November 18, 2019

Press release

Nippon Telegraph and Telephone Corporation

National Institute of Informatics, Research Organization of Information and Systems

AI achieved a score of 185 on the English written exam of the National Center Test for University Admissions in 2019.

The Nippon Telegraph and Telephone Corporation (NTT, CEO: Jun Sawada) took the English written exam of the 2019 National Center Test for University Admissions and achieved high marks of 185 points (64.1 T-score). This reflects its research and development of corevo※1, which is one of NTT's AI-related technologies, and its participation in the AI project called "Todai Robot Project – Can a robot get into the University of Tokyo?" ※2 with the Inter-University Research Institute Corporation Research Organization of Information and Systems, National Institute of Informatics (NII, director general: Masaru Kitsuregawa). This achievement was made possible by the English team and its insight into natural language processing technologies based on deep learning. The team improved its score by more than 30 points (corresponding to 7 T-score points) compared to the simple application of deep learning technologies.

Background and significance

NTT Communication Science Laboratories (NTT CS Labs.) ※3 have been conducting basic research on natural language processing and knowledge processing, including machine translation, information retrieval, and dialogue processing.

To solve practical problems, humans integrate various natural language processing abilities. For example, when preparing to write a report that summarizes previous literature, the writer must understand the content of the documents, summarize them, and write a cogent text. To increase the utility of future natural language processing technology for humans, natural language processing technologies must be evaluated in more practical settings, like those actually faced by humans.

The "Todai Robot Project – Can a robot get into the University of Tokyo?" (Torobo,

director: Noriko Arai) was started by NII to re-integrate AI-related fields that were segmented in the 80s. This project's aim was to open up this field's horizon and challenge young researchers. With the national admissions test and the University of Tokyo's second stage exams, the project pursues the extent to which AI can solve problems that humans themselves actually face. Among many types of tests, English exams contain problems that need to integrate both natural language and knowledge processing. Regarding the Torobo project as an appropriate benchmark for fueling the basic research of natural language processing and knowledge processing, NTT CS Labs. have been accumulating research results to automatically solve English exams that are typically included in the national admissions test.

Machine comprehension research based on deep learning continues to accelerate. XLNet※4, the cutting-edge technology, solves various problems from a relatively small dataset by combining a base model pre-trained from a huge amount of text data and a fine-tuning※5 process to adapt it to specific problems. However, when there is much scarcity in the training data and when encyclopedic knowledge is necessary, the performance of such new technology remains disappointing.

This time, our English team, comprised of members of the NTT CS Labs. and its research partners, applied their original technologies to sentence elimination problems, paragraph-titling problems, and pronunciation problems that were difficult for XLNet to handle. As a result, in the main test of the 2019 National Center Test for University Admissions, high marks of 185 points (full marks = 200 points, 154 points when our original technologies were not applied) and a T-score of 64.1 (57.0 when our original technologies were not applied). We also applied our technologies to the last three years of main and supplementary admission tests and achieved a T-score over 60 (Fig. 1).
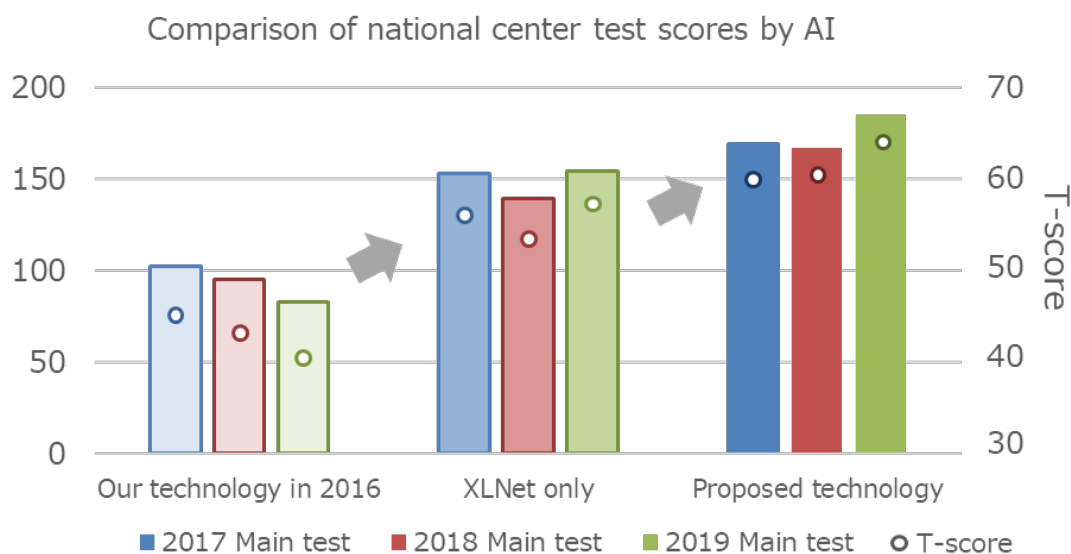
Fig. 1 Comparison of national center test scores by AI

