A MORE EFFICIENT AND OPTIMAL LLR FOR DECODING AND VERIFICATION

LAM Kwok Leung and Pascale FUNG

HKUST
Human Language Technology Center
Department of Electrical and Electronic Engineering
University of Science and Technology
Clear Water Bay, Hong Kong
{cpegeric,pascale}@ee.ust.hk

ABSTRACT

We propose a new confidence score for decoding and verification. Since the traditional log likelihood ratio (LLR) is borrowed from speaker verification technique, it may not be apropriate for decoding because we do not have a good modelling and definition of LLR for decoding/utterance verification. We have proposed a new formulation of LLR that can be used for decoding and verification task. Experimental results show that our proposed LLR can perform equally well compared with the result based on maximum likelihood in a decoding task. Also, we get an 5% improvement in decoding compared with traditional LLR.

1. INTRODUCTION

Nowadays, keyword spotting plays an important role in the speech recognition because it is useful for dealing with spontaneous speech. In the telephone application, most of the users speak naturally. It is impossible to define a set of rule to deal with different ways of speaking for continuous speech recognition task. Instead, keywords are extracted from the spontaneous speech. In this way, the system can perform user requests in an efficient manner. However, One of the disadvantages of keyword spotting is that its errors lead to degradation in understanding. Thus, a rejection/acceptance mechanism[2, 5] is essential for rejecting the incorrect keyword or accepting the correct keyword. The goal of utterance verification (UV) is to verify whether a hypothesized word or string of words correspond to actual occurences of those keyword[8, 6].

In our system, Speech Assisted onLine Search Agent (SALSA)[1], users need to speak the word "SALSA" as a pass-phrase. The detection and verification of the word "SALSA" becomes a critical task in terms of system response and performance. Rejection of the incorrect "SALSA" or acceptance of the correct "SALSA" is important since the system will only responses to user correctly if the performance of detection/verification is good.

There are two approaches to incorporate utterance verification into a speech recognition system. (1) Post-processor – First, we pass the speech to keyword spotter and get the recognition result. Then, the UV system to gives a confidence measure to the speech segment. This is called two-pass recognition/verification strategy. (2) a modified Viterbi decoder is used for both decoding and verification in one-pass strategy[4]. The one-pass strategy is more efficient than the two-pass algorithm.

In this paper, we propose a new formulation of LLR so that the decoding task using LLR can perform as good as Maximum Likelihood. Also, confidence score can be obtained in the one-pass Viterbi decoding algorithm.

The organization of this paper is as follows: In the next section, we present the problem of the traditional LLR. In section 3 we formulate a new LLR. In section 4, Decoder based on LLR is described and Verification will be presented in section 5. Experimental setup and experimental results are given in Section 6 and 7 followed by conclusions in section 8.

2. PROBLEM OF TRADITIONAL LLR

The general technique of verification is using the log likelihood ratio (LLR) as the confidence measure. The most commonly used confidence meausure is the discriminative function

$$LLR = \log \frac{P(O|H_0)}{P(O|H_1)}$$

H_0: null hypothesis, target model H_1: alternative hypothesis, alternative model

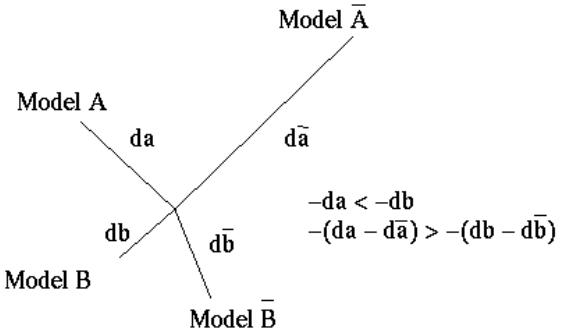
For implementation based on HMMs, LLR will become

$$LLR_{old} = \log \frac{b_j^c(o_t)}{\max_{m=1}^{M} b_j^m(o_t)},$$

where $b_j(o_t)$ is the observation probability in state j at frame t, c is the correct model and M is the number of model except the correct model.

However, this type of LLR may not be apropriate for decoding since alternative hypothesis is not modelled well. The problem is due to the fact that the alternative model always follows the same state as the target model. In some cases, the traditional LLR does not find the most representative alternative hypothesis, so the decoding task based on LLR can not perform as good as likelihood(Figure 1). In Figure 1, model B is more likely than model A. However, in terms of LLR, the model A is more likely than model B. For this motivation, we proposed a new LLR so that we want to have discriminative function that is consistent with likelihood in decoding task.

Figure 1: Disadvantage of traditional LLR



3. NEW FORMULATION OF LLR

Refer to (Figure 1), the traditional LLR is inconsistent with the likelihood. Since the alternative model always follows the same state as the target model, it does not always give the optimal score in the global observation space. Instead, the score is a local maximum in the observation space within the particular state.

