังคมและความบันเทิงตั้งแต่เดือนมิถุนายน 2555 ถึงเดือนกรกฎาคม 2555 และให้ข้อมูล URL และข้อมูลร่างกาย ข้อมูลแต่ละรายการอยู่ในประเภทเดียวเท่านั้นดังนั้นจึงเป็นปัญหาแบบป้ายเดี่ยวและการจำแนกหลายหมวดหมู่

ตารางที่ 1 ผลลัพธ์การเปรียบเทียบแบบจำลอง TextCNN, TextRNN, TextRNN + Attention

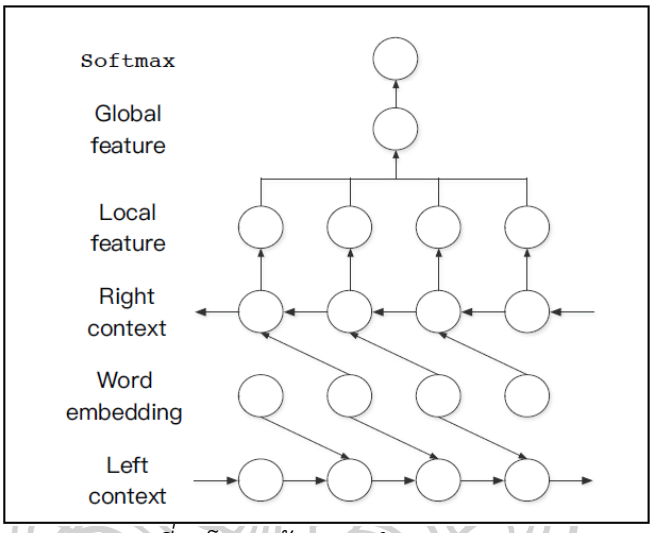
model	precision	Training time
TextCNN	0.8534	77s/epoch *2 epoch
TextRNN	0.8273	164s/epoch *2 epoch
TextRNN + Attention	0.8456	149s/epoch *2 epoch

โดยจากผลการทดลองนั้นแสดงให้เห็นในตารางที่ 1 ประเภทของข้อความสามารถแบ่งออกเป็นข้อความยาวและข้อความสั้นตามความยาวของประโยค เหตุผลคือการพิจารณาข้อจำกัดของรูปแบบการเรียนรู้เชิงลึกที่แตกต่างกัน ซึ่งข้อความที่เป็นลำดับและมีบริบทโมเดล RNN สามารถจัดการกับข้อมูลประเภทนี้ได้เป็นอย่างดี แต่ RNN ไม่สามารถทำงานแบบคู่ขนานได้ดี ใช้เวลาฝึกนานเมื่อประมวลผลข้อความที่มีคำมากกว่าสิบคำขึ้นไป ดังนั้นจึงเหมาะสำหรับการประมวลผลข้อความแบบสั้น ในขณะที่ CNN ถูกนำมาใช้ในการจัดหมวดหมู่ข้อความที่ยาวเพราะสามารถทำงานแบบคู่ขนานได้ดีและใช้เวลาน้อยกว่า

งานวิจัยดังกล่าวแสดงให้เห็นถึงความแตกต่างที่ชัดเจนของแบบจำลอง CNN และ RNN มีที่จุดเด่น จุดด้อยสำหรับข้อความในลักษณะต่างๆ แต่แบบจำลองที่นำมาแสดงในงานวิจัยดังกล่าวนั้นเป็นแบบดั้งเดิม ยังไม่ครอบคลุมแบบจำลองไฮบริดแบบสมัยใหม่

News Text Classification Based on Improved Bi-LSTM-CNN

ในปี 2018 Chenbin Li, Guohua Zhan และ Zhihua Li [2] นำเสนอการจำแนกประเภทข่าวด้วยเทคนิค Bi-LSTM-CNN โดยแบบจำลองในบทความนี้ส่วนใหญ่จะประกอบด้วย Bi-LSTM ประกอบด้วยคำเวกเตอร์และซ้ายและขวา บริบทคุณลักษณะท้องถิ่นคุณลักษณะระดับโลกและเลเยอร์ softmax ในขณะเดียวกันคำว่า vector ได้รับการฝึกอบรมโดย Word2Vec โดยจะแสดงให้เห็นโครงสร้างของแบบจำลองในภาพที่ 4



ภาพที่ 4 โครงสร้างแบบจำลอง Bi-LSTM

บทความนี้จะปรับแบบจำลองให้เหมาะสมเพื่อเปลี่ยนเลเยอร์ LSTM ทิศทางเดียวไปยังเลเยอร์ LSTM สองทิศทางเป็นเครือข่ายประสาทที่เกิดขึ้นอีก LSTM ไปข้างหน้าอ้างถึงคลังข้อมูลการอ่านตามลำดับและ LSTM ถอยหลังหมายถึงอ่านคลังข้อมูลในลำดับย้อนกลับ โครงสร้างดังกล่าวไม่เพียงแต่พิจารณาความหมายไปข้างหน้า แต่ยังพิจารณาความหมายของคำสั่งย้อนกลับซึ่งช่วยเพิ่มการแสดงออกของความหมายของข้อความ เวกเตอร์คำที่แต่ละตำแหน่งมีการแสดงออกที่ได้รับจาก LSTM ในสองทิศทางและบริบทด้านซ้ายและขวา ผ่านเลเยอร์ Bi-LSTM นี้โดยใช้ Convolution รวมบริบททางซ้ายและขวาของตำแหน่งปัจจุบันกับรูปแบบการแสดงใหม่ของคำที่รวมกัน หลังจากได้รับรูปแบบใหม่ของของคำจะส่งผลลัพธ์ไปยัง Max Pooling โดยจะแปลงข้อความที่มีความยาวต่างกันเวกเตอร์ที่มีความยาวคงที่และสามารถรวบรวมข้อมูลได้ตลอดข้อความ สุดท้ายหลังจากประมวลผลฟังก์ชั่น softmax แล้วจำนวนเอาต์พุตจะถูกแปลงเป็นความน่าจะเป็น

ชุดข้อมูลที่ใช้ในการทดลองนี้เป็นชุดย่อยของ THUC ข่าวการฝึกอบรมและการทดสอบ โดยเลือกข่าวจากสิบประเภทของกีฬา, การเงิน, อสังหาริมทรัพย์, บ้าน, การศึกษา,เทคโนโลยี, แฟชั่น,

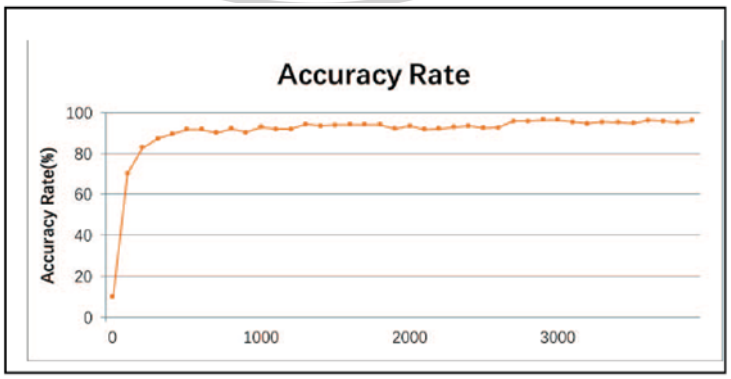
การเมือง, เกมและความบันเทิงเป็นข้อมูลการทดลอง คลังข้อมูลมี 65,000 ซึ่งชุดฝึกอบรมประกอบด้วย 50,000 ชุดการตรวจสอบรวมถึง 5,000 และชุดทดสอบรวม 10,000

ในบทความนี้วิธีการจำแนกประเภทของ TF-IDF, SVM, CNN, LSTM และ Bi-LSTM-CNN ตามตารางที่ 2 ผลการจำแนกประเภทจะแสดงในตาราง ตัวบ่งชี้หลักสำหรับการประเมินการจำแนกข้อความคือต่อไปนี้: อัตราความแม่นยำ, อัตราการสูญเสียและคะแนน F1

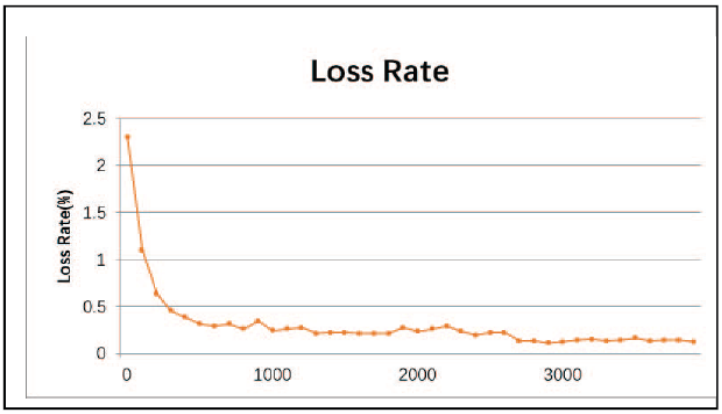
ตารางที่ 2 ผลลัพธ์การเปรียบเทียบแบบจำลอง TF-IDF, SVM, CNN, LSTM และ Bi-LSTM-CNN

Model	The comparison experiments results		
	Accuracy	Loss	F1
TF-IDF	90.27	0.31	0.86
SVM	93.49	0.25	0.92
LSTM	94.26	0.21	0.94
CNN	95.61	0.15	0.96
Bi-LSTM-CNN	96.45	0.11	0.99

เพื่อหลีกเลี่ยงการจำข้อมูลในกระบวนการทดลองจึงใช้วิธี Cross-validation เพื่อดำเนินการกับวัตถุทดสอบที่แตกต่างและวัตถุฝึกอบรม ความแม่นยำและอัตราการสูญเสียของการโหลดรูปแบบการฝึกอบรมล่วงหน้าในระหว่างกระบวนการฝึกอบรมจะแสดงในภาพที่ 2 และภาพที่ 3 มันสรุปได้จากรูปที่ความแม่นยำของแบบจำลองในบทความนี้ค่อยๆเพิ่มขึ้นและอัตราการสูญเสียลดลงอย่างต่อเนื่อง เมื่อฝึกเสร็จข้อมูลถึง 1,800 ความแม่นยำและอัตราการสูญเสียมีแนวโน้มที่จะมีเสถียรภาพ



ภาพที่ 5 อัตราความแม่นยำ



ภาพที่ 6 อัตราการสูญเสีย

จากการวิเคราะห์ข้อมูลข้างต้นของตารางที่ 2 แสดงให้เห็นว่าการเรียนรู้ลึกมีประสิทธิภาพเหนือกว่าวิธีการแบบดั้งเดิม แบบจำลองของงานวิจัยนี้ดีกว่าผลการเปรียบเทียบอื่นๆ การทดลองในแง่ของอัตราความแม่นยำอัตราการสูญเสียและ F1-score โดยใช้แบบจำลอง Bi-LSTM-CNN มีประสิทธิภาพของการจำแนกได้รับการปรับปรุงโดย 0.84% สามารถอธิบายได้ว่าเมื่อใช้ Bi-LSTM-CNN จะสามารถจับภาพคุณสมบัติได้อย่างแม่นยำยิ่งขึ้นและมีการรบกวนน้อยลง ยิ่งไปกว่านั้นแบบจำลองนี้ยังสามารถรักษาข้อมูลเชิงบริบท ความหมายของข้อความได้ดีและสามารถแสดงความแม่นยำมากขึ้นเพื่อให้ความถูกต้องของข้อความข่าวสามารถจัดหมวดหมู่ได้ดีขึ้น

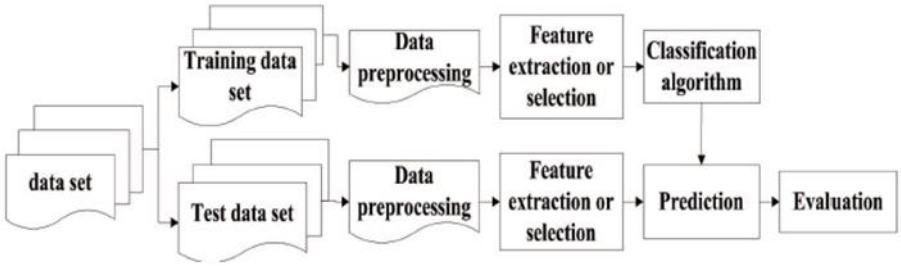
จากการทดลองดังกล่าวแสดงให้เห็นว่าแบบจำลองของผู้วิจัยนั้นให้ผลลัพธ์ที่ดีกว่าวิธีการอื่นจากภาพที่ 5 และภาพที่ 6 ทั้งอัตราความแม่นยำ อัตราการสูญเสีย และค่า F1 แต่เมื่อดูถึงเสถียรภาพของแบบจำลองนั้นจะต้องรอถึง 1,800 epoch จึงจะสามารถให้ผลลัพธ์ที่มีเสถียรภาพได้ ซึ่งหากแบบจำลองนี้มีการใช้งาน Batch Normalization ในแบบจำลองจะสามารถช่วยเร่งให้แบบจำลองมีเสถียรภาพรวดเร็วมากยิ่งขึ้น

Text classification based on hybrid CNN-LSTM hybrid model

ในปี 2018 Xiangyang She และ Di Zhang [3] นำเสนอการจำแนกข้อความโดยใช้แบบจำลองไฮบริดของ CNN-LSTM

กระบวนการการแบ่งหมวดหมู่ข้อความพื้นฐานจะแสดงตามภาพที่ 7

1. การเตรียมข้อมูลก่อนการประมวลผล คือ ลบคำหยุด เครื่องหมายวรรคตอน ฯลฯ ออกจากข้อความต้นฉบับทำการประมวลผลคำ การตัดคำภาษาจีนและดำเนินการแทนค่าคุณลักษณะเวกเตอร์บนข้อความ โดยทั่วไปมักใช้ VSM extraction
2. คุณสมบัติการสกัด (การแยกคุณลักษณะ) คือ แปลงข้อมูลมิติสูงให้ต่ำลง (การเลือกคุณสมบัติ) และเลือกคุณลักษณะที่ดีที่สุด
3. วิธีการจำแนกประเภท: ใช้อัลกอริทึมการจำแนก



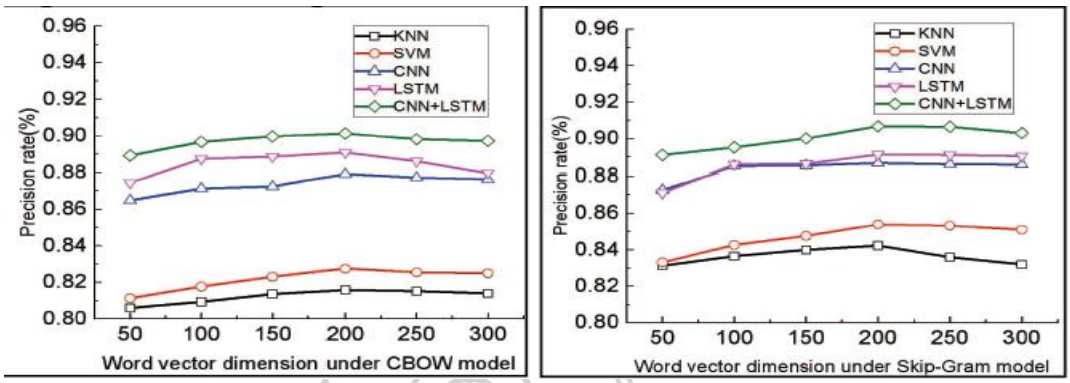
ภาพที่ 7 กระบวนการการแบ่งหมวดหมู่ข้อความพื้นฐาน

ข้อมูลการทดลองมาจากชุดข้อมูลข่าวจีนที่เผยแพร่โดย Sogou แล็บบนอินเทอร์เน็ต เนื้อหาคลังข้อมูลในชุดข้อมูลส่วนใหญ่มี 12 หมวดหมู่ ซึ่งมาจากข่าวคลังของแต่ละเว็บไซต์ข่าวแต่ละแห่งมี 2,000 ข้อความในแต่ละหมวดหมู่ โดยแสดงให้เห็นตามตารางที่ 3

ตารางที่ 3 ชุดข้อมูลข่าวจีนที่เผยแพร่โดย Sogou แล็บ

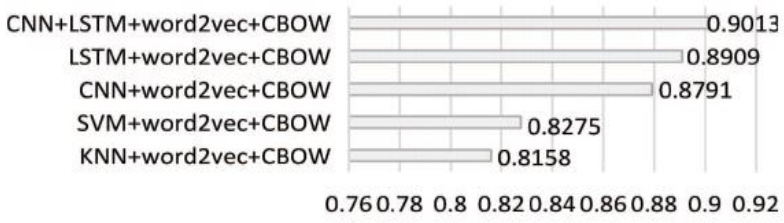
Category	Number of texts	Category	Number of texts
car	2000	education	2000
Finance	2000	Recruitment	2000
IT	2000	culture	2000
health	2000	military	2000
sports	2000	entertainment	2000
tourism	2000	fashion	2000

มีการทดลองใช้ CBOW และ Skip-Gram กับ Word2Vec เพื่อตั้งค่าขนาดคำต่างๆของเวกเตอร์เพื่อใช้เปรียบเทียบในการทดลองนี้

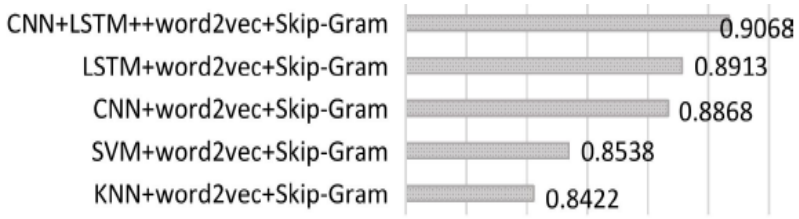


ภาพที่ 8 เปรียบเทียบการตั้งค่าขนาดเวกเตอร์

การทดลองใช้ห้าวิธีที่แตกต่างกันสำหรับการจัดหมวดหมู่ข้อความโดยมีการเปลี่ยนขนาดเวกเตอร์ของคำ จากภาพที่ 8 คุณภาพการจัดหมวดหมู่ก็มีการเปลี่ยนแปลงไปตามขนาดเวกเตอร์ของคำนั้นด้วย เมื่อเวกเตอร์ของคำมีขนาด เท่ากับ 200 จะให้ความแม่นยำสูงสุดและให้ผลการจำแนกดีที่สุด อาจเป็นเพราะขนาดเวกเตอร์ของคำนั้นมีขนาดที่เล็กมาก หากใช้ขนาดเวกเตอร์ที่มีขนาดใหญ่จะทำให้คุณลักษณะต่างๆของข้อความนั้นหายไป ดังนั้นเราจึงใช้ขนาดเวกเตอร์ของคำเท่ากับ 200 เพื่อดำเนินการทดสอบและใช้อัตราความแม่นยำในการประเมินผล



ภาพที่ 9 การเปรียบเทียบแบบจำลอง ที่ใช้ CBOW + WORD2VEC



ภาพที่ 10 การเปรียบเทียบแบบจำลอง ที่ใช้ SKIP-GRAM + WORD2VEC

จากภาพที่ 9 และภาพที่ 10 แสดงให้เห็นว่าความแม่นยำของแบบจำลอง CNN-LSTM สูงกว่าอีกสี่วิธีเพราะรุ่นไฮบริดไม่เพียง แต่จับความสามารถของ CNN แยกคุณลักษณะท้องถิ่น แต่ LSTM สามารถรักษาข้อมูลประวัติและแยกการอ้างอิงเชิงบริบทได้ ในขณะเดียวกันก็ใช้วิธีการเรียนรู้เชิงลึกเพื่อหลีกเลี่ยงการแยกคุณลักษณะด้วยตนเองและลดมิติของการดำเนินการ ทำให้ผลลัพธ์ของการจัดหมวดหมู่ดีกว่าวิธีอื่นๆ

นอกจากนี้ยังทำการทดลองเปรียบเทียบเวลาที่ต้องใช้สำหรับรุ่นไฮบริดในการฝึกอบรมในสภาพแวดล้อมของ GPU หน่วยเวลาแสดงหน่วยเป็นวินาที โดยค่าเฉลี่ยได้มาจากการทำซ้ำ 10 ครั้ง

ตารางที่ 4 การเปรียบเทียบระยะเวลาการฝึกอบรมของแบบจำลองที่ใช้ CBOW + WORD2VEC

Serial number	Method	Time
1	CNN+word2vec+CBOW	26
2	LSTM+word2vec+CBOW	60
3	CNN+LSTM+word2vec+CBOW	34

ตารางที่ 5 การเปรียบเทียบระยะเวลาการฝึกอบรมของแบบจำลองที่ใช้ SKIP-GRAM + WORD2VEC

Serial number	Method	Time
1	CNN+word2vec+Skip-Gram	26
2	LSTM+word2vec+ Skip-Gram	60
3	CNN+LSTM+word2vec+ Skip-Gram	34