Technical points

(1) Improvements in sentence elimination problems using automatically generated pseudo-problems

In machine learning, positive and negative examples are typically prepared to train models that can distinguish among such examples. Unfortunately, in sentence elimination problems, in which inappropriate sentences must be identified within a text, preparing training data is difficult since this type of problem was introduced recently and ordinary texts usually lack such unnecessary sentences. Even when applying deep learning technologies, in 2017-2019, we only correctly answered 6 of 15 problems in the main and supplementary tests. We addressed this problem by devising a method that automatically generates pseudo-problems by creating unnatural texts by changing the order of natural ones. By utilizing a large number of automatically generated problems for machine learning, we correctly answered all 15 problems. In addition, against the benchmark dataset comprised of past and our originally prepared exams, our accuracy improved from 60 to 86%.

(2) Answering method for paragraph-titling problem

In paragraph-titling problems, titles are selected that best capture their content. To solve such problems, each paragraph's content must be understood as well as the document's overall content and structure. Since all the title combinations need to be correctly answered, random guessing can only achieve 4% accuracy. Since the problem structure is complex, it cannot be solved by a simple application of such deep learning technologies as XLNet. Our English team calculated the similarity between the titles and paragraphs to find the optimal combination. All five problems in 2017-2019 were correctly answered. When combined with BERT※4, which is a deep learning technology for representing documents, we confirmed accuracy as high as 80% in our benchmark data.

(3) Using dictionaries for solving stress and pronunciation problems

In the national university admission tests, such various problems as stress and pronunciation, blank-filling, and document understanding comprehensively evaluate the English ability of examinees. NTT CS Labs. have been working to improve the technologies of these problems since 2014. For solving stress and pronunciation problems, since knowledge of pronunciation symbols is required, solving them with recent deep learning technologies is complicated. In such problems, we do not employ deep learning; we utilize pronunciation dictionaries with technologies that cope with orthographical variants and problem analysis (Fig. 3). As a result, almost all the stress and pronunciation problems were successfully answered.

The technological advancements for solving the English exam were achieved by a research partnership between NTT CS Labs. and Prof. Genichiro Kikui (Okayama Prefectural University), Prof. Kohji Dohsaka (Akita Prefectural University), Associate Prof. Hirotoshi Taira (Osaka Institute of Technology), Prof. Yasuhiro Minami (University of Electro-communications), and Prof. Junji Yamato (Kogakuin University). The members of NTT CS Labs. included Hiroaki Sugiyama (research scientist), Hiromi Narimatsu (research scientist), and Ryuichiro Higashinaka (senior distinguished researcher).

A 次の問い（問1～3）のパラグラフ（段落）には，まとまりをよくするために取り除いた方がよい文が一つある。取り除く文として最も適当なものを，それぞれ下線部①～④のうちから一つずつ選べ。

問1 ☐ 27

When flying across the United States, you may see giant arrows made of concrete on the ground. Although nowadays these arrows are basically places of curiosity, in the past, pilots absolutely needed them when flying from one side of the country to the other. ①The arrows were seen as being so successful that some people even suggested floating arrows on the Atlantic Ocean. ②Pilots used the arrows as guides on the flights between New York and San Francisco. ③Every 16 kilometers, pilots would pass a 21-meter-long arrow that was painted bright yellow. ④A rotating light in the middle and one light at each end made the arrow visible at night. Since the 1940s, other navigation methods have been introduced and the arrows are generally not used today. Pilots flying through mountainous areas in Montana, however, do still rely on some of them.

> A. In each of the following paragraphs, one sentence must be eliminated to improve the text's flow. Select the sentence that needs to be eliminated from ① to ④.

