Adjusting Occurrence Probabilities of Automatically-Generated Abbreviated Words in Spoken Dialogue Systems

Masaki Katsumaru, Kazunori Komatani, Tetsuya Ogata, and Hiroshi G. Okuno

Graduate School of Informatics, Kyoto University, Kyoto, Japan
{katsumaru, komatani, ogata, okuno}@kuis.kyoto-u.ac.jp

Abstract. Users often abbreviate long words when using spoken dialogue systems, which results in automatic speech recognition (ASR) errors. We define *abbreviated words* as sub-words of an original word and add them to the ASR dictionary. The first problem we face is that proper nouns cannot be correctly segmented by general morphological analyzers, although long and compound words need to be segmented in agglutinative languages such as Japanese. The second is that, as vocabulary size increases, adding many abbreviated words degrades the ASR accuracy. We have developed two methods, (1) to segment words by using conjunction probabilities between characters, and (2) to adjust occurrence probabilities of generated abbreviated words on the basis of the following two cues: phonological similarities between the abbreviated and original words and frequencies of abbreviated words in Web documents. Our method improves ASR accuracy by 34.9 points for utterances containing abbreviated words without degrading the accuracy for utterances containing original words.

Index Terms: Spoken dialogue systems, abbreviated words, adjusting occurrence probabilities.

1 Introduction

Users often omit parts of long words and utter abbreviated words [1]. For example, the abbreviated word *aoyamakan* (青山館), meaning Aoyama Hall, is said to indicate *aoyamaongakukinenkan* (青山音楽記念館), meaning Aoyama Memorial Hall of Music. They are apt to do this because users unfamiliar with a particular spoken dialogue system do not know much about how to use it and what content words are included in its vocabulary. In conventional system development, system developers manually add unknown words to an automatic speech recognition (ASR) dictionary by collecting and examining misrecognized words uttered by users. This manual maintenance requires a great deal of time and effort. Furthermore, a system cannot recognize these words until the manual maintenance has taken place. They continue to be misrecognized until the system developers find and add them to the system dictionary.

Our purpose is to automatically add abbreviated words users may utter at the initial time when an original dictionary in any domain has been provided. We define an *original dictionary* as the initial ASR dictionary for a system, *original words* as content words in an original dictionary, and *abbreviated words* as words that are sub-words of an original word and that indicate the same entity as the original word. We generate abbreviated words by omitting arbitrary sub-words of an original word. These abbreviated words are interpreted as their corresponding original words in a language understanding module. Automatic addition of vocabulary at the initial stage of system development alleviates manual maintenance time and effort. Furthermore, the system can recognize abbreviated words at an earlier stage, thus increasing its usability.

There are two problems when abbreviated words are added to an ASR dictionary.

1. **Segmenting proper nouns in order to generate abbreviated words**

Proper nouns cannot be correctly segmented by general morphological analyzers because they are domain-dependent words, such as regional names. To decide which sub-words to omit, proper nouns need to be segmented in agglutinative languages such as Japanese, while words in an isolating language such as English do not pose this problem.

2. **Reducing ASR errors caused by adding abbreviated words to an ASR dictionary**

ASR accuracy is often degraded by adding generated abbreviated words because the vocabulary size increases. Jan et al. merely added generated abbreviated words and did not take the degradation into account [2]. The following words tend to degrade ASR accuracy:

- (a) abbreviated words with phonemes close to those of other original words
- (b) abbreviated words that are not actually used

For the former, we segmented proper nouns by using conjunction probabilities between characters in addition to results of a morphological analyzer. For the latter, we manipulated occurrence probabilities of generated abbreviated words on the basis of phonological similarities between the abbreviated and original words [3]. We furthermore introduce a measure, *Web frequency*, for representing how much each generated abbreviated word is actually used. This measure is defined by using Web search results, and suppresses side effects caused by abbreviated words that are not used. These enable us to add abbreviated words to an ASR dictionary without increasing the ASR error rate.

2 Case Study of Deployed System

We preliminarily investigated gaps between users' utterances and the vocabulary of a system by analyzing words added by developers during the 5-year service of the Kyoto City Bus Information System [4]. Users stated their boarding stop as well as the destination or the bus route number by telephone, and the system informed them how long it would be before the bus arrived. There were 15,290 calls

to the system during the 58 months between May 2002 and February 2007, and the system developers added users' words that the system could not recognize¹.

