MTMT in QALab-3:

World History Essay Question Answering System that Utilizes Textbooks and Open Knowledge Bases

Takaaki Matsumoto[†‡] and **Teruko Mitamura**[†]

[†] Carnegie Mellon University
[‡] SOC Corporation

Outline

1. **Introduction and Baseline**

2. **Linked Open Data Assisted Machine Translation**

3. **Adding Open Knowledge Resource**

4. **Evaluation and Error Analysis**

5. **Conclusion**

Baseline System & Knowledge of the System

FelisCatus Zero-multilingual

- Developed by Sakamoto et al.

- Answer world history essay questions

- Multilingual: JA & EN

- Scala and UIMA

Knowledge: MT Textbooks

- Five machine translated JA textbooks

- Statistical MT in 2015 with Google Translate

FelisCatus Zero-multilingual

build passing   codecov 0%   License Apache 2.0   gitter join chat

This system answers world history essay questions in Japanese and/or Engli

(past version: FelisCatus Zero.)

Trial

A Fast Assumption

The baseline system uses five machine translated (MT) textbooks, but the translation technology is a bit old (statistical MT, in 2015).

Can we get a better result
if we use the latest neural MT?

Is Neural MT Better Than Statistical MT for the Task?

No. Neural MT made the ROUGE score worse. But **why?**

	ROUGE-1 Mean	ROUGE-2 Mean	Number of Questions (Long, Short)	ROUGE-1 Std. Dev.
Baseline with **Statistical** MT Textbooks	**0.063**	**0.009**	27 (5, 22)	0.081
Baseline with **Neural** MT Textbooks	0.056	0.004	27 (5, 22)	0.077

Dataset: NTCIR-13 QA Lab 3 phase-1

Neural MTs mis-translate nouns!

Since **ROUGE-1 is based on uni-gram matching**,
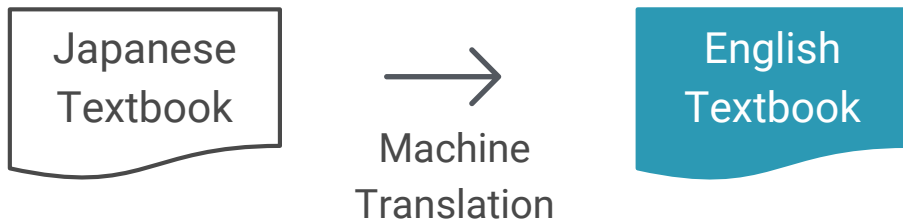mistranslating important nouns deteriorated answer quality.

Japanese Term	Google Translate	Bing Translator	Wikidata.org	Answer
林則徐	Hayashi Noriro	the zexu	Lin Zexu	Lín Zéxú
欽差大臣	Minister of Ginza	Minister of the Qin	Imperial Commissioner	Imperial Commissioner
キリスト教綱要	Christianity requirements	Christian elements	Institute of the Christian Religion	Institute of the Christian Religion

Linked Open Data Assisted Machine Translation

Applying Linked Open Data to Machine Translation

Conventional Method: Direct MT

Japanese Textbook → Machine Translation → English Textbook

Proposed Method: 2 Steps Translation

Japanese Textbook → Pre-translation → JA-EN Mixed Text → Machine Translation → English Textbook

Bilingual Term Corpus By LOD

Example of LOD Assisting MT

イギリスで増大しつづける中国茶（紅茶）の消費に対して，イギリス東インド会社はしだいに銀による支払いが追いつかなくなっていた。そこで，１８世紀末から，イギリスはインドでアヘンの専売制を始め，専売による財源の増加とアヘンを中国に売却することによって，茶の支払いにあてようとした。１８３９年，アヘン弛禁派をおさえ，厳禁派の林則徐が欽差大臣として広州に派遣され，アヘン密輸問題の解決にあたった。彼は外国商人がもつアヘンを没収してそれを廃棄した。アヘン貿易商人はこれに強く反発し，イギリス議会ではグラッドストンらによる"恥ずべき戦争"という反対にあったが，９票差で戦争を決定した。イギリス軍は沿岸の各地で清軍をやぶりながら北上し，１８４２年清朝と南京条約を結んだ。イギリスにつづき，１８４４年にアメリカ（望厦条約）とフランス（黄埔条約）も同様な条約を結び，清朝に条約の完全履行をせまった。しかし，華夷思想（中華思想）にもとづく朝貢外交の様式と異なるこの条約外交は，広州へのイギリス人の入城に対して地方官僚がそれを拒否するなど多くの摩擦を発生させることとなった。