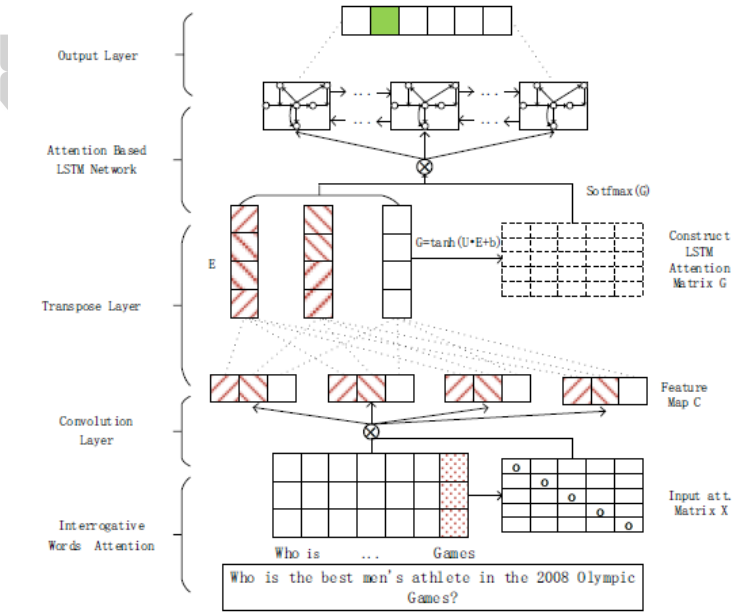
จากตารางที่ 4 และตารางที่ 5 แบบจำลองรูปแบบไฮบริดใช้เวลานานกว่า CNN ทั้งแบบ CBOW และ Skip-Gram ใน Word2vec อย่างไรก็ตามเวลาฝึกอบรมจะลดลงเมื่อเปรียบเทียบกับ LSTM ทั้งแบบ CBOW และ Skip-Gram ใน Word2vec เพราะ LSTM เก็บรักษาข้อมูลที่เป็นประวัติและข้อมูลที่มีลักษณะของเวลาซึ่งใช้เวลาฝึกอบรมนาน อย่างไรก็ตามเวลาฝึกอบรมของ CNN ค่อนข้างสั้น ดังนั้นเมื่อเทียบกับ LSTM แบบจำลองแบบไฮบริดก็ยังใช้เวลาน้อยกว่า

จากการทดลองดังกล่าวแสดงให้เห็นว่าแบบจำลองของผู้วิจัยนั้นให้ค่าความแม่นยำที่สูงที่สุดจากวิธีการอื่นๆ แต่การใช้ LSTM แบบทิศทางเดียวนั้นอาจยังไม่ครอบคลุมการเชื่อมโยงของเนื้อหาในข้อความนั้นๆ ซึ่งปัจจุบันมีวิธีการที่ใช้ LSTM แบบสองทิศทางที่สามารถช่วยให้รักษาคุณลักษณะเชิงบริบทของข้อความได้ครอบคลุมกว่า LSTM แบบทิศทางเดียว

Question Classification Based on MAC-LSTM

ในปี 2018 Bengong Yu, Qingtang Xu, และ Peihang Zhang [4] นำเสนอ Attention Convolution LSTM เครือข่ายประสาทเทียมแบบหลายระดับ (MAC-LSTM) สำหรับการจำแนกคำถาม วิธีการนี้ใช้กลไกความสนใจคำของคำถามเพื่อเน้นลักษณะคำของบริบทในคำถาม ในขณะเดียวกันก็ใช้กลไกความสนใจบวกกับข้อดีของ CNN และ LSTM โดย MAC-LSTM สามารถจับภาพทั้งคุณสมบัติท้องถิ่นของวลีตลอดจนคุณสมบัติโดยภาพรวมและอนุกรมเวลา การทดลองแสดงให้เห็นว่าแนวทางของเราได้ผลดีกว่าวิธีการเรียนรู้ของเครื่องแบบดั้งเดิมทั้ง

แบบจำลอง MAC-LSTM ที่เสนอในบทความนี้จะเรียนรู้เกี่ยวกับคำในคำถามมีลักษณะเป็นข้อความคำถามผ่านสองระดับของกลไกความสนใจ ปรับปรุงความถูกต้องของการจำแนกคำถามอย่างมีประสิทธิภาพ ประการแรกกลไกการเอาใจใส่คำถูกนำมาใช้ในการป้อนข้อมูลรูปแบบชั้นและเมทริกซ์ความสนใจคำในคำถาถูกสร้างขึ้นผ่านพจนานุกรมคำศัพท์และป้อนเมทริกซ์ความสนใจคำในคำถามลงในชั้น Convolutional เพื่อแยกคุณลักษณะในเครื่อง ประการที่สองแบบจำลองใช้เมทริกซ์ความสนใจคำในคำถามเพื่อคัดกรองคุณสมบัติการที่มีประโยชน์ที่สุดสำหรับคำถาม การจำแนกประเภทและโอนไปยังเลเยอร์ LSTM แบบสองทิศทางสำหรับการสกัดคุณสมบัติอนุกรมเวลาระดับสูงในที่สุด หลอมรวมเพิ่มเวกเตอร์คุณสมบัติลงในตัวแยกประเภท Softmax เพื่อให้การจำแนกคำถามได้สมบูรณ์



ภาพที่ 11 ภาพรวมแผนผังของ MAC-LSTM

เลเยอร์ของการรับข้อมูลเข้าขึ้นอยู่กับกลไกการสังเกตคำ ข้อแตกต่างจากการจัดประเภทข้อความทั่วไปประเภทการเลือกปฏิบัติของประโยคคำถามขึ้นอยู่กับคำในคำถาม เนื่องจากข้อความคำถามมักจะสั้นและข้อมูลความหมายและข้อมูลการเกิดคำร่วมกันไม่เพียงพอ คำในประโยคคำถามในประโยคส่งผลอย่างมากต่อผลการจัดหมวดหมู่คำถาม อย่างไรก็ตามเวกเตอร์คำที่ใช้กันทั่วไปในแบบจำลองการเรียนรู้เชิงลึกไม่เน้นข้อมูลคำซักถาม ดังนั้นงานวิจัยนี้เสนอกลไกความสนใจคำในคำถามโดยการสร้างพจนานุกรมคำในคำถาม คำในเวกเตอร์รวมกับเมทริกซ์ความสนใจในแนวทแยงของคำซักถาม เพื่อให้ตัวแบบมีความสามารถในการกำหนดว่าส่วนใดของประโยคมีความเกี่ยวข้องมากที่สุดกับคำในคำถาม วิธีการเฉพาะคือ

ประการแรก รวบรวมคำในคำถามในประโยคคำถามภาษาจีนและสร้างพจนานุกรมคำในคำถาม เช่น “Who”, “Where”, “How”

ประการที่สอง ค้นหาพจนานุกรมคำในคำถามหลังจากแบ่งคำแล้ว ให้หาคำที่เป็นคำถามในประโยคเพื่อนำไปฝึกคำในเวกเตอร์

ประการสุดท้าย เมทริกซ์ความสนใจในแนวทแยงจะได้รับการแนะนำให้รู้จักกับอธิบายลักษณะความเกี่ยวข้องตามบริบทและความแรงของการเชื่อมต่อระหว่างคำสามัญและคำในประโยค

Convolution Layer สามารถดึงข้อมูลคุณสมบัติในเครื่องได้จากข้อมูลดิบที่ผ่านการดำเนินการแปลง งานวิจัยนี้ใช้ตัวกรองมิติเดียวที่มีขนาดเท่ากันเพื่อเลื่อนผ่านลำดับเวกเตอร์คำที่เป็นคำถาม เพื่อที่จะแยกข้อความ คุณสมบัติในพื้นที่ต่าง ๆ การทำแผนที่ขึ้นกับคุณสมบัติได้ของคำซักถาม

การแยกคุณสมบัติลำดับตามความสนใจ สำหรับงานจำแนกคำถาม กฎการจัดหมวดหมู่ของคำถามขึ้นอยู่กับบริบทไปข้างหน้าและข้างหลังข้อมูล. ดังนั้นวานวิจัยนี้จะเข้าหน้าต่างการทำแผนที่จากชั้น Convolutional ไปยัง LSTM แบบสองทิศทางเพื่อแยกคุณลักษณะอนุกรมเวลาของข้อความ

จากการทดลองเพื่อตรวจสอบความถูกต้องของแบบจำลองนี้ เราได้ดำเนินการการทดลองเปรียบเทียบข้อมูลข้อความสั้นภาษาจีนชุดต่างๆ ใช้การฝึกแบบข้ามแกรมโหมดของ word2vec ขนาดหน้าต่างบริบทถูกตั้งค่าเป็น 5 และขนาดมิติเวกเตอร์ของคำถูกตั้งค่าเป็น 100 เพื่อรับข้อความสั้นต้นฉบับ อินพุตแสดงถึงการแมปวลีเมทริกซ์

นอกจากนี้ การสร้างการจำแนกคำถามระบบเป็นพื้นฐานของการจำแนกคำถาม สำหรับการจำแนกประเด็นภาษาอังกฤษ ส่วนใหญ่ใช้ UIUC ระบบการจัดหมวดหมู่สำหรับชุดข้อมูลมาตรฐาน TREC QA และแบ่งประเด็นออกเป็นหกประเภท: ABBR, DESC, ENTY,HUM, LOC และ NUM ไม่มีการจำแนกประเภทแบบครบวงจร

ระบบการจำแนกคำถามภาษาจีน ระบบการจัดหมวดหมู่ที่เสนอโดยสถาบันฮาร์บินแห่งเทคโนโลยีสำหรับการดึงข้อมูลและคอมพิวเตอร์โซเชียลศูนย์ได้รับการรับรองโดยนักวิชาการส่วนใหญ่ ให้เป็นไปตามลักษณะของภาษาจีน เสนอ 6 หมวด ได้แก่ DES, HUM, LOC, NUM, TIME และ OBJ

ตารางที่ 6 หมวดหมู่ของคำถาม

Category	Representative Example
DES	What is the main business of Wanda Plaza?
HUM	Who is the author of the “Basic Mechanical Design?”
LOC	Where was Andre born?
NUM	What is the total investment of Hefei Metro Line 1?
TIME	When was the construction of Huayan Temple?
OBJ	Which is the second largest nation in China?

ชุดข้อมูลที่เลือกในบทความนี้แบ่งออกเป็นสามส่วน: ชุดข้อมูล Baidu Labs 6,205 คำถาม CCF International การประมวลผลภาษาธรรมชาติและคอมพิวเตอร์จีนการประชุมประจำปี 2559 (NLPCC 2016) ชุดคำถามทบทวน QAทั้งหมด 9,604 คำถาม ชุดคำถามของ NLPCC 2017 QA review 9,518 คำถาม รวม 25,327 คำถาม ชุดข้อมูลสาธารณะโดย NLPCC QA มีรูปแบบที่ชัดเจนและคุณภาพสูง แต่ละชุดข้อมูลมีแต่คำถามและคำตอบ ดังนั้นเราจึงได้ดำเนินการบันทึกย่อด้วยตนเองกับข้อมูลนี้และคำถามแต่ละข้อจะถูกทำเครื่องหมายโดยบุคคลสามคนอย่างอิสระ

สำหรับข้อมูลที่มีข้อโต้แย้ง เราเจรจาร่วมกันเพื่อกำหนดป้ายกำกับในเวลาเดียวกัน เพื่อแยกแยะข้อผิดพลาดในชุดข้อมูล ชุดฝึก และชุดทดสอบเป็นแบบสุ่มแบ่งเป็นสัดส่วน 20% สำหรับแต่ละชุดข้อมูล

ตารางที่ 7 การกระจายหน่วยฝึกอบรมและหน่วยทดสอบ

Dataset	DES	HUM	LOC	NUM	TIME	OBJ	Total
Baidu train	478	1102	881	1007	915	622	5005
Baidu test	126	238	225	230	184	197	1200
NLPCC 2016 train	1430	1067	953	1493	1054	1687	7684
NLPCC 2016 test	373	351	238	357	266	335	1920
NLPCC 2017 train	1477	981	952	1270	1131	1803	7614
NLPCC 2017 test	359	255	228	327	272	463	1904

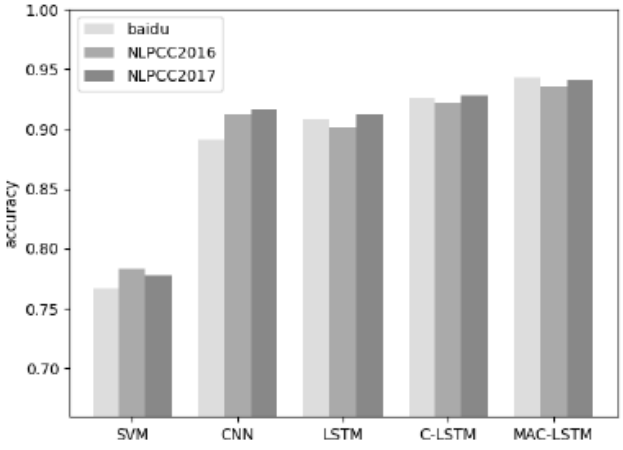
การตั้งค่าการทดสอบเปรียบเทียบ โมเดล MAC-LSTM ที่นำเสนอในบทความนี้ประกอบด้วยโครงข่ายประสาทเทียมที่มี LSTM ผ่านหลายระดับกลไกการเอาใจใส่และเน้นการซักถามคุณสมบัติคำสำหรับงานจำแนกคำถาม เพื่อตรวจสอบประสิทธิผลของแบบจำลอง เราตั้งค่าวิธีการหลายแบบสำหรับการทดลองเปรียบเทียบชุดข้อมูลสองชุด

การประเมินผลและผลลัพธ์ เราตรวจสอบการบังคับใช้และความเหนือกว่าของ MAC-LSTM แบบจำลองในการจัดประเภทข้อความสั้นโดยเปรียบเทียบกับที่แตกต่างกันโมเดลในชุดข้อมูลต่างๆ ผลลัพธ์จะแสดงในตารางที่ 8

ตารางที่ 8 การเปรียบเทียบประสิทธิภาพของแบบจำลอง MAC-LSTM

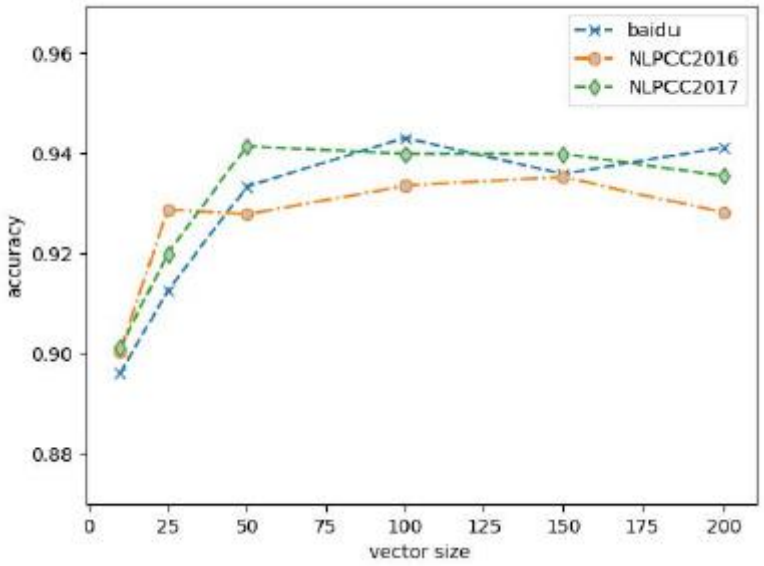
Methods	Baidu	NLPCC 2016	NLPCC 2017
SVM	76.69%	78.31%	77.82%
CNN	89.14%	91.28%	91.61%
LSTM	90.82%	90.16%	91.21%
C-LSTM	92.65%	92.19%	92.82%
MAC-LSTM	94.31%	93.52%	94.14%

จากตารางที่ 8 การกระจายข้อมูลในรูปแบบต่างๆ ชุดข้อมูลหลักแตกต่างกัน ความถูกต้องของแต่ละรุ่นผันผวนเป็นองศาที่แตกต่างกัน แต่ไม่ส่งผลกระทบต่อผลการเปรียบเทียบระหว่างรุ่นต่างๆ เราวาดฮิสโตแกรมเพื่อแสดงการเปรียบเทียบความแม่นยำของแบบจำลองหลายประเภทในชุดข้อมูลต่างๆ MAC-LSTM มีประสิทธิภาพดีกับชุดข้อมูลทั้งสามชุด



ภาพที่ 12 การเปรียบเทียบความแม่นยำในชุดข้อมูลสามชุด

จากภาพที่ 12 แสดงว่าแบบจำลองของเราดีกว่าแบบเดิมๆ SVM ที่มีคุณสมบัติการออกแบบที่ประดิษฐ์ขึ้นสูง คุณสมบัติการออกแบบต้องใช้แรงงานคนจำนวนมากและไม่สามารถทำได้ดีโดยทั่วไปกับชุดข้อมูลหรืองานอื่น ๆ MAC-LSTM ที่เสนอในบทความนี้มีความสามารถในการเรียนรู้การแสดงความหมายของประโยค จึงไม่จำเป็นที่ต้องใช้คุณสมบัติเทียมและมีความสามารถในการปรับขนาดได้ดีกว่า เราเปรียบเทียบผลลัพธ์ของโครงข่ายประสาทเทียมแบบเดียวกับ LSTM และโครงข่ายประสาทเทียมมีความแม่นยำสูงกว่า ในชุดข้อมูล NLPCC แสดงให้เห็นว่าอัลกอริธึมการบิดสามารถแยกคุณสมบัติข้อความได้อย่างมีประสิทธิภาพมากขึ้นเมื่อข้อความสั้นและปริมาณข้อมูลเพียงพอผลการทดลองแสดงให้เห็นว่าผลการจัดประเภทที่ได้รับจาก C-LSTM และ MAC-LSTM ที่เสนอในบทความนี้ดีมากกว่าที่ได้รับจาก CNN หรือ LSTMs แสดงว่าการรวมกันของโครงข่ายประสาทเทียมและ LSTM ไม่ได้แยกคุณสมบัติท้องถิ่นของข้อความอย่างเดียว แต่ยังสำรวจคุณลักษณะเพิ่มเติมของอนุกรมเวลาของปัญหาซึ่งเอื้อต่อการทำความเข้าใจในคำถาม เมื่อเทียบกับ C-LSTM แบบจำลองและวิธีการ MAC-LSTM ที่เสนอในบทความนี้ เราจะพบว่ากลไกความสนใจหลายระดับสามารถทำให้โมเดลให้ความสำคัญกับข้อมูลคุณสมบัติของเป้าหมายการจัดหมวดหมู่ในกระบวนการฝึกอบรมเพื่อให้ได้ประโยคคำถามมีคุณลักษณะที่ดีกว่าบนพื้นฐานของ C-LSTM วิทยานิพนธ์นี้นำเสนอกลไกการเอาใจใส่คำซักถามและเสริมสร้างความสามารถของตัวแบบในการแยกคุณสมบัติคำคำถาม



ภาพที่ 13 ประสิทธิภาพของเวกเตอร์ที่แตกต่างกันตามขนาด

โดยการสร้างเวกเตอร์คำในคำถาม คุณสมบัติเป็นตัวแทนของเวกเตอร์คำในคำถามมีความสัมพันธ์กับมิติของเวกเตอร์คำอย่างมาก งานวิจัยนี้ใช้เวกเตอร์คำที่มีมิติต่างกันเพื่อสร้างเวกเตอร์คำคำถามเพื่อทดสอบและตรวจสอบประสิทธิภาพและความมั่นคงของกลไกการเอาใจใส่คำซักถาม ดังแสดงในภาพที่ 13 ด้วยการเปลี่ยนแปลงของมิติเวกเตอร์ ความถูกต้องของแบบจำลองเพิ่มขึ้นในตอนเริ่มต้นและจะเสถียรเมื่อขนาดของมิติเกิน 50 งานวิจัยนี้สุดท้ายเลือก 100 เป็นมิติเวกเตอร์ซึ่งสามารถได้รับผลการทดลองที่ดีขึ้นโดยพิจารณาจากการคำนวณอย่างง่าย

งานวิจัยนี้วิเคราะห์ข้อบกพร่องของแบบดั้งเดิม วิธีการจำแนกคำถามและการประยุกต์เชิงลึกการเรียนรู้ในการจำแนกคำถาม ในขณะเดียวกันก็ผสานกลไกการเอาใจใส่ปรับปรุงส่วนลึก รูปแบบการเรียนรู้และเสนอความสนใจหลายระดับอย่างลึกซึ้ง รูปแบบการเรียนรู้ ประการแรก กลไกความสนใจถูกนำไปใช้กับทำให้ตัวแบบให้ความสำคัญกับคำซักถามมากขึ้น คุณสมบัติในข้อความคำถามประการที่สอง convolutionalโครงข่ายประสาทเทียมและ LSTM เชื่อมต่อผ่านกลไกความสนใจเพื่อให้ได้ข้อดีของ CNN และ LSTM เกี่ยวกับการจำแนกคำถาม MAC-LSTM เรียนรู้คุณสมบัติระดับวลีของคำซักถามเวกเตอร์ผ่านการดำเนินการบิดแล้วส่งการแสดงคุณสมบัติใน LSTM เพื่อเพิ่มความสามารถในการรับคุณสมบัติอนุกรมเวลา เราทำการทดลองโดยใช้แบบจำลอง MAC-LSTM เพื่อจัดการกับการจำแนกคำถามและบรรลุผลที่น่าพอใจ ปัจจุบันมีการจำแนกประเภทคำถามที่ขาดหายไปในคลังข้อมูล นอกจากนี้ยังคงต้องรวบรวมและติดป้ายกำกับข้อมูลใหม่ วิธีการใช้ข้อมูลขนาดใหญ่ จำนวนคลังข้อมูลที่ไม่มีป้ายกำกับเพื่อดำเนินการกึ่งควบคุมการเรียนรู้ยังเป็นแนวทางการวิจัยที่เป็นไปได้จากนี้ คุณภาพของการจำแนกคำถามจะส่งผลต่อคำตอบ นอกจากนี้การสกัดและการเลือกกลยุทธ์การสร้างคำตอบ ดังนั้นการวิจัยเรื่องการผสมผสานการสกัดคำตอบและการจำแนกคำถามก็เป็นงานวิจัยที่มีแนวโน้วในทิศทางเดียวกัน

