



ศึกษาการใช้งานเทคนิค Reinforcement ในการใส่คำบรรยายภาพอัตโนมัติภาษาไทย
Study of Reinforcement Techniques for Automatic Image Captioning in
Thai

พุทธรพงษ์ ด้านทองกลาง รหัส 62172110139-2

รายงานนี้เป็นส่วนหนึ่งของการศึกษาวิชาสหกิจศึกษา
สาขาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์และเทคโนโลยี
มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน
ปีการศึกษา 2565

ศึกษาการใช้งานเทคนิค Reinforcement ในการใส่คำบรรยายภาพอัตโนมัติ
ภาษาไทย

Study of Reinforcement Techniques for Automatic Image Captioning in
Thai

พุทธรพงษ์ ด้านทองกลาง รหัส 62172110139-2

รายงานนี้เป็นส่วนหนึ่งของการศึกษาวิชาสหกิจศึกษา
สาขาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์และเทคโนโลยี
มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน
ปีการศึกษา 2565

| | |
|------------------|--|
| ชื่อปริญญาบัตร | ศึกษาการใช้งานเทคนิค Reinforcement ในการใส่คำบรรยายภาพอัตโนมัติ ภาษาไทย |
| นักศึกษา | นายพุทธพงษ์ ด้านทองกลาง รหัส 62172110139-2 |
| ที่ปรึกษาโครงการ | นายสรรพฤทธิ์ มฤคทัต |
| สาขา | วิศวกรรมคอมพิวเตอร์ |
| ปีการศึกษา | 2565 |

บทคัดย่อ

เนื่องจากการพัฒนาอย่างรวดเร็วของปัญญาประดิษฐ์ (AI) ในช่วงไม่กี่ปีที่ผ่านมา งานวิจัยด้านหนึ่งของ AI ที่น่าสนใจคือการพัฒนาคำบรรยายภาพอัตโนมัติโดยใช้โครงข่ายประสาทเทียมเชิงลึก วิธีการนี้ได้รับแรงบันดาลใจจากกระบวนการเข้ารหัส/ถอดรหัส ซึ่งใช้โครงข่ายประสาทเทียมเชิงลึกเพื่อเข้ารหัสรูปภาพ และใช้การเรียนรู้ความจำระยะสั้นและระยะยาวเพื่อถอดรหัสเป็นคำบรรยาย อย่างไรก็ตาม ปัญหาคือระบบคำบรรยายภาพปัจจุบันให้คำบรรยายเป็นภาษาอังกฤษเท่านั้น เนื่องจากข้อมูลการฝึกอบรมที่ใช้สำหรับโมเดลเป็นภาษาอังกฤษ ดังนั้นคำบรรยายจำเป็นต้องแปลเป็นภาษาไทย ซึ่งอาจนำไปสู่ความไม่ถูกต้องในการแปลคำบรรยายโดยอัตโนมัติ เพื่อแก้ไขปัญหานี้ จึงมีการศึกษาการใช้เทคนิคการเสริมแรงเพื่อสร้างคำบรรยายภาพเป็นภาษาไทยโดยอัตโนมัติเพื่อให้คำบรรยายภาพอัตโนมัติเป็นภาษาไทยมีความแม่นยำและสะดวกในการใช้งานมากยิ่งขึ้น สิ่งนี้มีศักยภาพที่จะเป็นประโยชน์ต่อประชาชนทั่วไปที่ต้องการสร้างคำบรรยายภาพอัตโนมัติเป็นภาษาไทย เนื่องจากระบบมีความแม่นยำมากกว่า 90% ในการสร้างคำบรรยายภาพภาษาไทย โครงการนี้สามารถพัฒนาเพิ่มเติมเพื่อปรับปรุงความแม่นยำของคำบรรยายภาพอัตโนมัติเป็นภาษาไทยและเป็นประโยชน์ต่อประชาชนทั่วไป

คำสำคัญ : การเรียนรู้แบบเสริมกำลัง การฝึกลำดับความสำคัญด้วยตัวเอง ปัญญาประดิษฐ์ใส่คำบรรยายภาพอัตโนมัติ

Project title: Study of Reinforcement Techniques for Automatic Image Captioning in Thai

Candidates: Mr.Putthapong Danthonglang Code 62172110139-2

Adviser: Mr.Sanparith Marukatat

Department: Computer Engineering

Academic Year: 2022

Abstract

Due to the rapid development of Artificial Intelligence (AI) in recent years, one interesting area of AI research is the development of automatic image captioning using deep neural networks. This approach is inspired by the process of encoding/decoding, which uses deep neural networks to encode images and uses short- and long-term memory learning to decode them into captions. However, the problem is that the current image captioning systems only provide captions in English, as the training data used for the model is in English. Therefore, the captions need to be translated into Thai, which may lead to inaccuracies in the automatic translation of the captions. To address this issue, the use of Reinforcement techniques to automatically generate captions in Thai have studied to provide more accurate and convenient to use automatic image captioning in Thai. This has the potential to benefit the general public who wish to generate automatic image captions in Thai, as the system has achieved over 90% accuracy in generating Thai captions. This research can be further developed to improve the accuracy of automatic image captioning in Thai and benefit a wider audience.

Keywords : Reinforcement Learning Self-critical Sequence Training Artificial Intelligence automatic captioning.

กิตติกรรมประกาศ

การที่ข้าพเจ้าได้มาปฏิบัติงานสหกิจศึกษา ณ ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ NECTEC ตั้งแต่วันที่ 7 พฤศจิกายน พ.ศ.2565 ถึงวันที่ 24 กุมภาพันธ์ พ.ศ.2566 ส่งผลให้ข้าพเจ้าได้รับความรู้และประสบการณ์ต่าง ๆ ที่มีคุณค่ามากมาย สำหรับรายงานสหกิจศึกษานี้สำเร็จลงได้ด้วยดีจากความร่วมมือและการสนับสนุนจากหลายฝ่าย ดังนี้

1. อาจารย์ สุภัทรา เกิดเมฆ (อาจารย์ประสานงาน)
2. อาจารย์ กิระชาติ สุขสุทธิ (อาจารย์นิเทศ)
3. อาจารย์ รติพร จันทร์กลั่น (อาจารย์นิเทศ)

นอกจากนี้ขอขอบคุณบิดา-มารดา อาจารย์ และบุคคลท่านอื่น ๆ ที่ไม่ได้กล่าว ณ ที่นี้ซึ่งท่านเหล่านั้นได้กรุณาช่วยเหลือให้คำแนะนำในการจัดทำโครงการ และรายงานฉบับนี้ จึงขอขอบพระคุณทุกท่านที่ได้มีส่วนร่วมในการให้ข้อมูลคำแนะนำ และวิธีการปฏิบัติงาน รวมถึงเป็นที่ปรึกษาในการจัดทำโครงการ และรายงานฉบับนี้จนเสร็จสมบูรณ์

พุทธพงษ์ ด้านทองกลาง

สารบัญ

| | หน้า |
|--|------|
| บทคัดย่อไทย | ก |
| ABSTRACT | ข |
| กิตติกรรมประกาศ | ค |
| สารบัญ | ง |
| สารบัญตาราง | ช |
| สารบัญรูป | ซ |
| บทที่ 1 บทนำ | 1 |
| 1.1 ความเป็นมาและความสำคัญของปัญหา | 1 |
| 1.2 วัตถุประสงค์ของปริญญานิพนธ์ | 2 |
| 1.3 ขอบเขตของปริญญานิพนธ์ | 2 |
| 1.4 ประโยชน์ที่คาดว่าจะได้รับ | 2 |
| 1.5 ขั้นตอนการดำเนินงาน | 2 |
| 1.6 แผนการดำเนินงาน | 3 |
| บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง | 5 |
| 2.1 ภาษาไพธอน | 6 |
| 2.2 เฟลค | 8 |
| 2.3 ชุดข้อมูล MSCOCO | 8 |
| 2.4 การประมวลผลภาษาธรรมชาติ | 9 |
| 2.5 ระบบสืออาร์จีบี | 10 |
| 2.6 โครงข่ายประสาทเทียม | 12 |
| 2.7 การเรียนรู้เชิงลึก | 13 |
| 2.8 โครงข่ายประสาทเทียมแบบคอนโวลูชัน | 16 |
| 2.9 การเรียนรู้แบบหน่วยความจำระยะสั้นระยะยาว | 21 |
| 2.10 โครงข่ายประสาทเทียมของโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ | 27 |

สารบัญ (ต่อ)

| | หน้า |
|---|-----------|
| 2.11 การเรียนรู้โครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์สวิน | 29 |
| 2.12 การแทนตัวเข้ารหัสแบบสองทิศทางจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ | 31 |
| 2.13 การเรียนรู้แบบฝึกอบรมล่วงหน้า BERT ที่ปรับให้เหมาะสมอย่างมีประสิทธิภาพจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ | 33 |
| 2.14 การเรียนรู้เกี่ยวกับลำดับตัวเข้ารหัส-ตัวถอดรหัสเป็นรูปแบบลำดับ | 34 |
| 2.15 การเรียนรู้แบบเสริมกำลัง | 35 |
| 2.16 การฝึกลำดับความสำคัญด้วยตัวเอง | 37 |
| 2.17 การเพิ่มประสิทธิภาพนโยบายใกล้เคียง | 38 |
| 2.18 เกณฑ์ที่ใช้ในการวัดประสิทธิภาพของโมเดลของการใส่คำบรรยายภาพ | 40 |
| บทที่ 3 วิธีการดำเนิน | 43 |
| 3.1 การใช้งานเทคนิค Reinforcement ในการใส่คำบรรยายภาพอัตโนมัติภาษาไทย | 43 |
| 3.2 อัปโหลตไฟล์รูปภาพและการประมวลผลภาพ | 45 |
| 3.3 การใส่คำบรรยายภาพภาษาไทยด้วยเทคนิคฝึกลำดับความสำคัญด้วยตนเองโดยใช้โครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ | 46 |
| 3.4 ผลลัพธ์ของการใส่คำบรรยายภาพอัตโนมัติภาษาไทย | 53 |
| บทที่ 4 การทดลองและผลการดำเนินงาน | 54 |
| 4.1 ทดสอบความสามารถตามขอบเขตที่กำหนด | 54 |
| 4.2 ความต้องการขั้นต่ำของระบบในการทดลอง | 55 |
| 4.3 ทดสอบการอัปโหลตไฟล์ภาพที่มีนามสกุล JPEG JPG และ PNG | 56 |
| 4.4 ทดสอบการใส่คำบรรยายภาพอัตโนมัติภาษาไทย | 58 |
| 4.5 ทดสอบการใส่คำบรรยายภาพอัตโนมัติภาษาไทยครั้งละ 1 ภาพ | 59 |

สารบัญ (ต่อ)

| | หน้า |
|---|------|
| บทที่ 5 บทสรุปการดำเนินงานและข้อเสนอแนะ | 62 |
| 5.1 สรุปผลการทดลอง | 62 |
| 5.2 อุปสรรคและปัญหาในการดำเนินงาน | 63 |
| 5.3 แนวทางการแก้ไขปัญหา | 63 |
| 5.4 แนวทางในการพัฒนาและข้อเสนอแนะ | 63 |
| เอกสารอ้างอิง | 64 |
| ภาคผนวก | 67 |
| ภาคผนวก ก | 68 |
| ประวัติผู้จัดทำปริญญานิพนธ์ | 72 |

สารบัญตาราง

| ตารางที่ | หน้า |
|--|------|
| 1.1 ขั้นตอนการดำเนินงานวิชาสหกิจศึกษา | 4 |
| 4.1 การทดสอบความสามารถตามขอบเขตที่กำหนด | 54 |
| 4.2 การตั้งค่าการทดลองประสิทธิภาพของระบบ | 55 |

ตัวอย่าง

สารบัญรูป

| รูปที่ | หน้า |
|---|------|
| 2.1 ผลการรันจากการเขียนโค้ดในโปรแกรมที่ 2.1 | 7 |
| 2.2 ตัวอย่างชุดข้อมูลรูปภาพ | 8 |
| 2.3 ตัวอย่างการนำการประมวลผลภาษาธรรมชาติไปใช้งาน | 9 |
| 2.4 แสดงโมเดลระบบสืออาร์จีบี | 10 |
| 2.5 ตัวอย่างภาพที่มีการผสมของแสงสีทั้ง 3 แชนแนล | 11 |
| 2.6 ตัวอย่างโมเดลการจำแนกรูปภาพเอกซเรย์บริเวณทรวงอก | 12 |
| 2.7 เซลล์ประสาทซึ่งมันถูกแสดงด้วยสัญลักษณ์วงกลม | 14 |
| 2.8 เซลล์ประสาทการทำนายราคาตัวเครื่องบิน | 15 |
| 2.9 ตัวอย่างการเลือกพื้นที่ในการจำแนก | 16 |
| 2.10 ภาพเดียวกันแต่ถูกย่อให้มีขนาดเล็กลง | 17 |
| 2.11 โครงสร้างการทำงานของโครงข่ายประสาทเทียมแบบคอนโวลูชัน | 17 |
| 2.12 ตัวอย่างการกรองด้วยวิธีพูลลิงค่าสูงสุด (Max Pooling) | 18 |
| 2.13 ตัวอย่างของการสร้างฟังก์ชันลักษณะ โดยกำหนดค่าสไตรด์เท่ากับ 2 และตัวกรองเท่ากับ 3x3 | 19 |
| 2.14 ตัวอย่างการเพิ่มแพดดิ้ง โดยกำหนดค่าสไตรด์ เท่ากับ 1 และตัวกรอง เท่ากับ 3x3 | 20 |
| 2.15 ตัวอย่างการพูลลิง โดยกำหนดค่าสไตรด์เท่ากับ 1 และตัวกรอง เท่ากับ 3x3 | 21 |
| 2.16 ตัวอย่างกระบวนการโครงข่ายประสาทเทียมที่เกิดซ้ำ | 22 |
| 2.17 สมการโครงข่ายประสาทเทียมที่เกิดซ้ำ | 23 |
| 2.18 กระบวนการทำงานของการเรียนรู้แบบหน่วยความจำระยะสั้นระยะยาว | 24 |
| 2.19 ตัวอย่างสมการ forget gate | 24 |
| 2.20 ตัวอย่างสมการ input gate ที่ใช้ในควบคุมการอัปเดต cell state | 25 |
| 2.21 ตัวอย่างสมการ Input modulation gate ที่ใช้ในการอัปเดตค่า cell state | 25 |

สารบัญรูป (ต่อ)

| รูปที่ | หน้า |
|---|------|
| 2.22 ตัวอย่างสมการของการอัปเดตค่า cell state | 26 |
| 2.23 ตัวอย่างสมการของ output gate ที่ใช้ในการควบคุมการ read ใน sequence ถัดไป | 26 |
| 2.24 ตัวอย่างสมการของ ht ของ sequence ถัดไป | 27 |
| 2.25 การทำงานของโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ | 28 |
| 2.26 ตัวอย่างการเน้นที่ส่วนที่มีความเกี่ยวข้องของคำเฉพาะกับคำอื่น ๆ ในประโยคนั้น | 29 |
| 2.27 กระบวนการทำงานของโครงข่ายประสาทเทียมโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์สวิน | 30 |
| 2.28 การนำข้อมูลคุณลักษณะของภาพมาทำการฉายไปยังมิติ C | 30 |
| 2.29 กระบวนการทำงานของการแทนตัวเข้ารหัสแบบสองทิศทางจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ | 32 |
| 2.30 กระบวนการทำงานของการเรียนรู้แบบฝึกรอบม้วนหน้า BERT ที่ปรับให้เหมาะสมอย่างมีประสิทธิภาพจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ | 33 |
| 2.31 การทำงานของการเรียนรู้เกี่ยวกับลำดับตัวเข้ารหัส-ตัวถอดรหัสเป็นแบบลำดับ | 34 |
| 2.32 กระบวนการทำงานของการเรียนรู้แบบเสริมกำลัง | 35 |
| 2.33 กระบวนการทำงานของการฝึกรอบม้วนลำดับความสำคัญด้วยตนเอง | 37 |
| 2.34 กระบวนการของการเพิ่มประสิทธิภาพนโยบายไขว่คว้า | 39 |
| 2.35 รายละเอียดของการประเมินคุณภาพของ BLEU | 40 |
| 2.36 รายละเอียดของการประเมินคุณภาพของ CIDER | 41 |
| 3.1 ผังงานโดยรวมของระบบ | 44 |
| 3.2 ตัวอย่างหน้าโปรแกรมในการอัปโหลดภาพ | 45 |
| 3.3 ตัวอย่างการย่อขนาดของภาพให้เหลือ 384x384 | 46 |
| 3.4 ผังงานของการใช้เทคนิค Reinforcement ในการใส่คำบรรยายภาพอัตโนมัติภาษาไทย | 47 |
| 3.5 กระบวนเข้ารหัสด้วยการเรียนรู้โครงข่ายประสาทเทียมเชิงลึก | 48 |
| 3.6 กระบวนการถอดรหัสด้วยการเรียนรู้แบบหน่วยความจำระยะสั้นระยะยาว | 49 |

สารบัญรูป (ต่อ)

| รูปที่ | หน้า |
|--|------|
| 3.7 กระบวนเข้ารหัสด้วยการเรียนรู้โครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์สวิน | 50 |
| 3.8 กระบวนการเรียนรู้แบบฝึกอบรมล่วงหน้า BERT ที่ปรับให้เหมาะสมอย่างมีประสิทธิภาพจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ | 51 |
| 3.9 กระบวนการเรียนรู้การฝึกลำดับความสำคัญด้วยตนเอง | 52 |
| 3.10 แสดงผลลัพธ์คำบรรยายภาพอัตโนมัติภาษาไทยบนเว็บไซต์ | 53 |
| 4.1 ผู้ใช้งานอัปโหลดภาพไม่ตรงตามเงื่อนไข | 56 |
| 4.2 ผู้ใช้งานอัปโหลดไฟล์ภาพตรงตามเงื่อนไข | 57 |
| 4.3 ผลการทดลองการอัปโหลดไฟล์ภาพที่มีนามสกุล JPEG JPG และ PNG | 57 |
| 4.4 แสดงการใส่คำบรรยายภาพอัตโนมัติภาษาไทย | 58 |
| 4.5 ผลทดลองการใส่คำบรรยายภาพอัตโนมัติภาษาไทย | 59 |
| 4.6 แสดงการใส่คำบรรยายภาพอัตโนมัติภาษาไทยครั้งละ 1 ภาพ | 60 |
| 4.7 ผลการทดลองการใส่คำบรรยายภาพอัตโนมัติภาษาไทยครั้งละ 1 ภาพ | 61 |
| ก.1 แผนที่ตั้งสถานประกอบการ ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ NECTEC | 70 |
| ก.2 ตราสัญลักษณ์ของสถานประกอบการศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ NECTEC | 71 |

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

เนื่องจากในปัจจุบันองค์ความรู้ทางด้านปัญญาประดิษฐ์หรือเอไอ (Artificial Intelligence : AI) มีการพัฒนาที่รวดเร็ว ซึ่งมีงานวิจัยทางด้านปัญญาประดิษฐ์อยู่ตัวหนึ่งที่ค่อนข้างน่าสนใจ คือ ปัญญาประดิษฐ์ใส่คำบรรยายภาพอัตโนมัติ (Image Captioning) ได้รับแรงบันดาลใจจากการเข้ารหัส (Encoder)/ถอดรหัส (Decoder) ซึ่งใช้โครงข่ายประสาทเทียมเชิงลึก (Convolution Neural Network : CNN) ในการเข้ารหัสภาพ และใช้การเรียนรู้แบบหน่วยความจำระยะสั้นระยะยาว (Long Short Term Memory : LSTM) ในการถอดรหัสเพื่อใส่คำบรรยายภาพอัตโนมัติ ปัญหาที่พบก็คือ ปัญญาประดิษฐ์ใส่คำบรรยายภาพอัตโนมัตินั้นยังเป็นการใส่คำบรรยายภาษาอังกฤษอยู่เนื่องจากชุดข้อมูลที่ใช้ในการฝึกอบรม (Training) นั้นเป็นภาษาอังกฤษจึงทำให้เราต้องนำคำบรรยายที่เป็นภาษาอังกฤษนั้นนำไปแปลภาษาให้เป็นภาษาไทยซึ่งมีโอกาสทำให้เกิดความไม่ถูกต้องของการแปลคำบรรยายภาพอัตโนมัติ ดังนั้นคำบรรยายจำเป็นต้องแปลเป็นภาษาไทย ซึ่งอาจนำไปสู่ความไม่ถูกต้องในการแปลคำบรรยายโดยอัตโนมัติ

Steven J. Rennie, Etienne Marcheret และ Vaibhava Goel เป็นนักวิจัยอยู่ที่ IBM ได้ศึกษาและวิจัยเกี่ยวกับปัญญาประดิษฐ์ใส่คำบรรยายภาพอัตโนมัติภาษาอังกฤษซึ่งนำเสนอแนวทางใหม่ในการฝึกอบรมปัญญาประดิษฐ์ ซึ่งเราเรียกว่า การฝึกลำดับความสำคัญด้วยตัวเอง (Self-Critical Sequence Training : SCST) [1] โดยสามารถปรับปรุงประสิทธิภาพของปัญญาประดิษฐ์ใส่คำบรรยายภาพอัตโนมัติอังกฤษได้ดีเป็นอย่างมากซึ่ง การฝึกลำดับความสำคัญด้วยตัวเองเป็นอัลกอริธึมการเรียนรู้แบบเสริมกำลัง (Reinforcement Learning : RL) โดยจะใช้ผลลัพธ์ของตนเองเพื่อปรับรางวัลที่ได้รับให้เป็นมาตรฐานเพื่อใช้ในการตัดสินใจในใส่คำบรรยายภาพอัตโนมัติภาษาอังกฤษ

ดังนั้นทางผู้จัดทำจึงมีแนวคิดในการศึกษาการใช้งานเทคนิค Reinforcement ในการใส่คำบรรยายภาพอัตโนมัติภาษาไทย โดยใช้เข้ารหัสในการสกัดคุณลักษณะของภาพ/ถอดรหัสในการใส่คำบรรยายภาพที่เป็นภาษาไทย ซึ่งจะใช้วิธีการฝึกลำดับความสำคัญด้วยตัวเอง โดยใช้การเรียนรู้โครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์สวิน (Shifted Window Transformer or Swin Transformer) ในการเข้ารหัสภาพและใช้วิธีการเรียนรู้แบบฝึกอบรมล่วงหน้า BERT ที่ปรับให้เหมาะสมอย่างมีประสิทธิภาพจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ (Robustly Optimized BERT Pre-training; RoBERTa) ในการถอดรหัสออกเป็นคำบรรยาย ซึ่งเป็นกระบวนการประมวลผลภาษาธรรมชาติ (Natural Language Processing : NLP) โดยการเรียนรู้ลักษณะของคำเพื่อให้คำบรรยายภาพอัตโนมัติเป็นภาษาไทยมีความแม่นยำและสะดวกในการใช้งานมากยิ่งขึ้น

1.2 วัตถุประสงค์ของปริญญานิพนธ์

1.2.1 เพื่อศึกษาการใช้งานเทคนิคการเรียนรู้แบบเสริมกำลังในการใส่คำบรรยายภาพอัตโนมัติภาษาไทย

1.2.2 เพื่อช่วยให้ผู้ใช้งานปัญญาประดิษฐ์ใส่คำบรรยายภาพอัตโนมัติ สามารถใส่คำบรรยายภาพที่เป็นภาษาไทยได้

1.2.3 เพื่อพัฒนาปัญญาประดิษฐ์ใส่คำบรรยายภาพอัตโนมัติภาษาไทย

1.2.4 เพื่อสร้างโมเดลการเรียนรู้แบบเสริมกำลังในการใส่คำบรรยายภาพอัตโนมัติภาษาไทย

1.3 ขอบเขตของปริญญานิพนธ์

1.3.1 สามารถใส่คำบรรยายภาพอัตโนมัติภาษาไทยได้

1.3.2 สามารถอัปโหลดไฟล์ภาพที่มีนามสกุล JPEG JPG และ PNG

1.3.3 การใส่คำบรรยายภาพอัตโนมัติภาษาไทยครั้งละ 1 ภาพ

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 สามารถอำนวยความสะดวกในการใส่คำบรรยายภาพอัตโนมัติภาษาไทย

1.4.2 นำทฤษฎีการเรียนรู้โครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์สวิน การเรียนรู้แบบฝึกรอบมัลทิววิว BERT ที่ปรับให้เหมาะสมอย่างมีประสิทธิภาพจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ และการฝึกลำดับความสำคัญด้วยตัวเองมาประยุกต์ใช้ในการใส่คำบรรยายภาพอัตโนมัติภาษาไทย

1.5 ขั้นตอนการดำเนินงานปริญญานิพนธ์

1.5.1 ศึกษาข้อมูลการเขียนโปรแกรมภาษาไพธอนพื้นฐาน

1.5.2 ศึกษาและรวบรวมข้อมูลภาพที่จะมาทำการเรียนรู้และคำบรรยายภาษาไทย

1.5.3 ศึกษาทฤษฎีการเรียนรู้เชิงลึก

1.5.4 ศึกษาทฤษฎีการเรียนรู้โครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์สวิน

1.5.5 ศึกษาทฤษฎีการเรียนรู้แบบฝึกรอบมัลทิววิว BERT ที่ปรับให้เหมาะสมอย่างมีประสิทธิภาพจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์

1.5.6 ศึกษาทฤษฎีการใช้งานเทคนิคการเรียนรู้แบบเสริมกำลัง

1.5.7 ศึกษาทฤษฎีวิธีการฝึกรอบมัลทิววิวแบบการฝึกลำดับความสำคัญด้วยตัวเอง

1.5.8 เขียนโมเดลการเรียนรู้โครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์สวิน การเรียนรู้การแทนตัวเข้ารหัสแบบสองทิศทางจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ และการฝึกลำดับความสำคัญด้วยตัวเองมาประยุกต์ใช้ในการใส่คำบรรยายภาพอัตโนมัติภาษาไทย

1.5.9 ทดสอบความถูกต้องการเรียนรู้ของโมเดลการใส่คำบรรยายภาพอัตโนมัติภาษาไทย

1.5.10 สรุปผลการทดลอง

1.5.11 จัดทำเล่มปริญญานิพนธ์

1.6 แผนการดำเนินงานปริญญานิพนธ์

ตารางที่ 1.1 ขั้นตอนการดำเนินงานวิชาสหกิจศึกษา

[illegible]

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงหลักการ และทฤษฎีพื้นฐานของการใช้งานเทคนิค Reinforcement ในการใส่คำบรรยายภาพอัตโนมัติภาษาไทย โดยใช้การเข้ารหัสถอดรหัส เพื่อช่วยในการใส่คำบรรยายภาพที่เป็นภาษาไทย โดยภายในเนื้อหาจะประกอบไปด้วยทฤษฎีและความรู้เบื้องต้นเกี่ยวกับภาษาไพธอน ชุดข้อมูล MSCOCO การประมวลผลภาษาธรรมชาติ ระบบสีอาร์จีบี โครงข่ายประสาทเทียม การเรียนรู้เชิงลึก โครงข่ายประสาทเทียมแบบคอนโวลูชัน การเรียนรู้แบบหน่วยความจำระยะสั้นระยะยาว โครงข่ายประสาทเทียมของโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ การเรียนรู้โครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์สวิน การเรียนรู้แบบฝีกอบรมล่วงหน้า BERT ที่ปรับให้เหมาะสมอย่างมีประสิทธิภาพจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ การเรียนรู้เกี่ยวกับลำดับตัวเข้ารหัส-ตัวถอดรหัสเป็นรูปแบบลำดับ การเรียนรู้แบบเสริมกำลัง การฝึกลำดับความสำคัญด้วยตัวเอง การเรียนรู้แบบเสริมกำลัง การเพิ่มประสิทธิภาพนโยบายไกล์เคียง และเกณฑ์ที่ใช้ในการวัดประสิทธิภาพของโมเดลทางด้านคอมพิวเตอร์วิชันสามารถเข้าใจภาษาของมนุษย์ ดังรายละเอียดต่อไปนี้