The developers added 309 words to the system's vocabulary. Of these 91.6% were aliases for already known entities, while 8.4% were new entities of bus stops and landmarks. There were far fewer new entities added than aliases for the already known entities. This means that the developers had carefully prepared the vocabulary for bus stops and landmarks at the initial stage of system development. The reason the added words consisted almost exclusively of aliases is that, at the initial stage of system development, the system developers were unable to predict the wide range of other expressions that would be uttered by real users. Abbreviated words were the majority of the added aliases, which were 78.3% of all added words. This means that real users actually often utter abbreviated words. Of the 1,494 utterances collected from novices using the system, 150 utterances contained abbreviated words.

3 Generating and Manipulating Occurrence Probabilities of Abbreviated Words

The flow of our method for adding abbreviated words is shown in Figure 1. First, original words are segmented to identify sub-words to omit. For domain-dependent proper nouns, a *conjunction probability* is defined between each character as a measure of segmenting compound words. As described in section 3.1, proper nouns are segmented by using conjunction probabilities and a morphological analyzer. Abbreviated words are then generated by omitting some sub-words of the segmented words.

In section 3.2, we address how to suppress ASR errors caused by adding generated abbreviated words. We define the phonological similarities between the abbreviated and original words, and the Web frequencies of the abbreviated words. Then occurrence probabilities are manipulated on the basis of them.

3.1 Segmenting Words in ASR Dictionary and Generating Abbreviated Words

In our method, a compound word in the ASR dictionary is first segmented into a sub-word array, " $s_1 s_2 \dots s_n$ ". The segmentation is done at a part where either a morphological analyzer or conjunction probabilities would segment it. The morphological analyzer we use is MeCab [5]. Domain-dependent proper nouns are segmented by using conjunction probabilities between characters as follows. If a word in the ASR dictionary is expressed by the character string " $c_1 c_2 \dots c_{i-1} c_i \dots c_n$ ", a conjunction probability between c_{i-1} and c_i is formulated on the basis of the character N-gram probabilities in the ASR dictionary:

$$\min\{P(c_i | c_{i-1} c_{i-2} \dots c_1), P(c_{i-1} | c_i c_{i+1} \dots c_n)\}. \quad (1)$$

¹ The developers did not add all words users uttered during this period. Short words were not added because they could cause insertion errors. This was because the system's dialogue management is executed in a mixed-initiated manner, and its language constraint is not so strong.

Fig. 1. Flow of adding abbreviated words

This means that a conjunction probability is defined as smaller one of N-gram probabilities forward to c_i and backward to c_{i-1} . A word is segmented between c_{i-1} and c_i if the conjunction probability between them is lower than threshold θ . For example, the proper noun *shisekikoenmae* (史跡公園前), which means “in front of the Historical Park”, is segmented as shown in Figure 2. Using conjunction probabilities segments *shisekikoenmae* (史跡公園前) into *shisekikoen* (史跡公園) and *mae* (前), while using MeCab cannot. This segmentation is essential to generating various abbreviated words such as *shisekikoen* (史跡公園).

Next, an arbitrary number of sub-words are omitted and $(2^n - 1)$ abbreviated words from a sub-word array “ $s_1 s_2 \dots s_n$ ” are generated. The pronunciations of the generated abbreviated words are given by the pronunciations of the sub-words, which are detected by matching the pronunciation given by MeCab and the original pronunciation.

3.2 Reducing ASR Errors Caused by Adding Generated Abbreviated Words

Definition of Phonological Similarity. We define phonological similarity as a measure of confusion in ASR that is caused by generated abbreviated words. These words may cause ASR errors for utterances containing original words when the phonemes of the generated words are close to those of the original words or those of parts of the original words. We define the phonological similarity between generated abbreviated word w and vocabulary D_{org} of the original dictionary as

$$dist(w, D_{org}) = \min(e.d.(w, part(D'_{org}))). \quad (2)$$

Fig. 2. Segmentation result of “shisekikoenmae”

We denote D'_{org} as the vocabulary made by removing from D_{org} words from which w is generated. Partial sequences of all words of D'_{org} are given by $part(D'_{org})$. The edit distance between x 's and y 's phoneme strings is $e.d.(x, y)$; it is calculated by DP matching [6]. If we define S_1 as a phoneme set of vowels, a moraic obstruent and a moraic nasal, and S_2 as a phoneme set of consonants, we set costs of edit distance 2 when an element of S_1 is inserted, deleted, or substituted, and 1 when an element of S_2 is inserted, deleted, or substituted with one of S_2 .

