การแต่งวรรณกรรมประเภทเรื่องสั้นด้วยการเรียนรู้เชิงลึก

WRITING SHORT FICTION BY DEEP LEARNING

เขมชาติ ศิลาวรรณ

ธน นิ่มถาวร

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2563

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ปริญญานิพนธ์ปีการศึกษา 2563

ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง การแต่งวรรณกรรมประเภทเรื่องสั้นด้วยการเรียนรู้เชิงลึก

WRITING SHORT FICTION BY DEEP LEARNING

ผู้จัดทำ

1. นายเขมชาติ ศิลาวรรณ รหัสนักศึกษา 60010106

2. นายธน นิ่มถาวร รหัสนักศึกษา 60010379

_____ อาจารย์ที่ปรึกษา

(รศ. ดร. เกียรติกูล เจียรนัยธนะกิจ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

การแต่งวรรณกรรมประเภทเรื่องสั้นด้วยการเรียนรู้เชิงลึก

นายเขมชาติ ศิลาวรรณ 60010106

นายธน นิ่มถาวร 60010379

รศ. ดร. เกียรติกูล เจียรนัยธนะกิจ อาจารย์ที่ปรึกษา

ปีการศึกษา 2563

บทคัดย่อ

งานเขียนวรรณกรรมถือเป็นงานศิลปะที่แสดงออกด้วยภาษาเพื่อสื่อสารเรื่องราวให้เข้าใจกันระหว่างมนุษย์ เกิดมาจากความคิดสร้างสรรค์และผ่านการเรียบเรียงด้วยวิธีต่าง ๆ ในโครงงานชิ้นนี้ ผู้จัดทำได้ศึกษากระบวนการที่นำความสามารถของการเรียนรู้เชิงลึกมาใช้แต่งวรรณกรรมจากงานวิจัยที่มีอยู่แล้ว ผู้จัดทำพบว่าผลลัพธ์ที่ได้ออกมาจากงานวิจัยเก่านั้นยังมีข้อจำกัดอยู่หลายอย่างที่สามารถพัฒนาต่อเพื่อเพิ่มคุณภาพของผลลัพธ์ ผู้จัดทำจึงนำโครงสร้างโมเดลการเรียนรู้เชิงลึกที่มีชื่อว่า Convolutional Sequence to Sequence มาทำงานร่วมกับโมเดลรูปแบบอื่น ๆตามแนวคิดของงานวิจัย Strategies for Structuring Story Generation เพื่อให้สามารถแต่งวรรณกรรมประเภทเรื่องสั้นที่เนื้อเรื่องมีความต่อเนื่อง สัมพันธ์กัน ดำเนินเนื้อเรื่องได้ โมเดลการเรียนรู้เชิงลึกที่ผู้จัดทำนำมาพัฒนาต่อจะสามารถหาความสัมพันธ์ระยะยาวระหว่างเนื้อเรื่องที่แต่งขึ้นจากคำกำกับบทในแต่ละส่วน เพื่อไม่ให้เนื้อหาออกมาผิดไปจากเนื้อเรื่องเดิมที่แต่งไว้มาก และมีการศึกษาวิธีการควบคุมการดำเนินเรื่องจากพล็อตหรือรูปแบบการดำเนินเรื่องของวรรณกรรม เพื่อให้โมเดลสามารถแต่งวรรณกรรมประเภทเรื่องสั้นที่มีการดำเนินเรื่องราวต่อได้อย่างมีรูปแบบ ไม่วนซ้ำอยู่ที่จุดเดิม การพัฒนาโมเดลและการทำงานต่าง ๆจะทำขึ้นบน Google Colaboratory ที่มีหน่วยประมวลผลการเรียนรู้เชิงลึกให้ใช้ฟรี ผลลัพธ์ของโครงงานนี้จะช่วยเปิดโอกาสและสร้างความเป็นไปได้ใหม่ ๆ ด้วยการแต่งวรรณกรรมประเภทเรื่องสั้นที่มีคุณภาพโดยไม่จำเป็นต้องใช้มนุษย์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

Writing Short Fiction by Deep Learning

Mr. Kheamachat	Silawan	60010106
Mr. Tana	Nimtavorn	60010379
Assoc. Prof. Dr. Kietikul	Jearanaitanakij	Advisor

Academic Year 2020

ABSTRACT

Literature is the art of storytelling in human language created by creativity. In this project we study from existing research about the process of applying deep learning to compose short story. We found that the results from previous studies can be further developed and improved. We therefore adopt a deep learning model called Convolutional Sequence to Sequence along with other models from concept of Strategies for Structuring Story Generation to improve for better result. To be able to compose a short story with a continuation of the story, the deep learning model developed will be able to find a long-term relationship between the story written from the prompt in each parts The beginning of the story, The middle of the story and The end of the story in order not to make the story go out of the original, and we study on how to directed the storyline from the plot or the literary style of short story so the model can compose short story that can run in a new storyline, not repeating in the same place. Development will be done on a Google Colaboratory. By composing quality short stories without the use of humans, the results of this project can open many new possibilities.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

กิตติกรรมประกาศ

ขอขอบคุณ รศ.ดร.เกียรติกูล เจียรนัยธนะกิจ
อาจารย์ที่ปรึกษาโครงงานที่ให้คอยให้คำปรึกษา
แนะนำและคอยให้กำลังใจตลอดการทำโครงงานครั้งนี้

นายเขมชาติ ศิลาวรรณ
นายธน นิ่มถาวร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

สารบัญ

หน้า

การแต่งวรรณกรรมประเภทเรื่องสั้นด้วยการเรียนรู้เชิงลึก I

Writing Short Fiction by Deep Learning II

สารบัญ IV

สารบัญตาราง IV

สารบัญรูป VII

บทที่ 1 บทนำ 1

1.1 ความเป็นมาและความสำคัญ 1

1.2 วัตถุประสงค์ของโครงงาน 2

1.3 เป้าหมายของโครงงาน 2

1.4 ขอบเขตของโครงงาน 2

1.5 แผนการดำเนินงาน 2

บทที่ 2 งานวิจัยที่เกี่ยวข้อง 4

2.1 ทฤษฎีที่เกี่ยวข้อง 4

2.2 งานวิจัยที่เกี่ยวข้อง 8

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

สารบัญ(ต่อ)

หน้า

บทที่ 3 การออกแบบ 13

3.1 การวิจัยและภาพรวมของระบบ.....13

3.2 การออกแบบโมเดลและขั้นตอนการทำงาน.....14

3.3 การออกแบบการทดลอง.....18

3.4 การวัดผลการทดลอง.....18

บทที่ 4 การทดลองและผลการทดลอง.....19

4.1 การสร้างโมเดล.....19

4.2 การทดสอบโมเดลด้วยข้อมูลที่รวบรวมมา.....22

4.3 สรุปผลการทดสอบโมเดลโดยใช้ข้อมูลที่รวบรวมมา.....24

บทที่ 5 สรุป.....25

5.1 สรุปผลการทำงาน.....25

5.2 ปัญหาที่เกิดขึ้นและแนวทางแก้ไข.....25

5.3 แนวทางการพัฒนาต่อ.....27

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

สารบัญตาราง

ตาราง **หน้า**

ตาราง 4.1 ค่า Perplexity ที่ได้จากโมเดลการเรียนรู้เชิงลึกที่ใช้ในการแต่งเรื่องสั้นโดยมีการแบ่งเนื้อเรื่องออกเป็น 3 ส่วน เทียบกับโมเดล Hierarchical Neural Story Generation 24

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

สารบัญรูป

รูป		หน้า
รูป 1.1	แผนการดำเนินงานโปรเจกต์	3
รูป 2.1	การเปรียบเทียบคำที่ทำ Embeddings	4
รูป 2.2	Image Classification using CNN	5
รูป 2.3	โครงสร้างของ Self-Attention แบบ Scaled Dot-Product และแบบ Multi-head	6
รูป 2.4	แผนภาพกระบวนการสร้างเรื่องราว	9
รูป 2.5	Multihead self-attention mechanism	10
รูป 2.6	Fusion Model	11
รูป 2.7	แสดงการ batching ในระหว่าง training โมเดล ConvSeq2Seq	12
รูป 3.1	โมเดลการเรียนรู้เชิงลึกที่ใช้ในการแต่งเรื่องสั้น	14
รูป 3.2	การทำ Padding ของ Encoder	15
รูป 3.3	Self-Attention Mechanism ที่อยู่ในแต่ละ Head ของ Multi-head Self-attention	16
รูป 3.4	การทำ Padding ของ Decoder	17
รูป 4.1	กราฟแสดงผลลัพธ์ที่ได้ของ Decomposition model1 ที่ Epoch = 30	20
รูป 4.2	กราฟแสดงผลลัพธ์ที่ได้ของ Decomposition model2 ที่ Epoch = 30	20
รูป 4.3	กราฟแสดงผลลัพธ์ที่ได้ของ Decomposition model3 ที่ Epoch = 50	21

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

สารบัญรูป(ต่อ)

รูป		หน้า
รูป 4.4	กราฟแสดงผลลัพธ์ที่ได้ของ Fusion model ที่ Epoch = 30	21
รูป 4.5	กราฟแสดงผลลัพธ์ที่ได้ของ Decomposition model1 ที่ Epoch = 30	22
รูป 4.6	กราฟแสดงผลลัพธ์ที่ได้ของ Decomposition model2 ที่ Epoch = 30	23
รูป 4.7	กราฟแสดงผลลัพธ์ที่ได้ของ Decomposition model3 ที่ Epoch = 50	23
รูป 4.8	กราฟแสดงผลลัพธ์ที่ได้ของ Fusion model ที่ Epoch = 30	24

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญ

เรื่องสั้นเป็นเรื่องที่แต่งขึ้นมาในรูปแบบร้อยแก้วคล้ายคลึงกับนวนิยายแต่สั้นกว่า ได้รับอิทธิพลมาจากตะวันตก เรื่องสั้นมีหลากหลายประเภทหลากหลายแนว ตัวอย่างเช่น เรื่องสั้นแนวแฟนตาซี แนวไซไฟ แนวลึกลับ มีเหตุการณ์ในเรื่อง ตัวละครน้อย มักจะจบแบบเหนือความคาดหมายและอาจจะทิ้งท้ายให้คิดตามแบบฉบับของผู้แต่ง การอ่านเรื่องสั้นสามารถให้ความสนุกสนาน และอาจก่อให้เกิดคำถามจากความคิดที่ได้หลังอ่านเรื่องสั้นจบ สำหรับเด็ก การอ่านเรื่องสั้นสามารถพัฒนาทักษะการสื่อสารได้เป็นอย่างดี

ในปัจจุบันเทคโนโลยีปัญญาประดิษฐ์ได้พัฒนาไปอย่างมาก และเข้ามามีบทบาทสำคัญในการดำเนินชีวิต มีการนำปัญญาประดิษฐ์ การเรียนรู้ของเครื่อง และการเรียนรู้เชิงลึกมาใช้แก้ปัญหาที่มีความซับซ้อน ทั้งยังสามารถทำการทำนายอนาคตจากข้อมูลที่มีอยู่ได้อีกด้วย แต่ว่างานบางประเภทนั้นการใช้ปัญญาประดิษฐ์จะทำได้ไม่ดีเท่ากับที่มนุษย์ทำ เช่น งานที่ต้องใช้ความคิดและอารมณ์ การแต่งเพลง การวาดภาพ การเขียนนิยายหรือแต่งเรื่องราวขึ้นมา การเอาปัญญาประดิษฐ์มาใช้กับงานเหล่านี้นั้นผลลัพธ์ที่ได้จะไม่มีความเป็นธรรมชาติเท่าที่ควร