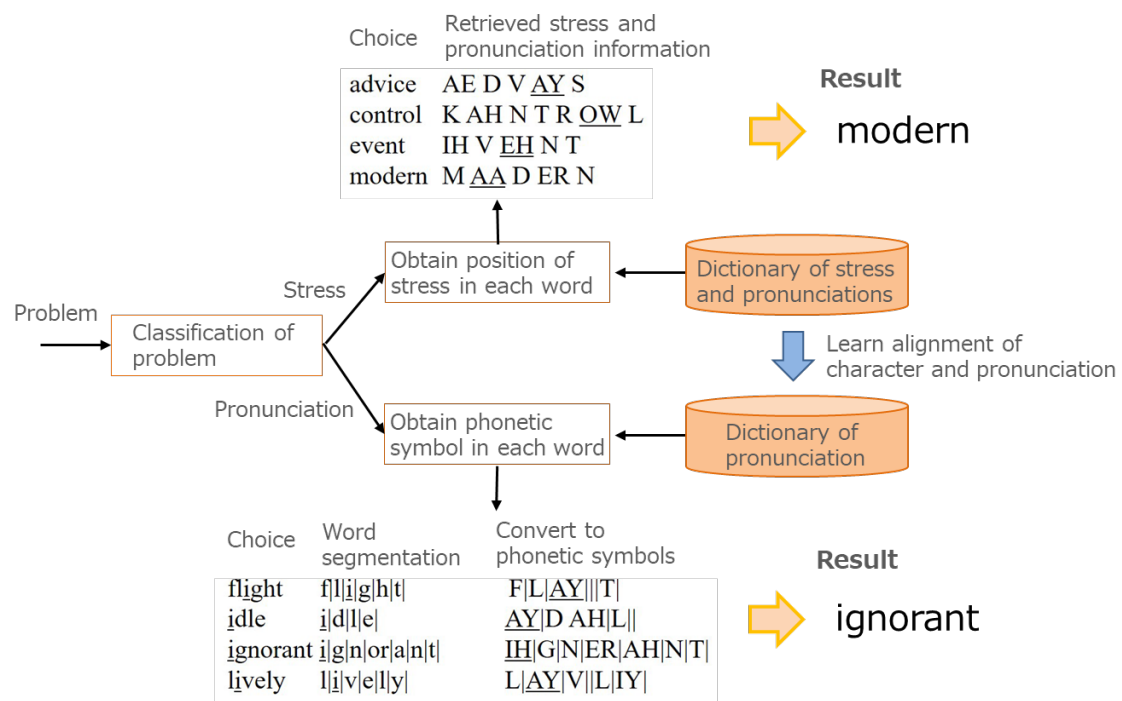Fig. 2 Example of sentence elimination problem (2019 main test, 3A)

Fig. 3 Method for solving stress and pronunciation problems: identifying position of stress in a word or pronunciation of an underlined part

Prospects

Through our research to date, we encountered many new research snags. For example, understanding such everyday documents as leaflets, advertising brochures, graphs and charts, and conversations remain complicated to solve automatically. In the future, we will continue our basic research and its integration to cope with such problems that require common sense and knowledge other than language. In addition, through our Torobo research, we aim to achieve dialogue and question-answering technologies that can understand context, common sense, and expert knowledge and apply them to various services.

Terms

※1 Corevo is a trademark of Nippon Telegraph and Telephone Corporation.

※2 Can a robot get into the University of Tokyo?

Started by Noriko Arai by NII in 2011, its aim is to re-integrate AI-related fields that were segmented in the 80s. The project opens up the field's horizons and challenges young researchers. As a concrete goal, it seeks to achieve high marks in the university admissions test by 2016 and pass the University of Tokyo's entrance exam in 2021.

Related link:
Torobo project website

http://21robot.org/

※3 NTT Communication Science Laboratories

For the future of information communication based on the nature of communication, NTT Communication Science Laboratories conduct basic research to create new

principles and concepts by an interdisciplinary approach that integrates human science and information science and produces innovative information communication services.

※4 XLNet, BERT

Deep learning technologies, such as XLNet and BERT can be used to obtain accurate representations of documents with a large amount of text data. Even if the target task's amount of training data is small, by using the acquired representations as a basis, high performance can be achieved in many natural language processing tasks.

※5 Transfer learning

Transfer learning adapts a model trained from a different dataset to new datasets. This process is also known as domain adaptation or fine-tuning. Utilizing the document representation learned from large text data by such technologies as XLNet and applying it to solve specific problems for university admissions exams is a good example of transfer learning.