Technology Reports | Natural Language Processing | Machine Translation | Document Classification
Special Articles on AI Supporting a Prosperous and Diverse Society

Automatic Domain Prediction in Machine Translation

Service Innovation Department Soichiro Murakami Atsuki Sawayama
Toshimitsu Nakamura Hosei Matsuoka Wataru Uchida

Machine translation can be applied to a variety of domains such as restaurants, lodging facilities, and transport agencies each of which differs in terms of conversation, vocabulary, phrasing, and their translation. It is therefore common to create a machine translation engine specialized for each domain using a corpus specific to that domain to improve translation performance. However, when faced with a translation task targeting multiple domains, the user must select multiple engines, which detracts from the convenience of machine translation. In response to this problem, NTT DOCOMO has developed technology for automatically predicting the domain of the machine translation engine from the text input by the user. This makes it possible to automatically select the optimal machine translation engine for the input text.

1. Introduction

In recent years, the number of foreign travelers visiting Japan has been increasing dramatically resulting in a sudden increase in "inbound demand." Against this background, voice translation services using speech recognition technology and machine translation technology are coming to be introduced for achieving smooth communication with foreign travelers in restaurants and other eating/drinking establishments, at lodging facilities, on public transportation, etc. Voice translation services

are also being introduced at medical institutions that will likely be used by ill or injured travelers to make the purpose of an examination or treatment understandable to the patient. In this way, voice translation services are coming to be used across a wide range of scenarios.

Amid this trend, Neural Machine Translation (NMT)*1 is attracting attention in the field of machine translation [1] [2]. "NMT" refers to the use of a bilingual corpus*2 to train a large-scale Neural Network (NN)*3, a scheme that has come to achieve more fluent and accurate translations than conventional statistical machine translation [3].

In NMT, using a large and high-quality bilingual corpus specific to a certain domain*4 can improve translation performance in that domain. It is therefore common to prepare a machine translation engine*5 specialized for each domain in voice translation services based on NMT. However, the sudden increase in inbound demand is being accompanied by an increase in domains that will likely require voice translation services. Furthermore, while a machine translation engine specific to each domain is needed, having to select which machine translation engine to use for each domain is troublesome for the user.

In response to these problems, NTT DOCOMO developed automatic domain prediction technology for automatically identifying the domain of input text. This technology predicts the domain of text input by the user by voice or keyboard so that a machine translation engine specific to that domain can be automatically selected for translation.

This article describes this domain prediction technology for automatically predicting usage scenarios in voice translation services.

2. Issues in Voice Translation Services

Voice translation services specific to overseas trips and customer service for foreign travelers include VoiceTra®*6, a voice translation app from the National Institute of Information and Communications Technology (NICT), ili®*7, an offline translation device for customer service from Logbar Inc., and POCKETALK®*8, a translation device from Sourcenext Corporation. NTT DOCOMO for its part provides "Jspeak" translation app for smartphones to facilitate face-to-face communication when making an overseas trip or when interacting with foreign travelers within Japan.

These examples show how voice translation services have been developed in diverse ways and how machine translation has come to be used in a wide range of domains. However, the content needing translation, the words and phrases used, and their translation depend on the domain such as restaurants, lodging facilities, public transportation, etc. For this reason, machine translation engines specialized for individual domains have been introduced to improve translation performance. Yet, for the user using a voice translation service, having to select a dedicated machine translation engine for each usage scenario takes time and effort. It is therefore considered that this troublesome task could be avoided if it were possible to predict the domain from the text input by the user and automatically select the optimal machine translation engine.

3. Automatic Domain Prediction

The automatic domain prediction technology that

*1 NMT: Machine translation technology using NNs (see *3), a machine-learning technique.

*2 Corpus: Language resource consisting of a large volume of text, utterances, etc. collected and stored in a database.

*3 NN: An entity that numerically models nerve cells within the human brain (neurons) and the connections between them. It is composed of an input layer, an output layer and hidden layers and is able to approximate complex functions by varying the number of neurons and layers and the strength of connections between layers.

*4 Domain: A usage scenario in machine translation.

we have developed extracts features from the text input by the user and performs document classification by machine learning*9 to predict the domain appropriate to the input text.

An overview of the system is shown in **Figure 1**. This figure shows the flow of classifying the text input by the user into one of several predetermined domains using a document classifier*10 and sending a translation request to the translation engine specialized for that domain. Here, "document classifier" refers to a device that classifies text into one of several predetermined classifications.

3.1 Automatic Domain Prediction as Document Classification

This technology performs document classification by predicting the domain of the input text. "Document classification" means the classification of text input to the voice translation service into one of several predefined labels. Here, "label" refers to a domain such as restaurants, lodging, or transportation. In the field of Natural Language Processing (NLP), it is common to construct a document classifier by training a machine-learning model using training data consisting of pairs of documents and labels.