Definition of Web Frequency. We define the Web frequency of a generated abbreviated word as the frequency it appears in Web documents. The Web frequencies of words indicate how often they are actually used. The frequency is obtained by performing a query on a Web search engine. We used Yahoo! Japan². We define the Web frequency of a generated abbreviated word as

$$WebFrequency(w) = \frac{count(w)}{count(original(w))}, \quad (3)$$

in which $count(\text{“word”})$ is the hit count of Web pages for query “word”, and $original(w)$ is the original word from which w was generated. We normalize $count(w)$ by $count(original(w))$ to give proper measures because $count(w)$ tends to be small (or high) when $count(original(w))$ is small (or high). The lower the Web frequency, the less frequently users may utter.

We generated abbreviated words from the vocabulary of the Kyoto City Bus Information System. The phonological similarities and Web frequencies of some of these words are shown in Table 1. The phonological similarity between the generated abbreviated word *horikawashi* and the vocabulary of the original dictionary is 0 because *horikawashi* is equal to part of *horikawashimodachiuri*. The similarity of *rokuhamitsuji* is same as that of *paresusaido*. However we should set the probability of *rokuhamitsuji* differently from that of *paresusaido* because *rokuhamitsuji* is generated by a segmentation error and not actually used, but *paresusaido* is used. The Web frequency of *rokuhamitsuji* is much lower than that

² <http://search.yahoo.co.jp>

Table 1. Phonological similarities and Web frequencies of generated abbreviated words

Abbreviated word (its original word)	P.S.	Closest original word	W.F.
horikawashi (horikawatakoyakushi) [Name of Area]	0	horikawashimodachiuri [Name of Area]	$5.77 * 10^{-2}$
shakadani (shakadaniguchi) [Name of Area]	2	haradani [Name of Area]	$2.75 * 10$
rokuhamitsuji (rokuharamitsuji) [Name of Temple]	6	kokuritsukindaibijutsukan [Name of Museum]	0.00
paresusaido (kyotoparesusaidohoteru) [Name of Hotel]	6	karasumashimochojamachi [Name of Town]	$3.86 * 10^2$

P.S.: phonological similarity
W.F.: Web frequency

of *paresusaido*. We can thus distinguish between *rokuhamitsuji* and *paresusaido* by considering Web frequency.

Manipulating Occurrence Probabilities on the basis of Phonological Similarity and Web Frequency. Degradation of ASR accuracy for utterances containing original words is avoided by manipulating the occurrence probabilities of the generated abbreviated words on the basis of their Web frequencies in addition to their phonological similarities. We define $P_{org}(w)$ as the occurrence probability of word w . The probabilities of the generated abbreviated words that meet two conditions,

$$dist(w, D_{org}) \leq d \quad (4)$$

$$WebFrequency(w) \leq e \quad (5)$$

(d, e : threshold)

are arranged as new occurrence probabilities:

$$P_{new}(w) = P_{org}(w) * \alpha^{dist(w, D_{org}) - d - 1} * WebFrequency(w). \quad (6)$$

Generated abbreviated words that meet only (4) are arranged as

$$P_{new}(w) = P_{org}(w) * \alpha^{dist(w, D_{org}) - d - 1}, \quad (7)$$

and those that meet only (5) are arranged as

$$P_{new}(w) = P_{org}(w) * WebFrequency(w). \quad (8)$$

We set α to 10. The lower the phonological similarity and Web frequency, the lower the occurrence probability. Generated abbreviated words with a Web frequency of 0 are removed from the ASR dictionary. $P_{new}(w)$ is calculated for all generated abbreviated words. We then normalize the probabilities of the original and generated abbreviated words to meet $\sum_{word \in W} P(word) = 1$, in which W is a set of all the original and abbreviated words.

4 Experimental Evaluation

We experimentally evaluated our method. We generated abbreviated words from a system’s ASR dictionary and added them to the dictionary. The metrics were the recall rate and the ASR accuracy for the collected utterances. The ASR accuracy is calculated as $(Cor - Ins) / Len * 100$ [%], in which *Cor*, *Ins*, and *Len* are the number of correction, insertion, and words in the manual transcription. To verify whether our method is independent of a particular domain, we also generated abbreviated words in another domain.

4.1 Target Data for Evaluation

We used real users’ utterances collected on the Kyoto City Bus Information System. We targeted users who were not familiar with the system’s vocabulary and collected utterances of users who were using the system for the first time by analyzing their telephone numbers. We collected 1,494 utterances by 183 users after removing utterances that were not relevant to the task. Of the 1,494 utterances, 150 contained 70 kinds of abbreviated words, and 1,142 contained only original words. The other 202 consisted of words that were neither abbreviated nor original such as, “Can this system tell me where to change buses?”.

