Automatic Generation of Japanese Question-Answering Pairs

Hiroki Tanioka, Tokushima University, Tokushima Japan,

Kaoru Kimura Tokushima University Tokushima Japan

Kazuma Takaoka, Works Applications, Tokushima Japan,

Ryohei Nakatani, Works Applications, Tokushima Japan,

Yoshitaka Uchida Works Applications Tokushima Japan

{tanioka.hiroki, c501506035}@tokushima-u.ac.jp
{takaoka_k, nakatani_r, uchida_yo}@worksap.co.jp

Abstract

There are large-scale English question-answer corpora available. However, there are no large-scale Japanese question-answer corpus. Furthermore, various types of question-answer corpora are needed for each domain. Hence, an automatic generation method of question-answer pairs is valuable. Our approach is composed of three steps. The first step is to gather target dataset. The second step is to extract question-answer pairs from the dataset. The third step is to learn question sentences to a Recurrent Neural Network(RNN). After these three steps, new question sentences can be generated automatically from another dataset using the RNN. In this research, we propose a generating method of Japanese question sentences, and report an evaluation result of generated question sentences by automatic evaluation metrics. The automatic evaluation includes BLEU, METEOR and sentence similarity. Here, it is well known that selection of appropriate case particles is difficult in Japanese. Thus, we propose Dependency Error Rate as another automatic evaluation metric.

Keywords: Question-Answer Corpus, Automatic Generation, Machine Learning, Evaluation Metrics

1. Introduction

Question-answering systems often need to question-answering index (Iulian V. S., et al., 2016; Jangho L., et al., 2016). Those question-answering systems are based on search engine. If many and high-quality question-answering pairs are indexed on the search engine, the question-answering system is expected to make a response to users. Thus, the question-answering system needs lots of sophisticated question-answering corpora. There are large-scale English question answer corpora available (Pranav R., et al., 2018). However, there are no large-scale Japanese question- answer corpus. The system needs more Japanese corpora. Therefore, we need to spend a bunch of costs for constructing corpora. A low-cost approach is needed for generating question-answering corpus. Furthermore, various types of question-answer corpora are needed for each domain. If some users need an information for specific domain. It is difficult to construct and index the user's target content by hand in advance. To make the index quickly, a machine learning technique can be used for generating question-answering pairs automatically.

2. Approach

Our approach includes two stages. For the first stage, Japanese question-answering pairs are manually extracted from some data sources. The schema of question-answering pairs is based on CMU corpus (Noah A. S., et al., 2008) as reference. For the second stage, these extracted question sentences are learned by RNN(Recurrent Neural Network) for generating question-answering pairs. In this paper, question sentence generation is focused before seeking answers in a target sentence. Then, new question sentences are made by RNN. Additionally, these question sentences are evaluated by some automatic evaluation metrics. Specifically, we propose the dependency error rate which is as one of metrics using a dependency parser for evaluating Japanese fluency.

2.1 Preparing Question-Answering Pairs

Our approach starts with extracting manually question-answering pairs from a target dataset. The procedures of extracting question-answering pairs are as follows.

Step 1. Choose a category in Table 1.
Step 2. Choose 10 topics in the category.
Step 3. Make 6 types of questions in each topic.
Step 4. Describe 1 or more answers for each question.

On the step 1, collaborators must choose one category. There are candidate categories including Wikipedia contents and Tokushima University's contents. On the step 2, those collaborators should choose ten topics depending on their interest. On the step 3, six types of questions include What, Who, Where, Whose, How and Yes/No questions. On the step 4, answers must be described in the target topic.

Table 2.1: A list of question-answering categories.

Category	Number of Themes	Number of Questions	Number of Answers
学術(Academia)	10	100	100
技術(Technology)	10	124	156
自然(Nature)	11	127	146
社会(Society)	11	108	132
地理(Geography)	10	102	100
人間(Human)	10	73	74
文化(Culture)	10	136	138
歴史(History)	10	54	61
マニュアル(Manual)	10	74	74
シラバス(Syllabus)	15	120	120

Table 2.2: Example of extracted candidate sentences.

