Intelligent learning assistants help oversea students ask questions

Yi Yu

University of Queensland, Brisbane QLD, 4072, Australia

Abstract. With the development of the economy and the advancement of technology, as well as the positive effects of economic globalization, more and more students are choosing to study abroad. However, for overseas students who are not very fluent in language communication, it will be troublesome, not only for the understanding of the lesson, but also in the communication with the instructors and professors there are great obstacles, in order to stimulate the economy of study abroad, in order to make the students have less difficulties with the teachers in academic research, to achieve the real sense of barrier-free communication, and to improve the academic performance, we use automatic speech recognition (ASR) technology as a template and basic framework. In order to stimulate the economy of study abroad and to make it less difficult for students to communicate with teachers in academic research and to improve their academic performance, we designed an application for automatic speech recognition and language conversion language simplification by using ASR as a template and a basic framework, respectively.

Keywords: mobile application, machine translation, Automatic Speech Recognition.

1. Introduction

The application of automatic translation after speech to text translation. One of its aspects is the accuracy of translation and the generalization of semantics. The quality of speech recognition directly affects the translation result, and translation result directly affects the user's using experience. This paper explores the application of based on ASR for converting speech to text and translating the text into the desired language. Some research studies have shown that with the economic development and social progress, more and more students choose to study abroad, (e.g., Williams, 1987) analyzed The International Market for Overseas Students in the English-Speaking World. Many studies have focused on the reasons for the increase in the number of international students, and there are many studies that have examined the help for overseas students in order to facilitate their lives, but few studies have examined the study help for international students, and few applications have been developed to help international students. Another research study (Arlin V. Peterson, George J. Peppas, 1988), which examined how Trained Peer Helpers Facilitate Student Adjustment In an Overseas School, examined how to help overseas students adjust to a new living environment and how to improve themselves. This study looked more at how to help overseas students adjust to a new living environment and how to improve their own conditions.

Professor Mostow has made outstanding contributions in the research of ASR and so on, from the principle to research to application of ASR, Professor Mostow has made a complete system, which is of great help to our translation application.

This article is based on Professor Mostow's research on ASR, which is a novel translation application to help overseas students make up for their language deficiencies by using automatic speech recognition (ASR) and converting speech to text in the target language. data analysis.

This application provides help at different aspects, such as speech recognition, text conversion, text translation or semantic refinement. When the system recognizes the user's question, the application searches for the same question in the question bank, then refines and sublimates the user's question to present it to the professor.

The rest of paper is organized as follows. Section 2 introduce the importance of Automatic Speech Recognition(ASR) in this application Section 3 explains what models are used and how to use them. Section 4 simply clarifies educational data mining and section 5 tells the evaluation & embedded experiments. And section 6 is the whole summary of this paper.

2. Automatic Speech Recognition(ASR)

First of all, this app will use the speech recognition system (ASR), which can record what people say, then recognize and compare it with the words in thesaurus whatever the language is , the system can convert their language into the corresponding desired text. in addition, this app can also input some common academic problems in advance like ASR, and match the most similar problem by comparing in the database. What's more, some similar problems are given to compensate for the shortcomings of inaccurate identification. At the same time, users can also allow the collection of personal information to continuously expand the database. When the user has barriers in language, the question can be automatically constructed or selected from the database in the light of the keyword given by the user, then translated into the specific language which user need to comfirm.

The most important part is how to convert speech into text, which has to be applied with automatic speech recognition (ASR). First of all, the speaker permissions of the mobile phone or computer will be allowed to access the voice to record the voice, and the second is to recognize the voice. The system can detect the duration of speech as well as the duration of silence, determine when to convert to text based on the time, use a timer to determine the response time, detect errors in the spoken words and automatically match the words to the corresponding text. By analyzing the temporal alignment of the ASR output, it can calculate the time it takes for the user to recognize each word and read it aloud. By extracting the pitch, amplitude and duration of the words read, it can calculate their intonation characteristics. (Mostow Why and How Our Automated Reading Tutor Listens Thanks to the help of ASR), speech recognition,, speech to text and translation into the desired language became more convenience and accurate.

In addition, two special variables are needed to be identified. First, who is speaking. How the system detects if someone is talking, Multimodal dialogue is utilized, which will have a timer to monitor the system in real time, as stated in 'Why and How Our Automated Reading Tutor Listens' Each variable has a timer that records when it last changed. Transitions between states occur when the student or Reading Tutor starts or stops speaking, or when one of the timers reaches a specified threshold value. This system uses the same principle as reading tutors (Mostow Why and How Our Automated Reading Tutor Listens). Like reading tutors, the app, through Multimodal dialogue, can identify whether a student or contributor is speaking, and can analyze what they are talking, to detect when they start and stop talking.

