Regular Articles

Taking the English Exam for the "Can a Robot Get into the University of Tokyo?" Project

Ryuichiro Higashinaka, Hiroaki Sugiyama, Hideki Isozaki, Genichiro Kikui, Kohji Dohsaka, Hirotoshi Taira, and Yasuhiro Minami

Abstract

NTT and its research partners are participating in the "Can a robot get into the University of Tokyo?" project run by the National Institute of Informatics, which involves tackling English exams. The artificial intelligence system we developed took a mock test in 2014 and achieved a better-than-human-average score for the first time. This was a notable achievement since English exams require English knowledge and also common sense knowledge that humans take for granted but that computers do not necessarily possess. In this article, we describe how our artificial intelligence system takes on English exams.

Keywords: artificial intelligence, natural language processing, Todai Robot Project

1. Introduction

The National Institute of Informatics (NII) has been promoting the Todai Robot Project (Can a robot get into the University of Tokyo?) [1]. By developing *artificial intelligence* that can pass the entrance exam of the University of Tokyo (Todai), the members of the project aim to clarify the limitations of artificial intelligence and the boundaries between humans and computers. The goal of the project is to achieve high marks on the National Center Test for University Admissions (NCTUA) by 2016 and to be admitted into the University of Tokyo by 2021.

NTT, together with its research partners, Okayama Prefectural University, Akita Prefectural University, Osaka Institute of Technology, and the University of Electro-Communications, joined the project to form an English team in 2014 and achieved a score of 95 out of 200 points) on a mock Center Test (Yozemi mock Center Test by the educational foundation Takamiya Gakuen) in the same year; the score we achieved with our computer program was above the

national average of 93.1. This is a big jump from the previous year, because the English score in 2013 was 52, which was a chance-level score considering that all problems are multiple choice, mostly consisting of four choices.

It may be a surprise to readers to learn that English is regarded as one of the most difficult subjects for computers compared with other subjects. This is because problems in English cannot be solved just by having English knowledge such as of vocabulary, grammar, and idioms. Rather, these problems require *common sense knowledge*. For example, there are problems in which it is necessary to estimate human emotion and the causal relationships of events. At NTT, we have been tackling problems that require common sense reasoning through the research and development of dialogue and machine translation systems. We successfully exploited our expertise in these fields to achieve a high score on the mock Center Test.

To pass the entrance exam of the University of Tokyo, it is necessary to achieve high marks on both

1	Pronunciation and accent	1A	Pronunciation
		1B	Accent
2	Grammar, dialogue, and ordering	2A	Grammar, wording, and vocabulary
		2B	Dialogue completion
		2C	Word ordering
3	Reading comprehension	3A	Word/phrase sense estimation
		3B	Selection of out-of-context sentence
		3C	Opinion summarization
4	Reading comprehension (processing information)	4A	Graphs, figures, and tables
		4B	Understanding of practical documents
5	Reading comprehension (understanding information)		
6	Reading comprehension (editorial)		

Short-sentence problems

Long-sentence problems

Fig. 1.   Basic structure of English exam of NCTUA.

the NCTUA and second-stage exams. However, since the second-stage exams contain difficult tasks involving writing, we are currently working on the NCTUA, where there are only multiple-choice problems. The typical structure of the English exam of the NCTUA is shown in **Fig. 1**. There are usually six categories of problems; the first three are short-sentence problems (problems related to words and short phrases), and the latter three are long-sentence problems (problems related to long documents that may contain illustrations, graphs, and tables).

In this article, we describe how *Torobo-kun* (*Torobo* stands for Todai Robot, and *kun* is an honorific title in Japanese normally used for boys; the members of the project affectionately refer to the exam-solving software this way) solves questions on the English exam, putting an emphasis on our approach on short-sentence problems. Unfortunately, the current level of artificial intelligence has difficulty solving long-sentence problems, and we have not yet established a concrete method for solving them. We discuss the reason for this difficulty in the last section. Moreover, Torobo-kun is currently a type of computer software. Therefore, it does not visually see and read questions on paper, nor does it fill in answer sheets using pencils. It obtains exams in electronic form and outputs choices that it thinks are correct.

