Knowledge Update in Full Text Triggered by a News Event

新聞事件觸發之全文知識更新

Author: Yu-Ting Lee

Department of Computer Science, National Chengchi University

Advised by Prof. Tsai-Yen Li and Prof. Hen-Hsen Huang

Background

- Events around the world happen within every second.
- Existing knowledge may need to be updated (e.g. books, web pages, etc.).
- Keeping the knowledge up-to-date is not a trivial task.
- A news event is considered as a trigger of knowledge update.

Background

Human editors are employed to perform knowledge update.

Disadvantages of updating knowledge by editors:

- Challenges of extracting overviews over time.
- Needing time to determine revised contents and follow update-patterns.
- Having sufficient domain knowledge to update knowledge.
- Hard to take advantage of long-term memory

Background

Main points of knowledge update

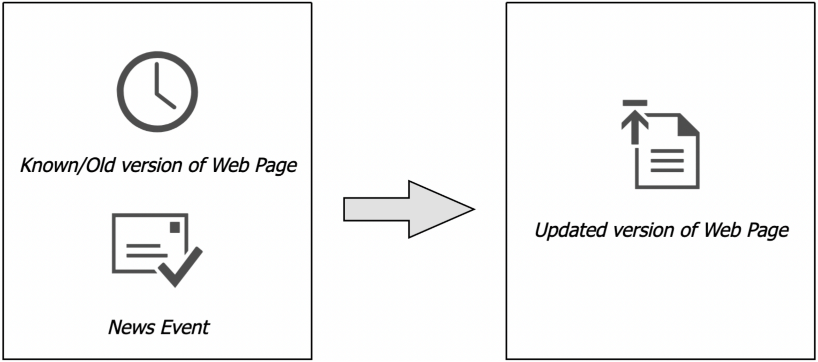
- Finding the salient information from revisions of knowledge sources
- Article updates are required to be predictable, and humans are able to discern.
- Updated patterns vary in news with different events/topics.
- Providing overview of previous revisions.

Background

Wikipedia

- Open to access
- Maintained by over 10,000 professional editors
- With more than 65 million English version contents
- Extending Wikipedia Current Event Portal dataset (WCEP) with Wikipedia pages.

Introduction



Introduction

Knowledge Update

Essentials

- Overviews of previous contents.
- Main points of triggered news event.

Paragraph

- May be longer than summary-level contents
- May not be revised during updating

Introduction

Example

Example
Non-updated Paragraph: On 2 August there were 15 new cases of COVID-19, 2 overseas acquired. Consequently, the South-east Queensland ’ s lockdown was extended until 4:00pm on 8 August (Sunday). The same day, because of the extension, the Ekka agricultural show was cancelled for the second year, 5 days before it was to be open to the public from 7 August (Saturday). <Timeline - Brisbane lockdowns>
Triggered News: Cairns and Yarrabah enter a snap three-day lockdown after an ”unexpected” case of COVID-19 was reported in a taxi driver from Kanimbla who was infectious in Far North Queensland for 10 days.
Updated Paragraph: On 2 August, South-east Queensland reported a spike in COVID-19 cases, leading to an extended lockdown until 8 August. This caused the Ekka agricultural show to cancel its 7 August public opening for the second consecutive year. Additionally, Cairns and Yarrabah faced a sudden three-day lockdown due to an unexpected case in a Kanimbla taxi driver, who was infectious for 10 days. These events underscore the persistent challenges in managing the pandemic. <Timeline - Brisbane lockdowns>

Introduction

Objectives

Balancing the information between new and old version contents:

- Determine whether paragraph is needed to be updated.
- Paragraphs in dataset should be labeled.
- Generate updated paragraphs for update-needed paragraphs.
- Merge the non-updated and the updated paragraphs to form an updated article.

Introduction

Contributions

Knowledge update for long input texts:

- Fine-grained content rewriting.
- Suggestions for paragraph rewriting.
- Automated paragraphs selection and updating.
- Unlimited length for full article input if all paragraphs are under 4,096 tokens.

