A Context-Sensitive Homograph Disambiguation in Thai Text-to-Speech Synthesis

Virongrong Tesprasit, Paisarn Charoenpornsawat and Virach Sornlertlamvanich
Information Research and Development Division
National Electronics and Computer Technology Center
112 Phahon Yohtin, Rd., Klong 1, Klong Luang, Pathumthani 12120 THAILAND
{virong, paisarn, virach}@nectec.or.th

Abstract

Homograph ambiguity is an original issue in Text-to-Speech (TTS). To disambiguate homograph, several efficient approaches have been proposed such as part-of-speech (POS) n-gram, Bayesian classifier, decision tree, and Bayesian-hybrid approaches. These methods need words or/and POS tags surrounding the question homographs in disambiguation. Some languages such as Thai, Chinese, and Japanese have no word-boundary delimiter. Therefore before solving homograph ambiguity, we need to identify word boundaries. In this paper, we propose a unique framework that solves both word segmentation and homograph ambiguity problems altogether. Our model employs both local and long-distance contexts, which are automatically extracted by a machine learning technique called Winnow.

1 Introduction

In traditional Thai TTS, it consists of four main modules: word segmentation, grapheme-to-phoneme, prosody generation, and speech signal processing. The accuracy of pronunciation in Thai TTS mainly depends on accuracies of two modules: word segmentation, and grapheme-to-phoneme. In word segmentation process, if word boundaries cannot be identified correctly, it leads Thai TTS to the incorrect pronunciation such as a string "ตากลม" which can be separated into two different ways with different meanings and pronunciations. The first one is "ตา(eye) กลม(round)", pronounced [ta:0 klom0] and the other one is "ตาก(expose) ลม(wind)", pronounced [ta:k1 lom0]. In grapheme-to-phoneme module, it may produce error pronunciations for a homograph which can be pronounced more than one way such as a word "เพลา" which can be pronounced [phlaw0] or [phe:0 la:0]. Therefore, to improve an accuracy of Thai TTS, we have to focus on solving the problems of word boundary ambiguity and homograph ambiguity which can be viewed as a disambiguation task.

A number of feature-based methods have been tried for several disambiguation tasks in NLP, including decision lists, Bayesian hybrids, and Winnow. These methods are superior to the previously proposed methods in that they can combine evidence from various sources in disambiguation. To apply the methods in our task, we treat problems of word boundary and homograph ambiguity as a task of word pronunciation disambiguation. This task is to decide using the context which was actually intended. Instead of using only one type of syntactic evidence as in N-gram approaches, we employ the synergy of several types of features. Following previous works [4, 6], we adopted two types of features: context words, and collections. Context-word feature is used to test for the presence of a particular word within +/- K words of the target word and collocation test for a pattern of up to L contiguous words and/or part-of-speech tags surrounding the target word. To automatically extract the discriminative features from feature space and to combine them in disambiguation, we have to investigate an efficient technique in our task.

The problem becomes how to select and combine various kinds of features. Yarowsky [11] proposed decision list as a way to pool several types of features, and to solve the target problem by applying a single strongest feature, whatever type it is. Golding [3] proposed a Bayesian hybrid method to take into account all available evidence, instead of only the strongest one. The method was applied to the task of context-sentitive spelling correction and was reported to be superior to decision lists. Later, Golding and Roth [4] applied Winnow algorithm in the same task and found that the algorithm performs comparably to the Bayesian hybrid method when using pruned feature sets, and is better when using unpruned sets or unfamiliar test set.

In this paper, we propose a unified framework in solving the problems of word boundary ambiguity and homograph ambiguity altogether. Our approach employs both local and long-distance contexts, which can be automatically extracted by a machine learning technique. In this task, we employ the machine learning technique called Winnow. We then construct our system based on the algorithm and evaluate them by comparing with other existing approaches to Thai homograph problems.

2 Problem Description

In Thai TTS, there are two major types of text ambiguities which lead to incorrect pronunciation, namely word boundary ambiguity and homograph ambiguity.