2.   Pronunciation and accent (word stress) problems

Problems regarding pronunciation and accent (word stress) are not that difficult for computers because they can consult electronic dictionaries.

With pronunciation, the task is to select a word that has a letter that is pronounced differently from the same letters in the other choices. An example of a pronunciation problem is shown in **Fig. 2**. To solve this problem, the program first consults an electronic dictionary. The dictionary here is one developed for speech recognition and synthesis research[*]; thus, it has pronunciation and accent information for words. Then, to obtain the pronunciation for a target letter (in this case, "I"), we calculate the most likely alignments between letters and pronunciations within the dictionary by using statistical methods. This tells us which letter corresponds to what pronunciation. Finally, since we know that the pronunciation for "I" in "ignorant" is different from the "I"s in the other choices, we can obtain the correct answer.

With accent, the task is to select, out of four choices, a word that has the most accented syllables in a different position compared with other choices. An example of an accent problem is shown in **Fig. 3**. The same dictionary is used to solve this type of problem.

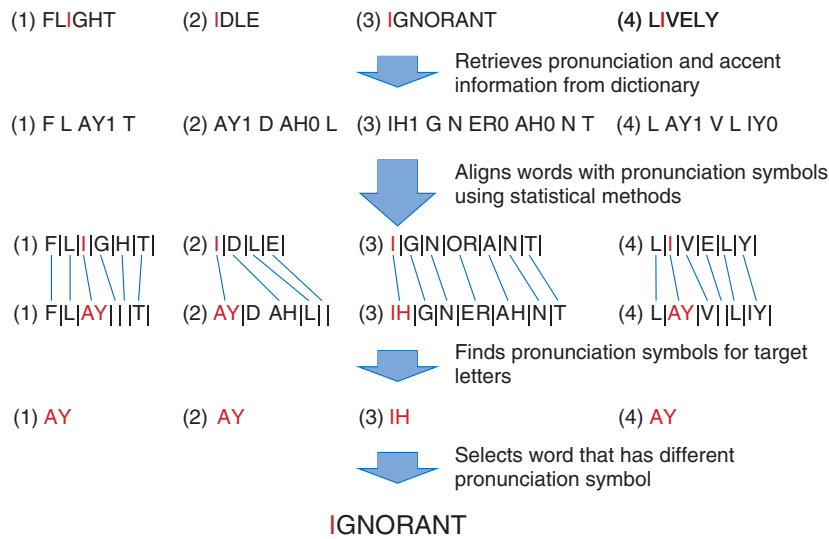[*]   We used the Carnegie Mellon University Pronouncing Dictionary.

(1) FLIGHT   (2) IDLE   (3) IGNORANT   (4) LIVELY

Retrieves pronunciation and accent information from dictionary

(1) F L AY1 T   (2) AY1 D AH0 L   (3) IH1 G N ER0 AH0 N T   (4) L AY1 V L IY0

Aligns words with pronunciation symbols using statistical methods

(1) F|L|I|G|H|T| (2) I|D|L|E| (3) I|G|N|OR|A|N|T| (4) L|I|V|E|L|Y|

(1) F|L|AY|||T| (2) AY|D AH|L|| (3) IH|G|N|ER|AH|N|T   (4) L|AY|V||L|IY|

Finds pronunciation symbols for target letters

(1) AY   (2) AY   (3) IH   (4) AY

Selects word that has different pronunciation symbol

IGNORANT

Fig. 2.   Pronunciation problem.