- 2.1 ภาษาไพธอน (Python Programming Language)
- 2.2 เฟลค (Flask)
- 2.3 ชุดข้อมูล MSCOCO (Microsoft Common Objects in Context Dataset)
- 2.4 การประมวลผลภาษาธรรมชาติ (Natural Language Processing : NLP)
- 2.5 ระบบสีอาร์จีบี (Red Green Blue : RGB)
- 2.6 โครงข่ายประสาทเทียม (Neural Network : NN)
- 2.7 การเรียนรู้เชิงลึก (Deep Learning : DL)
- 2.8 โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolution Neural Network : CNN)
- 2.9 การเรียนรู้แบบหน่วยความจำระยะสั้นระยะยาว (Long Short Term Memory : LSTM)
- 2.10 โครงข่ายประสาทเทียมของโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ (Transformer Neural Network)
- 2.11 การเรียนรู้โครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์สวิน (Shifted Window Transformer or Swin Transformer)
- 2.12 การแทนตัวเข้ารหัสแบบสองทิศทางจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ (Bidirectional Encoder Representations from Transformers : BERT)

- 2.13 การเรียนรู้แบบฝึกลำดับวลี BERT ที่ปรับให้เหมาะสมอย่างมีประสิทธิภาพจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ (Robustly Optimized BERT Pre-training: RoBERTa)
- 2.14 การเรียนรู้เกี่ยวกับลำดับตัวเข้ารหัส-ตัวถอดรหัสเป็นรูปแบบลำดับ (Encoder-Decoder Sequence to Sequence Model)
- 2.15 การเรียนรู้แบบเสริมกำลัง (Reinforcement Learning : RL)
- 2.16 การฝึกลำดับความสำคัญด้วยตัวเอง (Self-Critical Sequence Training : SCST)
- 2.17 การเพิ่มประสิทธิภาพนโยบายใกล้เคียง (Proximal Policy Optimization : PPO)
- 2.18 เกณฑ์ที่ใช้ในการวัดประสิทธิภาพของโมเดลของการใส่คำบรรยายภาพ (Model Evaluation of Caption Generation)

2.1 ภาษาไพธอน

คอมพิวเตอร์ที่ถูกพัฒนาขึ้นโดยมีจุดประสงค์หลักเพื่อไว้เป็นอุปกรณ์ในการคำนวณตัวเลขและสมการต่าง ๆ มีแนวการคำนวณอย่างเป็นระบบ คำนวณซ้ำ ๆ ไปเรื่อย ๆ ในยุคแรก ๆ แนวคิดเกี่ยวกับการคำนวณอย่างเป็นระบบนั้นมีมาตั้งแต่โบราณก่อนที่จะมีคอมพิวเตอร์ [4] เช่น วิธีของนิวตัน วิธีของออยเลอร์ ซึ่งเอาไว้คำนวณแบบวนซ้ำ ๆ เพื่อหาคำตอบของสมการหรือค่าที่ต้องการ ต่อมาจึงได้มีแนวคิดที่จะใช้เครื่องจักรเพื่อช่วยในการคำนวณ เรียกว่าคำนวณเชิงกล ถูกสร้างขึ้นโดยเบลส ปาสกาล (Blaise Pascal) ชื่อว่า ปาสกาลีน (Pascaline) ถัดมาก็ได้มีการนำวงจรอิเล็กทรอนิกส์มาใช้สร้างเป็นเครื่องคำนวณ ในช่วงสงครามโลกครั้งที่สอง และนั่นเป็นจุดเริ่มต้นของคอมพิวเตอร์ในปัจจุบัน ในยุคแรกของคอมพิวเตอร์นั้น จะใช้หลอดสุญญากาศ เป็นส่วนประกอบหลัก ซึ่งทำให้มีขนาดใหญ่ แล้วย่อย ๆ พัฒนาเป็นสารกึ่งตัวนำทำให้มีขนาดเล็กลง และยิ่งพัฒนาต่อมายิ่งเล็กลงเรื่อย ๆ กลายเป็นคอมพิวเตอร์ในปัจจุบันที่ใช้อยู่

เราได้ทราบกันแล้วว่าคอมพิวเตอร์นั้นถูกสร้างขึ้นเพื่อการคำนวณ โดยคอมพิวเตอร์นั้นเราต้องป้อนคำสั่งเพื่อให้มันทำงาน ตามชุดคำสั่งจำนวนมาก ๆ ที่ถูกกำหนดขึ้น เพื่อให้คอมพิวเตอร์ทำงานเป็นระบบ นั้นเรียกว่าโปรแกรมคอมพิวเตอร์ คำสั่งจำนวนมาก ๆ นั้นจะถูกกำหนดติดตั้งคำสั่งการทำงานตามที่ต้องการจึงเรียกว่าการเขียนโปรแกรม (Programming) คำสั่งที่นั้นที่ถูกติดตั้ง มีการทำงานของคอมพิวเตอร์ที่ซับซ้อน และมีตรรกะการทำงานที่ต่างจากมนุษย์ ภาษาที่ใช้สั่งการนั้นเรียกว่า ภาษาเครื่อง ซึ่งมนุษย์จะต้องทำความเข้าใจ ถัดมาเพื่อให้่ายต่อการใช้งานจึงได้มีการสร้างภาษาที่ใกล้เคียงกับมนุษย์ใช้ เรียกว่าภาษาแอสเซมบลี (Assembly) แต่ก็ยังยากต่อการใช้งานอยู่ จึงได้มีการพัฒนามาเป็นภาษาระดับสูง ภาษาระดับสูงนั้นเวลาทำงานต้องแปลงเป็นภาษาเครื่องอีกที เพื่อให้คอมพิวเตอร์เข้าใจ จึงทำให้มีการทำงานช้าลง แต่ก็สะดวกต่อการเขียนโปรแกรมมากขึ้น ภาษา

ระดับสูงในยุคแรก ๆ ได้แก่ ฟอรัทแรน (Fortran) ปาสกาล (Pascal) และ ภาษาซี (C) จากนั้นก็ได้มีการพัฒนาภาษาใหม่ ๆ ขึ้นมาเรื่อย ๆ จนถึงปัจจุบัน ในจำนวนนั้นก็มีภาษาที่เป็นยอดนิยมคือภาษาไพธอน ซึ่งก็ถูกพัฒนาจากภาษาระดับสูงในยุคแรก ๆ จึงเขียนได้ง่ายและทำงานได้ช้าเมื่อเทียบกับภาษาในยุคแรกๆ แต่ใช้งานได้ง่าย เหมาะกับคนทั่วไปที่ไม่ได้เชี่ยวชาญด้านคอมพิวเตอร์

จุดเด่นของภาษาไพธอนคือ มีโครงสร้างของภาษาไม่ซับซ้อน สนับสนุนการเขียนโปรแกรมเชิงวัตถุ มีชุมชนที่มีนักพัฒนาภาษาไพธอนอย่างมากมาย และมีไลบรารีให้ดาวน์โหลดเลือกใช้งานโดยไม่เสียค่าใช้จ่าย ไม่ว่าจะเป็นไลบรารีประมวลผลข้อมูลภาพ ไลบรารีสำหรับการวิเคราะห์ข้อมูล เช่น OpenCV, Tensorflow, Numpy และอื่น ๆ อีกมากมาย โดยมีตัวอย่างการใช้งานโค้ดภาษาไพธอนแสดง

ดังโปรแกรมที่ 2.1 และแสดงผลลัพธ์จากการทำงานดังรูปที่ 2.1

โปรแกรมที่ 2.1 ตัวอย่างการใช้งานโค้ดภาษาไพธอน

```
[1]: hi = "Hello World I'm AI"
[2]: print("( %s )" % hi, " length of string : %d" % (len(hi)), "character")
```

จากโปรแกรมที่ 2.1 สามารถอธิบายการทำงานได้ดังนี้

บรรทัดที่ 1 ประกาศสตริงชื่อ hi โดยเก็บค่า Hello World I'm AI

บรรทัดที่ 2 ใช้คำสั่งในการแสดงผลค่าของตัวแปร hi ฟังก์ชัน Length ที่สร้างขึ้นในไพธอนในการนับจำนวนอักขระในสตริง

```
( Hello World I'm AI ) length of string : 18 character
```

รูปที่ 2.1 ผลการรันจากการเขียนโค้ดในโปรแกรมที่ 2.1

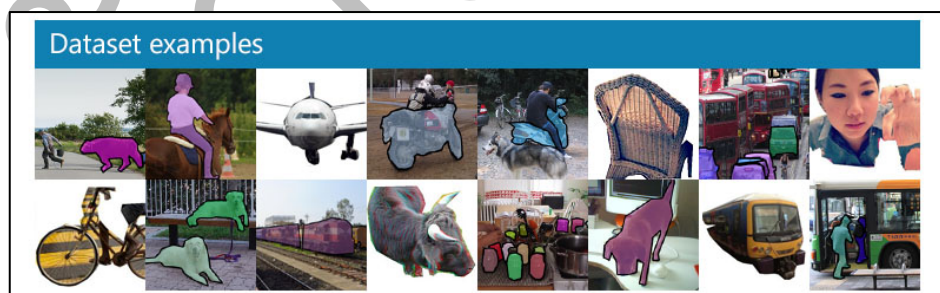
จากรูปที่ 2.1 สามารถอธิบายผลการรันจากโปรแกรมที่ 2.1 เป็นการแสดงผล (Hello World I'm AI) Length of string : 18 character ออกทางหน้าจอ

2.2 เฟลค

เฟลค [13] เป็นเว็บเฟรมเวิร์กขนาดเล็กที่เขียนด้วยไพธอนจัดเป็นไมโครเฟรมเวิร์กเนื่องจากไม่ต้องการเครื่องมือหรือไลบรารีเฉพาะ ไม่มีเลเยอร์นามธรรมของฐานข้อมูล การตรวจสอบความถูกต้องของแบบฟอร์ม หรือส่วนประกอบอื่นใดที่ไลบรารีของบุคคลที่สามที่มีอยู่แล้วมีฟังก์ชันทั่วไป อย่างไรก็ตามเฟลครองรับส่วนขยายที่สามารถเพิ่มคุณลักษณะของแอปพลิเคชันได้ เหมือนกับว่ามีการใช้งานใน เฟลคเอง ส่วนขยายเฟลคที่ได้รับความนิยมบางตัว ได้แก่ Flask-RESTful, Flask-SQLAlchemy และ Flask-Login ความเรียบง่าย ความยืดหยุ่น และความเรียบง่ายของเฟลคทำให้เป็นตัวเลือกยอดนิยมสำหรับเว็บแอปพลิเคชันและ API ขนาดเล็กถึงขนาดกลาง

2.3 ชุดข้อมูล MSCOCO

ชุดข้อมูล MSCOCO [5] เป็นชุดข้อมูลขนาดใหญ่สำหรับการตรวจจับวัตถุ การแบ่งส่วน และคำอธิบายภาพ เปิดตัวในปี 2014 และกลายเป็นชุดข้อมูลที่ใช้กันอย่างแพร่หลายที่สุดชุดหนึ่งสำหรับงานตรวจจับวัตถุและการจดจำภาพ ประกอบด้วยภาพมากกว่า 300,000 ภาพ แต่ละภาพมีคำอธิบายประกอบด้วยกล่องขอบและป้ายกำกับขอบเขตหลายรายการ ชุดข้อมูลยังรวมถึงคำบรรยายภาพ ซึ่งให้คำอธิบายที่เป็นข้อความของวัตถุและฉากในภาพ แบ่งออกเป็นชุดฝึกอบรม ชุดตรวจสอบ และชุดทดสอบ ถูกนำมาใช้อย่างแพร่หลายในการประเมินประสิทธิภาพของการตรวจจับวัตถุและอัลกอริทึมการจดจำภาพ นอกจากนี้ยังใช้เพื่อฝึกโมเดลแมชชีนเลิร์นนิงสำหรับงานต่างๆ รวมถึงการตรวจจับวัตถุ คำบรรยายภาพ และการตอบคำถามด้วยภาพ แสดงตัวอย่างชุดข้อมูลรูปภาพดังรูปที่ 2.2



รูปที่ 2.2 ตัวอย่างชุดข้อมูลรูปภาพ

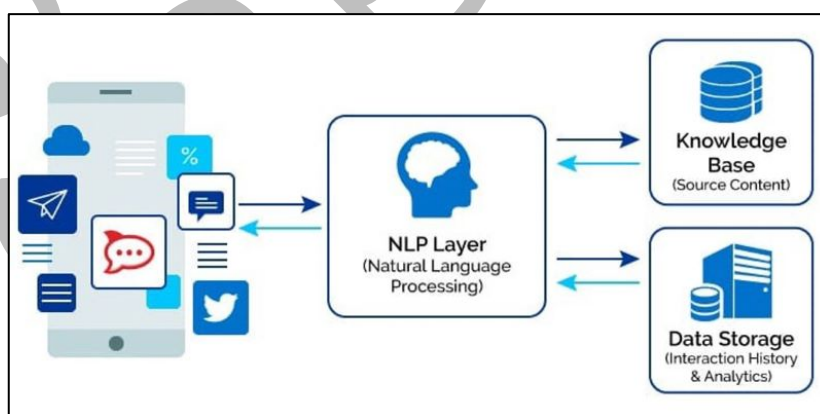
จากรูปที่ 2.2 สามารถอธิบายตัวอย่างชุดข้อมูลรูปภาพของ MSCOCO ที่ไว้ใช้สำหรับชุดฝึกอบรม ชุดตรวจสอบ และชุดทดสอบของโมเดลแมชชีนเลิร์นนิงสำหรับงานต่างๆ รวมถึงการตรวจจับวัตถุ คำบรรยายภาพ และการตอบคำถามด้วยภาพ

นอกจากชุดข้อมูล MSCOCO แล้ว ยังมีชุดข้อมูลเพิ่มเติมอีกหลายชุดที่ได้รับการเผยแพร่โดยเป็นส่วนหนึ่งของการรวบรวม MSCOCO ซึ่งรวมถึงชุดข้อมูล MSCOCO Keypoints ซึ่งรวมถึงคำอธิบายประกอบท่าทางของมนุษย์ และชุดข้อมูล MSCOCO Stuff ซึ่งรวมถึงคำอธิบายประกอบสำหรับคลาส "สิ่งของ" (เช่น หญ้า ท้องฟ้า น้ำ)

2.4 การประมวลผลภาษาธรรมชาติ

การประมวลผลภาษาธรรมชาติ [3] เป็นสาขาของปัญญาประดิษฐ์และภาษาศาสตร์เชิงคำนวณที่มุ่งเน้นไปที่การโต้ตอบระหว่างคอมพิวเตอร์และมนุษย์โดยใช้ภาษาธรรมชาติ มันเกี่ยวข้องกับการพัฒนาอัลกอริทึมและแบบจำลองที่สามารถวิเคราะห์ ทำความเข้าใจ และสร้างภาษามนุษย์

มีแอปพลิเคชันที่หลากหลาย รวมถึงการแปลภาษา การสรุปข้อความ การวิเคราะห์ความรู้สึก การสร้างแบบจำลองการสนทนา และการดึงข้อมูล เทคนิคทั่วไปบางอย่างในการประมวลผลภาษาธรรมชาติ รวมถึงการติดแท็กส่วนหนึ่งของคำพูด การรู้จำเอนทิตีที่มีชื่อ และการตัดคำ อาจเป็นสิ่งที่ท้าทายเนื่องจากภาษาธรรมชาติมีความซับซ้อนและคลุมเครือ และมีความแปรปรวนอย่างมากในวิธีที่ผู้คนใช้ภาษา ด้วยเหตุนี้ระบบการประมวลผลภาษาธรรมชาติมักจะอาศัยข้อมูลการฝึกอบรมที่มีคำอธิบายประกอบจำนวนมากและอัลกอริทึมการเรียนรู้ของเครื่องขั้นสูงเพื่อให้ทำงานได้ดี แสดงตัวอย่างการนำการประมวลผลภาษาธรรมชาติไปใช้งานดังรูปที่ 2.3



รูปที่ 2.3 ตัวอย่างการนำการประมวลผลภาษาธรรมชาติไปใช้งาน

จากรูปที่ 2.3 สามารถอธิบายตัวอย่างการนำการประมวลผลภาษาธรรมชาติไปใช้งานในชีวิตจริงซึ่งมีแอปพลิเคชันที่หลากหลายได้นำไปพัฒนาใช้งาน รวมถึงการแปลภาษา การสรุปข้อความ การวิเคราะห์ความรู้สึก การสร้างแบบจำลองการสนทนา และการดึงข้อมูลเป็นต้น

2.5 ระบบสีอาร์จีบี

ระบบสีอาร์จีบี [8] คือระบบสีของแสง เกิดจากการหักเหของแสงกลายเป็นสีรุ้ง (Spectrum) ด้วยกัน 7 สี ได้แก่ ม่วง คราม น้ำเงิน เขียว เหลือง ส้ม แดง ซึ่งเป็นช่วงแสงที่ตาของมนุษย์นั้นสามารถมองเห็นได้ โดยแสงสีม่วง (Ultraviolet) จะมีความถี่สูงสุด และแสงสีแดง (Infrared) จะมีความถี่ต่ำสุด คลื่นแสงที่มีความถี่สูงกว่า หรือต่ำกว่าในช่วง แสงสีม่วงและแสงสีแดงนั้น มนุษย์จะไม่สามารถมองเห็นได้

แสงสีทั้งหมดเกิดจากแสงสี 3 สี หรือ 3 แชนแนล (Channel) คือ สีแดง (Red) สีเขียว (Green) และสีน้ำเงิน (Blue) ทั้งสามสีถือเป็นแม่สีของแสง เมื่อแม่สีมาผสมกันจะได้สีดังนี้

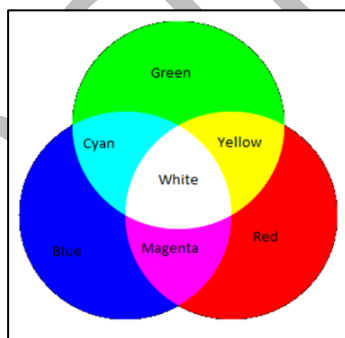
สีแดง + สีเขียว ได้ สีเหลือง (Yellow)

สีเขียว + น้ำเงิน ได้ สีฟ้า (Cyan)

สีแดง + สีน้ำเงิน ได้ สีแดงอมชมพู (Magenta)

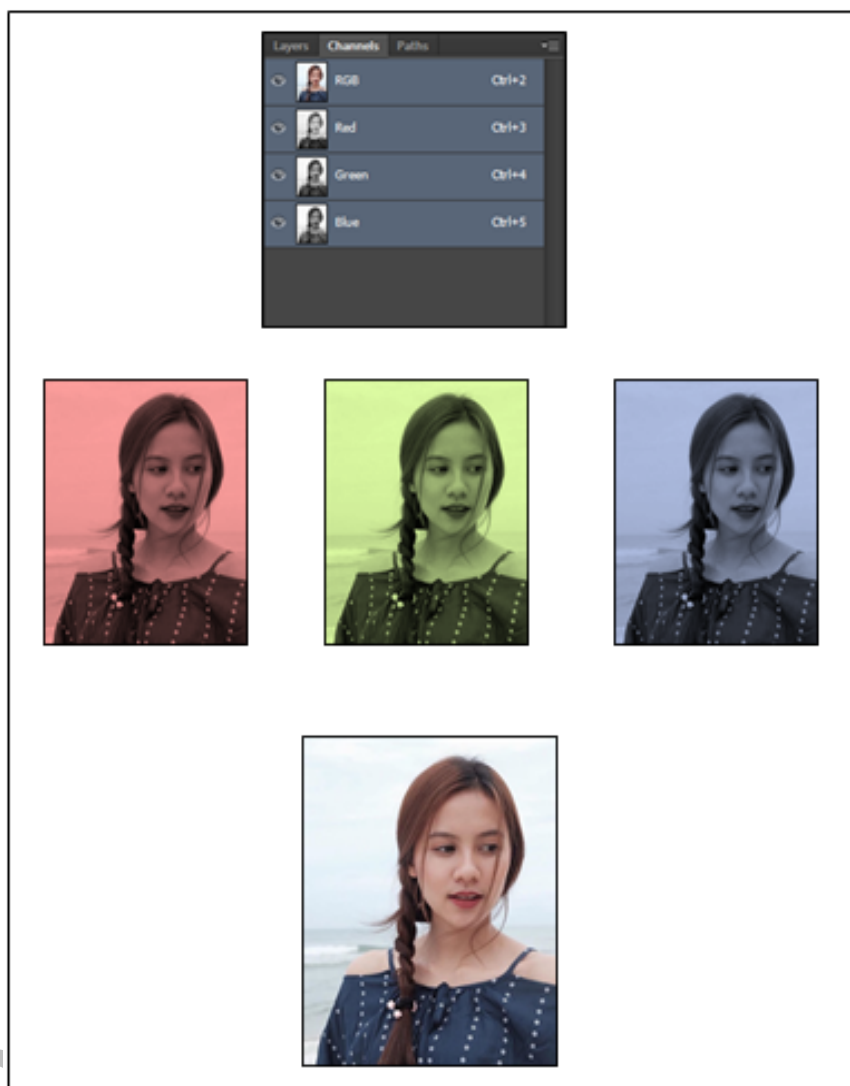
เมื่อแม่สีของแสงทั้ง 3 มาผสมกัน ด้วยปริมาณแสงสว่างที่เท่ากันจะได้เป็นแสงสีขาว แต่ถ้าผสมด้วยระดับความสว่างที่ต่างกันจะได้แสงสีจำนวนมากเกิดขึ้น แสดงดังรูปที่ 2.4 และ รูปที่ 2.5

ระบบสีอาร์จีบีนั้นจะถูกนำมาใช้กับการแสดงผลของอุปกรณ์อิเล็กทรอนิกส์ เช่น จอภาพของคอมพิวเตอร์ จอภาพของสมาร์ทโฟน รูปภาพจากกล้องดิจิทัล เป็นต้น



รูปที่ 2.4 แสดงโมเดลระบบสีอาร์จีบี

จากรูปที่ 2.4 สามารถอธิบายรูปการผสมแม่สีของแสงโมเดลระบบสีอาร์จีบีที่มีด้วยกันทั้งหมด 7 สี ซึ่งเกิดจากการหักเหของแสงกลายเป็นสีรุ้ง ได้แก่ ม่วง คราม น้ำเงิน เขียว เหลือง ส้ม แดง

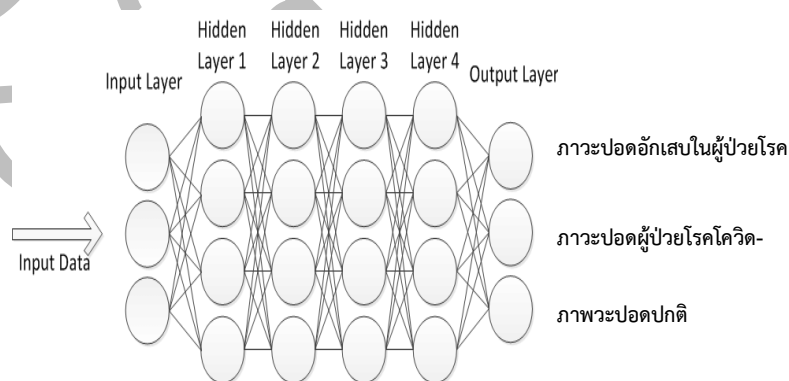


รูปที่ 2.5 ตัวอย่างภาพที่มีการผสมของแสงสีทั้ง 3 แชนแนล

จากรูปที่ 2.5 สามารถอธิบายรูปตัวอย่างที่มีการผสมของแสงทั้ง 3 แชนแนล สีแดง (Red) สีเขียว (Green) และสีน้ำเงิน (Blue) ทั้งสามสีถือเป็นแม่สีของแสง เมื่อทั้ง 3 สีผสมกัน ด้วยปริมาณแสงสว่างที่เท่ากันจะได้เป็นแสงสีขาว แต่ถ้าผสมด้วยระดับความสว่างที่ต่างกันจะได้แสงสีจำนวนมากเกิดขึ้น

2.6 โครงข่ายประสาทเทียม

โครงข่ายประสาทเทียม [19] เป็นระบบคอมพิวเตอร์จากโมเดลคณิตศาสตร์เพื่อจำลองการทำงานของโครงข่ายประสาทชีวภาพที่อยู่ในสมองของสิ่งมีชีวิต สามารถเรียนรู้ที่จะทำงานที่มอบหมายได้ จากการเรียนรู้ผ่านตัวอย่าง โดยไม่ถูกโปรแกรมด้วยกฎเกณฑ์ตายตัวแบบระบบอัตโนมัติ ซึ่งมีแนวคิดเริ่มต้นของเทคนิคนี้ได้มาจากการศึกษาโครงข่ายไฟฟ้าชีวภาพ (Bioelectric Network) ในสมอง ซึ่งประกอบไปด้วย เซลล์ประสาท (Neurons) และจุดประสานประสาท (Synapses) ข่ายงานประสาทเกิดจากการเชื่อมต่อระหว่างเซลล์ประสาทหนึ่งไปยังอีกเซลล์ประสาทหนึ่งต่อกันไปเรื่อย ๆ จนเกิดเป็นเครือข่ายที่ทำงานร่วมกัน การประมวลผลต่าง ๆ ของโครงข่ายประสาทเทียมเกิดขึ้นในหน่วยประมวลผลย่อย เรียกว่า โหนด (Node) ซึ่งโหนดเป็นการจำลองลักษณะการทำงานมาจากเซลล์การส่งสัญญาณ ระหว่างโหนดที่เชื่อมต่อกัน จำลองมาจากการเชื่อมต่อของใยประสาท และแกนประสาทในระบบประสาทของสมองมนุษย์ภายในโหนด จุดเชื่อมต่อแต่ละจุด มีความคล้ายคลึงกับจุดประสานประสาท (Synapses) ในสมอง มีความสามารถในการส่งสัญญาณไปยังเซลล์ประสาทเซลล์อื่น ๆ ที่เชื่อมต่อกับมันได้ ยกตัวอย่างเช่น การประมวลผลภาพ ซึ่งงานด้วยระบบโครงข่ายประสาทเทียมโดยจะเรียนรู้การจำแนกรูปภาพเอกซเรย์บริเวณทรวงอก ว่าอยู่ในภาวะปอดอักเสบในผู้ป่วยโรคโควิด-19 ภาวะปอดผู้ป่วยโรคโควิด-19 และภาวะปอดปกติ ได้มาจากการให้ตัวอย่างภาพที่กำกับโดยผู้เขียนโปรแกรมว่าภาพนี้อยู่ใน “เป็นภาวะปอดอักเสบในผู้ป่วยโรคโควิด-19 “ , “เป็นภาวะปอดผู้ป่วยโรคโควิด-19 “ หรือ “เป็นภาวะปอดปกติ” ซึ่งจะแสดงตัวอย่างโมเดลการจำแนกรูปภาพเอกซเรย์บริเวณทรวงอก ได้แสดงดังรูปที่ 2.6



รูปที่ 2.6 ตัวอย่างโมเดลการจำแนกรูปภาพเอกซเรย์บริเวณทรวงอก

จากรูปที่ 2.6 สามารถอธิบายตัวอย่างของโมเดลการจำแนกรูปภาพเอกซเรย์บริเวณทรวงอก ซึ่งจะอยู่ในกระบวนการโครงข่ายประสาทเทียม ในการทำการตัดสินใจเพื่อวิเคราะห์กลุ่มของรูปภาพถ่ายเอกซเรย์ทรวงอก

2.7 การเรียนรู้เชิงลึก

ปัญญาประดิษฐ์ (Artificial Intelligence : AI) เป็นการนำความฉลาดของมนุษย์ไปใส่ไว้ในคอมพิวเตอร์ในงานวิจัยปัญญาประดิษฐ์ ขึ้นแรกนั้น นักวิจัยพยายามที่จะทำให้คอมพิวเตอร์มีความฉลาดคล้ายคลึงกับมนุษย์ โดยให้คอมพิวเตอร์ลองทำงานที่มีความจำเพาะเจาะจงบางอย่าง เช่น การเล่นเกมส

การเรียนรู้ของเครื่อง (Machine Learning : ML) หมายถึงความสามารถของเครื่องที่สามารถเรียนรู้สิ่งต่าง ๆ ได้จากข้อมูลที่มีอยู่ โดยที่เราไม่ต้องเขียนโค้ดอธิบายกฎเกณฑ์ต่าง ๆ ของสิ่งนั้น การเรียนรู้ของเครื่องทำให้เครื่องคอมพิวเตอร์สามารถเรียนรู้สิ่งต่าง ๆ ได้ด้วยตนเอง โดยใช้ประโยชน์จากพลังในการประมวลผลของคอมพิวเตอร์ในยุคปัจจุบัน ทำให้สามารถประมวลผลข้อมูลจำนวนมากได้นั่นเอง

