Probabilistic Prediction in Multiclass Classification Derived for Flexible Text-Prompted Speaker Verification

著者	Kurogi Shuichi, Sakashita Shota, Takeguchi Satoshi, Ueki Takuya, Matsuo Kazuya
journal or publication title	Neural Information Processing
volume	9489
number	1
page range	216-225
year	2015-11-09
URL	http://hdl.handle.net/10228/5585

doi: info:doi/ 10.1007/978-3-319-26532-2

Probabilistic Prediction in Multiclass Classification Derived for Flexible Text-Prompted Speaker Verification

Shuichi Kurogi, Shota Sakashita, Satoshi Takeguchi, Takuya Ueki, and Kazuya Matsuo

Kyushu Institute of technology, Tobata, Kitakyushu, Fukuoka 804-8550, Japan
{kuro@,sakashita@kurolab.,takeguchi@kurolab.,matsuo@}cntl.kyutech.ac.jp
http://kurolab.cntl.kyutech.ac.jp/

Abstract. So far, we have presented a method for text-prompted multistep speaker verification using GEBI (Gibbs-distribution based extended Bayesian inference) for reducing single-step verification error, where we use thresholds for acceptance and rejection but the tuning is not so easy and affects the performance of verification. To solve the problem of thresholds, this paper presents a method of probabilistic prediction in multiclass classification for solving verification problem. We also present loss functions for evaluating the performance of probabilistic prediction. By means of numerical experiments using recorded real speech data, we examine the properties of the present method using GEBI and BI (Bayesian inverence) and show the effectiveness and the risk of probability loss in the present method.

Keywords: probabilistic prediction, text-prompted speaker verification, Gibbs-distribution-based extended Bayesian inference, loss functions in multiclass classification

1 Introduction

So far, we have presented a method for text-prompted multistep speaker verification [1, 2]. Here, from [3], text-prompted speaker verification has been developed to combat spoofing from impostors and digit strings are often used to lower the complexity of processing. From another perspective, the method focuses on reducing verification error by means of multistep verification using Gibbs-distribution-based Bayesian inference (GEBI) for rejecting unregistered speakers [2], where from the analysis of the properties, it is suggested that the tuning of the thresholds for acceptance and rejection is not so easy and affects the performance. Namely, we have tuned the thresholds by the method of EER (equal error rate) for FAR (false acceptance rate) and FRR (false rejection rate) to be almost the same. Furthermore, the obtained values of the thresholds are not so easy to be modified for different security or risk level of verification. To solve this problem, this paper presents probabilistic prediction. Here, note that from [4] and our experience, we can see that the probabilistic prediction in weather and climate forecasting allows the users to decide on the level of risk they are prepared and to take appropriate action within a proper understanding of the uncertainties. For introducing probabilistic prediction into verification, we first formulate multiclass classification problem, and then

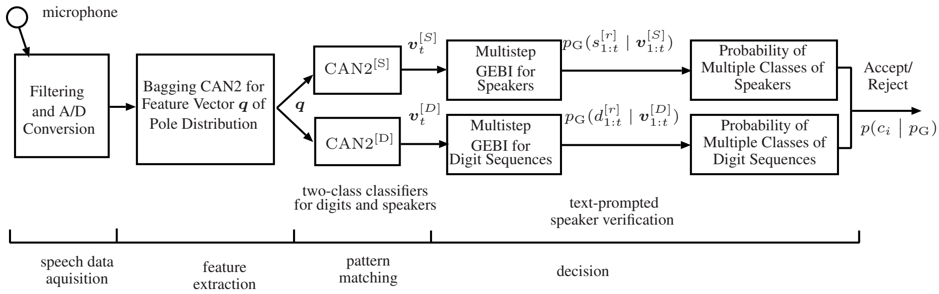


Fig. 1. Diagram of text-prompted speaker verification system using CAN2s

apply Bayesian inference (BI) to obtain the probability. We also present loss functions to evaluate the performance of the probabilistic prediction in multiclass classification derived for verification problem, and then examine the properties and effectiveness of the present method by means of using real speech signal.

Here, note that our speech processing system employs competitive associative nets (CAN2s). The CAN2 is an artificial neural net for learning efficient piecewise linear approximation of nonlinear function [5], and we have shown that feature vectors of pole distribution extracted from piecewise linear predictive coefficients obtained by the bagging (bootstrap aggregating) version of the CAN2 reflect nonlinear and time-varying vocal tract of the speaker [6]. Although the most common way to characterize speech signal in the literature is short-time spectral analysis, such as Linear Prediction Coding (LPC) and Mel-Frequency Cepstrum Coefficients (MFCC) [7], the bagging CAN2 learns more precise information than LPC and MFCC (see [6] for details).