บทที่ 3

ทฤษฎีที่เกี่ยวข้อง

การแปลงคำเป็นตัวเลข (Word Embedding)

ในการประมวลผลภาษาธรรมชาติ (NLP) การฝังคำเป็นคำที่ใช้แทนคำสำหรับการวิเคราะห์ข้อความ โดยทั่วไปแล้วจะอยู่ในรูปของเวกเตอร์ที่มีค่าจริงซึ่งเข้ารหัสความหมายของคำดังกล่าว โดยคาดว่าคำที่อยู่ใกล้เคียงในพื้นที่เวกเตอร์จะมีความหมายที่ใกล้เคียงกัน [5] การฝังคำได้โดยใช้ชุดของการสร้างแบบจำลองภาษาและเทคนิคการเรียนรู้คุณลักษณะ โดยที่คำหรือวลีจากคำศัพท์จะจับคู่กับเวกเตอร์ของตัวเลขจริง

การเรียนรู้เชิงลึก (Deep Learning)

แบบจำลองการคำนวณทางคอมพิวเตอร์ที่มีโครงสร้างการประมวลผลหลายชั้นนำไปสู่การเรียนรู้ด้วยข้อมูลที่มีความหลากหลายระดับ [6] วิธีการนี้ได้เพิ่มประสิทธิภาพของเทคโนโลยีในปัจจุบันอย่างมาก เช่น การเรียนรู้จากคำพูด, การเรียนรู้จากวัตถุที่มองเห็น รวมไปถึงการประมวลผลภาษาธรรมชาติและข้อความ ซึ่งเป็นการประมวลผลจากข้อมูลหลายแหล่งในรูปแบบต่างๆ โดยใช้การเรียนรู้ลึกแบบทำงานหลายอย่างพร้อมกัน [7]

โครงข่ายประสาทแบบคอนโวลูชัน (Convolutional Neural Network)

Convolutional Neural Network (ConvNet/CNN) [8] เป็นอัลกอริธึม Deep Learning ที่รับอิมเมจอินพุต กำหนดความสำคัญ (น้ำหนักและอคติที่เรียนรู้ได้) ให้กับแง่มุม/อ็อบเจกต์ต่างๆ ในภาพ และสามารถแยกความแตกต่างจากที่อื่นได้ การประมวลผลล่วงหน้าที่จำเป็นใน ConvNet นั้นต่ำกว่ามากเมื่อเทียบกับอัลกอริธึม การจำแนกประเภทอื่นๆ ในขณะที่ตัวกรองวิธีดั้งเดิมนั้นได้รับการออกแบบมาด้วยมือ แต่ด้วยการฝึกอบรมที่เพียงพอ ConvNets มีความสามารถในการเรียนรู้ตัวกรอง/ลักษณะเหล่านี้

โครงสร้างเครือข่ายประสาทเทียมหน่วยความจำระยะสั้นแบบยาวแบบสองทิศทาง (Bidirectional Long Short Term Memory)

โครงข่ายประสาทเทียมแบบสองทิศทาง (Bi-RNN) [9] เชื่อมต่อสองชั้นที่ซ่อนอยู่ของทิศทางตรงกันข้ามกับเอาต์พุตเดียวกัน ด้วยรูปแบบการเรียนรู้เชิงลึกเชิงสร้างสรรค์นี้ เลเยอร์เอาต์พุตสามารถ

รับข้อมูลจากสถานะในอดีต (ย้อนหลัง) และอนาคต (ไปข้างหน้า) ได้พร้อมกัน คิดค้นในปี 1997 โดย Schuster และ Paliwal

เทคนิคสำหรับการปรับปรุงความเร็วสมรรถนะและความเสถียรของเครือข่ายประสาทเทียม (Batch Normalization)

เป็นวิธีที่ใช้ในการทำให้โครงข่ายประสาทเทียมเร็วขึ้นและมีเสถียรภาพมากขึ้น [10] ผ่านการทำให้อินพุตของเลเยอร์เป็นมาตรฐานโดยการจัดกึ่งกลางใหม่และการปรับขนาดใหม่

การอธิบายคำทำนายของแบบจำลองด้วย SHAP (SHapley Additive exPlanations)

SHAP (SHapley Additive exPlanations) เป็นเครื่องมือในการอธิบายการทำนายของแบบจำลอง [11] โดยใช้ค่า Shapley ที่มาจากทฤษฎีเกม ซึ่งจะคำนวณค่าของตัวแปรแต่ละตัวว่าส่งผลต่อผลลัพธ์อย่างไร

ค่าแชปลีย์เป็นแนวคิดที่ยืมมาจากวรรณกรรมทฤษฎีเกมแบบร่วมมือและย้อนหลังไปถึงปี 1950 [12] รูปแบบดั้งเดิมค่าของแชปลีย์ถูกใช้เพื่อระบุถึงการมีส่วนร่วมของผู้เล่นในผลลัพธ์สุดท้ายของเกม สมมติว่าเรามีเกมแบบร่วมมือกันซึ่งกลุ่มผู้เล่นแต่ละคนร่วมมือกันเพื่อสร้างคุณค่าบางอย่าง หากเราสามารถวัดผลตอบแทนรวมของเกมได้ ค่าแชปลีย์จะจับส่วนที่เพิ่มขึ้นของผู้เล่นแต่ละคนในผลลัพธ์สุดท้าย ค่าแชปลีย์จะคำนวณส่วนเพิ่มโดยเฉลี่ยของจุดสนใจ x ต่อคะแนนแบบจำลอง โดยการคำนวณค่าแชปลีย์สามารถหาได้โดยใช้สมการดังนี้

$$\Phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(N - |S| - 1)!}{N!} \left(v(S \cup \{i\}) - v(S) \right)$$

ภาพที่ 14 สมการการคำนวณค่าแชปลีย์

บทที่ 4

วิธีการทดลอง

บทนี้จะกล่าวถึงลักษณะข้อมูลที่ใช้, การเตรียมข้อมูลก่อนประมวลผล, การเรียนรู้เชิงลึกของแบบจำลอง, การปรับพารามิเตอร์, การอธิบายคำทำนายของตัวแบบแยกประเภทคำถาม, การตรวจหาปัญหาที่อาจเกิดขึ้นจากข้อมูลฝึกและการปรับปรุงคุณภาพของข้อมูลฝึกจากปัญหาที่ตรวจพบ

ลักษณะข้อมูลที่ใช้

ข้อมูลที่นำมาวิจัยรวบรวมมาจากฐานข้อมูลคำถามงานอีเว้นท์ของบริษัทผู้ให้บริการแพลตฟอร์มด้านการจัดการอีเว้นท์ในไทยแห่งหนึ่งจำนวน 2,014 คำถาม ซึ่งแต่ละคำถามเป็นคำถามที่ไม่ซ้ำกัน ประเภทของคำถามแบ่งเป็น 8 ประเภท คือ Anatomy, Social, Finance, Work, Education, Lifestyle, Channel, Other โดยให้พนักงานของบริษัทผู้ให้บริการแพลตฟอร์มอีเว้นท์ที่รับผิดชอบในงานส่วนนี้เป็นผู้ใส่ป้ายกำกับประเภทข้อมูลในแต่ละคำถาม

ตารางที่ 9 จำนวนคำถามตามประเภท

Category	**Number**
Anatomy	155
Social	185
Finance	91
Work	605
Education	194
Lifestyle	212
Channel	138
Other	434

จากตารางที่ 9 เป็นจำนวนคำถามแต่ละประเภทที่นำมาใช้สำหรับการฝึกอบรมและทดสอบแบบจำลอง ซึ่งโดยเฉลี่ยข้อมูลแต่ละประเภทจะน้อยกว่า 400 คำถาม

ตารางที่ 10 ตัวอย่างคำถามตามประเภท

Category	**Question Example**
Anatomy	วันเดือนปีเกิด
Social	ช่องทางการติดต่อผู้สมัครอื่นๆ เช่น Facebook, Twitter / LINE :
Finance	รายได้ต่อเดือน
Work	Job Title / Job Position (ตำแหน่ง)
Education	การศึกษาของท่าน / Education
Lifestyle	คุณสนใจมาเที่ยวงาน Area Five ใน Zone ใด เป็นพิเศษ
Channel	คุณรับทราบข่าวสารงานคอนเสิร์ต David Foster จากช่องทางใด
Other	เหตุผลที่มาร่วมฟัง

จากตารางที่ 10 เป็นตัวอย่างคำถามตามประเภทที่นำมาใช้สำหรับสอนแบบจำลองให้เรียนรู้และทดสอบ

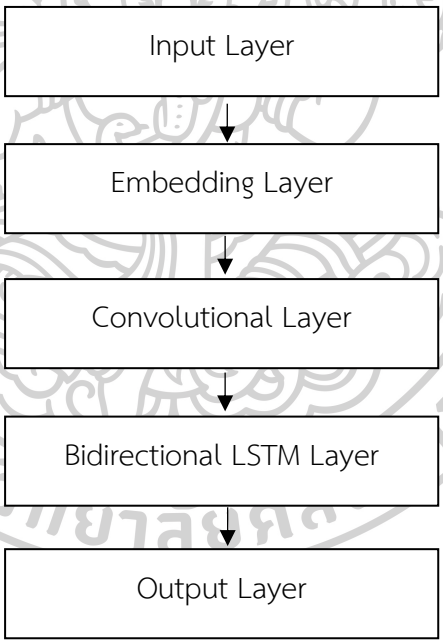
การเตรียมข้อมูลก่อนประมวลผล

ผู้วิจัยนำคำถามทั้งหมดเข้าสู่กระบวนการทำความสะอาดข้อมูลโดยนำคำถามที่ซ้ำกันออก เครื่องหมายวรรคตอนต่างๆออก ปรับตัวอักษรภาษาอังกฤษให้เป็นตัวเล็กและนำเครื่องหมายวรรคตอนและอักขระพิเศษต่างๆออกทั้งหมด จากนั้นนำมาตัดคำโดยใช้ newmm เป็น Library สำหรับการตัดคำภาษาไทย (Word tokenization) หลังจากผ่านกระบวนการตัดคำเรียบร้อยแล้วพบว่าคำถามที่มีความยาวมากที่สุดมีความยาวเท่ากับ 31 คำ คำถามที่มีความยาวน้อยที่สุดมีความยาวเท่ากับ 1 คำ และค่าเฉลี่ยความยาวของคำถามทั้งหมดคือ 4.86 คำ จากนั้นจึงสร้างคลังคำศัพท์ (Bag of words) โดยมีคำศัพท์ทั้งหมด 1,046 คำ

การเรียนรู้เชิงลึกของแบบจำลอง

ในงานวิจัยนี้เราออกแบบโครงข่ายประสาทเทียมแบบผสมโดยชั้นนำข้อมูลเข้าจะผ่าน Embedding Layer เพื่อแปลงคำศัพท์ต่างๆเป็นตัวเลขเพื่อให้ได้คุณลักษณะและมีการปรับขนาด

เวกเตอร์ให้ลดลงจากนั้นมีการใช้ Dropout เพื่อช่วยลดปัญหาการ Overfitting ของแบบจำลอง จากนั้นจะโอนไปยังชั้น Convolutional 1 มิติ เพื่อทำการดึงคุณลักษณะต่างๆของกลุ่มคำผ่านตัวกรองที่กำหนดและใช้ Activation ReLU จากนั้นใช้ Batch Normalization เพื่อปรับค่าจาก Layer ก่อนหน้าให้เป็นค่ามาตรฐานและโอนไปให้ Max Pooling 1 มิติ เพื่อทำการปรับขนาดข้อมูล โดยตัวกรองจะเลือกค่าที่มากที่สุดของแต่ละเมทริกซ์เพื่อทำฟีเจอร์แมพจากนั้นโอนไปยังชั้น Bidirectional Long-Short Term Memory ชั้นนี้จะทำการดึงคุณสมบัติของความเชื่อมโยงของเนื้อหาและลำดับการขึ้นต้นของคำก่อนหลังโดยมีการพิจารณาแบบสองทิศทาง จากนั้นทำ Batch Normalization เพื่อปรับค่าจาก Layer ก่อนหน้าให้เป็นค่ามาตรฐานจากนั้นปรับข้อมูลให้เรียบเพื่อส่งต่อไปยังชั้น Dense Layer และใช้ Activation ReLU และทำ Batch Normalization อีกครั้ง สุดท้ายชั้นข้อมูลนำออกมีการใช้ Activation Softmax



ภาพที่ 15 รายละเอียดโครงสร้างแบบจำลอง CNN-BiLSTM

การปรับพารามิเตอร์

ในการทดลองปรับค่าพารามิเตอร์เบื้องต้น ผู้วิจัยได้ทำการทดลองเพื่อปรับพารามิเตอร์ให้แบบจำลองมีประสิทธิภาพมากที่สุด โดยคำนึงถึงค่า Filter และ Kernel Size ของชั้น Convolution Layer และ Dropout โดยค่า Filter เป็นตัวกรองซึ่งกำหนดผังคุณลักษณะ Kernel Size เป็น

ตัวกำหนดความกว้างของตัวกรอง ซึ่งการปรับพารามิเตอร์ Filter และ Kernel Size นั้นส่งผลอย่างมากต่อความสามารถของแบบจำลองเพราะหากสามารถดึงคุณลักษณะที่สำคัญและระยะการมองเห็นได้เหมาะสมแล้วจะสามารถทำให้แบบจำลองทำนายประเภทของคำถามได้อย่างแม่นยำมากยิ่งขึ้น ส่วน Dropout นั้นสามารถช่วยในเรื่องของการลด Overfit ของแบบจำลองได้เป็นอย่างมาก

ตารางที่ 11 เปรียบเทียบ Filter และ Kernel Size

Filter	Kernel Size	Validation Accuracy	Test Accuracy
16	4	95.31%	94.39%
16	6	96.09%	95.44%
16	8	96.09%	95.44%
32	4	**96.88%**	**95.79%**
32	6	**96.88%**	95.09%
32	8	94.92%	94.74%
64	4	95.70%	94.74%
64	6	**96.88%**	95.09%
64	8	95.31%	95.44%

จากตารางที่ 11 เปรียบเทียบ Filter และ Kernel Size ในแต่ละขนาดพบว่าเมื่อใช้ Filter = 32, Kernel Size = 4 ให้ผลลัพธ์ดีที่สุด โดย Filter = 32, Kernel Size = 6 และ Filter = 64, Kernel Size = 6 ให้ผลลัพธ์รองลงมา

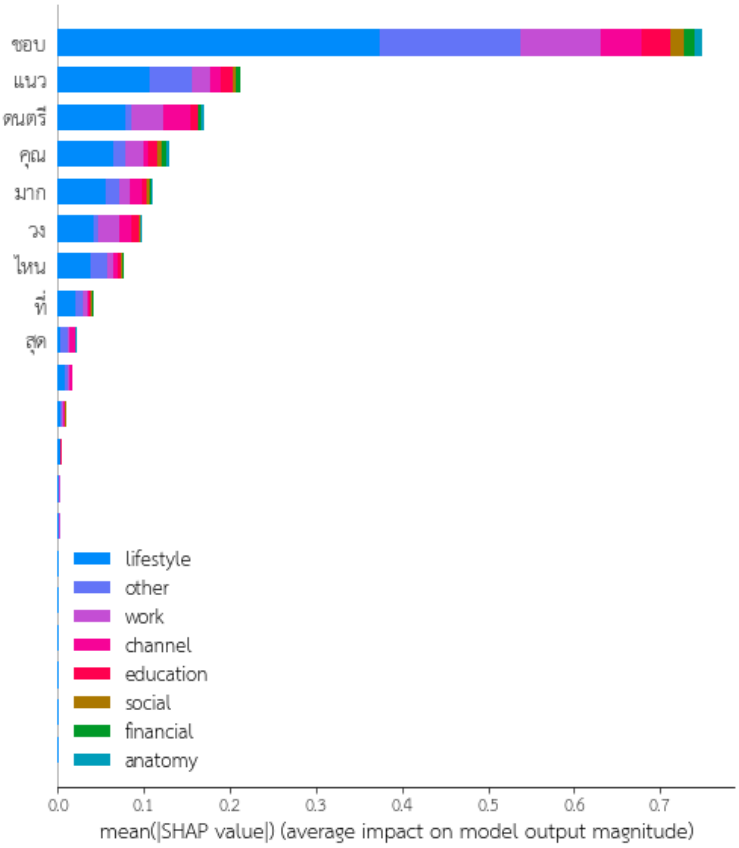
ตารางที่ 12 เปรียบเทียบการใช้งาน Dropout

Dropout	Validation Accuracy	Test Accuracy
None	96.09%	95.09%
0.25	**96.88%**	**95.79%**
0.5	96.48%	95.44%

จากตารางที่ 12 เปรียบเทียบการใช้งาน Dropout โดยพบว่าเมื่อใช้ Dropout = 0.25 จะให้ผลลัพธ์ที่ดีที่สุด โดยเมื่อใช้ Dropout = 0.5 จะให้ผลลัพธ์รองลงมา จากการทดลองนี้พบว่าการใช้ Dropout นั้นให้ผลลัพธ์ที่ดีกว่าการไม่ใช้ Dropout เลย

การอธิบายคำทำนายของตัวแบบแยกประเภทคำถาม

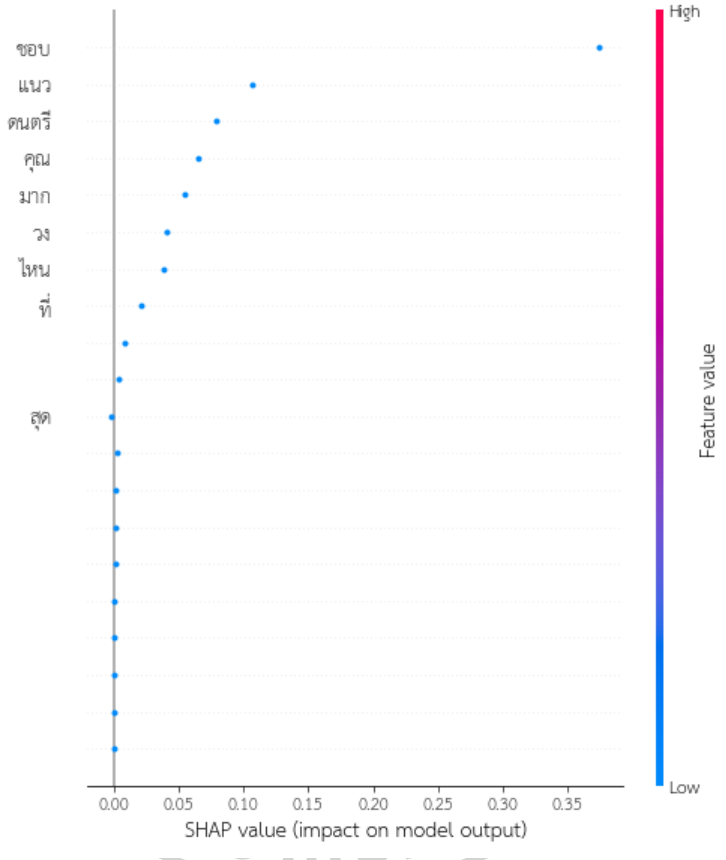
งานวิจัยนี้นำเสนอการอธิบายคำทำนายของแบบจำลองโดยใช้เครื่องมือ SHAP โดยแบบจำลองที่เรานำเสนอคือ CNN-BiLSTM ดังนั้นเราจึงใช้ SHAP Deep Explainer ซึ่งเป็นอัลกอริทึมที่เหมาะสมกับแบบจำลองการเรียนรู้เชิงลึก โดยเราจะนำแบบจำลองที่ถูกฝึกเรียบร้อยแล้วและนำข้อมูลคำถามที่ใช้สำหรับฝึกแบบจำลอง 10% ซึ่งมีจำนวน 201 คำถาม จาก 2,014 คำถาม เป็นพารามิเตอร์ให้ SHAP จากนั้นเรานำคำถาม 1 คำถามใช้ในการทดสอบให้ SHAP อธิบายคุณลักษณะของคำที่มีอิทธิพลต่อผลลัพธ์ เราจะขอยกตัวอย่างคำถาม “คุณชอบวงดนตรีแนวไหนมากที่สุด” ซึ่งคำถามนี้จัดเป็นคำถามประเภท Lifestyle แบบจำลองทายถูกด้วยค่าความเชื่อมั่น 99.90% เราจะให้ SHAP คำนวณค่า SHAP Value และแสดงกราฟอธิบายคำที่มีอิทธิพลในการทำนายของแบบจำลอง