การเรียนรู้แบบมีผู้ฝึกสอน (Supervised Learning) เกี่ยวข้องกับการใช้ข้อมูลนำเข้าที่ระบุผลลัพธ์ที่คาดหวังไว้ชัดเจนตัวอย่างของการเรียนรู้แบบมีผู้ฝึกสอน คือการพยากรณ์อากาศ ปัญญาประดิษฐ์จะเรียนรู้ข้อมูลจากสถิติ เช่น ความดัน, ความชื้น, แรงแลม โดยผลลัพธ์คืออุณหภูมิในแต่ละช่วงเวลา

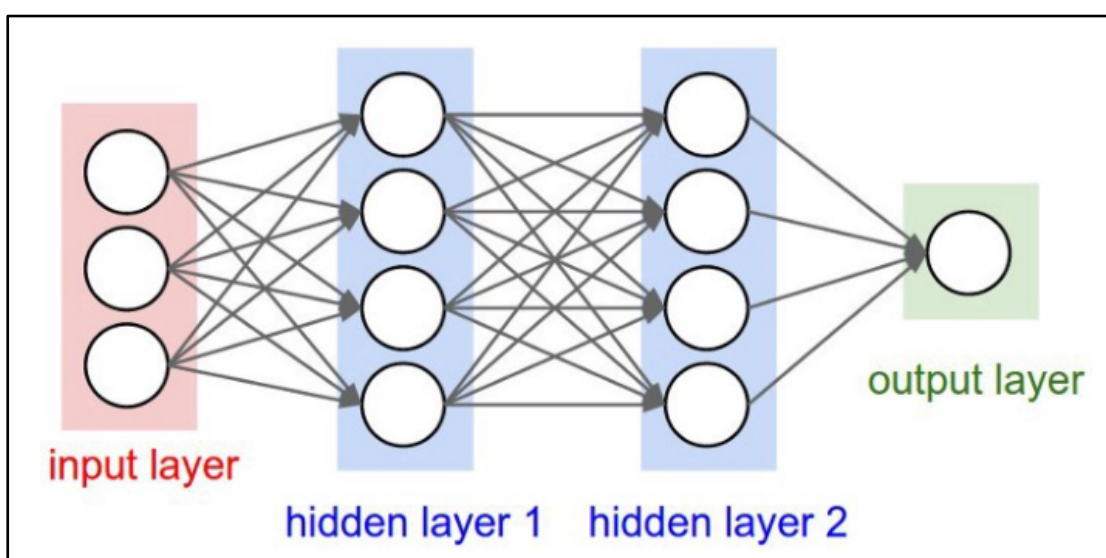
การเรียนรู้แบบไม่มีผู้ฝึกสอน (Unsupervised Learning) เป็นการเทรนโดยไม่มีการระบุผลลัพธ์ของข้อมูลนำเข้า เมื่อคุณเทรนปัญญาประดิษฐ์โดยใช้การเรียนรู้แบบไม่มีผู้ฝึกสอน ปัญญาประดิษฐ์จะตัดสินใจแบ่งกลุ่มข้อมูลด้วยตัวมันเอง ตัวอย่างหนึ่งของการเรียนรู้แบบไม่มีผู้ฝึกสอน คือการทำนายพฤติกรรมของผู้ใช้งานสำหรับเว็บ e-commerce การเรียนรู้ในลักษณะนี้ คุณไม่ต้องระบุผลลัพธ์ของข้อมูลนำเข้าปัญญาประดิษฐ์จะแบ่งกลุ่มข้อมูลโดยอัตโนมัติ และจะระบุว่าผู้ใช้งานแต่ละคนมีแนวโน้มที่จะซื้อสินค้าประเภทใดบ้าง

การเรียนรู้เชิงลึก [10] เป็นกระบวนการหนึ่งของการเรียนรู้ของเครื่องที่ทำให้เราสามารถเทรนปัญญาประดิษฐ์และทำนายผลลัพธ์จากข้อมูลนำเข้าได้ โดยใช้การเรียนรู้แบบมีผู้ฝึกสอน หรือ การเรียนรู้แบบไม่มีผู้ฝึกสอน เราจะเรียนรู้การทำงานของเครื่องเรียนรู้เชิงลึกผ่าน “Airplane Ticket Price Estimation” ซึ่งเป็นโมเดลปัญญาประดิษฐ์ที่ใช้ในการทำนายราคาตั๋วเครื่องบิน โดยเราจะเทรนปัญญาประดิษฐ์โดยใช้การเรียนรู้แบบมีผู้ฝึกสอนเราต้องการทำนายราคาตั๋วเครื่องบินจากข้อมูลเหล่านี้

- สนามบินต้นทาง (Origin Airport)
- สนามบินปลายทาง (Destination Airport)
- วันที่เดินทาง (Departure Date)
- สายการบิน (Airline)

โครงข่ายประสาทเทียมคล้ายกับสมองของมนุษย์ ในโครงข่ายประสาทเทียมจะมีเซลล์ประสาท

ถูกแสดงด้วยสัญลักษณ์วงกลมในแผนภาพ โดยเซลล์ประสาทเหล่านี้เชื่อมต่อถึงกันดังรูปที่ 2.7



รูปที่ 2.7 เซลล์ประสาทซึ่งมันถูกแสดงด้วยสัญลักษณ์วงกลม

จากรูปที่ 2.7 สามารถอธิบายเซลล์ประสาทที่มีลักษณะวงกลมซึ่งเรียกว่าเซลล์ประสาท ซึ่งจะเชื่อมต่อถึงกันเส้นที่เชื่อมต่อกันเรียกว่าจุดประสานประสาท เพื่อทำการส่งค่าไปให้ในแต่ละเซลล์ประสาทที่เชื่อมต่อกัน

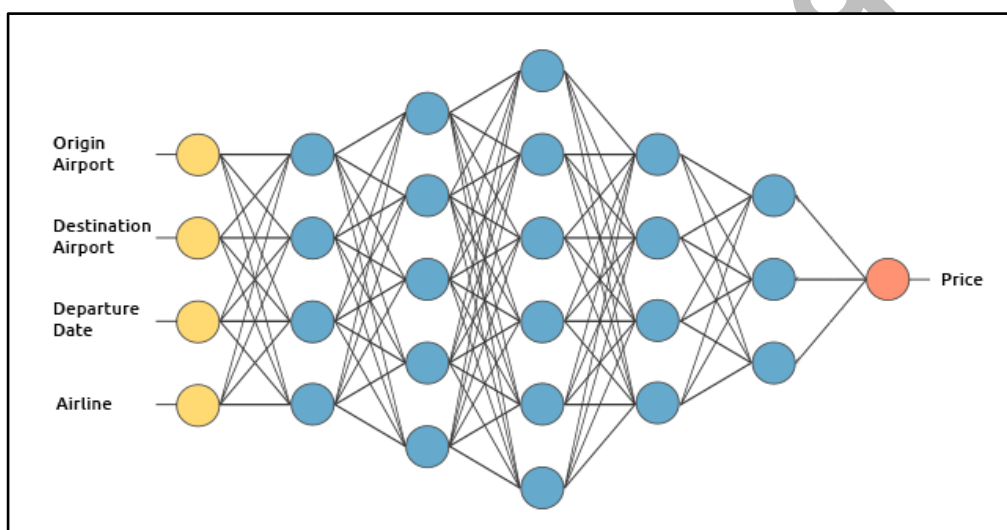
เซลล์ประสาท ถูกแบ่งเป็น 3 กลุ่มได้แก่

- ชั้นอินพุต (Input Layer)
- ชั้นซ่อนเร้น (Hidden Layer) - สามารถมีมากกว่า 1 ชั้น ได้
- ชั้นเอาต์พุต (Output Layer)

ชั้นอินพุทจะเป็นตัวรับข้อมูลนำเข้า ในกรณีของเรามีเซลล์ประสาท อยู่ 4 ตัวในชั้นอินพุท ได้แก่ สนามบินต้นทาง, สนามบินปลายทาง, วันที่เดินทาง และสายการบินโดยแต่ละเซลล์ประสาทในชั้นอินพุทจะทำการส่งข้อมูลไปให้เซลล์ประสาทในชั้นซ่อนเร้นชั้นแรก

ชั้นซ่อนเร้นจะทำการคำนวณข้อมูลที่รับมาจากชั้นก่อนหน้า โดยวิธีการทางคณิตศาสตร์ และส่งไปยังชั้นถัดไป ความท้าทายของการสร้างโครงข่ายประสาทเทียม คือการตัดสินใจว่าเราควรมีชั้นซ่อนเร้นกี่ชั้น และมีเซลล์ประสาทในแต่ละชั้นจำนวนเท่าใด คำว่าการเรียนรู้เชิงลึก หมายถึงการที่เรามีมากกว่า 1 ชั้นซ่อนเร้นนั่นเอง

ชั้นเอาต์พุทจะให้ผลลัพธ์ของการทำนายออกมา ในกรณีนี้คือราคาตั๋วเครื่องบิน ดังรูปที่ 2.8

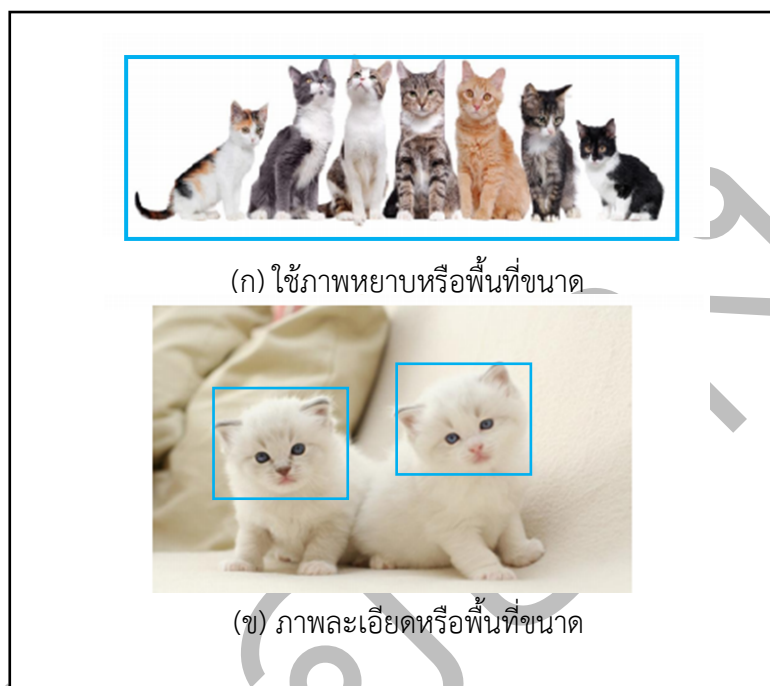


รูปที่ 2.8 เซลล์ประสาทการทำนายราคาตั๋วเครื่องบิน

จากรูปที่ 2.8 สามารถอธิบายการทำนายราคาของตั๋วเครื่องบินโดยใช้กระบวนการโครงข่ายประสาทเทียม เพื่อให้สามารถทำนายราคาของตั๋วเครื่องบินได้

2.8 โครงข่ายประสาทเทียมแบบคอนโวลูชัน

โครงข่ายประสาทเทียมแบบคอนโวลูชัน [9] เป็นโครงข่ายประสาทเทียมแบบหนึ่งที่มีการจำลองการมองเห็นของมนุษย์ ซึ่งการมองเห็นของมนุษย์นั้นจะมองพื้นที่เป็นบริเวณขนาดเล็กก่อน และนำกลุ่มของพื้นที่บริเวณขนาดเล็กมาผสมกันเป็นพื้นที่บริเวณขนาดใหญ่เพื่อดูว่าสิ่งที่เห็นอยู่คืออะไร



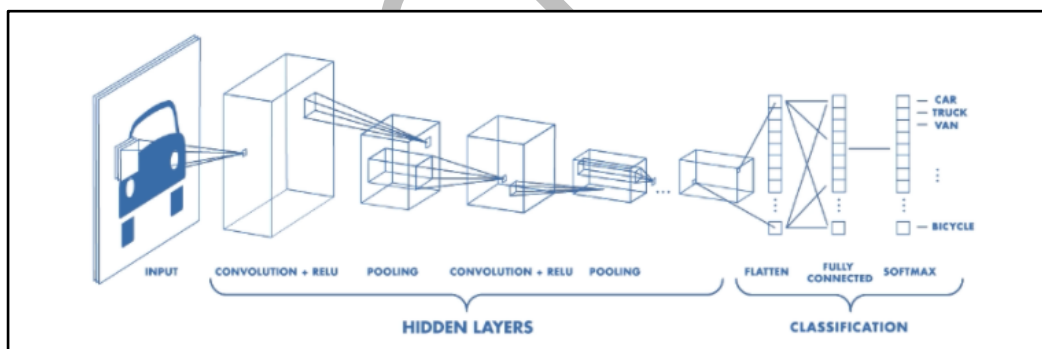
รูปที่ 2.9 ตัวอย่างการเลือกพื้นที่ในการจำแนก

เมื่อลองสังเกตจากการจำแนกแมวในรูปที่ 2.9 จะเห็นว่าภาพบนแม้จะมีแมวจำนวนหลายตัว แต่ก็สามารถแยกแมวออกจากกันได้ง่าย ๆ ด้วยการดูสี ดูลวดลาย บนตัวแมว และดูพื้นที่บริเวณขนาดใหญ่ ในขณะที่ภาพด้านล่างนั้นมีจำนวนแมวน้อยกว่าแต่กลับต้องดูพื้นที่เล็ก ๆ บริเวณตาและจมูก เนื่องจากไม่สามารถแยกแยะด้วยบริเวณพื้นที่ขนาดใหญ่เหมือนภาพด้านบนได้ เพราะสีของแมว หรือ ลวดลายมีความคล้ายคลึงกันมาก รวมถึงพื้นหลังของภาพที่กลมกลืนกับตัวแมวด้วย จะเห็นว่าภาพด้านบนนั้นถึงแม้จะดูเพียงสีและลวดลายแต่ก็ไม่ทำให้ทราบว่ามันเป็นแมว จึงต้องดูบริเวณ หาง หู และพื้นที่บริเวณขนาดใหญ่ประกอบไปด้วย ภาพด้านล่างก็ต้องดูพื้นที่บริเวณขนาดใหญ่ประกอบไปด้วยเช่นกัน ดังนั้นจึงจะต้องอาศัยทั้งภาพที่หยาบและภาพที่ละเอียดในการวิเคราะห์ ไม่สามารถใช้อย่างใดอย่างหนึ่งได้



รูปที่ 2.10 ภาพเดียวกันแต่ถูกย่อให้มีขนาดเล็กลง

การมองเห็นของมนุษย์นั้นจะต้องอาศัยการกวาดสายตาจากพื้นที่ขนาดเล็ก ไปรอบ ๆ รูปภาพ ร่วมกับการมองพื้นที่ขนาดใหญ่โดยรอบ จึงจะสามารถวิเคราะห์ภาพที่เห็นได้ ดังแสดงในรูปที่ 2.10 ถึงแม้ภาพจะมีขนาดที่แตกต่างกันแต่การที่ค่อย ๆ กวาดสายตามองภาพไปที่ละจุดจนครบทุกจุดบนภาพ จะทำให้สามารถวิเคราะห์ภาพออกมาได้ ไม่ว่าภาพจะมีขนาดเล็กหรือใหญ่ก็ตาม



รูปที่ 2.11 โครงสร้างการทำงานของโครงข่ายประสาทเทียมแบบคอนโวลูชัน

โครงสร้างการทำงานของโครงข่ายประสาทเทียมแบบคอนโวลูชันในรูปที่ 2.11 แบ่งเป็นสามส่วน ส่วนที่หนึ่ง คอนโวลูชัน (Convolution) กับพูลลิงเลเยอร์ (Pooling Layer) ส่วนที่สอง ฟูลลีคอนเนคเทดเลเยอร์ (Fully Connected Layer) และส่วนที่สามการทำนาย (Prediction)

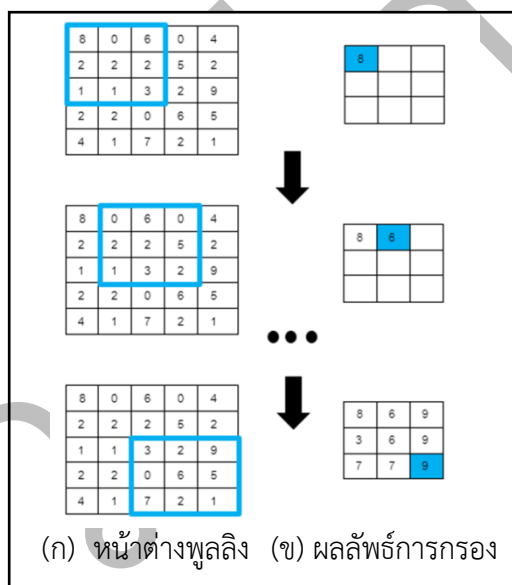
- ส่วนของคอนโวลูชัน กับพูลลิงเลเยอร์ มีข้อมูลเข้า (Input) แบบรูปภาพ ซึ่งจะมีตัวกรอง (Filter) ทำหน้าที่ในการเลื่อนภาพโดยจะเลื่อนไปที่ขั้นตอน (Step) ขึ้นอยู่กับสไตรด์ (Stride) เลื่อนไป

เรื่อย ๆ จนครบทุกพิกเซล (Pixels) ของภาพ ทำให้ได้ฟังก์ชันลักษณะ (Feature Map) ที่มีขนาดเล็กลง นั่นคือภาพได้ถูกย่อขนาดแล้ว และมีการดึงคุณลักษณะเด่นบริเวณต่าง ๆ ออกมาด้วย ซึ่งในบางครั้ง จะต้องมีการเพิ่มแพดดิ้ง (Padding) ให้กับรูปภาพ

- ส่วนของฟูลลีคอนเนคเทดเลเยอร์ จะนำเอาคุณลักษณะจากฟังก์ชันลักษณะที่ถูกดึงมาจากรูปภาพ ไปเรียนรู้ซึ่งกระบวนการเรียนรู้ภายในจะมีการตัดสินใจเกิดขึ้น

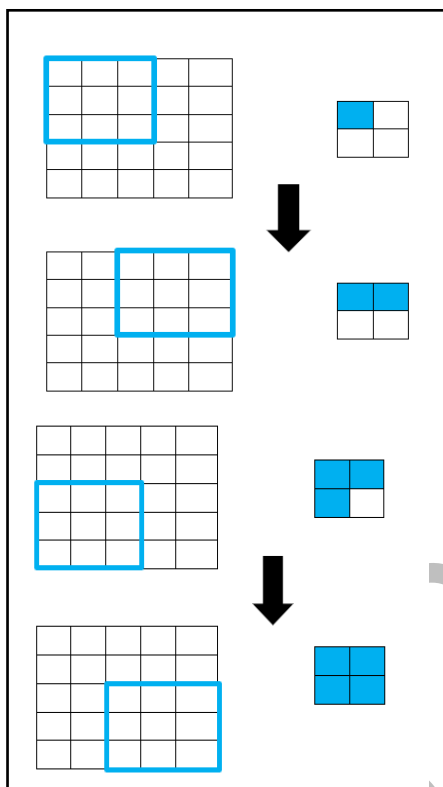
- ส่วนของการทำนาย จะเป็นการทำนายผลลัพธ์ที่ได้จากการเรียนรู้ จากส่วนของฟูลลีคอนเนคเทดเลเยอร์

ในเนื้อหาถัดไปจะอธิบายถึงรายละเอียดในการแยกคุณลักษณะ ประกอบไปด้วย ตัวกรอง สไลด์แพดดิ้ง ฟังก์ชันลักษณะ และพูลลิง (Pooling)



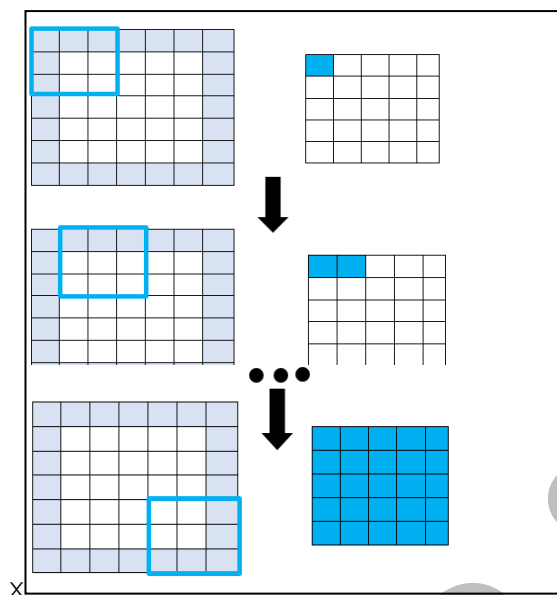
รูปที่ 2.12 ตัวอย่างการกรองด้วยวิธีพูลลิงค่าสูงสุด (Max Pooling)

2.9.1 ตัวกรอง สำหรับภาพดิจิทัลอนั้น ตัวกรองจะอยู่ในรูปแบบของตารางสองมิติที่มีขนาดพื้นที่เล็ก ๆ ยกตัวอย่างเช่น ตัวกรองขนาด 3x3 พิกเซล ที่อยู่บนรูปที่ 2.12 โดยตัวกรองจะถูกทาบบนพิกเซลแรกของภาพ จากนั้นจะเลื่อนไปทาบบนพิกเซลอื่นในภาพทีละพิกเซลจนครบทุกพิกเซลบนภาพ ผลลัพธ์ที่ได้จะเรียกว่าฟังก์ชันลักษณะ โดยตัวกรองหนึ่งตัวกรองจะดึงคุณลักษณะมาได้เพียงหนึ่งอย่าง จึงมีตัวกรองหลายตัวกรอง ซึ่งในบางกรณีตัวกรองนั้นจะไม่ถูกทาบบนบริเวณมุมของภาพเพราะตัวกรองจะล้นออกไปนอกภาพทำให้สูญเสียคุณลักษณะไปได้



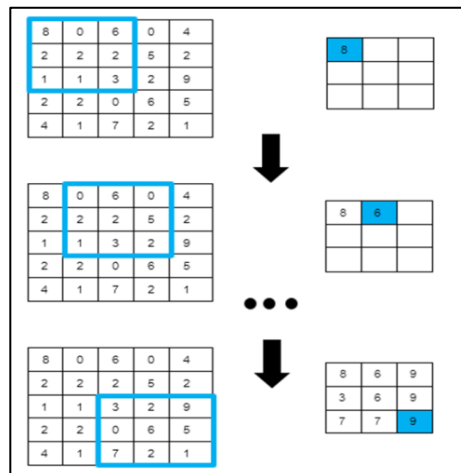
รูปที่ 2.13 ตัวอย่างของการสร้างฟังก์ชันลักษณะ โดยกำหนดค่าสไตรด์เท่ากับ 2 และตัวกรอง เท่ากับ 3×3

2.9.2 สไตรด์ เป็นตัวกำหนดว่าฟิลเตอร์ จะเลื่อนไปที่ขั้นตอน ซึ่งการกำหนดค่าสไตรด์มากขึ้น จะทำให้การหาคำนวณค่าคุณลักษณะมีพื้นที่ที่ทับซ้อนกันน้อยลง แต่นั่นก็ทำให้ฟังก์ชันลักษณะมีขนาดเล็กลง คุณลักษณะที่สำคัญอาจสูญหายไป ตัวอย่างการกำหนด สไตรด์เท่ากับ 2 และตัวกรอง เท่ากับ 3×3 แสดงดังรูปที่ 2.13



รูปที่ 2.14 ตัวอย่างการเพิ่มแพดดิ้ง โดยกำหนดค่าสไตรด์เท่ากับ 1 และตัวกรอง เท่ากับ 3×3

2.9.3 แพดดิ้ง เป็นการเพิ่มพื้นที่สีเทารอบ ๆ รูปภาพ โดยพื้นที่ที่จะถูกเติมด้วยค่า 0 หรือ ค่าต่าง ๆ เข้าไป เพื่อให้ฟังก์ชันลักษณะที่ได้ยังคงมีขนาดเท่ากับข้อมูลเข้า เหตุผลที่ต้องมีการเพิ่มแพดดิ้งเข้าไป เพราะในบางกรณีนั้น ที่บริเวณขอบภาพอาจมีคุณลักษณะที่สำคัญซึ่งอาจจะส่งผลต่อการตัดสินใจบางอย่าง จึงจำเป็นต้องเก็บคุณลักษณะตามขอบของรูปภาพ ตัวอย่างการเพิ่มแพดดิ้ง ซึ่งกำหนดค่าสไตรด์ เท่ากับ 1 และ ตัวกรอง เท่ากับ 3×3 ดังรูปที่ 2.14



รูปที่ 2.15 ตัวอย่างการพูลลิง โดยกำหนดค่าสไตรด์เท่ากับ 1 และตัวกรอง เท่ากับ 3×3

2.9.4 พูลลิง คือ ความสามารถในการย่อขนาดของรูปภาพแบบหนึ่ง ซึ่งหลักการที่นิยมใช้คือ พูลลิงค่าสูงสุด (Max Pooling) เป็นพูลลิงแบบหนึ่งที่หาค่าสูงสุดในบริเวณที่ตัวกรองนั้นทาบอยู่มาเป็นผลลัพธ์ จะทำให้ได้คุณลักษณะที่เด่นที่สุดของของพื้นที่นั้นออกมาดังแสดงในรูปที่ 2.15

สรุป กระบวนการทำงานของโครงข่ายประสาทเทียมแบบคอนโวลูชันนั้น จะเป็นการย่อภาพให้มีขนาดเล็กลงด้วยการพูลลิง แล้วได้คุณลักษณะเด่นออกมา จากนั้นคุณลักษณะเด่นจำนวนมากจะเข้าสู่การเรียนรู้ของโครงข่ายประสาทเทียมแล้วเกิดเป็นผลลัพธ์ในการทำนาย

2.9 การเรียนรู้แบบหน่วยความจำระยะสั้นระยะยาว

การเรียนรู้แบบหน่วยความจำระยะสั้นระยะยาว [14] เป็นประเภทแบบหนึ่งที่ถูกออกแบบมาสำหรับการประมวลผลลำดับ (sequence) ซึ่งพื้นฐานมาจากโครงข่ายประสาทที่เกิดซ้ำ (Recurrent Neural Network : RNN) ที่สามารถเรียนรู้การพึ่งพาระยะยาวในข้อมูล ซึ่งแตกต่างจากโครงข่ายประสาทเทียมที่เกิดซ้ำแบบดั้งเดิมซึ่งมีแนวโน้มที่จะลืมข้อมูลเมื่อเวลาผ่านไปเราเรียกปัญหานี้ว่า Gradient Vanishing

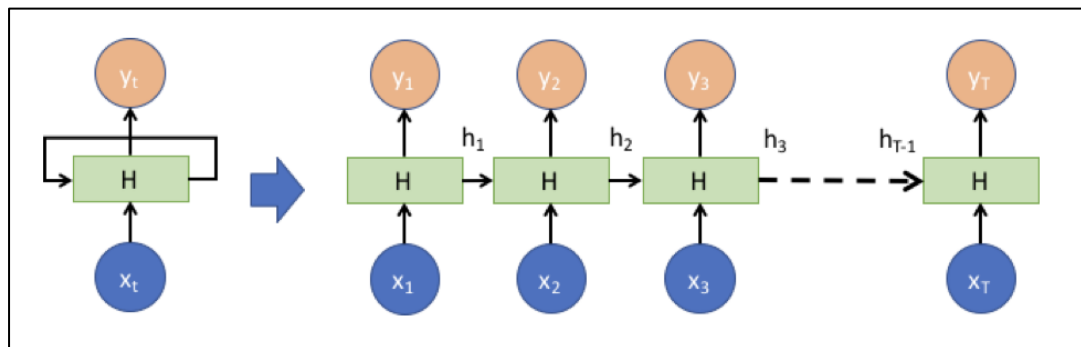
อธิบายกระบวนการโครงข่ายประสาทเทียมที่เกิดซ้ำ

- input data คือ x_1, x_2, \dots, x_t
- hidden state ที่เวลา $t = h_t$

input คือ ประโยคว่า “ฉันกินข้าว” จะอ่านประโยคนี้อะไรทีละคำ ๆ

$x_1 = \text{“ฉัน”}$, $x_2 = \text{“กิน”}$ และ $x_3 = \text{“ข้าว”}$

แสดงรูปตัวอย่างกระบวนการโครงข่ายประสาทเทียมที่เกิดซ้ำดังรูปที่ 2.16



รูปที่ 2.16 ตัวอย่างกระบวนการโครงข่ายประสาทเทียมที่เกิดซ้ำ

จากรูปที่ 2.16 สามารถอธิบายตัวอย่างกระบวนการโครงข่ายประสาทเทียมที่เกิดซ้ำโดยที่

- H = hidden layer
- y_t = output จาก RNN ที่เวลา t
- x_t = input data ที่เวลา t
- h_t = hidden state ที่เวลา t