We show the method of probabilistic prediction in **2**, experimental results and analysis in **3**, and the conclusion in **4**.

2 Probabilistic Prediction for Text-Prompted Speaker Verification

Fig. 1 shows an overview of the present text-prompted speaker verification system using CAN2s. In the same way as general speaker recognition systems [7], it consists of four steps: speech data acquisition, feature extraction, pattern matching, and making a decision. In this research study, we use a feature vector of pole distribution obtained from a speech signal (see [6] for details).

2.1 Multistep Speaker and Text Verification Using GEBI

Here, we show a brief explanation of multistep verification using GEBI (see [2] for details). In order to achieve text-prompted speaker verification using digits, let $S = \{s_i | i \in I^{[S]}\}$ and $D = \{d_i | i \in I^{[D]}\}$ denote a set of speakers $s \in S$ and digits $d \in D$, respectively, where $I^{[S]} = \{1, 2, \cdots, |S|\}$ and $I^{[D]} = \{1, 2, \cdots, |D|\}$. Furthermore, let $\mathrm{RLM}^{[M]}$ for $M = S$ and M be a set of regression learning machines $\mathrm{RLM}^{[m]}$ ($m \in I^{[M]}$), and each $\mathrm{RLM}^{[m]}$ learns to predict a single-step verification as $v^{[m]} = 1$ for the acceptance of a speech segment of a speaker $m = s_i$ or a digit $m = d_i$, and $v^{[m]} = 0$ for the rejection. Here, let us suppose that we have speech segments of spoken digits obtained by some appropriate segmentation method and this research focuses on the multistep verification of spoken digit sequences.

For multistep verification of input sequence of spoken digits, we have proposed Gibbs-distribution-based extended Bayesian inference (GEBI) as shown below for overcoming the problem of Bayesian inference (BI) in speaker verification of unregistered speakers (see [2] for details). Let $\boldsymbol{v}_{1:T}^{[M]} = \boldsymbol{v}_1^{[M]}\boldsymbol{v}_2^{[M]}\cdots\boldsymbol{v}_t^{[M]}$ be an output sequence of RLM$^{[m]}$ for the reference sequence $m_{1:T}^{[r]} = m_1^{[r]}m_2^{[r]}\cdots m_T^{[r]}$, we have recursive posterior probability for $t = 1, 2, \cdots, T$ as follows,

$$p_{\mathrm{G}}\left(m_{1:t}^{[r]} \mid \boldsymbol{v}_{1:t}^{[M]}\right) = \frac{1}{Z_t} p_{\mathrm{G}}\left(m_{1:t-1}^{[r]} \mid \boldsymbol{v}_{1:t-1}^{[M]}\right)^{\beta_t/\beta_{t-1}} p\left(\boldsymbol{v}_t^{[M]} \mid m_t^{[r]}\right)^{\beta_t} \quad , \tag{1}$$

$$p_{\mathrm{G}}\left(\overline{m_{1:t}^{[r]}} \mid \boldsymbol{v}_{1:t}^{[M]}\right) = \frac{1}{Z_t} p_{\mathrm{G}}\left(\overline{m_{1:t-1}^{[r]}} \mid \boldsymbol{v}_{1:t-1}^{[M]}\right)^{\beta_t/\beta_{t-1}} p\left(\boldsymbol{v}_t^{[M]} \mid \overline{m_t^{[r]}}\right)^{\beta_t} \quad . \tag{2}$$

where $\beta_t = \beta/t\ (t \geq 1)$ and $\beta_0 = 1$, and Z_t is the normalization constant. Note that the conventional BI is obtained for $\beta_t = 1 (t \geq 0)$ and we denote p_{B} instead of p_{G} for the probability obtained by the above equations with $\beta_t = 1 (t \geq 0)$, while p_{G} is obtained for $\beta_t = 1/t\ (t \geq 1)$ in the experiments shown below.