โครงงาน “แต่งวรรณกรรมประเภทเรื่องสั้นด้วยการเรียนรู้เชิงลึก (Create Short Fictional Story using Deep Learning)” นี้จึงถูกจัดทำขึ้น เพื่อทำการศึกษาวิธีใช้การเรียนรู้เชิงลึกในการเขียนวรรณกรรมประเภทเรื่องสั้นขึ้นมา ด้วยการศึกษาโครงสร้างจากงานเขียนฝีมือมนุษย์และทำการสร้างงานเขียนใหม่ ทั้งนี้เพื่อให้งานเขียนที่ได้มีความเป็นเหตุเป็นผลและมีเนื้อเรื่องที่สัมพันธ์กัน จึงหาเทคนิคละวิธีการต่าง ๆมาใช้ควบคู่กันไปด้วยเพื่อให้งานเขียนมีความเป็นธรรมชาติมากที่สุด เนื้อเรื่องที่ได้จะสามารถดำเนินต่อไปได้จากตัวแปรต่าง ๆที่กำหนด เป็นเหตุเป็นผล อ่านแล้วเข้าใจ โดยผลสำเร็จจะเป็นโมเดลที่สามารถแต่งวรรณกรรมประเภทเรื่องสั้นขึ้นมาด้วยตัวเองตามข้อความที่ป้อนเข้าไป พร้อมกับชุดข้อมูลที่เป็นตัวอย่างเรื่องสั้นสำหรับการประมวลผลในการเรียนรู้เชิงลึกและออกมาเป็นงานซึ่งมีความลื่นไหล อ่านแล้วสามารถเข้าใจได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

1.2 วัตถุประสงค์ของโครงงาน

เพื่อศึกษาวิธีใช้เทคนิคต่าง ๆควบคู่กับการเรียนรู้เชิงลึกในการแต่งวรรณกรรมประเภทเรื่องสั้นขึ้นมาจากชุดข้อมูลที่เป็นตัวอย่างวรรณกรรมประเภทเรื่องสั้น มีเนื้อหาตรงกับคำกำกับบทที่ป้อนเข้าไป และปรับปรุงงานวิจัยก่อนหน้า ประยุกต์ใช้กระบวนการอื่น ๆเพื่อให้วรรณกรรมประเภทเรื่องสั้นที่แต่งขึ้นมาสามารถดำเนินเรื่องได้ เนื้อเรื่องที่ดำเนินต่อไปมีความสอดคล้องกับเนื้อเรื่องก่อนหน้า

1.3 เป้าหมายของโครงงาน

ผู้ใช้สามารถให้คอมพิวเตอร์แต่งวรรณกรรมประเภทเรื่องสั้นขึ้นมาจากคำกำกับบทที่ป้อนเข้าไปจำนวน 3 ชุด คำกำกับบทแต่ละชุดจะมีความยาวขั้นต่ำชุดละ 3 คำ นำไปใช้เป็นตัวกำหนดเนื้อเรื่องในแต่ละส่วนได้แก่ ต้นเรื่อง กลางเรื่อง และปลายเรื่อง เนื้อเรื่องแต่ละส่วนจะออกมามีความยาวไม่เกิน 1,000 คำ รวมกันเป็นวรรณกรรมประเภทเรื่องสั้นไม่เกิน 3,000 คำ

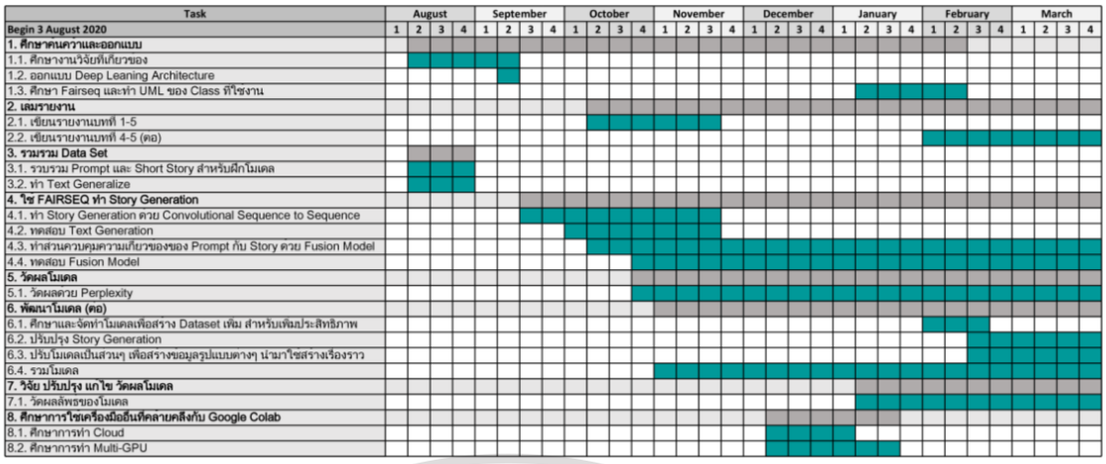
1.4 ขอบเขตของโครงงาน

1) สร้างโมเดลการเรียนรู้เชิงลึกที่สามารถแต่งเรื่องสั้นออกมาเป็นข้อความในภาษาอังกฤษจากเซ็ตข้อมูลเรื่องสั้นที่มีผู้แต่งไว้แล้ว มีเนื้อเรื่องที่ต่อเนื่อง และเกี่ยวข้องกับข้อความที่ระบุไว้ โดยมนุษย์สามารถอ่านแล้วทำความเข้าใจได้
2) เรื่องสั้นที่จะเอามาใส่ในเซ็ตข้อมูลควรจะมีจำนวนตัวอักษรอย่างน้อย 100 คำ ในหนึ่งเรื่อง
3) เนื่องจากการใช้การเรียนรู้เชิงลึกในการแต่งงานเขียนออกมา จำเป็นจะต้องใช้ชุดข้อมูลที่เป็นเรื่องสั้น หรือเรื่องแต่งโดยมนุษย์ในปริมาณที่เพียงพอเพื่อให้ได้งานเขียนที่สามารถอ่านแล้วเข้าใจได้
4) ผลลัพธ์ที่ได้จะอยู่ในรูปของโมเดล และโปรแกรมทดสอบเพื่อทำผลการทดลอง ทั้งนี้เพื่อให้งานมีประเด็นอยู่ในเรื่องการเรียนรู้เชิงลึก และการใช้เทคนิคต่าง ๆเพื่อปรับปรุงผลลัพธ์เป็นหลัก

1.5 แผนการดำเนินงาน

แผนการดำเนินงานของโปรเจกต์ 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.



รูป 1.1แผนการดำเนินงานโปรเจกต์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

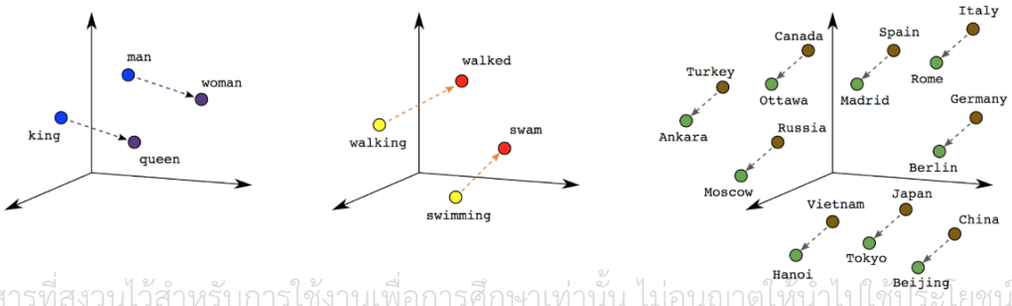
บทที่ 2
งานวิจัยที่เกี่ยวข้อง

ในการวิจัยครั้งนี้ ผู้จัดทำได้ศึกษางานวิจัยและทฤษฎีที่เกี่ยวข้องซึ่งนำเสนอในบทนี้ ทฤษฎีได้แก่ Word Embedding แปลงคำเป็นตัวเลขที่มีค่า, Convolutional Neural Network เป็นโครงข่ายประสาทเทียมเพื่อใช้ในการเรียนรู้, กระบวนการ Self-Attention ใช้ปรับค่าน้ำหนักของโมเดล, Gated Linear Unit (GLU) ควบคุมค่าของข้อมูลที่สนใจ, Residual Connection แก้ปัญหา Vanishing gradient และใช้ Perplexity ในการทดสอบผลลัพธ์ ส่วนงานวิจัยที่เกี่ยวข้องได้แก่ Hierarchical Neural Story Generation เป็นการใช้ Deep Learning แต่งเนื้อเรื่องขึ้นมา, Convolutional Sequence to Sequence Learning เป็นการใช้ Convolutional Neural Network ในการทำ Language Model ที่อยู่ในรูปแบบ Sequence to Sequence เป็นต้น โดยผู้จัดทำได้นำแนวคิดจากทฤษฎีและงานวิจัยเหล่านี้มาประยุกต์ใช้ในการทำงานวิจัย

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 Word Embedding

เป็นกระบวนการแปลงคำที่อยู่ในรูปแบบตัวอักษร ไปเป็นตัวเลขเพื่อให้สามารถนำไปเป็น Input ของ Neural Network ได้ เมื่อเปลี่ยนคำให้กลายเป็นค่าตัวเลขทางคณิตศาสตร์แล้ว ค่าที่ได้สามารถนำไปใช้ในการหาความสัมพันธ์ระหว่างคำอื่น ๆที่อยู่ในชุดข้อมูลเพื่อให้คอมพิวเตอร์ใช้ในการเรียนรู้และเลือกคำที่จะต้องใช้ในการทำงานเช่นการแปลภาษา หรือเลือกคำเพื่อแต่งประโยคขึ้นมาใหม่



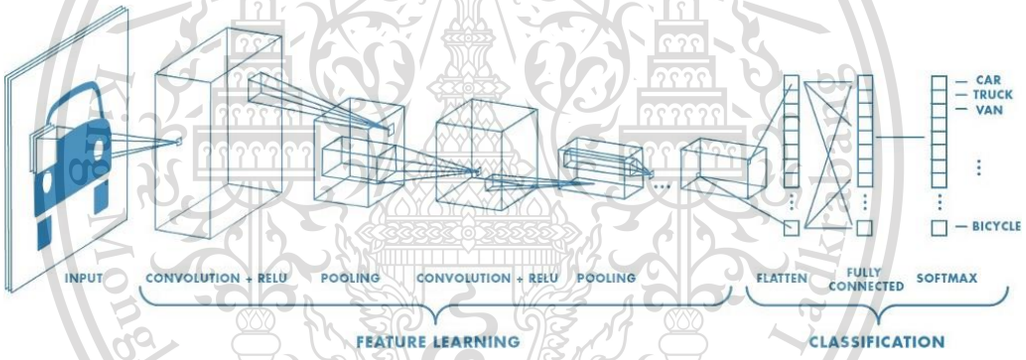
รูป 2.1การเปรียบเทียบคำที่ทำ Embeddings

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

กระบวนการทำ Word Embedding มีอยู่หลายวิธี เช่น TF-IDF, one hot encoding, word2vec เป็นต้น

