

Unsupervised Japanese-Chinese Opinion Word Translation Using Dependency Distance and Feature-Opinion Association Weight

*Guo-Hau Lai, Ying-Mei Guo and Richard Tzong-Han Tsai**

Department of Computer Science and Engineering, Yuan Ze University
135 Yuan-Tung Road, Chungli, Taoyuan, Taiwan, R.O.C.

*Corresponding Author

{s986005, s996148}@mail.yzu.edu.tw, thtsai@saturn.yzu.edu.tw

ABSTRACT

Online shoppers depend on customer reviews when evaluating products or services. However, in the international online marketplace, reviews in a user's language may not be available. Translation of online customer reviews is therefore an important service. A crucial aspect of this task is translating opinion words, key words that capture the reviewers' sentiments. This is challenging because opinion words often have multiple translations. We propose an unsupervised opinion word translation disambiguation scoring method using dependency distance and feature-opinion association as weighting factors. The scores of an opinion word's translation and its surrounding words' translations are estimated using Google snippets. We focus on Japanese-Chinese translation of hotel reviews from Rakutan Travel, using the 10 most common polysemous Japanese opinion words to evaluate system performance. Results show our weighting factors significantly improve translation accuracy compared to Google and Excite.

KEYWORDS : Opinion Word Translation Disambiguation, Dependency Distance, Feature-Opinion Association

1 Introduction

The development of Web 2.0 has made it easier for internet users to post their reviews or comments about products or services on structured websites. Online shoppers are increasingly likely to look at these reviews before deciding on a purchase. In recent years, the research field of sentiment analysis has focused on analyzing this form of textual-information, particularly opinions or sentiments expressed by internet users.

However, given the international nature of the web and online shopping, opinions in a user's mother language may not be available. Translation of online customer reviews is therefore an important service sought after in many markets. A crucial aspect of customer review translation is translating opinion words, key words that capture the sentiments. At present, machine translation (MT) systems can translate whole sentences or even complete paragraphs. This not a trivial task, however, because opinion words usually have multiple possible translations and the MT systems have low accuracy on polysemous words (Carpuat & Wu, 2005).

This paper proposes an unsupervised method of selecting the most appropriate Chinese translation for an opinion word in a given Japanese sentence. Candidate translations are retrieved from a bilingual dictionary. Consider the following Japanese sentence:

綺麗な夜景とともに食事を楽しむことができました。

(I was able to enjoy a nice meal with a beautiful night view.)

The target opinion word 綺麗 has three candidate translations: 漂亮 (beautiful), 乾淨 (clean), and 清楚 (clear) in Chinese. In this example, the most appropriate translation is 漂亮 (beautiful). This disambiguation problem is known as Word Translation Disambiguation (WTD).

One way to solve the WTD problem is to calculate the sum of association scores of pairs among translation of the target word and all its surrounding words' translations, and then select the one with the highest score. However, since the different surrounding words have different amounts of influence on the target word, it is necessary to add some weighting factors (e.g., word distance). Since our goal is disambiguating opinion words in opinionated sentences, the product features (or aspect expressions) should have direct influence on translation selection. In the above example, 夜景 (night view) and 食事 (meal) are product features, the former having the greatest influence on the opinion word 綺麗 (beautiful).

Our proposed unsupervised opinion word translation disambiguation scoring method uses the dependency distance and feature-opinion association as weighting factors. The scores of an opinion word's translation and its surrounding words' translations are estimated using Google search snippets. In our experiments, we focused on opinion word translation of hotel reviews from Japanese to traditional Chinese. From a dataset of hotel reviews compiled from Rakutan Travel, we selected the top-10 most common polysemous Japanese opinion words to evaluate the performance of our system. The results show that our weighting factors have significantly improved translation accuracy. Compared to Google Translate and Excite translation system, our system can translate opinions more accurately, which could be a boon for Chinese online shoppers seeking accommodations in Japan.

The remainder of this paper is organized as follows: Section 2 introduces some related work while Section 3 describes our proposed method in detail. The experimental results are given in Section 4. The error analysis discussed in Section 5. Finally, conclusion gives in the last section.

2 Related Work

In this section, we introduce some previous works related to our method.

Word translation disambiguation (WTD) (Marsi, Lynum, Bungum, & Gambäck, 2011), also called cross-lingual word sense disambiguation (CL-WSD), which is the task of selecting the most appropriate translation of a polysemous word in a given context. This task can be seen as a special variant of WSD.

The best-known open task in this specialty is the Multilingual Lexical Sample/CL-WSD task, held by the Senseval/SemEval workshop (Chklovski, Mihalcea, Pedersen, & Purandare, 2004; Jin, Wu, & Yu, 2007; Lefever & Hoste, 2010). This task provides a framework for the evaluation of systems that perform machine translation, with a focus on the translation of ambiguous words. Unlike in other lexical sample tasks, the sense inventory for CL-WSD is the set of translations from a bilingual dictionary or a parallel corpus instead of human-defined sense labels.

