Creation of a Japanese Adverb Dictionary that Includes Information on the Speaker's Communicative Intention Using Machine Learning

Toshiyuki Kanamaru[*][†]**, Masaki Murata**[*] **and Hitoshi Isahara**[*]

[*]National Institute of Information and Communications Technology (NICT)
3-5 Hikaridai, Seikacho, Sorakugun
Kyoto, 619-0289, Japan
{kanamaru, murata, isahara}@nict.go.jp
[†] Kyoto University
Yoshida Nihonmatsu-cho, Sakyo-ku
Kyoto, 606-8501, Japan
kanamaru@hi.h.kyoto-u.ac.jp

Abstract

Japanese adverbs are classified as either declarative or normal; the former declare the communicative intention of the speaker, while the latter convey a manner of action, a quantity, or a degree by which the adverb modifies the verb or adjective that it accompanies. We have automatically classified adverbs as either declarative or not declarative using a machine-learning method such as the maximum entropy method. We defined adverbs having positive or negative connotations as the positive data. We classified adverbs in the EDR dictionary and IPADIC used by Chasen using this result and built an adverb dictionary that contains descriptions of the communicative intentions of the speaker.

1. Introduction

Japanese adverbs are classified as either declarative (those that declare the communicative intention of the speaker) or normal (those that convey a manner of action, a quantity, or a degree by which the adverb modifies the verb or adjective that it accompanies). A sentence adverb is an example of a declarative adverb, and onomatopoeia is an example of a normal adverb. The EDR (EDR, 1993), which consists of an electronic dictionary and a corpus, uses the same classification system for adverbs.

Normally, only declarative adverbs are thought to concern the speaker's affection or judgment about a situation. However, several normal adverbs can also be considered to express the speaker's affection or judgment. For example, the Japanese adverb *ichi-ichi* (all the time) is regarded as a normal adverb; however, *ichi-ichi* in the phrase *ichi-ichi chuui suru* (to give orders to someone all the time) indicates that the speaker does not like being given orders. We call such affection, judgment, and so on the speaker's communicative intention, meaning that the speaker intends to express a particular affection or judgment by using the adverb.

We have created a Japanese adverb dictionary that includes information about the speaker's communicative intention by examining the intentions expressed through adverb use. To do this, we used machine learning, such as the maximum entropy method, instead of manually tagging all adverbs in the dictionary. We started by tagging some adverbs and then used machine learning to classify adverbs regarding whether they were used with communicative intention. Our goal was to semi-automatically add tags indicating the speaker's communicative intention to adverbs in the dictionary by using machine learning. In an existing corpus or dictionary which has been tagged manually, there is erroneous tag data due to the use of subjective judgment and the classification is not consistent. Our goal was to devise a way to correct these errors by applying machine learning to ensure the objectivity of the classification result.

We believe such a dictionary will enable the use of adverbs in opinion extraction, the classification of intentions, and so on; in such tasks, only adjectives, nouns, and verbs have been used in the past.

In this paper, we first describe the relationship between a speaker's communicative intention and the adverb used. After that, we provide an outline of the adverb dictionary we have created and explain how we applied dictionary correction to it using machine learning.

2. Adverbs and the Speaker's Communicative Intention

The classification of adverbs has been vague for some time, because the forms that adverbs can take are special. Although it needs further refinement, here we introduce Yamada's classification of adverbs which is widely used in research analyzing the relationship between speaker subjectivity and adverbs (Yamada, 1936).

Yamada classified adverbs into three groups: declarative adverbs, degree adverbs, and state adverbs. The characteristics of each group are summarized as follows.

- declarative adverb
    - An adverb which expresses the mental attitude of speakers, such as negative, supposition, or assumption.

- degree adverb
    - An adverb which modifies an adjective.
    - An adverb which modifies another adverb, especially a state adverb or adnominal noun.
    - An adverb which modifies a time or space noun.

- state adverb

– An adverb which means the manner of motion, aspect, modality, attitude, or state.

