Proceedings of 1993 International Joint Conference on Neural Networks

A Neural Network-Assisted Japanese-English Machine Translation System

Todd Law† Hidenori Itoh Hirohisa Seki

Nagoya Institute of Technology
Gokiso-cho, Showa-ku, Nagoya, Japan

Abstract: In this paper, we present a hybrid machine translation system which combines the strengths of logic programming, procedural programming, and neural networks. The system is designed for Japanese to English translation of natural language sentences found in daily newspaper weather reports. The system is trained on one full year of weather reports, or over 1000 sentences. A second full year of weather reports is used to empirically evaluate the system.

1 Introduction

We introduce a machine translation system specializing in the fully automatic translation of Japanese newspaper weather reports. Our system is a unique attempt to draw together fundamentally different artificial intelligence techniques in order to effectively translate difficult natural language sentences. These techniques include logic programming, procedural programming, and neural networks. This idea is illustrated in Figure 1.

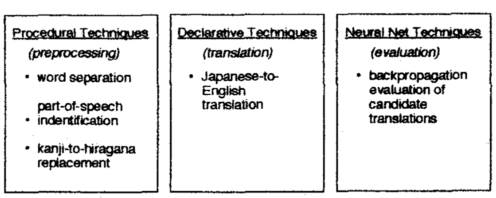

Figure 1: Applying Different Knowledge Processing Techniques to Machine Translation

The idea of applying neural network techniques to the machine translation problem has also been pursued by other researchers. For example, [1] uses feed forward neural networks to translate short sentences between English and Spanish, and [2] between English and Spanish. Our system is the first to incorporate neural networks with other knowledge processing techniques in a fully working system which can translate natural language sentences, such as found in a newspaper.

†This research is supported in part by the Japanese Ministry of Education.

2 System Overview

2.1 Integrated Techniques for Expert Systems

Our machine translation system attempts to draw together three different forms of knowledge representation in an integrated expert system. Procedural techniques are used for the initial preprocessing stage, where Japanese sentences are divided into words, tagged with a part-of-speech identifier, and expressed in a unique form. Declarative techniques are then used to perform the actual translation of sentences from Japanese to English. Neural techniques are used to evaluate and select the most appropriate translation from the many possible candidates.

2.2 System Specifications

The main characteristics of our machine translation system are as follows.

System Type Syntactic, neural evaluation

Human Languages Japanese to English

Domain newspaper weather reports

Grammar Type Relational Grammar

Machine Languages Prolog, C

Training sentences 1057 (one year of daily weather reports [3])

Translation rules 834

Total words in training sentences 14133

Average sentence length 12.3 words

Total vocabulary (Japanese) 800 words

Total vocabulary (English) 1079 words

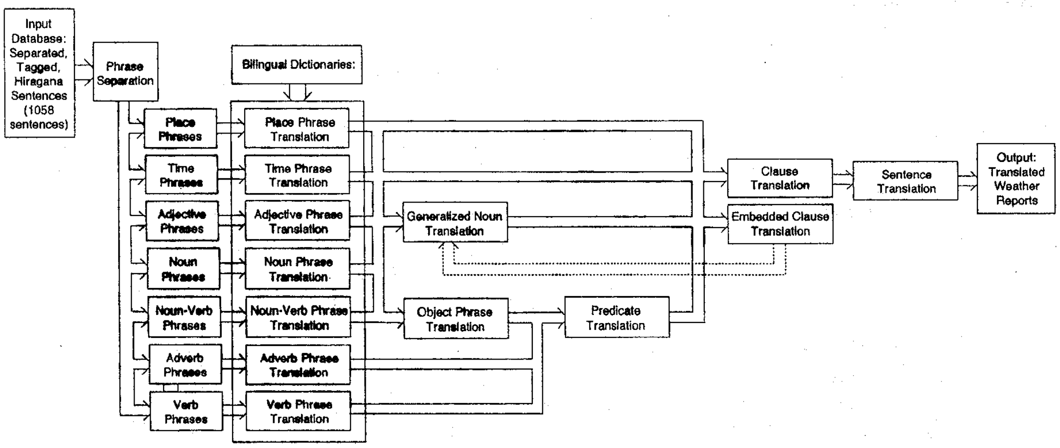

Figure 2: Translation Mechanism Block Diagram

3 Morphological Preprocessing

The nature of the Japanese language requires special techniques to be rendered in suitable form for input to a translation mechanism. First, unlike English, words in Japanese are not separated by spaces. Breaking sentences into their constituent words can reduce the complexity of the translation process considerably. Second, the Japanese language uses four different character sets, all of which may appear in a single sentence. Converting these character sets to a common representation greatly simplifies dictionaries and meaning representation. Third, there is a problem of multiple pronunciations of the same characters, and parallel representations, meaning that the language does not always appear sequentially (because of multiple *furigana*). Clearly, it is desirable to find a unique representation for words, and to have sentences appear as a consecutive sequence of characters.