รูปทางซ้ายแสดงให้เห็นว่ามี loop ที่วนกลับเข้ามาที่ hidden layer ของโครงข่ายประสาทซึ่งสิ่งของโครงข่ายประสาทเทียมที่เกิดซ้ำคือ hidden state ที่ $t-1$ และ input data ที่ t ประโยชน์ของ loop ก็คือเพื่อที่จะนำ hidden state ที่ $t-1$ เข้ามาใช้ (จะมองว่าโครงข่ายประสาทเทียมที่เกิดซ้ำมันก็คือโครงข่ายประสาทที่มี memory เพิ่มขึ้นมา เพื่อเก็บค่า hidden state ที่ได้คำนวณเอาไว้ที่ $t-1$) ส่วนในรูปทางขวาเป็นรูปที่คลี่รูปทางซ้ายออกมาแล้วเพื่อแสดงการทำงานเป็นทีละขั้นตอน (ทั้งหมดเป็นการแสดงโครงข่ายประสาทเทียมที่เกิดซ้ำตัวเดียวกันหมด แต่ว่าแสดงความเป็น sequence ออกมา) ในส่วนของสมการดังรูปที่ 2.17

$$\begin{aligned} & \bullet h_t = f_h(U_h h_{t-1} + W_h x_t + b_h) \\ & \bullet y_t = f_y(W_y h_t + b_y) \end{aligned}$$

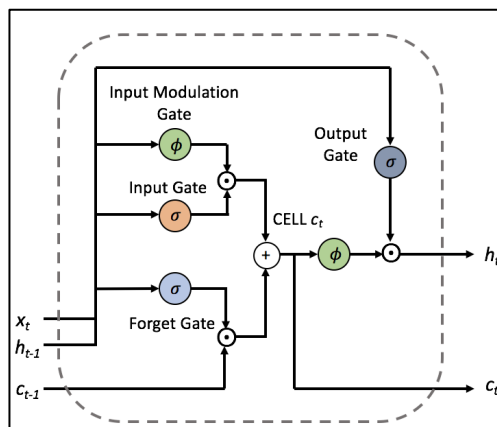
รูปที่ 2.17 สมการโครงข่ายประสาทเทียมที่เกิดซ้ำ

จากรูปที่ 2.17 สามารถอธิบายสมการโครงข่ายประสาทเทียมที่เกิดซ้ำโดยที่

- f_h คือ activation function ของ hidden layer (เช่น tanh หรือ ReLU หรือ sigmoid function)
- f_y คือ activation function ของ output layer (เช่น softmax function)
- W_h คือ weight matrix ของ hidden layer
- U_h คือ hidden-state-to-hidden-state matrix (หรือ transition matrix)

ในการคำนวณ hidden state ที่เวลา t ออกมาได้นั้น (h_t) จะต้องใช้ 2 ตัวแปรสำคัญ หรือ hidden state ก่อนหน้า (h_{t-1}) และ input data ณ ตอนนั้น (x_t)

การเรียนรู้แบบหน่วยความจำระยะสั้นระยะยาว มีความสามารถในการเก็บข้อมูลสำคัญเป็นระยะเวลานาน แสดงรูปกระบวนการทำงานของการเรียนรู้แบบหน่วยความจำระยะสั้นระยะยาวดังรูปที่ 2.18 ทำให้มีประโยชน์อย่างยิ่งสำหรับงานที่ต้องการการสร้างแบบจำลองของการพึ่งพาระยะยาว การเรียนรู้แบบหน่วยความจำระยะสั้นระยะยาว บรรลุเป้าหมายนี้โดยใช้ชุดประตู่ที่ควบคุมการไหลของข้อมูลผ่านเครือข่าย เกตเหล่านี้ช่วยให้ การเรียนรู้แบบหน่วยความจำระยะสั้นระยะยาว สามารถเลือกเก็บรักษาหรือลืมข้อมูล ตลอดจนอัปเดตสถานะภายในของเครือข่ายตามอินพุตใหม่ สิ่งนี้ช่วยให้ การเรียนรู้แบบหน่วยความจำระยะสั้นระยะยาว สามารถ “จดจำ” ข้อมูลสำคัญได้อย่างมีประสิทธิภาพเป็นระยะเวลานานขึ้น ในขณะเดียวกันก็สามารถปรับให้เข้ากับข้อมูลใหม่ตามที่น่าเสนอได้ การเรียนรู้แบบหน่วยความจำระยะสั้นระยะยาว ถูกนำไปใช้กับงานที่หลากหลายอย่างประสบความสำเร็จ รวมถึงการประมวลผลภาษาธรรมชาติ การรู้จำเสียง การแปลด้วยคอมพิวเตอร์ และการคาดคะเนอนุกรมเวลา พวกเขายังใช้ในแอปพลิเคชันต่างๆ ในด้านการเงิน เช่น การทำนายตลาดหุ้น และการตรวจจับการฉ้อโกง



รูปที่ 2.18 กระบวนการทำงานของการเรียนรู้แบบหน่วยความจำระยะสั้นระยะยาว

จากรูปที่ 2.18 สามารถอธิบายกระบวนการทำงานของการเรียนรู้แบบหน่วยความจำระยะสั้นระยะยาวซึ่ง

- Cell state เป็นตัวเก็บ state ของ memory cell ใน LSTM
- Gate เป็นตัวที่ควบคุมการไหลของข้อมูล ซึ่งก็คือ ค่า analog ที่คอยควบคุมว่าจะ read, write หรือ forget ซึ่งมันก็เหมือนกับประตูที่ควรเปิดให้ข้อมูลไหลเข้า ไหลออก หรือไหลหายไปเลย (forget)

Forget

Forget คือ การเคลียร์ cell state เดิมออกไป คือ เตรียมเคลียร์พื้นที่รับข้อมูลใหม่ โดยที่ forget gate มีหน้าที่ ตัดสินใจ (forget gate = 0 ลบ cell state , forget gate = 1 เก็บ cell state) ซึ่ง forget gate จะดูจาก input data ที่เข้ามา ประกอบกับ hidden state ก่อนหน้า (ตามสูตรของโครงข่ายประสาทเทียมที่เกิดซ้ำ) ประกอบการตัดสินใจ โดยจะใช้ sigmoid function เป็นตัวตัดสินใจ แสดงตัวอย่างสมการ forget gate ดังรูปที่ 2.19

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)$$

รูปที่ 2.19 ตัวอย่างสมการ forget gate

Write

Write คือเมื่อมี input data ใหม่เข้ามาจะต้องพิจารณาว่าการอัปเดต cell state ด้วย input data ใหม่หรือไม่

- การอัปเดต cell state จะควบคุมโดย input gate ซึ่งก็จะยังคงใช้ sigmoid function เป็นสมการที่ใช้ในตัดสินใจให้ทำการอัปเดต ซึ่งการคำนวณจะใช้ค่า input data ที่เข้ามา กับ hidden state ก่อนหน้านั้น แสดงตัวอย่างสมการ input gate ที่ใช้ในควบคุมการอัปเดต cell state ดังรูปที่ 2.20

$$i_t = \sigma(W_x^i x_t + W_h^i h_{t-1} + b_i)$$

รูปที่ 2.20 ตัวอย่างสมการ input gate ที่ใช้ในควบคุมการอัปเดต cell state

- ในการอัปเดตค่า cell state ใหม่ซึ่งเรียกว่า Input modulation gate เป็นตัวจัดการ โดยสมการก็จะคล้ายๆ กับ input gate แต่จะใช้เป็น tanh function แทน ซึ่งค่าที่ได้ จะมองว่าเป็น cell state candidate แสดงตัวอย่างสมการ Input modulation gate ที่ใช้ในการอัปเดตค่า cell state ดังรูปที่ 2.21

$$g_t = \tanh(W_x^c x_t + W_h^c h_{t-1} + b_c)$$

รูปที่ 2.21 ตัวอย่างสมการ Input modulation gate ที่ใช้ในการอัปเดตค่า cell state

อัปเดต cell state

การอัปเดตค่า cell state สามารถคำนวณจากจาก forget gate, input gate และ input modulation gate แสดงตัวอย่างสมการของการอัปเดตค่า cell state ดังรูปที่ 2.22

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t$$

รูปที่ 2.22 ตัวอย่างสมการของการอัปเดตค่า cell state

- ในส่วนแรกค่าถ้า f_t (forget gate) มีค่าเป็น 0 ให้ลบ cell state เดิม ก็จะนำค่า c_{t-1} มาประกอบการอัปเดต cell state แต่ถ้า f_t มีค่าเป็น 1 เราก็จะยังคงค่า c_{t-1} เอาไว้ประกอบการพิจารณาการอัปเดต
- ในส่วนที่สองของการอัปเดต cell state จากข้อมูลใหม่ถ้าค่า g_t (input modulation gate) จะถูกใช้งานก็ต่อเมื่อ output จาก i_t (input gate) มาใช้ในควมคุม ถ้า i_t เป็น 1 ก็ใช้ค่า g_t อัปเดต แต่ถ้า i_t เป็น 0 ก็ไม่ใช้ g_t อัปเดต cell state

4. Read

Read คือเป็นสิทธิ์ (permission) ในอนุญาตให้ sequence ถัดไป มา read ตัว h_t โดยจะใช้ output gate มาใช้ในควมคุม ซึ่งใช้สมการ forget gate และ input gate ก็คือ ใช้ sigmoid function กับค่า hidden state ตัวก่อนหน้า กับ input data ที่เข้ามา แสดงตัวอย่างสมการของ output gate ที่ใช้ในการควบคุมการ read ใน sequence ถัดไป ดังรูปที่ 2.23

$$o_t = \sigma(W_{x^o}x_t + W_{h^o}h_{t-1} + b_o)$$

รูปที่ 2.23 ตัวอย่างสมการของ output gate ที่ใช้ในการควบคุมการ read ใน sequence ถัดไป

ในการคำนวณ h_t ถ้า o_t (output gate) มีค่าเป็น 0 ค่าของ h_t ก็จะมีค่าเป็น 0 คือไม่ส่งค่าให้กับ sequence ถัดไป หาก o_t มีค่าเป็น 1 จะคำนวณค่า h_t และส่งออกไป sequence ถัดไป แสดงตัวอย่างสมการของ h_t ของ sequence ถัดไป ดังรูปที่ 2.24

$$h_t = o_t \odot \tanh(c_t)$$

รูปที่ 2.24 ตัวอย่างสมการของ h_t ของ sequence ถัดไป

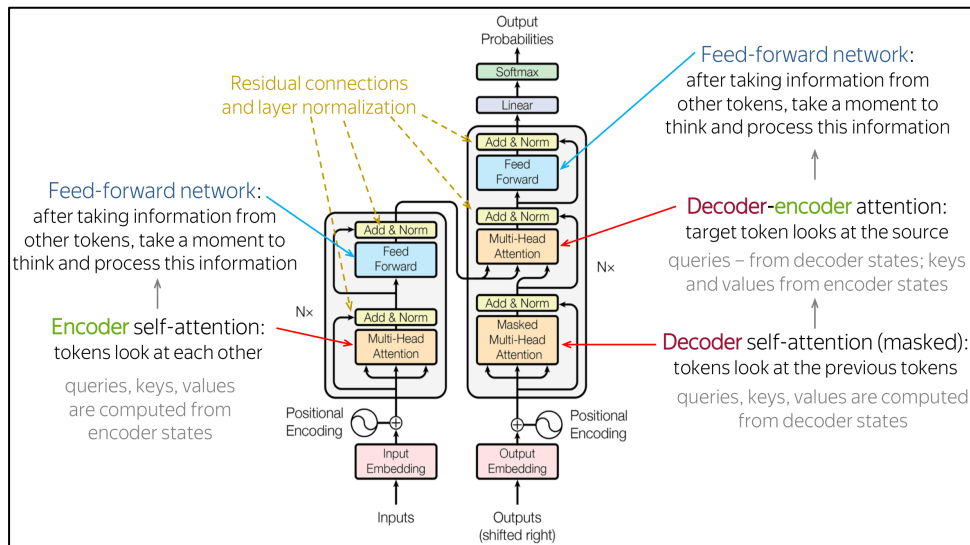
2.10 โครงข่ายประสาทเทียมโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์

โครงข่ายประสาทเทียมโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ [16] เป็นสถาปัตยกรรมเครือข่ายประสาทประเภทหนึ่งที่ได้รับการแนะนำในบทความเรื่อง "Attention Is All You Need" โดย Vaswani et al ในปีพ.ศ. 2560 เป็นแบบจำลองที่อิงตามกลไกความสนใจทั้งหมด และบรรลุผลอันล้ำสมัยในงานประมวลผลภาษาธรรมชาติมากมาย

ข้อได้เปรียบหลักของสถาปัตยกรรมโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์คือช่วยให้สามารถคำนวณแบบขนานได้โดยใช้กลไกการเอาใจใส่ตนเองแทนการเชื่อมต่อที่เกิดซ้ำ เช่นที่พบในโครงข่ายประสาทเทียมที่เกิดซ้ำสิ่งนี้ทำให้ฝึกได้เร็วขึ้นมากและช่วยให้ปรับขนาดเป็นลำดับอินพุตที่ใหญ่มากได้

สถาปัตยกรรมของโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ประกอบด้วย "ชั้น" หลายชั้น ซึ่งแต่ละชั้นประกอบด้วยกลไกการให้ความสนใจในตัวเองและโครงข่ายประสาทเทียมแบบฟีดฟอร์เวิร์ด กลไกการให้ความสนใจในตัวเองช่วยให้โมเดลสามารถเข้าร่วมส่วนต่างๆ ของลำดับอินพุตและคำนวณผลรวมถ่วงน้ำหนักของการแสดงอินพุต ในขณะที่เครือข่ายฟีดฟอร์เวิร์ดประมวลผลผลรวมถ่วงน้ำหนักและสร้างเอาต์พุต

โดยรวมแล้ว สถาปัตยกรรมโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์มีผลกระทบอย่างมากต่อการประมวลผลภาษาธรรมชาติ และถูกนำมาใช้ในแอปพลิเคชันที่หลากหลาย รวมถึงการแปลด้วยเครื่อง การสร้างแบบจำลองภาษา และการจำแนกข้อความ แสดงรูปการณ์ทำงานของโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ดังรูปที่ 2.25



รูปที่ 2.25 การทำงานของโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์

จากรูปที่ 2.25 สามารถอธิบายองค์ประกอบด้วยส่วนต่าง ๆ ของการทำงานของโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ดังนี้

- Input Embedding เป็นวิธีการแสดงข้อความอินพุต (หรือข้อมูลอื่นๆ) ในรูปแบบตัวเลขที่สามารถใช้เป็นอินพุตไปยังโมเดลแมชชีนเลิร์นนิง เช่น หม้อแปลงไฟฟ้า สิ่งนี้ทำได้โดยการแมปข้อความอินพุตกับเวกเตอร์ขนาดคงที่ โดยทั่วไปจะใช้เทคนิคต่างๆ เช่น การเข้ารหัสแบบใช้ครั้งเดียว (One-Hot Encoding) หรือการฝังคำ (Word Embeddings) จากนั้นโมเดลหม้อแปลงจะใช้อินพุตที่ฝังไว้นี้เพื่อการทำนายหรือทำงานอื่นๆ
- Multi-Head Attention (Encoder) เป็นการเน้นที่ส่วนที่มีความเกี่ยวข้องของคำเฉพาะกับคำอื่น ๆ ในประโยคนั้น มักจะถูกแสดงเป็นเวกเตอร์ความสนใจ (Attention Vector) ที่สร้างขึ้น ซึ่งรวบรวมความสัมพันธ์ตามบริบทระหว่างคำในประโยคนั้น แสดงตัวอย่างการเน้นที่ส่วนที่มีความเกี่ยวข้องของคำเฉพาะกับคำอื่น ๆ ในประโยคนั้นดังรูปที่ 2.26

| Attention : What part of the input should we focus? | | | Attention Vectors | | | |
|---|-------|-----------------|-------------------|------|------|--------------------|
| | Focus | | | | | |
| The | → | The big red dog | [0.71 | 0.04 | 0.07 | 0.18] ^T |
| big | → | The big red dog | [0.01 | 0.84 | 0.02 | 0.13] ^T |
| red | → | The big red dog | [0.09 | 0.05 | 0.62 | 0.24] ^T |
| dog | → | The big red dog | [0.03 | 0.03 | 0.03 | 0.91] ^T |

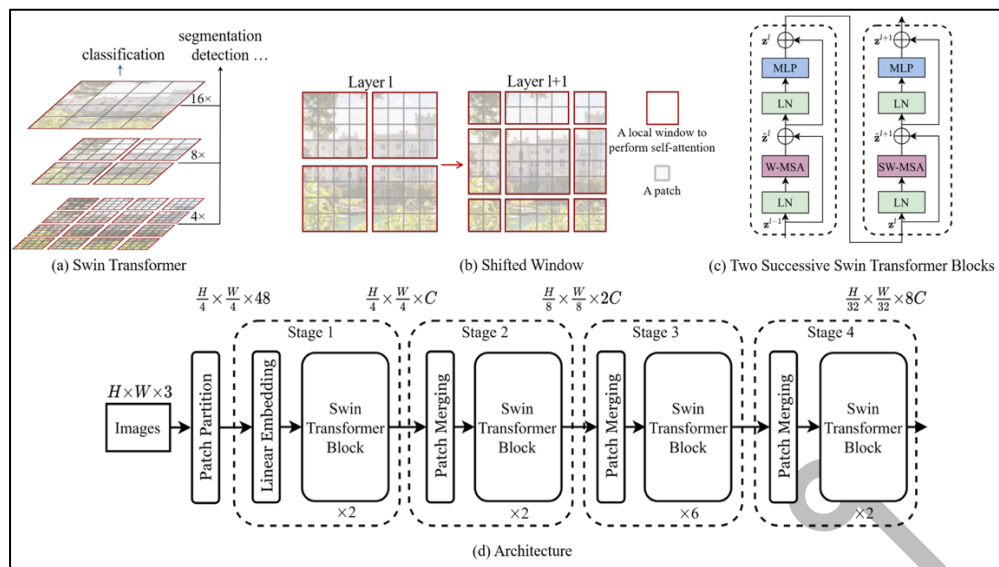
รูปที่ 2.26 ตัวอย่างการเน้นที่ส่วนที่มีความเกี่ยวข้องของคำเฉพาะกับคำอื่น ๆ ในประโยคนั้น

- Feed Forward Neural Network คือโครงข่ายประสาทเทียมแบบ feed-forward แบบง่ายที่ใช้กับเวกเตอร์ความสนใจทั้งหมด จุดประสงค์หลักคือการแปลงเวกเตอร์ความสนใจให้อยู่ในรูปแบบที่เลเยอร์เพื่อที่จะส่งข้อมูลระหว่างตัวเข้ารหัสไปยังตัวถอดรหัสได้ โดยเวกเตอร์ความสนใจแต่ละตัวเป็นอิสระจากกัน ดังนั้น สามารถใช้การทำให้ขนานกันได้สามารถส่งค่าทั้งหมดพร้อมกันไปยังบล็อกตัวเข้ารหัสและรับชุดข้อมูลเวกเตอร์ถอดรหัส สำหรับทุกคำ ๆ พร้อมกัน
- Masked Multi-Head Attention คือการเน้นที่ส่วนที่มีความเกี่ยวข้องของคำเฉพาะกับคำอื่น ๆ ในประโยคนั้นที่เป็นคำก่อนหน้า โดยทำการการซ่อนคำถัดไป เพื่อที่ในตอนแรกคำนั้นจะทำนายคำถัดไปเองจากลัฟร์ก่อนหน้า

2.11 การเรียนรู้โครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์สวิน

การเรียนรู้โครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์สวิน [6] เป็นรูปแบบหนึ่งของหม้อแปลงที่ได้รับการแนะนำในเอกสาร "Hierarchical Vision Transformer โดยใช้ Shifted Windows" โดย Liu และคณะ (2563). เป็นรูปแบบหนึ่งของสถาปัตยกรรมโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์สที่ออกแบบมาเพื่อใช้งานด้านการมองเห็นด้วยคอมพิวเตอร์ โดยเฉพาะสำหรับการจำแนกประเภทภาพ

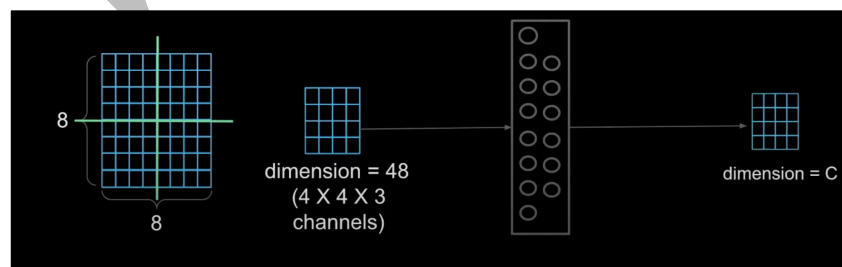
คุณสมบัติหลักประการหนึ่งของการเรียนรู้โครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์สวินคือการใช้หน้าต่างแบบเลื่อน ซึ่งช่วยให้แบบจำลองสามารถจับภาพคุณลักษณะทั้งในพื้นที่และส่วนกลางในอิมเมจอินพุต แบบจำลองนี้ยังมีโครงสร้างแบบลำดับชั้น ซึ่งช่วยให้สามารถประมวลผลภาพอินพุตที่หลากหลาย ปรับปรุงความสามารถในการจับภาพรายละเอียดที่ละเอียดยิ่งขึ้น แสดงตัวอย่างกระบวนการทำงานของโครงข่ายประสาทเทียมโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์สวิน ดังรูปที่ 2.27



รูปที่ 2.27 กระบวนการทำงานของโครงข่ายประสาทเทียมโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์สวิน

จากรูปที่ 2.27 สามารถอธิบายกระบวนการโครงข่ายประสาทเทียมโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์สวิน

- ทำการแยกภาพระบบสีแบบอาจีบีออกเป็นแพตช์ที่ไม่ทับซ้อนกัน ในแต่ละแพตช์จะถือว่าเป็น “โทเค็น” และคุณลักษณะของแพตช์จะถูกตั้งค่าเป็นการต่อเชื่อมของค่าระบบสีแบบอาจีบี มีการใช้ขนาดแพตช์ 4×4 ดังนั้นมิติคุณลักษณะของแต่ละแพตช์คือ $4 \times 4 \times 3 = 48$ เลเยอร์การฝังเชิงเส้น (Linear Embedding Layer) คือการนำข้อมูลคุณลักษณะของภาพมาทำการฉายไปยังมิติ C (มาตรฐานมิติ $C = 96$) แสดงดังรูปที่ 2.28



รูปที่ 2.28 การนำข้อมูลคุณลักษณะของภาพมาทำการฉายไปยังมิติ C

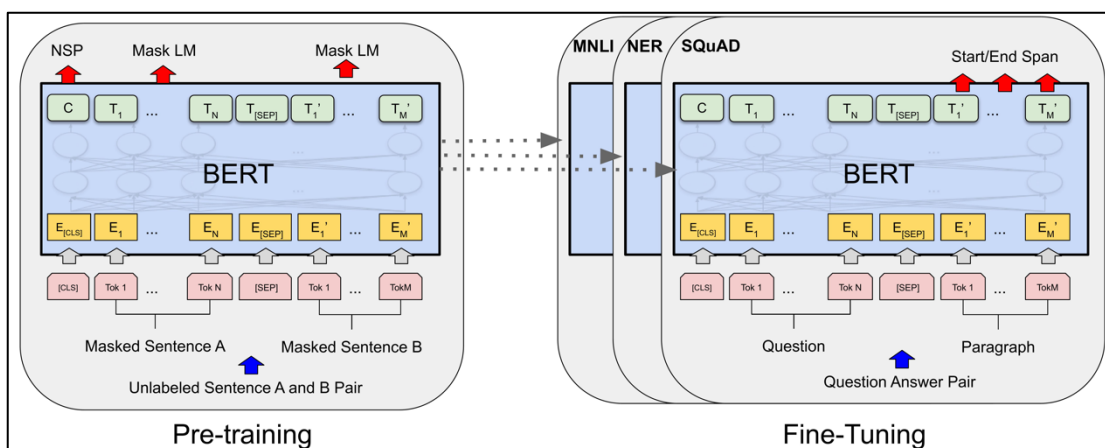
- บล็อกทรานส์ฟอร์มเมอร์หลาย ๆ บล็อกจะมีการคำนวณ Shifted Window Based Self-Attention เมื่อคำนวณเสร็จแล้วจะถูกนำไปใช้กับโทเค็นแพตช์ โดยบล็อกทรานส์ฟอร์มเมอร์จะเก็บจำนวนของโทเค็น $H/4 \times W/4$ และก็จะทำการฝังเชิงเส้นเรียกว่า “Stage 1”
- ในการสร้างการแสดงผลแบบลำดับขั้นจำนวนของโทเค็นจะลดลงโดยการผสานชั้นของแพตช์เมื่อเครือข่ายลึกมากขึ้น แรกจะผสานชั้นคุณลักษณะของแต่ละกลุ่มของแพตช์ที่อยู่ใกล้เคียงขนาด 2×2 และใช้เลเยอร์เชิงเส้นบนคุณลักษณะที่ต่อกัน 4 มิติ ซึ่งก็จะลดจำนวนโทเค็นโดยการคูณของ $2 \times 2 = 4$ (ลดความละเอียดลง 2 เท่า) ขนาดเอาต์พุตถูกตั้งค่าเป็น $2C$ และคงความละเอียดไว้ที่ $H/8 \times W/8$ บล็อกแรกของการรวมแพตช์และการแปลงคุณลักษณะนี้จะแสดงใน “State 2”
- ขั้นตอนซ้ำสองครั้ง ในชื่อ “State 3” ขนาดของเอาต์พุตเป็น $4C$ ความละเอียด $H/16 \times W/16$ และ “State 4” ขนาดของเอาต์พุตเป็น $8C$ และคงความละเอียด $H/32 \times W/32$

ผู้เขียนรายงานได้แสดงให้เห็นว่าการเรียนรู้โครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์สวินมีประสิทธิภาพเหนือกว่าโมเดลที่ทันสมัยหลายชุดในชุดข้อมูลมาตรฐานจำนวนหนึ่ง ซึ่งรวมถึง ImageNet และ COCO อย่างไรก็ตาม ควรสังเกตว่าการเรียนรู้โครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์สวินเป็นโมเดลที่ค่อนข้างใหม่ และจำเป็นต้องมีการวิจัยเพิ่มเติมเพื่อให้เข้าใจถึงความสามารถและข้อจำกัดของมันอย่างถ่องแท้

2.12 การแทนตัวเข้ารหัสแบบสองทิศทางจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์

การแทนตัวเข้ารหัสแบบสองทิศทางจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ [12] คือรูปแบบการประมวลผลภาษาธรรมชาติอันล้ำสมัยที่พัฒนาโดยนักวิจัยที่ Google มีพื้นฐานมาจากสถาปัตยกรรมทรานส์ฟอร์มเมอร์ และได้รับการฝึกให้ปฏิบัติงานได้หลากหลาย รวมถึงการแปลภาษาและการจำแนกข้อความ

คุณลักษณะสำคัญประการหนึ่งของการแทนตัวเข้ารหัสแบบสองทิศทางจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์คือประมวลผลลำดับอินพุตในลักษณะสองทิศทาง ซึ่งหมายความว่า จะพิจารณาบริบททั้งทางซ้ายและทางขวาของแต่ละคำในอินพุต ซึ่งช่วยให้การแทนตัวเข้ารหัสแบบสองทิศทางจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ เข้าใจบริบทและความหมายของคำในอินพุตได้ดีขึ้น ซึ่งเป็นสิ่งสำคัญสำหรับงานประมวลผลภาษาธรรมชาติจำนวนมาก แสดงกระบวนการทำงานของการแทนตัวเข้ารหัสแบบสองทิศทางจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ ดังรูปที่



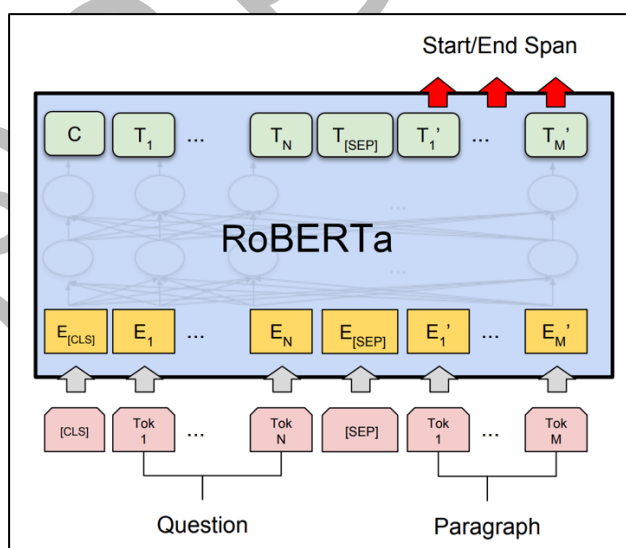
รูปที่ 2.29 กระบวนการทำงานของการแทนตัวเข้ารหัสแบบสองทิศทางจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์