There have been several studies that use cross-lingual evidence to deal with the WSD problem: (Chan & Ng, 2005; Chklovski et al., 2004; Ng, Wang, & Chan, 2003). These approaches rely on large parallel corpora to train a WSD classifier. However, for some language pairs (e.g., Japanese-Chinese), such corpora are not available. To overcome this problem, Dagan and Itai (1994) used a bilingual lexicon and statistical data from a monolingual corpus of the target language for disambiguation. Tsunakawa and Kaji (2010) proposed a method for using a bilingual dictionary with a correlation matrix to select an appropriate translation word. An item in the matrix is the correlation score between associated words and candidate translations.

The Web is increasingly being used as a data source in a wide range of natural language processing tasks including WSD. Liu and Zhao (2009) presented a fully unsupervised WTD method which selects the maximum sum of Web Bilingual Relatedness (WBR) between a translation and all context words. The WBR is calculated by four association measures based on mixed-language webpage counts from the Baidu search engine. Their WBR model outperformed the best unsupervised SemEval-2007 participant system in the Multilingual Chinese-English Lexical Sample Task. Another work using the Web as a knowledge resource is (Liu, Xue, Li, & Liu, 2010), which is based on minimum Normalized Google Distance and also outperformed the best unsupervised participant system in SemEval-2007.

Most of these studies use the corpus statistics to measure the association between a candidate translation and its context words to disambiguate polysemous words. They tend to consider that all context words have equal influence on a target word. However, since our work is a special case of WTD that focuses on translating polysemous opinion words for a given opinionated sentence, we need give more weight to words that are closely related to opinion words, such as the product features.

3 Opinion Word Translation Disambiguation

In this section, we introduce our method for selecting the translation of a given opinion word in a sentence. The procedure consists of six parts: (1) related word extraction, (2) related word translation, (3) translation corpus, (4) Japanese dependency analysis, (5) related product feature identification, and (6) word translation disambiguation scoring method. The procedures are described in detail in the following sections.

For related word extraction, we apply Japanese compound combination rules to extract nearby words related to the target word. Then, the extracted related words are translated from Japanese to Chinese by our dictionary-based system, which uses an online bilingual dictionary. The translation corpus is compiled from snippets returned by Google Search. To obtain dependency distances among the target word and related words, we feed the given sentence into a Japanese dependency analyzer. We also identify all product features appearing in the given sentence and estimate the association between the target word and each feature. Finally, the disambiguation scoring method calculates scores for each candidate translation to determine the appropriate one.

3.1 Related Word Extraction

We use the following procedure to extract nearby words related to the target word: First, the test sentences are analyzed by a Japanese part-of-speech (POS) and morphological analyzer, MeCab (Kudo, 2005). The MeCab output contains not only the segmented words with POS tags but also detailed information on the *katsuyou* form, root form, and pronunciation of each word. The *katsuyou* is the inflection of the *yougen* (a verb, an adjective, or an auxiliary verb). According to its tense and voice, the *yougen* may have different inflections. For example, 美味しい (delicious) should inflect to 美味しかった in the past tense and 美味しくない (not delicious) in the negative. However, dictionaries usually do not include all these inflections. So we use the root forms instead of the original segmented words in subsequent processing steps.

In addition, we found one difficulty using MeCab: due to the annotation standards of its training corpus, MeCab sometimes treats one compound or loanword as many morphemes. For example, 従業員 (staff) is separated into 従業 (work) and 員 (member), and the loanword エントランス (entrance) is separated into エン (dollar) and トランス (transformer). In both cases, the original meaning is lost. To solve this problem, we apply the Japanese compound combination rules shown in TABLE 1. Most of these morphological rules are based on POS tags. For example, a compound adjective (*comadj*) is composed of a verb (main) followed by an adjective (sub).

<i>noun</i> → noun-verbal noun-common noun-proper-misc <i>prefix</i> → prefix-nominal <i>suffix</i> → noun-suffix-misc <i>noun'</i> → <i>noun'</i> + <i>noun</i> <i>noun</i> + <i>noun</i> <i>commoun</i> → <i>prefix</i> + <i>noun</i> + <i>suffix</i> <i>prefix</i> + <i>noun'</i> <i>noun</i> + <i>suffix</i> <i>noun'</i> + <i>suffix</i> <i>noun'</i> <i>comadj</i> → verb-main + adjective-sub <i>location</i> → noun-proper-place-misc + noun-suffix-place <i>comword</i> → <i>commoun</i> <i>comadj</i> <i>location</i> <i>comadj</i> : compound adjective <i>commoun</i> : compound noun <i>comword</i> : compound word

TABLE 1 – Japanese Compound Combination Rules

Finally, we filter out irrelevant words using a list of stop words. From the remaining words, we retain only adjectives, verbs and nouns with the following POS tags, as shown in TABLE 2.

Type	POS tags
adjective	adjective-main, adjective-suffix, adjective-sub
verb	verb-main
noun	noun-verbal, noun-common, noun-adjective-base, noun-proper-misc, noun-proper-organization

TABLE 2 – The list of POS tags for retained words

3.2 Related Word Translation

The extracted related words are translated by our dictionary-based system, which uses the Sanseido Japanese-Chinese dictionary¹. Japanese words are input and Chinese translations are output. Given a Japanese word w_j , all translations of w_j in the dictionary are regarded as potential translations. The Sanseido dictionary includes a total of 28,000 entries and provides the majority of Chinese translations for this study. For few related words that do not appear in this dictionary, we use results from Google Translate.

