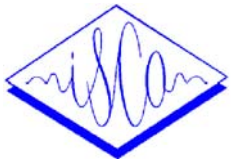

6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

A MONOLINGUAL SEMANTIC DECODER BASED ON WORD SENSE DISAMBIGUATION FOR MIXED LANGUAGE UNDERSTANDING

LIU Xiaohu, Pascale FUNG, CHEUNG Chi Shun

Human Language Technology Center
Department of Electronic and Electrical Engineering
University of Science and Technology, HKUST
Clear Water Bay, Hong Kong
{lxiaohu, pascale, eepercy}@ee.ust.hk

10.21437/Eurospeech.1999-444

ABSTRACT

In this paper, a new method for spoken mixed language understanding is presented. By mixed language, we mean that the words included in one sentence may come from different languages, a *primary language* and a *secondary language*. In conventional statistical semantic decoders, the conceptual structure is represented as a hidden Markov model, the decoding of the conceptual content of a sentence is carried out with the Viterbi algorithm. To handle mixed language, an unsupervised word sense disambiguation module is proposed to convert the secondary language words into the primary language. The approach is evaluated in the ATIS domain, where the primary language is English and we assume the secondary language is Chinese. The average accuracy of our extended semantic decoder is 26% higher than the accuracy of the baseline semantic decoder. The advantages of the extended semantic decoder are (1) it can handle mixed language input, and (2) it needs neither secondary language training data nor mixed language training data. The approach can be used for any main-secondary language pairs.

1. INTRODUCTION

For most spoken dialogue systems, each query or utterance is limited to a specific language. Most multilingual systems consist of a collection of monolingual systems behind a language-identification front-end. Such systems perform poorly when the input is in mixed language, as often is the case in Hong Kong (Cantonese/English, Mandarin/English) or some other countries. It is difficult to understand such kind of sentences using current monolingual semantic decoders.

The recognizer of our speech-assisted online search agent (SALSA)[1] is designed to accept mixed language input. The user can retrieve information in mixed language (English, Mandarin and Cantonese). The understanding module is in charge of understanding the query semantics.

The characteristics of mixed languages are (1) it is not mono-lingual, it involves two (or more) languages; (2) it is not multi-lingual either, because the words in a single sentence may come from different languages.

Conventional statistical semantic decoders use word sequence S to represent the sentence, and concept sequence C to represent the meaning of a sentence. To understand S is to find C^* that maximizes the posterior probability $P(C|S)$. Since the conceptual structure is represented as a hidden Markov model, the decoding of the conceptual content of a sentence can be carried out with the Viterbi algorithm.

To understand a mixed language query, we cannot use statistical parsers [2,3] because it is difficult to train a parser for mixed language. The traditional direct channel models for semantic decoder, which is used in many spoken language understanding systems [4,5] cannot handle mixed language either. In addition, we cannot use independent statistical natural language understanding models for each language as in multi-lingual understanding [6].

To extend the baseline semantic decoder to handle mixed language, we propose an unsupervised word sense disambiguation module to convert the secondary language words into the primary language. From a separate monolingual corpus in the primary language, we pre-compute the mutual information score between any two words. During semantic decoding, for a given secondary language word CW_x (next to a neighboring primary language word EW_y) in the mixed language

sentence *S*, the disambiguation module looks up *n* possible primary language word candidates from an online dictionary, $EW_{x1}, EW_{x2}, \ldots, EW_{xn}$. Among these candidates, it selects the word EW_{xi} that has the highest mutual information score with EW_y. The extended semantic decoder can transfer mixed language sentences into primary language sentences and annotate each word with a semantic class.

To evaluate this approach, a set of mixed language sentences is generated from the ATIS corpus, where the primary language is English and we assume the secondary language is Chinese. The experimental results between our extended semantic decoder and the baseline semantic decoder are compared. For mixed language sentences, the average accuracy of extended semantic decoder is 26% higher on the average.

Our methodology is described in section 2 in detail, and the experiments are shown in section 3. Finally, we conclude in section 4.

2. METHODOLOGY

There are four main modules in the SALSA system[1], speech recognizer, spoken language understanding module, dialogue module and verbalization module. The module of mixed language recognizer converts the speech uttered by the user into text string. The spoken language understanding module finds the underlying meaning of the text string and represents the meaning as a sequence of semantic class. The dialogue module performs the responding actions according to the meaning of input query and dialog history. Query result is sent back to the user by verbalization module via speech and text.

