ISCA Archive
http://www.isca-speech.org/archive

7th International Conference on Spoken Language Processing [ICSLP2002]
Denver, Colorado, USA
September 16-20, 2002

AUTOMATIC SIGN TRANSLATION

Ying Zhang, Bing Zhao, Jie Yang, Alex Waibel

Interactive Systems Laboratories, School of Computer Science
Carnegie Mellon University
{joy+, bzhao+, yang+,ahw+}@cs.cmu.edu

10.21437/ICSLP.2002-213

ABSTRACT

Large amounts of information is embedded in the natural scenes. Signs are good examples of objects in natural environments which have rich information content. In this paper, we present our efforts in the automatic sign translation. We describe the challenges in the automatic sign translation and introduce the architecture of our current system for automatic detection and translation of Chinese signs. Two data-driven machine translation methods: Example Based Machine Translation (EBMT) and Statistical Machine Translation (SMT) are compared for the task of translating Chinese signs into English. We report the experimental results of both methods that are trained from a small bilingual sign corpus combined with a bilingual glossary. The experiment results indicate that EBMT generates more correct translations while SMT is better at inferring unseen patterns. We are currently working on developing a multi-engine machine translation system that can incrementally learn from the data and combine the results from EBMT and SMT.

1. Introduction

Signs are everywhere in our lives. They make our lives easier when we are familiar with them, but sometimes they pose problems or even dangers if we do not understand their meanings correctly. For example, a tourist might get into trouble if he/she does not understand a sign that specifies military warnings or hazards in a foreign country. In this research, we are interested in signs that have direct influence upon a tourist from a different country or culture. These signs include, at least, the following categories:

- Names: street, building, company, etc.
- Information: designation, direction, safety advisory, warning, notice, etc.
- Commercial: announcement, advertisement, etc.
- Traffic: regulatory, warning and guide, etc.
- Propaganda: political slogan or religious expression, etc.
- Conventional symbols: symbols that are not used worldwide are especially confusing to a foreign tourist.

At the Interactive Systems Laboratories of Carnegie Mellon University, we are developing technologies for automatically detecting, recognizing, and translating signs (Yang, 2001; Gao, 2001). Sign translation, in conjunction with spoken language translation, can help international tourists to overcome the language barriers. It is part of our efforts in developing a tourist assistant system (Yang, 1999). The proposed systems are equipped with a unique combination of sensors and software. The hardware includes computers, GPS receivers, lapel microphones and earphones, video cameras and head-mounted displays. This combination enables a multimodal interface to take advantage of speech and gesture inputs in order to provide assistance to tourists. The software supports natural language processing, speech recognition, machine translation, handwriting recognition and multimodal fusion.

A successful sign translation system relies on three key technologies: sign detection, optical character recognition (OCR), and language translation. In this paper, we discuss the problem of automatic sign translation. The rest of this paper is organized as follows: In section 2, we describe challenges of automatic sign translation and the system for Chinese sign translation. In section 3 and 4, we compare two data-driven machine translation methods, i.e., Example Based Machine Translation (EBMT) and Statistical Machine Translation (SMT) for the task of translating Chinese signs into English. In Section 5, we report experiment results of both methods that are trained from a small bilingual sign corpus combined with a bilingual glossary. The results indicate that EBMT generates more correct translations while SMT works better at inferring unseen patterns. In Section 6, we conclude the paper.

2. Challenges and Approach

A sign is an object that suggests the presence of a fact, condition, or quality. The language used in the signs can be considered as of a different genre than other written text. This makes the sign translation different from translating the written text in many aspects. The lexical requirement of a sign translation system is different from an ordinary machine translation (MT) system, since signs are often filled with abbreviations, idioms, and named-entities, which do not usually appear in the formal languages. The physical constraints of a sign require the text in a sign to be short and concise. Yet shorter words and phrases are more likely to be ambiguous due to the insufficient context information to resolve the ambiguities. This makes the lexical and structural mismatch becomes more severe in sign translation. Furthermore, sign translation is sensitive to the domain and functionality of a sign: lexicon in different domains may have different meaning. However, domain identification using only the information from the text is difficult because signs are concise and can provide few contexts. For structural matching, the system needs to handle ungrammatical language, which is a common phenomenon in signs. Moreover, the imperfect sign recognition results make the sign translation even harder. Though in many cases human being can correctly “guess” the actual meaning of a sign containing noises with the help of the context knowledge, for MT systems, this is still a difficult problem. In summary, with the challenges mentioned above, sign translation is not a trivial problem that can be readily solved using the existing MT technology.