To resolve these issues, we developed a procedural algorithm and program which accomplished the following tasks:

1. word separation
2. kanji-to-hiragana replacement
3. part-of-speech identification

It was simple to perform part-of-speech identification at this stage, and reduces the complexity of the translation mechanism.

3.1 Algorithm

The preprocessing algorithm inputs sentences in the following natural format,

北部山間部は、終日雪が降り、
かなりの積雪となる恐れがある。

The algorithm consults a set of simple dictionaries containing multiple representations, pronunciations, part-of-speech identifiers, then outputs the sentences in *hiragana* form, with words separated and tagged with their part-of-speech markers, as follows,

[[pl, ほくぶさんかんぶ],[pe, は],[pu,、],
[t, しゅうじつ],[n, ゆき],[ne, が],[v, ふり],[pu,、],
[av, かなり],[p, の],[n, せきせつ],[ne, と],[v, なる],
[ve, おそれ],[ve, が],[v, ある],[pu,。]].

The symbols used in this representation demark part-of-speech ($n = noun$) and some simple semantic information ($t = time$). Full details of the preprocessing algorithm, the dictionaries and problems particular to the Japanese language can be found in [5].

4 Translating Mechanisms

Full details of the actual translating mechanism are outside the scope of this paper, but can be read in [5]. Briefly, however, the translation system consists of a set of bilingual Prolog dictionaries and a set of Prolog translation programs. The basic structure for the translation module is outlined in Figure 2. Translation is done hierarchically, with the top level for sentences, the second level for clauses, and the third level for various types of phrases.

4.1 Generating Multiple Translations

The important aspect of the translation mechanism for this paper is that for a single input Japanese sentence, multiple English translations can be produced by the expert system. An example of why this is necessary is explained as follows. Many words in Japanese can be multiply translated into English, for example, the expression *yoru ni* (夜に) translates to 'at night' while the expression *asa ni* (朝に) translates to 'in the morning'.

$$\left.\begin{array}{ll}\text{夜} & \text{に}\\ \textit{yoru} & \textit{ni}\\ \text{night} & \text{at}\end{array}\right\} \longrightarrow \text{at night}$$

$$\left.\begin{array}{ll}\text{朝方} & \text{に}\\ \textit{asagata} & \textit{ni}\\ \text{the morning} & \text{in}\end{array}\right\} \longrightarrow \text{in the morning}$$

In our system, we refer to such ambiguities in the translation as **choice points**. We represent the choices with a '/' in the output sentences.

5 Neural Evaluation

With the candidate parses produced, we then applied backpropagation neural networks to evaluate the translations.

5.1 Evaluating Choice Points

The output translation results contain various *choice points* which can resolved using a neural network. The choice points are marked with slashes in the translation output.

In our machine translation system, choice points occurred in four possible places: after the particles の (*no*) and には (*niwa*) following a time-expression, and after the particles は (*wa*) and に (*ni*) following a place expression. These particles may be translated variously as shown in Table 1.

Particle	Translation	Occurences
の (no)	for, on, _	180
には (niwa)	for, in, on, at, _	95
は (wa)	for, in, _	40
に (ni)	in, to	43

Table 1: Choice point particles and their translations.

The character '_' indicates a nil translation. In total, 358 choice points occurred in the original 1058 sentences.

The particles can be translated on the basis of the immediately surrounding words. Inputs into the network logically fall into three groups. The left side of the input layer inputs the word immediately to the left of the choice point. Similarily, the right side of the input layer inputs the word immediately to the right of the input layer. We then made a reduced dictionary of words that actually occur on either side of choice points, which was 138 words in total. If the word 'the' occurred after to a choice point, the following word was chosen instead. Punctuation marks and the sentence start marker were treated as individual words.

Since there are four possible particles (の, には, は, and に), there are four center inputs.

In translation, there are six possible outputs (at, for, in, on, to, and the null translation), so the output layer contains six outputs.

5.2 The Network

Using these specifications, we built a three-layer backpropagation neural network (one input layer, one hidden layer, one output layer).

We used no momentum, and α (the learning rate) was set to 1.0. Since there are four types of choice points, and six possible outcomes, we initially used a hidden layer containing 24 neurons.

We then trained our network on our 358 choice points. Input vectors were presented by setting the input of the corresponding neurons of the three input words (the choice point, and the words on either side of the choice point) to one. All other inputs were set to zero. The target vectors were all zeroes except for the output representing the correct word, which was set to one.

The network-produced output vector was taken to be the output whose values was closest to 1.0. Training was done by *sample*, meaning that after each input-output vector pair was presented, weights were updated immediately instead of at the end of each cycle.