Step.1 Pre-translation by the Bilingual Corpus

United Kingdom で増大しつづける China 茶（紅茶）の消費に対して，East India Company はしだいに銀による支払いが追いつかなくなっていた。そこで，１８世紀末から，United Kingdom は India でアヘンの専売制を始め，専売による財源の増加とアヘンを China に売却することによって，茶の支払いにあてようとした。１８３９年，アヘン弛禁派をおさえ，厳禁派の Lin Zexu が Imperial Commissioner として Guangzhou に派遣され，アヘン密輸問題の解決にあたった。彼は外国商人がもつアヘンを没収してそれを廃棄した。アヘン貿易商人はこれに強く反発し，United Kingdom 議会では Gladstone らによる"恥ずべき戦争"という反対にあったが，９票差で戦争を決定した。British Armed Forces は沿岸の各地で清軍をやぶりながら北上し，１８４２年清朝と Treaty of Nanjing を結んだ。United Kingdom につづき，１８４４年に United States of America（Treaty of Wanghia）と France（Treaty of Whampoa）も同様な条約を結び，清朝に条約の完全履行をせまった。しかし，Sinocentrism（Sinocentrism）にもとづく朝貢外交の様式と異なるこの条約外交は，Guangzhou への British people の入城に対して地方官僚がそれを拒否するなど多くの摩擦を発生させることとなった。

Step.2 Machine Translation (MS Bing Translator)

The East India Company was gradually unable to keep up with silver payments for the consumption of a growing China tea (tea) in United Kingdom. So, since the end of the eighteenth century, United Kingdom began to pay for tea by starting with the monopoly system of opium in India and by selling opium to China. In 1839, the prohibition of opium Sag, Lin Zexu was dispatched to Guangzhou as Imperial Commissioner to solve the problem of opium smuggling. He confiscated the opium of the foreign merchant and disposed of it. The opium trade merchant rallied strongly against this, and the United Kingdom Congress was opposed to the "shameful war" by Gladstone, but the war was decided by nine votes. British armed Forces north while veils the Qing army in various parts of the coast, and tied the Senyi morning and Treaty of Nanjing in 1842. United Kingdom, United States of America (Treaty of Wanghia) and France (Treaty of Whampoa) signed a similar treaty to the Qing Dynasty in 1844 to fully implement the Treaty. However, this treaty diplomacy which differs from the style of tribute diplomacy based on Sinocentrism (Sinocentrism) has caused a lot of friction that local bureaucrats refuse it for the entry of British people to Guangzhou.

Two translation errors; **British Parliament** (United Kingdom Congress) **& Qing Dynasty** (Senyi morning).

Without Pre-translation (Conventional Direct MT)

The British East India Company was gradually unable to keep up with silver payments for consumption of growing Chinese tea. So, since the end of the eighteenth century, the British tried to pay for the tea by starting the monopoly system of opium in India, increasing the financial resources and selling opium to China. In 1839, the prohibition of opium-sag, and the zexu of the Forbidden faction was dispatched to Guangzhou as Minister of the Qin, and the settlement of the opium smuggling problem was resolved. He confiscated the opium of the foreign merchant and disposed of it. The opium trade merchant rallied strongly against this, and the British Parliament was opposed to the ' shameful war ' by the Gladstone, but the war was decided by nine votes. The British Army veils the Qing army in various parts of the coast, and it tied the Nanjing Treaty with Senyi morning in 1842. In 1844, the United States (the Treaty) and France (Huang Tai Treaty) signed a similar treaty to the United Kingdom, and the Qing Dynasty concluded the full implementation of the Treaty. However, this treaty diplomacy, which differs from the style of tribute diplomacy based on Chinese thought, has caused a lot of friction, such as local bureaucrats refusing to enter the British into Guangzhou.

Six translation errors; some of them are nonsense or critical.

Comparison between LOD assisted MT and Simple MT

Five articles (approx. 250 words for each) were sampled and their translation quality were evaluated. For all articles, LOD assisted MT was better than simple MT.

	Number of Words Translated by LOD	LOD Failure	Bing Failure
Sample 1	33	1	2
Sample 2	40	3	10
Sample 3	22	1	3
Sample 4	21	1	9
Sample 5	42	3	7

Evaluation

LOD assisted neural MT is effective for this task.