An example of classification using a document classifier is shown in **Figure 2**. In this example, the document classifier extracts features from the text "Return visits to the clinic are received at counter 5" input into the voice translation service and predicts "medical care" from among the predefined domain labels using a machine-learning technique.

3.2 Training Data for Document Classifier

The training of a document classifier that uses a machine-learning technique requires the use of training data that pairs up input text of a voice translation service and domain labels.

In machine-learning techniques, model performance generally improves as the amount of training data increases. Furthermore, when constructing training data, care must be taken to prevent an imbalance in which data pairs in one domain

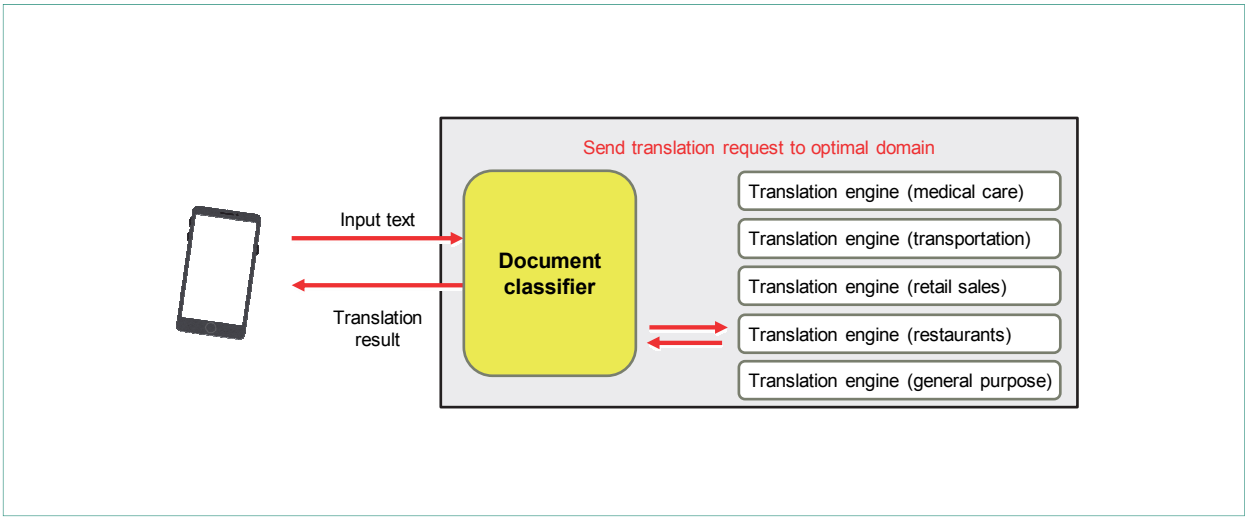

Figure 1 System overview

*5 Machine translation engine: Software for performing machine translation.
*6 VoiceTra®: A registered trademark of the National Institute of Information and Communications Technology (NICT).
*7 ili®: A registered trademark of Logbar Inc.
*8 POCKETALK®: A registered trademark of Sourcenext Corporation.
*9 Machine learning: Technology that enables computers to acquire knowledge, decision criteria, behavior, etc. from data, in ways similar to how humans acquire these things from perception and experience.

are many or few in number compared with that of another domain. This is to avoid the problem of over-fitting in which the accuracy of classification drops for text in a domain with a small amount of data.

Examples of document-classifier training data are listed in **Table 1**. Among these examples, the label "medical care" is attached to the text "When feeling dizzy, do you sweat or shiver with cold?" reflecting its domain.

3.3 Machine-learning Technique of Document Classifier

We here describe our system's document classifier that uses Long-Short Term Memory (LSTM) [4], which is a type of Recurrent NN (RNN) that introduces recurrent connections*11 in a NN. LSTM is widely used in the field of NMT that handles variable-length text. An overview of the feedforward NN*12 and RNN is shown in **Figure 3**.