Related Work

Long Context Summarization: PRIMERA model

- [Xiao et al. 2021]proposed sentence selection method to extract salient sentences for training sentence-prediction model.
- Being able to extract salient information from long text inputs to form summaries.
- More faster to extract salient information

Related Work

Dynamic Contents Generation on Headlines and Summaries

- Headline revisions is important to capture differences between article contents. [Panthaplackel et al. 2022]
- Capturing salient information from evolving news in different text lengths

Differences between headlines and summaries

- Lengths (average length: 10 v.s. 72.26; 175.56; 3,968.81)
- Completeness of sentences
- Key points of overall input articles or specific event update
- Format matching

Related Work

Wikipedia Current Events Portal (WCEP) dataset:

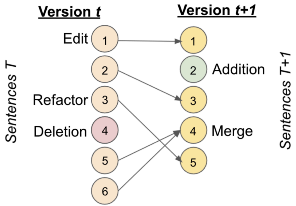
1. Consists of daily news and the topic that is related to news from 2006 to 2016.
2. Each portal contains human-written summary with cited pages, and a news article.
3. Types of events ranging from Disaster, Economic, Politics, Health, etc.
4. Wikipedia (English version) is maintained by over 10,000 editors everyday.

Reference: [Demian Gholipour Ghalandari et al. 2020]

Related Work

Asymmetrical sentence-matching algorithm

- Syntax in sentences may changed but the semantics is not changed.
- Sentences may be merged or splitted during updating.
- Sentences may be refactored according to changes on importance during updating.



Reference: [Spangher et al. 2022]

Dataset

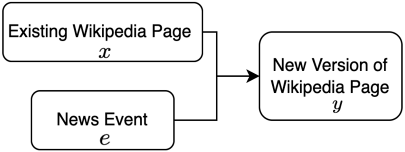
A Multi-grained Dataset for News Event Triggered Knowledge Update (NetKu)

Dataset consists of (e, x, y) triples at different granularities:

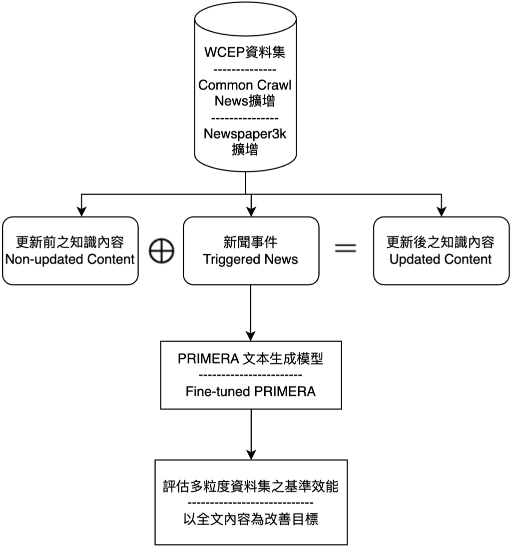
- Levels of news event (citation texts, first paragraph, full source article).

Revisions

- Some news events are not important enough for revisions.
- Some contents remain unchanged between revisions.