	ROUGE-1 Mean	ROUGE-2 Mean	Number of Questions (Long, Short)	ROUGE-1 Std. Dev.
Baseline + Statistical MT Textbooks	0.063	0.009	27 (5, 22)	0.081
Baseline + Neural MT Textbooks	0.056	0.004	27 (5, 22)	0.077
Baseline + LOD Neural MT Textbooks	**0.081**	**0.011**	27 (5, 22)	0.100

Adding Open Knowledge Resource

Another Motivation

LOD assisted neural MT for Japanese textbooks worked.
However, the advance was not very large.

Can we get a better result

by adding open English textbooks?

(Their contents are not equivalent to Japanese textbooks!)

Boundless.com

Replace your textbooks with

BOUNDLESS

Boundless offers openly licensed, high-quality, customizable digital courseware at a fraction of the cost of traditional textbooks.

Why Boundless.com?

Pros.

- Written in English (source language)

- High quality

- Openly licensed and free

- HTML based (easy to extract text)

  A few PDF based textbooks can be found in the web. However, they are not easy to retrieve text, and their copyrights are unclear.

Cons.

- Not Equivalent to JA textbooks
  Does it matter?

Evaluation and Error Analysis

End-to-end Evaluation

	CASE		STEM		STOPWORD	
	ROUGE-1	ROUGE-2	ROUGE-1	ROUGE-2	ROUGE-1	ROUGE-2
CMUQA1	0.1231	0.0159	0.1257	0.0164	0.0347	0.0065
CMUQA2	0.1251	0.0160	0.1278	0.0166	0.0371	0.0066
CMUQA3	0.0768	0.0105	0.0968	0.0184	0.0594	**0.0129**
Forst1	0.1057	0.0086	0.1057	0.0086	0.0183	0.0022
IMTKU1	0.1187	0.0133	0.1239	0.0156	0.0457	0.0013
IMTKU2	0.0065	0.0000	0.0101	0.0001	0.0126	0.0001
MTMT1	**0.1717**	0.0180	**0.1841**	**0.0207**	**0.0615**	0.0054
MTMT2	0.1546	**0.0190**	0.1630	0.0197	0.0465	0.0020

Long and Short Essays

System	Short Essay		Long Essay					
	ROUGE-1 (STEM)	ROUGE-2 (STEM)	ROUGE-1 (STEM)	ROUGE-2 (STEM)	Expert Score	Grammat ical Err.	Semantic Err.	Missing Keywords
CMUQA1	0.0152	0.0054	0.1204	0.0115	0 (-26.2)	7.8	0.2	3.6
CMUQA2	0.0173	0.0054	0.1241	0.0119				
CMUQA3	**0.0442**	**0.0131**	0.1263	0.0119				
Forst1	0.0225	0.0027	0	0				
IMTKU1	0.0262	0	0.1315	0.0069	0 (-29.0)	6.0	0	4.6
IMTKU2	0.0124	0	0.0132	0.0004				
DGLab1SumExP			0.1428	0.0139	0 (-22.0)	6.8	0.4	3.0
DGLab2SumExP			0.1458	0.0138				
MTMT1	0.0370	0.0023	**0.1692**	**0.0190**	0 (-36.4)	10.8	1.8	4.4
MTMT2	0.0255	0	0.1387	0.0107				

Why so many grammatical and semantic errors?

Original JA Sentence and LOD Pre-translation:

こうして各国で経済ナショナリズムが強まり，大戦後の国際協調の気運は消えさった。

こうして各国でeconomic nationalism が強まり，大戦後のinternationalism の気運は消えさった。

NMT for **the pre-translated Sentence**: **Ungrammatical**

Thus in strong economic nationalism, internationalism after World War momentum is gone left.

NMT for **the original sentence: Correct**

Thus, economic nationalism strengthened in each country, and the momentum of international cooperation after the war disappeared.

How can we fix it?

1. **TF-IDF based pre-translation decision:**

   To suppress JA-EN mixed text, too easy terms should not be pre-translated because NMT can handle them correctly.
   Probably TF-IDF is a good metric to determine whether the term should be pre-translated.

1. **Using "the latest" Bing Translator**

   The latest version of the Bing Translator can handle JA-EN mixed text better!

   "こうして各国でeconomic nationalism が強まり，大戦後のinternationalism の気運は消えさった。"
   "The economic nationalism strengthened in each country, and the momentum of internationalism after the war disappeared."

Conclusions

Conclusions

Neural MT doesn't work

Simple neural translation of knowledge resource does not work for domain specific cross-lingual question answering.

LOD works

Linked open data is effective to find correct translation for difficult terms in machine translation process.

Adding Open Knowledge

Adding source language (EN) open knowledge resource would help even if its content is not equivalent to the target language (JA) knowledge resources.