PSEUDO-MORPHEME AND CONFUSION NETWORK BASED KOREAN-ENGLISH STATISTICAL SPOKEN LANGUAGE TRANSLATION SYSTEM

Donghyeon Lee, Jonghoon Lee, Gary Geunbae Lee

Department of Computer Science & Engineering
Pohang University of Science and Technology, Pohang, South Korea
{semko, jh21983, gblee}@postech.ac.kr

ABSTRACT

In this demonstration, we present POSSLT (POSTECH Spoken Language Translation) for a Korean-English statistical spoken language translation (SLT) system using pseudo-morpheme and confusion network (CN) based technique. Like most other SLT systems, automatic speech recognition (ASR) and machine translation (MT) are coupled in a cascading manner in our SLT system. We used confusion network based approach to couple ASR and MT. It has better translation quality and faster decoding time than N-best approach. In the ASR and SMT for Korean, how to define processing units affects the performance. Pseudo-morpheme unit is a best choice for Korean-English SLT. Models used in SLT system are trained on a travel domain conversational corpus.

Index Terms— Spoken Language Translation, Statistical machine translation, Conversational corpus

1. INTRODUCTION

Recently, SLT has become increasingly important, because the technology enables two people to communicate across a language barrier. Currently, most of statistical SLT systems are achieved in a cascading method. In the cascading approach, SLT system is usually composed of three major components: a automatic speech recognition (ASR) part, a statistical machine translation (SMT) part and a speech synthesis (TTS) part. Simply, SLT systems can be developed by translating a single best recognizer output. But, translation quality can be improved using the N-best hypothesis, lattice or confusion network provided by ASR [1] [2] [3].

In the ASR and SMT, which are major components of the SLT, how to define recognition and translation units is most important. Especially case of agglutinative language such as Korean, performances of ASR and SMT are significantly different by the processing unit. Some researches on Korean ASR and Korean-English SMT have been carried out [4] [5].

In our POSSLT, we use a confusion network as an interface with ASR and SMT. We define the pseudo-morpheme and use it as the recognition and translation unit.

2. SYSTEM OVERVIEW

The POSSLT was developed by integrating ASR, SMT, and TTS. The system has a pipelined architecture. LM loader is added to increase the ASR coverage. Figure 1 shows the overall architecture of the POSSLT.

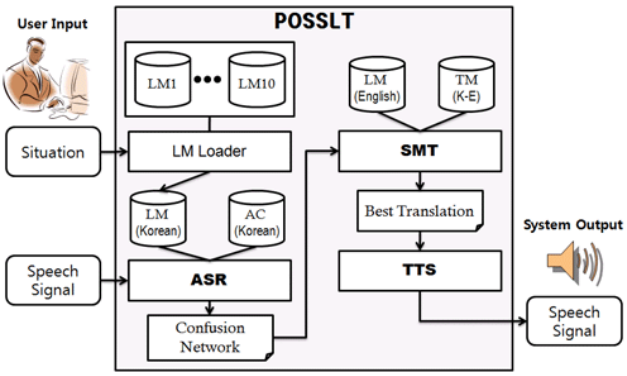

Figure 1. Overview of POSSLT

2.1. ASR

The system used HTK-based continuous speech recognition engine properly trained for Korean [6]. The phonetic set for Korean has 48 phoneme-like-units, and we used three-state tri-phone hidden Markov models and trigram language models. Pronunciation lexicons are automatically built by a Korean grapheme-to-phoneme (G2P) tool [7]. The ASR produces the N-best lists, lattices, confusion networks determined through the decoding process

2.2. SMT

We implemented a phrase-based SMT decoder based on Pharaoh [8]. The decoder needs a phrase translation model for the Korean-English pair and a language model for English. We used the Pharaoh training module and GIZA++

[9] to construct the phrase translation table. For language modeling, SRILM toolkit [10] was used to build a trigram language model.

This phrase-based SMT decoder can be expanded to CN-based decoder. CN-based SLT decoder is substantially the same as the phrase-based SMT decoder apart from the way the input is managed. Probabilities of CN are used as a feature in log linear model.

2.3. TTS

Synthesizer just pronounces the final translation result. Our systems used Microsoft SAPI 5.1 TTS engine for English TTS.