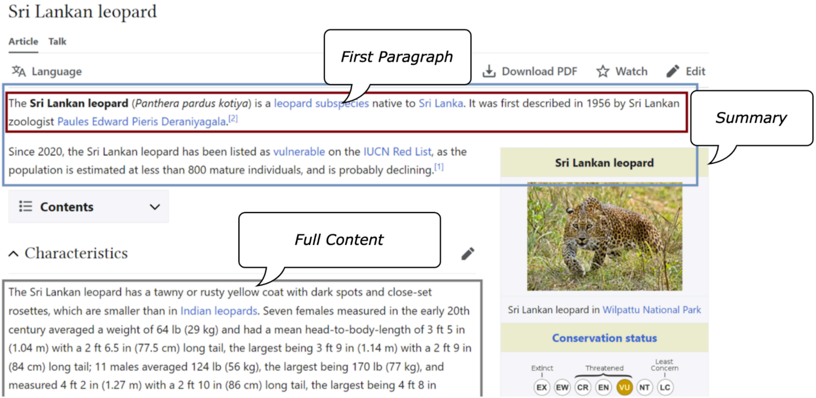
Dataset


WCEP資料集
Common Crawl News擴增
Newspaper3k擴增
更新前之知識內容
Non-updated Content
新聞事件
Triggered News
更新後之知識內容
Updated Content
PRIMERA 文本生成模型
Fine-tuned PRIMERA
評估多粒度資料集之基準效能
以全文內容為改善目標

Dataset

Multi-grained Knowledge Facts

Example of an Wikipedia article: Three levels of (x, y) construction



Dataset

Dataset Construction

- Pairs of (e, x, y).

Armed conflicts and attacks

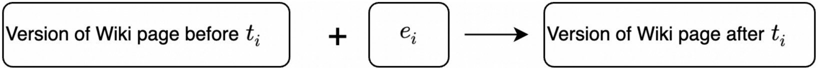
- Russo-Ukrainian War
  - Russian State Duma deputies approve the introduction of the concepts of "mobilization", "martial law", "wartime" and "armed conflict", as well as punishment for desertion, into the Criminal Code. (RT)

- Multiple source pages may pair with the same summary, and accumulating sources from Internet Archive Wayback Machine.
- News Event e:
  1. Citation Text
  2. First paragraph of the news article
  3. Full text of the news article

Dataset

Data Alignment

Given e_i listed on the Wikipedia Current Event Portal at time t_i



First version that cited e_i retrieved as y_i, and the previous version of y_i as x_i.

To avoid the irrelevant edits involved in the revisions, time window is restricted to one week between (e_i, y_i).

- x_i is considered as the original Wikipedia page before the occurrence of the event e_i.
- y_i is the updated version according to e_i because of its citation of e_i.

Dataset

Data Filtering

Overlapped instances: Some instances are overlapped with those in WCEP-10.

Preprocessing

- Keeping the instances with existing x_i.
- Keeping the instances with meaningful y_i (i.e. length>10).
- Keeping only English instances.

Dataset

Data Filtering

Statistics

Training data 1906 $\rightarrow 1602$
Testing data 239 $\rightarrow 201$
Validation data 238 $\rightarrow 192$

80% for training, 10% for testing, and 10% for validation.

Set	#Instances	#Paragraphs
Training	1,602	73,846
Validation	192	11,253
Test	201	10,425
Total	1,995	95,524

Dataset

Input: Three levels of knowledge facts

1. First Paragraph: The previous version of the first paragraph and *e*.
2. Summary: Summary of the previous version and *e*.
3. Full Content: The previous version of the full content and *e*.

Lengths of knowledge facts:

Level	Min	Max	Mean	Median
First Paragraph	17	244	72.26	62
Summary	17	813	174.36	94
Full Content	29	38,913	3,905.95	2,081

Dataset

Metrics

ROUGE (ROUGE-1, ROUGE-2, ROUGE-3, ROUGE-L)

The matching between n-gram, focus on recall.

BLEU (BLEU-1, BLEU-2, BLEU-3, BLEU-4)

The matching between n-gram, focus on precision.

METEOR

Weighted combination of ROUGE and BLEU scores.

Bert-Score (BS)

Measures the similarity between two pieces of texts by encoding with BERT.

Dataset

Baselines

Calculate the lexical overlapping

1. x_i and y_i
2. $x_i \bigoplus e_i$ and y_i

Inference: **Pre-trained PRIMERA without fine-tuning**

Generating the updated knowledge facts based on PRIMERA trained with WCEP-10 dataset.

