SpringerBriefs in Computer Science

SpringerBriefs present concise summaries of cutting-edge research and practical applications across a wide spectrum of fields. Featuring compact volumes of 50 to 125 pages, the series covers a range of content from professional to academic.

Typical topics might include:

- A timely report of state-of-the art analytical techniques
- A bridge between new research results, as published in journal articles, and a contextual literature review
- A snapshot of a hot or emerging topic
- An in-depth case study or clinical example
- A presentation of core concepts that students must understand in order to make independent contributions

Briefs allow authors to present their ideas and readers to absorb them with minimal time investment. Briefs will be published as part of Springer's eBook collection, with millions of users worldwide. In addition, Briefs will be available for individual print and electronic purchase. Briefs are characterized by fast, global electronic dissemination, standard publishing contracts, easy-to-use manuscript preparation and formatting guidelines, and expedited production schedules. We aim for publication 8–12 weeks after acceptance. Both solicited and unsolicited manuscripts are considered for publication in this series.

More information about this series at http://www.springer.com/series/10028

Yutaka Kidawara · Eiichiro Sumita ·
Hisashi Kawai
Editors

Speech-to-Speech Translation

Springer

Editors
Yutaka Kidawara
Advanced Speech Translation Research and Development Promotion Center, National Institute of Information and Communications Technology
Kyoto, Japan

Hisashi Kawai
Advanced Speech Technology Laboratory, Advanced Speech Translation Research and Development Promotion Center, National Institute of Information and Communications Technology
Kyoto, Japan

Eiichiro Sumita
Advanced Translation Technology Laboratory, Advanced Speech Translation Research and Development Promotion Center, National Institute of Information and Communications Technology
Kyoto, Japan

ISSN 2191-5768 ISSN 2191-5776 (electronic)
SpringerBriefs in Computer Science
ISBN 978-981-15-0594-2 ISBN 978-981-15-0595-9 (eBook)
https://doi.org/10.1007/978-981-15-0595-9

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

"Language barrier" is one of the biggest challenges that humans face in communication. In the Old Testament, it is said that God created different languages when humans acted in defiance and attempted to build the Tower of Babel against His wishes. Whether this is true or not, it is an unquestionable fact that smooth communication has been hindered because of the different words we speak. Speech translation device—often makes its appearance in science fiction—which immediately translates spoken words into different languages has been the target of research and development in Japan since 1986. After some 30 years of constant effort, this dream device has been made available for practical use in certain fields and under certain conditions.

The history of research and development on machine translation can be traced back to the 1940s, but it wasn't until the 1983 World Telecommunication Exhibition (TELECOM 83) that multilingual speech translation technology—speech recognition and speech synthesis combined with machine translation—really caught people's attention. In 1986, the Advanced Telecommunications Research Institute International (ATR) was established in Japan, followed by the launch of a research project on speech-to-speech translation. Researchers from around the world gathered to conduct research and development under this project. In the 1990s, machine translation was a dictionary- and rule-based technology, but by the 2000s, alongside the diffusion of the Internet and the Web, the focus on research and development shifted to a bilingual corpus-based statistical approach, which took advantage of not having to consider the rules such as grammar.

In July 2010, the National Institute of Information and Communications Technology (NICT) initiated the world's first field experiment of a network-based multilingual speech-to-speech translation system using a smartphone application. To support further languages, NICT then led the establishment of an international consortium to initiate a global field experiment in cooperation with as many as 32 research entities of the world who had been individually engaged in developing speech translation technologies for their own languages. In April 2014, a governmental project "Global Communication Plan (GCP)" was launched in Japan to promote research and development on multilingual speech translation technology

and its implementation to fields such as travel, medical care, disaster prevention, and daily living, with the aim to eliminate the language barriers for foreign nationals and inbound tourists. With NICT taking the initiative, the industry, academia, and government together are engaged in research and development of a high-precision multilingual speech translation technology, namely by applying deep-learning approaches. NICT's speech translation application "VoiceTra" has been released as part of the field experiment under the GCP and serves as the key platform for its promotion. Thanks to the efforts from the initial field experiment and the achievements from the international collaboration, the application is currently capable of handling speech translation between as many as 16 languages and further improvements are being made, especially for the 10 most frequently used languages[1] in Japan that have been set forth as the target under the GCP.

After Google announced the development of the Google Neural Machine Translation in September 2016, a significant paradigm shift from statistical to neural machine translation has been taking place. The latest version of VoiceTra has also applied neural machine translation between Japanese and English and achieved dramatic improvements in accuracy. Let us look at a couple of comparison examples of different translation applications:

(1) The English translation results to the Japanese input "今日は全国的に雨の予報です" were:

- "Today, it is forecasted to rain all over the country" (by VoiceTra);
- "Today's rain forecast nationwide" (by App X);
- "It is forecast of rain nationwide today" (by App Y); and

(2) The English translation results to the Japanese input "昨今のニューラル翻訳システムの性能は目を見張るものがある" were:

- "The performance of the recent neural translation system is eye-catching" (by VoiceTra);
- "The performance of recent neural translation systems is spectacular" (by App X);
- "The performance of recent neural translation systems is remarkable" (by App Y).

