CONTEXT DEPENDENT ANTI SUBWORD MODELING FOR UTTERANCE VERIFICATION

Padma Ramesh, Chin-Hui Lee, and Biing-Hwang Juang

Multimedia Communications Research Lab.
Bell Labs, Lucent Technologies, Murray Hill, NJ 07974-0636, USA

ABSTRACT

Utterance verification is used in spoken language dialog systems to reject the speech that does not belong to the task and to correctly recognize the sentences that do. Current verification systems use context dependent (CD) or context independent (CI) subword models and CI anti-subword models. We propose many methods of modeling the CD anti-subword models. We have compared these anti-models and show that the anti-models with the same context have the most separation between the speech that contains the subword and the speech that does not contain the subword. We have also conducted recognition/verification experiments with a two pass verifier and two one pass verification systems to compare the different types of anti-subword models. Our results show that the same context anti-subword models have the best recognition/verification performance.

1. INTRODUCTION

The performance of speech recognition systems have been improving steadily in the last decade. Accuracy of these systems has reached an acceptable range, making possible many applications, such as, operator services in telephony, database access in credit card information services, call centers and so on. These systems have perform reasonably well when the user speaks the words in the system's vocabulary, following the grammatical syntax implemented by the system. However, when the speech contains out of vocabulary words or does not conform to the task gram mar, there is usually a substantial performance degradation. As speech recognition systems proliferate reaching many users, they often have to deal with natural spontaneous speech. Users not familiar with the system may respond with unconstrained speech, hesitations, filler words, and out of vocabulary words. They may even respond with speech that does not belong to the task at hand (out of task sentences). A spoken dialog system has to handle gracefully all such speech input. In [1] Kawahara, Kitaoka and Doshita have explored one such system. Such a system has to correctly recognize the spoken sentences that conform to the implemented grammar, containing only the vocabulary words (in-grammar). This has to also correctly recognize the words even when the input sentences do not follow the grammar constraints and/or have other extraneous words, such as fillers, hesitations, etc. (out-of-grammar). Further, it has to reject the speech that does not belong to the task (out-of-task) [2]. In order to reject the out-of-task sentences, or recognize correctly when out-of-grammar sentences are spoken, utterance verification is used. This is a process that verifies that the recognized words are actually in the input speech. A two pass classifier for utterance rejection using a keyword/non-keyword classifier is discussed in [3]. Recent utterance verification systems use both the keyword and the anti-keyword models [4] in order to reject these sentences. The keyword models model the data from the keywords and the anti-keyword models model the data that is not the keyword.

2. Utterance Verification

Recent algorithms in utterance verification, [5] - [8] are often formulated as statistical hypothesis testing. At the subword level, this involves verifying that the subword is actually present in the speech segment using the subword and the anti-subword models. The subword likelihood ratio of the null and the alternate hypothesis is used for this purpose. The null hypothesis H_0 represents the input speech containing the given subword S_k. The alternate hypothesis H_1 represents the input speech not containing the given subword. The optimal hypothesis test involves the evaluation of the likelihood ratio

$$T(O;S_k) = \frac{L(O|H_0)}{L(O|H_1)} \tag{1}$$

where O is the sequence of speech observation vectors, $L(O|H_0)$ is the likelihood of the observation sequence given the subword hypothesis for the subword S_k, and $L(O|H_1)$ is the likelihood of the observation sequence given the non subword (anti-subword) hypothesis. The hypothesis test is performed by comparing the $\log T(O;S_k)$ to a predefined critical threshold r_k. The performance of such a system depends on how good the models and the anti models are. Currently utterance verification systems use context independent (CI) anti models for the phones, even though the subwords are modeled by context dependent (CD) phones. The goal of this paper is to expand the anti-subword models to improve the performance of the utterance verification system.

3. Context Dependent Anti-Subword Models

Context Independent (CI) subword models model the basic phone units of speech. The Context Dependent (CD) subword model for the phone a, XaY models the phone a with the left context X and the right context Y. The CD anti-subword model for the subword unit XaY should model all the data that does not have the label XaY. We propose using context dependent anti-models in which each subword unit XaY can be modeled in the following ways:

- using all the data with label other than XaY. In this case the CD anti-subword model for the CD subword unit XaY, is modeled by using the data for all the subword units $\hat{X}\hat{a}\hat{Y}$, where $\hat{X} \neq X$, or $\hat{a} \neq a$, or $\hat{Y} \neq Y$. These CD anti-subword models will be refered to as the type "ALL".