Since the Sanseido dictionary’s Chinese translations are in simplified Chinese, our system must convert the output to traditional Chinese characters. Direct conversion by table lookup may result in mistranslations since the usage of some terms is quite different in mainland China and Taiwan. We use a mapping table of common synonymous words provided by China Biz² to improve conversion accuracy.

One translation difficulty that is often encountered in informal-style online user reviews is varying use of common words or expressions. For example, the word すごい (very) can be expressed by the hiragana terms すごーい (terrible) or すっごい (terrific) and the katakana term スゴイ (amazing). Ikeda, Yanagihara, Matsumoto, and Takishima (2009) proposed a normalization algorithm to reduce the number of variant expressions. We apply some of their conversion rules to improve the recall of our translation system. For example, if words are written in all katakana (e.g., スゴイ, the above example), it may imply emphatic use. In this case, we convert the words to hiragana (e.g., スゴイ → すごい). In polite usage, the honorific prefixes (e.g., お, ご, and 御) are used in words such as お手洗い (toilet). In these cases, we remove the prefixes (e.g., お手洗い → 手洗い).

3.3 Translation Corpus

We compiled our translation corpus from snippets returned by Google Search for conjunctive queries of word pairs. These snippets provide useful clues related to the semantic relations that exist between two words. First, we take Chinese word pairs in which one word is a candidate translation of the target word and the other is a translation of a related word. Then, we submit each word pair joined by the Boolean operator “AND” to Google Search and collect the first 500 snippets for our text corpus.

¹ http://www.excite.co.jp/dictionary/japanese_chinese/

² <http://www.chinabiz.org.tw/>

3.4 Japanese Dependency Analysis

In order to obtain the dependency distance between the target word and a related word, we use CaboCha, a Japanese dependency structure analyzer based on Support Vector Machines (SVMs) and the most accurate publicly available system to date, with a reported accuracy of 89.29% (Kudo & Matsumoto, 2002). CaboCha uses a parsing algorithm based on the Cascaded Chunking Model.

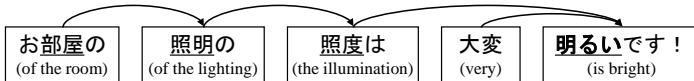


FIGURE 1 – An example of Japanese dependency parsing

The basic syntactic unit used in Japanese parsing is the *bunsetsu*, which consists of one or more words followed by either nothing or function words such as particles and auxiliary verbs. FIGURE 1 shows an example of Japanese dependency parsing for a sentence お部屋の照明の照度は大変明るいです! (The illumination of the lighting of the room is very bright!). In this example sentence, we first find the last node which contains the target word 明るい (bright). Then we calculate all dependency distances from other nodes which contain a related word. For example, the distance from the target word’s node to the third node containing the related word 照度 (illumination) is one and the distance to the second node containing 照明 (lighting) is two.

3.5 Related Product Feature Identification

In opinionated sentences, opinion words often describe product attributes or features, such as cleanliness, staff attitude, food quality, etc. in the hotel domain. We believe that considering the product feature(s) related to the target opinion word is helpful for disambiguation of the opinion word. To implement this feature, we enumerated product features³ for the hotel domain.

3.6 Word Translation Disambiguation Scoring Method

This section describes the word translation disambiguation scoring method. Assume the target opinion word is o . One way to select the appropriate translation of o is to first calculate the association scores for pairs of each candidate translation and each of its related words’ translations, and then select the translation with highest sum of these scores. Consider the following formula (Assume t is a candidate translation for o):

$$Translation(t | o) = \sum_{s_i \in S} \sum_{t' \in translations(s_i)} \frac{Association_T(t, t') \cdot Association_S(o, s_i)}{|translations(s_i)|} \quad (1)$$

where $Translation(t | o)$ is the score of the candidate translation t , S is the set of all related words, and $translations(s_i)$ is the set of s_i ’s translations (for convenience, hereafter referred to as related translations). For instance, o has two candidate translations (t_1 and t_2) and two related words (s_1 and s_2). The translations of s_1 are $t'_{1,1}$, $t'_{1,2}$ and $t'_{1,3}$, and s_2 can be translated to $t'_{2,1}$. Using the above notation, $translations(s_1) = \{t'_{1,1}, t'_{1,2}, t'_{1,3}\}$ and $translations(s_2) = \{t'_{2,1}\}$. $Association_T(t, t')$

³ The product feature list is available at <http://iisr.cse.yzu.edu.tw/~guohau/coling/feature.list>

is the association score between a candidate translation t and a related translation t' . $Association_S(o, s_i)$ is the association score between the target opinion word o and a related word s_i —this weighting factor is primarily designed to model their association in the Japanese context. Notice that the final score is divided by the size of $translations(s_i)$. The purpose of adding this term is to balance out the extra influence of s_i 's multiple possible translations. In the next two parts, we will introduce the terms $Association_T(t, t')$ and $Association_S(o, s_i)$ in detail.