As you may see, the difference in the level of performance between different translation systems for general use is hard to tell and choosing one over another is a difficult task. However, while further evaluation is required, we may see from a few more examples below that systems targeted for specific purposes, i.e., VoiceTra, as mentioned earlier, achieve better results in terms of accuracy within the targeted domain due to the sufficient number of corpora and dictionaries prepared for the respective domains.

[1]Japanese, English, Chinese, Korean, Thai, French, Indonesian, Vietnamese, Spanish, and Myanmar.

(1) [Domain: Daily living] The English translation results to the Japanese input "トイレが流れません" were:

- "The toilet doesn't flush" (by VoiceTra, Evaluation[2]: A);
- "The toilet does not flow" (by App X and Y, Evaluation: B);

(2) [Domain: Travel] the English translation results to the Japanese input "富士山五合目まではバスで行けます" were:

- "You can go to the fifth station of Mount Fuji by bus" (by VoiceTra, Evaluation A);
- "It is possible to go by bus to Mt. Fuji" (by App X, Evaluation: C); and
- "You can go by bus to the 5th of Mt. Fuji" (by App Y, Evaluation: B); and lastly,

(3) [Domain: Medical care] the English translation results to the Japanese input "認知症というと、治らないものばかりだと思っていました" were:

- "I thought dementia is not curable" (by VoiceTra, Evaluation: A);
- "When thinking of dementia, I thought that only things did not go away" (by App X, Evaluation: C); and
- "I thought that dementia was the only thing that was not cured" (by App Y, Evaluation: C).

This book explores the fundamentals of the research and development on multilingual speech-to-speech translation technology and its social implementation process which have been conducted as part of the GCP. The practicability of such technology is rapidly increasing due to the latest developments in deep-learning algorithms. Multilingual speech translation applications for smartphones are very handy, however, concerns do exist on the interactivity of the operation that may keep smooth communication from being ensured in real use; therefore, the development of translation devices for specific purposes is also ongoing. Namely, easy-to-use, high-precision translation services that are distinctive from those for smartphones and tablets are being developed by industries, and the final sections of this book will present one example of such practical technology specifically designed for hospitals. In such way, the GCP is an unprecedented type of scheme allowing open innovation to utilize multilingual translation technology in every corner of society.

Yutaka Kidawara
Director General
Advanced Speech Translation Research and
Development Promotion Center (ASTREC), NICT Kyoto Japan

[2]The samples were evaluated on a scale of A (perfectly accurate) – D (inaccurate).

Acknowledgements We would like to express our sincere gratitude to the researchers at ASTREC, NICT for their cooperation in writing this book and to Kensuke Ishii for his tremendous editorial support. We would also like to express our appreciation to the Ministry of Internal Affairs and Communications for their cooperation in promoting the GCP. Finally, we would like to thank all the participants of the Council for Global Communication Development Promotion and everyone who has taken part in the ongoing field experiments.

Contents

Contributors

Yutaka Ashikari System Development Office, Advanced Speech Translation Research and Development Promotion Center, National Institute of Information and Communications Technology, Kyoto, Japan

Masakiyo Fujimoto Advanced Speech Technology Laboratory, Advanced Speech Translation Research and Development Promotion Center, National Institute of Information and Communications Technology, Kyoto, Japan

Shoji Harada Fujitsu Laboratories Ltd., Kanagawa, Japan

Kenji Imamura Advanced Translation Technology Laboratory, Advanced Speech Translation Research and Development Promotion Center, National Institute of Information and Communications Technology, Kyoto, Japan

Hiroaki Kato Advanced Speech Technology Laboratory, Advanced Speech Translation Research and Development Promotion Center, National Institute of Information and Communications Technology, Kyoto, Japan

Hisashi Kawai Advanced Speech Technology Laboratory, Advanced Speech Translation Research and Development Promotion Center, National Institute of Information and Communications Technology, Kyoto, Japan

Tasuku Kitade Biometrics Research Laboratories, NEC Corporation, Kanagawa, Japan

Sheng Li Advanced Speech Technology Laboratory, Advanced Speech Translation Research and Development Promotion Center, National Institute of Information and Communications Technology, Kyoto, Japan

Xugang Lu Advanced Speech Technology Laboratory, Advanced Speech Translation Research and Development Promotion Center, National Institute of Information and Communications Technology, Kyoto, Japan

Jinfu Ni Advanced Speech Technology Laboratory, Advanced Speech Translation Research and Development Promotion Center, National Institute of Information and Communications Technology, Kyoto, Japan

Takuma Okamoto Advanced Speech Technology Laboratory, Advanced Speech Translation Research and Development Promotion Center, National Institute of Information and Communications Technology, Kyoto, Japan

Yoshinori Shiga Advanced Speech Technology Laboratory, Advanced Speech Translation Research and Development Promotion Center, National Institute of Information and Communications Technology, Kyoto, Japan

Fumiaki Sugaya MINDWORD, Inc., Tokyo, Japan

Eiichiro Sumita Advanced Translation Technology Laboratory, Advanced Speech Translation Research and Development Promotion Center, National Institute of Information and Communications Technology, Kyoto, Japan

Kentaro Tachibana AI System Department, AI Unit, DeNA Co., Ltd., Tokyo, Japan

Keiji Yasuda Data Science Center, Nara Institute of Science and Technology, Nara, Japan