ภาพที่ 16 กราฟแสดงความถี่คำของผลลัพธ์การทำนายที่มีอิทธิพลต่อคำถามทุกประเภท

จากภาพที่ 16 แสดงค่าเฉลี่ยของคำในประโยคคำถาม “คุณชอบวงดนตรีแนวไหนมากที่สุด” ที่มีอิทธิพลต่อประเภทของคำถามทั้งหมดที่แบบจำลองสามารถทำนายได้ โดยที่แบบจำลองทำนายว่า

คำถามนี้อยู่ในคำถามประเภท Lifestyle โดยคำว่า “ชอบ” ปรากฏอยู่ในคำถามประเภท Lifestyle มากที่สุด และพบได้ในคำถามประเภท Other, Work, Channel, Education, Social, Financial ตามลำดับ แต่พบได้น้อยมากในคำถามประเภท Anatomy ในขณะที่คำว่า “แนว” และ “ดนตรี” ที่ปรากฏอยู่ในคำถามประเภท Lifestyle เป็นลำดับที่สองและสามไม่ปรากฏเลยในคำถามประเภท Anatomy



ภาพที่ 17 กราฟแสดงอิทธิพลของคำของผลลัพธ์การทำนายที่มีต่อประเภทคำถามของผลลัพธ์

จากภาพที่ 17 แสดงคำในประโยคคำถาม “คุณชอบวงดนตรีแนวไหนมากที่สุด” ที่มีอิทธิพลต่อคำตอบของแบบจำลอง โดยคำว่า “ชอบ” มีอิทธิพลต่อคำตอบของแบบจำลองมากที่สุด รองลงมาคือคำว่า “แนว” และ “ดนตรี” ตามลำดับ

การตรวจหาปัญหาที่อาจเกิดขึ้นจากข้อมูลฝึก

เนื่องจากเป้าหมายของงานค่อนข้างเฉพาะทางชุดข้อมูลจึงมีขนาดเล็ก การสำรวจปัญหาในชุดข้อมูลฝึกจึงเป็นสิ่งที่สำคัญเพื่อที่จะสร้างข้อมูลด้วยการสังเคราะห์ตัวอย่างด้วยตนเอง (Manual

example synthesis) และการเพิ่มตัวอย่างเพิ่มเติม (Additional example) โดยกระทำโดยบุคคลที่สามที่รับผิดชอบในการจัดการกับข้อมูลเพื่อให้เป็นประโยชน์กับการฝึกโดยเลี่ยงการเกิด bias ที่อาจเกิดจากการผูกกับคำศัพท์ในงานด้านใดด้านหนึ่งมากเกินไป

ผู้วิจัยได้แบ่งประเภทของปัญหาที่อาจเกิดขึ้นในชุดข้อมูลฝึกทั้งหมด 3 ประเภท

- ตัวอย่างที่หายาก (Rare sample)
- การติดป้ายกำกับผิด (Label conflict)
- คำศัพท์เหมือนกันแต่เรียงประโยคไม่เหมือนกัน (Feature conflict)

ผู้วิจัยได้พิจารณาวิธีการหาปัญหาที่อาจเกิดขึ้นจากข้อมูลฝึกโดยมีกระบวนการดังนี้

- การค้นหาตัวอย่างที่คล้ายกัน: วิธีการที่เสนอจะค้นหาตัวอย่างอื่นๆ ในชุดการฝึกที่มีความคล้ายคลึงกันกับตัวอย่างที่แบบจำลองจำแนกผิด โดยจะพิจารณาตัวอย่างที่คล้ายกันมากกว่าหรือเท่ากับ 80% ด้วย Thai2Vec โดยใช้ฟังก์ชัน Sentence similarity ของ PythaiNLP
- การคำนวณการระบุแหล่งที่มาของฟีเจอร์: วิธีการนี้ใช้ SHAP Explainer (XAI) เพื่อคำนวณค่าคุณลักษณะของคำแต่ละคำที่ส่งผลกระทบต่อการตัดสินใจจำแนกประเภทของแบบจำลอง
- การตรวจจับความขัดแย้ง: วิเคราะห์ความขัดแย้งระหว่างค่าคุณลักษณะที่มีอิทธิผลสูงต่อประเภทของคำถาม วิธีการนี้สามารถระบุสาเหตุที่เป็นไปได้ของการจำแนกที่ไม่ถูกต้อง

การปรับปรุงคุณภาพของข้อมูลฝึกจากปัญหาที่ตรวจพบ

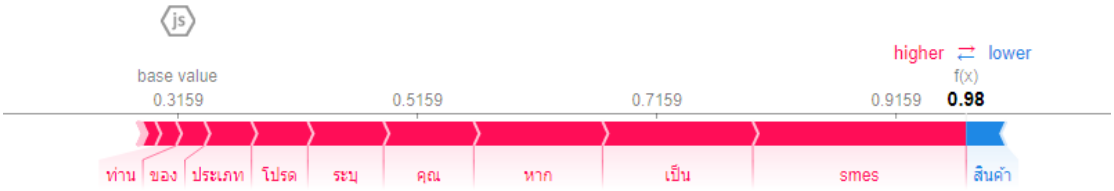
เมื่อผู้วิจัยได้นำเสนอขั้นตอนการตรวจหาปัญหาที่อาจเกิดขึ้นจากข้อมูลฝึก ผู้วิจัยจะพิจารณาเพื่อปรับปรุงชุดข้อมูลฝึก โดยทำตามขั้นตอนที่สอดคล้องกับปัญหาที่อาจเกิดขึ้นในชุดข้อมูลฝึกได้ดังนี้

ตัวอย่างที่มีความหายาก (Rare sample)

- ขั้นตอนที่หนึ่ง: นำตัวอย่างที่มีปัญหาเข้าสู่กระบวนการตรวจสอบความคล้ายกัน
- ขั้นตอนที่สอง: ผลลัพธ์มักพบตัวอย่างเพียงเล็กน้อยหรือไม่มีเลย
- ขั้นตอนที่สาม: หากตัวอย่างมีจำนวนน้อยมากหรือไม่มีตัวอย่างที่คล้ายกัน ให้เพิ่มตัวอย่างที่คล้ายกันในชุดการฝึกเพื่อเพิ่มจำนวนตัวอย่างในกลุ่มนั้น เพราะเป็นไปได้ว่าประโยคคำถามนั้นแบบจำลองมีการเรียนรู้ได้น้อยเพราะข้อมูลนั้นหายากหรือมีการกระจายตัวที่ไม่ดีพอ

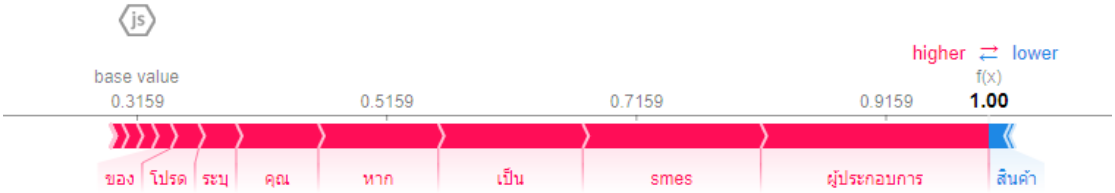
การติดป้ายกำกับผิด (Label conflict)

- ขั้นตอนที่หนึ่ง: นำตัวอย่างที่มีปัญหาเข้าสู่กระบวนการตรวจสอบความคล้ายกัน
- ขั้นตอนที่สอง: พบตัวอย่างอื่นที่คล้ายคลึงกัน โดยตัวอย่างที่พบมีความคล้ายคลึงกันสูงแต่ป้ายกำกับมักแตกต่างกัน
- ขั้นตอนที่สาม: SHAP แสดงค่าคุณลักษณะของคำมีความสัมพันธ์กันสูงระหว่างสองตัวอย่าง



แต่ป้ายกำกับขัดแย้งกัน สิ่งนี้บ่งชี้ว่าป้ายกำกับต้องได้รับการแก้ไขเพิ่มเติม

ภาพที่ 18 SHAP แสดงคุณสมบัติของคำสำหรับคำถามที่แบบจำลองจำแนกประเภทผิด "หากคุณเป็น SMEs โปรดระบุประเภทสินค้าและบริการของท่าน"



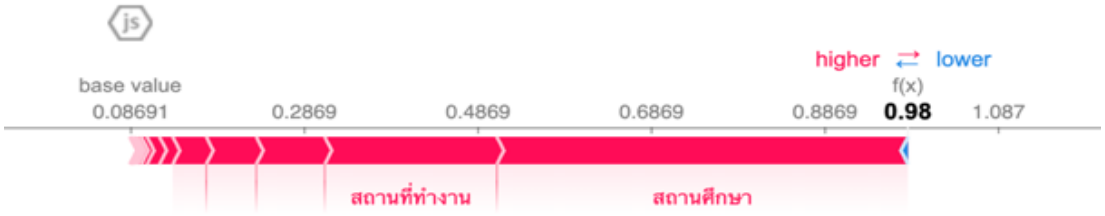
ภาพที่ 19 SHAP แสดงคุณสมบัติของคำสำหรับคำถามที่คล้ายที่สุดกับคำถามที่แบบจำลองจำแนกประเภทผิด "หากคุณเป็นผู้ประกอบการ SMEs โปรดระบุประเภทสินค้าและบริการของท่าน"

จากภาพที่ 18 คำถามที่แบบจำลองทายผิดติดป้ายกำกับเป็นคลาส "Other" แต่แบบจำลองทายเป็นคลาส "Work" ในขณะที่ภาพที่ 19 เป็นคำถามที่คล้ายกันมากที่สุด และติดป้ายกำกับเป็นคลาส "Work" ในสถานการณ์เช่นนี้แทบจะเป็นไปไม่ได้เลยที่ตัวอย่างทั้งสองนี้จะอยู่ในคลาสที่แตกต่างกัน กรณีนี้มีการติดป้ายกำกับผิดแต่แรก โดยที่แบบจำลองทำนายผลลัพธ์ได้ถูกต้องแล้ว

คำศัพท์เหมือนกันแต่เรียงประโยคไม่เหมือนกัน (Feature conflict)

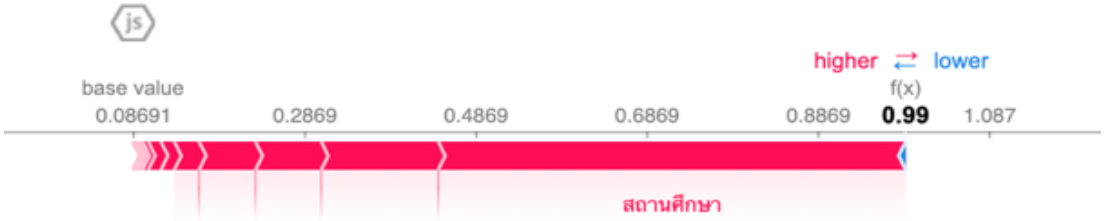
- ขั้นตอนที่หนึ่ง: นำตัวอย่างที่มีปัญหาเข้าสู่กระบวนการตรวจสอบความคล้ายกัน
- ขั้นตอนที่สอง: พบตัวอย่างอื่นที่คล้ายคลึงกัน ตัวอย่างทดสอบติดป้ายกำกับถูกต้องแล้วแต่แบบจำลองทำนายผลลัพธ์ออกมาผิด

- ขั้นตอนที่สาม: SHAP สามารถระบุได้ว่าปัญหาอาจเกิดขึ้นจากปัจจัยหลัก 2 ประการ คือ



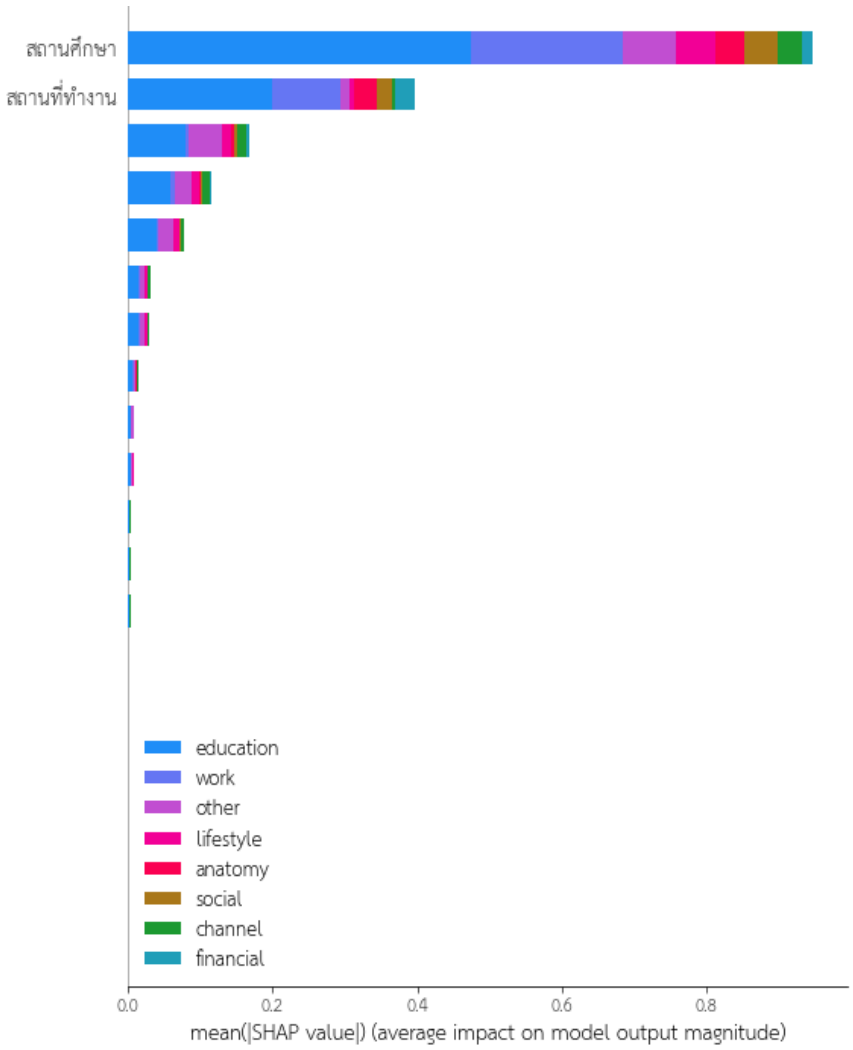
ความขัดแย้งในคำศัพท์ และความขัดแย้งในการจัดเรียงลำดับของคำศัพท์

ภาพที่ 20 SHAP แสดงคุณสมบัติของคำสำหรับคำถามที่แบบจำลองจำแนกประเภทผิด “สถานที่ทำงาน/สถานศึกษา”



ภาพที่ 21 SHAP แสดงคุณสมบัติของคำสำหรับคำถามที่คล้ายที่สุดกับคำถามที่แบบจำลองจำแนกประเภทผิด “สถานศึกษา”

จากภาพที่ 20 คำถามติดป้ายกำกับเป็น Work ซึ่งผู้วิจัยพิจารณาแล้วว่าติดป้ายกำกับถูกต้อง แต่แบบจำลองทายผิดเป็น Education ในขณะที่ภาพที่ 21 เป็นคำถามที่คล้ายกันมากที่สุด และติดป้ายกำกับเป็น Education



ภาพที่ 22 SHAP แสดงอิทธิพลของคำทั้งสองที่ส่งผลกระทบต่อคลาสทั้งหมด “สถานศึกษา” และ “สถานที่ทำงาน”

จากภาพที่ 22 คำว่า “สถานที่ทำงาน” และ “สถานศึกษา” พบในคลาส “Education” มากที่สุด และที่สำคัญเราพบว่า “สถานที่ทำงาน” พบในคลาส “Education” มากกว่าคลาส “Work” อีกด้วย ปัญหานี้ได้รับการแก้ไขโดยการเพิ่มข้อมูลด้วยตนเองด้วยการจัดเรียงคำใหม่ให้ครอบคลุมประเภทต่างๆมากยิ่งขึ้น หากขึ้นต้นด้วย “สถานที่ทำงาน” จะจัดเป็นคลาส “Work” และหากขึ้นต้นด้วย “สถานศึกษา” จะจัดเป็นคลาส “Education” สิ่งนี้ช่วยให้แบบจำลองเชื่อมโยงน้ำหนักของคำที่มีบริบท ซึ่งจะพบในสถานการณ์จริงในชุดข้อมูลได้มากขึ้น

การวัดประสิทธิภาพ

ในการทดลองนี้ใช้ TensorFlow Framework, Keras library และ Google Colab Pro โดยเลือกใช้ TPU runtime, การกำหนดค่า Loss นั้นใช้ Sparse Categorical Crossentropy อัลกอริทึมเพื่อเพิ่มประสิทธิภาพนั้นใช้ Adam Optimizer กำหนดค่า Learning rate เท่ากับ 0.001, epoch = 100, มีการแบ่งข้อมูลฝึก 80% และข้อมูลทดสอบ 20% จากนั้นทำ Class weight เพื่อปรับค่าน้ำหนักของคำถามแต่ละประเภทให้สมดุลกัน และการวัดประสิทธิภาพของแบบจำลองนั้นจะพิจารณาจากตาราง Confusion Matrix ซึ่งประกอบด้วยค่า Accuracy, Precision และ Recall

บทที่ 5

ผลการดำเนินการวิจัย

บทนี้จะกล่าวถึงผลการทดสอบแบบจำลอง, การเปรียบเทียบการใช้ Batch Normalization, การตามหาตัวอย่างจากข้อมูลทดสอบที่เครื่องทายผิดในลักษณะที่มั่นใจมาก, การตามหาประโยคที่มีความคล้ายกันสูงระหว่างตัวอย่างที่ทายผิด, การปรับปรุงข้อมูลชุดฝึกจากการหาประโยคที่มีความคล้ายกันสูงระหว่างตัวอย่างที่ทายผิด และการเปรียบเทียบประสิทธิภาพของแบบจำลองหลังปรับปรุงคุณภาพของชุดข้อมูลฝึก

ผลการทดสอบแบบจำลอง

ในงานวิจัยนี้วิธีการจำแนกประเภทคำถามของงานอีเว้นท์นั้นได้เลือกแบบจำลอง CNN, LSTM, BiLSTM, CNN-LSTM เพื่อนำมาเปรียบเทียบกับแบบจำลองที่เรานำเสนอในการทดลองนี้ จากผลการทดลองโดยใช้วิธีการประเมินผลโดยพิจารณาจากตาราง Confusion matrix แสดงในตารางที่ 13 โดยตัวชี้วัดที่นำมาประเมินผลคือค่า Accuracy, Precision และ Recall ที่ได้จากการทดลองของแต่ละแบบจำลองด้วยชุดข้อมูลทดสอบ

ตารางที่ 13 การเปรียบเทียบประสิทธิภาพของแบบจำลองทั้งหมด

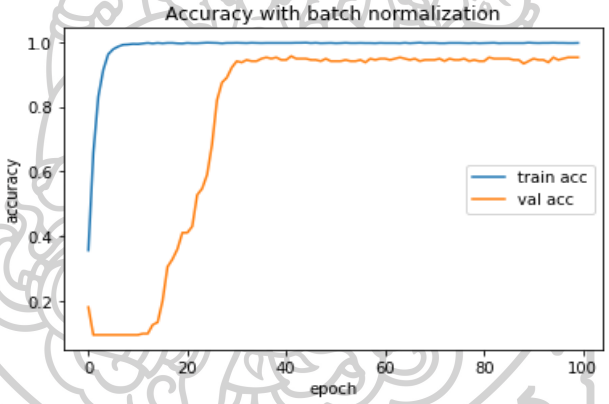
Model	Performance on Test Set		
	Accuracy	Precision	Recall
CNN	92.06	92.66	91.68
LSTM	92.31	92.59	90.96
BiLSTM	93.30	93.43	92.46
CNN-LSTM	93.55	94.04	93.07
CNN-BiLSTM	**94.54**	**94.11**	**93.87**

จากตารางที่ 13 แสดงผลลัพธ์การเปรียบเทียบประสิทธิภาพของแบบจำลองมทั้งหมด โดยจะเห็นได้ว่าแบบจำลอง CNN-BiLSTM นั้นให้ผลลัพธ์ค่าความแม่นยำเหนือกว่าแบบจำลองอื่นๆ เนื่องจากการใช้แบบจำลองแบบผสมที่นำ CNN ซึ่งมีประสิทธิภาพในการสกัดคุณลักษณะของข้อความได้เป็นอย่างดี โดยสังเกตได้จากลักษณะของคำถามส่วนใหญ่จะมีความยาวไม่มาก การกำหนดตัวกรอง