Type	SVM	Human	Target Sentence
True Positive	✓	✓	東京都の首長は、東京都知事である。 The head of Tokyo is the governor of Tokyo.
	✓	✓	地震に対して、地殻が非常にゆっくりとずれ動く現象を地殻変動と呼ぶ。 Crustal deformation is a phenomenon in which the crust moves very slowly relative to an earthquake.
False Negative		✓	魚料理や肉料理などの主菜など、イタリア料理のコースの流れがサブタイトルになっている。 A flow of course of Italian cuisine such as main dish such as fish dish and meat dish is subtitle.
		✓	1 メガパーセクは 326 万光年。 1 Mega parsec is 3.26 million light-years.
False Positive	✓		2012 年現在、国際的な本初子午線として IERS 基準子午線が使用されている。 As of 2012, the IERS standard meridian is used as the international prime meridian.
	✓		東京都の議決機関は東京都議会である。 The decision-making body in Tokyo is the Tokyo Metropolitan Assembly.
True Negative			ツンドラ・タウン（Tundratown） 寒冷地域の動物たちが暮らすエリア。 Tundratown, an area is where cold region animals live.
			ディズニーのアニメーション映画で 10 億ドルを突破したのは、『トイ・ストーリー3』、『アナと雪の女王』に次いで史上 3 番目である。 It was the third one in history, after "Toy Story 3" and "Anna and the Snow Queen" that broke through $ 1 billion in Disney's animated film.

2.2 Extracting Candidate Sentences

The second step, candidate sentences are extracting using Support Vector Machine (SVM) (C.-C. Chang and C.-J. Lin., 2011). Here, every sentence is segmented as terms by Sudachi (Kazuma T., et al., 2018), in analysing mode A, with the full version dictionary. The kernel of SVM is RBF. The parameters are "c=1000 -w0 10 -w1 1". Train data contains 5,000 sentences including 188 sentences chosen manually by human. Test data contains 2,000 sentences. Then, the number of sentences chosen by SVM is 70. The number of sentences chosen by both SVM and human is 29, that means True Positive. False Negative is 41, False Positive is 158, and True Negative is 1772. Hence, the accuracy is 90.1%, and the precision is 41.4%.

2.3 Learning Prepared Question Sentences

The third step, prepared question sentences are learned using sequence-to-sequence model (Minh-Thang L., et al., 2015). Every prepared question sentence is parsed to segmented terms by Sudachi in analysing mode C. The model of sequence-to-sequence is shown in Figure 2.3. The model trains the extracted candidate sentences for input, and manually generated questions for output.

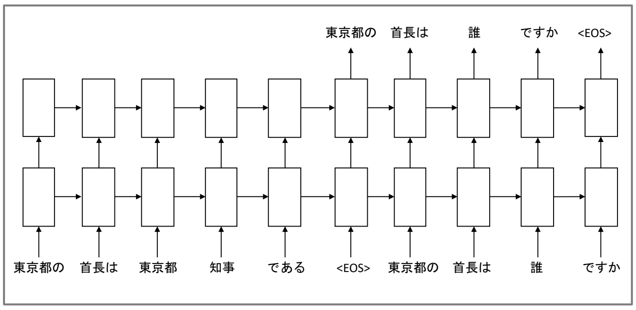

Figure 2.3: Neural Machine Translation using soft attention sequence-to-sequence model.

After training, the sequence-to-sequence model generates question sentences automatically derived from candidate sentences. This means the model is a question sentence generator.

3. Evaluation

Generated question sentences should be evaluated by some metrics, BLEU (Papineni, K., et al., 2002), METEOR (Banerjee S. and Lavie A., 2005), Position independent word Error Rate (PER) and Dependency Error Rate (DER). DER is one of metrics using a dependency parser for evaluating Japanese fluency.

3.1 BLEU and METEOR

BLEU is an algorithm for the evaluation machine translation output. Here, N is set as 4. S_i is the number of N-grams in a generated sentence i, with the matching reference cooccurrence in a correct sentence. T_i is the number of N-grams in a generated sentence i. BP_{BLEU} means a penalty when a sentence is too shorter than a correct sentence. For simplicity, BP_{BLEU} is set as 1.

$$p_n = \frac{\sum_i S_i}{\sum_i T_i},$$

$$BLEU = BP_{BLEU} * \exp(\frac{1}{N}\sum_{n=1}^{N}\log p_n). \tag{3.1.1}$$

METEOR is a metric for the evaluation machine translation output. Here, P means Precision, R means Recall, and three parameters; $\alpha = 0.9$, $\beta = 3$ and $\gamma = 0.5$. c means the number of words in both a generated sentence and a correct sentence, and u means the number of words in the generated sentence, with matching reference words in the correct sentence.

Parsed sentence:

「東京都の首長は、東京都知事である。」
（The head of Tokyo is the governor of Tokyo.)