Much research has been done on declarative adverbs which concern the speaker's mental attitude and subjectivity; for example, the work of Watanabe (Watanabe, 1983) and Nakau (Nakau, 1994). In contrast, research on state adverbs concerning the proposition of a sentence has only recently been done, commencing with the work of Nitta (Nitta, 2002). Most research has analyzed the relationship between declarative adverbs and the proposition of a sentence.

Although most research analyzing the relationship between the speaker's mental attitude and the adverb has focused on declarative adverbs, some degree adverbs or state adverbs also reflect the speaker's communicative intention. The example given in the Introduction of how the frequency adverb *ichiichi* (all the time) can be used illustrates this.

A frequency adverb is usually used to express that the same situation is repeated. However, the use of *ichiichi* in the example below indicates not only that the situation occurs repeatedly, but that the speaker also intends to express dissatisfaction with the situation.

1. *Ichiichi "home" wo kurikku shinakereba naranai.* (I must click "home" all the time.)

2. *"home" wo kurikku shinakereba naranai.* (I must click "home".)

As described above, when a speaker tries to express a subjective judgment or feelings about the situation by using a certain word, we regard such a word as expressing the "speaker's communicative intention". Several normal adverbs can be considered to express the speaker's affection or judgment about the situation.

We therefore decided to create a Japanese adverb dictionary that includes information about the speaker's communicative intention. The target adverbs include degree adverbs and state adverbs as well as declarative adverbs.

3. Construction of the adverb dictionary

3.1. About the adverb dictionary

We wanted to create a dictionary which indicated the relation between the speaker's understanding of a situation and the adverb used regarding this. Therefore, we attempted to tag the adverbs with the speaker's communicative intention. Regarding the understanding of a situation, there can be various viewpoints, such as the variable judgment viewpoint or the certainty judgment viewpoint.

We intended to create a dictionary containing these viewpoints because such a dictionary will be useful for various forms of text analysis.

The cost of analyzing all viewpoints at one time is high, though, so we decided to analyze only the expressed desirability of the situation. We assumed that an adverb could be used to express a positive, negative, or neutral feeling. Each assumed use was added to the adverb as a tag. The tag "p" indicated that the adverb had a positive connotation, "n" indicated a negative connotation, and "0" indicated neither a positive nor a negative connotation.

The target adverbs were gathered from among the words classified as adverbs in the EDR and IPADIC; for this we used (Matsumoto et al., 1999), a Japanese morphological analyzer. In this way, we obtained 4,759 adverbs.

When adding the tags, we referred to *Gendai Fukushi Youhou Jiten*, an adverb usage dictionary where adverbs are classified as having a positive, negative, or neutral connotation(Hida and Asada, 1994). We added tags to 833 adverbs.

3.2. Automatic classification of adverbs using machine learning

In this section, we describe the method used to calculate the probabilities in our study. We used the maximum-entropy method because it can be used to calculate the probabilities of tags [1].

- Method based on the maximum-entropy method (Ristad, 1997; Ristad, 1998)

  In this method, the distribution of probabilities $p(a, b)$ is calculated for the case where Equation (1) is satisfied and Equation (2) is maximized; the desired probabilities $p(a|b)$ are then calculated using the distribution of probabilities $p(a, b)$:

$$\sum_{a \in A, b \in B} p(a, b) g_j(a, b) = \sum_{a \in A, b \in B} \tilde{p}(a, b) g_j(a, b) \quad (1)$$
$$for \ \forall f_j \ (1 \leq j \leq k)$$

$$H(p) = - \sum_{a \in A, b \in B} p(a, b) \, log \, (p(a, b)), \quad (2)$$

  where $A, B,$ and F are, respectively, sets of categories, contexts, and features $f_j (\in F, 1 \leq j \leq k)$; $g_j(a, b)$ is a function defined as 1 when context b has feature f_j and the category is a, or defined as 0 otherwise; and $\tilde{p}(a, b)$ is the occurrence rate of (a, b) in the training data.

