10th Annual Conference of the International Speech Communication Association

Interspeech 2009 Brighton

6-10 September 2009

Abstracts

Efficient Combination of Confidence Measures for Machine Translation

Sylvain Raybaud (LORIA)
David Langlois (LORIA)
Kamel Smaili (LORIA)

Abstract We present in this paper a twofold contribution to Machine Translation. First, we present a method to automatically build training and testing corpora for confidence measures containing realistic errors. Errors introduced into reference translation simulate classical machine translation errors (word deletion and word substitution), and are supervised by Wordnet. Second, we use SVM to combine original and classical confidence measures both at word- and sentence-level. We show that the obtained combination outperform by 14% (absolute) our best single word-level confidence measure, and that sentence-level combination of confidence measures produces meaningful scores.

Iterative Sentence-Pair Extraction from Quasi-Parallel Corpora for Machine Translation

Ruhi Sarikaya (IBM T.J. Watson Research Center)
Sameer Maskey (IBM T.J. Watson Research Center)
Rong Zhang (IBM T.J. Watson Research Center)
Ea-Ee Jan (IBM T.J. Watson Research Center)
Dagen Wang (IBM T.J. Watson Research Center)
Bhuvana Ramabhadran (IBM T.J. Watson Research Center)
Salim Roukos (IBM T.J. Watson Research Center)

Abstract This paper addresses parallel data extraction from the quasi-parallel corpora generated in a crowd-sourcing project where ordinary people watch tv shows and movies and transcribe/translate what they hear, creating document pools in different languages. Since they do not have guidelines for naming and performing translations, it is often not clear which documents are the translations of the same show/movie and which sentences are the translations of the each other in a given document pair. We introduce a method for automatically pairing documents in two languages and extracting parallel sentences from the paired documents. The method consists of three steps: i) document pairing, ii) sentence pair alignment of the paired documents, and iii) context extrapolation to boost the sentence pair coverage. Human evaluation of the extracted data shows that 95% of the extracted sentences carry useful information for translation. Experimental results also show that using the extracted data .....

Using Syntax in Large-Scale Audio Document Translation

Jing Zheng (SRI International)
Necip Fazil Ayan (SRI International)
Wen Wang (SRI International)
David Burkett (UC Berkeley)

Abstract Recently, the use of syntax has very effectively improved machine translation (MT) quality in many text MT tasks. However, using syntax in speech MT poses additional challenges because of disfluencies and other spoken language phenomena, and of errors introduced by automatic speech recognition (ASR). In this paper, we investigate the effect of using syntax in a large-scale audio document translation task targeting broadcast news and broadcast conversations. We do so by comparing the performance of three synchronous context-free grammar based translation approaches: 1) hierarchical phrase-based translation, 2) syntax-augmented MT, and 3) string-to-dependency MT. The results show a positive effect of explicitly using syntax when translating broadcast news, but no benefit when translating broadcast conversations. The results indicate that improving the robustness of syntactic systems against conversational language style is important to their success and requires future effort.

Grapheme to phoneme conversion using an SMT system

Antoine Laurent (Laboratoire Informatique Université du Maine (LIUM))
Paul Deléglise (Laboratoire Informatique Université du Maine (LIUM))
Sylvain Meignier (Laboratoire Informatique Université du Maine (LIUM))
Abstract This paper presents an automatic grapheme to phoneme conversion system that uses statistical machine translation techniques provided by the Moses Toolkit. The generated word pronunciations are employed in the dictionary of an automatic speech recognition system and evaluated using the ESTER 2 French broadcast news corpus. Grapheme to phoneme conversion based on Moses is compared to two other methods: G2P, and a dictionary look-up method supplemented by a rule-based tool for phonetic transcriptions of words unavailable in the dictionary. Moses gives better results than G2P, and have performance comparable to the dictionary look-up strategy.

Localization of Speech Recognition in Spoken Dialog Systems: How Machine Translation Can Make Our Lives Easier

David Suendermann (SpeechCycle, Inc.)
Jackson Liscombe (SpeechCycle, Inc.)
Krishna Dayanidhi (SpeechCycle, Inc.)
Roberto Pieraccini (SpeechCycle, Inc.)
Abstract The localization of speech recognition for large-scale spoken dialog systems can be a tremendous manual exercise. Usually though, a vast number of transcribed and annotated utterances exists for the source language. In this paper, we propose to use such data and translate it into the target language using machine translation. The translated utterances and their associated (original) annotations are then used to train statistical grammars for all contexts of the target system. As an example, we localize an English spoken dialog system for Internet troubleshooting to Spanish by translating more than 4 million source utterances without any human intervention. In an application of the localized system to more than 10,000 utterances collected on a similar Spanish Internet troubleshooting system, we show that the overall accuracy was only 5.7% worse than that of the English source system.

Using Same-Language Machine Translation to Create Alternative Target Sequences for Text-To-Speech Synthesis

Peter Cahill (University College Dublin)
Jinhua Du (Dublin City University)
Andy Way (Dublin City University)
Julie Carson-Berndsen (University College Dublin)
Abstract Modern speech synthesis systems attempt to produce speech utterances from an open domain of words. In some situations, the synthesiser will not have the appropriate units to pronounce some words or phrases accurately but it still must attempt to pronounce them. This paper presents a hybrid machine translation and unit selection speech synthesis system. The machine translation system was trained with English as the source and target language. Rather than the synthesiser only saying the input text as would happen in conventional synthesis systems, the synthesiser may say an alternative utterance with the same meaning. This method allows the synthesiser to overcome the problem of insufficient units in runtime.

Human Translations Guided Language Discovery for ASR Systems

Sebastian Stüker (Institut für Anthropomatik, Universität Karlsruhe (TH))
Laurent Besacier (Laboratory of Informatics of Grenoble (LIG), University J. Fourier)
Alex Waibel (Institut für Anthropomatik, Universität Karlsruhe (TH))
Abstract The traditional approach of collecting and annotating the necessary training data is due to economic constraints not feasible for most of the 7,000 languages in the world. At the same time it is of vital interest to have natural language processing systems address practically all of them. Therefore, new, efficient ways of gathering the needed training material have to be found. In this paper we continue our experiments on exploiting the knowledge gained from human simultaneous translations that happen frequently in the real world, in order to discover word units in a new language. We evaluate our approach by measuring the performance of statistical machine translation systems trained on the word units discovered from an oracle phoneme sequence. We improve it then by combining it with a word discovery technique that works without supervision, solely on the unsegmented phoneme sequences.

Language Identification for Speech-to-Speech Translation

Daniel Chung Yong Lim, Ian Lane (Carnegie Mellon University, USA)
Abstract This paper investigates the use of language identification (LID) in real-time speech-to-speech translation systems. We propose a framework that incorporates LID capability into a speech-tospeech translation system while minimizing the impact on the system s real-time performance. We compared two phone-based LID approaches, namely PRLM and PPRLM, to a proposed extended approach based on Conditional Random Field classifiers. The performances of

these three approaches were evaluated to identify the input language in the CMU English-Iraqi TransTAC system, and the proposed approach obtained significantly higher classification accuracies on two of the three test sets evaluated.