3.6.1 Association Score between Translations in the Target Language

In this section, we describe how we estimate $Association_T(t, t')$. There are many ways to measure the correlation between two words: one simple way is to calculate the mutual information score in the corpus. First, the snippets described in Section 3.3 are split into segments using punctuation. Their similarity is estimated by Pointwise Mutual Information (PMI) (Church & Hanks, 1990), which is defined as:

$$Association_T(t, t') = PMI(t, t') = \log_2 \frac{p(t, t')}{p(t) \cdot p(t')} \quad (2)$$

where $p(t)$ and $p(t')$ are the probability of word t and t' appearing separately in the corpus, and $p(t, t')$ is the probability of the co-occurrence of word t and t' in the corpus, which is estimated by the number of co-occurring segments for t and t' divided by the total number of segments. However, using PMI does not yield the expected $Association_T(t, t')$ measurements, because the corpus is compiled from the search snippet results of sending all candidate-/related-translation pairs to Google Search. The very low or zero co-occurrence frequencies of some pairs cause difficulty in calculating their values. For instance:

$$\text{if } \tilde{p}(t, t') \rightarrow 0, \text{ then } PMI(t, t') \rightarrow -\infty.$$

In addition, according to Formula 1, each term in the summation is the product of two scores: $Association_T$ and $Association_S$. For two related words s_1 and s_2 and any of their translations $t'_{1,j}$ and $t'_{2,k}$, if $Association_T(t, t'_{1,j})$ is greater than $Association_T(t, t'_{2,k})$ and $Association_S(o, s_1)$ is greater than $Association_S(o, s_2)$, but $Association_T(t, t'_{1,j})$ and $Association_T(t, t'_{2,k})$ are both negative (PMI score), $Association_T(t, t'_{1,j}) * Association_S(o, s_1)$ is not guaranteed to be larger than $Association_T(t, t'_{2,k}) * Association_S(o, s_2)$. Such a result is not appropriate for our application here.

To solve this problem, PMI is mapped to an exponential function where the value is always positive, which is defined as:

$$PMI_{Exp}(t, t') = e^{PMI(t, t')} \quad (3)$$

3.6.2 Association Score between the Opinion Word and Related Words in the Source Language

We aim to determine the translation of an opinion word by scoring its association to its related words in the source language. Different related words have different influence on the target word, so added weighting factors are necessary. There are two factors that we consider: dependency distance and feature-opinion association.

Distance weighting has been used in several studies (Beeferman, Berger, & Lafferty, 1997; Brosseau-Villeneuve, Nie, & Kando, 2010; Gao, Zhou, Nie, He, & Chen, 2002) as a means of estimating the association between two words. The exponential model, in which association between two words decreases exponentially when the distance between them increases, is a commonly used approach. We employ Beeferman et al. (1997)'s distance weighting approach. Therefore, the association score of o and s_i is defined as:

$$Association_s(o, s_i) = \mu \cdot e^{-\mu \cdot distance(o, s_i)} \quad (4)$$

where $distance(o, s_i)$ is the dependency distance (see Section 3.4 for the details) between o and s_i ; and μ is the parameter for decay rate determined by maximum likelihood estimate:

$$\mu = \log_2 \left(1 + \frac{1}{E_{\tilde{p}}[k]} \right) = E_{\tilde{p}}[k] \sum_{k \geq 0} k \tilde{p}(k) \quad (5)$$

where $\tilde{p}(k)$ is the probability of the distance between o and s_i being k .

Since our goal is disambiguating opinion words in opinionated sentences, we should give product-feature words more influence on the translation selection than normal words. To do this automatically, we modify Formula 4 by introducing the feature-opinion association (FOA) score, which is defined as:

$$FOA(o, s_i) = \begin{cases} \min \left[J(o, s_i), \frac{1}{\lambda} \right], & \text{if } s_i \text{ is a predefined product feature} \\ \frac{1}{\lambda} & \text{if } s_i \text{ is not a predefined product feature} \end{cases} \quad (6)$$

where the constant value λ is determined empirically to be 1500. The purpose of introducing the minimal function is to set the lower bound of FOA to be $1 / \lambda$, which is a reasonably small value. $J(o, s_i)$ is the Jaccard coefficient, which is defined as:

$$J(o, s_i) = \frac{|o \cap s_i|}{|o \cup s_i|} = \frac{freq(o, s_i)}{freq(o) + freq(s_i) - freq(o, s_i)} \quad (7)$$

Then, we can modify $Association_s(o, s_i)$ as follows:

$$Association_s(o, s_i) = \mu \cdot e^{-\mu \cdot IFOA(o, s_i) \cdot distance(o, s_i)}, \quad (8)$$

$$IFOA(o, s_i) = \frac{1}{\lambda \cdot FOA(o, s_i)}$$

where IFOA is the abbreviation of the inversed FOA score.