Dependency pairs:

東京 (Tokyo)	→ 都 (city)
都 (city)	→ の (of)
東京都の (of Tokyo city)	→ 首長は、 (The head,)
首長 (head)	→ は <subject>
は <subject>	→ 、 (,)
首長は、 (The head,)	→ 東京都知事である。 (is the governor of Tokyo.)
東京 (Tokyo)	→ 都知事 (the governor of Tokyo)
都知事 (the governor of Toyo)	→ で (is)
で (is)	→ ある (is)
ある (is)	→ 。 (.)

Figure 3.3: Example of dependency pairs, includes underlined phrases are chunks, and other phrases are tokens.

$$F_{mean} = \frac{P \cdot R}{\alpha \cdot P + (1-\alpha) \cdot R},$$
$$METEOR = F_{mean} \cdot (1 - \gamma(\frac{c}{u})^{\beta}). \quad (3.1.2)$$

3.2 Position independent word Error Rate

Position independent word Error Rate (PER) (Christoph T., et al., 1997) is an evaluation metrics based on sentence similarity.

$$PER = 1 - \frac{\sum_i d_i}{\sum_i m_i} \quad (3.2.1)$$

Where d means the number of words in a generated sentence, with matching reference words in a correct sentence, and m means the number of words in a correct sentence.

3.3 Dependency Error Rate

Dependency Error Rate (DER) is one of metrics using a dependency parser for evaluating Japanese fluency. Here, CaboCha (Taku K. and Yuji M., 2000) is employed for dependency parser. The parser analyses a sentence and outputs dependency pairs as shown in Figure 3.3.

$$DER = 1 - \frac{\sum_i g_i}{\sum_i f_i} \quad (3.3.1)$$

Where g means the number of dependency pairs in a generated sentence, with matching reference dependency pairs in a correct sentence, and f means the number of dependency pairs in a correct sentence.

4. Conclusion

The evaluation results of 196 candidate sentences are BLEU = 0.095, METEOR = 0.235, PER = 0.388 and DER = 0.137. The results of generated question sentences and evaluation for each sentence with four metrics are shown in Table 4. For qualitative evaluation, the generated sentences have grammatical errors and unnatural particles in Japanese. When the errors and particles are included in a sentence, DER has larger reduction rate than PER. Hence, further research to increase generating accuracy and detect answers is necessary with other examples and machine learning models.

5. Bibliographical References

Banerjee S. and Lavie A. (2005). METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments, Proceedings of Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization at the 43th Annual Meeting of the Association of Computational Linguistics (ACL-2005), pp.65--72, June.

Christoph T., Stephan V., Hermann N., Alexander Z., and Hassan S. (1997). Accelerated DP based search for statistical translation. In Fifth European Conf. on Speech Communication and Technology, pp.2667--2670, Rhodos, Greece, September.

C.-C. Chang and C.-J. Lin. (2011). LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology, Vol.2, No.3, pp.27:1--27:27, May.

Iulian V. S., Alberto G.-D., Çağlar G., Sungjin A., Sarath C., Aaron C. and Yoshua B. (2016). Generating Factoid Questions With Recurrent Neural Networks: The {30M} Factoid Question-Answer Corpus, ACL 2016, pp.588.

Jangho L., Gyuwan K., Jaeyoon Y., Changwoo J., Minseok K., and Sungroh Y.. (2016). Training IBM Watson using Automatically Generated Question-Answer Pairs, CoRR 2016.

Kazuma T., Sorami H., Noriko K., Miho S., Yoshitaka U. and Yuji M., (2018). Sudachi: a Japanese Tokenizer for Business, Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), European Language Resources Association (ELRA), Miyazaki, Japan, May.

Minh-Thang L., Hieu P. and Christopher D. M. (2015). Effective Approaches to Attention-based Neural Machine Translation, EMNLP, Association for Computational Linguistics, Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp.1412—1421, Lisbon, Portugal, September.

Noah A. S., Michael H. and Rebecca H. (2008). Question Generation as a Competitive Undergraduate Course Project, In Proceedings of the NSF Workshop on the Question Generation Shared Task and Evaluation Challenge, Arlington, VA, September.

Papineni K., Salim R., Todd W. and Wei-Jing Z. (2002). BLEU: a method for automatic evaluation of machine translation', 40th Annual meeting of the Association for Computational Linguistics, pp.311--318.