2.1.2 Convolutional Neural Network

เป็นรูปแบบหนึ่งของโครงข่ายประสาทเทียมชนิดหนึ่ง ประกอบไปด้วย Convolutional layer และ Pooling layer ส่วนใหญ่มักจะใช้กับโมเดลประมวลผลรูปภาพ แต่สามารถนำมาใช้กับการประมวลผลคำได้ด้วยด้วยการประยุกต์ใช้ลักษณะของ Convolutional Neural Network ที่ต่อ ๆ กันเหมือนพีระมิด ทำให้ใช้คำนวณประโยคหลาย ๆประโยคพร้อมกันเป็นแบบขนานได้ ลดระยะเวลาที่ใช้ลงเป็นอย่างมาก ต่างกับการใช้ Recurrent Neural Network หรือ Long Short-Term Memory ที่มีลักษณะเป็น Sequential

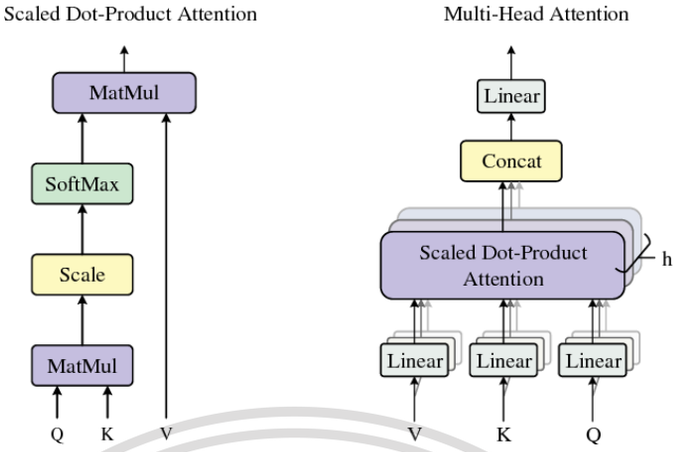


รูป 2.2Image Classification using CNN

2.1.3 Self-Attention Mechanism

self-attention ประกอบไปด้วย Encoder และ decoder ฝั่ง encoder จะทำการแปลง Input ที่เป็นคำเป็น attention score หรือ context vector แล้วส่งไปให้ decoder ฝั่ง decoder จะรู้ว่าคำแต่ละคำในประโยคมีความเกี่ยวข้องกันแค่ไหน ทำให้คำที่สร้างออกมาจากมีความสอดคล้องกันมากขึ้น และมีการนำ output จาก decoder ส่งกลับเข้าไปเป็น input ของ decoder ใน Timestep ต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.



รูป 2.3โครงสร้างของ Self-Attention แบบ Scaled Dot-Product และแบบ Multi-head

2.1.4 Gated Linear Unit

การ Gating ลักษณะนี้ทำให้โมเดลสามารถเลือกฟีเจอร์ที่สำคัญของคำที่เป็น input เพื่อใช้ในการสร้างคำถัดไป ทำโดยการ นำ input ที่ได้จาก convolutional layer แล้วแบ่งเป็น 2 ส่วน A และ B นำ A เข้า sigmoid function จะทำหน้าที่เป็น Gate จะทำการตัดสินว่าแต่ละของ B มีความเกี่ยวข้องกับบริบทปัจจุบัน แล้วทำการคูณกับ B แบบ element-wise หรือที่เรียกว่า Hadamard product โดย Gated Linear Unit มีสมการดังสมการที่ 2.1

$$h(X) = (X * W + b) \otimes \sigma(X * V + c) \tag{2.1}$$

โดย m เป็นขนาดของ input feature maps, n เป็นขนาดของ output feature maps, k เป็นขนาดของ kernel $X \in \mathbb{R}^{n \times m}, W \in \mathbb{R}^{k \times m \times n}, b \in \mathbb{R}^{n}, V \in \mathbb{R}^{k \times m \times n}, c \in \mathbb{R}^{n}$

2.1.5 Residual Connection

เมื่อมีการซ้อนกันเยอะๆ ของ convolutional layer จะเกิดปัญหา vanishing gradient จะทำให้โมเดล optimize ยาก และประสิทธิภาพในการเทรนจะแย่ลง เนื่องจาก gradient ที่ทำการปรับด้วยวิธี backpropagation นั้นเกิดมีค่าน้อยมาก ๆทำให้ weight ไม่ถูกปรับ โมเดลจึงไม่เกิดการเรียนรู้ใหม่ สามารถแก้ได้โดนการนำ input ของ convolution block (x) ก่อนหน้า มาบวกกับ output ของ block (F(x)) ปัจจุบัน ดังนั้น input ของ block ลำดับต่อไปคือ F(x)+x

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

2.1.6 Perplexity

เป็นวิธีการวัดผล neural network ว่ามีความสับสนแค่ไหนในการสร้างคำถัดไป เป็นค่าการกระจายของความน่าจะเป็นในการสร้างคำถัดไป ยิ่งค่าต่ำ ก็จะยิ่งมีความสับสนน้อย

$$PP(W) = P(w_1 w_2 \dots w_N)^{-\frac{1}{N}} = \sqrt[N]{\frac{1}{P(w_1 w_2 \dots w_N)}} \quad (2.2)$$

โดยที่ P(w) คือความน่าจะเป็นของคำที่อยู่ใน test set

2.1.7 Semantic Role Labeling

เป็นกระบวนการทำเครื่องหมายระบุบทบาททางความหมายของคำ เพื่อทำให้โมเดลสามารถเข้าใจหน้าที่ของคำในประโยคได้มากยิ่งขึ้น ซึ่งจะช่วยเพิ่มประสิทธิภาพในการเรียนรู้ของโมเดล

A rare herb that grows once a millenium is said to grant immortality	A rare<quality> herb<what> that grows<action> once a millenium<when> is said to grant<action> immortality<property>

ตัวอย่างประโยคจากชุดข้อมูล ก่อน(ซ้าย)และหลัง(ขวา) การทำ Semantic Role Labeling

2.1.8 Entity Anonymization

เป็นกระบวนการแทนที่คำที่หมายถึง คน สัตว์ สิ่งของหรือสถานที่ ด้วยสัญลักษณ์พื้นฐาน เพื่อให้โมเดลเรียนรู้ได้ง่ายยิ่งขึ้น

A rare herb that grows once a millenium is said to grant immortality. You aren't sure about that but you do know that herb is very tasty, and you don't know why everyone keeps trying to raid your garden once every thousand years	<ent0> that grows once a millenium is said to grant immortality. <ent1> aren't sure about that but <ent1> do know that <ent0> is very tasty, and <ent1> don't know why <ent2> keeps trying to raid your garden once every thousand years

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตัวอย่างประโยคจากชุดข้อมูล ก่อน(ซ้าย)และหลัง(ขวา) การทำ Entity Anonymization

2.1.9 Entity Reference Generation

เป็นกระบวนการเปลี่ยนเอ็นทิตี้ ให้กลับเป็นคำที่หมายถึง คน สัตว์ สิ่งของหรือสถานที่

<ent0> that grows once a millenium is said to grant immortality. <ent1> aren't sure about that but <ent1> do know that <ent0> is very tasty, and <ent1> don't know why <ent2> keeps trying to raid your garden once every thousand years.	A rare herb that grows once a millenium is said to grant immortality. You aren't sure about that but you do know that herb is very tasty, and you don't know why everyone keeps trying to raid your garden once every thousand years

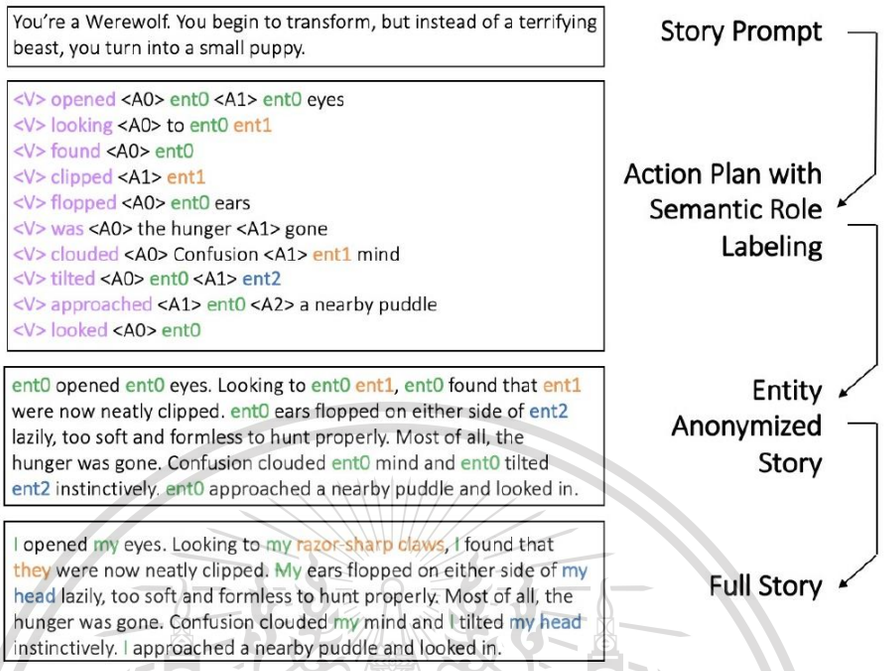
ตัวอย่างประโยคจากชุดข้อมูล ก่อน(ซ้าย)และหลัง(ขวา) การทำ Entity Reference Generation

2.2 งานวิจัยที่เกี่ยวข้อง

จากการศึกษางานวิจัยที่เกี่ยวกับ Text Generation มีงานวิจัยที่เกี่ยวข้องดังนี้

Angela Fan, Mike Lewis, Yann Dauphin (2019) งานวิจัยการสร้างเรื่องราวโดยมีแบบ ซึ่งได้ศึกษาเทคนิคต่างๆ ในการสร้างเรื่องราวในแต่ละขั้นตอน ขั้นตอนการทำงานของโมเดลคือ เมื่อป้อนคำกำกับบทเข้าไป โมเดลจะทำการสร้างแผนการกระทำ ซึ่งเป็นข้อความที่ที่มีการระบุบทบาททางความหมายไว้ และเอ็นทิตี้จะถูกแทนที่ด้วยสัญลักษณ์พื้นฐาน ต่อมาจะทำการสร้างเรื่องราวต้นแบบจากแผนการกระทำ ขั้นตอนตอนสุดท้ายจะทำการแทนที่เอ็นทิตี้ด้วยคำนามหรือสรรพนามโดยใช้วิธี Coreference-based เพื่อให้กลายเป็นเรื่องราวที่สมบูรณ์ในที่สุด

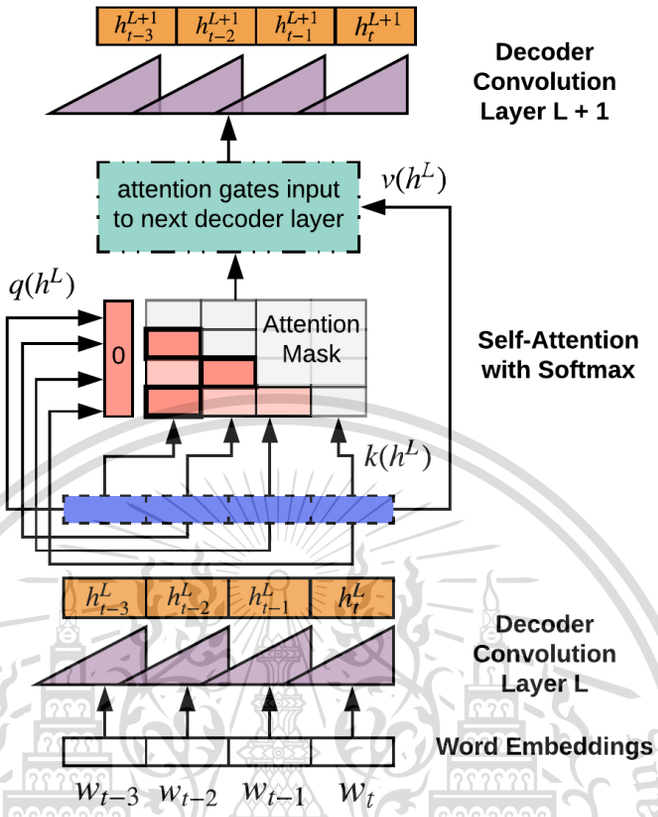
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.