- using data with the same phone label but in different context. Here the CD anti-subword model for the subword unit XaY is modeled by the data from the subword label $\hat{X}a\hat{Y}$, where $\hat{X} \neq X$, or $\hat{Y} \neq Y$. These anti-models will be known as of the type "SAMEPH".
- using data with label of other phones in the same context. The anti-models for the unit XaY is modeled by the data with label $X\hat{a}Y$, where $\hat{a} \neq a$. These anti-models are designated as of the type "SAMECTX".
- using data from other phones in any context. The anti-models for the unit XaY is modeled by the data from $\hat{X}\hat{a}\hat{Y}$, where $\hat{a} \neq a$. These will be refered to as of the type "OTHER".

Each of the above four types of CD anti-subword models can be modeled in the following two ways.

- using all the data available;
- using only the data from the most confusable subword units.

If only the most confusable subword units are used, we will designate them as the type mentioned above. If all of the data is used for the type "ALL", it will be very much the general speech garbage model. For the "OTHER" type, when all of the data is used, it will be the same as the Context Independent (CI) anti-subword model. When all of the data is used, the same context anti-model will be known as of the type "ALSMCTX". For the same phone type , if all the data is used we will refer to it as "ALSMPH". In this paper we will study all of these types of CD anti-subword models. We will present their analysis and the recognition/verification results with these models.

4. Comparison of Context Dependent Anti-Subword Models

In this section, we compare the different types of CD anti-subword models with respect to how well they separate the data from the non-data using the null and the alternate hypothesis. The log likelihood ratio of the null and the alternate hypothesis,

$$LLR = \log T(O; S_k) = \log L(O|H_0) - \log L(O|H_1) \quad (2)$$

is used as the mis-classification measure. The probability density of this measure is plotted in the Figures 1 - 6.

A comparison of these figures show that the anti-models of the type "ALSMCTX" have the smallest overlap and hence the largest separation of the data (speech containing the subword) from the non data (speech not containing the subword). The "SAMECTX" and the "OTHER" type have smaller overlaps than the types "ALSMPH", "SAMEPH", and "ALL". The types, "ALSMCTX", "SAMECTX", and "OTHER" contain the central subword unit different from the subword unit for which the anti-models are being modeled. The types, "ALSMPH", "SAMEPH", and "ALL" contain the same central subword unit as the the subword unit for which the anti-models are being modeled. Thus the confusion between the subword model and the anti-subword model is much more for these types leading to a bigger overlap. This is also shown to be the case when these models are used in recognition/verification as will be shown in the next section.

5. Utterance Verification Experiments

The various types of CD anti-subword models have been tested using a spoken language dialog system [2] for a car reservation task [9], [10]. Here the input speech is first recognized using the sentence grammar and the subword HMM models. Then the individual word segments are verified using the anti-subword models. The subword level scores are combined to yield the the word level scores. A log confidence measure is defined as

$$\log \delta(CM_p) = \log \frac{1}{1 + exp(-LLR_p)} \quad (3)$$

where $\delta(CM_p)$ is the confidence measure for the subword S_p and LLR_p is the log likelihood ratio for the subword S_p. The word level Confidence score is given by

$$\log \delta(CM_w) = 1/N \sum_n \delta(CM_p(n)) \quad (4)$$

where N is the number of subwords in the word w. A threshold based rejection on the word level confidence score is used to reject the sentences.