  $p(a|b)$ can be calculated by the above equations and given as

$$p(a|b) = \frac{\prod_{j=1}^{k} \alpha_{a,j}^{g_j(a,b)}}{\sum_{a \in A} \prod_{j=1}^{k} \alpha_{a,j}^{g_j(a,b)}} \quad (3)$$
$$(0 \leq \alpha_{a,j} \leq \infty)$$

  where

$$\alpha_{a,j} = e^{\lambda_{a,j}} \quad (4)$$

$\lambda_{a,j}$ is a parameter of $g_j(a, b)$. The parameter $\lambda_{a,j}$ or $\alpha_{a,j}$ indicates the weight of feature f_j when category a occurs in context b. This parameter is calculated by using numerical

---

[1]We initially used the support vector machine method (Cristianini and Shawe-Taylor, 2000) for corpus correction, because it is known to be very effective. The support vector machine method cannot calculate tag probabilities, though, so we had to abandon this method when we decided to use the tag probabilities for corpus correction.

methods such as *improved iterative scaling* (Pietra et al., 1995). In our experiments, we deleted the combinations of category a and feature f_j that occurred only once and decreased the number of calculated features, because the system could handle only a limited number of features for calculation.

In general, the distribution of $\tilde{p}(a, b)$ is very sparse. We cannot use it directly, so we must estimate the true distribution of $p(a, b)$ from the distribution of $\tilde{p}(a, b)$. We assume that the estimated values of the frequency of each pair of category and feature calculated from $\tilde{p}(a, b)$ are the same as those calculated from $p(a, b)$. (This corresponds to Equation (1).) These estimated values are less sparse. We can thus use the above assumption for calculating $p(a, b)$. Furthermore, we maximize the entropy of the distribution of $\tilde{p}(a, b)$ to obtain one solution for $\tilde{p}(a, b)$, because using only Equation (1) produces many solutions for $\tilde{p}(a, b)$. Maximizing the entropy has the effect of making the distribution more uniform and is known to be a good solution for data sparseness problems.

Next, we describe the features which provide the context when the probabilities are calculated. In this paper, we used the last 10 words of five sentences which included the target adverb. The five sentences were randomly selected from Web texts in advance.

Four data sets which consisted of tagged adverbs were prepared as learning data. We used all the tagged adverbs (883 words) in examination 1. We used the 170 adverbs classified as declarative adverbs in the EDR in examination 2. We used the 713 words not used in examination 2 for examination 3. We used 219 words not listed in the EDR for examination 4.

3.3. Correction of erroneous tags using machine learning

In this section, we briefly describe the method we used to correct the dictionary created using the machine learning method. This method was proposed by Murata et al (Murata et al., 2005). An outline of our corpus correction method is given below:

1. We first calculate the probabilities for each tag category (including the tag category originally assigned to the sentence).

2. We then use these probabilities to judge whether the tag is correct.

    (a) We consider the tag to be correct when its category has the highest probability.

    (b) We consider the tag to be incorrect when one of the other categories has the highest probability.

3. Finally, we correct a tag judged to be incorrect. This correction is done by changing the tag to the tag of the category with the highest probability. (In practice this correction is confirmed by annotators.)

With this method, we can estimate the likelihood of each tag being incorrect and begin by correcting the errors where the value is highest. This is convenient for actual corpus correction.

We calculated the probabilities in an open experiment using 10-fold cross-validation.

4. Experimental results

The results from the classification experiment are shown in Table 1. Further examination of the features used for the classification seems necessary since the overall accuracy for the open data was lower than desired.

Next, we consider the results of error tag correction using the machine learning method. We sorted the incorrect outputs in order of the probability calculated in the classification experiment. We then examined the results manually. Table 2 shows adverbs whose tags we had to correct because of the results of the open experiment.

For example, for the adverb "*furutte* (willingly)", the following are typical of the sentences extracted from the Web corpus.

- *Furutte*

    – *Otomodachi mo osasoi no ue, furutte gosanka kudasai* (Please also invite your friend, and care to join us willingly.)