In Formula 6, we mentioned that when s_i is not a predefined product feature, $FOA(o, s_i)$ is $1 / \lambda$, making the $IFOA(o, s_i)$ 1, which implies that $IFOA(o, s_i)$ does not have any effect. This increases the influence of related product-feature words while having no influence on regular related words.

3.7 Application of Our WTD Formulae

Let us consider the following example to demonstrate our proposed WTD scoring method:

お部屋もお風呂も綺麗に掃除がされている。

(The room and bathroom have been swept clean.)

The target opinion word 綺麗 has three candidate translations: 漂亮 (beautiful), 乾淨 (clean), and 清楚 (clear). For convenience, we consider only the first two: 漂亮 (beautiful) and 乾淨 (clean). The steps of our WTD method are as follows:

1. Related Word Extraction

Three related words are extracted from the input sentence: 部屋 (room), 風呂 (bathroom), and 掃除 (sweep).

2. Related Word Translation

The related words are translated into Chinese (called related translations): 房間 (room) and 屋子 (house) for 部屋, 浴室 (bathroom) for 風呂, 打掃 (sweep) for 掃除. TABLE 3 lists Japanese words and their Chinese translations used in this application.

Japanese word	Chinese translations
綺麗	漂亮 (beautiful), 乾淨 (clean), 清楚 (clear)
部屋	房間 (room), 屋子 (house)
風呂	浴室 (bathroom)
掃除	打掃 (sweep)

TABLE 3 – Japanese word and its Chinese translations

3. Translation Corpus

We look up the Chinese word pairs (e.g., “漂亮” AND “房間”, “乾淨” AND “房間”) in our Google Search translation corpus (Section 3.3) and collect the snippets.

4. Japanese Dependency Analysis

After dependency analysis, the dependency distances between each target opinion word and the related word are acquired:

{綺麗-部屋: 3, 綺麗-風呂: 2, 綺麗-掃除: 2}

5. Related Product Features Identification

The product features are 部屋 (room) and 風呂 (bathroom).

6. Word Translation Disambiguation Scoring Method

Consider Formula 1:

$$Translation(t | o) = \sum_{s_i \in S} \sum_{t' \in translations(s_i)} \frac{Association_t(t, t') \cdot Association_s(o, s_i)}{|translations(s_i)|}$$

o : the target opinion word 綺麗

s_i : any related word in {部屋, 風呂, 掃除}

t : any candidate translation in {漂亮, 乾淨}

t' : any related translation in {房間, 屋子, 浴室, 打掃}

Consider Formula 2:

Assume the $Association_t(t, t')$ for each candidate translation-related translation pair is:

{漂亮-房間: 0.5, 漂亮-屋子: 0.55, 漂亮-浴室: 0.4, 漂亮-打掃: 0.35,

乾淨-房間: 0.6, 乾淨-屋子: 0.45, 乾淨-浴室: 0.7, 乾淨-打掃: 0.75}

Consider Formula 5:

$$\mu = \log_2 \left(1 + \frac{3}{3+2+2} \right) = 0.51$$

Consider Formulae 6 to 8:

We assume the feature-opinion association scores [$FOA(o, s_i)$] are:

$$FOA(\text{綺麗, 部屋}) = 0.0025, IFOA(\text{綺麗, 部屋}) = 0.27$$

$$FOA(\text{綺麗, 風呂}) = 0.0032, IFOA(\text{綺麗, 風呂}) = 0.21$$

Then, the $Association_S(o, s_i)$ between the target opinion word and each related word is:

$$Association_S(\text{綺麗, 部屋}) = 0.51 * e^{-0.51 * 0.27 * 3} = 0.34$$

$$Association_S(\text{綺麗, 風呂}) = 0.51 * e^{-0.51 * 0.21 * 2} = 0.41$$

$$Association_S(\text{綺麗, 掃除}) = 0.51 * e^{-0.51 * 1 * 2} = 0.18$$

Now, we can calculate the weighted score for each candidate translation:

$Translation(\text{漂亮} | \text{綺麗})$

$$\begin{aligned} &= Association_S(\text{綺麗, 部屋}) * Association_T(\text{漂亮, 房間}) / |translations(\text{部屋})| + \\ & \quad Association_S(\text{綺麗, 部屋}) * Association_T(\text{漂亮, 屋子}) / |translations(\text{部屋})| + \\ & \quad Association_S(\text{綺麗, 風呂}) * Association_T(\text{漂亮, 浴室}) / |translations(\text{風呂})| + \\ & \quad Association_S(\text{綺麗, 掃除}) * Association_T(\text{漂亮, 打掃}) / |translations(\text{掃除})| \\ &= 0.34 * 0.5 / 2 + 0.34 * 0.55 / 2 + 0.41 * 0.4 / 1 + 0.18 * 0.35 / 1 \\ &= 0.41 \end{aligned}$$

$Translation(\text{乾淨} | \text{綺麗})$

$$\begin{aligned} &= Association_S(\text{綺麗, 部屋}) * Association_T(\text{乾淨, 房間}) / |translations(\text{部屋})| + \\ & \quad Association_S(\text{綺麗, 部屋}) * Association_T(\text{乾淨, 屋子}) / |translations(\text{部屋})| + \\ & \quad Association_S(\text{綺麗, 風呂}) * Association_T(\text{乾淨, 浴室}) / |translations(\text{風呂})| + \\ & \quad Association_S(\text{綺麗, 掃除}) * Association_T(\text{乾淨, 打掃}) / |translations(\text{掃除})| \\ &= 0.34 * 0.6 / 0.5 + 0.34 * 0.45 / 0.5 + 0.41 * 0.7 / 1 + 0.18 * 0.75 / 1 \\ &= 0.6 \end{aligned}$$

So, the chosen translation is 乾淨 (clean).