In the hidden layer of an RNN such as LSTM,

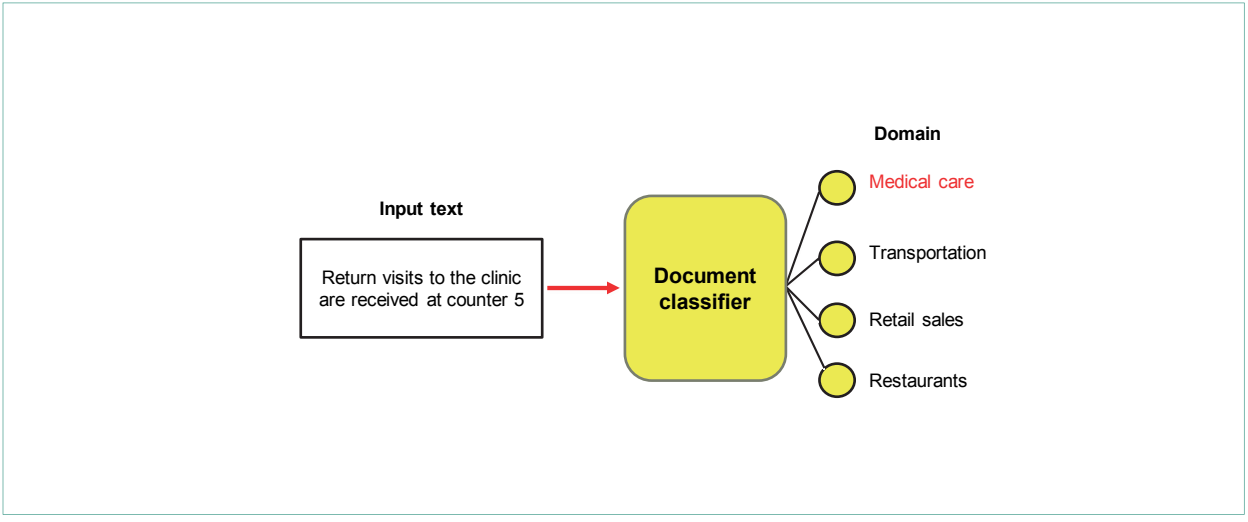

Figure 2 Overview of document classifier

Table 1 Examples of training data for a document classifier

Text	Domain
Can I see a doctor?	Medical care
When feeling dizzy, do you sweat or shiver with cold?	Medical care
This smart card cannot be charged here, so please do it beforehand.	Transportation
Arrival time may differ from the timetable depending on the weather or road conditions.	Transportation
Please bring your receipt to return or exchange any items.	Retail sales
Where is the toothpaste?	Retail sales
All items on the menu except for Japanese sake and shochu are all-you-can-drink.	Restaurants
All juices are 100% with no sugar added.	Restaurants

*10 Classifier: A device that classifies input into one of several predetermined classifications based on its feature values.

*11 Recurrent connections: Connections that are made in a recurrent manner.

*12 Feedforward NN: A NN that propagates signals only in a single direction in the order of input layer, hidden layers, and output layer without any recurrent connections in the network.

the inner state vector at the immediately previous time point t–1 can be taken over at the next time point t enabling flexible handling of variable-length input such as text. Furthermore, since text can be input in a time-series manner, the context information of that text can be expressed as a fixed-length vector called a context vector. In short, the use of an RNN enables the extraction of feature values[*13] that represent context from the input text.

An example of a decision made by a document classifier using LSTM is shown in **Figure 4**. In this example, the number of dimensions of the LSTM vector is 200. A document classifier using LSTM determines which domain the input text conforms to most based on the fixed-length context vector created from the input text using a NN. First, the

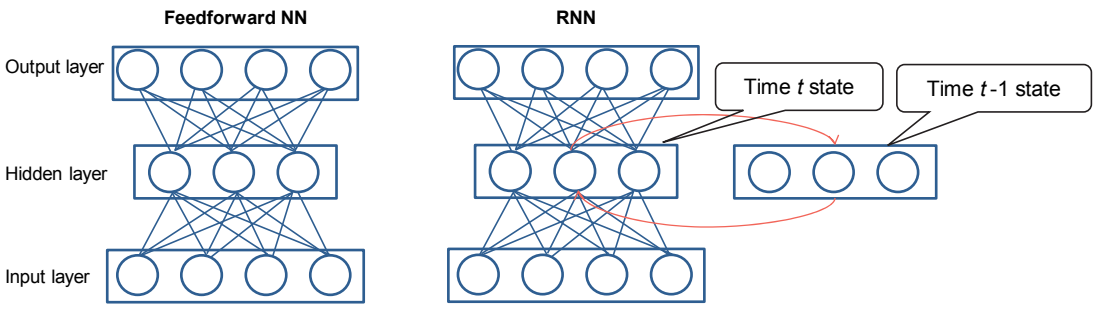

Figure 3 Feedforward NN and RNN

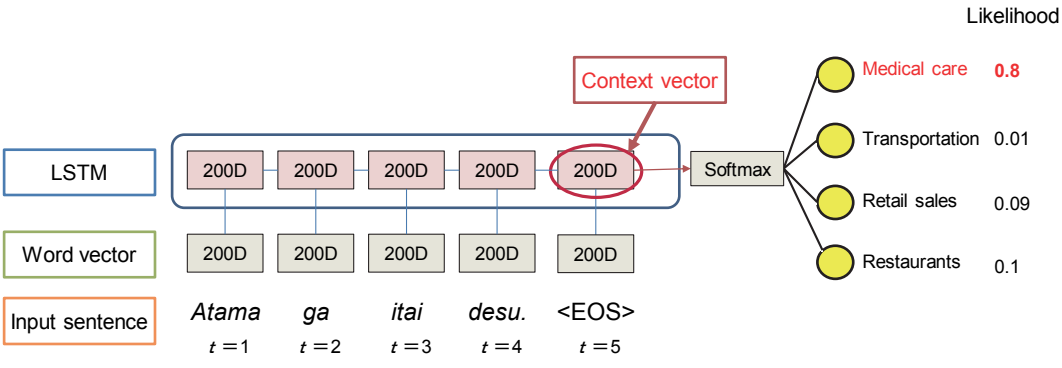

Figure 4 Document classifier using LSTM

*13 Feature values: Values extracted from data, and given to that data to give it features.