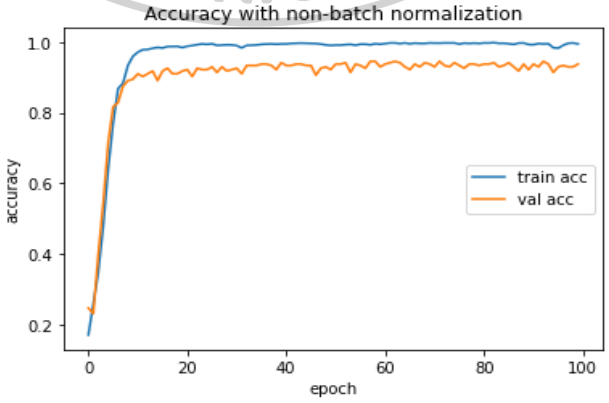
ของ CNN สามารถสกัดคุณลักษณะของคำถามได้อย่างมีประสิทธิภาพ ในส่วนของ BiLSTM สามารถทำงานได้ดีกับข้อความที่มีบริบทของเนื้อหาซึ่งมีส่วนสำคัญต่อความหมายของคำถาม เมื่อนำวิธีการทั้งสองแบบมาผสมกันจึงทำให้มีประสิทธิภาพมากยิ่งขึ้น โดยจะเห็นว่า CNN-LSTM ให้ผลลัพธ์ที่ดีมาเป็นอันดับที่สองเป็นเพราะว่า LSTM นั้นมีการมองเห็นข้อความแบบทิศทางเดียวในขณะที่ BiLSTM นั้นจะมองเห็นทั้งสองทิศทาง ซึ่งจะเห็นได้จากที่ BiLSTM นั้นก็มีผลลัพธ์ที่ดีกว่า LSTM เช่นเดียวกัน

การเปรียบเทียบการใช้ Batch Normalization

ในงานวิจัยนี้มีการใช้ Batch Normalization เพื่อเพิ่มประสิทธิภาพที่ดีที่สุด โดยนำแบบจำลองที่เรานำเสนอคือ CNN-BiLSTM เพื่อทดสอบความแตกต่างของการใช้ Batch Normalization โดยกำหนดให้แบบแรกใช้ Batch Normalization แบบที่สองไม่ใช้ Batch Normalization จากนั้นพิจารณาผลลัพธ์ค่าความแม่นยำและค่าความสูญเสียของชุดข้อมูลตรวจสอบ



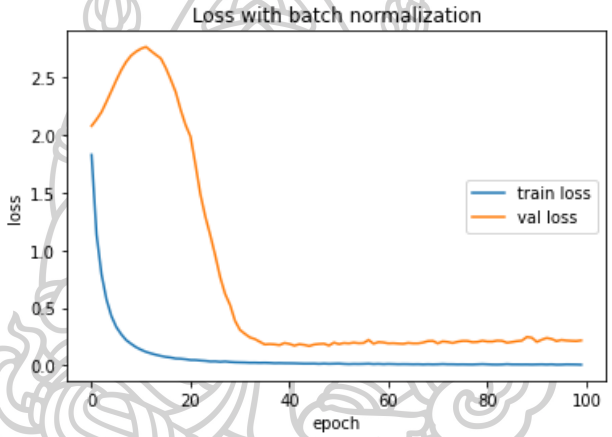
ภาพที่ 23 กราฟแสดงค่าความแม่นยำของแบบจำลองที่ใช้ Batch Normalization



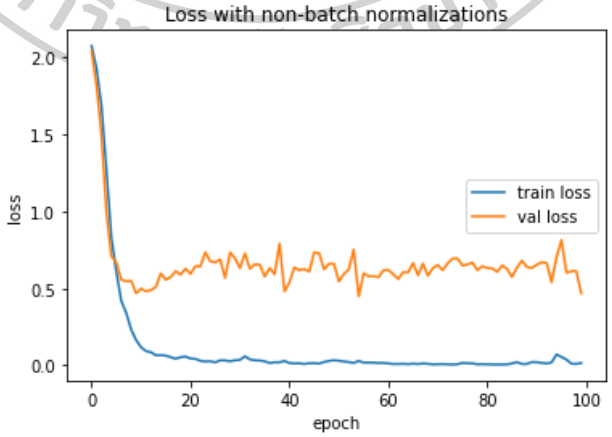
ภาพที่ 24 กราฟแสดงค่าความแม่นยำของแบบจำลองที่ไม่ใช้ Batch Normalization

จากภาพที่ 23 แบบจำลองที่ใช้ Batch Normalization นั้นเริ่มมีการเปลี่ยนแปลงของค่าความแม่นยำของชุดข้อมูลตรวจสอบใน epoch ที่ 14 จากนั้นค่าความแม่นยำของชุดข้อมูลตรวจสอบค่อยๆเพิ่มขึ้นเรื่อยๆโดยสามารถให้ค่าความแม่นยำที่สุงที่สุดคือ 95.70%

จากภาพที่ 24 แบบจำลองที่ไม่ใช้ Batch Normalization นั้นเริ่มมีการเปลี่ยนแปลงของค่าความแม่นยำของชุดข้อมูลตรวจสอบตั้งแต่ epoch แรกแต่สามารถให้ค่าความแม่นยำสูงสุดได้เพียง 94.53% ซึ่งถึงแม้ว่าในช่วงแรกแบบจำลองที่ใช้ Batch Normalization จะใช้เวลานานกว่าที่จะมีการเปลี่ยนแปลงค่าความแม่นยำของชุดข้อมูลตรวจสอบ แต่เมื่อค่าความแม่นยำมีการเปลี่ยนแปลงแล้วสามารถรีดประสิทธิภาพความแม่นยำได้สูงกว่าแบบไม่ใช้ Batch Normalization



ภาพที่ 25 กราฟแสดงค่าความสูญเสียของแบบจำลองที่ใช้ Batch Normalization



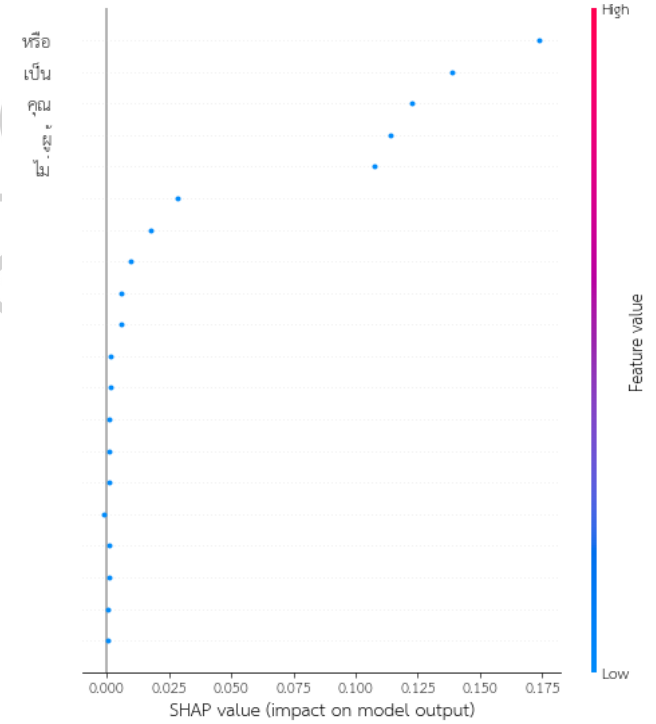
ภาพที่ 26 กราฟแสดงค่าความสูญเสียของแบบจำลองที่ไม่ใช้ Batch Normalization

จากภาพที่ 25 แบบจำลองที่ใช้ Batch Normalization นั้นเริ่มมีการเปลี่ยนแปลงของค่าความสูญเสียของชุดข้อมูลตรวจสอบใน epoch ที่ 21 จากนั้นค่าความสูญเสียของชุดข้อมูลตรวจสอบจึงค่อยๆลดลง

จากภาพที่ 26 แบบจำลองที่ไม่ใช้ Batch Normalization นั้นเริ่มมีการเปลี่ยนของค่าความสูญเสียของชุดข้อมูลตรวจสอบตั้งแต่ epoch แรก แต่เมื่อเปรียบเทียบแล้วจะเห็นได้ว่าแบบจำลองที่ใช้ Batch Normalization ให้ผลลัพธ์ของค่าความสูญเสียที่น้อยกว่าแบบไม่ใช้ Batch Normalization และยังมีความเสถียรดีที่มากอีกด้วย โดยค่าความสูญเสียนั้นยิ่งมีค่าน้อยยิ่งดีเพราะเป็นการที่สามารถบอกได้ว่าแบบจำลองนั้นมีการทำนายคำตอบที่ถูกต้องด้วยความมั่นใจที่สูงและมีความผิดพลาดน้อย

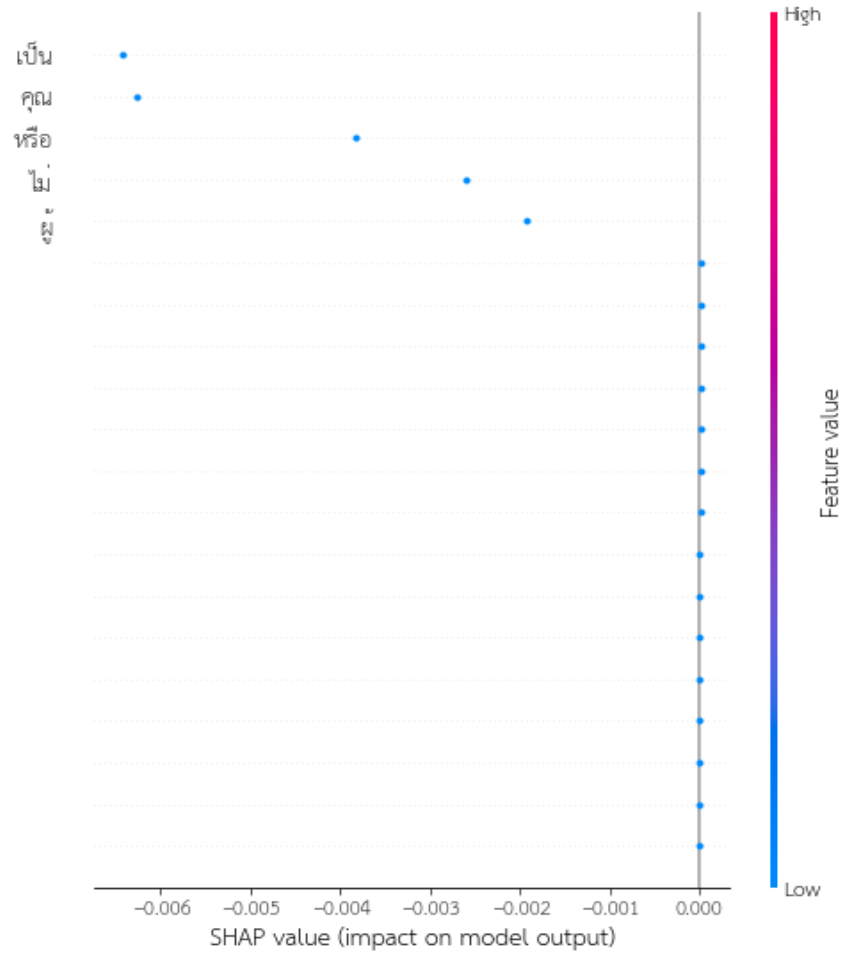
การตามหาตัวอย่างจากข้อมูลทดสอบที่เครื่องทายผิดในลักษณะที่มั่นใจมาก

ตัวอย่างที่ 1 “คุณเป็นผู้พิการหรือไม่” คำถามนี้ติดป้ายกำกับเป็น Anatomy แต่แบบจำลองทำนายผลลัพธ์เป็น Other ด้วยค่าความเชื่อมั่น 99.51% ซึ่งผลลัพธ์ดังกล่าวนั้นไม่ถูกต้อง



ภาพที่ 27 กราฟแสดงคำที่มีอิทธิพลต่อการทำนาย “คุณเป็นผู้พิการหรือไม่” ในประเภท Other

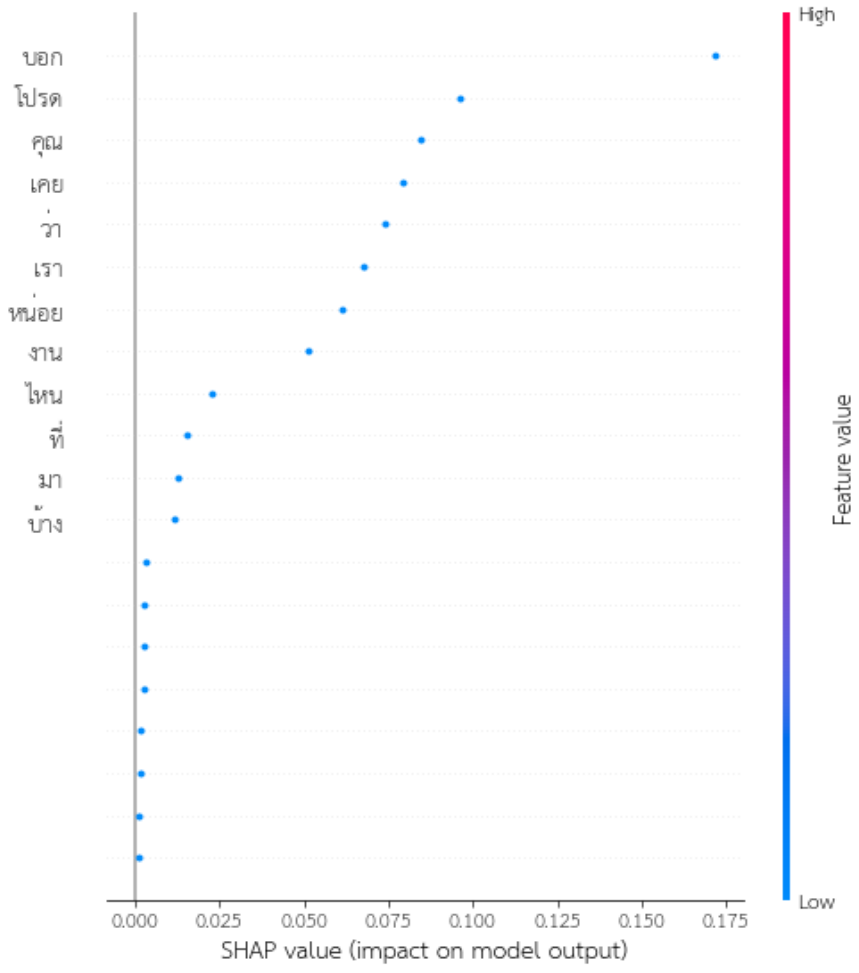
จากภาพที่ 27 แสดงให้เห็นว่าคำว่า “หรือ”, “เป็น”, “คุณ”, “ผู้”, “ไม่” มีอิทธิพลต่อการทำนายว่า “คุณเป็นผู้พิการหรือไม่” เป็น Other สูง



ภาพที่ 28 กราฟแสดงคำที่มีอิทธิพลต่อการทำนาย “คุณเป็นผู้พิการหรือไม่” ในประเภท Anatomy

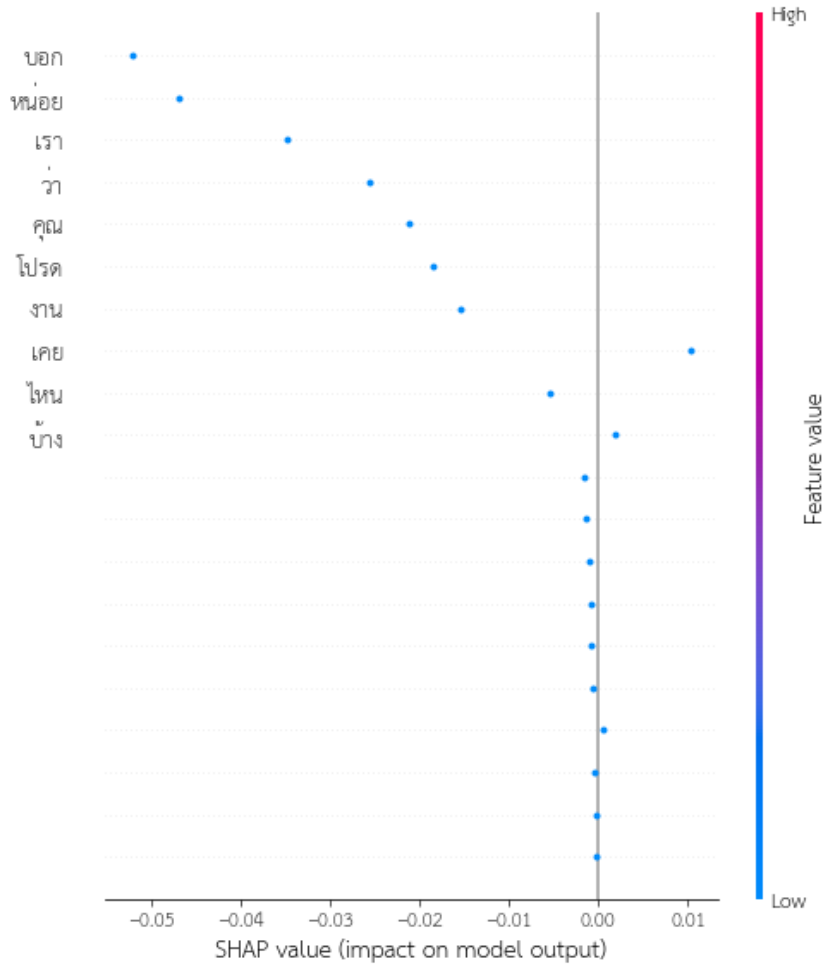
จากภาพที่ 28 แสดงให้เห็นว่าคำว่า “หรือ”, “เป็น”, “คุณ”, “ผู้”, “ไม่” ไม่มีอิทธิพลต่อการทำนายว่า “คุณเป็นผู้พิการหรือไม่” เป็น Anatomy เลย โดยที่จะสังเกตได้ว่าคำว่า “พิการ” นั้นหายไปเนื่องจากคำนี้ไม่เคยปรากฏอยู่ในคำถามใดเลยในชุดข้อมูลฝึก

ตัวอย่างที่ 2 “โปรดบอกเราหน่อยว่าคุณเคยสัมภาษณ์งานที่ไหนมาบ้าง” คำถามนี้ติดป้ายกำกับเป็น Work แต่แบบจำลองทำนายผลลัพธ์เป็น Finance ด้วยค่าความเชื่อมั่น 79.44% ซึ่งผลลัพธ์ดังกล่าวนั้นไม่ถูกต้อง



ภาพที่ 29 กราฟแสดงคำที่มีอิทธิพลต่อการทำนายของ “โปรดบอกเราหน่อยว่าคุณเคยสัมภาษณ์งานที่ไหนมาบ้าง” ในประเภท Finance

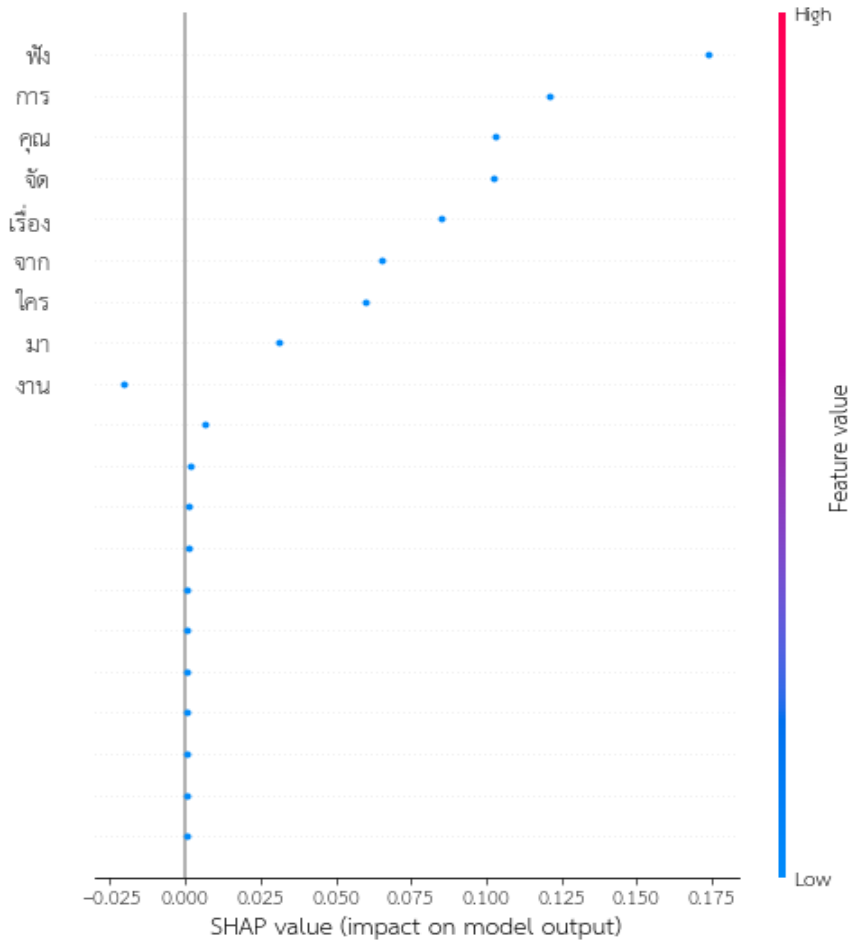
จากภาพที่ 29 แสดงให้เห็นว่าคำว่า “บอก”, “โปรด”, “คุณ”, “เคย”, “ว่า” มีอิทธิพลต่อการทำนายว่า “โปรดบอกเราหน่อยว่าคุณเคยสัมภาษณ์งานที่ไหนมาบ้าง” เป็นคำถามประเภท Finance สูง