4 Experiments

We conducted experiments with our Japanese hotel review corpus to empirically evaluate the translation accuracy of our WTD scoring method using different sets of weighting factors and the modified PMI formula. We also compared our system's performance to that of Google Translate and the Excite translation system.

4.1 Dataset

Our dataset consists of 956,892 reviews of 15,291 hotels from the Rakutan Travel website⁴, the largest hotel-booking/review website in Japan. The sentences are segmented and duplicate content is removed. After processing, the dataset contains 4,341,266 sentences. We also use this data to create the product feature list for Section 3.5.

⁴ <http://travel.rakuten.co.jp/>

4.2 Experiment Design

We selected the top-10 most common polysemous opinion words and annotated each of their occurrences in the dataset with their translations. For each of the ten opinion words, we constructed test examples by randomly selecting 1,200 sentences from the dataset. Each test example contains only one opinion word. The ground truth of each translation was assigned by two human annotators. Statistics for the gold standard dataset are presented in TABLE 4.

Word	#Sense	#Instance	Avg length	Min length	Max length
明るい (bright)	2	992	36.3	7	135
甘い (sweet)	2	808	40.3	7	145
暖かい (warm)	2	979	41.6	6	147
丁寧 (polite)	2	1,057	37.9	11	125
冷たい (cool)	2	957	44.0	6	174
薄い (thin)	2	1,041	38.3	6	141
綺麗 (beautiful)	3	736	35.2	9	113
きつい (tiring)	3	755	41.9	10	136
寂しい (lonely)	3	794	38.7	8	120
厳しい (strict)	4	506	45.0	7	141
Avg.	2.5	862.5	39.9	7.7	137.7

TABLE 4 – Statistics for the gold standard dataset

As shown in TABLE 4, our gold standard dataset contains a total of 8,625 test examples with an average of 2.5 senses per word. The average length of the test examples is 39.9 Japanese characters.

In order to measure the impact of different sets of weighting factors and modified PMI formula on translation accuracy, we ran a set of 30 experiments for each configuration. For each experiment, we randomly chose 85% of the test examples for each opinion word. After running all the experiments, we performed a two-tailed paired T-test on the average accuracies of different sets of weighting factors to prove our weighting approach significantly improved translation accuracy. We also compared our system’s performance with that of Google Translate and the Excite translation system using the same test method. The online translation systems’ performance was checked by two annotators (only for the correctness of the opinion word translation, but ignoring translation of surrounding words).

4.3 Evaluation Metrics

For each opinion word, the results are given in terms of translation accuracy, which is defined as:

$$accuracy = \frac{\# \text{ of correct translations}}{\# \text{ of test sentences}} \quad (9)$$

We also calculated macro and micro averages to measure the overall performance across all opinion words. The macro average is simply the average of the accuracies of all ten opinion words. In contrast, micro average is the sum of correct occurrences divided by the sum of all occurrences.

4.4 Results

TABLE 5 shows the experimental results for the different configurations of our WTD scoring method. The value in each cell indicates the average accuracy. For our baseline system, we used the most frequent sense (MFS) method:

$$MFS = \frac{\#most\ frequent\ sense}{\#test\ sentence} \quad (10)$$

In TABLE 5, $A_T(\text{PMI})$ stands for the configuration in which $Association_T(t, t')$ is estimated by the original PMI (Formula 2) and $Association_S(o, s_i)$ is set to 1. $A_T(\text{PMI}_{Exp})$ means that the modified PMI formula (Formula 3) is used to replace Formula 2. $A_T(\text{PMI}_{Exp})+A_S(D)$ means that $Association_S(o, s_i)$ is estimated by Formula 4, which considers the dependency distance. $A_T(\text{PMI}_{Exp})+A_S(D+F)$ means that Formula 8, which considers both dependency distance and feature-opinion association, is used to replace Formula 4. TABLE 5 shows that $A_T(\text{PMI}_{Exp})$ significantly improves overall accuracy by about 7.6% over $A_T(\text{PMI})$. $A_T(\text{PMI}_{Exp})+A_S(D)$ improves overall performance by about 9.6% from $A_T(\text{PMI})$, and $A_T(\text{PMI}_{Exp})+A_S(D+F)$ achieved the best result, improving overall accuracy by about 11% over $A_T(\text{PMI})$. It should be noted that $A_T(\text{PMI}_{Exp})+A_S(D+F)$ also has a positive impact on performance of every opinion word individually.