นอกเหนือจากการประมวลผลแบบสองทิศทางแล้วการแทนตัวเข้ารหัสแบบสองทิศทางจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ ยังใช้กลไกการให้ความสนใจแบบหลายหัว ซึ่งช่วยให้สามารถเข้าร่วมส่วนต่าง ๆ ของลำดับอินพุตพร้อมกันได้ สิ่งนี้ช่วยให้โมเดลจับการอ้างอิงระยะยาวและความสัมพันธ์ระหว่างคำในอินพุตได้อย่างมีประสิทธิภาพ

โดยรวมแล้วการแทนตัวเข้ารหัสแบบสองทิศทางจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ ได้รับผลลัพธ์ที่น่าประทับใจในงานประมวลผลภาษาธรรมชาติที่หลากหลาย และได้กลายเป็นรูปแบบที่ใช้กันอย่างแพร่หลายในสาขานี้ ได้กำหนดเกณฑ์มาตรฐานใหม่สำหรับการปฏิบัติงาน เช่น การแปลภาษา การตอบคำถาม และการจัดประเภทข้อความ และได้แสดงให้เห็นถึงประสิทธิภาพของสถาปัตยกรรมทรานส์ฟอร์มเมอร์สำหรับงานประเภทนี้

2.13 การเรียนรู้แบบฝึกอบรมล่งหน้า BERT ที่ปรับให้เหมาะสมอย่างมีประสิทธิภาพจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์

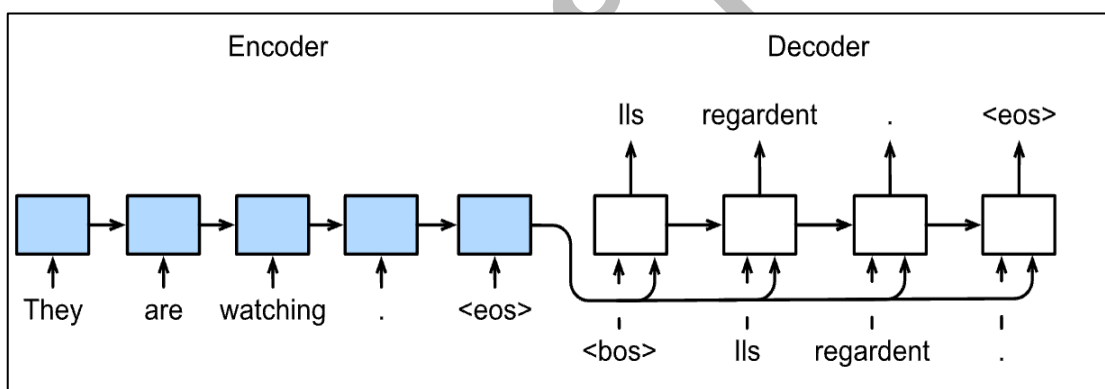
การเรียนรู้แบบฝึกอบรมล่งหน้า BERT ที่ปรับให้เหมาะสมอย่างมีประสิทธิภาพจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ [18] คือโมเดลภาษาที่พัฒนาโดย Facebook AI Research (FAIR) ซึ่งเปิดตัวในปี 2019 โมเดลนี้เป็นส่วนเสริมของโมเดล BERT ที่ได้รับความนิยมและได้รับการฝึกบนชุดข้อมูลขนาดใหญ่กว่ามาก ซึ่งรวมถึงความหลากหลาย ชุดหนังสือ บทความ และเว็บไซต์การเรียนรู้แบบฝึกอบรมล่งหน้า BERT ที่ปรับให้เหมาะสมอย่างมีประสิทธิภาพจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ ใช้เทคนิคที่เรียกว่าไดนามิกแมสก์ ซึ่งหมายความว่าระหว่างการฝึกล่งหน้า โมเดลจะสุ่มมาส์กโทเค็นที่แตกต่างกันภายในอินพุต ทำให้โมเดลมีประสิทธิภาพมากขึ้นสำหรับการมาส์กประเภทต่างๆ นอกจากนี้การเรียนรู้แบบฝึกอบรมล่งหน้า BERT ที่ปรับให้เหมาะสมอย่างมีประสิทธิภาพจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ใช้ขนาดแบทช์ที่ใหญ่ขึ้นและขั้นตอนการฝึกอบรมที่มากขึ้น ซึ่งเป็นการเพิ่มประสิทธิภาพให้ดียิ่งขึ้น การเรียนรู้แบบฝึกอบรมล่งหน้า BERT ที่ปรับให้เหมาะสมอย่างมีประสิทธิภาพจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ได้รับการพิสูจน์แล้วว่ามีประสิทธิภาพดีกว่า BERT รุ่นก่อนในงานทำความเข้าใจภาษาธรรมชาติหลายอย่าง เช่น การตอบคำถาม การวิเคราะห์ความรู้สึก และการจัดประเภทข้อความ แสดงกระบวนการเรียนรู้แบบฝึกอบรมล่งหน้า BERT ที่ปรับให้เหมาะสมอย่างมีประสิทธิภาพจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ดังรูปที่ 2.30



รูปที่ 2.30 กระบวนการทำงานของการเรียนรู้แบบฝึกอบรมล่งหน้า BERT ที่ปรับให้เหมาะสมอย่างมีประสิทธิภาพจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์

2.14 การเรียนรู้เกี่ยวกับลำดับตัวเข้ารหัส-ตัวถอดรหัสเป็นรูปแบบลำดับ

การเรียนรู้เกี่ยวกับลำดับตัวเข้ารหัส-ตัวถอดรหัสเป็นรูปแบบลำดับ [15] คือสถาปัตยกรรมโครงข่ายประสาทชนิดหนึ่งที่ใช้กันทั่วไปสำหรับงานประมวลผลภาษาธรรมชาติ เช่น การแปลด้วยเครื่อง การสรุป และการสร้างข้อความ ประกอบด้วยสองส่วนแยกกัน: ตัวเข้ารหัสและตัวถอดรหัส ตัวเข้ารหัสจะประมวลผลลำดับอินพุต ซึ่งโดยปกติจะเป็นประโยคหรือเอกสาร และแปลงเป็นการแสดงความยาวคงที่ ซึ่งเรียกว่าเวกเตอร์บริบท เวกเตอร์นี้จะถูกส่งผ่านไปยังตัวถอดรหัส ซึ่งสร้างลำดับเอาต์พุต ทีละโทเค็น โดยทั่วไปแล้วตัวถอดรหัสได้รับการฝึกฝนเพื่อเพิ่มความเป็นไปได้ของลำดับเอาต์พุตสูงสุดที่กำหนดลำดับอินพุตและเวกเตอร์บริบท สถาปัตยกรรมตัวเข้ารหัส-ตัวถอดรหัสเป็นตัวอย่างของโมเดลภาษาแบบมีเงื่อนไข ซึ่งลำดับเอาต์พุตถูกสร้างขึ้นตามลำดับอินพุตและเวกเตอร์บริบท มีการใช้กันอย่างแพร่หลายในงานประมวลผลภาษาธรรมชาติเนื่องจากสามารถจัดการลำดับอินพุตและเอาต์พุตที่มีความยาวผันแปรได้ และสามารถจัดการกับการขึ้นต่อกันที่ซับซ้อนระหว่างลำดับอินพุตและเอาต์พุต แสดงการทำงานของการทำงานของการเรียนรู้เกี่ยวกับลำดับตัวเข้ารหัส-ตัวถอดรหัสเป็นแบบลำดับดังรูปที่ 2.31



รูปที่ 2.31 การทำงานของการเรียนรู้เกี่ยวกับลำดับตัวเข้ารหัส-ตัวถอดรหัสเป็นแบบลำดับ

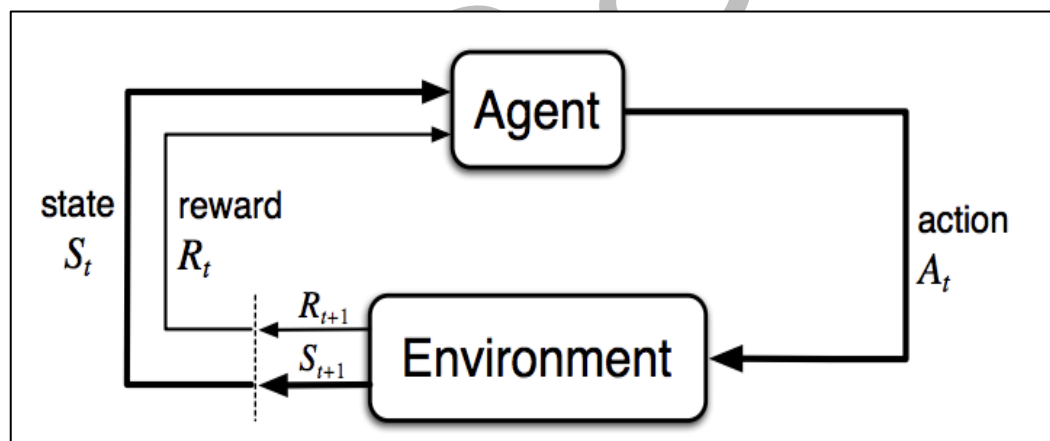
จากรูปที่ 2.31 สามารถอธิบายการทำงานของการทำงานของการเรียนรู้เกี่ยวกับลำดับตัวเข้ารหัส-ตัวถอดรหัสเป็นแบบลำดับ สาธิตการประยุกต์ใช้สถาปัตยกรรมตัวเข้ารหัส-ตัวถอดรหัส ซึ่งทั้งตัวเข้ารหัสและตัวถอดรหัสถูกปรับใช้เป็นโครงข่ายประสาทเทียมที่เกิดซ้ำเพื่องานการแปลด้วยคอมพิวเตอร์ โดยตัวเข้ารหัสโครงข่ายประสาทเทียมที่เกิดซ้ำจะใช้ลำดับความยาวผันแปรเป็นอินพุต (Variable-Length Sequence) และแปลงเป็นสถานะซ่อนรูปร่างคงที่ (Fixed-Shape) ซึ่งจะมีกลไกความสนใจ (Attention Mechanisms) ช่วยให้เราสามารถเข้าถึงอินพุตที่เข้ารหัสโดยไม่ต้องบีบอัดอินพุตทั้งหมด

ให้เป็นตัวแทนความยาวคงที่เดียว จากนั้นเพื่อสร้างลำดับเอาต์พุต ทีละหนึ่งโทเค็น โมเดลตัวถอดรหัสจะประกอบด้วยโครงข่ายประสาทเทียมที่เกิดซ้ำแยกต่างหาก จะคาดการณ์โทเค็นเป้าหมายที่ต่อเนื่องกันโดยให้ทั้งลำดับอินพุตและโทเค็นก่อนหน้าในเอาต์พุต ในระหว่างการฝึกอบรม โดยทั่วไปแล้วตัวถอดรหัสจะถูกกำหนดเงื่อนไขตามโทเค็นก่อนหน้าในปายกำกับ "ความจริงพื้นฐาน" (ground-truth) เวลาทดสอบ จะต้องกำหนดเงื่อนไขเอาต์พุตของตัวถอดรหัสแต่ละตัวบนโทเค็นที่จะทำนาย

2.15 การเรียนรู้แบบเสริมกำลัง

การเรียนรู้แบบเสริมกำลัง [11] คือการเรียนรู้ของเครื่องประเภทหนึ่งที่ตัวแทนเรียนรู้ที่จะโต้ตอบกับสภาพแวดล้อมเพื่อเพิ่มรางวัลสูงสุด ตัวแทนเรียนรู้โดยการดำเนินการและรับข้อเสนอแนะในรูปแบบของรางวัลหรือบทลงโทษ

เป้าหมายของตัวแทนคือการเรียนรู้นโยบายที่จะเพิ่มรางวัลสะสมสูงสุดเมื่อเวลาผ่านไป สิ่งนี้ทำผ่านการลองผิดลองถูก ขณะที่ตัวแทนพยายามดำเนินการต่างๆ และได้รับรางวัลหรือบทลงโทษตามผลลัพธ์ แสดงตัวอย่างกระบวนการทำงานของ ดังรูปที่ 2.32



รูปที่ 2.32 กระบวนการทำงานของการเรียนรู้แบบเสริมกำลัง

จากรูปที่ 2.32 สามารถอธิบายกระบวนการทำงานของการเรียนรู้แบบเสริมกำลังได้ดังนี้

1. ระบุปัญหา : กำหนดงานที่ตัวแทนควรดำเนินการให้ชัดเจน รวมถึงสภาพแวดล้อมและระบบรางวัล
2. สร้างแบบจำลองสภาพแวดล้อม : สร้างแบบจำลองทางคณิตศาสตร์ของสภาพแวดล้อมที่เอเจนต์จะโต้ตอบด้วย
3. เลือกอัลกอริทึม การเรียนรู้แบบเสริมกำลัง : เลือกอัลกอริทึม การเรียนรู้แบบเสริมกำลังที่เหมาะสมกับปัญหาที่มีอยู่
4. ฝึกอบรมตัวแทน : ใช้อัลกอริทึมที่เลือกเพื่อฝึกอบรมตัวแทนในสภาพแวดล้อมแบบจำลอง
5. ประเมินเอเจนต์ : ทดสอบเอเจนต์ในสภาพแวดล้อมเพื่อประเมินประสิทธิภาพ
6. ปรับแต่งเอเจนต์อย่างละเอียด : ปรับแต่งพารามิเตอร์ของเอเจนต์อย่างละเอียดตามการประเมิน และทำซ้ำขั้นตอนที่ 4 และ 5 จนกว่าเอเจนต์จะทำงานได้ดี
7. ปรับใช้ตัวแทน: ปรับใช้ตัวแทนที่ผ่านการฝึกอบรมในสภาพแวดล้อมจริงเพื่อดำเนินการ

การเรียนรู้แบบเสริมกำลังสามารถนำไปใช้กับงานที่หลากหลาย เช่น การควบคุมหุ่นยนต์ การปรับกระบวนการผลิตให้เหมาะสม และการเล่นเกม ประสบความสำเร็จในการแสดงความสามารถเหนือมนุษย์ในบางงาน เช่น การเล่นเกมหมากรุกและโกะ

ความท้าทายอย่างหนึ่งในการเรียนรู้แบบเสริมกำลังคือการสร้างสมดุลระหว่างการสำรวจ ซึ่งเจ้าหน้าที่พยายามดำเนินการใหม่ๆ เพื่อเรียนรู้เพิ่มเติมเกี่ยวกับสภาพแวดล้อม กับการแสวงหาประโยชน์ ซึ่งเจ้าหน้าที่ใช้สิ่งที่ได้เรียนรู้เพื่อเพิ่มรางวัลสูงสุด การค้นหาสมดุลที่เหมาะสมระหว่างการสำรวจและการแสวงหาประโยชน์เป็นสิ่งสำคัญสำหรับเจ้าหน้าที่ในการเรียนรู้อย่างมีประสิทธิภาพ

ความท้าทายอีกประการหนึ่งคือการจัดการกับรางวัลที่ล่าช้า ซึ่งรางวัลสำหรับการกระทำอาจไม่ได้รับจนกว่าจะผ่านไปหลายขั้นตอนในภายหลัง ซึ่งอาจทำให้ตัวแทนเรียนรู้คุณค่าของการกระทำต่างๆ ได้ยาก และอาจนำไปสู่นโยบายที่ไม่เหมาะสม

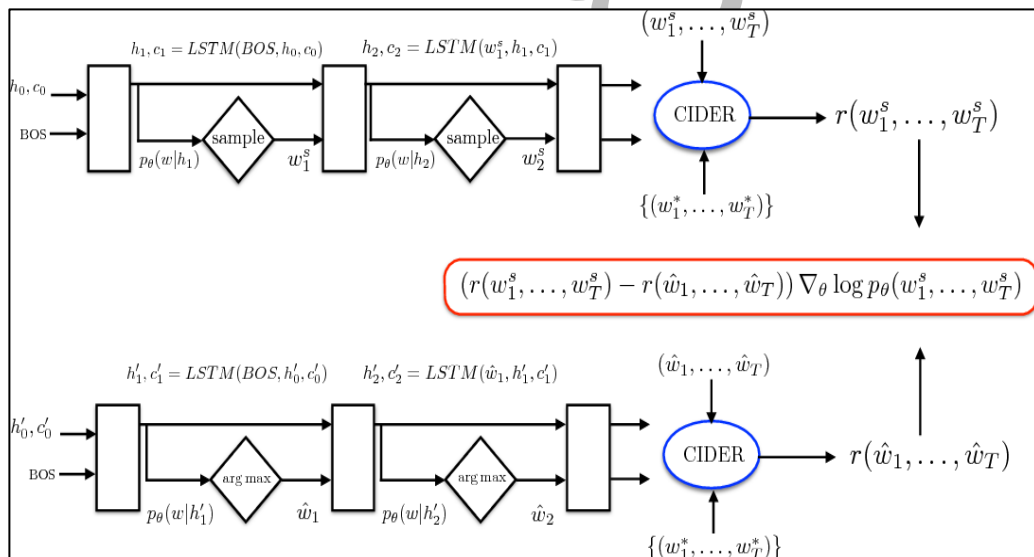
โดยรวมแล้ว การเรียนรู้แบบเสริมกำลังเป็นเครื่องมือที่ทรงพลังสำหรับการแก้ปัญหาที่หลากหลาย แต่อาจเป็นเรื่องยากที่จะนำไปใช้จริง เนื่องจากความซับซ้อนของกระบวนการเรียนรู้และความจำเป็นในการสร้างสมดุลระหว่างการสำรวจและการแสวงหาผลประโยชน์

2.16 การฝึกลำดับความสำคัญด้วยตัวเอง

การฝึกอบรมลำดับความสำคัญด้วยตนเอง [1] เป็นวิธีการฝึกอบรมแบบจำลองลำดับต่อลำดับ เช่น แบบจำลองการแปลด้วยเครื่อง โดยใช้การเรียนรู้แบบเสริมกำลัง เป้าหมายของการฝึกอบรมลำดับความสำคัญด้วยตนเองคือการปรับปรุงคุณภาพของผลลัพธ์ของแบบจำลองโดยใช้อัลกอริธึมการเรียนรู้แบบเสริมกำลังเพื่อเพิ่มประสิทธิภาพการทำงานของแบบจำลองในงานเฉพาะ

ในการฝึกอบรมลำดับความสำคัญด้วยตนเองโมเดลจะถือว่าเป็นตัวแทนที่ดำเนินการ (สร้างลำดับของคำ) และได้รับรางวัลตามคุณภาพของผลลัพธ์ โดยทั่วไป รางวัลจะคำนวณโดยใช้ฟังก์ชันการให้รางวัลที่กำหนดไว้ล่วงหน้า ซึ่งจะให้รางวัลที่สูงขึ้นสำหรับเอาต์พุตที่ใกล้เคียงกับลำดับเป้าหมาย และรางวัลที่ต่ำกว่าสำหรับเอาต์พุตที่อยู่ห่างออกไป

จากนั้น โมเดลจะได้รับการฝึกโดยใช้วิธีการไล่ระดับนโยบาย ซึ่งจะปรับพารามิเตอร์ของโมเดลเพื่อเพิ่มรางวัลที่คาดหวังให้ได้สูงสุด ในระหว่างการฝึกอบรม แบบจำลองจะสร้างลำดับของคำ และรางวัลจะใช้เพื่ออัปเดตพารามิเตอร์ของแบบจำลองในลักษณะที่จะปรับปรุงความน่าจะเป็นในการสร้างลำดับที่คล้ายคลึงกันในอนาคต แสดงกระบวนการทำงานของการฝึกอบรมลำดับความสำคัญด้วยตนเอง ดังรูปที่ 2.33



รูปที่ 2.33 กระบวนการทำงานของการฝึกอบรมลำดับความสำคัญด้วยตนเอง

จากรูปที่ 2.33 สามารถอธิบายการทำงานของการทำงานของการฝึกอบรมลำดับความสำคัญด้วยตนเองได้ดังนี้

1. ฝึกโมเดลคำบรรยายภาพล่วงหน้าโดยใช้การเรียนรู้แบบมีผู้สอน โดยใช้ชุดข้อมูลของภาพและคำบรรยายที่มนุษย์อธิบายประกอบที่เกี่ยวข้อง
2. ระหว่างการฝึก ใช้โมเดลที่ฝึกไว้ล่วงหน้าเพื่อสร้างคำบรรยายสำหรับภาพที่กำหนด
3. ใช้การคาดคะเนของโมเดลเองเป็นรางวัลสำหรับการฝึกอบรม แทนที่จะใช้คำอธิบายประกอบโดยมนุษย์ รางวัลจะคำนวณโดยใช้ความเป็นไปได้ของคำอธิบายภาพที่สร้างขึ้น ซึ่งประเมินโดยใช้แบบจำลองที่ได้รับการฝึกอบรมล่วงหน้า
4. ใช้การเรียนรู้แบบเสริมกำลังเพื่อปรับพารามิเตอร์ของแบบจำลองเพื่อเพิ่มรางวัลที่คาดหวังให้ได้สูงสุด
5. ทำซ้ำขั้นตอนที่ 2-4 ซ้ำหลายๆ ครั้ง โดยอัปเดตพารามิเตอร์ของโมเดลในแต่ละขั้นตอน
6. เมื่อสิ้นสุดการฝึกอบรม โมเดลสุดท้ายควรได้เรียนรู้วิธีสร้างคำบรรยายที่มีแนวโน้มว่าจะตรงกับคำอธิบายประกอบที่มนุษย์สร้างขึ้น เนื่องจากโมเดลได้รับการฝึกอบรมโดยใช้การคาดคะเนของตัวเองเป็นรางวัล
7. โมเดลสุดท้ายสามารถใช้เพื่อสร้างคำบรรยายสำหรับรูปภาพใหม่ได้

ข้อดีอย่างหนึ่งของการฝึกอบรมลำดับความสำคัญด้วยตนเองคือช่วยให้โมเดลได้รับการฝึกอบรมโดยตรงเกี่ยวกับงานที่จะใช้ แทนที่จะอาศัยการเรียนรู้ภายใต้การดูแลจากชุดข้อมูลขนาดใหญ่ของการแปลที่มนุษย์สร้างขึ้น สิ่งนี้สามารถทำให้โมเดลมีประสิทธิภาพมากขึ้นและสามารถสรุปอินพุตใหม่ได้ดีขึ้น

อย่างไรก็ตามการฝึกอบรมลำดับความสำคัญด้วยตนเองอาจเป็นสิ่งที่ท้าทายในการนำไปใช้จริง เนื่องจากต้องมีการออกแบบฟังก์ชันการให้รางวัลอย่างรอบคอบและความสามารถในการสุ่มตัวอย่างจากการกระจายเอาต์พุตของแบบจำลอง นอกจากนี้ยังต้องมีการคำนวณอย่างเข้มข้นเนื่องจากต้องสร้างเอาต์พุตหลายรายการสำหรับแต่ละอินพุตระหว่างการฝึกอบรม

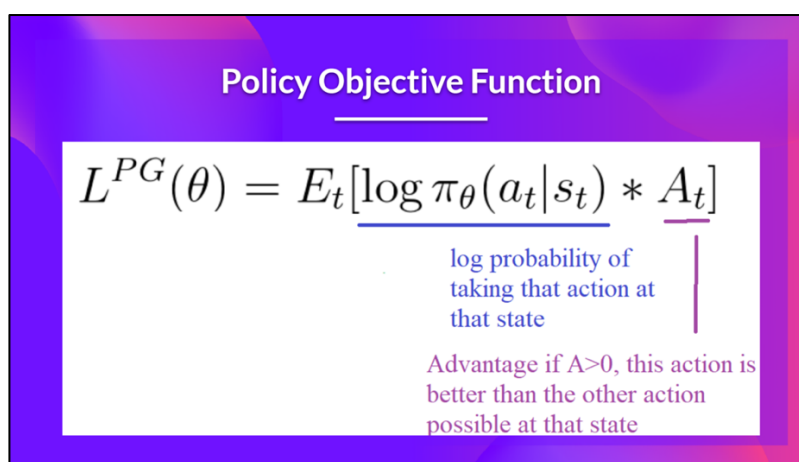
2.17 การเพิ่มประสิทธิภาพนโยบายใกล้เคียง

การเพิ่มประสิทธิภาพนโยบายใกล้เคียงเป็นอัลกอริทึมสำหรับการเรียนรู้แบบเสริมกำลัง ซึ่งเป็นประเภทหนึ่งของการเรียนรู้ของเครื่องที่เกี่ยวข้องกับตัวแทนการฝึกอบรมในการตัดสินใจในสภาพแวดล้อมเพื่อเพิ่มรางวัลสูงสุดการเพิ่มประสิทธิภาพนโยบายใกล้เคียงเป็นอัลกอริทึมตามนโยบาย หมายความว่าใช้ข้อมูลเดียวกันเพื่อปรับปรุงนโยบายและประเมินนโยบาย สิ่งนี้ตรงกันข้ามกับอัลกอริทึมนอกนโยบายซึ่งใช้ข้อมูลแยกต่างหากสำหรับการประเมินและปรับปรุง

คุณสมบัติหลักประการหนึ่งของการเพิ่มประสิทธิภาพนโยบายใกล้เคียงคือใช้วิธีภูมิภาคที่เชื่อถือในการอัปเดตนโยบาย ซึ่งจะช่วยป้องกันการเปลี่ยนแปลงขนาดใหญ่ที่อาจทำให้ไม่เสถียรใน

นโยบาย นอกจากนี้ยังใช้ฟังก์ชันการตัดเพื่อให้แน่ใจว่าการเปลี่ยนแปลงในนโยบายอยู่ในช่วงที่กำหนด ซึ่งช่วยให้การฝึกอบรมมีเสถียรภาพมากขึ้น

การเพิ่มประสิทธิภาพนโยบายใกล้เคียงได้รับการพิสูจน์แล้วว่ามีประสิทธิภาพในงานที่หลากหลายและใช้กันอย่างแพร่หลายในการเรียนรู้แบบเสริมกำลัง นอกจากนี้ยังค่อนข้างง่ายในการติดตั้งและมีไฮเปอร์พารามิเตอร์ค่อนข้างน้อย ซึ่งทำให้เป็นตัวเลือกยอดนิยมสำหรับนักวิจัยและผู้ปฏิบัติงาน แสดงกระบวนการของการเพิ่มประสิทธิภาพนโยบายใกล้เคียงดังรูปที่ 2.34



The diagram shows the Policy Objective Function equation:
$$L^{PG}(\theta) = E_t[\log \pi_{\theta}(a_t|s_t) * A_t]$$
 The term $\log \pi_{\theta}(a_t|s_t)$ is annotated as 'log probability of taking that action at that state'. The term A_t is annotated as 'Advantage if A>0, this action is better than the other action possible at that state'.

รูปที่ 2.34 กระบวนการของการเพิ่มประสิทธิภาพนโยบายใกล้เคียง

จากรูปที่ 2.34 สามารถอธิบายกระบวนการของการเพิ่มประสิทธิภาพนโยบายใกล้เคียงได้ดังนี้

1. เริ่มต้นเครือข่ายนโยบายของตัวแทน RL ด้วยน้ำหนักแบบสุ่ม
2. รวบรวมข้อมูลเป็นชุดโดยการโต้ตอบกับสภาพแวดล้อม โดยใช้นโยบายปัจจุบันเพื่อเลือกการดำเนินการ
3. คำนวณความน่าจะเป็นของแต่ละการดำเนินการในชุดงานภายใต้้นโยบายปัจจุบัน
4. คำนวณการสูญเสียนโยบาย ซึ่งเป็นค่าลบของความเป็นไปได้ของการดำเนินการในชุดงานคุณด้วยฟังก์ชันความได้เปรียบที่วัดมูลค่าสัมพัทธ์ของการดำเนินการแต่ละรายการ
5. ใช้อัลกอริทึมการปรับให้เหมาะสมใกล้เคียงเพื่อปรับน้ำหนักของเครือข่ายนโยบายเพื่อลดการสูญเสียนโยบาย อัลกอริทึมการปรับให้เหมาะสมใกล้เคียงสามารถนำไปใช้ได้โดยใช้วิธีการระดับความน่าเชื่อถือหรือวิธีการตัด
6. ทำซ้ำขั้นตอนที่ 2-5 ซ้ำหลายๆ ครั้ง โดยอัปเดตเครือข่ายนโยบายในแต่ละขั้นตอน

7. เมื่อสิ้นสุดการฝึกอบรม เครือข่ายนโยบายของตัวแทนควรได้เรียนรู้นโยบายที่เพิ่มผลตอบแทนที่คาดหวังให้ได้สูงสุดเมื่อเวลาผ่านไป
8. นโยบายขั้นสุดท้ายสามารถใช้เพื่อควบคุมพฤติกรรมของตัวแทนในสภาพแวดล้อม
- 9.