Pranav R., Robin J. and Percy L. (2018). Know What You Don't Know: Unanswerable Questions for SQuAD, Association for Computational Linguistics (ACL), 2018.

Taku K. and Yuji M., (2000). Japanese Dependency Analysis Based on Support Vector Machines, EMNLP/VLC.

6. Language Resource References

JQAC Library. (2018). Japanese Question-Answering Corpus. https://taniokah.github.io/jqac/

Table 4: Example of automatically generated question sentences and evaluation for each sentence with four metrics.

Candidate sentence	Generated question sentence	Correct Sentence	BLEU	METEOR	PER*	DER*
世界遺産：世界遺産は、「顕著な普遍的価値」を有する文化遺産や自然遺産などであり、1972 年に成立した世界遺産条約に基づき、世界遺産リストに登録された物件を指す。 World Heritage: The World Heritage are cultural heritage and natural heritage which have "remarkable universal value," and it refers to the property registered in the World Heritage list based on the World Heritage Convention established in 1972.	世界遺産は何ですか？ What is the World Heritage Site?	世界遺産は、「顕著な普遍的価値」を有する文化遺産や自然遺産のことですか？ Is the world heritage a cultural heritage or natural heritage having "Remarkable Universal Value"?	0.275	0.353	0.273	0.182
世界遺産：世界遺産は、政府間委員会である世界遺産委員会の審議を経て決定される。 World Heritage Site: The World Heritage Site is decided after deliberation by the World Heritage Committee which is an intergovernmental committee.	世界遺産は政府の委員会ですか？ Is the World Heritage a government committee?	世界遺産は、どこの審議を経て決定されますか？ Which deliberation is the World Heritage decided after?	0.156	0.309	0.375	0.200
東京都：東京都の首長は、東京都知事である。 Tokyo: The head of Tokyo is the governor of Tokyo.	東京都の首長は何ですか？ What is the head of Tokyo?	東京都の首長を何といいますか？ What do you call the head of Tokyo?	0.181	0.366	0.600	0.300
地震：地震に対して、地殻が非常にゆっくりとずれ動く現象を地殻変動と呼ぶ。 Earthquake: A phenomenon in which the crust moves very slowly against earthquake is called crustal deformation.	地震が地殻には何を出すに何と言いますか？ What does the earthquake say to what to put out on the crust?	地震に対して，地殻が非常にゆっくりとずれ動く現象を何と呼ぶか？ What do you call a phenomenon that the crust moves very slowly against earthquakes?	**0.041**	0.325	0.526	**0.063**
本初子午線：「本初」とは「最初・首位」という意味である。 Prime meridian: "prime" means "the first and the first place".	本初子午線の本初たる最初は何を出す意味ですか？ What does the first as the prime of prime meridian mean to put out?	本初子午線の「本初」とは「最初・首位」をという意味ですか？ Does "Prime" of prime meridian mean "the first and the first place"?	0.288	0.377	0.550	0.222
動物園：水中の動物を中心として扱うものは特に水族館とされ、動物園の特殊な形態としてサファリパークや移動動物園、鳥類園、クマ牧場などがある。 Zoo: The ones that keep mainly with animals underwater are considered to be particularly aquariums, and special forms of zoos include safari parks, mobile zoos, bird gardens, bear pastures and so on.	水中の動物を中心として、形態のサファリパークが移動動物園されているか？ Mainly underwater animals, are safari parks of the form mobile zoos?	一般に陸上の動物を中心として扱う動物園に対して，水中の動物を中心として扱うものは特に何と言われるか？ In general, what are especially those keep mainly underwater animals against zoos keep mainly land animals?	0.319	0.280	0.382	0.214
日本：連合国の指令のもと、国制の改革が進められ、大日本帝国憲法の改正手続きによって日本国憲法を制定し、1947 年発効の同憲法によって「国民主権」「平和主義」「基本的人権の尊重」の三大原則を確立した。 Japan: Under the direction of the Allied Powers, reform of the national system was promoted, the Constitution of Japan was enacted by the amendment procedure of the Constitution of the Japan Imperial Empire, and the Constitution which came into effect in 1947, entitled National Sovereignty, Pacifism, and Respect for Human Rights .	指令のもとの国制の改革を大日本帝国憲法していますか？ Does the Constitution of the Japan Imperial Empire reform the national system under the directive?	日本国憲法が施行されたのは何年ですか？ What year was the Constitution of Japan enforced?	**0.029**	0.119	0.267	**0.077**

* PER* = 1 - PER, DER* = 1 - DER.