The second point is the recognition of languages, there are many languages in the world, so the system can automatically identify the type of language after recording, and match the correct words, to achieve synchronous real-time conversion, truly accomplish barrier-free communication.

3. Model

To focus on speech input and text output, both acoustic and lexical models are employed. For the acoustic model, most of the target groups of the application are students older than high school age, so data from adults are more suitable for this application than data from children. Therefore, the TED-LIUM 3 dataset is utilized for project, which is a large collection of 452 hours of TED talks from 2295 speakers, and divide it into two parts. The first part is used to train the initial acoustic model for ASR, and the second part is used for evaluation. The project will use an application to identify acoustic recordings and calculate their accuracy. To train the data, the project would use Markov Chain Monte Carlo methods(MCMC) which require to specify the prior distributions and constraints for every parameter. Markov Chain Monte Carlo(MCMC):

$\sim \quad Normal(0,1)$

$bk \quad \sim Norm(0, 1)$

$ak \quad \sim Unifo(0, 2.5)$

$learnk \quad \sim Be(1, 1)$

$gues \quad \sim Uniform(0, 0.4)$

$sli \quad \sim Uniform(0, 0.4)$

We use the following conditional distributions for each node:

$\alpha(0)|\theta n \sim Bernoulli(\{1 + \exp{-1.7}\ ak(\theta n - bk)\}{-1})$

$\alpha(t)|\ \alpha nk(t{-}1) = 0 \sim Bernoulli(learnk)$

$\alpha(t)|\ \alpha nk(t{-}1) = 1 \sim Bernoulli(1)$

$Y(t)|\eta nj(t) = 0 \sim Bernoulli(guessj)$

$Y(t)|\eta nj(t) = 1 \sim Bernoulli(1 - slipj)$

Given η as a conjunction of α, the likelihood of Y given η, the conditional independence of α(0) given θ, and of α(t) given α(t-1), the posterior distribution of θ, a, b, α, η, learn (l), guess (g) and slip(s) given Y is

$(\boldsymbol{\theta}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{l}, \boldsymbol{g}, \boldsymbol{s}|\boldsymbol{Y}) \propto (\boldsymbol{Y}|\boldsymbol{g}, \boldsymbol{s}, \boldsymbol{\eta}, \boldsymbol{\alpha}) P\boldsymbol{\alpha}(0)\boldsymbol{\theta}, \boldsymbol{a}, \boldsymbol{b}$

T

$(\Pi \quad P\boldsymbol{\alpha}(t)\boldsymbol{\alpha}(t{-}1), \boldsymbol{l}) P(\boldsymbol{\theta}) P(\boldsymbol{a}) P(\boldsymbol{b}) P(\boldsymbol{l}) P(\boldsymbol{g}) P(\boldsymbol{s})$

$t = 1$

(Xu & Mostow)

The requirement is to specify a priori distributions and constraints for all parameters.

This project will set the accuracy Θn from 0 to 1, the content divergence from 0 to 2.5, and the content difficulty from 0 to 1. While collecting and centralizing the training data, this project considers the protection of user privacy. The gender and identity of the speaker can be detected. The acoustic model is dispersed to many devices and fine-tuning includes local user data (Mdhaffar, 2022). When fine-tuning occurs, one finds that leakage of user information is possible, even if no transmission occurs (Mdhaffar, 2022). How to protect the privacy of speakers is a problem that the project will focus on in the future.

For the vocabulary model, the project will autocorrect the text after the application has converted what the input text. The project hopes to find different accuracy results based on different user learning times by using a learning decomposition method, i.e., finding a learning curve. Based on accuracy and simplicity, the project will choose an exponential curve instead of a power curve (Beck & Mostow). The project creates two variables t1 and t2. t1 represents the number of learning opportunities when the user uses the application for the first time and t2 represents the number of practice opportunities when the user has already used the application that day (Beck & Mostow). The project will also estimate the parameter B, which represents the correlation between users who used the application for the first time and those who used it later.