4.2 Recall Rate of Generated Abbreviated Words

We generated abbreviated words from 1,481 words (bus stops and landmarks) of the 1,668 original words in the Kyoto Bus Information System dictionary. The threshold θ of segmentation by conjunction probabilities was set at 0.12 after preliminary experiments. We segmented the words by using both conjunction probabilities and MeCab, omitted sub-words, and generated 11,936 abbreviated words. To prove the effectiveness of our segmentation, we also generated 2,619 abbreviated words by segmentation using only conjunction probabilities and 8,941 using only MeCab. We evaluated three methods of segmentation:

- conjunction probabilities only
- MeCab (morphological analyzer) only
- both conjunction probabilities and MeCab (our method).

The recall rates for abbreviated words generated by each method are shown in Table 2. For 70 different abbreviated words uttered by real users in the collected data, our method generated 66 (94%) while 51 (73%) were generated by only conjunction probabilities and 60 (86%) by using only MeCab. The recall rate with our method was 8 points higher than that with only MeCab. Using conjunction probabilities led to this improvement.

4.3 Evaluation of ASR Accuracy

We constructed a statistical language model in order to manipulate the occurrence probabilities for each word because the Kyoto City Bus Information System’s ASR is a grammar-based one. First, content words were assigned to

Table 2. Recall rate for each segmentation method

Method of segmentation	Number of generated abbreviated words	Recall rate [%]
conjunction probabilities only	2,619	73
MeCab (morphological analyzer) only	8,941	86
MeCab + conjunction probabilities (our method)	11,936	94

the class of bus stops, landmarks, and bus route numbers. Next, we constructed a class N-gram model from all kinds of sentences that the grammar-based language model generated. We used a CMU Toolkit [7] to construct the statistical language model. We added the abbreviated words generated to the class of bus stops and landmarks in addition to original bus stops or landmarks. The acoustic model was a triphone model with 2,000 states and 16 mixture components for telephone speech. The ASR engine was Julius [8]. We set d to 5 and e to 400,000 by trial and error. The following are the experimental conditions:

Cond. 1: original dictionary (baseline)

Use the system's original ASR dictionary before adding abbreviated words (vocabulary size: 1,668)

Cond. 2: Cond. 1 + generated abbreviated words

Add generated abbreviated words to original dictionary (13,604)

Cond. 3: Cond. 2 + manipulating occurrence probabilities on the basis of only phonological similarity

Add generated abbreviated words to original dictionary and manipulate occurrence probabilities on the basis of only phonological similarity (13,604)

Cond. 4: Cond. 2 + manipulating occurrence probabilities on the basis of only Web frequency

Add generated abbreviated words to the original dictionary and manipulate occurrence probabilities on the basis of only Web frequency (7,203)

Cond. 5: Cond. 2 + manipulating occurrence probabilities on the basis of both phonological similarity and Web frequency (our method)

Add generated abbreviated words to the original dictionary and manipulate occurrence probabilities on the basis of both phonological similarity and Web frequency (7,203)

Table 3 shows the ASR accuracy of content words for 150 utterances with abbreviated words, 1,142 utterances with only original words, and all 1,494 utterances.

Comparing Cond. 1 and 2, the ASR accuracy for all utterances in Cond. 2 degraded by 12.3 points although that for utterances with abbreviated words in Cond. 2 improved by 23.6 points. This is because we only added abbreviated words generated by our method. This result shows that ASR accuracy degrades by merely adding these words. In Cond. 3, the ASR accuracy for utterances with original words improved by 15.0 points compared with Cond. 2 and degraded only 0.1 points compared with Cond. 1. This came from manipulating the probabilities of generated abbreviated words on the basis of phonological similarities. This result shows that phonological similarity based manipulation reduces ASR

Table 3. ASR accuracy [%] for content words of utterances for each condition

Condition	Utterances with abbreviated words	Utterances with original words	All utterances
1: original dictionary (baseline)	1.1	74.9	52.5
2: 1 + generated abbr.	24.7	59.8	40.2
3: 2 + (a) phonological similarity	25.3	74.8	56.5
4: 2 + (b) Web frequency	25.3	44.5	27.6
5: 2 + (a) + (b) (our method)	36.0	76.0	58.9

errors caused by adding abbreviated words. Comparing Cond. 2 and 4, the ASR accuracy for utterances with abbreviated words in Cond. 4 was slightly higher. This is because we arranged probabilities on the basis of Web frequency. This indicates that Web frequency based manipulation of the probabilities suppresses ASR errors caused by generated abbreviated words not actually used. However the ASR accuracy for all utterances in Cond. 4 was degraded compared with the accuracy in Cond. 2. This was because high occurrence probabilities were given to short words that frequently appeared in the Web, and accordingly insertion errors increased. In Cond. 5, the ASR accuracy for utterances with abbreviated words increased by 34.9 points compared with that in Cond. 1, and increased by 10.7 points compared with that in Cond. 3 or 4. The ASR accuracy for utterances with original words in Cond. 5 did not degrade when compared with Cond. 1. This was because we used both phonological similarity and Web frequency to adjust occurrence probabilities. These results demonstrate the effectiveness of our method of manipulating occurrence probabilities on the basis of both phonological similarities and Web frequencies for reducing ASR errors caused by adding abbreviated words.