In the existing MT techniques, an Knowledge Based MT system (KBMT) works well with grammatical sentences, but it requires a great amount of human efforts to construct the knowledge base, and it is difficult for such a system to handle the ungrammatical texts which appear frequently in signs. On the other hand, Statistical MT and Example Based Machine Translation (EBMT) enhanced with domain detection are more appropriate for sign translation.

We are currently developing an automatic Chinese sign translation system. The system utilizes a video camera to capture the image with signs, detects signs in the image, recognizes signs, and translates the sign recognition results into English. We choose Chinese for several reasons. Firstly, Chinese is the most spoken languages in the world and it is very different from European languages. Secondly, a foreign tourist might have serious language barrier in China because English is not commonly used there. Thirdly, statistics have shown that more people will visit China in the near future. Finally, technologies developed for Chinese sign translation can be extended to other languages. Figure 1 shows the system architecture (Yang 2001) and Figure 2 illustrates the current user interface. For those readers who are interested in automatic sign detection and recognition please refer to (Gao 2001, Zhang 2002). This paper focuses on sign translation only.

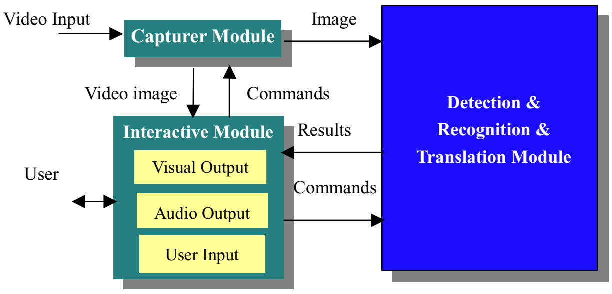
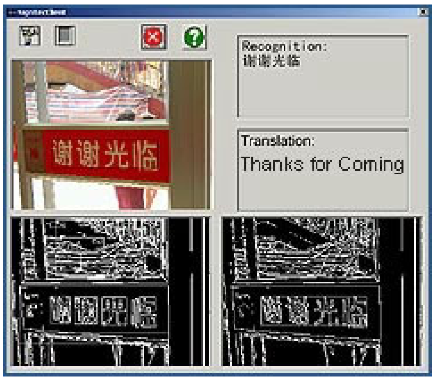


Figure1: System Architecture


Recognition:
谢谢光临

Translation:
Thanks for Coming

Figure 2. System Interface

We have investigated the Data-Driven Machine Translation (DDMT) technologies, including Example-Based Machine Translation (EBMT) and Statistical Machine Translation (SMT) for sign translation. Trained on a bilingual sign corpus, neither EBMT nor SMT needs any human written rules, which makes it very easy to adapt the system to other language pairs rapidly. In this paper, we compared the translation results using these two methods and proposed augmenting the translation with semantics and other domain knowledge to interpret the signs. We will discuss these two methods in more details in the following two sections.

3. DATA-DRIVEN MACHINE TRANSLATION

In recent years, corpora of multilingual translated texts have become widely available for a number of languages. Notwithstanding the seminal paper by Nagao (1984), it is primarily since the early 90's that such bilingual texts have been exploited in the area of Machine Translation (MT). The two main paradigmatic approaches that have been proposed are Statistics-based Machine Translation (SMT) and the Example-Based Machine Translation (EBMT).

We transcribed about 670 Chinese signs with their translations from (Kubler 93). We also had about 1300 signs from photos taken in China translated by the native Chinese speakers. This is a very small training data set for any data-driven systems. Although we have some large bilingual Chinese-English broadcasting news corpora, experiments have shown that adding corpus of different genre cannot help too much.