Fine-tuned PRIMERA:
Three levels of knowledge facts (Summary, First Paragraph, Full Content)

Dataset

Baselines

Method	ROUGE-1	ROUGE-2	ROUGE-L	BLEU-1	BLEU-2	BLEU-3	BLEU-4	METEOR	BS
Generating the First Paragraph									
Original x	0.9560	0.9520	0.9559	0.5536	0.5936	0.5743	0.5478	0.7193	0.988
Concatenation of x and e	0.9613	0.9537	0.9608	0.5540	0.5941	0.5747	0.5482	0.7144	0.956
PRIMERA w/o fine-tuning	0.4534	0.3713	0.4422	0.5705	0.4679	0.4237	0.3992	0.5517	0.910
PRIMERA (fine-tuned)	0.9601	0.9585	0.9599	0.8454	0.9098	0.8905	0.8600	0.8519	0.990
Generating the Summary									
Original x	0.9467	0.9371	0.9466	0.7786	0.8758	0.8557	0.8186	0.7629	0.985
Concatenation of x and e	0.9505	0.9391	0.9499	0.5608	0.6398	0.6230	0.5920	0.6592	0.962
PRIMERA w/o fine-tuning	0.3554	0.2790	0.3469	0.6383	0.5343	0.4783	0.4467	0.5689	0.902
PRIMERA (fine-tuned)	0.9565	0.9525	0.9564	0.7854	0.8904	0.8780	0.8480	0.7743	0.989
Generating the Full Content									
Original x	0.9117	0.8883	0.9115	0.4496	0.7172	0.7634	0.7351	0.3447	0.979
Concatenation of x and e	0.9173	0.8918	0.9166	0.4184	0.6800	0.7263	0.6996	0.3259	0.979
PRIMERA w/o fine-tuning	0.0695	0.0372	0.0676	0.8256	0.6961	0.6007	0.5411	0.6101	0.870
PRIMERA (fine-tuned)	0.5185	0.4705	0.5179	0.5999	0.7785	0.7899	0.7641	0.4942	0.966

Dataset

Statistics of three-level contents generation

Summary	MIN	MAX	MEAN	MEDIAN
Hypothesis	15	806	170.7612	94.0
Reference	17	813	174.3632	94.0

First Paragraph	MIN	MAX	MEAN	MEDIAN
Hypothesis	15	244	71.7264	62.0
Reference	17	244	72.2637	62.0

Full Content	MIN	MAX	MEAN	MEDIAN
Hypothesis	33	858	610.7861	731.0
Reference	29	38913	3905.9453	2081.0

Dataset

Insights

Readability

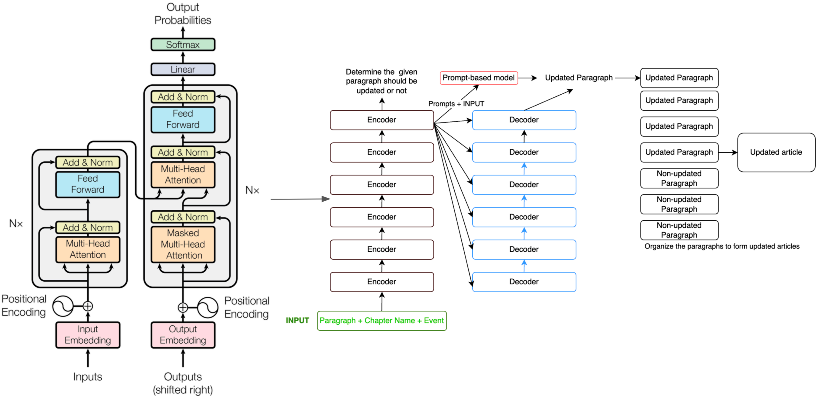
- Summary level and First-Paragraph level (better)
- Full-Content level (worse)

Updates (sentences, paragraphs)