รูป 2.4แผนภาพกระบวนการสร้างเรื่องราว

Angela Fan, Mike Lewis, Yann Dauphin (2018) งานวิจัยเกี่ยวกับการแต่งเนื้อเรื่องที่มีความต่อเนื่องและลื่นไหลด้วยการประยุกต์ใช้ Hierarchical Neural Network โดยมีชุดข้อมูลที่ใช้ในการวิจัยเป็นงานเขียนโดยมนุษย์จำนวน 300,000 เรื่องจากเว็ปไซต์ Hierarchical โมเดลในการแต่งเนื้อเรื่องจะทำการแต่งคำกำกับบทก่อน แล้วนำคำกับบทที่ได้ไปแต่งเนื้อเรื่องที่มีความเกี่ยวข้องกับกำกับบท งานวิจัยนี้ทำให้เนื้อเรื่องที่แต่งมีความต่อเนื่องโดยการใช้โมเดล convolutional sequence to sequence ที่เพิ่ม gated multi-scale self-attention mechanism เพื่อให้แต่งเนื้อเรื่องที่ยาวได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

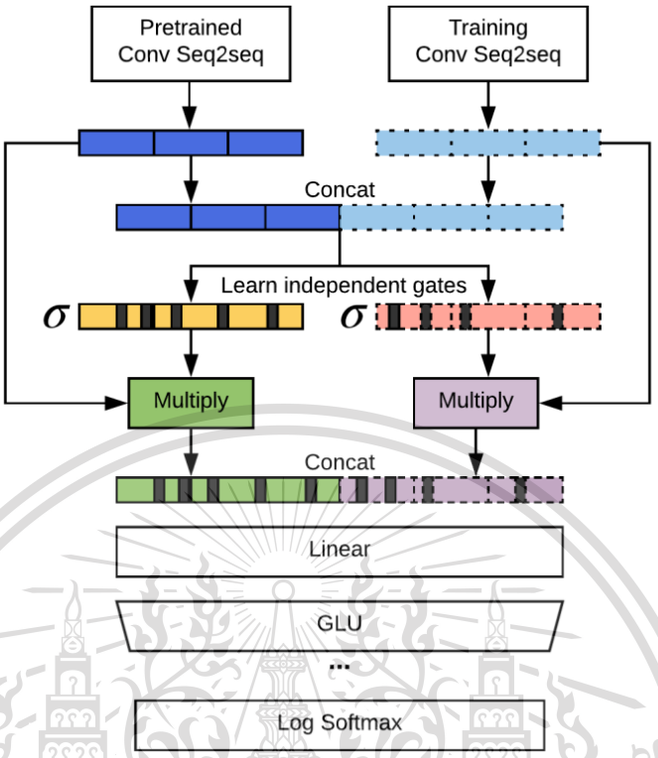


รูป 2.5Multihead self-attention mechanism

จากนั้นผู้ทำวิจัยได้นำเลเยอร์สุดท้ายที่ได้จาก โมเดล convolutional sequence to sequence แบบปรับปรุง มาเข้ากระบวนการ cold fusion กับโมเดลลักษณะแบบเดียวกันนี้ แต่ทำการเทรนไว้ก่อนแล้วเรียกว่า pretrain model กระบวนการนี้ทำขึ้น

เพื่อเสริมความสอดคล้องระหว่างเนื้อเรื่องกับหัวเรื่องโดย cold fusion จะนำเลเยอร์สุดท้ายของโมเดลทั้งสองตัวมาเข้า gates เพื่อหาความสำคัญและเลือกว่าส่วนใดระหว่างสองโมเดลนั้นที่จะนำมาใช้เป็นผลลัพธ์ ผลลัพธ์จะมีความสัมพันธ์ระหว่างเนื้อเรื่องกับหัวเรื่อง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.



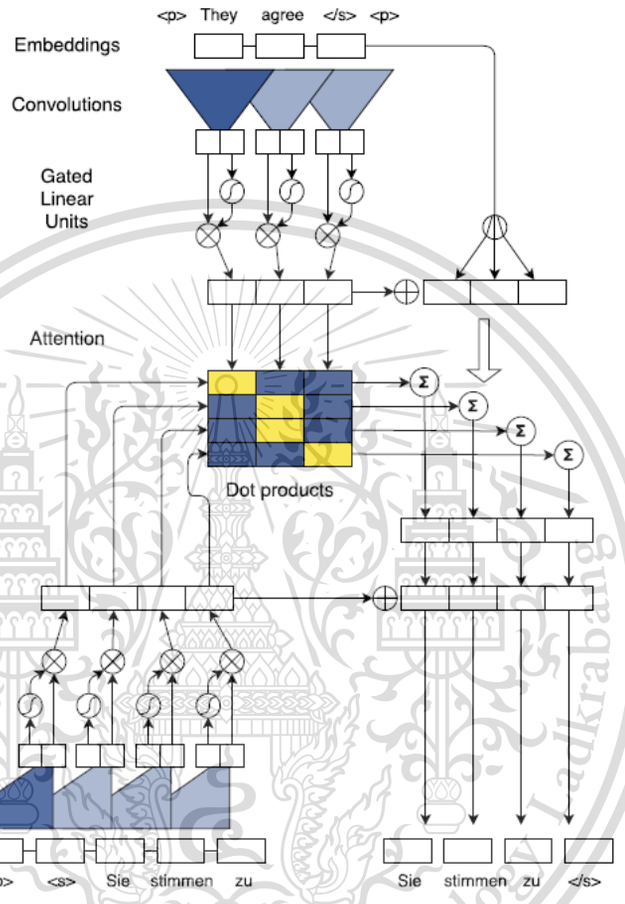
รูป 2.6Fusion Model

การทดสอบใช้ทั้งมนุษย์และโปรแกรมที่สร้างมาวัดผลเนื้อเรื่องที่ได้ ผลการวิจัยมนุษย์ที่เป็นผู้ตัดสินชอบเนื้อเรื่องที่แต่งโดย Hierarchical โมเดล มากกว่าเนื้อเรื่องที่แต่งโดยโมเดลที่เป็น Non Hierarchical

Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, Yann N. Dauphin (2017) งานวิจัยนี้เกี่ยวกับการทำ Neural machine translation โดยใช้ CNN แทนที่ RNN ด้วยข้อเสียของการใช้ RNN คือ การที่จะทำกระบวนการ NMT ได้ จะต้องนำประโยคมาใส่ในเวกเตอร์ที่กำหนดขนาดไว้ตายตัว แม้ว่าขนาดของประโยคจะสั้นหรือยาวแค่ไหนขนาดของเวกเตอร์ก็ไม่มีการเปลี่ยนแปลง และ RNN มีรูปแบบการทำงานเป็นลำดับ การจะเปลี่ยนรูปแบบการทำงานให้เป็นแบบขนานนั้นทำได้โดยยาก งานวิจัยนี้ได้นำโมเดลรูปแบบ CNN มาใช้ในกระบวนการ NMT ทำให้โมเดลนี้สามารถคำนวนแบบขนานได้ ทำให้สามารถใช้ประโยชน์จาก GPU ได้ดีขึ้นและทำการ optimization ง่ายขึ้น และนอกจากนี้ยังใช้งานคู่กับ Self-Attention และ Gated Linear Units ในการทำ NMT ของงานวิจัยนี้ ผู้วิจัยได้ปรับเปลี่ยนโครงสร้างของ Sequence to Sequence Model ส่วนของ encoder และ decoder มีการใช้ CNN มารับ input แต่ละ batch ผลลัพธ์ที่ออกมาจะเป็นความสัมพันธ์ของคำเหล่านั้นผ่านการคำนวณแบบ parallel ต่างจากการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ทำ sequence to sequence แบบปกติ จากนั้นผลลัพธ์ที่ได้ในแต่ละชุดจะนำไปเข้า GLU ปรับค่าใน residual connection และทำ attention



รูป 2.7แสดงการ batching ในระหว่าง training โมเดล Convolutional Sequence to Sequence

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บทที่ 3
การออกแบบ

โครงงานนี้ได้มีการออกแบบการทำงานโดยนำความรู้ทางด้านการเรียนรู้เชิงลึกของคอมพิวเตอร์มาประยุกต์ใช้เพื่อพัฒนาโมเดลการเรียนรู้เชิงลึกสำหรับแต่งงานเขียนประเภทเรื่องสั้นที่มีการดำเนินเนื้อเรื่องได้ โดยการศึกษาและพัฒนาจากงานวิจัยที่มีอยู่แล้วรวมกับความรู้เกี่ยวกับการเรียนรู้เชิงลึกที่ได้รับจากการเรียนในสถาบัน การออกแบบแบ่งส่วนออกเป็นทั้งหมดดังนี้