Compared with the online translation systems, TABLE 5 shows our system outperforms Excite and Google for opinion word translation. The two online translation systems perform even worse than MFS on average.

Word	MFS	Excite	Google	$A_T(\text{PMI})$	$A_T(\text{PMI}_{Exp})$	$A_T(\text{PMI}_{Exp})+A_S(D)$	$A_T(\text{PMI}_{Exp})+A_S(D+F)$
明るい	0.7323	0.7311	0.6460	0.8656	0.9067	0.9198	0.9212
甘い	0.7679	0.7679	0.7291	0.8062	0.8431	0.8608	0.8876
暖かい	0.5234	0.4766	0.7103	0.6685	0.7988	0.8183	0.8600
丁寧	0.8284	0.0439	0.7854	0.5045	0.8287	0.8679	0.8711
冷たい	0.9098	0.9119	0.6790	0.9382	0.9119	0.9210	0.9386
薄い	0.9033	0.9033	0.8198	0.8734	0.9088	0.9293	0.9549
綺麗	0.4845	0.5313	0.5408	0.7341	0.7804	0.7875	0.7889
きつい	0.5436	0	0.0408	0.6563	0.7458	0.7649	0.7729
寂しい	0.5275	0.1157	0.0725	0.6263	0.6201	0.6352	0.6666
厳しい	0.4886	0.4846	0.2039	0.5571	0.5592	0.6026	0.6209
micro	0.6932	0.5102	0.5622	0.7327	0.8081	0.8280	0.8456
macro	0.6709	0.4966	0.5227	0.7230	0.7903	0.8107	0.8283

TABLE 5 – Comparison of different translation systems and configurations

In addition, we performed a two-tailed paired T-test on the average accuracies of different weighting factors. The T-test results in TABLE 6 show that different weighting factors can have a statistically significant impact (bold text results) on system performance.

Word	$A_T(\text{PMI}_{\text{Exp}})$	$A_T(\text{PMI}_{\text{Exp}})$ + $A_S(\text{D})$	T-test p-value	$A_T(\text{PMI}_{\text{Exp}})$ + $A_S(\text{D})$	$A_T(\text{PMI}_{\text{Exp}})$ + $A_S(\text{D}+\text{F})$	T-test p-value
明るい	0.9067	0.9198	1.86E-24	0.9198	0.9212	0.00886
甘い	0.8431	0.8608	2.33E-25	0.8608	0.8876	7.05E-33
暖かい	0.7988	0.8183	3.15E-23	0.8183	0.8600	6.85E-34
丁寧	0.8287	0.8679	3.16E-34	0.8679	0.8711	2.51E-06
冷たい	0.9119	0.9210	7.07E-16	0.9210	0.9386	1.42E-26
薄い	0.9088	0.9293	2.03E-29	0.9293	0.9549	3.5E-36
綺麗	0.7804	0.7875	9.85E-10	0.7875	0.7889	0.120067
きつい	0.7458	0.7649	2.77E-20	0.7649	0.7729	1.16E-13
寂しい	0.6201	0.6352	7.97E-21	0.6352	0.6666	4.46E-35
厳しい	0.5592	0.6026	3.16E-30	0.6026	0.6209	1.02E-26
micro	0.8081	0.8280	1.63E-38	0.8280	0.8456	2.33E-40
macro	0.7903	0.8107	2.62E-38	0.8107	0.8283	2.31E-39

TABLE 6 – T-tests on the average accuracies of different weighting factors

5 Discussion

In this section, we discuss the causes of some common errors that our system made.

5.1 Errors caused by Japanese homonyms

Japanese has many homonyms—words that share the same pronunciation but have different meanings and kanji. For example, the words 蜜柑 (orange), 未完 (unfinished), and 未刊 (unpublished) all share the same hiragana spelling みかん (orange) but are represented by different kanji. When calculating the $Association_T$ of hiragana words, the co-occurrence frequency of opinion word/related word pairs may be overestimated.

5.2 Limitations of single word-pair association scores

In the sentence, 暑い時期は駅からの距離がきついです (When it is hot, walking the distance from the station is tiring.), which describes the hotel location, the two most likely Chinese translations for the opinion word きつい are 累人 (tiring) and 擁擠 (crowd), the former being the correct choice. The $Association_T$ between 累人 (tiring) and 距離 (distance) is incorrectly calculated as being lower than that between 擁擠 (crowd) and 距離 (distance). To determine the correct translation in cases like this, we should calculate $Association_T$ between pairs of related words and the opinion word. In the above example, if we consider 暑い (hot) and きつい (tiring) as one entity and calculate the $Association_T$ between 距離 (distance) and 暑い-きつい (hot-tiring), we get the correct translation.

5.3 Target opinion word associated with multiple feature words

If the target opinion word can apply to multiple product feature words in a sentence, the incorrect pairing may end up with the highest FOA. For example, in the sentence エントランスが明るくて従業員の対応も素敵でした。 (The entrance was bright and the staff’s attitude was also friendly.), the opinion word 明るい (bright) can describe both feature words エントランス

(entrance) and 従業員 (staff). In this case our system calculates a higher *FOA* for the 明るい-従業員 (staff-bright) pair.