document classifier applies morphological analysis[*14] to the input Japanese sentence "頭が痛いです." ("*Atama ga itai desu.*" or "My head hurts.") to get the word-partitioned input word string "頭が 痛い です <EOS>" (*atama-ga itai desu* <EOS>). Here, "<EOS>" is a pseudo token[*15] that expresses the end of the sentence. Next, the classifier inputs the 200-dimension word vectors obtained by vectorizing each word of the input word string into the LSTM one-by-one and calculates the context vector expressing the context information of the input sentence. Finally, it uses the Softmax function[*16] based on this context vector to calculate the likelihood that the input sentence conforms to any one domain and uses those likelihood values to predict the domain most suitable for that input sentence.

3.4 Accuracy of Domain Prediction

We performed training of an LSTM-based document classifier using paired data consisting of text and labels and measured the accuracy of classification with respect to text data. In the experiment, we defined medical care, transportation, retail sales, and restaurants as the target domains and prepared 1,000 sentences of text data for each domain.

Domain prediction accuracy is summarized in Table 2. Examining the classification accuracy (*F* value[*17]) of each domain, it can be seen that the accuracy of this document classifier is generally high. In addition, the average processing time of domain prediction per sentence was approximately 12 ms, which indicates that domain prediction could be performed within a realistic processing time in actual use.

3.5 Application Example

Next, we describe an example of applying this system to machine translation using domain prediction technology (Fig. 1). The input text (in Japanese) was "このレストランではカリフォルニア産の高級ワインが召し上がれます." ("*Kono resutoran de wa kariforunia san no kokyu wain ga meshiagaremasu.*"). Using the document classifier, the system performed automatic domain prediction of this text and predicted the domain to be "restaurants." The system then translated the input text using the machine-translation engine for the restaurants domain resulting in the following translation:

"You can enjoy California high-quality wine at this restaurant."

However, on translating the input text using a

Table 2 Domain prediction accuracy

Domain	LSTM			No. of examples
	Precision	Recall	*F* value	
Medical care	0.99	0.92	0.95	1,000
Transportation	0.95	0.95	0.95	1,000
Retail sales	0.87	0.94	0.91	1,000
Restaurants	0.92	0.92	0.92	1,000

*Average processing time: 12 ms per sentence

*14 Morphological analysis: The task of dividing text written in natural language into morphemes—the smallest units of meaning in a language—and determining the part of speech of each.

*15 Token: A character or character string treated as the smallest unit of text.

*16 Softmax function: A function used to calculate probability values when normalizing the total output of a NN to 1.0.

*17 *F* value: A scale used for comprehensive evaluation of accuracy and exhaustiveness, and it is calculated as the harmonic mean of precision and recall.

general-purpose machine-translation engine, the following result was obtained:

"This restaurant has a high quality wine in California."

On comparing these translation results using a machine-translation engine dedicated to the restaurants domain and a general-purpose machine-translation engine, it can be seen that the dedicated translation engine can translate in a more fluent manner using phrases typical of that domain.

In this way, the use of automatic domain prediction technology enables higher quality translation by translating with a machine-translation engine optimal to the domain of the input text.

4. Conclusion

Given a voice translation service used in a variety of domains, this article described technology for automatically selecting the optimal machine translation engine using automatic domain prediction so that translation can be performed with an engine matching the user's domain.

Future plans include the development of domain prediction technology with even higher levels of accuracy and the development of domain prediction technology using information other than text.

REFERENCES

[1] I. Sutskever, O. Vinyals and Q. V. Le: "Sequence to Sequence Learning with Neural Networks," Advances in neural information processing systems, pp.3104–3112, 2014.

[2] D. Bahdanau, K. Cho and Y. Bengio: "Neural Machine Translation by Jointly Learning to Align and Translate," In Proc. of the 3rd International Conference on Learning Representations, 2014.

[3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N Gomez, L. Kaiser and I. Polosukhin: "Attention is All You Need," In Proc. of Advances in Neural Information Processing Systems 30, pp.5998–6008, 2017.

[4] S. Hochreiter and J. Schmidhuber: "Long Short-Term Memory," Neural computation, Vol.9, No.8, pp.1735–1780, 1997.