3.1 การวิจัยและภาพรวมของระบบ

3.1.1 Dataset preparation

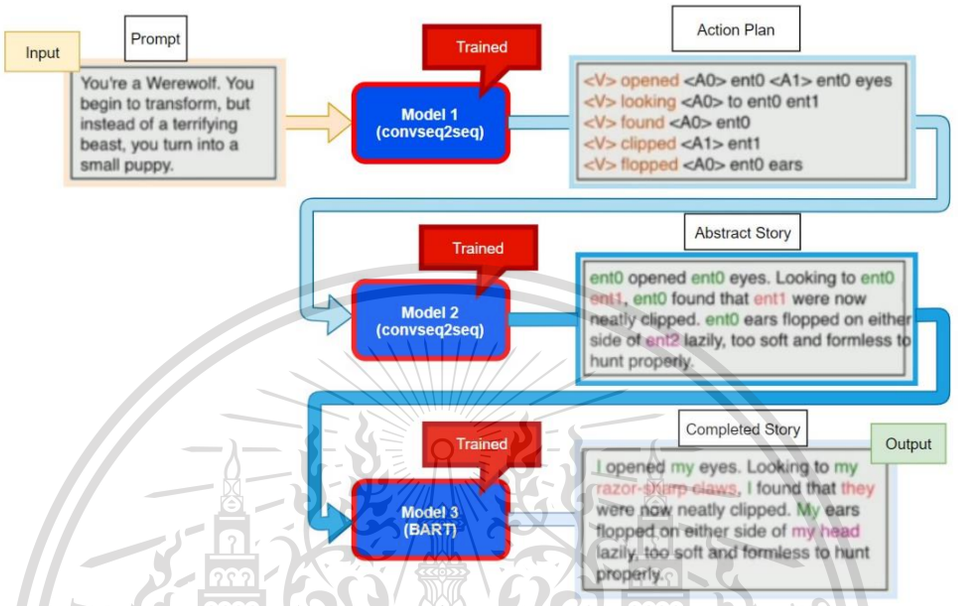
ชุดข้อมูล(Dataset) ที่ใช้ในงานโครงงานนี้ เป็นชุดข้อมูลที่รวบรวมมาจาก r/Writingprompt ซึ่งหนึ่งในเป็นสังคมออนไลน์บนเว็บไซต์ Reddit เป็นที่ที่คนจะเข้ามาเขียนเรื่องแต่งจากหัวข้อที่มีคนตั้งไว้ที่เรียกว่า คำกำกับบท(prompt) เราได้ใช้ praw และ pushshift ในการดาวน์โหลดข้อมูลจาก Reddit เราได้รวบรวมทั้งคำกำกับบทและเนื้อเรื่องที่คนแต่งจากคำกำกับบทนั้น ๆ มาใช้ในการสอนโมเดล โดยวรรณกรรมประเภทเรื่องสั้นที่แต่งขึ้นมาจะมีความยาวอย่างน้อย 100 คำ เราได้จัดเก็บชุดข้อมูลไว้ในรูปแบบตัวอักษร จากนั้นนำข้อมูลที่เป็นเรื่องสั้นไปผ่านโมเดลการเรียนรู้เชิงลึกที่ทำหน้าที่ Entities prediction เพื่อให้ได้ออกมาเป็นเรื่องสั้นที่แทนที่ตัวละครหรือสิ่งของในเรื่องนั้นด้วยคำว่า ent ตามด้วยเลขกำกับเพื่อให้สามารถจำแนกได้ เช่น ent0 ent1 เรียกชุดข้อมูลนี้ว่า Entities Anonymized Story ต่อมาจะนำ Entities Anonymized Story ไปผ่านโมเดลการเรียนรู้เชิงลึกที่ทำหน้าที่ Semantic Role Labeling prediction เพื่อหา Verb predicates สำหรับทำชุดข้อมูลที่ระบุว่า ent ตัวไหนทำกริยาอะไรกับ ent ตัวไหนบ้าง เรียกชุดข้อมูลที่ได้นี้ว่า Action plan สามารถบ่งบอกสิ่งที่เกิดขึ้นในเรื่องผ่านการกระทำของแต่ละตัวละครเป็นลำดับๆ

3.1.2 Preprocessing

เราได้ทำประมวลผลชุดข้อมูลของโครงงานเพื่อให้ชุดข้อมูลอยู่ในรูปแบบที่ง่ายต่อการป้อนเป็นอินพุตเพื่อสอนโมเดล โดยการตัดเครื่องหมายต่าง ๆ ออกไปให้เหลือแค่เพียงตัวอักษรภาษาอังกฤษ เปลี่ยนคำให้อยู่ในรูปเดียวกัน และตัดคำ(tokenize) เพื่อให้ง่ายต่อการป้องเข้าสู่โมเดลเพื่อใช้ในการเทรน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

3.2 การออกแบบโมเดลและขั้นตอนการทำงาน



รูป 3.1โมเดลการเรียนรู้เชิงลึกที่ใช้ในการแต่งเรื่องสั้น

(รูปที่ 3.1) เป็นภาพโมเดลการเรียนรู้เชิงลึกที่ผู้จัดทำใช้ในการแต่งเรื่องสั้น ผู้จัดทำได้นำโครงสร้างโมเดลมาจากงานวิจัย Angela Fan, Mike Lewis, Yann Dauphin (2018) โดยทำการดัดแปลงให้สามารถแต่งเรื่องสั้นขึ้นมาจาก Writing prompt หลายอันได้

โมเดลมีองค์ประกอบหลักสองส่วน คือ ส่วน Story Generation และ Prompt Relevance โดยขั้นตอนการทำงานของโมเดลสามารถอธิบายได้ดังนี้

การทำงานในส่วนของ Convolutional Sequence to Sequence (รูปที่ 2.6) ประกอบไปด้วย

3.2.1 Encoder

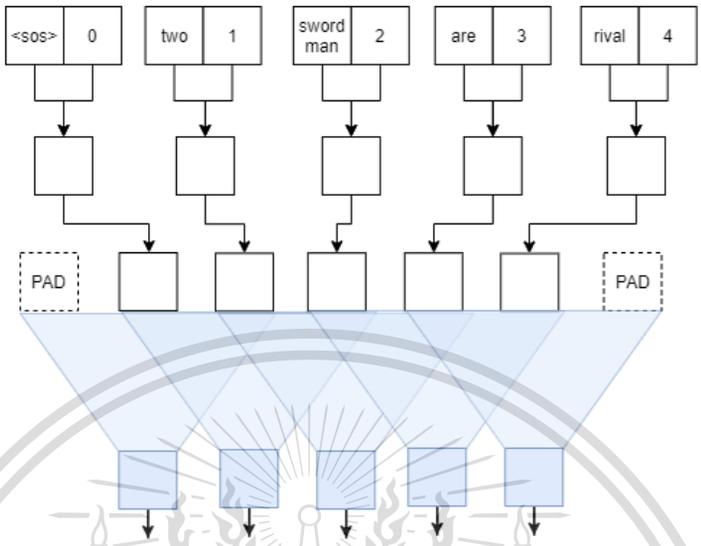
3.2.1.1 Embeddings

รับประโยคคำที่อยู่ในรูป word vector ที่ Embed ด้วยวิธี Word2vector เข้ามา แล้วทำ Positional Embedding โดยนำ word vector แต่ละตัวไปบวกรวมกับค่าตำแหน่งของมันในประโยคเองที่อยู่ในรูปเมทริกขนาดเท่ากัน

3.2.1.2 Convolutions

เป็น convolution แบบ 1 มิติ รับผลรวมของ word vector กับตำแหน่งของมันในประโยค จาก Embedding layer และมีการทำ padding ทางซ้ายและขวาของ vector ด้วยจำนวน (kernel - 1) / 2 เพื่อคงขนาดไว้หลังจากเข้า convolutional layer (รูปที่ 3.2)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.



รูป 3.2การทำ Padding ของ Encoder

3.2.1.3Gated Linear Units

รับ vector เข้ามาแบ่งเป็น 2 ส่วน A และ B นำ A เข้า sigmoid function จะทำหน้าที่เป็น Gate จะทำการตัดสินว่าแต่ละส่วนของ B มีความเกี่ยวข้องกับบริบทปัจจุบัน แล้วทำการคูณกับ B แบบ element-wise

3.2.1.4Residual connections

เป็นการป้องกันไม่ให้โมเดลเกิด vanishing gradient โดยนำ input ของ convolution block (x) ก่อนหน้า มาบวกกับผลที่ได้จาก GLU ที่ผ่านมา ของ block (F(x)) ปัจจุบัน ดังนั้น input ของ block ลำดับต่อไปคือ F(x)+x

3.2.2 Decoder

การทำงานหลักๆจะเหมือนกันกับ Encoder แต่มีรายละเอียดบางส่วนทีต่างกัน

3.2.2.1Embeddings

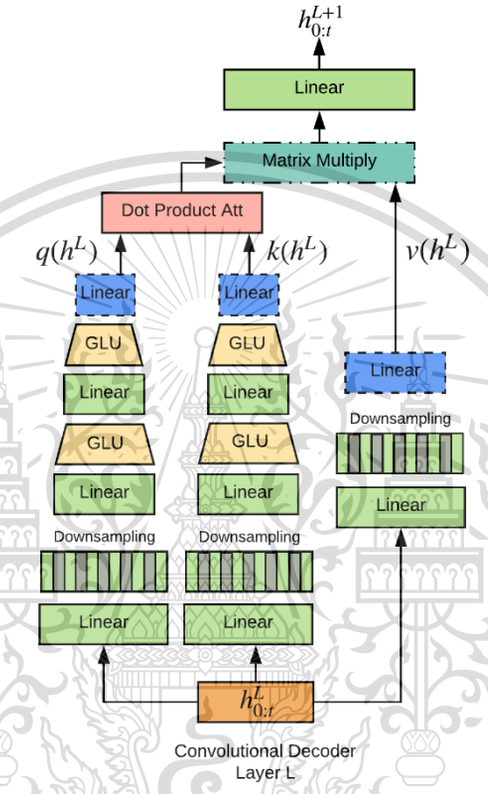
รับประโยคคำที่เป็น output ของ convolutional sequence to sequence จาก timestep ครั้งก่อนเข้ามา แล้วทำ Positional Embedding โดยนำ word vector แต่ละตัวไปบวกรวมกับค่าตำแหน่งของมันในประโยคเองที่อยู่ในรูปเมทริกขนาดเท่ากัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational purposes only. It is not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

3.2.2.2Multi-head Self-attention

กระบวนการนี้ (รูปที่ 2.4) จะรับ output จาก convolutional layer เข้ามา แล้วทำการ Downsampling แบ่งออกเป็นส่วนๆ ให้แต่ละ head ในแต่ละ head จะมี self-attention mechanism

(รูปที่ 3.3) head จะได้รับ input ที่ไม่เหมือนกัน head ที่ n จะได้รับ input ทุก ๆ n timestep head ที่1 จะได้ input ทุก timestep head ที่ 2 จะได้ input ทุก ๆ 2 timestep head ที่ 3 จะได้รับ input ทุก ๆ 3 timestep เป็นต้น

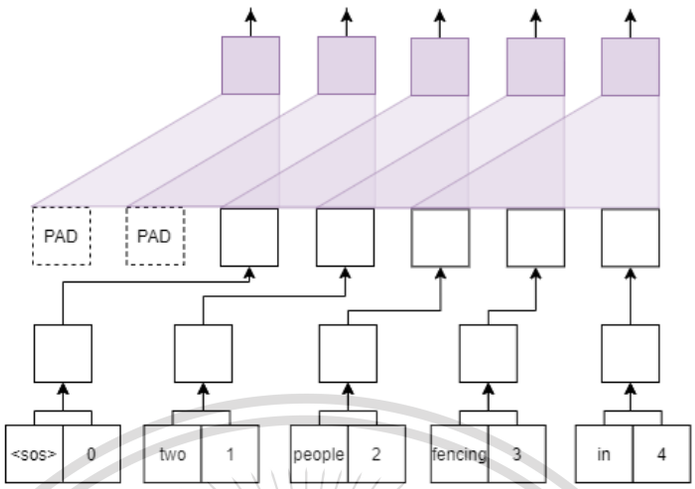


รูป 3.3Self-Attention Mechanism ที่อยู่ในแต่ละ Head ของ Multi-head Self-attention

3.2.2.3Convolutions

เป็น convolution แบบ 1 มิติ รับผลรวมของ word vector กับตำแหน่งของมันในประโยค จาก Embedding layer และมีการทำ padding ทางซ้ายของ vector ด้วยจำนวน kernel - 1 เพื่อกันไม่ให้ convolutional layer รับผลรวม word vector อันถัดไปเพราะต้องการให้โมเดลทำการเรียนรู้ที่จะเลือกคำถัดไป ดังรูป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.