The ASR accuracy is still low throughout the experiment. A reason for this low level of accuracy is a mismatch between the acoustic model and the users' circumstances. Actually, there were several cases in which acoustic scores for correct word sequences were lower than those for others. We have addressed how to improve the language model. Improving the acoustic model will lead to a higher level of ASR accuracy.

4.4 Generating Abbreviated Words in Another Domain

We also generated abbreviated words for the restaurant domain to verify whether our method is independent of a particular domain. We check only generated abbreviated words because we have no dialogue data in this domain and cannot evaluate ASR accuracy. In this domain as well, domain-dependent proper nouns were correctly segmented by using conjunction probabilities, and several appropriate abbreviated words were generated although the morphological analyzer could not segment some of them. For example, our method could segment *bisutorokyatorudoru* into *bisutoro* (*bistro*) and *kyatorudoru* (*name of restaurant*) by detecting the high frequency of *bisutoro* in the dictionary, although MeCab could not segment it. This segmentation enabled us to generate the abbreviated word *kyatorudoru*, which is often used.

5 Conclusion

We generated abbreviated words and added them to an ASR dictionary to enable a dialogue system to recognize abbreviated words uttered by users. To increase the recall rate of the generated abbreviated words, we segment proper nouns by introducing conjunction probabilities between characters in the system's dictionary. To add abbreviated words without increasing the ASR error rate, we manipulate their occurrence probabilities on the basis of their Web frequency (the frequency of their use in Web documents) in addition to the phonological similarity between the abbreviated and original words.

Experimental evaluations using real users' utterances demonstrated that our method is effective. The recall rate was higher than that using only a morphological analyzer. The ASR accuracy for utterances with abbreviated words was 34.9 points higher than that when only the original dictionary was used without degrading the accuracy for utterances with original words. These results show that our method for vocabulary expansion enables a dialogue system to recognize user's abbreviated words without increasing the ASR error rate. Future work includes collecting utterances in another domain and using them to evaluate our method.

Acknowledgments. We are grateful to Dr. Shun Shiramatsu of Kyoto University for allowing us to use the Web page counting program he developed.

References

1. Zweig, G., Nguyen, P., Ju, Y., Wang, Y., Yu, D., Acero, A.: The Voice-Rate Dialog System for Consumer Ratings. In: Proc. Interspeech, pp. 2713–2716 (2007)
2. Jan, E.E., Maison, B., Mangu, L., Zweig, G.: Automatic Construction of Unique Signatures and Confusable Sets for Natural Language Directory Assistance Applications. In: Proc. Eurospeech, pp. 1249–1252 (2003)
3. Katsumaru, M., Komatani, K., Ogata, T., Okuno, H.G.: Expanding Vocabulary for Recognizing User's Abbreviations of Proper Nouns without Increasing ASR Error Rates in Spoken Dialogue Systems. In: Proc. Interspeech, pp. 187–190 (2008)
4. Komatani, K., Ueno, S., Kawahara, T., Okuno, H.G.: User Modeling in Spoken Dialogue Systems for Flexible Guidance Generation. In: Proc. Eurospeech, pp. 745–748 (2003)
5. Kudo, T., Yamamoto, K., Matsumoto, Y.: Applying conditional random fields to Japanese morphological analysis. In: Proc. EMNLP, pp. 230–237 (2004), <http://mecab.sourceforge.net/>
6. Navarro, G.: A Guided Tour to Approximate String Matching. ACM Computing Surveys 33(1), 31–88 (2001)
7. Clarkson, P.R., Rosenfeld, R.: Statistical Language Modeling Using the CMU-Cambridge Toolkit. In: Proc. ESCA Eurospeech, pp. 2707–2710 (1997), <http://svr-www.eng.cam.ac.uk/~prc14/toolkit.html>
8. Kawahara, T., Lee, A., Takeda, K., Itou, K., Shikano, K.: Recent progress of open-source LVCSR Engine Julius and Japanese model repository. In: Proc. ICSLP, pp. 3069–3072 (2004)