(1) ADVICE   (2) CONTROL   (3) EVENT   (4) MODERN

Retrieves pronunciation and accent information from dictionary

(1) AE0 D V AY1 S   (2) K AH0 N T R OW1 L   (3) IH0 V EH1 N T   (4) M AA1 D ER0 N

Converts into accent patterns

(1)   0   1   (2)   0   1   (3)   0   1   (4)   1   0

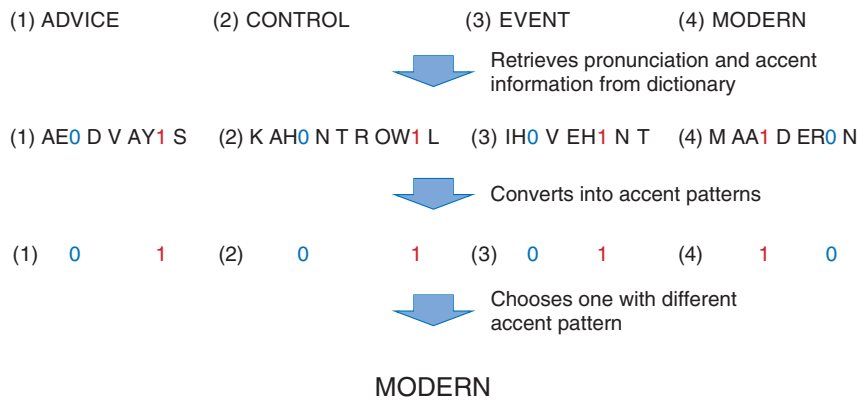Chooses one with different accent pattern

MODERN

Fig. 3.   Accent problem.

In the figure, alphabetic characters (letters) denote pronunciation symbols, and the numbers 0 and 1 respectively indicate whether a vowel is a non-accented or an accented one. From these 0s and 1s, we can find the accented vowels in the words and immediately determine that *modern* is the odd one out.

3.   Grammar, wording, and vocabulary problem

The task here is to select one of the four choices that best fits in the blank in a sentence. Below is an example:

I had a severe toothache, so I made [A] with the dentist.
(1) a promise, (2) a reservation, (3) an appointment, (4) an arrangement.

In this example, the answer that best fits [A] is (3), an appointment. To solve this type of problem, we turn to a technology called statistical language models, which have been widely used in machine translation to produce fluent translations. Language models are created by statistically processing a large number of documents; they contain the probability information about how one expression follows another. Given a question, our program first creates sentences by
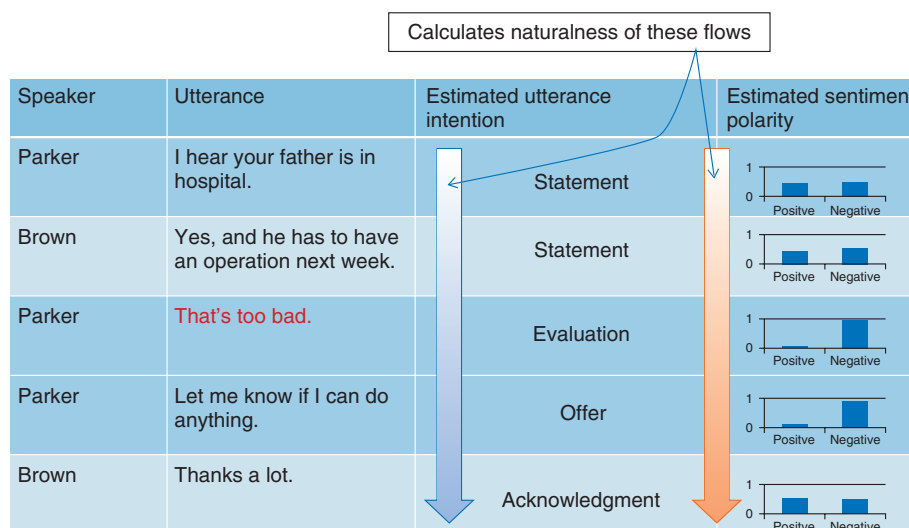
Fig. 4.   Flow of utterance intentions and sentiment polarities.

filling in the blanks with each of the choices and calculates the fluency of each sentence by using the language models. Then, we select the choice that realizes the most fluent (probable) sentence.

4.   Dialogue completion problem and opinion summarization problem

In the dialogue completion problem, it is necessary to select one utterance from four choices to fill in the blank in a given dialogue. In the example below, the utterance that best fits [B] is choice (4).

Parker: I hear your father is in hospital.
Brown: Yes, and he has to have an operation next week.
Parker: [B] Let me know if I can do anything.
Brown: Thanks a lot.
(1) Exactly, yes., (2) No problem., (3) That's a relief., (4) That's too bad.