การเพิ่มเติมประสิทธิภาพนโยบายใกล้เคียง ถูกนำมาใช้ในการแก้ปัญหาทางงานต่างๆ รวมถึงการเล่นเกม Atari การควบคุมหุ่นยนต์ และการปรับปรุงปฏิกริยาเคมีให้เหมาะสม นอกจากนี้ยังถูกนำมาใช้ในแอปพลิเคชันในโลกแห่งความเป็นจริง เช่น การควบคุมรถยนต์ที่ขับเคลื่อนด้วยตัวเองและการปรับการใช้พลังงานให้เหมาะสมในศูนย์ข้อมูล

2.18 เกณฑ์ที่ใช้ในการวัดประสิทธิภาพของโมเดลของการใส่คำบรรยายภาพ (Model Evaluation of Caption Generation)

เกณฑ์ที่ใช้ในการวัดประสิทธิภาพของโมเดลของการใส่คำบรรยายภาพมีหลายวิธีในการประเมินประสิทธิภาพของโมเดลการใส่คำบรรยายโดยขึ้นอยู่กับงานเฉพาะและเป้าหมายของโมเดลต่อไปนี้เป็นวิธีการประเมินทั่วไปบางประการ

1. คะแนน BLEU (Bilingual Evaluation Understudy) [7] : นี่เป็นตัวชี้วัดทั่วไปสำหรับการประเมินคุณภาพของข้อความที่เครื่องสร้างขึ้น โดยจะเปรียบเทียบข้อความที่สร้างขึ้นกับการแปลอ้างอิงและวัดการทับซ้อนของ n-gram ระหว่างทั้งสอง คะแนน BLEU ที่สูงขึ้นบ่งชี้ว่าข้อความที่สร้างขึ้นมีความคล้ายคลึงกับการแปลอ้างอิงมากกว่า แสดงรายละเอียดของการประเมินคุณภาพของ BLEU ดังรูปที่ 2.35

$$BLEU = BP * \exp\left(\sum_{k=1}^n w_k \log(p_k)\right)$$

$$BP = e^{\min\left(1 - \frac{\text{len}(\text{reference})}{\text{len}(\text{prediction})}, 0\right)}$$

รูปที่ 2.35 รายละเอียดของการประเมินคุณภาพของ BLEU

จากรูปที่ 2.35 สามารถอธิบายรายละเอียดของการประเมินคุณภาพของ BLEU ได้ดังนี้

- BP คือบทลงโทษความสั้น ซึ่งเป็นบทลงโทษสำหรับข้อความที่แปลด้วยคอมพิวเตอร์ที่สั้นกว่าการแปลอ้างอิง
- $w(k)$ เป็นปัจจัยถ่วงน้ำหนักสำหรับแต่ละระดับของ n -กรัม โดยที่ n อยู่ในช่วงตั้งแต่ 1 ถึง 4 ปัจจัยถ่วงน้ำหนักถูกใช้เพื่อให้ความสำคัญมากขึ้นกับ n -กรัมระดับสูง (เช่น 4 กรัม) มากกว่า n ระดับล่าง - กรัม (เช่น 1 กรัม)
- $p(k)$ คือความแม่นยำ n -gram ซึ่งเป็นจำนวน n -grams ในข้อความที่แปลด้วยเครื่องซึ่งตรงกับ n -grams ในการแปลอ้างอิง หาคำนวณด้วยจำนวน n -grams ทั้งหมดในการแปลด้วยเครื่องข้อความ.

คะแนน BLEU อยู่ระหว่าง 0 ถึง 1 โดย 1 แสดงถึงการจับคู่ที่สมบูรณ์แบบระหว่างข้อความที่แปลด้วยคอมพิวเตอร์และการแปลอ้างอิง โปรดทราบว่ากระดาษ BLEU ดั้งเดิมเสนอให้ใช้ 1-gram ถึง 4-gram แต่บางงานเพิ่มเติมเสนอให้ใช้ถึง 6-grams

2. คะแนน CIDER (Consensus-based Image Description Evaluation) [2] : ตัวชี้วัดนี้ได้รับการออกแบบมาโดยเฉพาะสำหรับการประเมินรูปแบบคำบรรยายภาพ โดยจะเปรียบเทียบคำบรรยายที่สร้างขึ้นกับชุดของคำบรรยายอ้างอิง และคำนวณคะแนนตามความเห็นพ้องของคำบรรยายอ้างอิง แสดงรายละเอียดของการประเมินคุณภาพของ CIDER ดังรูปที่ 2.36

$$\text{CIDEr} = \sum_i (1/n) * \text{IDi} * \log(c/r)$$

รูปที่ 2.36 รายละเอียดของการประเมินคุณภาพของ CIDER

จากรูปที่ 2.36 สามารถอธิบายรายละเอียดของการประเมินคุณภาพ CIDER คือ เป็นการวัดที่ใช้ในการประเมินคุณภาพของคำบรรยายภาพที่สร้างโดยอัลกอริทึมการเรียนรู้ของเครื่อง ตัวชี้วัด

ไฮเดอร์เปรียบเทียบคำบรรยายที่สร้างขึ้นกับชุดของคำบรรยายอ้างอิงสำหรับรูปภาพเดียวกัน และคำนวณคะแนนตามความคล้ายคลึงกันของคำบรรยายที่สร้างขึ้นกับคำบรรยายอ้างอิง

n = จำนวนคำบรรยายอ้างอิง

ID_i = จุดตัดของคำบรรยายที่สร้างขึ้นและคำบรรยายอ้างอิง (วัดจากจำนวน n -กรัมที่ไม่ซ้ำกัน)

c = จำนวนรวมของ n -grams ที่ไม่ซ้ำกันในคำบรรยายที่สร้างขึ้น

r = จำนวนรวมของ n -grams ที่ไม่ซ้ำกันในคำบรรยายอ้างอิง

เมตริกไฮเดอร์ใช้การผสมผสานระหว่างความแม่นยำและการเรียกคืนเพื่อคำนวณคะแนน โดยคะแนนที่สูงกว่าจะบ่งบอกถึงการจับคู่ที่ดีกว่าระหว่างคำบรรยายที่สร้างขึ้นและคำบรรยายอ้างอิง

3. การประเมินโดยมนุษย์ : อีกทางเลือกหนึ่งคือการให้ผู้ประเมินที่เป็นบุคคลประเมินคุณภาพของคำบรรยายที่สร้างขึ้น ซึ่งสามารถทำได้หลายวิธี เช่น ให้ผู้ประเมินให้คะแนนคำบรรยายตามขนาดหรือให้พวกเขาเลือกคำบรรยายที่ดีที่สุดจากชุดตัวเลือก การประเมินโดยมนุษย์สามารถให้ข้อเสนอแนะที่เหมาะสมและละเอียดยิ่งขึ้นเกี่ยวกับจุดแข็งและจุดอ่อนของแบบจำลอง แต่อาจใช้เวลาและค่าใช้จ่ายสูงในการดำเนินการ

สิ่งสำคัญคือต้องระลึกไว้เสมอว่าไม่มีเมตริกการประเมินใดที่สมบูรณ์แบบ และมักจะเป็นประโยชน์หากใช้วิธีเหล่านี้ร่วมกันเพื่อทำความเข้าใจประสิทธิภาพของโมเดลอย่างครอบคลุมมากขึ้น

บทที่ 3

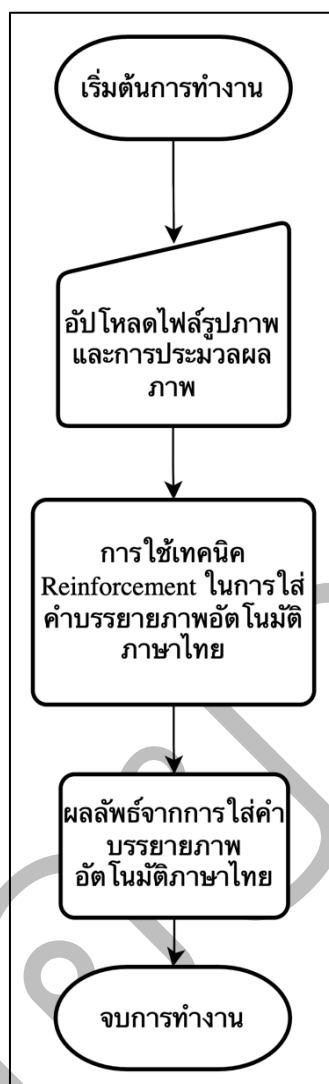
วิธีการดำเนินงาน

ในการจัดทำปฏิญานิพนธ์ฉบับนี้หลังจากผู้จัดทำได้ทำการศึกษารายละเอียด และรวบรวมเนื้อหาข้อมูลต่าง ๆ ที่เกี่ยวข้องแล้ว ได้ดำเนินการออกแบบ และพัฒนาระบบซึ่งมีขั้นตอนในการดำเนินงานดังต่อไปนี้

- 3.1 การใช้งานเทคนิค Reinforcement ในการใส่คำบรรยายภาพอัตโนมัติภาษาไทย
- 3.2 อัปโหลดไฟล์รูปภาพและการประมวลผลภาพ
- 3.3 การใส่คำบรรยายภาพภาษาไทยด้วยเทคนิคฝึกลำดับความสำคัญด้วยตนเองโดยใช้โครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์
- 3.4 ผลลัพธ์ของการใส่คำบรรยายภาพอัตโนมัติภาษาไทย

3.1 การใช้งานเทคนิค Reinforcement ในการใส่คำบรรยายภาพอัตโนมัติภาษาไทย

จากปัญหาที่กล่าวมาในบทก่อนหน้านี้ทางคณะผู้จัดทำจึงได้มีแนวคิดในการแก้ไขปัญหาดังกล่าวโดยใช้การเข้ารหัสถอดรหัส เพื่อช่วยในการใส่คำบรรยายภาพที่เป็นภาษาไทย ซึ่งผู้ใช้นั้นสามารถทราบคำบรรยายภาพที่เป็นภาษาไทยได้อย่างถูกต้อง โดยชุดข้อมูลที่น่าสนใจในเรียนรู้ของโมเดลถูกรวบรวมจากชุดข้อมูล MSCOCO ซึ่งภาพรวมของขั้นตอนการดำเนินงานในการจัดทำการใช้เทคนิค Reinforcement ในการใส่คำบรรยายภาพอัตโนมัติภาษาไทยแสดงโดยใช้ผังงาน (Flow Chart) ดังรูปที่ 3.1

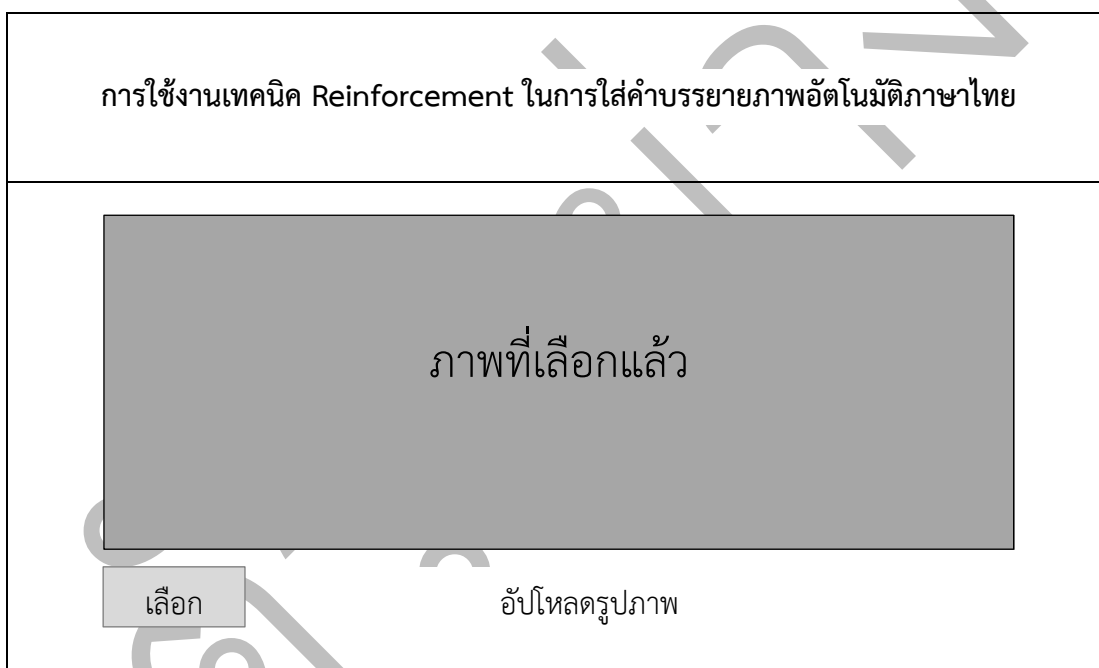


รูปที่ 3.1 ผังงานโดยรวมของระบบ

จากรูปที่ 3.1 สามารถอธิบายขั้นตอนการทำงานโดยรวมของระบบได้ดังนี้ เมื่อเริ่มต้นการทำงานจะต้องอัปโหลดรูปที่มีนามสกุล JPEG JPG และ PNG หลังจากนั้นจะทำการประมวลผลภาพด้วยวิธีการสร้างรูปร่างใหม่เพื่อให้ได้รูปร่างของภาพที่สอดคล้องกับรูปร่างของโมเดลที่จะทำการใช้เทคนิค Reinforcement ในการใส่คำบรรยายภาพอัตโนมัติภาษาไทย เมื่อกระบวนการของขั้นตอนที่ผ่านมาเสร็จสิ้น จะได้ผลลัพธ์จากการจำแนกกลุ่มจะแสดงผ่านหน้าจอโปรแกรมให้ผู้ใช้ทราบและจบการทำงาน

3.2 การอัปโหลดรูปภาพและการประมวลผลภาพ

ผู้ใช้งานโปรแกรมสามารถอัปโหลดได้เฉพาะไฟล์ภาพที่อัปโหลดจะต้องมีนามสกุลเป็น JPEG JPG และ PNG เท่านั้น เพราะไฟล์ JPEG JPG และ PNG จะเก็บรายละเอียดของภาพ และสีได้มากถึง 16 ล้านสี ซึ่งเหมาะในการนำมาใช้กับภาพที่มีสีสัน รายละเอียดจำนวนมากของภาพ และภาพที่ถูกอัปโหลด เมื่อผู้ใช้ได้อัปโหลดไฟล์ภาพที่มีนามสกุลถูกต้องตามเงื่อนไข แล้วจึงเข้าสู่ขั้นตอนนี้ คือ ขั้นตอนการประมวลผลภาพโดยทำการแปลงค่าสีให้อยู่ในระบบสีอาจีบี เพื่อให้สอดคล้องกับโมเดลที่จะทำการเรียนรู้ หลังจากนั้นจะถูกย่อขนาดให้เหลือ 384x384 เพราะขนาดของภาพจะได้สอดคล้องกับขนาดของโมเดลที่จะทำการเรียนรู้ แสดงหน้าต่างโปรแกรมอัปโหลดรูปภาพ ได้ดังรูปที่ 3.2 แสดงรูปที่ถูกย่อขนาดของภาพให้เหลือ 384x384 ได้ดังรูปที่ 3.3



รูปที่ 3.2 ตัวอย่างหน้าต่างโปรแกรมในการอัปโหลดภาพ

จากรูปที่ 3.2 สามารถอธิบายตัวอย่างหน้าต่างโปรแกรมในการอัปโหลดรูปภาพ โดยรูปที่จะสามารถอัปโหลดได้นั้นจะต้องมีนามสกุลเป็น JPEG JPG และ PNG เท่านั้น

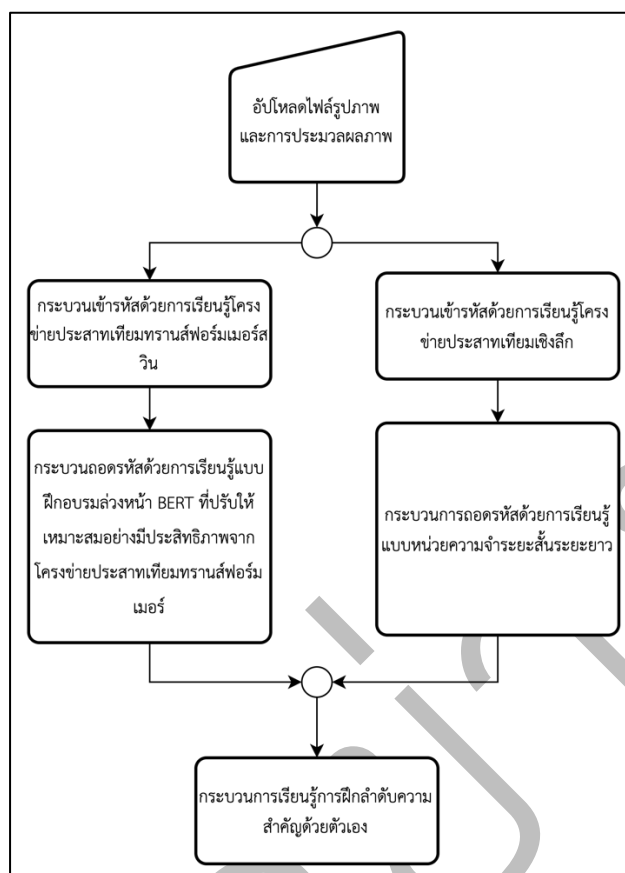


รูปที่ 3.3 ตัวอย่างการย่อขนาดของภาพให้เหลือ 384x384

จากรูปที่ 3.3 สามารถอธิบายตัวอย่างในการย่อขนาดรูปที่มีขนาด 500x500 ให้เป็นขนาด 384x384 เพื่อให้ได้ขนาดของภาพจะที่สอดคล้องกับขนาดของโมเดลที่จะทำการเรียนรู้

3.3 การใช้เทคนิค Reinforcement ในการใส่คำบรรยายภาพอัตโนมัติภาษาไทย

การใช้เทคนิค Reinforcement ในการใส่คำบรรยายภาพอัตโนมัติภาษาไทย เป็นขั้นตอนที่มีกระบวนการวิเคราะห์ภายในอย่างลึกซึ้งด้วยอัลกอริทึมที่ซับซ้อนโดยแบ่งการทำงาน 2 แบบ คือ แบบแรกใช้ขั้นตอนกระบวนการเข้ารหัสด้วยการเรียนรู้โครงข่ายประสาทเทียมเชิงลึกและขั้นตอนกระบวนการถอดรหัสด้วยการเรียนรู้แบบหน่วยความจำระยะสั้นระยะยาว แบบที่สองใช้ขั้นตอนกระบวนการเข้ารหัสด้วยการเรียนรู้โครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์สวินและขั้นตอนกระบวนการเรียนรู้แบบฝึกรวมล่วงหน้า BERT ที่ปรับให้เหมาะสมอย่างมีประสิทธิภาพจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์หลังจากนั้นก็จะใช้กระบวนการเรียนรู้การฝึกลำดับความสำคัญด้วยตนเองในการปรับการใส่คำบรรยายภาพอัตโนมัติภาษาไทยให้เป็นมาตรฐาน ซึ่งทำการแสดงผังงานดังรูปที่ 3.4



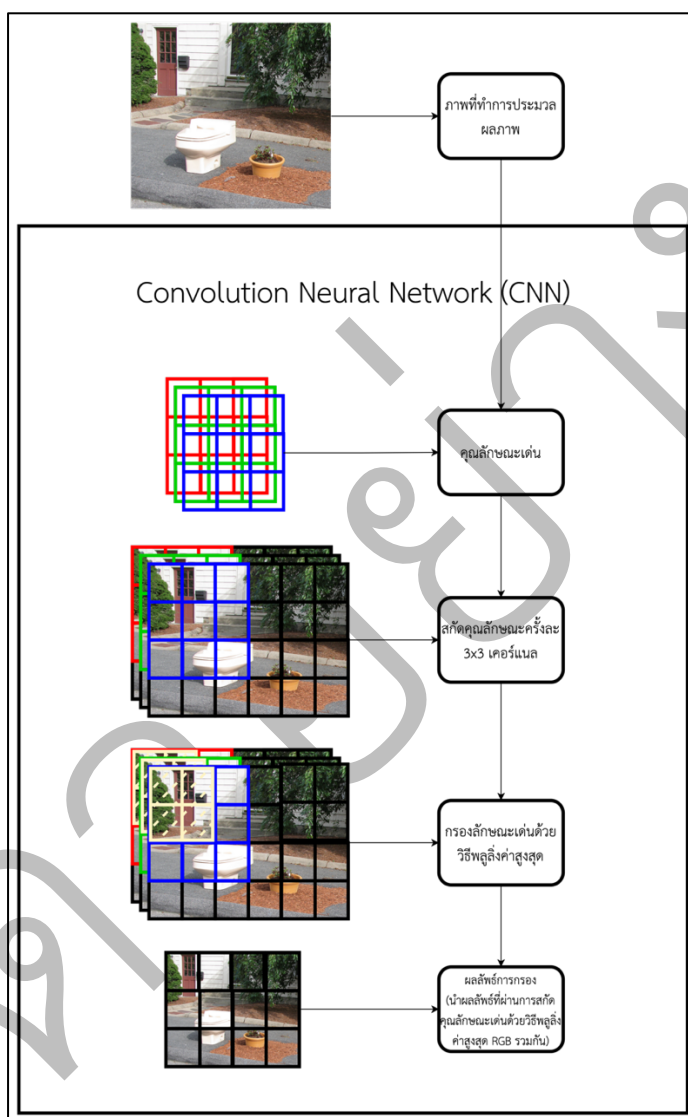
รูปที่ 3.4 ผังงานของการใช้เทคนิค Reinforcement ในการใส่คำบรรยายภาพอัตโนมัติภาษาไทย

จากรูปที่ 3.4 สามารถอธิบายการทำงานหลังจากการประมวลผลภาพออกเป็น 2 แบบ คือ แบบแรกจะใช้กระบวนการคอนโวลูชันและการเรียนรู้แบบหน่วยความจำระยะสั้นระยะยาว แบบสองจะใช้การเรียนรู้โครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์สวินและการเรียนรู้แบบฝึกรอบรณวล่งหน้า BERT ที่ปรับให้เหมาะสมอย่างมีประสิทธิภาพจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ส หลังจากนั้นก็จะใช้กระบวนการเรียนรู้การฝึกลำดับความสำคัญด้วยตนเองในการปรับการใส่คำบรรยายภาพอัตโนมัติภาษาไทยให้เป็นมาตรฐาน

3.3.1 กระบวนการเข้ารหัสด้วยการเรียนรู้โครงข่ายประสาทเทียมเชิงลึก

เป็นกระบวนการที่นำเอาภาพที่อัปโหลดไฟล์รูปและทำการประมวลผลภาพแล้วมาหาคุณลักษณะเด่นของภาพเพื่อทำการเข้ารหัส ซึ่งจะต้องสกัดคุณลักษณะเด่นหรือเลือกคุณลักษณะเด่นบนภาพ หลังจาก

นั้นคุณลักษณะเด่นที่ถูกเลือกจะผ่านการกรองด้วยวิธีฟูลลิงค่าสูงสุด เหมือนกับเป็นการย่อขนาดของภาพลง 2 เท่า แต่ยังคงลักษณะเด่นเดิมไว้ ซึ่งถ้าคุณลักษณะเด่นมีจำนวนมากข้อมูลในการวิเคราะห์ก็จะมีรายละเอียดมากยิ่งขึ้น แสดงกระบวนการเข้ารหัสด้วยการเรียนรู้โครงข่ายประสาทเทียมเชิงลึกแสดงดังรูปที่ 3.5

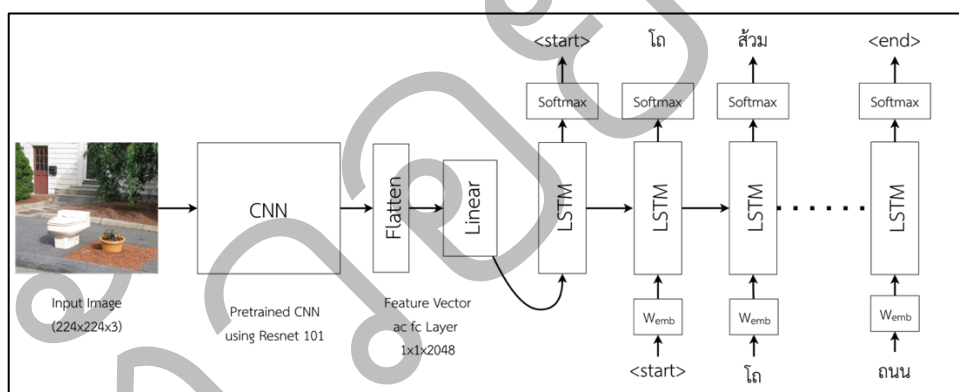


รูปที่ 3.5 กระบวนการเข้ารหัสด้วยการเรียนรู้โครงข่ายประสาทเทียมเชิงลึก

จากรูปที่ 3.5 สามารถอธิบายกระบวนการเข้ารหัสด้วยการเรียนรู้โครงข่ายประสาทเทียมเชิงลึก ซึ่งจะทำให้การสกัดคุณลักษณะเด่น ซึ่งในตัวกรองคุณลักษณะเด่นนั้นจะมีขนาดของตัวกรอง 3×3 เคอร์เนล โดยจะกรอกทั้ง 3 ขานแนลสีได้แก่ สีแดง สีเขียวและสีน้ำเงินเพื่อให้ได้คุณลักษณะเด่นของภาพออกมาในแต่ละสี ต่อมาทำการกรองลักษณะเด่นด้วยวิธีพูลลิงค่าสูงสุดขนาดของตัวกรอง 2×2 เคอร์เนล เพื่อทำการลดขนาดภาพลงแต่ยังคงคุณลักษณะเด่นเดิมอยู่ สุดท้ายก็จะได้ผลลัพธ์จากการกรองคุณลักษณะของภาพเพื่อส่งไปให้กระบวนการเรียนรู้แบบหน่วยความจำระยะสั้นระยะยาวทำการเรียนรู้ลักษณะของตัวเลขที่ผ่านการกรองลักษณะเด่น

3.3.2 กระบวนการถอดรหัสด้วยการเรียนรู้แบบหน่วยความจำระยะสั้นระยะยาว

เป็นกระบวนการที่จะนำคุณลักษณะเด่นที่เข้ารหัสจากกระบวนการเข้ารหัสด้วยการเข้ารหัสด้วยการเรียนรู้โครงข่ายประสาทเทียมเชิงลึก เข้ามาเรียนรู้คุณลักษณะเด่นที่ถูกเข้ารหัส แล้วทำการตัดสินใจในการใส่คำบรรยายภาพอัตโนมัติภาษาไทย แสดงกระบวนการถอดรหัสด้วยการเรียนรู้แบบหน่วยความจำระยะสั้นระยะยาวแสดงดังรูปที่ 3.6

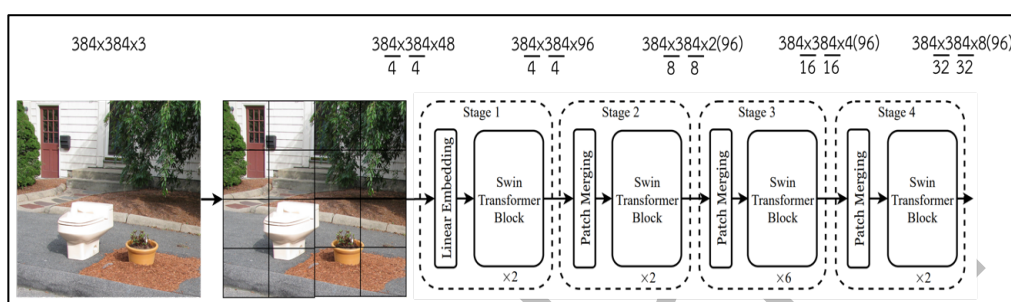


รูปที่ 3.6 กระบวนการถอดรหัสด้วยการเรียนรู้แบบหน่วยความจำระยะสั้นระยะยาว

จากรูปที่ 3.6 สามารถอธิบายกระบวนการถอดรหัสด้วยการเรียนรู้แบบหน่วยความจำระยะสั้นระยะยาว ซึ่งจะทำต่อจากการบวนการเข้ารหัสด้วยการเรียนรู้โครงข่ายประสาทเทียมเชิงลึก ที่ทำการสกัดคุณลักษณะเด่นมาให้ หลังจากหน่วยความจำระยะสั้นระยะยาวจะทำการเรียนรู้ลักษณะเด่นของภาพที่ผ่านการเข้ารหัสแล้วทำการสร้างคำบรรยายภาพลำดับเริ่มต้นขึ้นมาหลังจากนั้นก็ส่งทั้งลักษณะเด่นภาพและคำบรรยายก่อนหน้าให้กับลำดับถัดไปเพื่อทำการถอดรหัสสำหรับสร้างคำบรรยายภาพออกมา

