SPEECH RECOGNITION USING GENERIC SHORTEST DISTANCE ALGORITHM

Mahak Dureja*

Department of CSE, THE NORTHCAP University, Address,
Gurgaon, Haryana122017, India
mdureja7@gmail.com

Sumanlata Gautam

Department of CSE, THE NORTHCAP University, Address,
Gurgaon, Haryana122017, India
sumanlatagautam@ncuindia.edu

Abstract

The first major contribution of this paper may well be a dialog model support propositional content and intentions. Two distinct components of the model square measure mentioned in depth: (a) associate approach to the modeling and trailing of propositional content that support description logics and default unification. (b) Associate approach to the modeling and trailing of intentions that is supported dialogue acts, dialogue moves and dialog games alongside dialogue phases. These building blocks could also be organized in a way the best that the complete dialogue is delineate on five utterly completely different levels terribly very language-independent way.

A replacement characterization referred to as dialogue moves is introduced that cover several dialogue acts. The system supports zero verbal communications with foreign interlocutors in mobile things. It acknowledges spoken input, analyses and interprets it, and ultimately utters the interpretation. This polyglot system handles dialog in three business-oriented domains, with duplex translation between two languages (English, and Hindi).

Keywords: Speech Recognition; speech to speech translation; machine translation; text to speech synthesis; mathematics linguistics.

1. Introduction

Recently, speech to speech translation (s2s) technology has competed associate a lot of and a lot of necessary part in reducing the roadblock in cross lingua social communication. The enhancements in automatic speech recognition (ASR), mathematics linguistics (MT) and text-to-speech synthesis (TTS) technology has enabled the serial binding of these individual modules to realize S2S translation of acceptable quality. Earlier work on S2S translation has targeted on giving either one-way or two-way translation on one device (Waibel et al., 2003; Zhou et al., 2003). The program wants the user to select the availability and target language a priori. “The character of communication, either single user talking or flip taking between a pair of users might find you throughout a one-way or cross-lingual dialog interaction. In most of the systems, the need to select the imbalance of translation for each flip can take away from a natural dialog flow. Moreover, single interface primarily based S2S translation is not suitable for cross lingual communication. In such a state of affairs, it’s imperative to supply amount and low latency communication. Throughout a typical language between a pair of sneakers of identical language, the interaction is amount. Likewise, cross lingual dialog between a pair of remote users will critically profit through progressive translation. Whereas progressive secret writing for text translation was addressed earlier in Furuse and lida, 1996; Sankarn et al., 2010, tend to handle the matter throughout speech to speech translation setting for facultative amount cross-lingual dialog. They tend to handle the matter of incrementally throughout a completely unique session initiation protocol (SIP) primarily based on S2S translation system that permits a pair of users to maneuver and move in cross lingual dialog over mobile phone or landline. Our system implements progressive speech recognition and translation, holding latency of low interaction which includes a good setting for remote dialog targeted towards achieving a task.

The system delineate on high of provides a natural due to collect cross lingual dialog data. We tend to use the system to assemble forty written dialogs in English as well as Hindi. A bilingual speaker generated dialog things inside the travel domain and so the written dialog was used as a reference inside the choice, a pair of subjects inside the data assortment, a female English speaker and Hindi speaker. For scanning the lives verbatim, the tools were schooled. Because of ASR errors, the topics need to improvise few turns (about 150) to withstand the dialog. The common style of rounds per state of affairs inside the collected corpus 7 turns per state of affairs for English and Hindi severally.

2. Proposed Framework

In order to create voice recognition software in C# programming language we are using Visual Studio 2015 RC. In that application we will call MATLAB functions from C# client, so we must use COM (Component Object Model). To creating a program with speech to speech recognition in C# System Speech library must be used. Speech Recognition Engine class is available in .NET 4.5, 4, 3.5, 3.0 and .NET 4 Client Profile.

3. Training For Speech Recognition Engine

Following the tutorial, after coding, the speech recognition engine must be trained. Unfortunately, this is impossible to accomplish from the code itself. It can only be done using the Windows Speech Recognition:

- Open Control Panel
- Go to Ease of Access
- Choose Speech Recognition
- Then choose "Train your computer to better understand you