To solve this type of problem, we first fill in the blank by using each of the choices and, for each case, estimate the naturalness of the conversation; the choice that achieves the highest degree of naturalness is selected as the answer. We estimate the naturalness of a conversation from two viewpoints; one is the flow of utterance intentions (also known as dialogue acts or illocutionary acts, e.g., statement, evaluation, question, etc.) and the flow of sentiment polarities, e.g., positive and negative.

The flow of utterance intentions and sentiment polarities when the blank has been filled in with (4) is shown in **Fig. 4**. Here, the utterance intentions and sentiment polarities are those that have been automatically estimated from dialogue data using statistical methods. The conversation begins with a *statement* about Brown's father being in hospital followed by another *statement* about the fact that he has to have an operation. This is followed by an *evaluation* (evaluative response), *offer*, and *acknowledgment*. The flow of these utterance intentions seems as natural as that in human conversation. As for sentiment polarities, the polarities regarding the admission to hospital and the operation match that of (4); that is, they all have negative polarities. Therefore, this flow of sentiment is also as natural as that in human conversation. On the basis of the naturalness of these two flows, Torobo-kun can select (4) as the most appropriate answer.

The task with opinion summarization problems is similar to that of the dialogue completion problem. The task in this case is to select an utterance that best summarizes a speaker's opinion in a multi-party discussion. Since a discussion is a kind of conversation, we use the same technique as just described; that is, we select an utterance that creates the most natural flow. In this type of problem, the choices are usually of the same utterance intention. Therefore, we only leverage the flow of sentiment polarities.
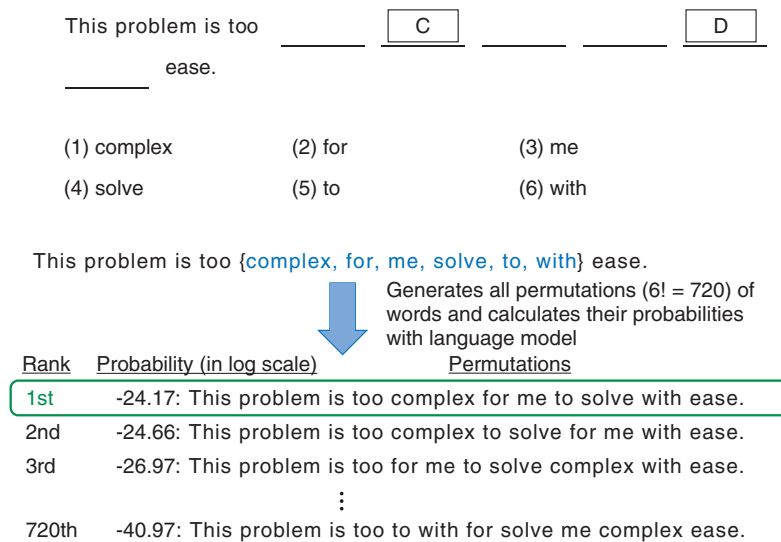
This problem is too [____] [C] [____] [____] [D]
[____] ease.

(1) complex   (2) for   (3) me

(4) solve   (5) to   (6) with

This problem is too {complex, for, me, solve, to, with} ease.

Generates all permutations (6! = 720) of words and calculates their probabilities with language model

Rank	Probability (in log scale)	Permutations
1st	-24.17:	This problem is too complex for me to solve with ease.
2nd	-24.66:	This problem is too complex to solve for me with ease.
3rd	-26.97:	This problem is too for me to solve complex with ease.
⋮		
720th	-40.97:	This problem is too to with for solve me complex ease.

Fig. 5.   Solving the word ordering problem.

5. Word ordering problem

The task in these problems is to fill in blanks in a sentence by correctly ordering given words in order to create a valid grammatical sentence. An example is shown in **Fig. 5**. In the example, the correct ordering for the given words is "complex for me to solve with" (the answer for this problem consists of the words that correspond to blanks [C] and [D]). For this type of problem, our program creates all possible permutations of word orderings (in the example, 720 permutations are created) and chooses the most appropriate one. To calculate the appropriateness, we turn to the statistical language models we used for the grammar, wording, and vocabulary problems.