The verification by our previous method shown in [2] at $t = T$ is given by

$$V_{1:T}^{[M]} = \begin{cases} 1 \text{ if } p_{\mathrm{G}}\left(m_{1:T}^{[r]} \mid \boldsymbol{v}_{1:T}^{[M]}\right) \geq p_\theta^{[M]} \\ -1 \text{ otherwise} \end{cases} \tag{3}$$

for speaker $(m, M) = (s, S)$ and text $(m, M) = (d, D)$, respectively. Here, $p_\theta^{[M]}$ for $M = S$ and D are thresholds, and $V_{1:T}^{[M]} = 1$ and -1 indicates acceptance and rejection, respectively. The verification of text-prompted speaker is executed by $V_{1:T}^{[SD]} = V_{1:T}^{[S]} \wedge V_{1:T}^{[D]} = 1$ and -1 for acceptance and rejection, respectively. The performance of verification depends on the thresholds $p_\theta^{[M]}$ for $M = S$ and D. To execute more flexible verification than using thresholds, we introduce probabilistic probability into the verification problem in the next section.

2.2 Probabilistic Prediction for Speaker and Text Verification

We introduce multiple classes to classify the verification results, and then introduce probabilistic prediction for speaker and text verification.

Multiclass Classification for Speaker and Text Verification For speaker verification, we consider the following three classes, where we suppose all elements in each input and reference speaker sequence, respectively, consists of the same speaker;

$c_{+1}^{[S]}$ (Class of correct speakers): class of speakers satisfying $s_{1:T} = s_{1:T}^{[r]}\ (\in S_{1:T})$ for the input $s_{1:T}$ and the reference $s_{1:T}^{[r]}$, where S is the set of registered speakers, and $S_{1:T}$ denotes the set of $s_{1:T}$ whose all elements $s_t\ (t = 1, 2, \cdots, T)$ are registered speaker $s_t \in S$.

$c_{-1}^{[S]}$ (Class of incorrect speakers): class of speakers satisfying $s_{1:T} \neq s_{1:T}^{[r]}$ for $s_{1:T}, s_{1:T}^{[r]} \in S_{1:T}$.

$c_0^{[S]}$ (Class of unregistered speakers) : class of speakers satisfying $s_{1:T} \neq s_{1:T}^{[r]}$ for $s_{1:T} \notin S^{[T]}$.

Here, note that these classes are determined for the pair of input and reference sequences.

For text (or digit sequence) verification, we consider the following $N+1$ classes of $T(=mN)$-length digit sequence consisting of m times of N-length subsequences:

$c_i^{[D]}$ for $i = 0, 1, 2, \cdots, N$ (Class of digit sequences with correct ratio being i/N): class of input $d_{1:T}$ and reference $d_{1:T}^{[r]}$ digit sequences, which consist of m times of N-length subsequence whose i digits are the same.

In order to simplify the explanation, let $C^{[S]} = \{c_i^{[S]} \mid i \in I^{[C^{[S]}]})\}$ be the set of speaker verification classes, $C^{[D]} = \{c_i^{[D]} \mid i \in I^{[C^{[D]}]}\}$ be the set of text verification classes, C denote $C^{[S]}$ or $C^{[D]}$, and $I^{[C]}$ denote $I^{[C^{[S]}]} = \{-1, 0, 1\}$ or $I^{[C^{[D]}]} = \{0, 1, 2, \cdots, N\}$.

Note that these classes have the ordered indices which we utilize for probabilistic prediction of multiclass classification derived for the verification. Namely, we can divide two sets of classes, where one consists of the classes with the indices from $i = i_\theta^{[C]}$ to $i_{\max}^{[C]}$ and the other consists of the remaining classes, where $i_\theta^{[C]}$ and $i_{\max}^{[C]}$ indicate the threshold for verification and the maximum index of the classes in C, respectively. Furthermore, as shown in **3.2**, we have a possibility to have a class with a large classification error but a sum of adjacent classes has smaller error. Thus, in order to achieve a reliable probabilistic prediction, we will combine some adjacent classes so that every combined class has smaller classification error.

Probabilistic Prediction in Multiclass Classification In order to formulate the probabilistic prediction of multiclass classification, let $X^{[\text{test}]} = \{(\boldsymbol{x}_j, t_j) \mid j \in I^{[\text{test}]}\}$ be a test dataset, where $\boldsymbol{x}_j$ is the jth data of the pair $\left(m_{1:T}^{[r]}, \boldsymbol{v}_{1:T}^{[M]}(m_{1:T})\right)$ determined by the sequences of reference $m_{1:T}^{[r]}$ and input $m_{1:T}$, $t_j \in C$ indicates target class to be classified, and $I^{[\text{test}]} = \{1, 2, \cdots, |I^{[\text{test}]}|\}$. Furthermore, let $p_{\mathrm{G}}(\boldsymbol{x}_j)$ denote the GEBI probability $p_{\mathrm{G}}\left(m_{1:T}^{[r]} | \boldsymbol{v}_{1:T}^{[M]}\right)$ given by (1). Then, from BI, we have the following posterior probability