1: if B>1
2: the learning opportunities of t1 are better
3: else if B<1
4: vice versa
5: else if B=1
6: neither is better

(Beck & Mostow)

It was found that short sentences are difficult to detect false positives, and the distraction strategy increases the false positive rate, although it improves the false positive detection rate (Mostow). Therefore, this strategy will not be used in this project. This project will use Bayesian method :

$$\mathrm{P(A|B)} = \mathrm{P(B|A)} \times \mathrm{P(A)/P(B)}$$

Bayes Rule created by Thomas Bayes). A Bayesian approach is used to evaluate the relationship between the questions posed by students and the questions in the database to provide students with the best relevant questions. For instance, when a student's question contains artificial intelligence, models, and voices, what possibilities the different questions contain are then matched to the most appropriate question.

4. Educational Data Mining(EDM)

An important goal of educational data mining is to discover what is helpful to students. This application can collect a large amount of data from students as they use it (e.g. asking questions to the teacher). With this data, we can not only optimize the performance of the software and expand the database of the application, but also personalize the features for different users and personalize it for them through big data to improve their experience. This section analyzes and discusses two kinds of data mining related to this application. They are the keywords in the questions asked by the users and the subsequent actions of the users for different types of questions.

First are the keywords in the questions. Students and teachers come from two different countries, native languages are not the same, and students may study different specialties such as medicine, computer science, biology, etc. Different disciplines have many unique specialized vocabularies, and the frequency of these specialized vocabularies in the questions asked is quite high. Collecting and analyzing students' spoken questions to find keywords that distinguish different fields of study has many benefits for categorizing users by their professional orientation. Teachers, students, administrators, technicians, content authors, programmers, and instructors themselves differ in the form of content and information they are interested in and can understand. Teachers may need a brief summary of student usage and learning progress; administrators need evidence of tutor effectiveness; technicians need alerts of problems; content authors need usability metrics; programmers need accurate bug reports; and tutors need parameters that can be used, rules that can be interpreted, or knowledge that can be leveraged. These information goals are usually not clear from the outset, but are built up and refined over the course of use. (Mostow & Beck)

First, the accuracy of speech recognition can be improved in future used by recognizing the vocabulary in a question based on the user's area of expertise. Second, alternative answers can be given that better match the user's needs based on the area of expertise. One of the issues involved is the need to build mainstream databases of specialized vocabulary and questions, as well as the ability to quickly access subject-specific language via the Internet.

Another piece of data that needs to be explored is what users do after asking different types of questions. The operation is to re-ask the question or use the answer given by the intelligent learning assistant. We can start by categorizing the questions, for example, by their introductory words. When user *A* asks "What...", he goes straight to the question. ", then he goes straight to the advice given by the Intelligent Learning Assistant and asks the next question "Why .....' . If this happens often, we can reduce the number of alternative answers given and the number of possible next questions. In other cases, user *B* asks "Why...", but he does not choose the intelligent answer. ", but instead of choosing the answer given by the intelligent learning assistant, he repeats the question. In this case, when User *B* asks "Why...' , the Intelligent Learning Assistant tends to provide more approximate answers for you to choose from.

Because students suffer from language barriers, they do not know how to ask the right question, or they do not know how to express themselves, so it is necessary to provide alternative answers. In this case, the user can have a wider range of choices. The alternative answer can be a user question approximation question or it can be beneficial for the user to predict the next question, which is necessary because people's cognitive needs range from the most basic concept to why and how to use it. When intelligent learning assistants need to analyze the operational user behavior records, they can help users to have a better user experience and improve the efficiency of using the software.

5. Evaluation&Embedded Experiments

The evaluation of the model is an important process. One of the criteria we judge is whether the model is over-fitted. Only with the right size and the right amount of tracking can machine learning make the model work as expected. Specific examples will be given shortly below. For example, if the student's words are 'Hi professor, I'm having trouble completing the reading material you gave us last week, can you walk me through it? I may have questions about it', then, to work properly, the

output of the model should be 'The student is having trouble completing the work and needs guidance', and the model should be judged by whether it conveys the appropriate meaning. If the model results in "student has a problem", there is a serious lack of information, leading to misjudgment by the professor and a waste of time. If the result is "Hi Professor, I'm having trouble completing the reading material you gave us last week, can you walk me through it? I may have questions about it', then it over-fits because it uses every word, so the problem with over-fitting means to repeat everything. However, the sentence is still too long for the professor to understand the student's claim quickly and simply. Therefore, the model function will only work as expected if it has the proper size and threshold. A second example can be found in asking students to assess. For example, what a student usually says in front of a professor is "Good morning professor, the research you asked us to do was sent via email last week, can you help me revise it now if possible? Just reply to my email after revising it still. "If you put the above sentence into the model, what you usually get is "need to revise the study by email". If the result is "needs to be revised", then it is not appropriate because it again misses important information about the sentence. What the model should do is shorten it while keeping all the information in the sentence without missing anything. if the output is "Good morning, Professor, the study you requested from us was sent by email last week, can you fix it for me if possible? Then the problem of overlap reappears, because it uses almost every word in the entire sentence, losing the quality of the results obtained in advance from our model that will be simplified.