    – *Goannai ga iki mashitara, furutte gosanka onegai itashimasu.* (When you receive the guide, and please care to join us willingly.)

"*Furutte*" is described as follows in the *Gendai Fukushi Youhou Jiten*:

> This is an example of a word used as a modifier for a proposition, where the speaker looks forward to active action on the part of the listener. This word is used for objective expression and does not suggest a specific affection.

Although "*furutte*" has neither a positive nor a negative connotation, the speaker's positive call is found in the extracted Web texts, and this word suggests the speaker's hope. For this reason, it is fitting that the tag of "*furutte*" be "p" to represent a positive connotation.

The second example, "*imahitotsu*", is marked by an asterisk in Table 2. The original tag was "n", but we changed the tag to "p" when we found the tag was inaccurate. This is an example of a tag whose connotation was different from what we thought and the error was extracted.

As explained above, we were able to find adverbs where the initial classification was uncertain and review the classification to confirm it or remove the error. This is why the classification of adverbs using machine learning is advantageous for creating a dictionary that includes such information.

5. Future work

In our future work, we will compare the accuracy achieved using our dictionary to that obtained using applications such as opinion extraction to classify adverbs.

6. References

Nello Cristianini and John Shawe-Taylor. 2000. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press.

Data	Closed data	Open data
Examination 1	97.7	59.5
Examination 2	98.8	58.2
Examination 3	97.9	60.0
Examination 4	96.8	60.7

Table 1: Results for each data set

Data	Probability	Output tag	Original tag	Adverb
Examination 1	0.552	p	0	*ichihayaku* (Chinese character) (fastest of all)
	0.545	p	0	*ichihayaku* (hiragana)
	0.414	n	p	*imahitotsu* (quite)∗
Examination 2	0.835	n	0	*chittomo* (at all)
Examination 3	0.591	n	p	*imahitotsu* (quite)∗
Examination 4	0.993	p	0	*furutte* (willingly)

Table 2: Examples of errors extracted from open data

EDR, 1993. *EDR Electronic Dictionary Technical Guide*. EDR (Japan Electronic Dictionary Research Institute, Ltd.).

Yoshifumi Hida and Hideko Asada. 1994. *Gendai Fukushi Youhou Jiten (Adverb usage dictionary)*. Tokyoudou Shuppan. (in Japanese).

Yuji Matsumoto, Akira Kitauchi, Tatsuo Yamashita, Yoshitaka Hirano, Hiroshi Matsuda, and Masayuki Asahara. 1999. Japanese morphological analysis system ChaSen version 2.0 manual, 2nd edition.

Masaki Murata, Masao Utiyama, Kiyotaka Uchimoto, Qing Ma, and Hitoshi Isahara. 2005. Correction of errors in a verb modality corpus used for machine translation with a machine-learning method. *ACM Transactions on Asian Language Information Processing*.

Minoru Nakau. 1994. *Ninchi Imiron no Genri (Principles of Cognitive Semantics)*. Taishukan Shoten. (in Japanese).

Yoshio Nitta. 2002. *Fukushiteki Hyogen no Shosou (Aspects of Adverbial Expression)*. Kuroshio Shuppan. (in Japanese).

Stephen Della Pietra, Vincent Della Pietra, and John Lafferty. 1995. Inducing features of random fields. Technical report, Carnegie Mellon University. CMU-CS-95-144.

Eric Sven Ristad. 1997. Maximum Entropy Modeling for Natural Language. ACL/EACL Tutorial Program, Madrid.

Eric Sven Ristad. 1998. Maximum Entropy Modeling Toolkit, Release 1.6 beta. http://www.mnemonic.com/software/memt.

Minoru Watanabe. 1983. *Fukuyougo no Kenkyu (Adverb Research)*. Meiji Shoin. (in Japanese).

Yoshio Yamada. 1936. *Nihon Bunpou Gaku Gairon (Survey of Japanese Grammar)*. Houbun Kan. (in Japanese).