$$p\left(c_i \mid p_{\mathrm{G}}(\boldsymbol{x}_j)\right) = \frac{p\left(p_{\mathrm{G}}\left(\boldsymbol{x}_j\right) \mid c_i\right) \, p(c_i)}{\sum_{c_l \in C} p\left(p_{\mathrm{G}}(\boldsymbol{x}_l) \mid c_l\right) \, p(c_l)}, \tag{4}$$

where $p(c_i)$ is the prior probability of $c_i \in C$, and $p\left(p_{\mathrm{G}}(\boldsymbol{x}_j) \,\middle|\, c_i\right)$ denotes the likelihood of the value of $p_{\mathrm{G}}(\boldsymbol{x}_j)$ being for c_i. Here, $p\left(p_{\mathrm{G}}(\boldsymbol{x}_j) \,\middle|\, c_i\right)$ can be estimated from a training dataset $X^{[\text{train}]} = \{(\boldsymbol{x}_j, t_j) \mid j \in I^{[\text{train}]}\}$ involving $\boldsymbol{x}_j$ independent and identically distributed (i.i.d) with respect to the data in the test dataset, and we usually use $p(c_i)$ to be equal for all c_i, while we can use specific values depending on the situation, e.g., we can use $p\left(c_0^{[S]}\right) = 0$ for the situation where there is no unregistered speaker expected.

With the above probability $p(c_i \mid p_{\mathrm{G}}(\boldsymbol{x}_j))$ for $c_i \in C$, the user or a decision maker is expected to make flexible decision for verification as shown in **3.2**.

2.3 Loss functions for Evaluating the Performance

We use the following loss functions to evaluate the performance of the probabilistic prediction in multiclass classification extended from the loss functions for two-class classification shown in [8]. First, we divide the multiple classes into two sets of classes: one consists of a class with the maximum probability and the other of remaining classes, where the index of the class in the former set is given by

$$i_M(j) = \underset{i \in I^{[C]}}{\text{argmax}}\, p(c_i \mid p_{\mathrm{G}}(\boldsymbol{x}_j)). \tag{5}$$

Now, the average classification error (ACE) for $i_M(j)$ is given by

$$L_{\mathrm{ACE}} = \frac{1}{n}\left[\sum_{j \in I^{[\text{test}]}} \mathbf{1}\{t_j \neq c_{i_M(j)}\}\right] = \frac{1}{n}\left[\sum_{\{j \mid t_j \neq c_{i_M(j)}\}} 1\right] \tag{6}$$

Here, $\mathbf{1}\{z\}$ indicates an indicator function, equal to 1 if z is true, and to 0 if z is false, $\{j|t_j \neq c_{i_M(j)}\}$ indicates the set of indices satisfying $t_j = c_{i_M(j)}$ for $j \in I^{[\text{test}]}$.

The negative log probability loss (NLP) for $i_M(j)$ is given by

$$L_{\mathrm{NLP}} = -\frac{1}{n}\left[\sum_{\{j \mid t_j = c_{i_M(j)}\}} \log p\left(c_{i_M(j)} \mid p_{\mathrm{G}}(\boldsymbol{x}_j)\right) + \sum_{\{j \mid t_j \neq c_{i_M(j)}\}} \log\left(1 - p\left(c_{i_M(j)} \mid p_{\mathrm{G}}(\boldsymbol{x}_j)\right)\right)\right] \tag{7}$$

The first term of the right hand side becomes smaller for larger probability of correct classification and the second term becomes smaller for smaller probability of incorrect classification.

The negative log predictive density loss (NLPD) for evaluating regression performance given by

$$L_{\mathrm{NLPD}} = -\frac{1}{n}\left[\sum_{j \in I^{[\text{test}]}} \log p\left(t_j \mid p_{\mathrm{G}}(\boldsymbol{x}_j)\right)\right] \tag{8}$$

is considered to be applicable for evaluating the performance of probabilistic prediction in multiclass classification.