3.1. Example Based Machine Translation

The EBMT system (Brown 1996, Brown 1999) we used is a shallow matching system that can function using nothing more than sentence-aligned plain text and a bilingual dictionary. Given sufficient parallel text, the dictionary can be extracted statistically from the corpus (Brown 1997). In the translation process, the system looks up all matching phrases in the source language and performs a word-level alignment on the entries containing matches to determine a (usually partial) translation. Portions of the input for which there are no matches in the corpus do not generate a translation.

Because the EBMT system does not generate translations for 100% of its input text, a bilingual dictionary and a phrasal glossary are used to fill any gaps. Selection of a "best"' translation is guided by a trigram model of the target language and a chart table.

The segmentation of Chinese words from character sequences is important for translation of Chinese signs. This is because the meaning of a Chinese sentence is based on words, but unlike English, there are no explicit boundaries around Chinese words. A module for Chinese word segmentation is included in the system. This segmenter uses a word-frequency list to make segmentation decisions. The word-frequency list can be augmented with new lexicons discovered by an automatic tokenization process (Zhang 2001).

3.2. Statistical Machine Translation

The statistical machine translation systems, such as the model described in (Brown 1993) are usually based on the assumption of a noise channel between the source and the target language. The translation of a sentence f is chosen by the maximum conditional probability:

$$\begin{aligned} trans(f) &= \arg\max_e p(e \mid f) \\ &= \arg\max_e \frac{p(f \mid e)p(e)}{p(f)} \\ &= \arg\max_e p(f \mid e)p(e) \end{aligned}$$

The statistical approach tries to infer several statistical models, including the language models and the translation models from the bilingual corpus. Given a source sentence *f*, SMT will use these statistical models to generate (decoding) a sentence in the target language *e*, in our case, generate the English translation for the Chinese signs.

For sign translation in general, the sentences are very short. The length ranges from 1 to 10 Chinese words per sentence. Also, we have only a very small bilingual training corpus (about 2000 parallel sign sentences in this case). To obtain better alignment results, we modified the IBM Model-1 using a bi-direction training process.

IBM Mode-1 calculates the conditional probability *t(f|e)* of translating a word *e* in the target language to a word *f* in the source language for every word in the training data. Note that if we apply the Bayesian rules to the noise channel model, the translation model calculates the translation probability from a word in the target language to a word in the source language. Given the parallel training data: $\{(f^{(i)}, e^{(i)}), i = 1..S\}$, Model-1 can be trained using the EM algorithm:

$$t(f \mid e) = \lambda_e^{-1} \sum_{s=1}^{S} c(f \mid e; f^{(s)}, e^{(s)}) \qquad (1)$$

where λ_e^{-1} is a normalization factor such that $\sum_f t(f \mid e) = 1$

$$c(f \mid e; f^{(s)}, e^{(s)}) = \frac{t(f \mid e)}{\sum_{k=1}^{l} t(f \mid e_k)} \sum_{j=1}^{m} \delta(f, f_j) \sum_{i=1}^{l} \delta(e, e_i) \qquad (2)$$

where $l = |f^s|, m = |e^s|$. For detailed information, see (Brown 1993).

From equation (2), Model-1 counts the co-occurrence of the potential translation pairs of *(f,e)* in the training corpus. We here modified this count by a bi-direction training process. First we trained the Model-3 (Brown 1993) from the source language to the target language, which generated an initial alignment of the training corpus from Chinese to English.

$$A_1 = \{(f_k^{A_1}, e_k^{A_2}), k = 1..K^{A_1}\}$$

Then we trained the system from the target language to the source, which generated an alignment from English to Chinese.

$$A_2 = \{(f_k^{A_{21}}, e_k^{A_2}), k = 1..K^{A_2}\}$$

We used the intersection of these two alignment results. The alignments in $A = A_1 \cap A_2$ are the alignments agreed by both directions of alignment (from Chinese to English and from English to Chinese).