3.3.3 กระบวนเข้ารหัสด้วยการเรียนรู้โครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์สวิน

เป็นกระบวนการที่นำเอาภาพที่อัปโหลดไฟล์รูปและทำการประมวลผลภาพแล้วมาหาคุณลักษณะของภาพโดยจำกัดความสนใจภายในพื้นที่เพื่อทำการเข้ารหัส แต่คงไว้ซึ่งการเชื่อมต่อข้ามหน้าต่าง แสดงกระบวนการเข้ารหัสด้วยการเรียนรู้โครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์สวินแสดงดังรูปที่ 3.7

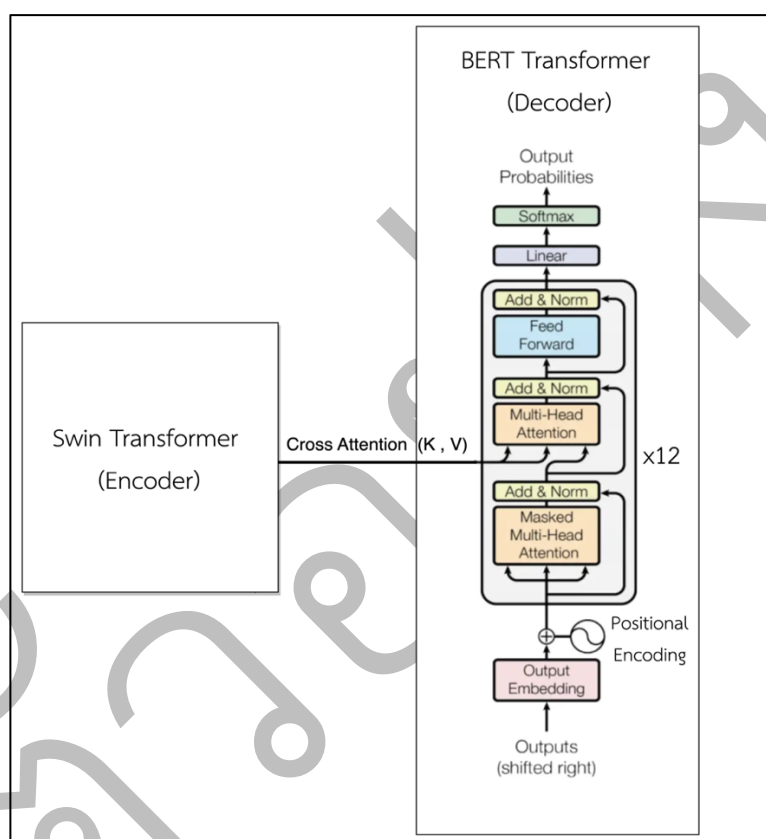


รูปที่ 3.7 กระบวนเข้ารหัสด้วยการเรียนรู้โครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์สวิน

จากรูปที่ 3.7 สามารถอธิบายกระบวนการเข้ารหัสด้วยการเรียนรู้โครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์สวิน ซึ่งจะหาคุณลักษณะของภาพโดยจำกัดความสนใจภายในพื้นที่เพื่อทำการเข้ารหัสภาพ โดยทำการแยกภาพระบบสี่แบบออกเป็นแพตช์ที่ไม่ทับซ้อนกัน ในแต่ละแพตช์จะถือว่าเป็นลำดับของโทเค็น และคุณลักษณะของแพตช์เริ่มต้นจะมีการใช้ขนาดแพตช์ 4×4 ดังนั้นมิติคุณลักษณะของแต่ละแพตช์คือ $4 \times 4 \times 3 = 48$ หลังจากนั้นนำข้อมูลคุณลักษณะของภาพมาทำการฉายไปยังมิติ $C = 96$ ซึ่งในบล็อกทรานส์ฟอร์มเมอร์สหลายๆ บล็อกจะมีการคำนวณ Shifted Window Based Self-Attention ก็คือการคำนวณหาความสำคัญของแต่ละหน้าต่างตามหน้าต่างที่เลื่อนโดยจะถูกนำไปใช้ในแต่ละแพตช์ โดยขนาดหน้าต่างจะค่อยๆ เพิ่มขึ้นในแต่ละ State ตามลำดับ

3.3.4 กระบวนการเรียนรู้แบบฝึกอบรมล่งหน้า BERT ที่ปรับให้เหมาะสมอย่างมีประสิทธิภาพจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์

เป็นกระบวนการที่จะนำคุณลักษณะเด่นที่เข้ารหัสจากกระบวนการเข้ารหัสด้วยการเรียนรู้โครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์สวิน เข้ามาเรียนรู้คุณลักษณะเด่นที่ถูกเข้ารหัส แล้วทำการตัดสินใจในการใส่คำบรรยายภาพอัตโนมัติภาษาไทย แสดงกระบวนการเรียนรู้แบบฝึกอบรมล่งหน้า BERT ที่ปรับให้เหมาะสมอย่างมีประสิทธิภาพจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์แสดงดังรูปที่ 3.8



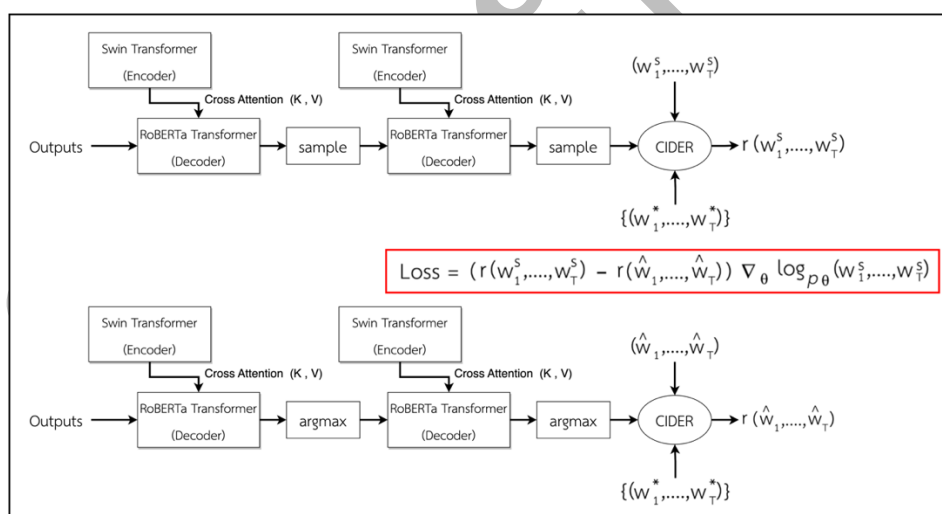
รูปที่ 3.8 กระบวนการเรียนรู้แบบฝึกอบรมล่งหน้า BERT ที่ปรับให้เหมาะสมอย่างมีประสิทธิภาพจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์

จากรูปที่ 3.8 สามารถอธิบายกระบวนการถอดรหัสด้วยการแทนตัวเข้ารหัสแบบสองทิศทางจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ ซึ่งจะทำต่อจากการกระบวนการเข้ารหัสด้วยการเรียนรู้โครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์สวิน โดยจะใช้การทำ Cross Attention เพื่อหาความสัมพันธ์ตามบริบท

ระหว่างคำในประโยคนั้น โดยนำข้อมูลที่ผ่านมาผ่านการเข้ารหัสเพื่อหาคุณลักษณะของภาพโดยจำกัดความสนใจภายในพื้นที่แล้วนำไปเข้าร่วมกับข้อมูลที่ผ่านมาผ่านการถอดรหัสจาก Masked Multi-Head Attention หลังจากส่งไปให้ Feed Forward ทำการแปลงเวกเตอร์ความสนใจให้อยู่ในรูปแบบที่เลเยอร์เพื่อที่จะส่งข้อมูลระหว่างตัวเข้ารหัสไปยังตัวถอดรหัสได้โดยเวกเตอร์ความสนใจแต่ละตัวเป็นอิสระจากกันดังนั้นสามารถใช้การทำให้ขนานกันได้สามารถส่งค่าทั้งหมดพร้อมกันไปยังบล็อกตัวเข้ารหัสและรับชุดข้อมูลเวกเตอร์ถอดรหัส สำหรับทุกคำ ๆ พร้อมกัน แล้วก็ส่งไปยัง Softmax เพื่อทำการหาความน่าจะเป็นในการใส่คำบรรยายภาพ

3.3.5 กระบวนการเรียนรู้การฝึกลำดับความสำคัญด้วยตนเอง

เป็นกระบวนการที่จะนำคำบรรยายที่ได้จากจากกระบวนการถอดรหัสด้วยการแทนตัวเข้ารหัสแบบสองทิศทางจากโครงข่ายประสาทเทียมทรานส์ฟอร์มเมอร์ มาคำนวณลำดับความสำคัญเพื่อหาค่า Loss Function ในการปรับค่าการตัดสินใจในการใส่คำบรรยายภาพอัตโนมัติภาษาไทยให้สามารถใส่คำบรรยายภาพดีขึ้นในแต่ละรอบ แสดงกระบวนการเรียนรู้การฝึกลำดับความสำคัญด้วยตนเองดังรูปที่ 3.9



รูปที่ 3.9 กระบวนการเรียนรู้การฝึกลำดับความสำคัญด้วยตนเอง

จากรูปที่ 3.9 สามารถอธิบายการกระบวนการเรียนรู้การฝึกลำดับความสำคัญด้วยตนเอง ซึ่งจะทำได้จากกระบวนการเข้ารหัสและถอดรหัสภาพ โดยจะทำการหารางวัลจากการสร้างคำบรรยายภาพที่สร้างขึ้นแบบ Sample และรางวัลจากการจากการสร้างคำบรรยายภาพที่สร้างขึ้นแบบ Argmax โดย

คำนวณด้วยวิธีการของ CIDER score หลังจากนั้นก็จะใช้การเรียนรู้แบบเสริมกำลังเพื่อปรับพารามิเตอร์โมเดลให้ดีขึ้นซึ่งสมการที่คำนวณค่า Loss ก็จะมีอยู่ในรูปที่ 3.9

3.4 ผลลัพธ์ของการใส่คำบรรยายภาพอัตโนมัติภาษาไทย

เมื่อกระบวนการของขั้นตอนที่ผ่านมาเสร็จสิ้น จะได้ผลลัพธ์ออกมา ผลลัพธ์ที่ได้ คือคำบรรยายภาพอัตโนมัติภาษาไทย ผลลัพธ์ของคำบรรยายภาพอัตโนมัติภาษาไทยแสดงผลคำบรรยายภาพอัตโนมัติภาษาไทยบนเว็บไซต์ดังรูปที่ 3.10

| การใช้งานเทคนิค Reinforcement ในการใส่คำบรรยายภาพอัตโนมัติภาษาไทย |
|---|
|  <p>คำบรรยายภาพอัตโนมัติภาษาไทย : โถส้วมสีขาวถูกวางทิ้งไว้ข้างกระถางต้นไม้สีน้ำตาลริมถนน</p> |

รูปที่ 3.10 แสดงผลลัพธ์คำบรรยายภาพอัตโนมัติภาษาไทยบนเว็บไซต์

จากรูปที่ 3.10 สามารถอธิบายหลังจากที่ทำกระบวนการคอนโวลูชันและกระบวนการโครงข่ายประสาทเทียมเสร็จ จะได้ผลลัพธ์จากการจำแนกกลุ่มของภาพะปอดผ่านหน้าจอโปรแกรมให้ผู้ใช้ทราบและจบการทำงาน

บทที่ 4

การทดลองและผลการดำเนินงาน

ในการประเมินประสิทธิภาพการทำงานของการศึกษาการใช้งานเทคนิค Reinforcement ในการใส่คำบรรยายภาพอัตโนมัติภาษาไทย ต้องมีการทดลองและวัดผลการทดลองในกรณีต่าง ๆ ให้สอดคล้องกับขอบเขตที่ได้ระบุไว้ โดยได้ทำการทดสอบและประเมินประสิทธิภาพการทำงานของระบบตามลำดับดังนี้

- 4.1 ทดสอบความสามารถตามขอบเขตที่กำหนด
- 4.2 ความต้องการขั้นต่ำของระบบในการทดลอง
- 4.3 ทดสอบการอัปโหลดไฟล์ภาพที่มีนามสกุล JPEG JPG และ PNG
- 4.4 ทดสอบการใส่คำบรรยายภาพอัตโนมัติภาษาไทย
- 4.5 ทดสอบการใส่คำบรรยายภาพอัตโนมัติภาษาไทยครั้งละ 1 ภาพ

4.1 ทดสอบความสามารถตามขอบเขตที่กำหนด

การทดสอบความสามารถตามขอบเขตที่กำหนด เพื่อตรวจสอบและประเมินผลความสามารถว่าระบบที่พัฒนามีความสามารถทำตามขอบเขต โดยแสดงรายละเอียดของขอบเขตที่กำหนด และความสามารถของระบบพร้อมทั้งระบุหัวข้อการทดลองเพื่อแสดงผลการทดลองได้ดังตารางที่ 4.1

ตารางที่ 4.1 การทดสอบความสามารถตามขอบเขตที่กำหนด

| ขอบเขต | ความสามารถ | หมายเหตุ |
|---|------------|--------------------|
| 1.3.1.สามารถใส่คำบรรยายภาพอัตโนมัติภาษาไทย | ✓ | หัวข้อทดลองที่ 4.4 |
| 1.3.2 สามารถอัปโหลดไฟล์ภาพที่มีนามสกุล JPEG JPG และ PNG | ✓ | หัวข้อทดลองที่ 4.3 |
| 1.3.3 การใส่คำบรรยายภาพอัตโนมัติภาษาไทยครั้งละ 1 ภาพ | ✓ | หัวข้อทดลองที่ 4.5 |

จากตารางที่ 4.1 จะเห็นได้ว่า การศึกษาการใช้งานเทคนิค Reinforcement ในการใส่คำบรรยายภาพอัตโนมัติภาษาไทยสามารถทำงานได้ตามขอบเขตที่ระบุไว้ทุกประการ โดยการทดลองในแต่ละหัวข้อการทดลองหรือตารางผลการทดลองแต่ละตารางนั้นสะท้อนถึงความสามารถที่ตอบสนองต่อการทำงานตามขอบเขตที่กำหนดไว้

4.2 ความต้องการขั้นต่ำของระบบในการทดลอง

การทดลองนี้ได้ตั้งค่าการทดลองต่าง ๆ เพื่อใช้สำหรับการทดลองและประเมินประสิทธิภาพของระบบโดยละเอียดและพารามิเตอร์ที่ใช้ในการทดลอง แสดงดังตารางที่ 4.2

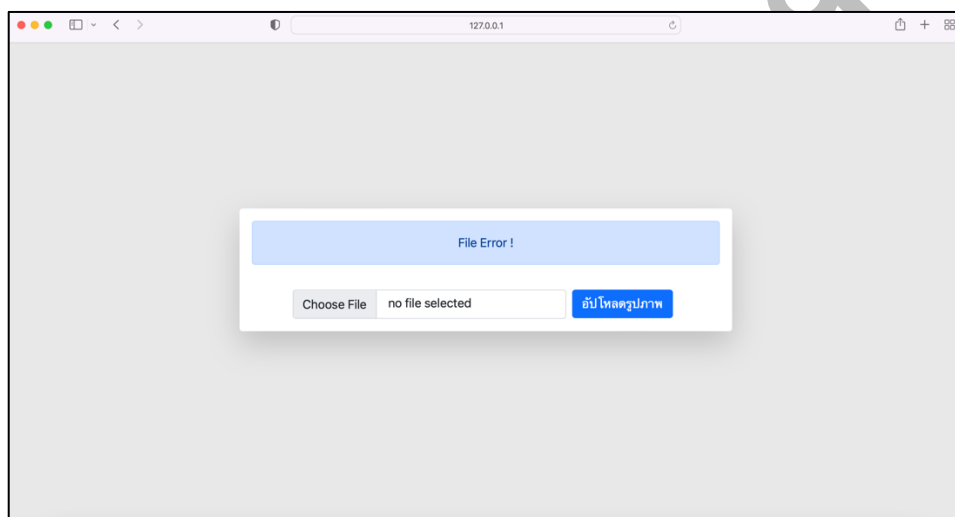
ตารางที่ 4.2 การตั้งค่าการทดลองประสิทธิภาพของระบบ

| อุปกรณ์ / ซอฟต์แวร์ | เวอร์ชัน / ขนาด |
|---------------------|--------------------|
| ระบบปฏิบัติการ | MacOS Version 12.0 |
| CPU | Apple M1 |
| Ram | 8 GB |
| Hard Disk | SSD 256 GB |

จากตารางที่ 4.2 การตั้งค่าการทดลองประสิทธิภาพของระบบ จะใช้ MacOS Version 12.0 เป็นระบบที่ใช้สำหรับทำงานโดยอุปกรณ์คอมพิวเตอร์มีคุณสมบัติดังนี้ หน่วยประมวลผลกลางของคือ Apple M1 มีความเร็วในการประมวลผลที่ 3.2 GHz ซึ่งมีหน่วยความจำหลักขนาด 8 GB มีหน่วยความถาวรประเภท Solid State Drive ขนาด 256 GB

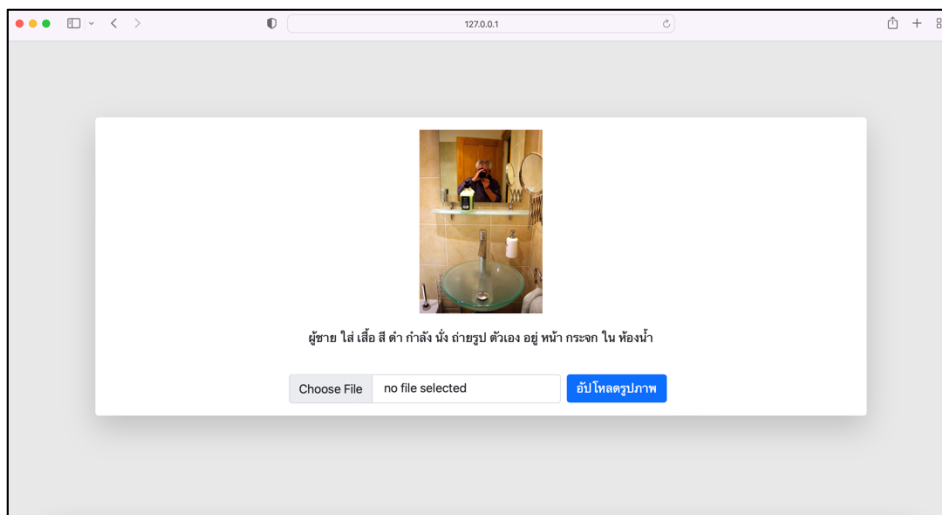
4.3 ทดสอบการอัปโหลดไฟล์ภาพที่มีนามสกุล JPEG JPG และ PNG

การทดสอบการอัปโหลดไฟล์ภาพที่มีนามสกุล JPEG, JPG และ PNG เป็นการทดลองเพื่อตรวจสอบการทำงานตรงตามเงื่อนไขที่กำหนดของไฟล์ภาพที่มีนามสกุล JPEG, JPG และ PNG โดยจะทำการอัปโหลดไฟล์รูปภาพเข้าสู่เว็บไซต์ เมื่อผู้ใช้งานอัปโหลดภาพไม่ตรงตามเงื่อนไขจะแสดงดังรูปที่ 4.1 ผู้ใช้งานอัปโหลดไฟล์ภาพตรงตามเงื่อนไขจะแสดงดังรูปที่ 4.2 และพิจารณาการทำงานตรงตามเงื่อนไขและไม่ตรงตามเงื่อนไข โดยอัปโหลดภาพ 1 ภาพ ต่อ 1 ครั้ง ทดสอบทั้งหมด 20 ครั้ง แสดงผลการทดลองดังรูปที่ 4.3



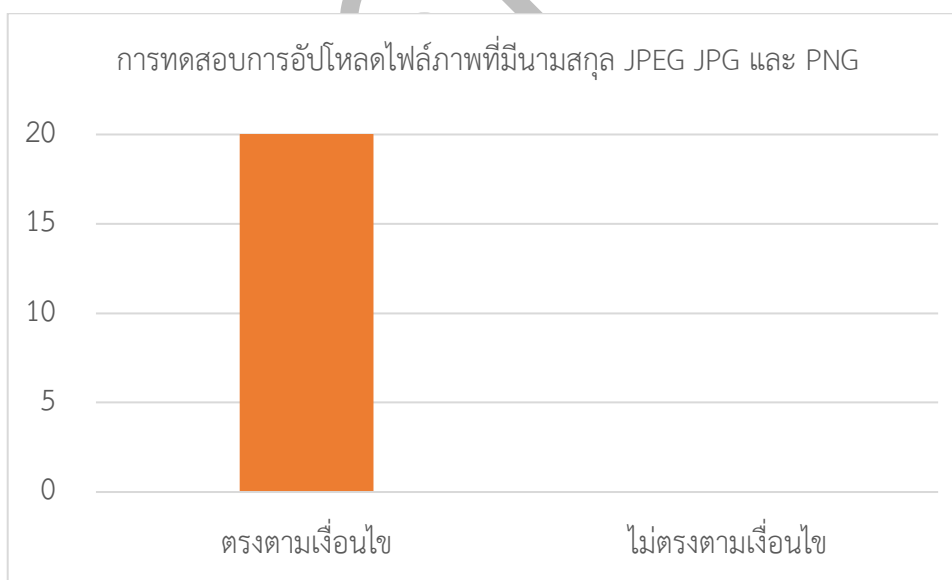
รูปที่ 4.1 ผู้ใช้งานอัปโหลดภาพไม่ตรงตามเงื่อนไข

จากรูปที่ 4.1 สามารถอธิบายการที่ผู้ใช้อัปโหลดภาพไม่ตรงตามเงื่อนไขที่กำหนดไว้ เนื่องจากนามสกุลไฟล์ที่อัปโหลดนั้นไม่ใช่นามสกุล JPEG, JPG และ PNG จึงมีการแจ้งเตือนขึ้นมาว่า “ File Error !”



รูปที่ 4.2 ผู้ใช้งานอัปโหลดไฟล์ภาพตรงตามเงื่อนไข

จากรูปที่ 4.2 สามารถอธิบายการที่ผู้ใช้อัปโหลดไฟล์ภาพตรงตามเงื่อนไข ซึ่งหากนามสกุลไฟล์ที่อัปโหลดนั้นมีนามสกุล JPEG, JPG และ PNG แล้วรูปที่อัปโหลดนั้นจะถูกประมวลผลแล้วใส่คำบรรยายภาพภาษาไทยอัตโนมัติ

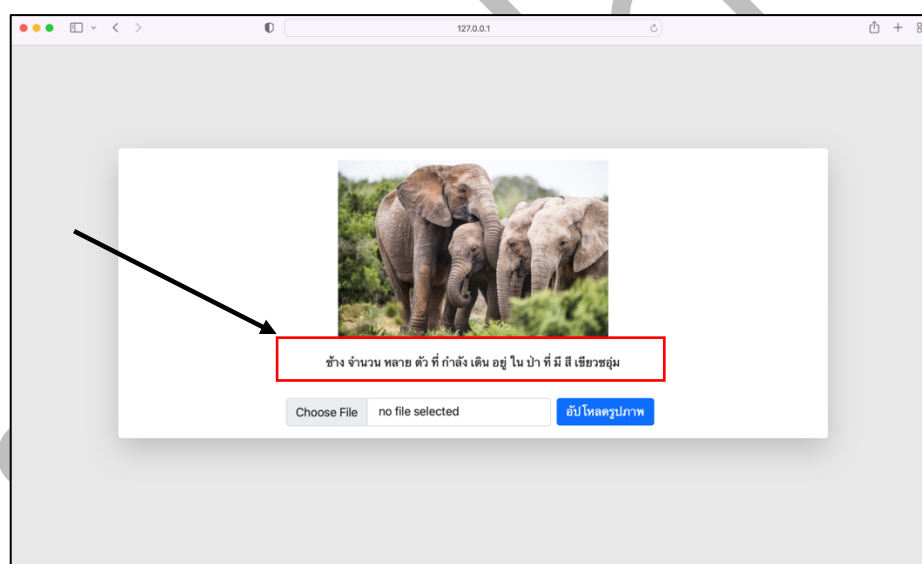


รูปที่ 4.3 ผลการทดลองการอัปโหลดไฟล์ภาพที่มีนามสกุล JPEG JPG และ PNG

จากรูปที่ 4.3 สามารถอธิบายผลการทดลองการอัปโหลดไฟล์ภาพที่มีนามสกุล JPEG JPG และ PNG โดยอัปโหลดภาพ 1 ภาพ ต่อ 1 ครั้ง ทดสอบทั้งหมด 20 ครั้ง สามารถทำงานได้ตรงตามเงื่อนไขทั้งหมดร้อยละ 100 ซึ่งสรุปได้ว่าการทำงานในส่วนนี้ สามารถทำงานได้อย่างถูกต้องตรงตามเงื่อนไข

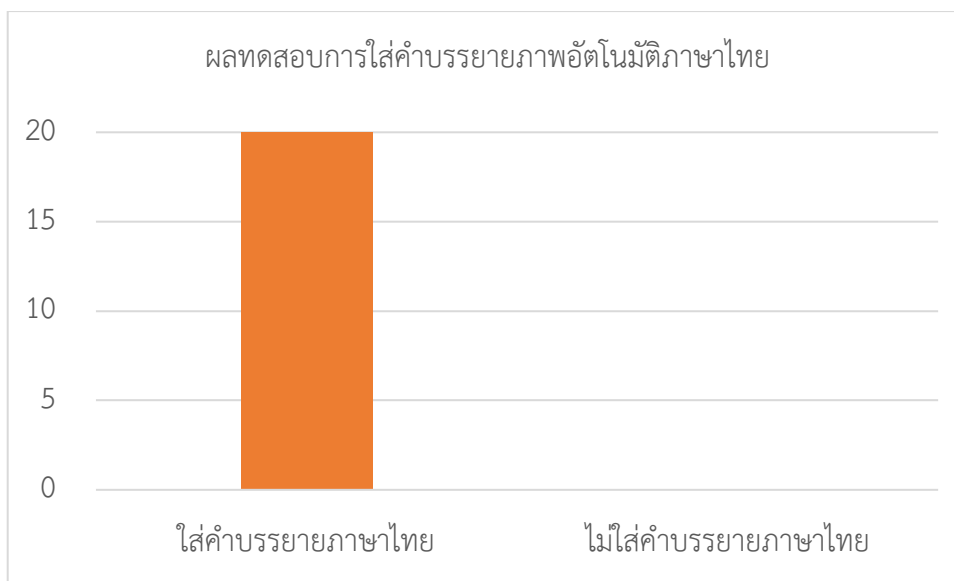
4.4 ทดสอบการใส่คำบรรยายภาพอัตโนมัติภาษาไทย

การทดสอบการใส่คำบรรยายภาพอัตโนมัติภาษาไทย เป็นการทดลองเพื่อทดสอบถูกต้องของการใส่คำบรรยายภาพอัตโนมัติที่เป็นภาษาไทยว่าทำงานได้ โดยจะทำการอัปโหลดไฟล์ภาพซึ่งภาพนั้นจะต้องผ่านการตรวจสอบนามสกุลไฟล์ที่ตรงตามเงื่อนไขแล้วหลังจากนั้นทำการใส่คำบรรยายภาพอัตโนมัติภาษาไทยแสดงดังรูปที่ 4.4 และพิจารณาความถูกต้องของการใส่คำบรรยายภาพอัตโนมัติภาษาไทย โดยอัปโหลดภาพ 1 ภาพ ต่อ 1 ครั้ง ทดสอบทั้งหมด 20 ครั้ง แสดงผลการทดลองดังรูปที่ 4.5