ภาพที่ 30 กราฟแสดงคำที่มีอิทธิพลต่อการทำนายของ “โปรดบอกเราหน่อยว่าคุณเคยสัมภาษณ์งานที่ไหนมาบ้าง” ในประเภท Work

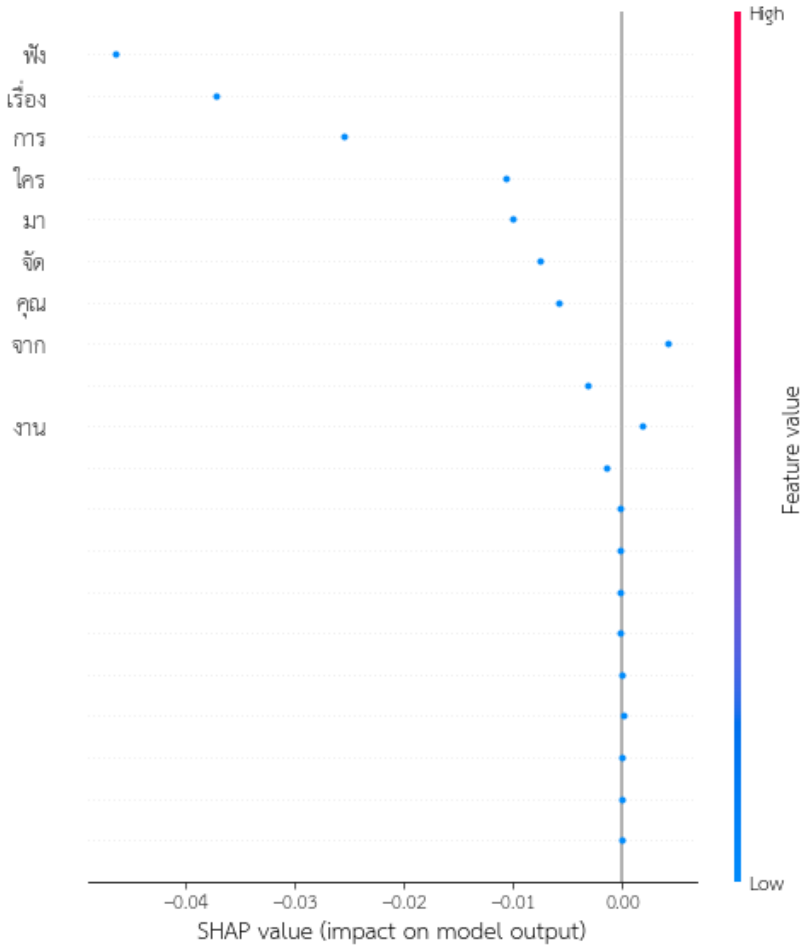
จากภาพที่ 30 แสดงให้เห็นว่าคำว่า “บอก”, “โปรด”, “คุณ”, “เคย”, “ว่า” ไม่มีอิทธิพลต่อการทำนายว่า “โปรดบอกเราหน่อยว่าคุณเคยสัมภาษณ์งานที่ไหนมาบ้าง” เป็นคำถามประเภท Work ในขณะที่คำว่า “เคย”, “บ้าง” มีอิทธิพลสูงต่อคำถามประเภท Work โดยจะสังเกตได้ว่าคำว่า “สัมภาษณ์” นั้นหายไปเนื่องจากคำนี้ไม่เคยปรากฏอยู่ในคำถามใดเลยในชุดข้อมูลฝึก

ตัวอย่างที่ 3 “คุณฟังเรื่องการจัดงานมาจากใคร” คำถามนี้ติดป้ายกำกับเป็น Channel แต่แบบจำลองทำนายผลลัพธ์เป็น Other ด้วยค่าความเชื่อมั่น 99.72% ซึ่งผลลัพธ์ดังกล่าวนั้นไม่ถูกต้อง



ภาพที่ 31 กราฟแสดงคำที่มีอิทธิพลต่อการทำนายของ “คุณฟังเรื่องการจัดงานมาจากใคร” ในประเภท Other

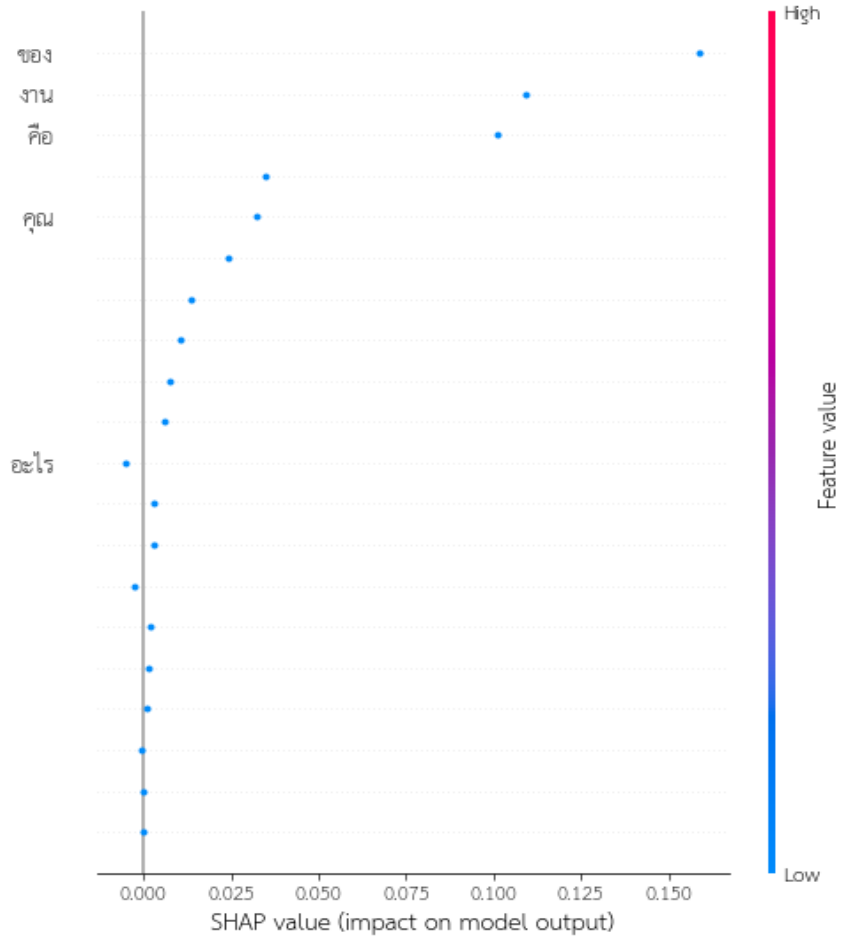
จากภาพที่ 31 แสดงให้เห็นว่าคำว่า “ฟัง”, “คุณ”, “การ”, “จัด”, “เรื่อง” มีอิทธิพลต่อการทำนายว่า “คุณฟังเรื่องการจัดงานมาจากใคร” เป็นคำถามประเภท Other ในขณะที่คำว่า “งาน” นั้นไม่มีอิทธิพลต่อคำถามประภท Other เลย



ภาพที่ 32 กราฟแสดงคำที่มีอิทธิพลต่อการทำนายของ “คุณฟังเรื่องการจัดงานมาจากใคร” ในประเภท Channel

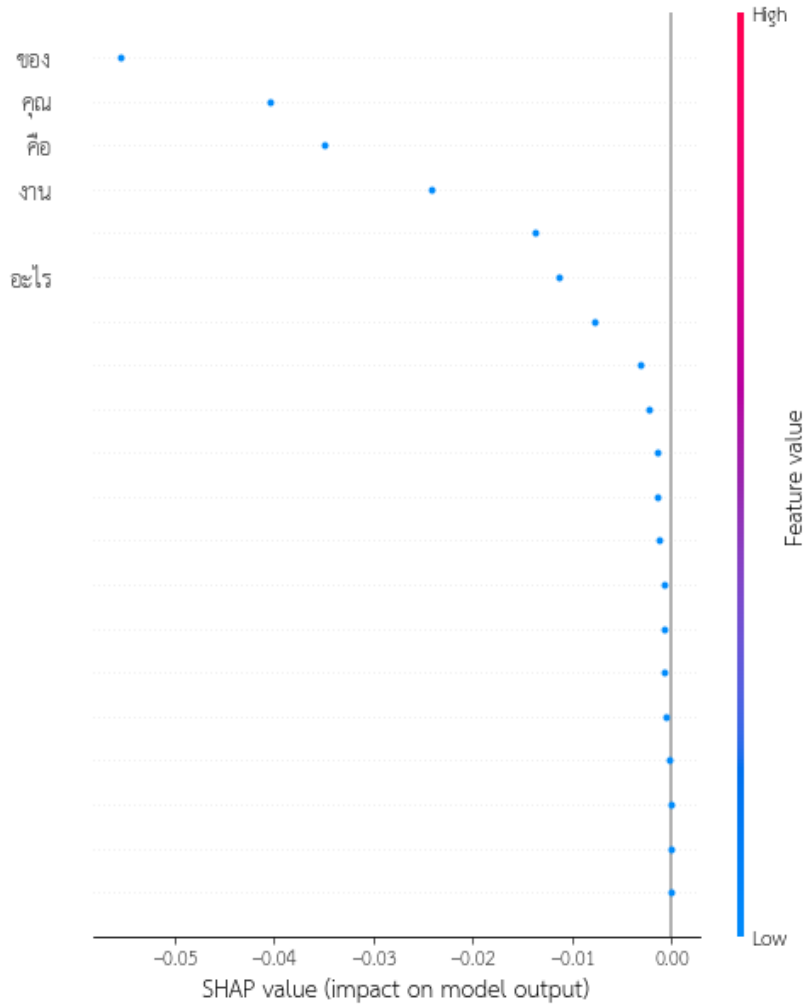
จากภาพที่ 32 แสดงให้เห็นว่าคำว่า “ฟัง”, “คุณ”, “การ”, “จัด”, “เรื่อง” ไม่มีอิทธิพลต่อการทำนายว่า “คุณฟังเรื่องการจัดงานมาจากใคร” เป็นคำถามประเภท Channel เลย ในขณะที่คำว่า “จาก”, “งาน” นั้นมีอิทธิพลต่อคำถามประภท Channel สูง

ตัวอย่างที่ 4 “งานอดิเรกของคุณคืออะไร” คำถามนี้ติดป้ายกำกับเป็น Lifestyle แต่แบบจำลองทำนายผลลัพธ์เป็น Work ด้วยค่าความเชื่อมั่น 72.34% ซึ่งผลลัพธ์ดังกล่าวนั้นไม่ถูกต้อง



ภาพที่ 33 กราฟแสดงคำที่มีอิทธิพลต่อการทำนายของ “งานอดิเรกของคุณคืออะไร” ในประเภท Work

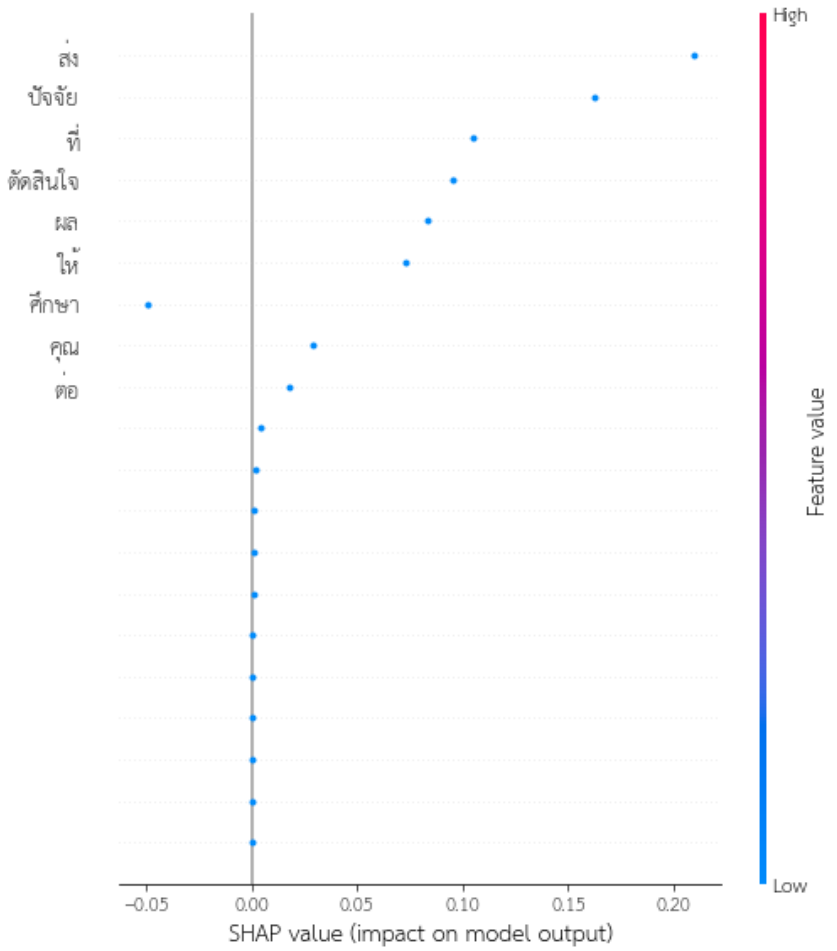
จากภาพที่ 33 แสดงให้เห็นว่าคำว่า “ของ”, “งาน”, “คือ”, “คุณ”, “อะไร” มีอิทธิพลต่อการทำนายว่า “งานอดิเรกของคุณคืออะไร” เป็นคำถามประเภท Work สูง ในขณะที่คำว่า “อะไร” นั้นไม่มีอิทธิพลต่อคำถามประภท Work เลย



ภาพที่ 34 กราฟแสดงคำที่มีอิทธิพลต่อการทำนายของ “งานอดิเรกของคุณคืออะไร” ในประเภท Lifestyle

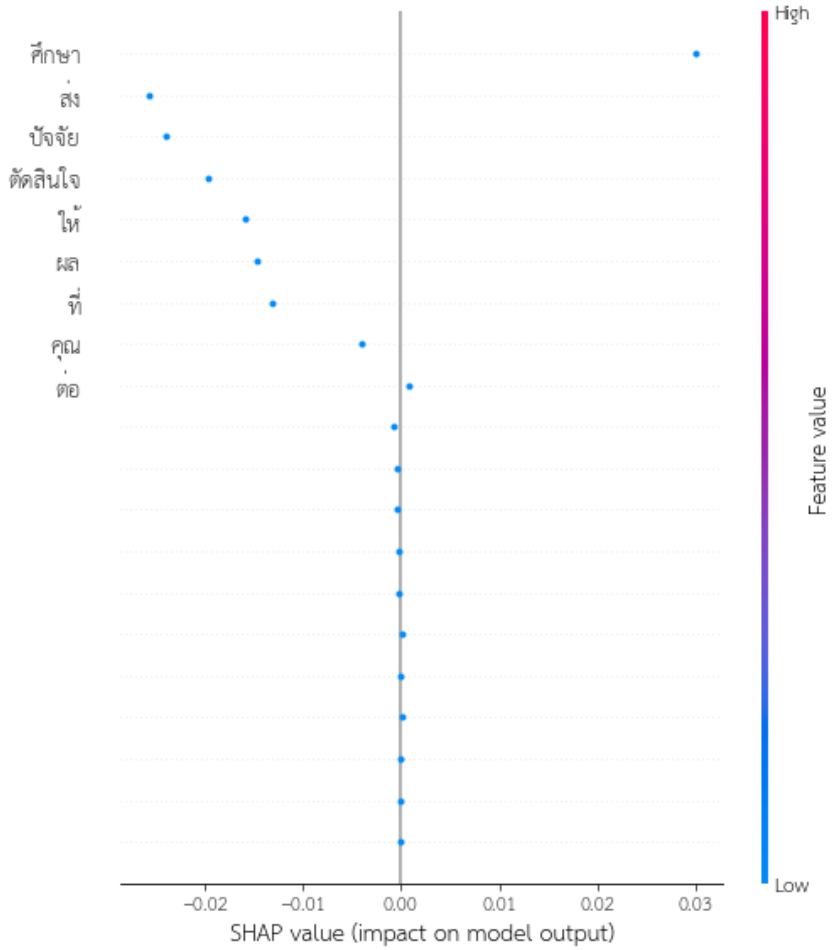
จากภาพที่ 34 แสดงให้เห็นว่าคำว่า “ของ”, “งาน”, “คือ”, “คุณ”, “อะไร” ไม่มีอิทธิพลต่อการทำนายว่า “งานอดิเรกของคุณคืออะไร” เป็นคำถามประเภท Lifestyle เลย โดยที่จะสังเกตได้ว่าคำว่า “อดิเรก” นั้นหายไปเนื่องจากคำนี้ไม่เคยปรากฏอยู่ในคำถามใดเลยในประโยคคำถามชุดทดสอบ

ตัวอย่างที่ 5 “ปัจจัยที่ส่งผลให้คุณตัดสินใจศึกษาต่อ” คำถามนี้ติดป้ายกำกับเป็น Education แต่แบบจำลองทำนายผลลัพธ์เป็น Other ด้วยค่าความเชื่อมั่น 99.87% ซึ่งผลลัพธ์ดังกล่าวนั้นไม่ถูกต้อง



ภาพที่ 35 กราฟแสดงคำที่มีอิทธิพลต่อการทำนายของ “ปัจจัยที่ส่งผลให้คุณตัดสินใจศึกษาต่อ” ในประเภท Other

จากภาพที่ 35 แสดงให้เห็นว่าคำว่า “ส่ง”, “ปัจจัย”, “ที่”, “ตัดสินใจ”, “ผล” มีอิทธิพลต่อการทำนายว่า “ปัจจัยที่ส่งผลให้คุณตัดสินใจศึกษาต่อ” เป็นคำถามประเภท Other สูง ในขณะที่คำว่า “ศึกษา” นั้นไม่มีอิทธิพลต่อคำถามประภท Other เลย



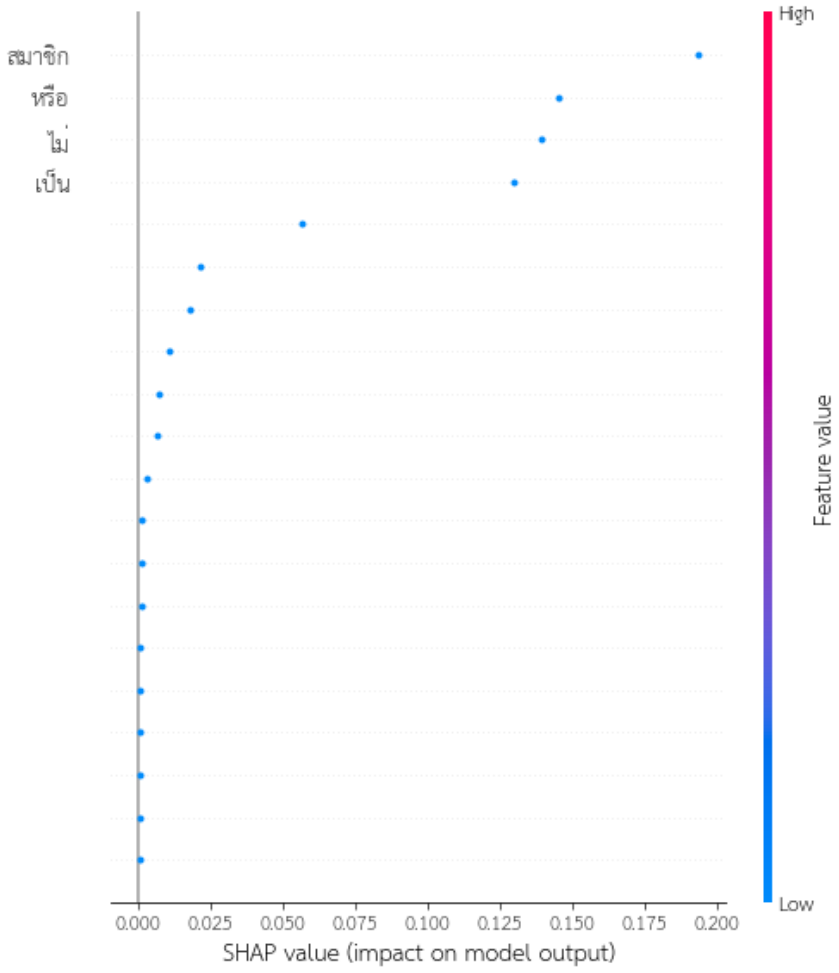
ภาพที่ 36 กราฟแสดงคำที่มีอิทธิพลต่อการทำนายของคำถาม “ปัจจัยที่ส่งผลให้คุณตัดสินใจศึกษาต่อ” ในประเภท Education

จากภาพที่ 36 แสดงให้เห็นว่าคำว่า “ส่ง”, “ปัจจัย”, “ที่”, “ตัดสินใจ”, “ผล” ไม่มีอิทธิพลต่อการทำนายว่า “ปัจจัยที่ส่งผลให้คุณตัดสินใจศึกษาต่อ” เป็นคำถามประเภท Education เลย ในขณะที่คำว่า “ศึกษา” นั้นมีอิทธิพลต่อคำถามประภท Education สูง