2.1 Word Boundary Ambiguity (WBA)

Thai as well as some other Asian languages has no word boundary delimiter. Identifying word boundary, especially in Thai, is a fundamental task in Natural Language Processing (NLP). However, it is not a simple problem because many strings can be segmented into words in different ways. Word boundary ambiguities for Thai can be classified into two main categories defined by [6]: Context Dependent Segmentation Ambiguity (CDSA), and Context Independent Segmentation Ambiguity (CISA).

CISA can be almost resolved deterministically by the text itself. There is no need to consult any context. Though there are many possible segmentations, there is only one plausible segmentation while other alternatives are very unlikely to occur, for example, a string "ไปหามเหสี" which can be segmented into two different ways: "ไป(go) หาม(carry) เห(deviate) สี(color)" [paj0 ha:m4 he:4 si:4] and "ไป(go) หา(see) มเหสี(queen)" [paj0 ha:4 ma:3 he:4 si:4]. Only the second choice is plausible. One may say that it is not semantically ambiguous. However, simple algorithms such as maximal matching [6, 9] and longest matching [6] may not be able to discriminate this kind of ambiguity. Probabilistic word segmentation can handle this kind of ambiguity successfully.

CDSA needs surrounding context to decide which segmentation is the most probable one. Though the number of possible alternatives occurs less than the context independent one, it is more difficult to disambiguate and causes more errors. For example, a string "ตากลม" can be segmented into "ตา กลม" (round eye) and "ตาก ลม" (to expose wind) which can be pronounced [ta:0 klom0] and [ta:k1 lom0] respectively.

2.2 Homograph Ambiguity

Thai homographs, which cannot be determined the correct pronunciation without context, can be classified into six main categories as follows:

1. Number such as 10400 in postcode, it can be pronounced [nvng1 su:n4 si:1 su:n4 su:n4] or [nvng1 mv:n1 si:1 r@:ji3] in amount.

2. Abbreviation such as ก.พ. can be pronounced [sam4 nak2 nga:n0 kha:2 ra:t2 cha:3 ka:n0 phon0 la:3 rv:an0] (Office Of The Civil Service Commission) or [kum0 pha:0 phan0] (February).

3. Fraction such as 25/2 can be pronounced [yi:2 sip1 ha:2 thap3 s@:ng4] (for address) or [yi:2 sip1 ha:2 su:an1 s@:ng4] (for fraction).

4. Proper Name such as "สมพล" is pronounced [som4 phon0] or [sa1 ma3 phon0].

5. Same Part of Speech such as "เพลา" (time) can be pronounced [phe:0 la:0], while "เพลา" (axe) is pronounced [phlaw0].

6. Different Part of Speech such as "แหน" is pronounced [nx:4] or [hx:n4].

3 Previous Approaches

POS n-gram approaches [7, 10] use statistics of POS bigram or trigram to solve the problem. They can solve only the homograph problem that has different POS tag. They cannot capture long distance word associations. Thus, they are inappropriate of resolving the cases of semantic ambiguities.

Bayesian classifiers [8] use long distance word associations regardless of position in resolving semantic ambiguity. These methods can successful capture long distance word association, but cannot capture local context information and sentence structure.

Decision trees [2] can handle complex condition, but they have a limitation in consuming very large parameter spaces and they solve a target problem by applying only the single strongest feature.

Hybrid approach [3, 12] combines the strengths of other techniques such as Bayesian classifier, n-gram, and decision list. It can be capture both local and long distance context in disambiguation task.

4 Our Model

To solve both word boundary ambiguity and homograph ambiguity, we treat these problems as the problem of disambiguating pronunciation. We construct a *confusion set* by listing all of its possible pronunciations. For example, C = {[ma:0 kwa:1], [ma:k2 wa:2]} is the confusion set of the string "มากกว่า" which is a boundary-ambiguity string and C={[phe:0 la:0] ,[phlaw0]} is the confusion set of the homograph "เพลา". We obtain the features that can discriminate each pronunciation in the set by Winnow based on our training set.

4.1 Winnow

Winnow algorithm used in our experiment is the algorithm described in [1]. Winnow is a neuron-like network where several nodes are connected to a target node [4, 5]. Each node called *specialist* looks at a particular value of an attribute of the target concept, and will vote for a value of the target concept based on its specialty; i.e. based on a value of the attribute it examines. The global algorithm will then decide on weighted-majority votes receiving from those specialists. The pair of (attribute=value) that a specialist examines is a candidate of features we are trying to extract. The global algorithm updates the weight of any specialist based on the vote of that specialist. The weight of any specialist is initialized to 1. In case that the global algorithm predicts incorrectly, the weight of the specialist that predicts incorrectly is halved and the weight of the specialist that predicts correctly is multiplied by 3/2. The weight of a specialist is halved when it makes a mistake even if the global algorithm predicts correctly.

