

Text Summarization

1. Full experiment on ThaiSum dataset
2. Models: Text rank, TF-IDF, mBART, MT-5
3. Experimental results : ROUGE-1/2/L, F1-Score , Train/Valid. Loss, 10 concrete examples.
4. Machine: Google Colab **Pro**

Proposed method

Preprocess: Finetune pretrained LLMs (mBART) โดยใช้ ThaiSum dataset

1. หา Text rank (หรือ TF-IDF หรือ Name Entity) ของข่าวที่ต้องการสรุป
2. คัดเอาส่วนของข่าวที่คำในนั้นมีค่า Text rank (หรือ TF-IDF หรือส่วนของข่าวที่มี Name Entity เป็นชื่อคน, สถานที่, องค์กร, วันที่, จำนวนเงิน) รวมกันน้อยกว่าค่า Threshold (ต้องมีการทดลองหาค่า Threshold ที่เหมาะสม)
3. นำข่าวที่กรองแล้วไปเข้า Finetuned LLMs

Chatbot

1. Full experiment on ThaiQA dataset
2. Models: LSTM, Flex GPT-2-base (PadThai), WangchanBERTa, WangchanGLM
3. Needed experimental results : ROUGE-1/2/L, BLEU, F1-Score, Train/Valid. Loss, 10 concrete examples.
4. Machine: Google Colab **Pro**

Proposed method

Preprocess: Finetune pretrained LLMs (WangchanBERTa) โดยใช้ ThaiQA dataset

1. หา vector ของ Question และ documents แต่ละ document โดยใช้ doc2vec
2. เปรียบเทียบ similarity ของ Question vector กับแต่ละ document vector
3. เลือก document vectors ที่คล้ายกับ Question vector มากที่สุด 3 document vectors
4. เรียกใช้งาน QuestionAnsweringModel ที่ finetuned แล้ว 3 ครั้ง โดยแต่ละครั้งให้ส่ง document (text) แต่ละตัวที่สอดคล้องกับ document vector ที่เลือกมา 3 ตัว ไปเป็น context และ ส่ง Question (text) ไปเป็น question
5. เลือกคำตอบจาก QuestionAnsweringModel ที่มีค่าความน่าจะเป็นสูงที่สุดมาเป็นคำตอบ