6. Word/phrase sense estimation problem

The task in this case is to estimate the meaning of an unknown word/phrase (the unknown word/phrase may not be unknown to the computer program because it has access to dictionaries; here, *unknown* means *unknown or unfamiliar to examinees*) and to select a choice whose meaning is the most similar to the given word or phrase. Below is an example:

George: I must get this paper finished by next Monday.
Paul: So you can't go to dinner with me this weekend?
George: No, but I'll take a rain check.

Paul: Sure, how about next weekend.
Unknown word (phrase): *take a rain check*
(1) accept your offer later, (2) change my mind, (3) go with you, weather permitting, (4) refuse the invitation.

In this problem, *take a rain check* is the unknown phrase, and we need to select the choice that has the meaning closest to the phrase. In this example, because *take a rain check* means to request a postponement of an offer, the correct answer is (1). For this type of problem, we utilize a technique called word2vec [2] and an idiom dictionary. Word2vec enables us to calculate the similarity between words on the basis of their usages in large text data. The idea used here is called distributional similarity; that is, the more similar the contexts of words, the closer their meanings are.

Using word2vec, we calculate the similarities between an unknown word/phrase and the choices and select the one that has the highest similarity score. Note that when the unknown word/phrase is listed in the idiom dictionary, that word/phrase is replaced by its gloss (definition statement) before calculating the similarity. This heightens the accuracy in calculating similarities.

7. Long-sentence problems and future directions

With long-sentence problems, we apply the method used for textual entailment. In textual entailment, two

documents are given, and the task is to recognize whether the content of one document is contained in the other. In long-sentence problems, there are many questions that ask whether the content of a choice is contained in a given document. Therefore, the technology for textual entailment can be directly applied. However, this approach achieved very poor results in the mock Center Test. In particular, it was difficult to absorb the differences in expressions with the same meaning and to find reasonable referents for referring expressions. It is therefore necessary to further improve the accuracy of textual entailment. In addition, it is necessary to read the minds of characters in stories, understand illustrations, graphs, and tables, and grasp the logical and rhetorical structures of documents. As sentences get longer and longer, the need to *read between the lines* becomes more important. Between the lines, there is common sense, and to solve long-sentence problems, it is necessary to tackle the problem of common sense head on. In the future, we will focus on solving long-sentence problems and creating programs that can achieve high marks on both the NTCUA and the second-stage exams of the University of Tokyo.

Acknowledgment

References

[1] N. H. Arai and T. Matsuzaki, "The Impact of A.I. on Education—Can a Robot Get into the University of Tokyo?", Proc. of the 22nd International Conference on Computers in Education (ICCE 2014), pp. 1034–1042, Nara, Japan, Nov./Dec. 2014.

[2] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed Representations of Words and Phrases and their Compositionality," Proc. of the 27th Annual Conference on Neural Information Processing Systems (NIPS 2013), pp. 3111–3119, Stateline, NV, USA, Dec. 2013.

Ryuichiro Higashinaka
Senior Research Scientist, Interaction Research Group, Innovative Communication Laboratory, NTT Communication Science Laboratories.
Senior Research Scientist, Audio, Speech, and Language Media Project, NTT Media Intelligence Laboratories.
He received a B.A. in environmental information, Master of Media and Governance, and Ph.D. from Keio University, Kanagawa, in 1999, 2001, and 2008, respectively. He joined NTT in 2001. His research interests include building question answering systems and spoken dialogue systems. From November 2004 to March 2006, he was a visiting researcher at the University of Sheffield in the UK. He received the Maejima Hisoka Award from the Tsushinbunka Association in 2014. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE), the Japanese Society for Artificial Intelligence (JSAI), the Information Processing Society of Japan (IPSJ), and the Association for Natural Language Processing (NLP).

Hiroaki Sugiyama
Researcher, Interaction Research Group, Innovative Communication Laboratory, NTT Communication Science Laboratories.
He received the B.E. and M.E. in information science and technology from the University of Tokyo in 2007 and 2009, respectively. He joined NTT Communication Science Laboratories in 2009 and studied chat-oriented dialogue systems and language development of human infants. He is currently working on his Ph.D. in information science at Nara Institute of Science and Technology. He is a member of JSAI.

Hideki Isozaki

Professor, Department of Systems Engineering, Okayama Prefectural University.