Then we run the final training of SMT using Model-1 where we assigned a higher weight θ to those alignments in *A*. By doing this, we can obtain a reliable alignment even though the training set is very small.

The modified Model-1 Training for the third pass is listed as below:

$$c(f \mid e; f^{(s)}, e^{(s)}) = \frac{t(f \mid e)}{\sum_{k=1}^{l} t(f \mid e_k)} \lambda_{f,e} \sum_{j=1}^{m} \delta(f, f_j) \sum_{i=1}^{l} \delta(e, e_i)$$

Where $\lambda_{f,e} = \begin{cases} \theta & if \quad (f,e) \in A \\ 1.0 & \quad\quad else \end{cases}$

θ is an arbitrary number greater than 1.0

We also used a Chinese-English glossary in the training of Model-1. The glossary is provided by the Language Data Consortium (LDC). It contains about 128,366 entries of Chinese words. If an aligned word pair also occurred in the bilingual glossary, we boost the frequency of this pair by an arbitrary number similar to the way we did in the modified Model-1 training.

For example, there is a bilingual sentence pair in the training data: 慢 行 /man xing/ slowly drive / → *"Drive slowly"*, where 慢 can be aligned to either "*drive*" or "*slowly*" with equal probability using the training data only. As there is an entry in the glossary for 慢→ "slowly", we boost the weight of aligning 慢 to "slowly", thus they are more likely to be aligned than the pair 慢 with "drive".

4. EXPERIMENTAL RESULTS

4.1. Training and Testing

It takes SMT several hours to train the model on a training set with about 2000 bilingual sentences, while EBMT needs only several seconds to index the bilingual corpus. The decoding time of 70 testing sentences is about several seconds for both systems.

4.2. Translation Quality

We didn't use the popular evaluation methods such as the IBM Bleu score or the NIST score, which calculate the n-gram match of the translation result against the reference translations. One of the reasons is because signs are always short. Many signs have only 2 or 3 words, thus they do not fit in the scenario of using n-grams. Secondly, we care about how the translation results affect the user's understanding and behavior. For example, if the system translates 严禁吸烟*/yan jin xi yan/ Prohibit Smoking/* into *"Do Smoking*" instead of "*Do not Smoking*", it should be considered as "incorrect", although it has 2 out 3 unigrams correctly matched against the reference.

We evaluated the results by manually labeling the translations as "Correct Translation", "Partially Correct Translation" and "Wrong Translation". As the sentences are short, it is not expensive to do this human evaluation from time to time. Table 1 shows the results of EBMT and SMT tested on 70 signs.

We have analyzed the results and classified the cause of errors to 5 categories: errors caused by the Chinese segmentation mistakes (SEG); errors caused by lacking domain knowledge (DK); named-entities are not recognized (NE); or because the training set is not large enough (Data); and other reasons (Other). Figure 3 shows a comparison on the error sources between EBMT and SMT.

***Table 1*: Results of EBMT and SMT**

System	Correct	Partial Correct	Wrong
EBMT	31 (44%)	21 (30%)	18 (26%)
SMT	22 (31%)	28 (40%)	20 (29%)

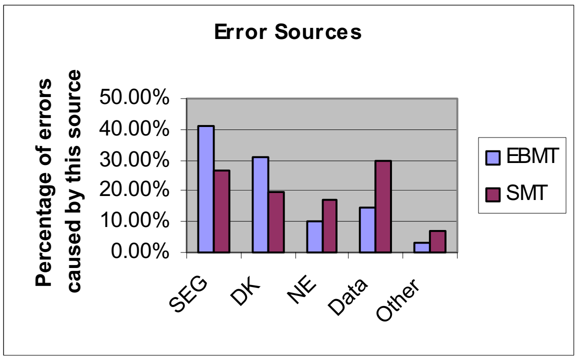


Figure 3:Error Sources.

Some discussions are in order:

1. It is obvious to notice that if a sign in the testing set occurred at least once in the training data, EBMT can translate it correctly, whereas SMT might not. On the other hand, SMT can capture more general patterns in the bilingual corpus given enough training data. The more complicated model of SMT can handle some syntactic features such as word reordering.