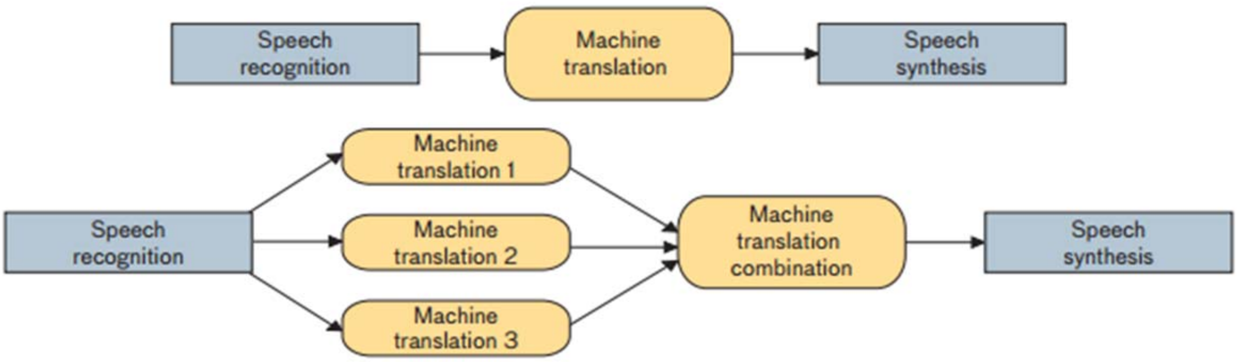

FIG 1: BASIC FRAMEWORK

To acknowledge the spoken words from associate degree of acoustic speech signal recognition system uses HMMs that is "Hidden Markov Models". HMMs measures the way of math modeling for the system with the mathematician processes with the state of unobservant speech. Number of acoustic signals observations is taken by HMM with some associate degree for modeling the probabilities of observations given potential for the unobservant states that the spoken words are inside the acoustic signal. The model unit of measurement developed victimization work data which offers the concept of developing math distribution that represents observed data and the transitions potential. They looked for W* that is the translated speech, which performs considering associate degree acoustic signal and computes the foremost attainable words sequence that generates the discovered signal. This methodology is the combination of word generating associate degree and the words showing a particular sequence of words. Various ASR systems uses numerous parameters for modeling P(o|W) P(W) where P(O|W) denotes probabilistic operation that used to model W to the acoustic observation of the acoustic model O. and P(W) is likelihood of generating the word W providing the language mode.

Modern systems that are fully optimized, will use quite 10 times the period of time laptop resources throughout coding once increasing accuracy. Solely small scale ASR systems will be running on terribly restricted laptop resources such as mobile processors. Moreover, the system has a major starting time in loading model and precipitation. Speech to speech applications uses both large as well as small scales ASR, looking for the target application. In 2009, distributed design for speech-to-speech translation is given in "Lincoln laboratory

Journal", that have a tendency to build use of associate degree internally developed, giant vocabulary system further as many industrial off-the-peg (COTS) systems. Computational linguistics presently uses the applied mathematics models for most effective MT systems which find the translations in target words. This method produces an outsized memorized table of translation pairs, referred to as phrase table that links word from source language to the desired target language. Throughout translation, the most effective hypothesis T* is outlined by: T S * a = rgmax P(S|T) P(T). P(T) like ASR, P (T) could be a language model shaped via n-grams multiple words and P(S|T) is the translation model that translates source language S to target language Tor simply the phrase table. The modeling and coding processes for MT systems, except for the character of the info itself, square measure parallel processes of coaching associate degreed coding an HMM. In giant applications, the phrase table and language model will contain many millions or maybe billions of entries. As a result, the search method accustomed rewrite the most effective translation for a given input sentence are often computationally costly.

4. Algorithm

Generic-Single-Source-Shortest-Distance (G, s)

1. for i ← 1 to |Q|
2. do d[i] ← r[i] ← 0
3. d[s] ← r[s] ← 1
4. S ← {s}
5. while S $= ∅
6. do q ← head(S)
7. Dequeue(S)
8. r" ← r[q]
9. r[q] ← 0
10. for each e ∈ E[q]
11. do if d[n[e]] $= d[n[e]] ⊕ (r" ⊗ w[e])
12. then d[n[e]] ← d[n[e]] ⊕ (r" ⊗ w[e])
13. r[n[e]] ← r[n[e]] ⊕ (r" ⊗ w[e])
14. if n[e] $∈ S 15 then Enqueue(S, n[e])

5. Conclusion

Innovative ideas and the coordination of multiple approaches that constructs the knowledge bases manually and also learning methods were applied to the constructed corpus that yields the successful 'dialogue module'. Robust techniques were considered for the huge amount of data besides of focusing on high precision algorithms. This huge amount of data has been interpreted and elucidated with different linguistic models data. Our first provocation for the system is speech in one language gets translated in another language but what we done so far is capable speech recognition system with innovative generic single source shortest distance algorithm which has its own advantages over the other speech recognition algorithms. Although different phenomena like hesitations and deliberations in human speech make the system little away from perfect. Future scope for the thesis is that the system is capable of running in real time.

REFRENCES

[1] R. Frederking, A. Rudnicky, C. Hogan, and K. Lenzo, “Interactive speech translation in the diplomat project,” Machine Translation Journal, vol. special issue on spoken language translation, 2000.

[2] A. Black, R. Brown, R. Frederking, R. Singh, J. Moody, and E. Steinbrecher, “Tongues: Rapid development of a speechto-speech translation system,” in HLT2002, San Diego, 2002, pp. 2051–2054.

[3] R. Frederking, A. Black, R. Brown, J. Moody, and E. Steinbrecher, “Field testing the tongues speech-to-speech machine translation system,” in LREC, 2002.

[4] X. Huang, F. Alleva, H.-W.Hon, K.-F.Hwang, M.-Y. Lee, and R. Rosenfeld, “The SPHINX-II speech recognition system: an overview,” Computer Speech and Language, vol. 7(2), pp. 137–148, 1992.

[5] T. Schultz and A. Waibel, “The globalphone project: Multilingual lvcsr with janus-3,” in Multilingual Information Retrieval Dialogs: 2nd SQEL Workshop, Plzen, Czech Republic, 1997, pp. 20–27.

[6] A. Black and K. Lenzo, “Optimal data selection for unit selection synthesis,” in 4rd ESCA Workshop on Speech Synthesis, Scotland 2001.

[7] A. Black and K. Lenzo, “Building voices in the Festival speech synthesis system,” http://festvox.org, 2000

[8] A. Black, P. Taylor, and R. Caley, “The Festival speech synthesis system,” http://festvox.org/festival, 1998

[9] A. Black and K. Lenzo, “Limited domain synthesis,” in ICSLP2000, Beijing, China., 2000.

[10] F. Malfrere and T. Dutoit, “High quality speech synthesis for phonetic speech segmentation,” in Eurospeech97, Rhodes, Greece, 1997, pp. 2631–2634

[11] R. Frederking and R. Brown, “The Pangloss-Lite Machine Translation System,” in AMTA, Montreal, Quebec, Canada, October 1996, pp. 268–272.