5.3 Results

After running our network several times, we found that training was equally effective with ten hidden-layer neurons instead of twenty-four. It is also faster to train a network with fewer hidden-layer neurons. We also tried various values for α and found that values between 0.5 and 1.5 gave acceptable results in reasonable lengths of time. After 50 cycles of training, we could successfully recall 354 of the original 358 pattern pairs, or 98.9%.

6 Testing

To test our machine translation system, we used a second year's worth of newspaper weather reports [4], taken from the Yomiuri Newspaper for the whole of 1992 (training data was from all of 1991). The basic characteristics of the test data are as follows.

Number of sentences 993

Number of words 14082

Average Sentence Length 14.2 words

6.1 Preprocessing

We then fed the raw data through our preprocessing program, and evaluated the results. Of the original 993 sentences, 458 sentences (46%) passed successfully through the preprocessing program. All failures can be attributed to new words used by the test sentences that were not in the dictionaries.

6.2 Translation

We then took those sentences which successfully passed the preprocessing stage, and input them

Original Japanese	English Translation
冷たい季節風も収まり、日中は暖かくなる。	Cold seasonal winds will die down, and for the daytime it will be warm.
再び冬型の気圧配置になるが、かんきの南下は遅れそう。	Again there will be winter pressure pattern, but cold air southward movement likely will be delayed.
来週は晴天に恵まれるが、寒気が入ってくるため、寒くなる。	For next week fine weather be enjoyable cold air will enter in, and consequently it will be cold.

Table 2: A Sampling of Some Translation Results

to our machine translation system. Of the 458 input sentences, translations were successfully produced for 318 sentences (69.4%).

6.3 Neural Evaluation

These 318 successfully translated sentences contained 136 choice points. Thirteen of these test points were neighboring words which our network had never seen before, but these words were added to the neural network dictionary, and given zero training weights. This means that the number of inputs on the left and right side of our neural network increased from 152 to 164 (one of the thirteen cases was a duplicate). The 136 choice points were then input to the trained neural evaluation system. 122 (89.7%) of the choice points were correctly evaluated. Of those 133 choice points which contained only previously-trained words, 122, or 91.7% of the choice points were correctly evaluated. Of those thirteen choice points (including the duplicate) containing never-seen-before surrounding words, nine were translated correctly (70%).

6.4 Results

The entire test body included over 1000 sentences, so it would be impossible to reproduce them all here. However, a few sentences translated by our system are shown in Table 2. These are, of course, some of our best results. The worst results are when no translation output at all was produced because either (a) the pre-processor failed to recognize words in the input sentence, or (b) the translator either (i) had inappropriate grammatical structures to deal with the input sentences, or (ii) words in the bilingual dictionaries were not represented in the form used by the test sentences. To explain point (ii) further, the words 動き (*ugoki*) may occur either as a noun ('movement') or as a verb-stem ('to move'). Because our system could only recognize it as a verb, it failed when this word was suddenly used as a noun.

As for the neural portion of our system, we find it encouraging that never-seen-before combinations of surrounding words can be successfully translated in close to 90% of the observed cases. Although the sample was small, combinations containing never-seen-before words were also translated successfully in a majority of cases.

7 Discussion

Weather reports may seem a very mundane form of natural language, but contain a surprising amount of variety. This variety is statistically illustrated by the graph in Figure 3, which shows that the rate of new word introduction, even after one year of weather reports, can be as high as 9 new words per 100 words used. On average, this value is approximately 2.3 new words per 100 words used. Even at this rate, with average sentence length at 12.3 words, there is only a $(1 - 0.023)^{12.4} = 75\%$ chance that the average sentence will contain only previously seen words.

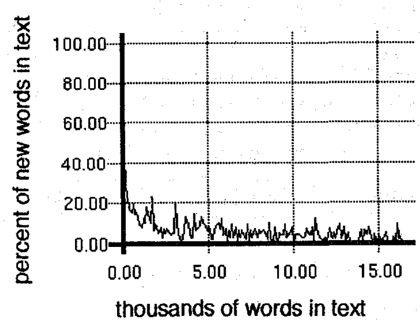

Figure 3: Percentage new words in weather reports

8 Conclusions

We have shown that integrating diverse knowledge techniques can be an effective strategy for the difficult problem of machine translation. These methods include procedural, declarative, and neural network techniques.

References

[1] Several Studies on Natural Language and Backpropagation, R. Allen, Proc. Int. Conf. on Neural Networks, 1987.

[2] Descending Epsilon in Backpropagation: A Technique for Better Generalization, Y. Yu et al., Proc. Int. Jt. Conf. Neur. Net., 1990.

[3] Yomiuri Shinbum, Jan. 1 to Dec. 31, 1991.

[4] Yomiuri Shinbum, Jan. 1 to Dec. 31, 1992.

[5] A Neural Network-Assisted Japanese-English Machine Translation System, Master's Thesis, T. Law, H. Itoh, Nagoya Institute of Technology, March 1993.