And how to improve the ITS is important. In general, improving an intelligent tutoring system (ITS) requires that it be properly instrumented. The Instrument Data Flow view is used to determine what data to record, how to modify ITS to generate additional data, and calculations in the raw data, which variables can generate the required analysis for improving educational outcomes, and to generate the desired visualization view. It is first necessary to determine what information each user will need to achieve this goal. These information goals motivate instrumentation decisions. Teachers, students, administrators, technical support staff, content authors, software developers, researchers, and ITS itself differ in the content and form of information they are interested in and can understand. Teachers may need simple summaries of student usage and progress; administrators need evidence of ITS effectiveness; technical support staff need problem alerts; content authors need usability metrics; developers need accurate bug reports; researchers need detailed examples and information analysis; and ITSs need parameters that can be used, rules that can be interpreted, or knowledge that can be modified to exploit. (Mostow)

The second criterion for evaluating the model is about the model's translation process. The reason for designing the translation function is to help international students understand English and communicate more effectively with others. However, the experience of using it depends on the quality of the translation. In the case of Chinese, for example, a word may have many meanings in different contexts, so if the translation messes up the meaning, it may be difficult for the professor to understand what the student is asking for. A concrete example can clarify this issue. If the real message the student wants to say is "Oh my gosh, having trouble with the paper and needing guidance (obviously the student is having trouble understanding English grammar)" in Chinese, then a good translation would be "The student is having trouble with the academic paper and needs help ", while a bad translation would be "God is having trouble with the academic paper, in which case the translator has carefully understood the student's lament and put the student in a difficult situation. The solution to this problem is still to resize the batch and perform machine learning on the model. This is entirely about the evaluation of the model and on the next topic we will discuss the experiments that can be performed on the model.

A specific experiment that can be done on recognizing context. The machine is required to learn the understanding of the context, either in a positive or negative tone, whether the student is urgent or not be expressed in the results of the model. We designed the experiment to use several sentences with different tones of voice and to determine the meaning of each sentence, which is important because the results may affect the way and order of translation. The professor hopes that scheduling his teaching with all students will make the teaching schedule more flexible and the students' learning

as well as the professor's work more relaxing.For example, "I'm having a hard time understanding all of this because this assignment is coming up tomorrow!" implies that the student is really anxious to ask the professor for help, while "I may be in trouble in the next few days, so I may need your help" implies that the student is not anxious.

6. Conclusion

ASR models play an indispensable role in APP design research, whether it is for inputting speech, converting speech to text, or for language recognition of text, target language conversion, or semantic refinement. ASR provides a great convenience for inputting speech, converting speech to text, language recognition of text, conversion of target language, and semantic refinement. And Professor Moscow has created irreplaceable value and importance in research on these four areas. This application will be of great help to overseas exchange and study, not only to promote the increase of international students, but also to promote the development of education, academic exchange and study skills, which are necessary.

Acknowledgements

This work was supported by Professor Mostow, theNational Science Engineering. Anyopinions, findings and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science.

References

[1] Jack Mostow, & Joseph E. Beck (2006). Some useful tactics to modify, map and mine data from intelligent tutors Natural Language Engineering.

[2] Jack Mostow , Why and How Our Automated Reading Tutor Listens , Project LISTEN (www.cs.cmu.edu/~listen), School of Computer Science Carnegie Mellon University

[3] Salima Mdhaffar, Jean-Fran\\c{c}ois Bonastre, Marc Tommasi, Natalia Tomashenko, & Yannick Est\\`eve (2022). Retrieving Speaker Information from Personalized Acoustic Models for Speech Recognition

[4] Yanbo Xu and Jack Mostow , A Unified 5-Dimensional Framework for Student Models , Carnegie Mellon University Project LISTEN

[5] Joseph E. Beck, & Jack Mostow (2008). How Who Should Practice: Using Learning Decomposition to Evaluate the Efficacy of Different Types of Practice for Different Types of Students intelligent tutoring systems.

[6] Jack Mostow,Some Useful Design Tactics for Mining ITS D ata , Project LISTEN, School of Computer Science Carnegie Mellon University

[7] Yanbo Xu, & Jack Mostow (2014). A Unified 5-Dimensional Framework for Student Models. EDM (Workshops).