2.4. LM Loader

In cascading SLT system, SMT coverage depends on the used ASR. In order to increase the ASR coverage, our system loads and unloads ASR language models dynamically. In our system which uses a travel corpus, language models are built for ten domain situation categories such as an airport, a hotel, a shopping, etc. Besides user utterances, user selection of the situation is needed as an input to decide which language model have to be loaded in advance. By using the divided language models, many benefits such as fast decoding, higher accuracy and more coverage can be obtained.

3. PROCESSING UNITS

In Korean-English CN-based decoder, translation unit must be the same as recognition unit because CN-based decoder directly uses CN in decoding process. We should define the processing unit before building models used in SLT. Various units can be the recognition and translation units. Figure 2 shows various processing units for SLT. In figure 2, the 'V' mark means a space.

< Eojeol >
- 제∨말을∨이해하시겠어요

< Morpheme >
- 제∨말∨을∨이해∨하∨시∨겠∨어요

< Pseudo-morpheme >
- 제∨말∨을∨이해하∨시겠어요

Figure 2. Various Processing Units for SLT

Pseudo-morpheme is produced by coupling pairs of short and frequent morphemes into larger units. There are trade-offs between length of unit and performance in ASR and SMT. Pseudo-morpheme based unit has the better SLT quality than eojeol based and morpheme based one because it helps to obtain relatively high accuracy on recognition results despite a little loss of translation performance.

4. POSSLT IMPLEMENTATION

The POSSLT framework was implemented using a C++ and visual studio. The screenshot of the POSSLT operation is shown in Fig. 3.

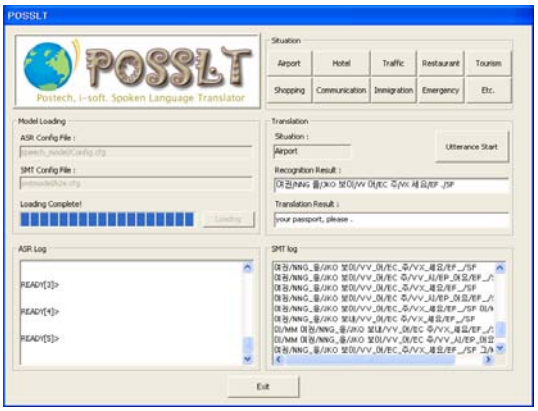

Figure 3. Screen shot of POSSLT

5. REFERENCES

[1] R. Zhang, G. Kikui, H. Yamamoto, T. Watanabe, F. Soong, and W. K. Lo, "A unified approach in speech-to-speech translation: Integrating features of speech recognition and machine translation," in Proc. of Coling 2004, Geveva, 2004.

[2] S. Saleem, S. Chen Jou, S. Vogel, and T.Schultz, "Using word lattice information for a tighter coupling in speech translation systems," in Proc. of ICSLP 2004, Jeju, Korea, 2004.

[3] N. Bertoldi, M. Federico, "A New Decoder for Spoken Language Translation based on Confusion Networks", In Proceedings of the Automatic Speech Recognition and Understanding Workshop (ASRU), San Juan, Puerto Rico, November-December 2005.

[4] O.W. Kwon, J. Park, "Korean large vocabulary continuous speech recognition with morpheme-based recognition units," Speech Communication, Vol. 39, No. 3-4, pp. 287-300, 2003.

[5] Jonghoon Lee, Donghyeon Lee, Gary Geunbae Lee, "Improving Phrase-based Korean-English Statistical Machine Translation," Proceedings of Interspeech-ICSLP, 2006.

[6] http:// htk.eng.cam.ac.uk

[7] Jinsik Lee, Seungwon Kim, Gary Geunbae Lee, "Grapheme-to-Phoneme Conversion Using Automatically Extracted Associative Rules for Korean TTS System," Proceedings of Interspeech-ICSLP, 2006.

[8] P. Koehn, "Pharaoh: A Beam Search Decoder for Phrase-based Statistical Machine Translation Models," in Proc. of AMTA, Washington DC, 2004.

[9] F. J. Och and H. Ney, "Improved statistical alignment models," in Proc. of 38th Annual Meeting of the ACL, page 440-447, Hongkong, China, October 2000.

[10] A. Stolcke, "SRILM – An Extensible Language Modeling Toolkit," in Proc. of ICSLP, 2002.