Anti-Model	FA	ING	OOG	OOT	ACC
ALSMCTX	30.98	80.32	56.84	47.62	77.28
ALSMPH	43.59	75.57	42.63	12.70	70.81
SAMECTX	36.23	79.37	45.26	25.40	74.77
SAMEPH	43.90	75.41	42.11	12.70	70.62
OTHER	41.57	77.41	41.58	12.70	72.35
ALL	43.49	75.78	41.58	12.70	70.90
CI	37.54	79.42	49.47	31.75	75.37

Table 1: Utterance Verification Results

In this paper we present the utterance verification results on the TIME subtask of the car reservation system. All the data for this task was collected over the telephone and spoken by the general public. There are 818 in grammar (ING), 110 out of grammar (OOG), 63 out of task (OOT), and 991 total sentences in this database. This subtask has 51 key words and many out of vocabulary words in the database. The semantic slots are 1895 ING,

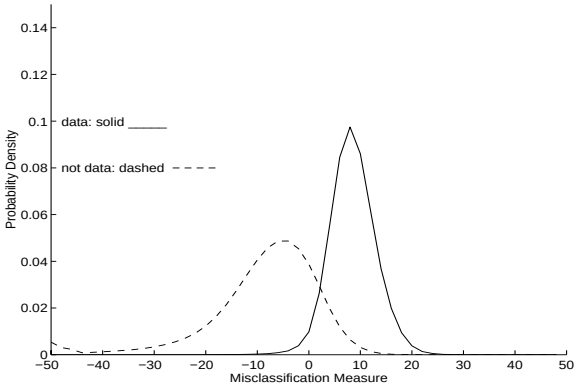

Figure 1: Probability density of the Misclassification Measure for the CD Anti-Subword Models of the type ALSMCTX.

190 OOG, 63 OOt and 2148 total. The verification results are shown in Table 1. From the table, it can be seen that the ALSM-CTX type of CD anti-models has the best performance for all the categories, including False acceptance (OOT results are OOT rejection numbers, the rest are correct recognition numbers). The SAMECTX type and the CI (Context Independent) are the next best in performance. The types ALSMPH, SAMEPH, and ALL types perform at levels less than the other types. These results confirm the analysis presented in Section 4.

6. Utterance Verification using one pass strategy

Next we used a one pass recognition/verification system to compare the CD anti-models. This is the Hybrid Decoder system developed by Koo, Lee, and Juang [9], [10]. This system uses verification in the forward Viterbi decoding itself. The ordinary hybrid decoder uses only the frame level confidence measure. The extended hybrid decoder uses the word level confidence score along with the frame level score during Viterbi decoding. The Utterance verification results for the ordinary hybrid decoder are presented in Table 2.

The results for the extended hybrid decoder are shown in Table 3. Here three sets of results are given. The results for decoding are when only the top candidate from recognition is used. The detection results are with multiple candidates. Even though, a verification scheme is used in the forward recognition path, the performance can still be improved by using a post process rejection. The Post Proc Rejection results are with this scheme. It can be seen from the tables, that the post process rejection gives the best results for all the categories such as the false acceptance, OOT rejection, etc. The extended hybrid decoder performance is very similar to that of the hybrid decoder.

7. Conclusions

In this paper we have extended the anti-subword modeling to CD anti-models. We have presented different ways of obtaining the CD anti-subword models. We have given a comparison of the different types of models, and shown how they separate the speech from the data from that of non data. We have also presented utterance verification results from a two pass recognition/verification system and from two one pass hybrid verification systems. These results show us that modeling the CD anti-models using data from the same context is better than any other type of anti-model. We are currently extending this work to model the anti-models using minimum verification techniques to further improve their performance.