3 Experiments

3.1 Experimental Setting

We have recorded speech data sampled with 8kHz of sampling rate and 16 bits of resolution in a silent room of our laboratory. They are from seven speakers (2 female and 5 mail speakers): $S = \{\mathrm{fHS}, \mathrm{fMS}, \mathrm{mKK}, \mathrm{mKO}, \mathrm{mMT}, \mathrm{mNH}, \mathrm{mYM}\}$ for ten Japanese digits $D = \{/\mathrm{zero}/, /\mathrm{ichi}/, /\mathrm{ni}/, /\mathrm{san}/, /\mathrm{yon}/, /\mathrm{go}/, /\mathrm{roku}/, /\mathrm{nana}/, /\mathrm{hachi}/, /\mathrm{kyu}/\}$. For each speaker and each digit, ten samples are recorded on different times and dates among two months. We denote each spoken digit by $x = x_{s,d,l}$ for $s \in S$, $w \in W$ and $l \in L = \{1, 2, \cdots, 10\}$, and the given dataset by $X = (x_{s,d,l} | s \in S, d \in D, l \in L)$.

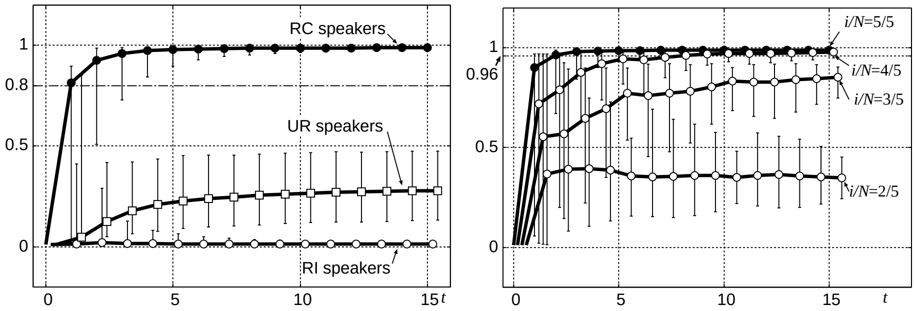
By meas of random selection from X, we have generated training dataset $X^{[\text{train}]} = \{(\boldsymbol{x}_j, t_j) \mid j \in I^{[\text{train}]}\}$ for making the likelihood $p\left(p_{\mathrm{G}}(\boldsymbol{x}_j) \mid c_i\right)$ given in (4) and test dataset $X^{[\text{test}]} = \{(\boldsymbol{x}_j, t_j) \mid j \in I^{[\text{test}]}\}$ for evaluating the performance of probabilistic prediction. A data $\boldsymbol{x}_j$ indicates the jth data of $\left(m_{1:T}^{[r]}, \boldsymbol{v}_{1:T}^{[M]}(m_{1:T})\right)$ consists of reference and input sequences of $T(=15)$-length spoken digits for $T = m \times N = 15$ with $m(=3)$ times of $N(=5)$-length digit sequences indicating some ID numbers. Of course, $\boldsymbol{x}_j \in X^{[\text{train}]}$ and $\boldsymbol{x}_j \in X^{[\text{test}]}$ are not the same but should be independent and identically distributed (i.i.d). To have this done, for each of training and test datasets, we have generated 1,000 data for each combination of 3 classes of speaker sequences involving correct, incorrect and unregistered speakers and 6 classes of digit sequences involving i/N correct digits for $i = 0, 1, \cdots, N = 5$. Thus, we have 18,000 data for training and test datasets.

In order to evaluate the performance of learning machines $\mathrm{RLM}^{[M]}$ for predicting unknown (untrained) data and the data of unregistered speaker, we employ a combination of LOOCV (leave-one-out cross-validation) and OOB (out-of-bag) estimate (see [2] for details). For the regression learning machines, we have used CAN2s for learning piecewise linear approximation of nonlinear functions (see [6] for details).