Conclusion

In this paper, we propose an unsupervised opinion word translation disambiguation scoring method which uses dependency distance and feature-opinion association as weighting factors. The scores of an opinion word's translation and its surrounding words' translations are estimated using Google search snippets. In our experiments, we focused on translation of hotel reviews from Japanese to Chinese. From a dataset of hotel reviews compiled from Rakutan Travel, we selected the top-10 most common polysemous Japanese opinion words to evaluate the performance of our system. The results show that our scoring method for representing the influence of product features and dependency distance improves translation accuracy effectively. Compared to Google Translate and Excite translation system, our system can translate opinion words more accurately, which could be of benefit to Chinese online shoppers seeking accommodations in Japan.

Acknowledgments

We would like to thank the two anonymous reviewers for their valuable comments, which have greatly improved the presentation of this paper. This work was supported in part by the Taiwan National Science Council under Grants NSC 98-2221-E-155-047 and NSC 99-2628-E-155-004.

Reference

- Beeferman, D., Berger, A. and Lafferty, J. (1997). *A model of lexical attraction and repulsion*. Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics, Madrid, Spain.
- Brosseau-Villeneuve, B., Nie, J.-Y. and Kando, N. (2010). *Towards an optimal weighting of context words based on distance*. Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010), Beijing, China.
- Carpuat, M. and Wu, D. (2005). *Evaluating the Word Sense Disambiguation Performance of Statistical Machine Translation*. Proceedings of the 2nd International Joint Conference on Natural Language Processing (IJCNLP 2005), Jeju Island, Korea.
- Chan, Y. S. and Ng, H. T. (2005). *Scaling Up Word Sense Disambiguation via Parallel Texts*. Proceedings of the 20th National Conference on Artificial Intelligence (AAAI 2005), Pittsburgh, Pennsylvania.
- Chklovski, T., Mihalcea, R., Pedersen, T. and Purandare, A. (2004). *The SENSEVAL-3 Multilingual English-Hindi Lexical Sample Task*. Proceedings of the third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text (SENSEVAL-3), Barcelona, Spain.
- Church, K. W. and Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1), 22-29.
- Dagan, I. and Itai, A. (1994). Word Sense Disambiguation Using a Second Language Monolingual Corpus. *Computational Linguistics*, 20(4), 563-596.

- Gao, J., Zhou, M., Nie, J.-Y., He, H. and Chen, W. (2002). *Resolving Query Translation Ambiguity using a Decaying Co-occurrence Model and Syntactic Dependence Relations*. Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2002), Tampere, Finland.
- Ikeda, K., Yanagihara, T., Matsumoto, K. and Takishima, Y. (2009). *Unsupervised Text Normalization Approach for Morphological Analysis of Blog Documents*. Proceedings of the 22nd Australasian Joint Conference on Advances in Artificial Intelligence, Melbourne, Australia.
- Jin, P., Wu, Y. and Yu, S. (2007). *SemEval-2007 Task 5: Multilingual Chinese-English Lexical Sample*. Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval 2007), Prague, Czech Republic.
- Kudo, T. (2005). MeCab: Yet Another Part-of-Speech and Morphological Analyzer. Retrieved May, 2010, from <http://mecab.sourceforge.net/>
- Kudo, T. and Matsumoto, Y. (2002). *Japanese Dependency Analysis using Cascaded Chunking*. Proceedings of the 6th Conference on Natural Language Learning (CoNLL 2002), Taipei, Taiwan.
- Lefever, E. and Hoste, V. (2010). *SemEval-2010 Task 3: Cross-lingual Word Sense Disambiguation*. Proceedings of the 5th International Workshop on Semantic Evaluation (SemEval 2010), Los Angeles, California.
- Liu, P., Xue, Y., Li, S. and Liu, S. (2010). *Minimum Normalized Google Distance for Unsupervised Multilingual Chinese-English Word Sense Disambiguation*. Proceedings of the 4th International Conference on Genetic and Evolutionary Computing (ICGEC 2010), Shenzhen, China.
- Liu, P. and Zhao, T. (2009). *Unsupervised Translation Disambiguation Based on Maximum Web Bilingual Relatedness: Web as Lexicon*. Proceedings of the 6th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2009), Tianjin, China.
- Marsi, E., Lynum, A., Bungum, L. and Gambäck, B. (2011). *Word Translation Disambiguation without Parallel Texts*. Proceedings of the International Workshop on Using Linguistic Information for Hybrid Machine Translation (LIHMT 2011), Barcelona, Spain.
- Ng, H. T., Wang, B. and Chan, Y. S. (2003). *Exploiting Parallel Texts for Word Sense Disambiguation: An Empirical Study*. Proceedings of the 41st Annual Meeting on Association for Computational Linguistics (ACL 2003), Sapporo, Japan.
- Tsunakawa, T. and Kaji, H. (2010). *Augmenting a Bilingual Lexicon with Information for Word Translation Disambiguation*. Proceedings of the 8th Workshop on Asian Language Resources (ALR 2010), Beijing, China.