- Update specific paragraphs

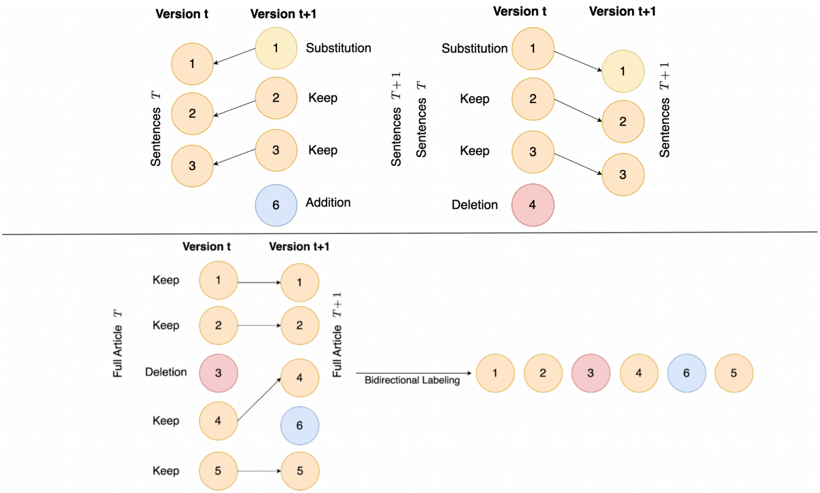
Model

Architecture



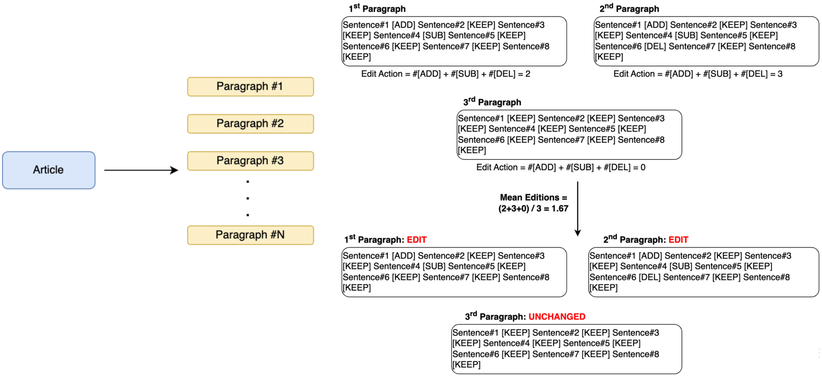
Model

Bidirectional Sentence-matching Algorithm



Model

Paragraph Labeling



Model

Generate Updated Paragraphs

Generate updated paragraphs with pre-trained language models

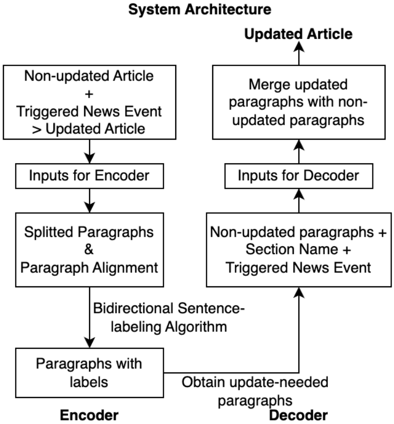
- Generate paragraphs with summarization methods with input format $x \oplus e$
- Collaborating with prompt-based models (GPT-3.5, GPT-4, Alpaca, Vicuna., etc) to test our model architecture with state-of-the-art conversational LLMs for general purposes.

Prompt 輸入格式
As an article writer, your task is to provide an updated paragraph based on the given non-updated paragraph and a triggered news. **Non-updated paragraph:** {paragraph} **Triggered News:** {trigger}

Model

Full Model Architecture

Demo Link[1]



[1] http://140.119.164.212:7840/

Experiments

Settings

- Fine-tuning LLaMA-13B with the Alpaca dataset using LoRA.
- Fine-tuning the model based on Alpaca with the NetKu dataset using LoRA.
- Vicuna: Fine-tuning ShareGPT with the dataset on fine-tuned LLaMA using LoRA.
- Fine-tuning Vicuna with the NetKu dataset using LoRA.
- Collaborating with prompt-based models.