2. In general, both methods suffer from the small size of bilingual sign corpus, which may be relieved by introducing generalization approaches, such as word clustering and grammar or template induction. Given this small training data set, EBMT outperforms SMT on average. Table 1 shows that EBMT has more "Correct" translations than SMT because EBMT can translate a test sentence correctly as long as it appears in the training set. SMT provides more generalization than EBMT. This explains why SMT has more "Partial correct" translations than EBMT.

3. For EBMT, mistakes in Chinese segmentation causes more problem than SMT, because wrong word segmentation affects the source language match between test and training sentences. Figure 3 also shows that when training data is small, SMT suffers more than EBMT.

5. CONCLUSIONS

In this paper, we presented our research efforts in automatic sign translation. We have developed a Chinese sign translation system. The system can automatically detect signs from an image and translate them into English. We have compared two Data-Driven MT techniques, i.e., EBMT and SMT for a Chinese sign translation task. We have evaluated both method on the same task and analyzed the errors. Although experiment results are encouraging, none of the methods alone can solve the problem very well. A multi-engine approach can be a better solution.

6. ACKNOWLEDGEMENT

We would like to thank Ralf Brown, Stephan Vogel, Xilin Chen, Jing Zhang, and other members in the Interactive Systems Laboratories for their help in the research work of this paper.

7. REFERENCES

[1] Brown, R.D., Adding Linguistic Knowledge to a Lexical Example-Based Translation System. Proceedings of the Eighth International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-99), pp. 22-32, Chester, England, August, 1999.

[2] Brown, R.D., Automated Dictionary Extraction for "Knowledge-Free" Example-Based Translation. Proceedings of the Seventh International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-97), pp. 111-118, Santa Fe, New Mexico, July, 1997.

[3] Brown, P.F., Della Pietra, S.A., Della Pietra, V.J., and Mercer, R.L. The Mathematics of Statistical Machine Translation: Parameter Estimation. Computational Linguistics 19(2): 263-311. 1993

[4] Brown, R.D., Example-based machine translation in the pangloss system. Proceedings of the 16th International Conference on Computational Linguistics, pp. 169-174, 1996.

[5] Gao, J. and Yang, J., "An Adaptive Algorithm for Text Detection from Natural Scenes," Proceedings of Computer Vision and Pattern Recognition (CVPR 2001).

[6] Jain, A.K. and Yu, B., Automatic text location in images and video frames. Pattern Recognition, vol. 31, no. 12, pp. 2055-2076, 1998.

[7] Kubler, Cornelius C., "Read Chinese Signs". Published by Chheng & Tsui Company, 1993.

[8] Nagao, M. (1984), "A framework of a mechanical translation between Japanese and English by analogy principle", in Artificial and Human Intelligence: edited review papers at the International NATO Symposium on Artificial and Human Intelligence sponsored by the Special Programme Panel held in Lyon, France, October, 1981', Elsevier Science Publishers, Amsterdam, chapter 11, pp. 173-180.

[9] Yang, J., Yang, W., Denecke, M., and Waibel, A., Smart sight: a tourist assistant system. Proceedings of Third International Symposium on Wearable Computers, pp. 73-78. 1999.

[10] Yang, J., Gao, J., Yang, J., Zhang, Y., Waibel, A., Towards Automatic Sign Translation, Proceedings of Human Language Technology 2001.

[11] Zhang, Y., Brown, Ralf.D., and Frederking, Robert E., "Adapting an Example-Based Translation System to Chinese". Proceedings of Human Language Technology Conference 2001

[12] Zhang, Y., Brown, Ralf D., Frederking, Robert E. and Lavie, Alon. "Pre-processing of Bilingual Corpora for Mandarin-English EBMT". Proceedings of MT Summit VIII. 2001

[13] Zhang, J., Chen, X., Hanemann, A., Yang, J., and Waibel, A., "A Robust Approach for Recognition of Text Embedded in Natural Scenes," Proceedings of ICPR 2002.