4.2 Features

To train the algorithm to resolve pronunciation ambiguity, the context around a homograph or a boundary-ambiguity string is used to form features. The features are the context words, and collocations. Context words are used to test for the presence of a particular word within +10 words and –10 words from the target word. Collocations are patterns of up to 2 contiguous words and part-of-speech tags around the target word. Therefore, the total number of features is 10; 2 features for context words, and 8 features for collocations.

5 Preliminary Experiment

To test the performance of the different approaches, we select sentences containing Thai homographs and boundary ambiguity strings from our 25K-words corpus to use in benchmark tests. Every sentence is manually separated into words. Their parts of speech and pronunciations are manually tagged by linguists. The resulting corpus is divided into two parts; the first part, about 80% of corpus, is utilized for training and the rest is used for testing.

In the experiment, we classify the data into three group depending on types of text ambiguity according to section 2: CDSA, CISA and Homograph, and compare the results from different approaches; Winnow, Bayseian hybrid [3] and POS trigram. The results are shown in Table 1.

	Trigram	Bayseian	Winnow
CDSA	73.02%	93.18%	95.67%
CISA	98.25%	99.67%	99.70%
Homograph	52.46%	94.25%	96.45%

Table1: The result of comparing different approaches

6 Conclusion

In this paper, we have successfully applied Winnow to the task of Thai homograph disambiguation. Winnow shown its ability to construct networks that extract the features in data effectively. The learned features, which are context words and collocations, can capture useful information and make the task of Thai homograph disambiguity more accurate. The experimental results show that Winnow outperform trigram model and Bayesian hybrid. Our future works will investigate other machine learning techniques such as SNoW and SVM.

References

[1] Blum, A. Empirial Support for Winnow and Weighted-Majority Algorithm: Results on a Calendar Scheduling Domain, Machine Learning. 1997, 26:5-23.

[2] Brieman, L. et al. Classification and Regression Trees. Wadsworth & Brooks, Monterrey CA.1984.

[3] Golding, A. R. A Bayesian Hybrid Mehod for Context-sentitive Spelling Correction. *In Proceedings of the Third Workshop on Very Large Corpora.* 1995.

[4] Golding, A. R. & Roth, D. Applying Winnow to Context-Sensitive Spelling Correction. In Lorenza Saitta, editor, Machine Learning*: Proceedings of the 13th International Conference on Machine Learning*. 1996.

[5] Littlestone, N. Learning Quickly when Irrelevant Attributes Bound: A New Linear-Threshold Algorithm. Machine Learning. 1988, 2:285-318.

[6] Meknavin, S., Charoenpornsawat P. and Kijsirikul, B. Feature-based Thai Word Segmentation. Proceeding of of the Natural Language Processing Pacific Rim Symposium. 1997.

[7] Merialdo, B. Tagging text with a probabilistic model. In Proceedings of the IBM Natural Language ITL, Paris, France. 1990.

[8] Mosteller, F. and Wallace, D. Inference and Disputed Authorship: The Federalist Addision-Wesley, Reading, Massachusetts. 1964.

[9] Sornlertlamvanich, V. Word Segmentation for Thai. *Machine Translation System*. National Electronics and Computer Technology Center (in Thai). 1993.

[10] Sproat, R.,Hirschberg, J. and Yarowsky, D. A corpus-based synthesizer. In Proceedings, International Conference on Spoken Language Processing, Banff. 1992

[11] Yarowsky, D. Decision Lists for Lexical Ambiguity Resolution. In Proceeding of 32nd Annual Meeting of the Association for Computational Linguistics. 1994

[12] Yarowsky, D. Homograph Disambiguation in Speech Synthesis. In J. van Santen, R. Sproat, J. Olive and J. Hirschberg (eds.), Progress in Speech Synthesis. Springer-Verlag, pp. 159-175, 1996.