Without losing generality and to make it easy to explain our method, the primary language is assumed to be English and the secondary language is Chinese. The kernel of understanding module semantic decoder, is trained from a English corpus, where the words are tagged with semantic class by hand. For any new input sentence, we convert the sentence into English using statistical word sense disambiguation. This converted sentence is then processed by the semantic decoder.

2.1 Baseline Semantic Decoder

Our baseline semantic decoder is similar to [4]. We assume that the understanding problem is a noisy channel transfer problem and represent the conceptual structure as a hidden Markov model. To understand a sentence is to maximize the posterior probability of conceptual sequence given a sentence.

The problem of understanding a sentence can be expressed in these terms: given a word sequence *S* recognized by speech recognizer, we want to find the concept sequence *C* most likely produced it, namely the one for which the posterior probability $P(C \mid S)$ is maximum.

Using a sequence of words to represent the sentence, and a sequence of concept *C* to represent the meaning of a sentence, we represent word sequence *S* and concept sequence *C* as

$$S = W_1, W_2, \ldots, W_n \quad (1)$$

$$C = C_1, C_2, \ldots, C_n \quad (2)$$

The goal is to find *C** that maximize the posterior probability *P(C|S)*. Using the Bayes inversion formula, *C** is computed with the following equation

$$\arg\max_C P(C \mid S) = \arg\max_C P(S \mid C) P(C) \quad (3)$$

The parameters in the above formula can be computed from a training corpus annotated with conceptual labels. And given a text string, the decoding of the conceptual content of a sentence can be carried out by the Viterbi algorithm.

We use a partial parser for preprocessing. The partial parser is in charge of grouping phrases, handling number, date, time phrases, etc. We argue that the statistic data learned from the corpus where the basic unit is phrase is more meaningful than from the corpus where the basic unit is word.

2.2 Word Sense Disambiguation

Ideally, we want to collect a mixed language corpus annotated with semantic labels (concept), to train a mixed language semantic decoder. The problem is that it is difficult and time consuming to collect such kind of corpora. Therefore, our challenge is to learn a semantic decoder without collecting multi-lingual or mixed language corpora.

The relationship between modules is shown in Fig. 1. The word sense disambiguation module converts the mixed language query recognized by speech recognizer into mono-lingual query. We use an online dictionary and the statistics obtained from primary language corpus to disambiguate mixed language query. The monolingual query is then processed by the baseline semantic decoder.

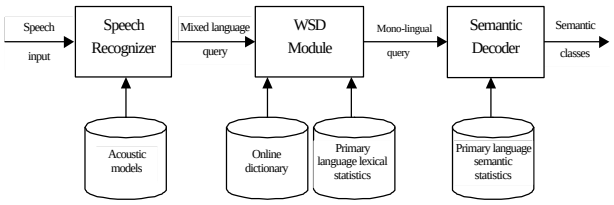

Fig.1. Relationship between modules

Conventional machine translation systems convert text in one language into another language. All words in the source language need to be converted into words in the target language with the same meaning. In our case, what we need to do is to translate each secondary language word into primary language word. Therefore for mixed language understanding, we will focus on the most important technique word sense disambiguation.

We can find all translation candidates of a Chinese word in the query by looking up an online Chinese-English dictionary. The aim of the word sense disambiguation module is to select the correct English word from the candidate set. We adopt an unsupervised statistical method. To make it easy to explain our approach, the primary language is assumed to be English and the secondary language is Chinese.

Co-occurrence information is used to weight all translation candidates for a Chinese word. We postulate that the correct translation of Chinese word *C* should co-occur frequently with the contextual words of *C*, and incorrect translation of *C* should co-occur rarely with the contextual words.

Mutual information is a good measure of the co-occurrence relationship between two words[7]. From an English corpus, we compute the mutual information between any two words using the following formula, where *c* is a constant, *EW* is an English word and *f()* is counting function.

$$MI(EW_1, EW_2) = \log \frac{P(EW_1, EW_2)}{P(EW_1) * P(EW_2)} = \log \frac{c * f(EW_1, EW_2)}{f(EW_1) * f(EW_2)} \quad (4)$$

Based on the mutual information, the translation candidate for a Chinese ambiguous word in a sentence is selected as follows: for a given Chinese word CW_x in the sentence *S*, we want to choose the correct target word from *n* target English words, EW_{x1}, EW_{x2},…, EW_{xn}. Suppose the nearest neighboring English word in query *S* is EW_y, we select the target word EW_{xi}, so that the mutual information between EW_{xi} and EW_y is maximum.

$$i = \arg\max_j MI(EW_{xj}, EW_y) \quad (5)$$

3. EXPERIMENTS

3.1 Training Data and Testing Data

We perform experiments for the ATIS understanding task. All the training and testing sentences are extracted from the ARPA ATIS corpus[8]. We train the baseline semantic decoder with 1,000 sentences annotated with 43 different concepts an additional 500 testing sentences are also extracted from ATIS corpus. Since ATIS corpus does not include Chinese words, we replace some English words randomly in the testing set with Chinese words to construct the final mixed language testing data.