8. REFERENCES

1. T. Kawahara, N. Kitaoka, and, S. Doshita, "Concept Based Phrase Spotting Approach for Spontaneous Speech Understanding," ICASSP96.
2. T. Kawahara, C. Lee, and B. Juang, "Combining Keyphrase detection and Subword-based Verification for Flexible Speech Understanding," ICASSP97.
3. R. Sukkar, and J. Wilpon, "A Two Pass Classifier for Utterance Rejection in Keyword Spotting," ICASSP93.
4. M. Rahim, C. Lee, and B. Juang, "Robust Utterance Verification for Connected Digits Recognition," ICASSP95.
5. R. Rose, B. Juang, and C. Lee, "A Training Procedure for Verifying String Hypothesis in Continuous Speech Recognition," ICASSP95.
6. R. Sukkar, A. Setlur, M. Rahim, and C. Lee, "Utterance Verification of keyword Strings Using Word-Based Minimum Verification Error (WB-MVE) Training," ICASSP96.
7. R. Sukkar, and C. Lee, "Vocabulary Independent Discriminative Utterance Verification for Nonkeyword Rejection in Subword Based Speech Recognition," IEEE Trans. Speech and Audio Processing, Vol. 4, No. 6, pp.420-429, Nov. 1996.
8. R. Sukkar, "Subword-Based Minimum Verification Error (SB-MVE) Training for Task Independent Utterance Verification," ICASSP98.
9. M. Koo, C. Lee, and B. Juang, "A New Hybrid Decoding Algorithm for Speech Recognition and Utterance Verification," Proc. ASRU Workshop, pp. 303-310, 1997.
10. M. Koo, C. Lee, and B. Juang, "A New Decoder Based on a Generalized Confidence Score," ICASSP98.

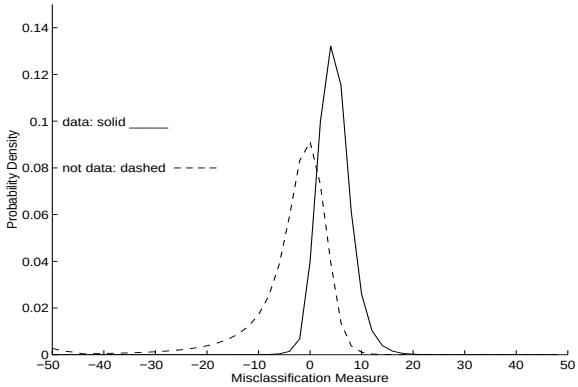

Figure 2: Probability density of the Misclassification Measure for the CD Anti-Subword Models of the type ALSMPH.

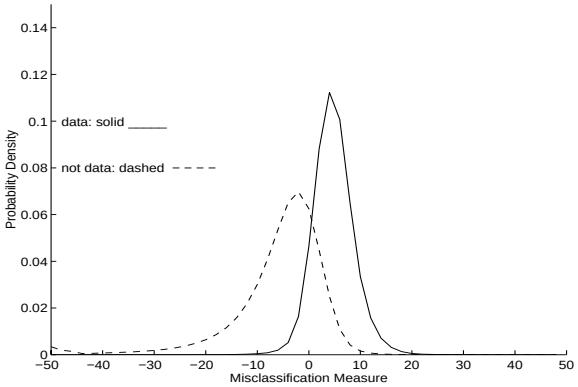

Figure 3: Probability density of the Misclassification Measure for the CD Anti-Subword Models of the type SAMECTX.

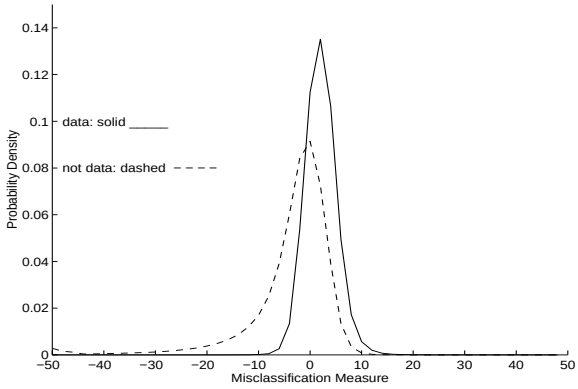

Figure 4: Probability density of the Misclassification Measure for the CD Anti-Subword Models of the type SAMEPH.

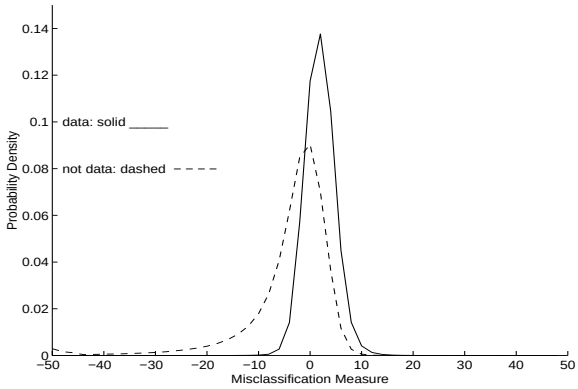

Figure 5: Probability density of the Misclassification Measure for the CD Anti-Subword Models of the type ALL.