การตามหาประโยคที่มีความคล้ายกันสูงระหว่างตัวอย่างที่ทายผิด

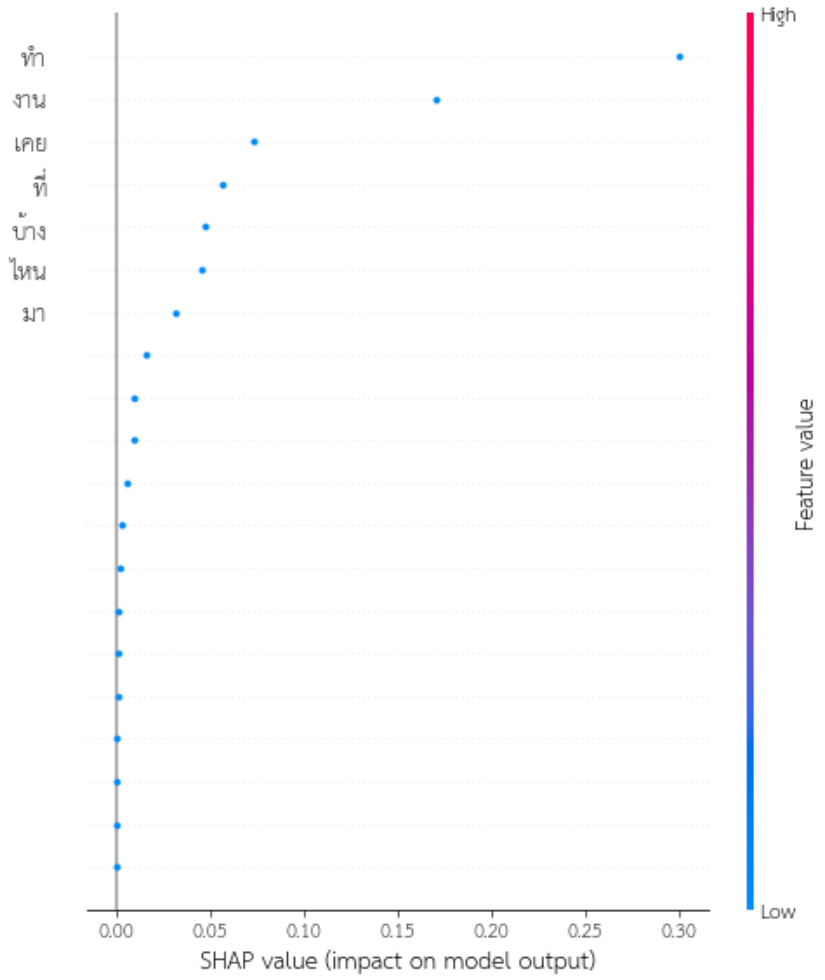
ผู้วิจัยจะตามหาประโยคที่มีความคล้ายกันสูงระหว่างตัวอย่างที่ทายผิด โดยใช้ Thai2Vec ซึ่งได้ประโยคคำถามทั้งหมดที่คล้ายกับคำถามตัวอย่างซึ่งประกอบไปด้วย “เป็นสมาชิกหรือไม่”, “เคยทำงานที่ไหนมาบ้าง”, “คุณฟังเรื่องการจัดงานมาจากใคร”, “งานที่ทำอยู่ตอนนี้คืองานอะไร”, “ปัจจัยที่ส่งผลให้คุณตัดสินใจศึกษาต่อ”

ประโยคคำถาม “เป็นสมาชิกหรือไม่” ซึ่งคล้ายกับ “คุณเป็นผู้พิการหรือไม่” ในคำถามตัวอย่างที่ 1 โดยคำถามนี้จัดเป็นคำถามประเภท Other ซึ่งมีความคล้ายคลึงกัน 83.87%



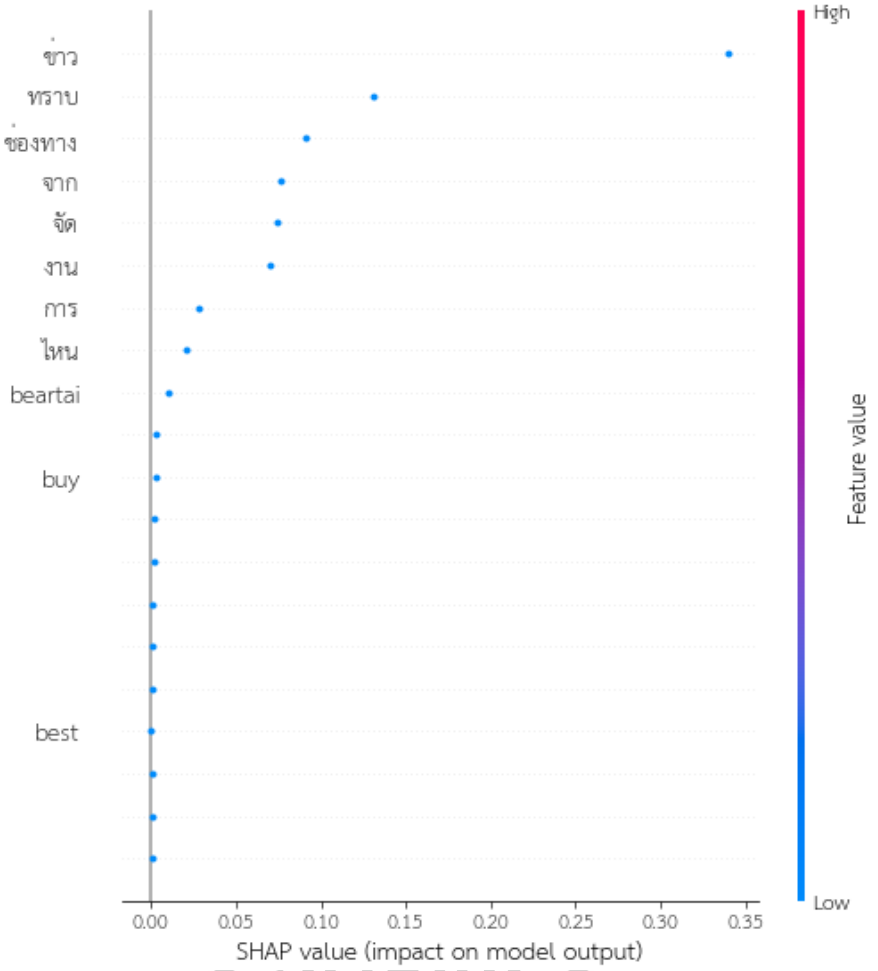
ภาพที่ 37 กราฟแสดงคำที่มีอิทธิพลต่อการทำนายของคำถาม “เป็นสมาชิกหรือไม่” ในประเภท Other

ประโยคคำถาม “เคยทำงานที่ไหนมาบ้าง” ซึ่งคล้ายกับ “โปรดบอกเราหน่อยว่าคุณเคยสัมภาษณ์งานที่ไหน มาบ้าง” ในคำถามตัวอย่างที่ 2 โดยคำถามนี้จัดเป็นคำถามประเภท Work ซึ่งมีความคล้ายคลึงกัน 82.50%



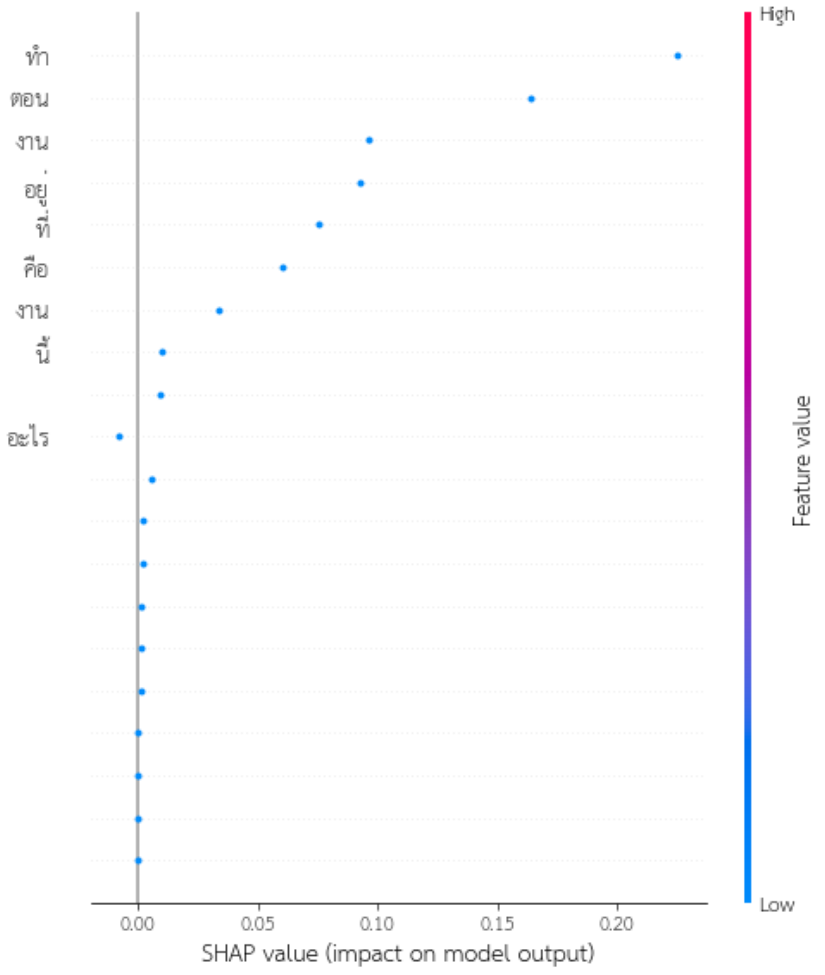
ภาพที่ 38 กราฟแสดงคำที่มีอิทธิพลต่อการทำนายของคำถาม “เคยทำงานที่ไหนมาบ้าง” ในประเภท Work

ประโยคคำถาม “ทราบข่าวการจัดงาน beartai best buy จากช่องทางไหน” ซึ่งคล้ายกับ “คุณฟังเรื่องการจัดงานมาจากใคร” ในคำถามตัวอย่างที่ 3 โดยคำถามนี้จัดเป็นคำถามประเภท Channel ซึ่งมีความคล้ายคลึงกัน 76.90%



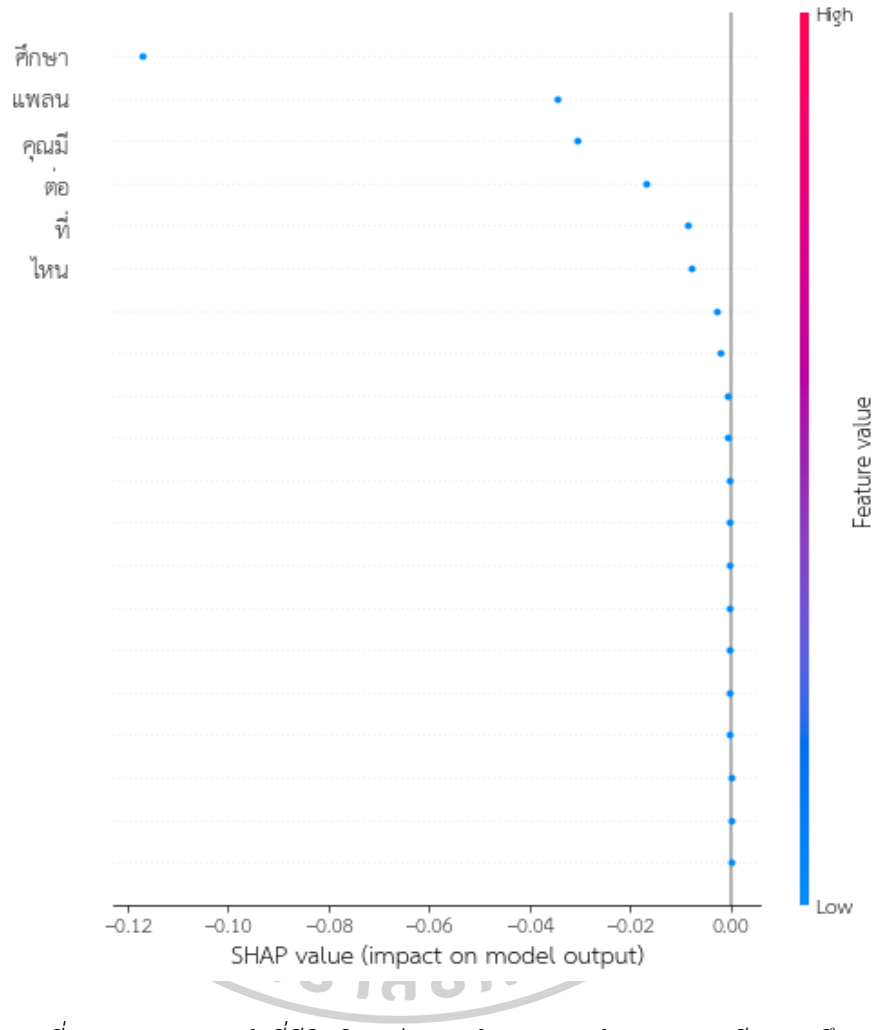
ภาพที่ 39 กราฟแสดงคำที่มีอิทธิพลต่อการทำนายของคำถาม “ทราบข่าวการจัดงาน beartai best buy จากช่องทางไหน” ในประเภท Channel

ประโยคคำถาม “งานที่ทำอยู่ตอนนี้คืองานอะไร” ซึ่งคล้ายกับ “งานอดิเรกของคุณคืออะไร” ในคำถามตัวอย่างที่ 4 โดยคำถามนี้จัดเป็นคำถามประเภท Work ซึ่งมีความคล้ายคลึงกัน 78.71%



ภาพที่ 40 กราฟแสดงคำที่มีอิทธิพลต่อการทำนายของคำถาม “งานที่ทำอยู่ตอนนี้คืองานอะไร” ในประเภท Work

ประโยคคำถาม “คุณมีแพลนศึกษาต่อที่ไหน” ซึ่งคล้ายกับ “ปัจจัยที่ส่งผลให้คุณตัดสินใจศึกษาต่อ” ในคำถามตัวอย่างที่ 5 โดยคำถามนี้จัดเป็นคำถามประเภท Education ซึ่งมีความคล้ายคลึงกัน 79.33%



ภาพที่ 41 กราฟแสดงคำที่มีอิทธิพลต่อการทำนายของคำถาม “คุณมีแพลนศึกษาต่อที่ไหน” ในประเภท Education

การเปรียบเทียบประโยคที่มีความคล้ายกันสูงระหว่างตัวอย่างที่ทายผิด

ผู้วิจัยจะนำคำถามตัวอย่างที่ทายผิดมาเปรียบเทียบกับคำถามที่มีความคล้ายคลังกันกับคำถามตัวอย่าง โดยจะพิจารณาว่ามีอะไรที่เป็นจุดแตกต่างกันและคล้ายกันบ้าง

จากภาพที่ 27 ซึ่งเป็นคำถามตัวอย่างที่ 1 คือ “คุณเป็นผู้พิการหรือไม่” มีกลุ่มคำ 5 อันดับแรกที่มีอิทธิพลสูงที่สุดได้แก่ “หรือ”, “เป็น”, “คุณ”, “ผู้”, “ไม่” ตามลำดับและภาพที่ 37 ซึ่งเป็นคำถามที่คล้ายกันกับตัวอย่างที่ 1 คือ “เป็นสมาชิกหรือไม่” มีกลุ่มคำ 5 อันดับแรกที่มีอิทธิพลสูงที่สุด

ได้แก่ “สมาชิก”, “หรือ”, “ไม่”, “เป็น” ตามลำดับ โดยเมื่อพิจารณาดูแล้วนั้นคำว่า “หรือ”, “เป็น”, “ไม่” เป็นคำที่มีอิทธิพลต่อการทำนายว่าเป็นคำถามประเภท Other สูง และเป็นคำที่มีความสอดคล้องกันระหว่าง 2 คำถาม โดยคำถาม “เป็นสมาชิกหรือไม่” นั้นจัดอยู่ในคำถามประเภท Other ซึ่งเมื่อผู้วิจัยได้ตรวจสอบข้อมูลดูแล้วพบว่าคำว่า “พิการ” ไม่ปรากฏอยู่ในคลังคำศัพท์เลย ซึ่งคำว่า “พิการ” เป็นคำที่ผู้วิจัยใช้พิจารณาให้เป็นคำถามประเภท Anatomy ซึ่งด้วยสาเหตุนี้แบบจำลองอาจพิจารณาว่าคำว่า “หรือ”, “เป็น”, “ไม่” นั้นมีอิทธิพลต่อคำถามประเภท Other มาก ดังนั้นจึงมีการทำนายคำถาม “คุณเป็นผู้พิการหรือไม่” ออกมาเป็นคำถามประเภท Other

จากภาพที่ 29 ซึ่งเป็นคำถามตัวอย่างที่ 2 คือ “โปรดบอกเราหน่อยว่าคุณเคยสัมภาษณ์งานที่ไหนมาบ้าง” มีกลุ่มคำ 5 อันดับแรกที่มีอิทธิพลสูงที่สุดได้แก่ “บอก”, “โปรด”, “คุณ”, “เคย”, “ว่า”, “เรา” ตามลำดับและภาพที่ 38 ซึ่งเป็นคำถามที่คล้ายกันกับตัวอย่างที่ 2 คือ “เคยทำงานที่ไหนมาบ้าง” มีกลุ่มคำ 5 อันดับแรกที่มีอิทธิพลสูงที่สุดได้แก่ “ทำ”, “งาน”, “เคย”, “ที่”, “บ้าง” ตามลำดับ โดยเมื่อพิจารณาดูแล้วนั้นคำว่า “เคย” เป็นคำเพียงคำเดียวที่มีความสอดคล้องกันระหว่าง 2 คำถาม และเราพบว่าคำว่า “บอก”, “โปรด”, “คุณ” นั้นมีอิทธิพลต่อการทำนายว่าเป็นคำถามประเภท Finance สูง ในขณะที่คำว่า “ทำ”, “งาน” เป็นคำที่มีอิทธิพลต่อการทำนายว่าเป็นคำถามประเภท Work สูง โดยคำถาม “เคยทำงานที่ไหนมาบ้าง” นั้นจัดอยู่ในคำถามประเภท Work ซึ่งเมื่อผู้วิจัยได้ตรวจสอบข้อมูลดูแล้วพบว่าคำว่า “สัมภาษณ์” ไม่ปรากฏอยู่ในคลังคำศัพท์เลยและคำว่า “บอก”, “โปรด”, “คุณ”, “เคย” นั้นปรากฏอยู่ในคำถามประเภท Finance ด้วยความถี่สูงมากกว่าคำว่า “งาน” ที่ปรากฏอยู่ในคำถามประเภท Work ด้วยสาเหตุนี้อาจทำให้แบบจำลองพิจารณาให้คำว่า “บอก”, “โปรด”, “คุณ” นั้นมีอิทธิพลต่อคำถามประเภท Finance มาก ดังนั้นจึงมีการทำนายคำถาม “โปรดบอกเราหน่อยว่าคุณเคยสัมภาษณ์งานที่ไหนมาบ้าง” ออกมาเป็นคำถามประเภท Finance

จากภาพที่ 31 ซึ่งเป็นคำถามตัวอย่างที่ 3 คือ “คุณฟังเรื่องการจัดงานมาจากใคร” มีกลุ่มคำ 5 อันดับแรกที่มีอิทธิพลสูงที่สุดได้แก่ “ฟัง”, “การ”, “คุณ”, “จัด”, “เรื่อง”, “จาก” ตามลำดับและภาพที่ 39 ซึ่งเป็นคำถามที่คล้ายกันกับตัวอย่างที่ 3 คือ “ทราบข่าวการจัดงาน beartai best buy จากช่องทางไหน” มีกลุ่มคำ 5 อันดับแรกที่มีอิทธิพลสูงที่สุดได้แก่ “ข่าว”, “ทราบ”, “ช่องทาง”, “จาก”, “จัด” ตามลำดับ โดยเมื่อพิจารณาดูแล้วนั้นคำว่า “จัด”, “จาก” เป็นคำที่มีอิทธิพลสูงต่อการทำนายของทั้ง 2 คำถามและเป็นคำที่มีความสอดคล้องกันระหว่าง 2 คำถาม โดยที่คำว่า “จัด”, “จาก” นั้นมีอิทธิพลสูงต่อคำถามประเภท Channel แต่มีอิทธิพลน้อยกว่า “ฟัง”, “การ”, “คุณ” ที่มี

อิทธิพลสูงต่อคำถามประเภท Other ซึ่งเมื่อผู้วิจัยได้ตรวจสอบข้อมูลดูแล้วพบว่าคำว่า “ฟัง” นั้นปรากฏอยู่ในคำถามประเภท Other ด้วยความถี่สูงมาก แต่ปรากฏได้น้อยกว่าในคำถามประเภท Channel ด้วยสาเหตุนี้อาจทำให้แบบจำลองพิจารณาให้คำว่า “ฟัง” นั้นมีอิทธิพลต่อคำถามประเภท Other มาก ดังนั้นจึงมีการทำนายคำถาม “คุณฟังเรื่องการจัดงานมาจากใคร” ออกมาเป็นคำถามประเภท Other