รูปที่ 4.4 แสดงการใส่คำบรรยายภาพอัตโนมัติภาษาไทย

จากรูปที่ 4.4 สามารถอธิบายแสดงการใส่คำบรรยายภาพอัตโนมัติภาษาไทย ซึ่งในรูปหลังจากทำการอัปโหลดไฟล์ภาพซึ่งภาพนั้นจะต้องผ่านการตรวจสอบนามสกุลไฟล์ที่ตรงตามเงื่อนไขแล้วหลังจากนั้นก็จะการใส่คำบรรยายภาพอัตโนมัติภาษาไทยก็คือ “ช้าง จำนวน หลาย ตัว ที่ กำลัง เดิน อยู่ ใน ป่า ที่ มี สี เขียวชอุ่ม “

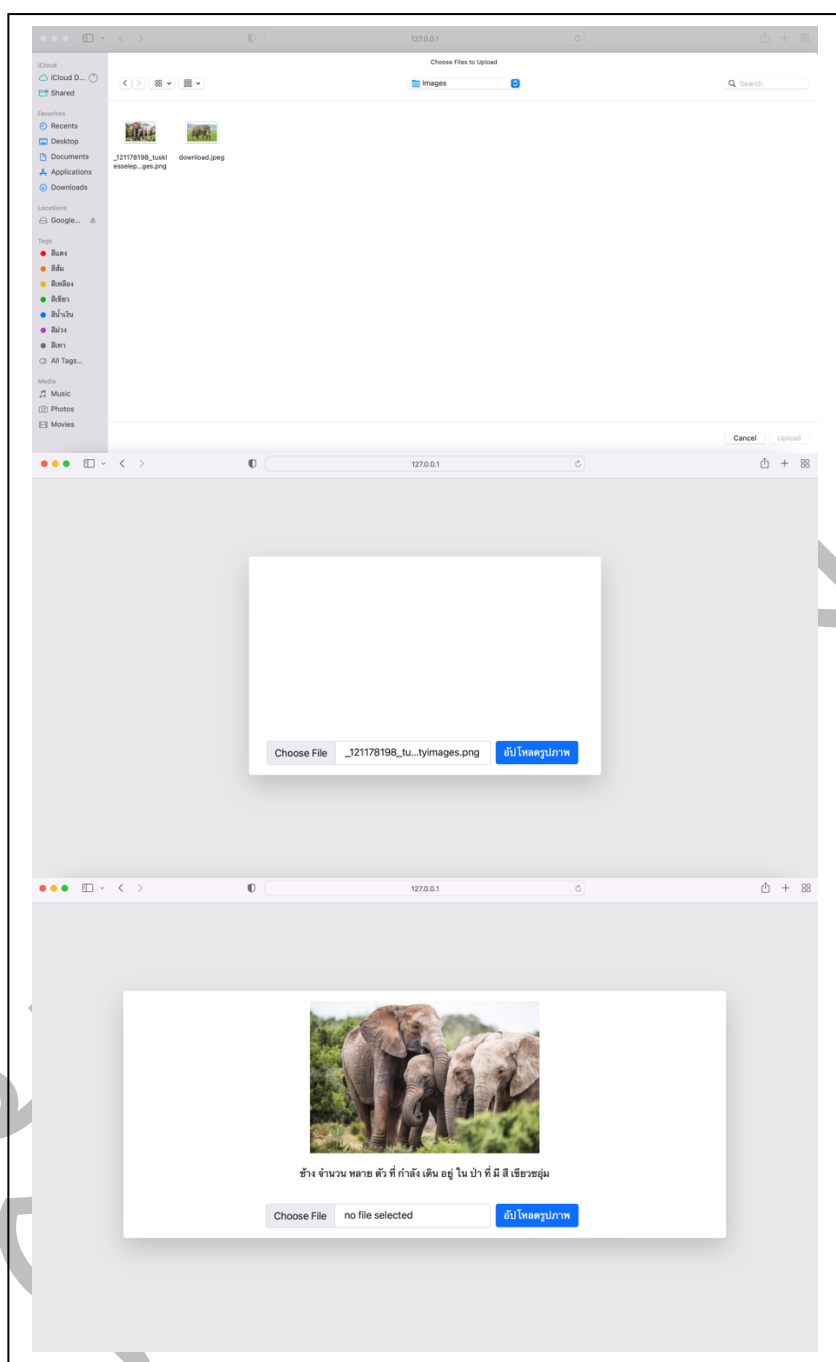


รูปที่ 4.5 ผลทดลองการใส่คำบรรยายภาพอัตโนมัติภาษาไทย

จากรูปที่ 4.5 สามารถอธิบายผลการทดลองการใส่คำบรรยายภาพอัตโนมัติภาษาไทย โดยอัปโหลดภาพ 1 ภาพ ต่อ 1 ครั้ง ทดสอบทั้งหมด 20 ครั้ง สามารถทำงานได้ถูกต้องทั้งหมดร้อยละ 100 ซึ่งสรุปได้ว่าการทำงานในส่วนนี้ สามารถทำงานได้อย่างถูกต้อง

4.5 ทดสอบการใส่คำบรรยายภาพอัตโนมัติภาษาไทยครั้งละ 1 ภาพ

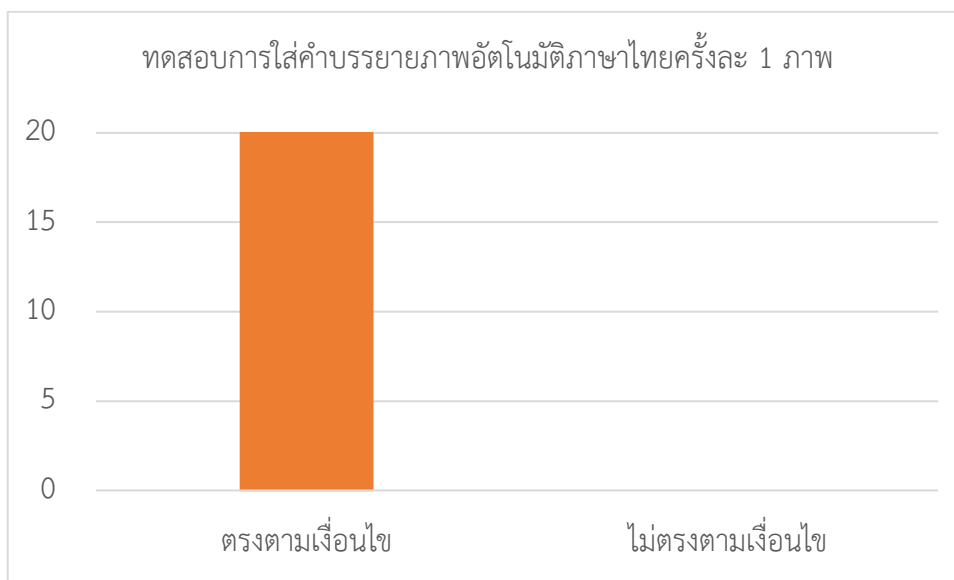
การทดสอบการใส่คำบรรยายภาพอัตโนมัติภาษาไทยครั้งละ 1 ภาพ เป็นการทดลองเพื่อตรวจสอบการทำงานตรงตามเงื่อนไขของการใส่คำบรรยายภาพอัตโนมัติภาษาไทยครั้งละ 1 ภาพ โดยทำการอัปโหลดภาพซึ่งภาพนั้นจะต้องผ่านการตรวจสอบนามสกุลไฟล์ที่ตรงตามเงื่อนไขแล้วหลังจากนั้น ทดสอบการใส่คำบรรยายภาพอัตโนมัติภาษาไทยแสดงดังรูปที่ 4.6 และพิจารณาการทำงานได้ตรงตามเงื่อนไข โดยทำการอัปโหลดภาพเพื่อทำการใส่คำบรรยายภาพอัตโนมัติภาษาไทย 1 ภาพ ต่อ 1 ครั้ง ทดสอบทั้งหมด 20 ครั้ง แสดงผลการทดลองดังรูปที่ 4.7



รูปที่ 4.6 แสดงการใส่คำบรรยายภาพอัตโนมัติภาษาไทยครั้งละ 1 ภาพ

จากรูปที่ 4.6 สามารถอธิบายการใส่คำบรรยายภาพอัตโนมัติภาษาไทยครั้งละ 1 ภาพ ซึ่งผู้ใช้ทำการเลือกกด Choose File เพื่อทำการอัปโหลดภาพซึ่งภาพ ซึ่งสามารถเลือกได้เพียงหนึ่งภาพหลังจากนั้น

ไฟล์ภาพนั้นจะต้องผ่านการตรวจสอบนามสกุลไฟล์ที่ตรงตามเงื่อนไขแล้วเมื่อตรงตามเงื่อนไขจะทำการใส่คำบรรยายภาพอัตโนมัติภาษาไทย



รูปที่ 4.7 ผลการทดลองการใส่คำบรรยายภาพอัตโนมัติภาษาไทยครั้งละ 1 ภาพ

จากรูปที่ 4.7 สามารถอธิบายผลการทดลองการใส่คำบรรยายภาพอัตโนมัติภาษาไทยครั้งละ 1 ภาพ โดยทำการอัปโหลดภาพเพื่อทำการใส่คำบรรยายภาพอัตโนมัติภาษาไทย 1 ภาพ ต่อ 1 ครั้ง ทดสอบทั้งหมด 20 ครั้ง สามารถทำงานได้ตรงตามเงื่อนไขทั้งหมดร้อยละ 100 ซึ่งสรุปได้ว่าการทำงานในส่วนนี้สามารถทำงานได้อย่างถูกต้องตรงตามเงื่อนไข

บทที่ 5

บทสรุปการดำเนินงานและข้อเสนอแนะ

5.1 สรุปผลการทดลอง

ในการจัดทำปฏิญานิพนธ์เรื่องการศึกษาการใช้งานเทคนิค Reinforcement ในการใส่คำบรรยายภาพอัตโนมัติภาษาไทย เพื่อเข้ามาช่วยแก้ปัญหาเรื่องปัญญาประดิษฐ์ใส่คำบรรยายภาพอัตโนมัติ นั้นยังเป็นการใส่คำบรรยายภาษาอังกฤษอยู่เนื่องจากชุดข้อมูลที่ใช้ในการฝึกอบรม นั้นเป็นภาษาอังกฤษ จึงทำให้เราต้องนำคำบรรยายที่เป็นภาษาอังกฤษนั้นนำไปแปลภาษาให้เป็นภาษาไทย ซึ่งมีโอกาสทำให้เกิดความไม่ถูกต้องของการแปลคำบรรยายภาพอัตโนมัติ ซึ่งการศึกษาการใช้งานเทคนิค Reinforcement ในการใส่คำบรรยายภาพอัตโนมัติภาษาไทย จะช่วยให้สามารถใส่คำบรรยายภาพนั้นมีความถูกต้องในการใส่คำบรรยายภาพอัตโนมัติภาษาไทย โดยที่ระบบจะมีการใส่คำบรรยายภาพอัตโนมัติจากผู้ใช้ได้อัปโหลดภาพไปยังเว็บไซต์ ระบบจะใส่คำบรรยายภาพอัตโนมัติภาษาไทยที่มีโอกาสเป็นไปได้มากที่สุดผ่านทางเว็บไซต์ เพื่อให้ประชาชนทั่วไปที่ต้องการนำภาพมาใส่คำบรรยายภาพอัตโนมัติภาษาไทยได้ จากผลการดำเนินงานพบว่าระบบสามารถใส่คำบรรยายภาพอัตโนมัติภาษาไทยมีความถูกต้องตรงตามเงื่อนไขมากกว่าร้อยละ 90

5.1 สรุปผลการทดสอบการอัปโหลดไฟล์ภาพที่มีนามสกุล JPEG JPG และ PNG ซึ่งผลการทดลองการอัปโหลดไฟล์ภาพที่มีนามสกุล JPEG JPG และ PNG โดยอัปโหลดภาพ 1 ภาพ ต่อ 1 ครั้ง ทดสอบทั้งหมด 20 ครั้ง สามารถทำงานได้ตรงตามเงื่อนไขทั้งหมดร้อยละ 100 พบว่าการทำงานในส่วนนี้สามารถทำงานได้อย่างถูกต้องตรงตามเงื่อนไข

5.2 สรุปผลการทดสอบการใส่คำบรรยายภาพอัตโนมัติภาษาไทย ซึ่งผลการทดลองการใส่คำบรรยายภาพอัตโนมัติภาษาไทย โดยอัปโหลดภาพ 1 ภาพ ต่อ 1 ครั้ง ทดสอบทั้งหมด 20 ครั้ง สามารถทำงานได้ถูกต้องทั้งหมดร้อยละ 100 พบว่าการทำงานในส่วนนี้สามารถทำงานได้อย่างถูกต้อง

5.3 สรุปผลการทดสอบการใส่คำบรรยายภาพอัตโนมัติภาษาไทยครั้งละ 1 ภาพ ซึ่งผลการทดลองการใส่คำบรรยายภาพอัตโนมัติภาษาไทยครั้งละ 1 ภาพ โดยทำการอัปโหลดภาพเพื่อทำการใส่คำบรรยายภาพอัตโนมัติภาษาไทย 1 ภาพ ต่อ 1 ครั้ง ทดสอบทั้งหมด 20 ครั้ง สามารถทำงานได้ตรงตามเงื่อนไขทั้งหมดร้อยละ 100 พบว่าการทำงานในส่วนนี้สามารถทำงานได้อย่างถูกต้องตรงตามเงื่อนไข

5.2 อุปสรรคและปัญหาในการดำเนินงาน

5.2.1 เครื่องที่ใช้ในการ train model มีประสิทธิภาพที่ค่อนข้างต่ำสำหรับงานระดับใหญ่ๆ ชุดข้อมูลจำนวนมากจึงทำให้ต้องใช้ระยะเวลาในการประมวลผลนาน

5.2.2 ตัวอย่างเครื่องมือจาก Github ที่ช่วยในการ train model (Self-Critical Sequence Training : SCST) ซึ่งเป็นเวอร์ชัน Python 2.7 พบปัญหาเรื่องความผิดพลาดของซอสโค้ดและไลบรารีเวอร์ชันเก่าทั้งหมด

5.3 แนวทางการแก้ไขปัญหา

5.3.1 ทำการ Remote เข้าเครื่องเซิร์ฟเวอร์ของแลปเพื่อทำการ train model ซึ่งมีประสิทธิภาพมากทำให้ช่วยลดระยะเวลาในการประมวลผล

5.3.2 ทำการแก้ไขเครื่องมือจาก Github ที่ช่วยในการ train model ซึ่งแก้ไขให้เป็น Python 3.x และไลบรารีเวอร์ชันใหม่

5.4 แนวทางในการพัฒนาและข้อเสนอแนะ

5.4.1 ทำการศึกษาการใช้งานเทคนิค Reinforcement ในการใส่คำบรรยายภาพอัตโนมัติภาษาไทย ที่ใช้สถาปัตยกรรมนอกเหนือจาก Convolution Neural Network, Long Short Term Memory, Swin Transformer, RoBERTa Transformer

5.4.2 ทำการเพิ่มชุดข้อมูลสำหรับ train model เพื่อให้การใส่คำบรรยายภาพอัตโนมัติภาษาไทยมีความถูกต้องมากยิ่งขึ้น

เอกสารอ้างอิง

บทความการประชุมวิชาการ

- [1] Rennie, S. J., Marcheret, E., Mroueh, Y., Ross, J., & Goel, V. “Self-critical Sequence Training for Image Captioning”. 2017 IEEE Conference on Computer Vision and Pattern Recognition. 21-26 July 2017. Honolulu: pp. 7008-7024.
- [2] Vedantam, R., Lawrence Zitnick, C., & Parikh, D. “Cider: Consensus-based Image Description Evaluation”. 2015 IEEE Conference on Computer Vision and Pattern Recognition. 7-12 June 2015. Boston: pp. 4566-4575.

เว็บไซต์

- [3] ชากิยะห์ ตาเยะ. (2564, ธันวาคม. 3). *NLP คืออะไร*. [ออนไลน์], แหล่งที่มา: <https://www.mindphp.com/%E0%B8%84%E0%B8%B9%E0%B9%88%E0%B8%A1%E0%B8%B7%E0%B8%AD/73%E0%B8%84%E0%B8%B7%E0%B8%AD%E0%B8%AD%E0%B8%B0%E0%B9%84%E0%B8%A3/8859-nlp.html>
- [4] บล็อกของ phyblas. (2564, ธันวาคม. 9). *ภาษา python เบื้องต้น*. [ออนไลน์], แหล่งที่มา: <https://phyblas.hinaboshi.com/tsuchinoko01>
- [5] COCO. (2564, ธันวาคม. 3). *What is COCO?*. [ออนไลน์], แหล่งที่มา: <https://cocodataset.org/#home>
- [6] James Loy. (2564, ธันวาคม. 3). *A Comprehensive Guide to Microsoft’s Swin Transformer?*. [ออนไลน์], แหล่งที่มา: <https://towardsdatascience.com/a-comprehensive-guide-to-swin-transformer-64965f89d14c>
- [7] Joseph Gatto. (2564, ธันวาคม. 3). *Understanding the SEMBLEU (and BLEU) Metric*. [ออนไลน์], แหล่งที่มา: <https://medium.com/analytics-vidhya/understanding-the-sembleu-and-bleu-metric-2e8a01cedcb3>
- [8] Koratink. (2564, ธันวาคม. 9). *ระบบสืออาร์จีบี*. [ออนไลน์], แหล่งที่มา: <https://shorturl.asia/e2SWQ>
- [9] Medium. (2564, ธันวาคม. 3). *Convolutional Neural Network (CNN) คืออะไร*. [ออนไลน์], แหล่งที่มา: <https://shorturl.asia/tGYsQ>

เอกสารอ้างอิง (ต่อ)

- [10] Medium. (2564, ธันวาคม. 3). *Deep Learning มันทำงานอย่างไรกันนะ?*. แหล่งที่มา: <https://medium.com/@pakpoom.thawee/deep-learning-%E0%B8%A1%E0%B8%B1%E0%B8%99%E0%B8%97%E0%B8%B3%E0%B8%87%E0%B8%B2%E0%B8%99%E0%B8%AD%E0%B8%A2%E0%B9%88%E0%B8%B2%E0%B8%87%E0%B9%84%E0%B8%A3%E0%B8%81%E0%B8%B1%E0%B8%99%E0%B8%99%E0%B8%B0-8eef57561d19>
- [11] Medium. (2564, ธันวาคม. 3). *มาทำความรู้จักกับ Reinforcement Learning แบบเบากันเถอะ?*. [ออนไลน์], แหล่งที่มา: <https://medium.com/@nutorbitx/%E0%B8%A1%E0%B8%B2%E0%B8%97%E0%B8%B3%E0%B8%84%E0%B8%A7%E0%B8%B2%E0%B8%A1%E0%B8%A3%E0%B8%B9%E0%B9%89%E0%B8%88%E0%B8%B1%E0%B8%81%E0%B8%81%E0%B8%B1%E0%B8%9A-reinforcement-learning-%E0%B9%81%E0%B8%9A%E0%B8%9A%E0%B9%80%E0%B8%9A%E0%B8%B2%E0%B9%86%E0%B8%81%E0%B8%B1%E0%B8%99%E0%B9%80%E0%B8%96%E0%B8%AD%E0%B8%B0-d36e71237b8>
- [12] Pakawat Nakwijit. (2564, ธันวาคม. 3). *ทำความเข้าใจ BERT?*. [ออนไลน์], แหล่งที่มา: <https://medium.com/@chameleontk/%E0%B8%97%E0%B8%B3%E0%B8%84%E0%B8%A7%E0%B8%B2%E0%B8%A1%E0%B9%80%E0%B8%82%E0%B9%89%E0%B8%B2%E0%B9%83%E0%B8%88-bert-98589715545>
- [13] Saixiii. (2564, ธันวาคม. 3). *Flask คืออะไร*. [ออนไลน์], แหล่งที่มา: <https://saixiii.com/python-flask-web-application/>
- [14] Sanparith Marukatat. (2564, ธันวาคม. 3). *LSTM เท่าที่เข้าใจ?*. [ออนไลน์], แหล่งที่มา: <https://sanparithmarukatat.medium.com/lstm%E0%B9%80%E0%B8%97%E0%B9%88%E0%B8%B2%E0%B8%97%E0%B8%B5%E0%B9%88%E0%B9%80%E0%B8%82%E0%B9%89%E0%B8%B2%E0%B9%83%E0%B8%88-75027db3167f>
- [15] Simeon Kostadinov. (2564, ธันวาคม. 3). *Understanding Encoder-Decoder Sequence to Sequence Model*. [ออนไลน์], แหล่งที่มา: <https://towardsdatascience.com/understanding-encoder-decoder-sequence-to-sequence-model-679e04af4346>

เอกสารอ้างอิง (ต่อ)

- [16] Surapong Kanoktipsatharporn. (2564, ธันวาคม. 3). *Transformer คืออะไร Self-Attention คืออะไร สอน Neural Machine Translation แปลภาษาฝรั่งเศส เป็นภาษาอังกฤษด้วย Transformer - NLP?*. [ออนไลน์],แหล่งที่มา: <https://www.bualabs.com/archives/3186/what-is-transformer-machine-learning-teach-neural-machine-translation-translate-french-to-english-with-transformer-model-nlp-ep-12/>
- [17] Thomas Simonini. (2564, ธันวาคม. 3). *Proximal Policy Optimization (PPO)*. [ออนไลน์], แหล่งที่มา: <https://huggingface.co/blog/deep-rl-ppo>
- [18] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luka Zettlemoyer, Veselin Stoyanov. (2564, ธันวาคม. 3). *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. [ออนไลน์], แหล่งที่มา: <https://arxiv.org/abs/1907.11692>

ภาคผนวก

ตัวอย่าง

ภาคผนวก ก.

ประวัติสถานประกอบการสหกิจศึกษา

ประวัติสถานประกอบการสหกิจศึกษา

ข้อมูลสถานประกอบการ

ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (National Electronics and Computer Technology Center : NECTEC หรือเนคเทค) ก่อตั้งขึ้นโดยมติคณะรัฐมนตรี เมื่อวันที่ 16 กันยายน พ.ศ. 2529 โดยในระยะเริ่มต้นมีสถานะเป็นโครงการภายใต้ศูนย์ถ่ายทอดเทคโนโลยี สำนักงานปลัดกระทรวงวิทยาศาสตร์ เทคโนโลยีและการพลังงาน (ชื่อในขณะนั้น) ต่อมาในวันที่ 30 ธันวาคม 2534 เนคเทคได้เปลี่ยนแปลง สถานะเป็นศูนย์แห่งชาติเฉพาะทาง และเปลี่ยนการจัดรูปแบบองค์กรใหม่ เพื่อให้มีความคล่องตัวขึ้นกว่าเดิม ตามพระราชบัญญัติพัฒนาวิทยาศาสตร์และเทคโนโลยี พ.ศ. 2534

ภารกิจหลักของเนคเทค ได้แก่การให้ทุนสนับสนุนการวิจัยในภาครัฐ การดำเนินการวิจัยเอง เพื่อเร่งให้ผลงานวิจัยเกิดผลจริงในภาคอุตสาหกรรม การให้บริการเพื่อสร้าง ความแข็งแกร่งให้แก่อุตสาหกรรมอิเล็กทรอนิกส์ คอมพิวเตอร์ โทรคมนาคม และสารสนเทศ และการทำหน้าที่เป็นสำนักงานเลขานุการคณะกรรมการเทคโนโลยีสารสนเทศแห่งชาติ

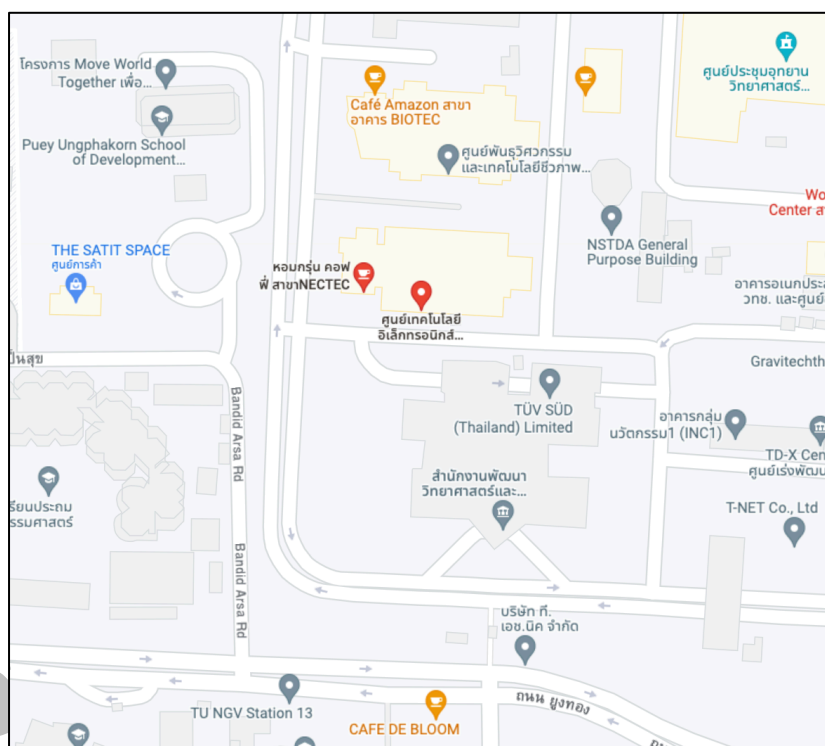
ตั้งแต่ พ.ศ. 2539 เป็นต้นมา เนคเทคได้รับพระมหากรุณาธิคุณจากสมเด็จพระเทพรัตนราชสุดาฯ สยามบรมราชกุมารี ให้ดำเนินโครงการเครือข่ายกาญจนาภิเษก เพื่อกระจายความรู้ แก่ประชาชน และเป็นสำนักงานเลขานุการโครงการเทคโนโลยีสารสนเทศตามพระราชดำริฯ เพื่อประยุกต์ใช้ไอทีกับสังคมไทย โดยเน้นนักเรียนในชนบท ผู้พิการ และเด็กที่ป่วยในโรงพยาบาล

เนคเทคได้ดำเนินงานโครงการที่จัดได้ว่าเป็นโครงการสร้างพื้นฐานระดับชาติหลายโครงการ เช่น เครือข่ายไทยสารอินเทอร์เน็ต เครือข่ายคอมพิวเตอร์เพื่อโรงเรียนไทย เครือข่ายกาญจนาภิเษก เขตอุตสาหกรรมซอฟต์แวร์ (ซอฟต์แวร์พาร์ค) ศูนย์พัฒนาพาณิชย์อิเล็กทรอนิกส์ (Electronic Commerce Resource Center) การพัฒนากฎหมาย เทคโนโลยีสารสนเทศ การเป็นศูนย์กลางประสานงานในการแก้ไขปัญหาคอมพิวเตอร์ปี ค.ศ. 2000 และสำนักบริการเทคโนโลยีสารสนเทศภาครัฐ (สบทร.)

1. ข้อมูลสถานประกอบการ

1.1 สถานที่ตั้งสถานประกอบการ

112 อุทยานวิทยาศาสตร์ประเทศไทย ถนนพหลโยธิน ตำบลคลองหนึ่ง อำเภอ
คลองหลวง จังหวัดปทุมธานี 12120



รูปที่ ก.1 แผนที่ตั้งสถานประกอบการ ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ NECTEC

1.2 ตราสัญลักษณ์สถานประกอบการ



รูปที่ ก.2 ตราสัญลักษณ์ของสถานประกอบการศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ

NECTEC

1.3 ข้อมูลทั่วไปการติดต่อสื่อสาร

โทร.02-564-6900 ต่อ 2346 - 2355

ประวัติผู้จัดทำปริญญานิพนธ์



ชื่อ-สกุล นายพุทพงษ์ ด้านทองกลาง
 ภูมิลำเนา 69 หมู่ 4 ตำบล จักราช อำเภอ จักราช
 จังหวัด นครราชสีมา 30230

ประวัติการศึกษา

พ.ศ. 2558 จบการศึกษาระดับมัธยมศึกษาปีที่ 3
 วิทยาศาสตร์ - คณิตศาสตร์
 โรงเรียนจักราชวิทยา

พ.ศ. 2561 จบการศึกษาระดับมัธยมศึกษาปีที่ 6
 วิทยาศาสตร์ - คณิตศาสตร์
 โรงเรียนจักราชวิทยา

พ.ศ. 2565 ศึกษาในระดับปริญญาตรี
 วิศวกรรมคอมพิวเตอร์
 คณะวิศวกรรมศาสตร์และเทคโนโลยี

Email putthapong.da@rmuti.ac.th