We propose a new LLR to make it more consistent with likelihood and more optimal in the observation space. At the same time, performance can be improved for verification. To achieve this goal, the new LLR is:

$$LLR_{new} = \log \frac{b_j^c(o_t)}{\max_{m=1}^{M} \max_{k=1}^{N} b_k^m(o_t)},$$

where N is the number of state and M is the number of model except the target model

However, this type of LLR is computational expensive since the computation time is N times more than the traditional LLR. For this reason, a subword-class anti-model has been used instead of using M subword models[7]. For details of the antisubword class models, please refer to [7].

The proposed LLR is now as follows:

$$LLR_{new} = \log \frac{b_j^c(o_t)}{\max_{k=1}^{N} b_k^a(o_t)},$$

where N is the number of state and a is the alternative model

In this case, our proposed LLR is more efficient and more optimal. time.

4. DECODER BASED ON NEW LLR

In order to use LLR for decoding, a modified Viterbi decoder is implemented[4, 3].

We define $\delta_t(j)$ as the best score along a single path at frame t, which accounts for the first t observations and ends in state j, the modified Viterbi algorithm will be as follows:

$$\delta_t(j) = \max_i [\delta_{t-1}(i) + \log a_{ij}] + \log \frac{b_j^c(o_t)}{\max_{k=1}^{N} b_k^a(o_t)},$$

The accumulated path score $\delta_t(j)$, obtained in the Viterbi algorithm corresponds to the confidence measure in the path time t. This implied that we can do one-pass recogniton/verification algorithm instead of two-pass algorithm[4].

5. VERIFICATION BASED ON NEW LLR

Since our task is based on subword units HMMs. The confidence measure for the word string is computed based on the confidence score of the subword units.

$$LLR_{subword} = \sum_{t=1}^{T} \log \frac{b_j(O_t)}{\max_{k=1}^{N} b_k^a(o_t)},$$

where N is the number of states of each model and T is the duration of the subword model

The normalized LLR_{word} is used as confidence measure for verification.

$$NormalizedLLR_{word} = \frac{1}{T} \sum_{n=1}^{N} LLR_n,$$

where T - the duration of the word string and N is the number of subword units for the word string

In order to compute the threshold of keyword likelihood ratio based on subword units, an individual threshold of each subword units is computed based on the pdf_s of incorrect and correct recogntion of each subword units (Figure 5, Figure 3). The correct recognition score for subword units is computed based on the decoding score using the true lexicon. Incorrect recognition score for subword units is computed based on the random lexicon. In the figure, dotted line is the *pdf* of the correct recognition and solid line is the *pdf* of the incorrect recognition.

Figure 2: Probability density of /v/ based on traditional LLR. Large overlap between the pdf of the correct and incorrect recognition is shown

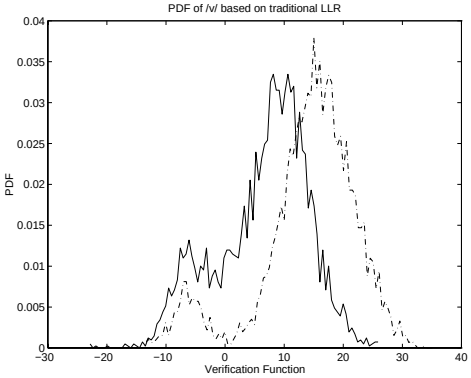
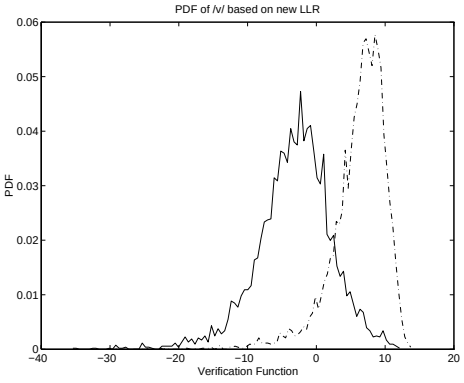


Figure 3: Probability density of /v/ based on new LLR. Better separation between the pdf of correct and incorrect recognition is shown



6. EXPERIMENTAL SETUP

Our system is a continuous speech recogntion system based on phoneme continuous density hidden Markov models. Mixture Gaussian state observation density has a maximum of 10 mixture components per state. Each subword unit is modeled by a 3-state left-to-right HMM with no state skips. We use the set of context-

Figure 4: Probability density of /r/ based on traditional LLR. Large overlap between the pdf of the correct and incorrect recognition is shown