3.2 Experimental Results

Although all Chinese words in a testing sentence are translated from English words, we neglect the original English words when we translate the Chinese words into English at the testing stage. Therefore a Chinese word *CW* translating from the original English word *EW* may be translated into another English word *EW'* to annotate the semantic class for *CW*.

If the semantic class for *EW'* is the same as the semantic class for *EW* in the original ATIS corpus, the label for *CW* is regarded as correct.

Table 1. The accuracy of semantic decoder for different testing data

Ratio of English words	0.4	0.5	0.6	0.7	0.8	0.9	1.0	Average
Baseline semantic decoder accuracy	21.3%	23.1%	37.5%	59.1%	63.1%	75.0%	81.5%	**51.4%**
Extended semantic decoder accuracy	75.0%	75.2%	76.1%	77.0%	77.0%	81.4%	81.5%	**77.6%**

The experimental results are shown in Tab. 1. We test the accuracy of the semantic decoder with different testing data and different size of testing data. The ratio of English words is the number of English words over the number of all words in the testing sentence. When the ratio is one, there is no Chinese word in the testing data. The accuracy of semantic decoder is the ratio of the number of words annotated correctly over the total number of words in the testing set.

The average accuracy of the extended semantic decoder is 26% higher than that of the baseline semantic decoder. It shows our approach is viable and effective.

4. CONCLUSION

A new method for spoken mixed language understanding is presented. By mixed language we mean that the words included in one sentence may come from different languages, such as Chinese and English. We assume that the understanding problem is a noisy channel transfer problem and represent the conceptual structure as a hidden Markov model. The semantic decoder is trained from English training corpora. To use the semantic decoder, the mixed language sentence is converted into English sentence with the help of a word sense disambiguation module.

Assuming the primary language is English, our experiments on the task of English/Chinese mixed language understanding show that our new approach is effective. The advantage is that we need neither Chinese training data nor mixed language training data.

Theoretically, our method is not only useful for mixed language of Chinese and English, but also applicable to other pair of language, e.g. English and French, even several languages mixed together.

References

[1] Pascale Fung, Cheung Chi Shun, Lam Kwok Leung, Liu Wai Kat, and Lo Yuen Yee. SLASA Version 1.0: A Speech-based Web Browser for Hong Kong English. 5th International Conference on Spoken language Processing. Sydney. pp.1615-1617

[2] W. Ward, Understanding Spontaneous Speech: The PHOENIX System, ICASSP'91, Vol. 1, pp.365-367.

[3]John Dowding, Jean Mark Gawron, Doug Appelt, et al. GEMINI: A Natural Language System For Spoken-language Understanding. 31st ACL, Columbus, Ohio, June 1993, pp. 54-61

[4] Robberto Pieraccini, Esther Levin. A learning Approach to Natural Language Understanding. NATO-ASI, New Advances & Trends in Speech Recognition & Coding, Spinger-Verlag, Bulion, Spain,1993

[5] M. Epstein, K. Papineni, S. Roukos, T. Ward and S. Della Pietra. Statistical Natural Language Understanding Using Hidden Clumping. IC 96.

[6] Ti. Ward, S. Roukos, C. Neti, J. Gros, M. Epstein, S. Dharanipragada. Towards Speech Understanding Across Multiple Languages. 5th International Conference on Spoken Language Processing. pp. 2243-2246

[7]Liu Xiaohu, Li Sheng. Statistic-based Target Word Selection in English-Chinese Machine Translation. Journal of Harbin Institute of Technology, May, 1997.

[8] Hemphill, C. T., Godfrey, J.J., Doddington, G. R., The ATIS Spoken Language Systems, pilot Corpus," Proc. of 3rd DARPA Workshop on Speech and Natural Language, pp. 102-108, Hidden Valley(PA), June 1990.