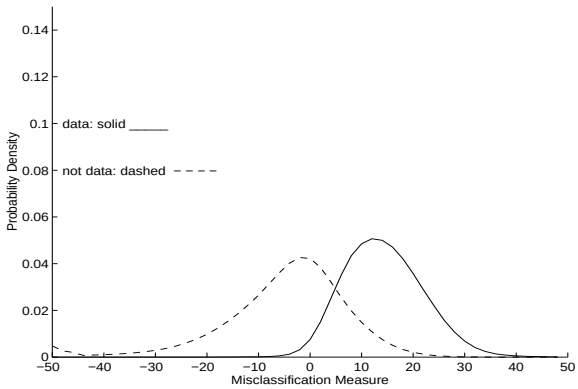

Figure 6: Probability density of the Misclassification Measure for the CD Anti-Subword Models of the type OTHER.

Anti-Model	FA	ING	OOG	OOT	ACC
Decoding					
ALSMCTX	26.64	86.65	36.84	65.08	81.61
ALSMPH	29.06	83.38	44.21	66.67	79.42
SAMECTX	29.47	82.43	47.89	65.08	78.86
SAMEPH	30.47	78.63	48.42	69.84	75.70
OTHER	35.52	82.59	29.47	52.38	77.00
ALL	31.18	78.52	50.53	68.25	75.74
CI	26.03	85.28	50.00	66.67	81.61
Detection					
ALSMCTX	24.02	88.02	58.42	65.08	84.73
ALSMPH	27.14	84.85	56.32	65.08	81.75
SAMECTX	26.44	84.12	61.05	65.08	81.52
SAMEPH	28.15	80.53	57.37	69.84	78.17
OTHER	33.10	83.96	47.37	52.38	79.80
ALL	28.25	80.16	57.89	68.25	77.84
CI	23.31	87.02	62.63	66.67	84.26
Post Process Rejection					
ALSMCTX	23.31	87.97	63.16	69.84	85.24
ALSMPH	26.74	84.59	61.58	65.08	81.98
SAMECTX	25.03	83.11	65.26	69.84	81.15
SAMEPH	27.95	80.42	59.47	69.84	78.26
OTHER	32.59	83.85	51.05	52.38	80.03
ALL	27.95	80.16	61.58	68.25	78.16
CI	22.20	86.49	66.32	69.84	84.22

Table 2: Utterance Verification Results with Ordinary Hybrid Decoder

Anti-Model	FA	ING	OOG	OOT	ACC
Decoding					
ALSMCTX	27.24	86.60	40.53	65.08	81.89
ALSMPH	30.07	83.11	47.89	63.49	79.42
SAMECTX	27.85	82.96	48.95	66.67	79.47
SAMEPH	31.38	78.42	51.05	69.84	75.74
OTHER	35.32	83.38	31.05	52.38	77.84
ALL	29.77	79.00	51.05	68.25	76.21
CI	25.33	86.12	52.63	66.67	82.59
Detection					
ALSMCTX	24.72	88.02	56.32	63.49	84.50
ALSMPH	27.24	84.85	57.37	63.49	81.80
SAMECTX	24.62	84.96	60.53	65.08	82.22
SAMEPH	29.77	79.95	55.79	69.84	77.51
OTHER	32.80	84.64	48.42	52.38	80.49
ALL	28.05	80.42	56.32	68.25	77.93
CI	23.01	87.44	64.74	66.67	84.82
Post Process Rejection					
ALSMCTX	23.71	88.07	60.00	68.25	85.01
ALSMPH	26.84	84.91	62.10	63.49	82.26
SAMECTX	24.32	83.64	64.21	65.08	81.38
SAMEPH	29.67	79.68	57.37	69.84	77.42
OTHER	32.19	84.54	52.63	52.38	80.77
ALL	27.75	80.47	57.89	68.25	78.12
CI	21.90	86.86	67.37	71.43	84.68

Table 3: Utterance Verification Results with Extended Hybrid Decoder