Experiments

	MIN	MAX	MEAN	MEDIAN
GPT-3.5 (2023/03/15)	77	86710	7859.5060	4399.5
GPT-4 (2023/08/26)	77	86928	7928.8214	4423.0
Alpaca-13B	53	86431	7740.3036	4375.5
Alpaca-13B Fine-Tune NetKu	77	86450	7715.0536	4365.5
Vicuna-13B	77	87265	8071.9167	4392.5
Vicuna-13B Fine-Tune NetKu	65	86528	7748.6845	4370.5
BART (two-staged)	76	86560	7774.1845	4387.5
Reference	34	67597	5766.7857	3167.0

Table: 經 prompt-based 模型更新之全文長度評估

Experiments

Method 全文內容更新	ROUGE-1	ROUGE-2	ROUGE-L	BLEU-1
Original $x + e$	0.9867	0.9831	0.9862	0.4496
PRIMERA w/o fine-tuning	0.0695	0.0372	0.0676	0.8256
PRIMERA (fine-tuned)	0.5185	0.4705	0.5179	0.5999
GPT-3.5 (2023/03/15)	**0.8815**	**0.8268**	**0.8778**	**0.9431**
GPT-4 (2023/08/26)	0.8736	0.8552	0.8672	0.9374
Alpaca-13B	0.8745	0.8559	0.8691	0.9204
Alpaca-13B Fine-Tune NetKu	0.8779	0.8584	0.8704	0.9172
Vicuna-13B	0.8684	0.8539	0.8626	0.9032
Vicuna-13B Fine-Tune NetKu	0.8760	0.8573	0.8679	0.9140
BART (two-staged)	**0.8468**	**0.7949**	**0.8439**	**0.9613**

Table: 實驗結果 (ROUGE, BLEU)

Experiments

Method 全文內容更新	BLEU-2	BLEU-3	BLEU-4	METEOR	BS
Original $x + e$	0.7172	0.7634	0.7351	0.3447	0.979
PRIMERA w/o fine-tuning	0.6961	0.6007	0.5411	0.6101	0.870
PRIMERA (fine-tuned)	0.7785	0.7899	0.7641	0.4942	0.966
GPT-3.5 (2023/03/15)	**0.9076**	**0.9363**	**0.9145**	**0.7119**	**0.921**
GPT-4 (2023/08/26)	0.9046	0.9233	0.9126	0.7128	0.922
Alpaca-13B	0.9007	0.9140	0.9081	0.7119	0.920
Alpaca-13B Fine-Tune NetKu	0.9005	0.9104	0.9144	0.7116	0.921
Vicuna-13B	0.8939	0.8992	0.8835	0.7091	0.916
Vicuna-13B Fine-Tune NetKu	0.8989	0.9073	0.9056	0.7115	0.920
BART (two-staged)	**0.9327**	**0.9596**	**0.9573**	**0.7128**	**0.922**

Table: 實驗結果 (BLEU, METEOR, BERTScore)

Conclusion

Challenges

- Lack of knowledge update dataset.
- Max input length of pre-trained models.
- Human subjective opinions may not necessarily identify paragraphs with significance.

Conclusion

Proposed Solutions

- We propose “ A Multi-grained Dataset for News Event Triggered Knowledge Update ” dataset.
- Paragraphs are divided and labeled to train the encoder.
- Encoder is able to determine whether paragraphs should be updated.
- Collaborating with prompt-based models is enabled.

Conclusion

Applications

- Documents related to the legal changes can be updated at a more convenient speed.
- Technical documents must keep pace with state-of-the-art methods.
- Company's strategic planning is mainly influenced by social trends and business competitors.

Future Work and Improvements

- Knowledge understanding from specific domains.
- Training with non-English contexts if needed.

Thank You

Data: https://github.com/hhhuang/NetKu

Model: https://github.com/theQuert/Event-Triggered-Article-Updating-System

Demo: http://140.119.164.212:7840/