จากภาพที่ 33 ซึ่งเป็นคำถามตัวอย่างที่ 4 คือ “งานอดิเรกของคุณคืออะไร” มีกลุ่มคำ 5 อันดับแรกที่มีอิทธิพลสูงที่สุดได้แก่ “ของ”, “งาน”, “คือ”, “คุณ”, “อะไร” ตามลำดับและภาพที่ 40 ซึ่งเป็นคำถามที่คล้ายกันกับตัวอย่างที่ 4 คือ “งานที่ทำอยู่ตอนนี้คืองานอะไร” มีกลุ่มคำ 5 อันดับแรกที่มีอิทธิพลสูงที่สุดได้แก่ “ทำ”, “ตอน”, “งาน”, “อยู่”, “ที่” ตามลำดับ โดยเมื่อพิจารณาดูแล้วนั้นคำว่า “งาน” เป็นคำที่มีอิทธิพลต่อการทำนายว่าเป็นคำถามประเภท Work สูง และเป็นคำที่มีความสอดคล้องกันระหว่าง 2 คำถาม โดยคำถาม “งานที่ทำอยู่ตอนนี้คืองานอะไร” นั้นจัดอยู่ในคำถามประเภท Work ซึ่งเมื่อผู้วิจัยได้ตรวจสอบข้อมูลดูแล้วพบว่าคำว่า “อดิเรก” ไม่ปรากฏอยู่ในคลังคำศัพท์เลย ด้วยสาเหตุนี้อาจทำให้แบบจำลองพิจารณาว่า คำว่า “งาน” นั้นมีอิทธิพลต่อคำถามประเภท Work มาก ดังนั้นจึงมีการทำนายคำถาม “งานอดิเรกของคุณคืออะไร” ออกมาเป็นคำถามประเภท Work

จากภาพที่ 35 ซึ่งเป็นคำถามตัวอย่างที่ 5 คือ “ปัจจัยที่ส่งผลให้คุณตัดสินใจศึกษาต่อ” มีกลุ่มคำ 5 อันดับแรกที่มีอิทธิพลสูงที่สุดได้แก่ “ส่ง”, “ปัจจัย”, “ที่”, “ตัดสินใจ”, “ผล” ตามลำดับและภาพที่ 41 ซึ่งเป็นคำถามที่คล้ายกันกับตัวอย่างที่ 5 คือ “คุณมีแพลนศึกษาต่อที่ไหน” มีกลุ่มคำ 5 อันดับแรกที่มีอิทธิพลสูงที่สุดได้แก่ “ศึกษา”, “แพลน”, “คุณ”, “มี”, “ต่อ” ตามลำดับ โดยเมื่อพิจารณาดูแล้วนั้นคำที่มีอิทธิพลสูงต่อการทำนายของทั้ง 2 คำถามนั้นไม่มีคำใดเกี่ยวข้องกันเลย ในขณะที่คำว่า “ศึกษา” ของทั้ง 2 คำถามนั้นมีอิทธิพลสูงต่อการทำนายว่าเป็นคำถามประเภท Education ด้วยสาเหตุนี้อาจทำให้แบบจำลองพิจารณาว่า คำว่า “ส่ง”, “ปัจจัย” นั้นมีอิทธิพลต่อคำถามประเภท Other มาก ดังนั้นจึงมีการทำนายคำถาม “ปัจจัยที่ส่งผลให้คุณตัดสินใจศึกษาต่อ” ออกมาเป็นคำถามประเภท Other

การปรับปรุงข้อมูลชุดฝึกจากการหาประโยคที่มีความคล้ายกันสูงระหว่างตัวอย่างที่ทายผิด

ผู้วิจัยได้นำข้อสังเกตจากหัวข้อการเปรียบเทียบประโยคที่มีความคล้ายกันสูงระหว่างตัวอย่างที่ทายผิดมาวิเคราะห์เพื่อปรับปรุงข้อมูลชุดฝึกเพื่อให้แบบจำลองได้เรียนรู้ใหม่และติดตามผลการ

ปรับปรุงข้อผิดพลาดที่เกิดขึ้นว่ามีประสิทธิภาพที่ดีขึ้นหรือไม่

จากตัวอย่างที่ 1 “คุณเป็นผู้พิการหรือไม่” ผู้วิจัยได้เพิ่มประโยคคำถามที่มีคำว่าพิการเข้าไปในชุดฝึกเพิ่มเติมจำนวน 10 คำถาม เนื่องจากก่อนหน้านี้ไม่มีคำถามที่มีคำว่า “พิการ” อยู่ในชุดฝึกเลย โดยก่อนปรับปรุงข้อมูลชุดฝึกนั้นแบบจำลองทำนายผลลัพธ์ว่าเป็นคำถามประเภท Other ด้วยค่าความเชื่อมั่น 99.51% ซึ่งเป็นผลลัพธ์ที่ผิด แต่หลังจากที่มีการเพิ่มข้อมูลเข้าไป แบบจำลองทำนายผลลัพธ์ว่าเป็นคำถามประเภท Anatomy ด้วยค่าความเชื่อมั่น 99.84% ซึ่งเป็นผลลัพธ์ที่ถูกต้อง

จากตัวอย่างที่ 2 “โปรดบอกเราหน่อยว่าคุณเคยสัมภาษณ์งานที่ไหนมาบ้าง” ผู้วิจัยได้เพิ่มประโยคคำถามที่มีคำว่า “สัมภาษณ์” เข้าไปในชุดฝึกเพิ่มเติมจำนวน 10 คำถาม เนื่องจากก่อนหน้านี้ไม่มีคำถามที่มีคำว่า “สัมภาษณ์” อยู่ในชุดฝึกเลย โดยก่อนที่จะปรับปรุงข้อมูลชุดฝึกนั้นแบบจำลองทำนายผลลัพธ์ว่าเป็นคำถามประเภท Finance ด้วยค่าความเชื่อมั่น 99.51% ซึ่งเป็นผลลัพธ์ที่ผิด แต่หลังจากที่มีการเพิ่มข้อมูลเข้าไปแล้วนั้น แบบจำลองทำนายผลลัพธ์ว่าเป็นคำถามประเภท Work ด้วยค่าความเชื่อมั่น 92.00% ซึ่งเป็นผลลัพธ์ที่ถูกต้อง

จากตัวอย่างที่ 3 “คุณฟังเรื่องการจัดงานมาจากใคร” ผู้วิจัยได้เพิ่มประโยคคำถามที่มีคำว่า “ฟัง” ในคำถามประเภท Channel เข้าไปในชุดฝึกเพิ่มเติมจำนวน 10 คำถาม เนื่องจากก่อนหน้านี้คำว่า “ฟัง” ส่วนใหญ่จะอยู่ในคำถามประเภท Other และแทบไม่ปรากฏในคำถามประเภท Channel เลย โดยก่อนที่จะปรับปรุงข้อมูลชุดฝึกนั้นแบบจำลองทำนายผลลัพธ์ว่าเป็นคำถามประเภท Other ด้วยค่าความเชื่อมั่น 99.72% ซึ่งเป็นผลลัพธ์ที่ผิด แต่หลังจากที่มีการเพิ่มข้อมูลเข้าไปแล้วนั้น แบบจำลองทำนายผลลัพธ์ว่าเป็นคำถามประเภท Channel ด้วยค่าความเชื่อมั่น 99.92% ซึ่งเป็นผลลัพธ์ที่ถูกต้อง

จากตัวอย่างที่ 4 “งานอดิเรกของคุณคืออะไร” ผู้วิจัยได้เพิ่มประโยคคำถามที่มีคำว่า “อดิเรก” เข้าไปในชุดฝึกเพิ่มเติมจำนวน 10 คำถาม เนื่องจากก่อนหน้านี้ไม่มีคำถามที่มีคำว่า “อดิเรก” อยู่ในชุดฝึกเลย โดยก่อนที่จะปรับปรุงข้อมูลชุดฝึกนั้นแบบจำลองทำนายผลลัพธ์ว่าเป็นคำถามประเภท Work ด้วยค่าความเชื่อมั่น 72.34% ซึ่งเป็นผลลัพธ์ที่ผิด แต่หลังจากที่มีการเพิ่มข้อมูลเข้าไปแล้วนั้น แบบจำลองทำนายผลลัพธ์ว่าเป็นคำถามประเภท Work ด้วยค่าความเชื่อมั่น 99.81% ซึ่งเป็นผลลัพธ์ที่ถูกต้อง

จากตัวอย่างที่ 5 “ปัจจัยที่ส่งผลให้คุณตัดสินใจศึกษาต่อ” ผู้วิจัยได้ทดลองตัดคำว่า “ปัจจัย” และ “ส่ง” ออกจากคำถามในข้อมูลชุดฝึกเนื่องจากปรากฏเฉพาะบางคำถามและมีจำนวนไม่มากใน

ข้อมูลชุดฝึกทั้งหมด โดยก่อนที่จะปรับปรุงข้อมูลชุดฝึกนั้นแบบจำลองทำนายผลลัพธ์ว่าเป็นคำถามประเภท Other ด้วยค่าความเชื่อมั่น 99.87% ซึ่งเป็นผลลัพธ์ที่ผิด แต่หลังจากการที่ผู้วิจัยได้ทำความสะอาดข้อมูลแล้วนั้น แบบจำลองทำนายผลลัพธ์ว่าเป็นคำถามประเภท Education ด้วยค่าความเชื่อมั่น 98.42% ซึ่งเป็นผลลัพธ์ที่ถูกต้อง

การแสดงสถิติการปรับปรุงคุณภาพของชุดข้อมูลฝึกหลังจากตรวจพบปัญหา

ผู้วิจัยได้มีการปรับปรุงชุดข้อมูลฝึกหลังจากตรวจพบปัญหาที่เกิดขึ้นเพื่อให้ชุดข้อมูลมีความถูกต้องและมีตัวอย่างครอบคลุมกับปัญหาที่เกิดขึ้นในชุดข้อมูล

ตารางที่ 14 สถิติการปรับปรุงคุณภาพของชุดข้อมูลฝึกหลังจากตรวจพบปัญหา

Method	Sample	Percent
Edit label	6	+0.30%
Manual example synthesis & Additional example	120	+5.96%

จากตารางที่ 14 แสดงสถิติการปรับปรุงคุณภาพของชุดข้อมูลฝึกหลังจากตรวจพบปัญหา โดยกรณีปัญหาที่มีการติดป้ายกำกับผิดจะมีการแก้ไขป้ายกำกับใหม่เป็นจำนวน 6 ตัวอย่าง ซึ่งคิดเป็น 0.30% กรณีปัญหาที่ตัวอย่างมีจำนวนน้อยมากจะมีการสังเคราะห์ตัวอย่างด้วยตนเองกับการเพิ่มตัวอย่างเพิ่มเติมเป็นจำนวน 120 ตัวอย่าง โดยคิดเป็น 5.96% จากคำถามทั้งหมด 2,014 ตัวอย่าง

การเปรียบเทียบประสิทธิภาพของแบบจำลองก่อนและหลังปรับปรุงคุณภาพของชุดข้อมูลฝึก

ผู้วิจัยได้มีการเปรียบเทียบประสิทธิภาพของแบบจำลองก่อนและหลังปรับปรุงคุณภาพของชุดข้อมูลฝึกเพื่อประเมินผลแบบจำลองว่ามีประสิทธิภาพที่ดีขึ้นหรือไม่

ตารางที่ 15 ประสิทธิภาพของแบบจำลองก่อนปรับปรุงข้อมูลฝึก

Model	Validation		
	Accuracy	Precision	Recall
CNN	92.06	92.66	91.68
LSTM	92.31	92.59	90.96
BiLSTM	93.30	93.43	92.46
CNN-LSTM	93.55	94.04	93.07
CNN-BiLSTM	**94.54**	**94.11**	**93.87**

ตารางที่ 16 ประสิทธิภาพของแบบจำลองหลังปรับปรุงชุดข้อมูลฝึก

Model	Validation		
	Accuracy	Precision	Recall
CNN	93.91 (+1.85%)	94.75 (+2.09%)	93.55 (+1.87%)
LSTM	95.08 (+2.77%)	95.74 (+3.15%)	95.40 (+4.44%)
BiLSTM	95.32 (+2.02%)	96.46 (+3.03%)	95.19 (+2.73%)
CNN-LSTM	96.25 (+2.70%)	**97.53 (+3.49%)**	96.23 (+3.16%)
CNN-BiLSTM	**96.96 (+2.42%)**	96.75 (+2.64%)	**97.17 (+3.3%)**

จากตารางที่ 15 และ 16 แสดงการเปรียบเทียบประสิทธิภาพของแบบจำลองก่อนและหลังปรับปรุงชุดข้อมูลฝึกด้วยวิธีการที่เรานำเสนอ ผู้วิจัยพบว่าประสิทธิภาพการทำนายผลลัพธ์ของแบบจำลองทุกแบบดียิ่งขึ้นทั้งหมดหลังการปรับปรุงชุดข้อมูลด้วยวิธีการที่เรานำเสนอ โดยประสิทธิภาพของแบบจำลอง CNN-BiLSTM นั้นยังคงให้ผลลัพธ์ที่ดีที่สุด โดยค่า Accuracy เพิ่มขึ้น 2.42% ค่า Precision เพิ่มขึ้น 2.64% และค่า Recall เพิ่มขึ้น 3.3% นอกจากนี้แบบจำลองอื่นก็มีประสิทธิภาพที่ดีขึ้นทั้งหมดอีกด้วย สิ่งนี้แสดงให้เห็นถึงความสัมพันธ์กันระหว่างคุณภาพของข้อมูลชุดฝึกกับประสิทธิภาพของแบบจำลอง เมื่อข้อมูลที่เราใช้ฝึกอบรมแบบจำลองมีคุณภาพสูงขึ้นจะทำให้แบบจำลองมีประสิทธิภาพในการทำนายผลลัพธ์ที่ดียิ่งขึ้นด้วย

มหาวิทยาลัยศิลปากร

บทที่ 6

สรุปผลการดำเนินงานวิจัยและข้อเสนอแนะ

งานวิจัยนี้นำเสนอวิธีการจำแนกประเภทคำถามของงานอีเว้นท์ด้วยโครงข่ายประสาทเทียมแบบผสม CNN-BiLSTM เพื่อแก้ปัญหาการจำแนกคำถามของงานอีเว้นท์ที่ถูกจัดเก็บในฐานข้อมูลโดยไม่มีการระบุประเภทของคำถาม โดยเทคนิค CNN จะสามารถสกัดคุณลักษณะของคำถามได้อย่างรวดเร็วและมีประสิทธิภาพและเทคนิค BiLSTM จะสามารถพิจารณาความหมาย บริบทและความเชื่อมโยงของเนื้อหาในคำถามแบบสองทิศทาง โดยจะนำข้อความจากอดีต และอนาคตมาพิจารณาเข้าด้วยกัน ทำให้สามารถพิจารณาประโยคคำถามได้อย่างครอบคลุม ซึ่งเมื่อนำเทคนิคทั้ง 2 เทคนิคมาผสมกันจะช่วยให้จำแนกคำถามออกมาได้อย่างแม่นยำ และเราได้นำเครื่องมือ SHAP ซึ่งเป็นการใช้เอไอเชิงอธิบายช่วยในการอธิบายผลลัพธ์การทำนายของแบบจำลอง วิธีการของเราช่วยให้เข้าใจพฤติกรรมการทำนายผลลัพธ์ของแบบจำลองได้มากยิ่งขึ้น โดยเราสามารถอธิบายได้ว่าปัจจัยที่มีอิทธิพลต่อการทำนายผลลัพธ์ของแบบจำลองเกิดจากอะไรได้บ้าง นำไปสู่การค้นพบปัญหาที่อาจเกิดขึ้นในชุดข้อมูลฝึก และพิจารณาเพื่อปรับปรุงชุดข้อมูลให้มีคุณภาพครอบคลุมเหตุการณ์ใหม่ๆที่จะเกิดขึ้นได้เมื่อข้อมูลมีความหลากหลายมากยิ่งขึ้น อีกทั้งยังช่วยให้กระบวนการทดลองมีประสิทธิภาพและเมื่อชุดข้อมูลฝึกมีคุณภาพย่อมส่งผลให้แบบจำลองมีประสิทธิภาพการทำนายผลลัพธ์ที่ดีมากยิ่งขึ้นอีกด้วย

ผลลัพธ์การทดลองของเราแสดงให้เห็นว่าแบบจำลองที่เรานำเสนอ CNN-BiLSTM มีประสิทธิภาพเหนือกว่าแบบจำลองอื่นๆ และวิธีการของเราตรวจพบปัญหาในชุดข้อมูลช่วยให้เกิดการปรับปรุงชุดข้อมูลให้มีคุณภาพที่ดียิ่งขึ้น โดยปัญหาส่วนใหญ่ที่พบในชุดข้อมูลจะเป็นปัญหาตัวอย่างที่หายาก เนื่องจากข้อมูลต่อคลาสมีจำนวนค่อนข้างน้อย สิ่งเหล่านี้เป็นเรื่องปกติสำหรับชุดข้อมูลขนาดเล็กที่มีหลายหมวดหมู่ เมื่อเราทำการปรับปรุงชุดข้อมูล เราพบว่ามีเพียง 6 ตัวอย่างที่ติดป้ายกำกับใหม่ และมี 120 ตัวอย่างที่เพิ่มเข้ามาใหม่ วิธีการของเราลดการจำแนกประเภทผิดจาก 22 ตัวอย่างเหลือเพียง 13 ตัวอย่าง จาก 403 ตัวอย่าง โดยใช้แบบจำลองที่ดีที่สุดคือ CNN-BiLSTM ในการทดสอบ เราประเมินประสิทธิภาพของแบบจำลองทั้งหมดอีกครั้ง เราพบว่าแบบจำลอง CNN-BiLSTM ยังคงมีประสิทธิภาพที่ดีที่สุด และแบบจำลองอื่นๆก็มีประสิทธิภาพดีขึ้นเมื่อเทียบกับก่อนที่เราปรับปรุงคุณภาพชุดข้อมูล ซึ่งสิ่งเหล่านี้แสดงให้เห็นว่าประสิทธิภาพของแบบจำลองที่ดีขึ้นนั้นมีความสัมพันธ์กันกับคุณภาพของชุดข้อมูลที่มีความถูกต้องที่ได้รับการปรับปรุงด้วยวิธีการของเรา

รายการอ้างอิง

1. Cai, J., et al. *Deeplearning Model Used in Text Classification*. in 2018 15*th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*. 2018. Chengdu, China: IEEE.
2. Li, C., G. Zhan, and Z. Li. *News Text Classification Based on Improved Bi-LSTM-CNN*. in 2018 9*th International Conference on Information Technology in Medicine and Education (ITME)*. 2018. Hangzhou, China: IEEE.
3. She, X. and D. Zhang. *Text Classification Based on Hybrid CNN-LSTM Hybrid Model*. in 2018 11*th International Symposium on Computational Intelligence and Design (ISCID)*. 2018. Hangzhou, China: IEEE.
4. Yu, B., Q. Xu, and P. Zhang. *Question Classification Based on MAC-LSTM*. in 2018 *IEEE Third International Conference on Data Science in Cyberspace (DSC)*. 2018. Guangzhou, China: IEEE.
5. Jurafsky, D. and J.H. Martin, *Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition*. 2000, Upper Saddle River, New Jersey: Prentice Hall.
6. LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning.* Nature 2015. 521: p. 436–444.
7. Deng, L. and D. Yu, *Deep Learning: Methods and Applications.* Foundations and Trends® in Signal Processing, 2014. 7(3-4): p. 197-387.
8. Saha, S. *A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way*. 2018, Dec 16 Oct 12, 2021]; Available from: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53.
9. Schuster, M. and K.K. Paliwal, *Bidirectional recurrent neural networks.* IEEE Transactions on Signal Processing, 1997. 45(11): p. 2673 - 2681.
10. Collis, J. *Glossary of Deep Learning: Batch Normalisation*. 2017, Jun 28 Oct 12, 2021]; Available from: https://medium.com/deeper-learning/glossary-of-deep-learning-batch-normalisation-8266dcd2fa82.
11. Lundberg, S. *SHAP (SHapley Additive exPlanations)*. 2018 Oct 12, 2021]; Available

from: https://shap.readthedocs.io/en/latest/index.html.

12. Gopinath, D. *The Shapley Value for ML Models*. 2021, Oct 26 19 Nov, 2021]; Available from: https://towardsdatascience.com/the-shapley-value-for-ml-models-f1100bff78d1.

มหาวิทยาลัยศิลปากร

ประวัติผู้เขียน

ชื่อ-สกุล	นายอภินันท์ พีรชัยเดโช
วัน เดือน ปี เกิด	02 สิงหาคม 2537
สถานที่เกิด	เขตบางกอกน้อย กรุงเทพมหานคร
วุฒิการศึกษา	สำเร็จการศึกษาปริญญาวิทยาศาสตร์บัณฑิต สาขาเทคโนโลยีสารสนเทศ มหาวิทยาลัยศิลปากร พ.ศ.2559
ที่อยู่ปัจจุบัน	39 หมู่ 3 ต.สวนกล้วย อ.บ้านโป่ง จ.ราชบุรี 70110