รูป 3.4การทำ Padding ของ Decoder

3.2.3 Attention

ค่า attention เกิดจากการนำ output จากฝั่ง encoder และ decoder มา dot product กัน

3.2.4 Conditional decoder input

หาผลบวกระหว่างค่า output ของ encoder และ ค่า input ที่ผ่าน positional embedding แล้วนำค่าที่ได้ไปหาผลรวมถ่วงน้ำหนักกับค่า attention แล้วนำไปบวกกับ output ของ decoder

3.2.5 Softmax layer

นำ output จากขั้นที่แล้วมาผ่านฟังก์ชัน Softmax เพื่อเลือกคำที่จะสร้าง

การทำงานในส่วนของ Prompt Relevance (รูปที่) ประกอบไปด้วย

3.2.6 Concatenate

รับ vector จากสองที่ คือ layer สุดท้ายของ Pretrained Model กับ Training Model นำมา concatenate เข้าด้วยกัน

3.2.7 Learn independent gates

เป็น gate ที่จะเลือกว่า vector ส่วนไหนมีความเกี่ยวข้องกับบริบทปัจจุบันและควรนำมาใช้เป็นคำตอบ เมื่อเกิดการเรียนรู้โมเดลก็จะสามารถให้ความสำคัญกับจุดที่สำคัญจริง ๆได้ โดยนำ vector ที่มา concatenate กันแล้ว ไปผ่าน linear projection แล้วเข้า sigmoid function จากนั้นนำค่าที่ได้ไปคูณกับ vector ของแต่ละที่ นำมา concatenate ต่อด้วย linear layer network เพื่อทำ GLU และ log Softmax ตามลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

3.3 การออกแบบการทดลอง

เราได้ทำการสุ่มแบ่งชุดข้อมูล เป็น 3 ส่วน คือ Train set, Valid set, Test set โดยแบ่งออกตามอัตราส่วน 60:20:20 เราจะทำการทดลองโดยการป้อนชุดของ Writing Prompt เข้าไปเป็น Input ของ Fusion Model และ Model1 ใน Decomposition Model (รูป 3.1) ชุดข้อมูล Action Plan จะถูกป้อนให้ Model2 ใน Decomposition Model (รูป 3.1) และ Generic Story จะถูกป้อนให้ Model3 ใน Decomposition Model (รูป 3.1)

3.4 การวัดผลการทดลอง

ผลการทดลองจะทำการวัด Perplexity เป็นค่าความประหลาดใจ ใช้ในการวัดประสิทธิภาพของโมเดลในการเลือกคำออกมาใช้ได้อย่างเหมาะสม โดยผลลัพธ์ที่ได้จะทำการเทียบ Fusion Model ของงานวิจัย Hierarchical Neural Story Generation ที่ใช้อ้างอิง Angela Fan, Mike Lewis, Yann Dauphin (2018) เทียบกับ Decomposition Model แต่ละ Model ของงานวิจัย Angela Fan, Mike Lewis, Yann Dauphin (2019)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บทที่ 4
การทดลองและผลการทดลอง

ผู้จัดทำได้ตั้งเป้าหมายไว้ที่การทำโมเดลการเรียนรู้เชิงลึกที่ใช้ในการแต่งเรื่องสั้นโดยทำการสร้างโมเดลขึ้นมาสองแบบ แบบแรกคือโมเดล Convolutional Sequence to Sequence จำนวนสองตัว เชื่อมกันด้วย Fusion mechanism แบบที่สองคือ Decomposition model เป็นการแบ่งโมเดลออกเป็น 3 ส่วน (รูปที่ 3.1) ส่วนแรกและส่วนที่สองคือ Convolutional Sequence to Sequence ส่วนที่สามคือ BART ทำงานร่วมกันด้วยการรับ Output จากส่วนก่อนหน้ามาเป็น Input การทดสอบจะทำโดยใช้วิธีการวัดค่า Perplexity ของโมเดล เพื่อหาว่าโมเดลที่ได้มีความสับสนแค่ไหนในการสร้างคำถัดไป

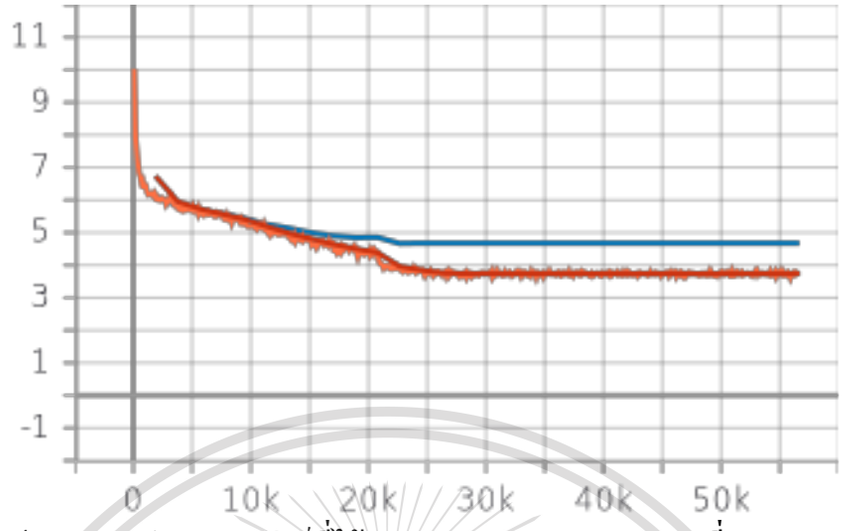
4.1 การสร้างโมเดล

ในการสร้างโมเดล ทุกโมเดลได้ใช้ Epoch = 30 รอบ ตามคำบอกของผู้จัดทำงานวิจัยที่ใช้อ้างอิง จากการเทรนและทดสอบโมเดล พบว่าหลังจาก Epoch ที่10 โมเดลหลายตัวจะเริ่มเกิด Overfitting เนื่องจากจำนวนข้อมูลที่ใช้เทรนโมเดลนั้นคือ 10,000 เรื่อง น้อยกว่ามากเมื่อเทียบกับงานวิจัยที่ใช้อ้างอิงซึ่งมีจำนวน 300,000 เรื่อง ชุดข้อมูลนำมาจากแหล่งข้อมูลเดียวกับงานวิจัยที่ใช้อ้างอิง Angela Fan, Mike Lewis, Yann Dauphin (2018) โดยรวบรวมมาประมาณ 50,000 เรื่อง ด้วยเวลาที่มีจำกัดทำให้สามารถนำข้อมูลไปประมวลผลเพื่อใช้ในการเทรนโมเดลได้เพียงแค่ 10,000 เรื่องเท่านั้น Fusion Model จะรับข้อมูลที่รวบรวมมาโดยตรง แต่ Decomposition model จะต้องสร้างโมเดล Entities prediction และ SRL prediction มาใช้สร้างชุดข้อมูลใหม่ก่อนนำไปใช้เทรน ได้ผลลัพธ์ดังนี้

4.1.1 Decomposition Model1

เป็นโมเดล Convolutional Sequence to Sequence ที่รับ input เป็นคำกำกับบท และมี output เป็น Action plan มีค่า loss อยู่ที่ Train = 3.737 Valid = 4.677

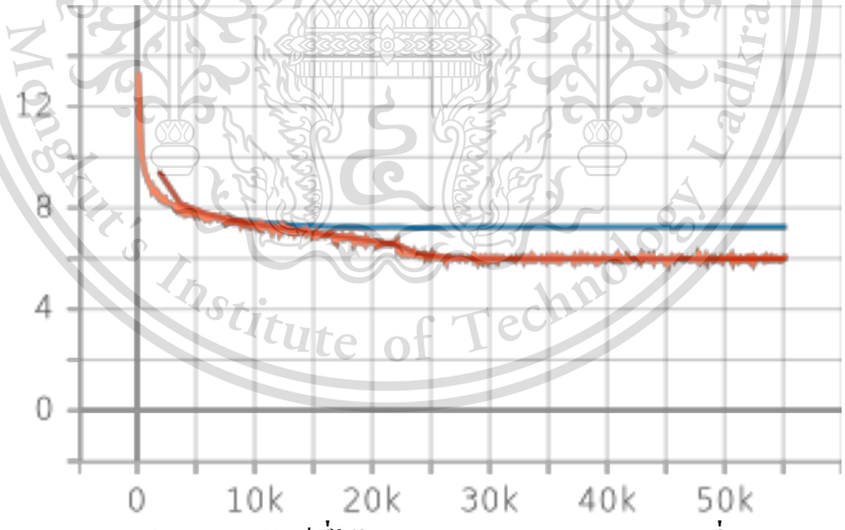
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.



รูป 4.1 กราฟแสดงผลลัพธ์ที่ได้ของ Decomposition model1 ที่ Epoch = 30

4.1.2 Decomposition Model2

เป็นโมเดล Convolutional Sequence to Sequence ที่รับ input เป็น Action plan และมี output เป็น Generic story มีค่า loss อยู่ที่ Train = 5.983 Valid = 7.242

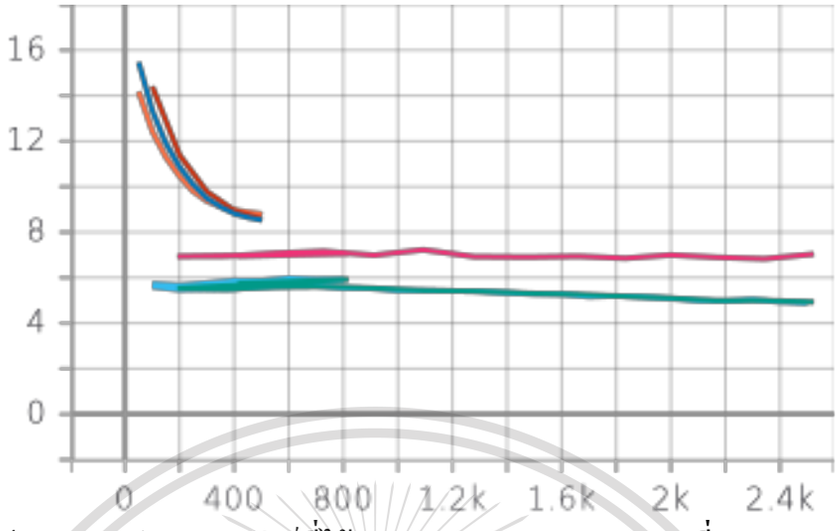


รูป 4.2 กราฟแสดงผลลัพธ์ที่ได้ของ Decomposition model2 ที่ Epoch = 30

4.1.3 Decomposition Model3

เป็นโมเดล BART ที่รับ input เป็น Generic story และมี output เป็น Story มีค่า loss อยู่ที่ Train = 4.931 Valid = 7.037

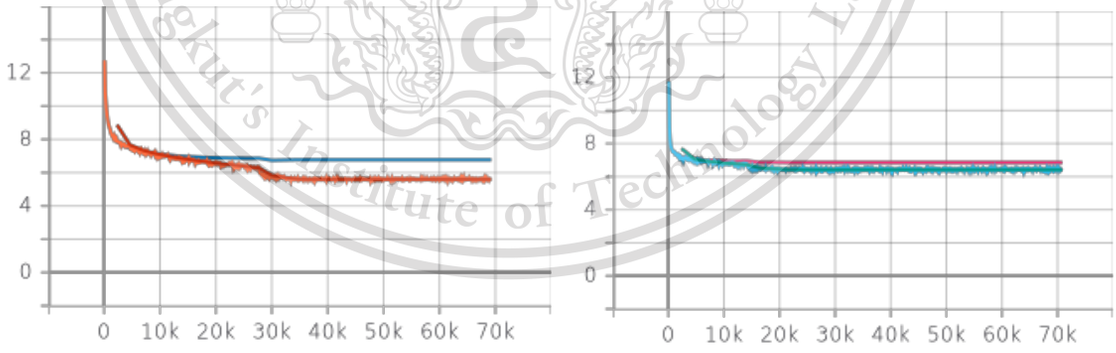
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.