He received the B.E., M.E., and Ph.D. in engineering from the University of Tokyo in 1983, 1986, and 1997, respectively. He joined NTT Basic Research Laboratories in 1986 and studied artificial intelligence including logical inference. From 1990 to 1991, he was a guest researcher at Stanford University, USA. In 1995, he began studying natural language processing and soon after joined NTT Communication Science Laboratories. He is currently studying machine translation and scientific literature question answering including image processing. He received the 2004 IPSJ Best Paper Award. He joined Okayama Prefectural University in 2011. He is a member of the Association of Computational Linguistics, NLP, IPSJ, JSAI, and IEICE.

Genichiro Kikui

Professor, Department of Computer Science and Systems Engineering, Okayama Prefectural University.

He received the B.S. and M.E. in electrical engineering and the Ph.D. in informatics from Kyoto University in 1984, 1986, and 2007, respectively. He joined NTT in 1986. He was a visiting researcher at the German Research Centre for Artificial Intelligence (DFKI) in 1993 and the Center for the Study of Language and Information (CSLI) at Stanford University, USA, in 1997. He worked for ATR Spoken Language Processing Labs from 2001 to 2005 as the head of Department 2, responsible for integrating speech and natural language processing. He has been a professor with Okayama Prefectural University since 2011. He has worked on written and spoken natural language processing including text/speech translation, automatic text revision, and cross-language information retrieval. He is currently interested in information extraction and its applications. He received the 1997 IPSJ Best Author Award, the 2005 IEICE Information and Systems Society Best Paper Award, and the 22nd Telecom Technology Award. He is a member of IEICE, IPSJ, JSAI, and NLP.

Kohji Dohsaka

Professor, Department of Electronics and Information Systems, Akita Prefectural University.

He received the B.S. and M.E. in information and computer science from Osaka University in 1984 and 1986, respectively, and the Ph.D. in information science from Japan Advanced Institute of Science and Technology, Ishikawa, in 2004. He joined NTT in 1986. He was a visiting researcher at the Massachusetts Institute of Technology (MIT) in Cambridge, MA, USA, in 2002 and an invited professor at Osaka University from 2008 to 2011. He has been a professor with Akita Prefectural University since 2012. He has worked on dialogue systems, human-robot interaction, and natural language processing; he is now interested in communication robots for human support. He received the 1996 IPSJ Best Paper Award and the 2010 and 2011 JSAI Annual Conference Award. He is a member of IEICE, IPSJ, JSAI, and NLP.

Hirotoshi Taira

Associate Professor, Faculty of Information Science and Technology, Osaka Institute of Technology.

He received the B.S. in chemistry and the M.S. in solid-state chemistry from the University of Tokyo and the Ph.D. in information science from Nara Institute of Science and Technology in 1994, 1996, and 2002, respectively. He joined NTT Communication Science Laboratories in 1996 and researched text categorization with machine learning and bioinformatics. He was with the Business Intelligence Deployment Center, Research and Development Headquarters, NTT DATA, where he studied and developed business intelligence techniques from 2005 to 2007. He was a research scientist in the Natural Language Group, Innovative Communication Laboratory, NTT Communication Science Laboratories, where he researched Japanese predicate argument structure analysis and recognition of textual entailment from 2007 to 2014. He joined Osaka Institute of Technology in 2014. His research interests include natural language processing, machine learning, and artificial intelligence. He is a member of IPSJ, JSAI, and NLP.

Yasuhiro Minami

Professor, Department of Information Systems Fundamentals, Graduate School of Information Systems, the University of Electro-Communications.

He received the M.E. in electrical engineering and the Ph.D. in electrical engineering from Keio University, Kanagawa, in 1988 and 1991, respectively. He joined NTT in 1991. He was a visiting researcher at MIT, USA, from 1999 to 2000 before rejoining NTT. He has been a professor with the University of Electro-Communications in Tokyo since 2014. His interest is modeling for speech recognition. He received the 2006 IPSJ Best Paper Award, the 2005 IEICE Information and Systems Society Best Paper Award, and the 2005 and 2008 Telecom System Technology Award. He is a member of IEICE, the Acoustical Society of Japan, and the Institute of Electrical and Electronics Engineers.