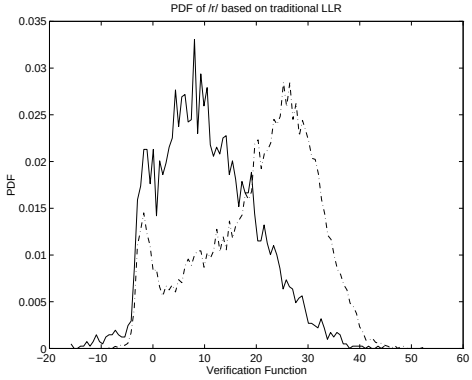
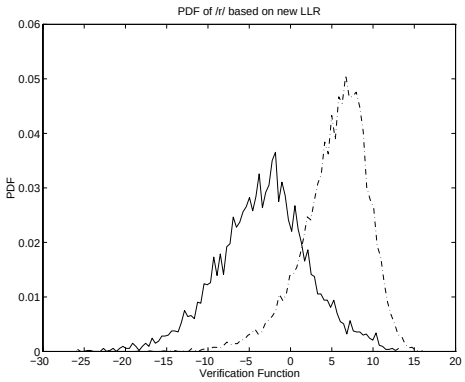


Figure 5: Probability density of /r/ based on new LLR. Better separation between the pdf of correct and incorrect recognition is shown



independent phone units as a universal phone set. The total units we used are 45 context-indepedent phones. The recognzier feature vector consisted of the following 39 parameters: 12 MFCC, 12 delta cepstral coefficients, 12 delta-delta cepstral coefficients, energy, and the delta and delta-delta of the energy parameters.

For anti-models, 6 antisubword class models are used.

7. EXPERIMENTAL RESULTS

Two experiments are conducted to test the overall performance of the proposed LLR, which compare the decoding and verification performance between LLR_{old} and LLR_{new}.

7.1. Decoding

For decoding performance, the testing data consists of defined phrases only and are covered by a finite-state

Figure 6: Recognition Accuracy

Likelihood	LLR_{new}	LLR_{old}
97.6%	96.9%	91.84%

Figure 7: Utterance verification performance

	False Rejection (%)	False Alarm (%)
LLR_{new}	5%	54%
	10%	45.6%
	15%	37%
LLR_{old}	5%	77.8%
	10%	71%
	15%	64.6%

sentence grammars. The likelihood ratio decoder has been implemented[4] so that we can compare the result of recognition based on different decoding criterion. The result shows that the proposed LLR_{new} approaches the performance of the Maximum-Likelihood-based system whereas the performance of LLR_{old} is worse. The recogntion performance based on new LLR has about 5% improvement compared with the performance based on traditional LLR(Figure 6).

7.2. Verification

For utterance verification, there are two types of error for evaluation. Type I: False Rejection – The correctly decoded keyword is rejected by UV. Type II:False Alarms – The incorrectly decoded keyword is accepted by UV.

For our results, we fix the false rejection rate at 5.0%, 10% and 15% (Figure 7), our proposed LLR has a low false acceptance rate than the traditional LLR.

When $LLR_{subword}$ is used as the subword-level verification function, the likelihood function are modeled by the probability density functions of $LLR_{subword}$ (Figure 3, Figure 5). In the figure, solid line corresponds to incorrect subword recognition and dotted line corresponds to correct subword recognition.

In Figures 2,3,4,5, there are some comparisions between the pdf_s of subword /v/ and /r/ based on traditional and new LLR. From the figures, the pdf_s based on new LLR has less ovarlap between the *pdf* of correct recogntion score and incorrect recogntion score.

8. CONCLUSION

In this paper, we formulate a new LLR for utterance verification. From the experiment results, we find that new LLR gives 5% improvement when compared to the traditional LLR. It also gives a better performance in verification than traditional LLR. We also show that our proposed LLR is more optimal in the whole observation space and more performant than the traditional LLR.

9. REFERENCES

[1] Pascale FUNG, CHEUNG Chi Shuen, LAM Kwok Leung, LIU Wai Kat, and LO Yuen Yee. A speech assisted online search agent (salsa). In *ICSLP*, 1998.

[2] Taktoshi JITSUHIRO, Satoshi Takehasi, and Kiyoaki Aikawa. Rejection of out-of-vocabulary words using phoneme confidence liklihood. In *ICASSP*, 1998.

[3] Myoung-Wan Koo, Chin-Hui Lee, and Biing-Hwang Juang. A new decoder based on a generalized confidence score. In *ICASSP*, 1998.

[4] E. Lleida and R. C. Rose. Efficient decoding and training procedures for utterance verification in continuous speech recogntion. In *ICASSP*, 1996.

[5] Ze'ev Rivlin, Michael Cohen, Victor Abrash, and Thomas Chung. A phone-dependent confidence measure for utterance rejection. In *ICASSP*, 1996.

[6] R. C. Rose, H. Yao, G. Roccardi, and J. Wright. Integration of utterance verification with statistical language modeling and spoken language understanding. In *ICASSP*, 1998.

[7] Rafid A. Sukkar and Chin-Hui Lee. Vocabulary independent discriminative utterance verification for nonkeyword rejection in subword based speech recognition. In *ICASSP*, 1996.

[8] Sheryl R. Young. Detecting misrecognitions and out-of-vocabulary words. In *ICASSP*, 1994.