รูป 4.3 กราฟแสดงผลลัพธ์ที่ได้ของ Decomposition model3 ที่ Epoch = 50

4.1.4 Fusion model

เป็นโมเดล Convolutional Sequence to Sequence 2 ตัว ที่เชื่อมกันด้วย Fusion mechanism รับ input เป็นคำกำกับบท และมี output เป็น Story โมเดลทำการเทรนเป็นลำดับ ลำดับแรกคือ Pretrained model มีค่า loss อยู่ที่ Train = 5.602 Valid = 6.773 ลำดับต่อมาทำการเทรน Training model ประกอบกับใช้ Pretrained model มีค่า loss อยู่ที่ Train = 6.417 Valid = 6.854



ก) ข)

รูป 4.4 กราฟแสดงผลลัพธ์ที่ได้ของ Fusion model ที่ Epoch = 30

ก) กราฟแสดงผลลัพธ์ของ Pretrained Model

ข) กราฟแสดงผลลัพธ์ของ Training Model

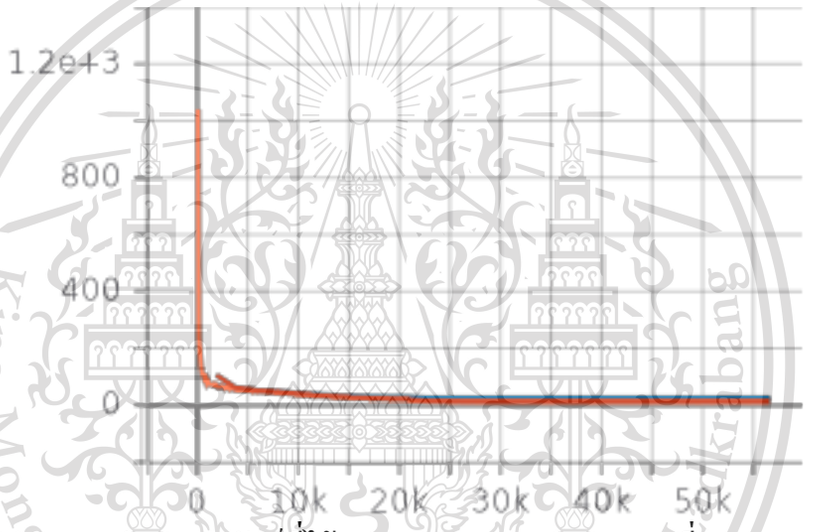
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

4.2 การทดสอบโมเดลด้วยข้อมูลที่รวบรวมมา

ในงานวิจัยที่ใช้อ้างอิง Angela Fan, Mike Lewis, Yann Dauphin (2018) ได้มีการนำ Test set ที่เป็นชุดข้อมูลของคำกำกับบทมาป้อนให้กับ Fusion model ที่ผ่านการเทรนแล้ว เพื่อวัดค่า Perplexity หรือค่าความสับสนที่ได้จากโมเดล เมื่อโมเดลทำการเลือกคำที่จะใช้สร้าง ในโครงงานนี้จะทำการหาค่า Perplexity ของทุกโมเดล จากข้อมูล Valid set ได้ผลลัพธ์ดังนี้

4.2.1 Decomposition Model1

มีค่า Perplexity อยู่ที่ Train = 13.33 Valid = 25.58

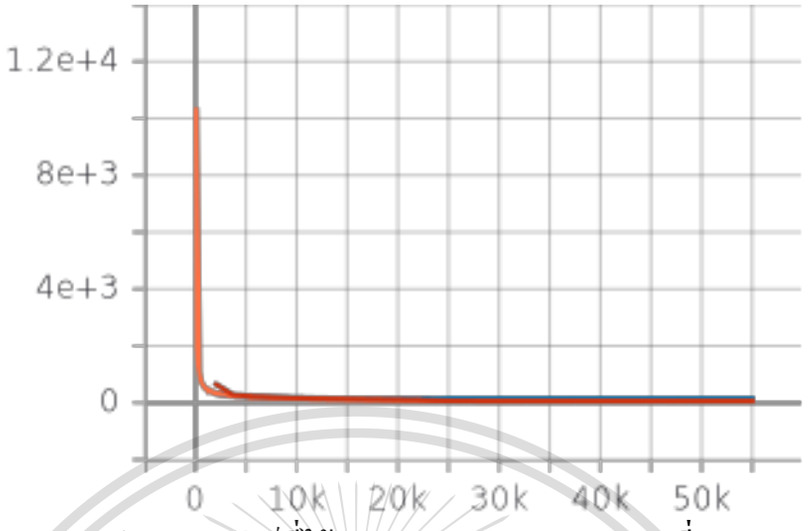


รูป 4.5 กราฟแสดงผลลัพธ์ที่ได้ของ Decomposition model1 ที่ Epoch = 30

4.2.2 Decomposition Model2

มีค่า Perplexity อยู่ที่ Train = 63.27 Valid = 151.4

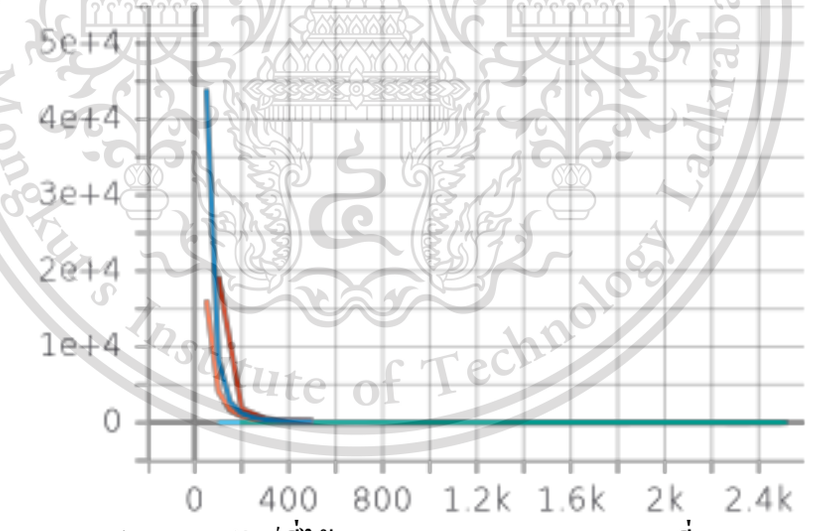
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.



รูป 4.6 กราฟแสดงผลลัพธ์ที่ได้ของ Decomposition model2 ที่ Epoch = 30

4.2.3 Decomposition Model3

มีค่า Perplexity อยู่ที่ Train = 10.4 Valid = 51.35

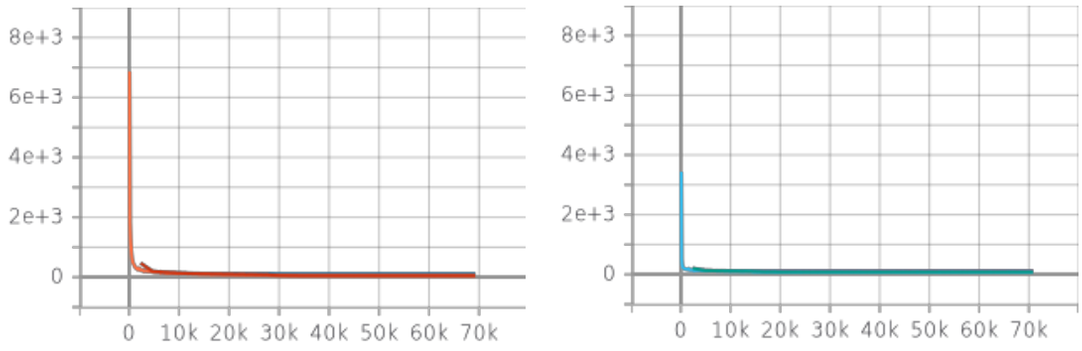


รูป 4.7 กราฟแสดงผลลัพธ์ที่ได้ของ Decomposition model3 ที่ Epoch = 50

4.2.1 Fusion model

Pretrained model มีค่า Perplexity อยู่ที่ Train = 85.43 Valid = 115.7 ลำดับต่อมาทำการเทรน Training model ประกอบกับใช้ Pretrained model มีค่า Perplexity อยู่ที่ Train = 48.59 Valid = 109.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.



ก) ข)

รูป 4.8 กราฟแสดงผลลัพธ์ที่ได้ของ Fusion model ที่ Epoch = 30

ค) กราฟแสดงผลลัพธ์ของ Pretrained Model

ง) กราฟแสดงผลลัพธ์ของ Training Model

4.3 สรุปผลการทดสอบโมเดลโดยใช้ข้อมูลที่รวบรวมมา

ผลการทดสอบวัดค่า Perplexity โมเดล ได้ผลลัพธ์ดังนี้

ตาราง 4.1ค่า Perplexity ที่ได้จากโมเดลการเรียนรู้เชิงลึกที่ใช้ในการแต่งเรื่องสั้นโดยมีการแบ่งเนื้อเรื่องออกเป็น 3 ส่วน เทียบกับโมเดล Hierarchical Neural Story Generation

Model	**Valid Perplexity**
Decomposition Model1: Conv seq2seq + self-attention	25.58
Decomposition Model2: Conv seq2seq + self-attention	151.4
Decomposition Model3: BART	51.35
Fusion: Conv seq2seq + self-attention (Pretrained model)	109.3
Fusion: Conv seq2seq + self-attention (Training model)	115.7

จากตาราง 4.1 Valid Perplexity คือค่าความสับสนที่ได้จากการทดสอบด้วยชุดข้อมูล Valid

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บทที่ 5
สรุป

5.1 สรุปผลการทำงาน

การดำเนินงานในระยะเวลา 2 ภาคเรียนที่ผ่านมานั้น ประกอบไปด้วย การศึกษางานวิจัยเกี่ยวกับการแต่งวรรณกรรมประเภทเรื่องสั้นและการสร้างคำโดยอาศัยการเรียนรู้เชิงลึกควบคู่กับกระบวนการอื่น ๆ หลักจากนั้นจึงทำการรวบรวมข้อมูลและสร้างชุดข้อมูลเพื่อใช้ในการเทรนโมเดล และได้นำชุดข้อมูลที่มีอยู่เดิมสร้างออกมาเป็นชุดข้อมูลใหม่ด้วยการนำข้อมูลไปประมวลผลกับโมเดลการเรียนรู้เชิงลึกคือ Entities prediction ได้ออกมาเป็นชุดข้อมูล Entities Anonymized และ Semantic Role Labeling prediction ได้ออกมาเป็นชุดข้อมูล Action plan จากนั้นนำชุดข้อมูลทั้งหมดที่มีไปใช้เทรนโมเดลแต่ละส่วน โมเดลสร้างขึ้นมาสองรูปแบบ แบบแรกเป็น Decomposition คือการแบ่งหน้าที่ออกเป็นส่วนๆ และเทรนโมเดลตามหน้าที่เหล่านั้น แบบที่สองคือ Fusion model จากงานวิจัยก่อนคือ Hierarchical Neural Story Generation หลังจากเทรนเสร็จก็จะมาทำการทดสอบเทียบผลลัพธ์ด้วยการวัดค่า Perplexity ผลที่ได้แสดงให้เห็นว่าจำนวนชุดข้อมูลที่มีอยู่นั้นน้อยเกินไป โมเดลจะเกิดการ overfitting ได้เร็ว และได้ผลลัพธ์ออกมาไม่น่าประทับใจนัก แต่ผลลัพธ์บางส่วนที่ออกมาแสดงให้เห็นว่าโมเดล Decomposition มีการดำเนินเรื่องต่างจาก Fusion model อยู่เล็กน้อย เนื่องจาก input ของ Decomposition model มีความเป็นโครงสร้างมากกว่า Fusion model

5.2 ปัญหาที่เกิดขึ้นและแนวทางแก้ไข

1. การรวบรวมข้อมูลเพื่อทำ Data Set ด้วย Reddit API มีข้อจำกัดในการดึงข้อมูลอยู่ที่ 100 ชุดต่อการเรียกใช้ 1 ครั้ง
   วิธีแก้ปัญหาคือ ใช้แพ็คเกจที่มีชื่อว่า PRAW (The Python Reddit API Wrapper) ด้วยแพ็คเก็จนี้ ทำให้สามารถดึงข้อมูลจาก Reddit โดยระบุช่วงเวลาของกระทู้ที่ต้องการได้ ทำให้สามารถรวบรวมข้อมูลได้เพียงพอต่อความต้องการ
2. การสร้างโมเดลการเรียนรู้เชิงลึกโดยใช้โครงสร้างโมเดลจากงานวิจัยที่ใช้อ้างอิง Angela Fan, Mike Lewis, Yann Dauphin (2018) นั้นทำได้ยากและต้องใช้เวลานาน เนื่องจากโมเดลมีความซับซ้อน ความยากในการเขียนโปรแกรมเพื่อสร้างโมเดลรูปแบบนี้จึงมากตามความซับซ้อน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

วิธีแก้ปัญหาคือ ใช้ Sequence Modelling Toolkit ที่มีชื่อว่า Fairseq เป็นชุดเครื่องมือที่พัฒนาขึ้นโดยนักวิจัยจาก Facebook Research กลุ่มเดียวกับที่ทำงานวิจัยที่ใช้อ้างอิงในโครงงานนี้ ชุดเครื่องมือถูกพัฒนาขึ้นเพื่อช่วยเหลือในการทำวิจัยและพัฒนาโมเดลรูปแบบต่าง ๆที่ใช้ทำงานเกี่ยวกับ Translation, Summarization, Language modeling และ Text generation

3. การประมวลผลในการเทรนโมเดลและจัดทำชุดข้อมูลเป็นภาระต่อหน่วยประมวลผลเป็นอย่างมาก หน่วยประมวลผล GPU ที่มีอยู่นั้นไม่สามารถตอบโจทย์ในการทำงานได้ ผู้จัดทำจึงใช้บริการ Cloud ของ Google Colab Pro เพราะมีหน่วยประมวลผล GPU ที่มีประสิทธิภาพและค่าใช้จ่ายไม่แพง
4. โมเดลการเรียนรู้เชิงลึกได้แก่ Entities prediction และ SRL prediction ที่นำมาใช้ในการสร้างข้อมูลสำหรับเทรนโมเดล ไม่มีการเขียนเอกสารการใช้งานที่ชัดเจนและเป็นเวอร์ชั่นเก่า ทำให้มีปัญหากับการใช้งานกับซอฟต์แวร์เวอร์ชั่นใหม่ตัวอื่นๆหรือมีปัญหากับ environment ใน cloud
   วิธีแก้ปัญหา คือ อ่านและสืบค้นวิธีแก้ปัญหาจากกระทู้เก่าๆ ที่เคยมีคนถามไว้แล้วใน GitHub repository ของซอฟต์แวร์นั้น ๆ และใช้วิธี Downgrade library และ tool ที่ใช้เพื่อให้ตรงกับความต้องการของซอฟแวร์เหล่านี้
5. โมเดลการเรียนรู้เชิงลึกที่ใช้ผ่านซอฟแวร์เวอร์ชั่นเก่าที่ทำการ Downgrade library และ tool ไม่สามารถใช้งานหน่วยประมวลผล GPU ได้ เนื่องจาก environment ใน cloud บางส่วนที่จำเป็นกับการใช้ GPU ไม่สามารถทำการ downgrade ตามลงมาได้ ส่งผลให้การจัดทำชุดข้อมูลกินเวลานานเกินไป แก้ปัญหาด้วยการใช้โมเดลการเรียนรู้เชิงลึกกับซอฟแวร์ตัวอื่นที่ใหม่กว่า เพื่อให้สามารถจัดทำชุดข้อมูลได้เร็วกว่าเดิม
6. ผู้เขียนงานวิจัย Decomposition model Angela Fan, Mike Lewis, Yann Dauphin (2019) นั้น ยังไม่ได้ทำการเผยแพร่ Source code ของโมเดล ทำให้ผู้จัดทำต้องหาวิธี implement กันเองตามคำอธิบายของผู้เขียนงานวิจัยนั้น แต่ทว่าโมเดลมีความซับซ้อนมากเกินไป และไม่มีข้อมูลอธิบายประกอบโครงสร้างการทำงานของโค้ดใน Toolkit อย่างเพียงพอ ด้วยเวลาทำงานที่มีอยู่อย่างจำกัด ผู้จัดทำจึงแก้ปัญหาโดยการใช้โมเดลรูปแบบธรรมดาไปก่อน ซึ่งผลลัพธ์ที่ได้ออกมาจะไม่ค่อยดีตามการคาดการณ์ ในอนาคตเมื่อผู้เขียนงานวิจัยได้ทำการเผยแพร่ source code ของ Decomposition model แล้ว จะสามารถเอามาใช้งานแทนที่โมเดลเดิมได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ ดัดแปลง เผยแพร่ หรือนำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

7. ในการใช้งาน Google Colab นั้น ผู้จัดทำได้ทำการ mount Google Drive เพื่อใช้เป็นแหล่งเก็บข้อมูล เมื่อใช้งานเป็นปริมาณมากจึงทำให้ทราบว่า Google Drive มีการจำกัดโควต้าปริมาณข้อมูลที่ทำการ Download หรือ Upload ในแต่ละวัน โดยปริมาณของข้อมูลอยู่ที่เท่าไหร่นั้น ทีมงาน Google ไม่ได้แจ้งไว้อย่างชัดเจน ผู้ใช้งานท่านอื่นๆที่แลกเปลี่ยนความคิดเห็นในเรื่องนี้ระบุว่าในแต่ละวันจำนวนโควต้าจะไม่เท่ากัน เมื่อโควตาเต็มจะต้องรอเป็นเวลา 24 ชั่วโมง ถึงจะใช้งาน Google Drive ได้อีกครั้ง เพื่อไม่ให้เกิดปัญหานี้ขึ้นอีก ผู้จัดทำจึงทำการย้ายไฟล์เท่าที่จำเป็น ไปประมวลผลที่ VM ของ Google Colab แล้วจึงย้ายผลลัพธ์ที่ได้ลงใน Google Drive อีกทีหนึ่ง และทำการใช้คำสั่งใน Toolkit ให้บันทึกข้อมูล Checkpoint ของโมเดลเท่าที่จำเป็นเท่านั้น เพื่อลดปริมาณข้อมูลที่ใช้จาก Google Drive

5.3 แนวทางการพัฒนาต่อ

จากการสังเกตชุดข้อมูล ผู้จัดทำพบว่าข้อมูลที่รวบรวมมาในชุดข้อมูลนั้นยังมีปัญหาอยู่หลายจุด เช่น ชุดข้อมูล Entities Anonymized ที่ได้จากโมเดล Entities prediction มีการจับตัวละครในเรื่องได้ไม่สมบูรณ์ หน้าที่ของโมเดลคือการเปลี่ยนคำที่บ่งบอกถึงคน สัตว์ สิ่งของ หรือสถานที่นั้นๆ ให้กลายเป็นคำว่า ent ผลลัพธ์ที่ออกมา ยังมีบางคำไม่ได้รับการแปลงเป็น ent และบางคำเป็นการแปลผิดอีกด้วย หากนำโมเดลนี้มาพัฒนาต่อก็จะสามารถสร้างชุดข้อมูลที่มีคุณภาพมากขึ้นได้ และทำให้ผลลัพธ์ของโมเดลที่ได้รับการเทรน ดีขึ้นอีกด้วย นอกจากนี้ โมเดลที่ใช้สร้างชุดข้อมูลใช้เวลาสร้างโดยประมาณ 1500 เรื่อง ต่อ 3 ชั่วโมง หากใช้หน่วยประมวลผลที่มีประสิทธิภาพมากกว่านี้ ก็จะทำให้มีชุดข้อมูลสำหรับเทรนโมเดลเพิ่มขึ้นเพราะใช้เวลาในการสร้างชุดข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บรรณานุกรม

[1] Angela Fan, Mike Lewis, Yann Dauphin. 2018. Hierarchical Neural Story Generation. **arXiv preprint.** arXiv:1805.04833.

[2] Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann Dauphin. 2017. Convolutional sequence to sequence learning. **arXiv preprint.** arXiv:1705.03122.

[3] Chamanth mvs. 2020. **Word Embedding (What? and Why?)**. [Online]. Available : https://medium.com/analytics-vidhya/word-embedding-why-and-what-dcf5c42724d2.

[4] Sumit Saha. 2018. **A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way**. [Online]. Available : https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53.

[5] Alibaba Cloud. 2018. **Self-Attention Mechanisms in Natural Language Processing**. [Online]. Available : https://alibaba-cloud.medium.com/self-attention-mechanisms-in-natural-language-processing-9f28315ff905.

[6] Lei Mao. 2020. **Gated Linear Units (GLU) and Gated CNN**. [Online]. Available : https://leimao.github.io/blog/Gated-Linear-Units/.

[7] Natthawat Phongchit. 2020. **มาทำความรู้จัก ResNet กันดีกว่า**. [Online]. Available : https://medium.com/@natthawatphongchit/มาทำความรู้จัก-resnet-กันดีกว่า-aec3a8c10793.

[8] Dan Jurafsky. 2019. **Language Modeling**. [Online]. Available :.124/lec/languagemodeling.pdf.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

[9] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, Luke Zettlemoyer.2019 .BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension **arXiv preprint.** arXiv:1910.13461

[10] Angela Fan, Mike Lewis, Yann Dauphin. 2019.Strategies for Structuring Story Generation. **arXiv preprint.** arXiv: 1902.01109

[11] Kenton Lee, Luheng He, Luke Zettlemoyer. Higher-order Coreference Resolution with Coarse-to-fine Inference. **arXiv preprint.** arXiv: 1804.05392

[12] Luheng He, Kenton Lee, Mike Lewis, Luke Zettlemoyer. 2017. Deep Semantic Role Labeling: What Works and What's Next. 473–483 in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Vancouver: Association for Computational Linguistics

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.