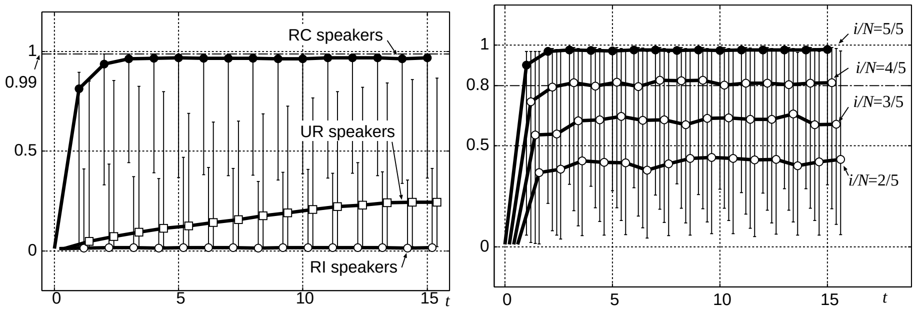
3.2 Experimental Results and Analysis

Experimental Result of Probabilistic Prediction First of all, we show the multi-step probabilities in Fig. 2. As explained in [2], we have tuned the thresholds to be $(p_\theta^{[S]}, p_\theta^{[D]}) = (0.80, 0.96)$ for GEBI and $(0.99, 0.80)$ for BI to achieve EER (equal error rate) at $t = T = 15$ for FAR (false acceptance rate) and FRR (false rejection rate) to be almost the same. For this tuning, we also have employed thresholds $(i_\theta^{[S]}, i_\theta^{[D]}) = (1, 4)$ for deciding the security level of correct verification, i.e., we assume that the data in $c_i^{[S]}$ for $i \geq i_\theta^{[S]} = 1$ and $c_i^{[D]}$ for $i \geq i_\theta^{[D]} = 4$ should to be accepted in speaker and text verification, respectively, and the other data should be rejected. In Fig. 2, we can see that these threshold values seem reasonable but not so easy to be tuned.

By means of the probability prediction by (4), we have the probability $p\left(c_i \mid p_{\mathrm{G}}\right) = p\left(c_i \mid p_{\mathrm{G}}(\boldsymbol{x}_j)\right)$ and $p\left(c_i \mid p_{\mathrm{B}}\right) = p\left(c_i \mid p_{\mathrm{B}}(\boldsymbol{x}_j)\right)$ as shown in Fig. 3. From Fig. 3(a), we can estimate the probability of the classes depending on p_{G}. For example, from the left hand side of Fig. 3(a) for speaker verification, the probability of correct speaker



(a) GEBI

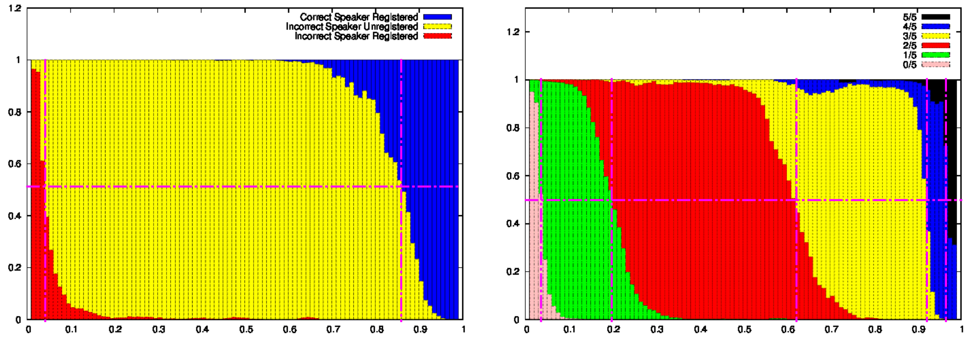


(b) BI

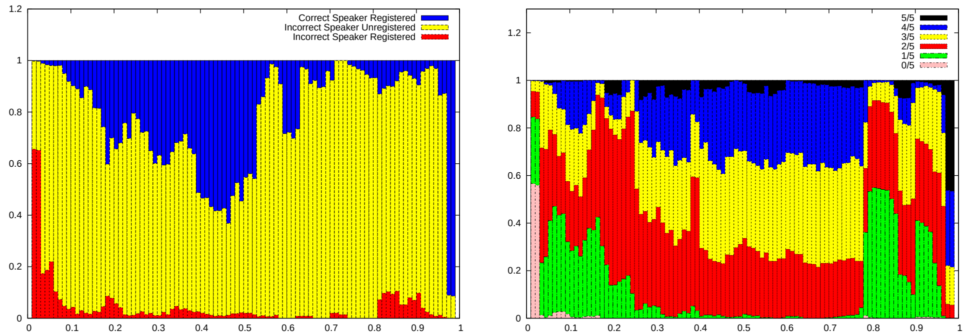
Fig. 2. Experimental result of multistep probability of (a) GEBI and (b) BI for speakers (left) and digits (right), where the curves of speakers denote RC (registered correct), UR (unregistered), RI (registered incorrect). The plus and minus error bars indicate RMS (root mean square) of positive and negative errors from the mean, respectively. The curves for different datasets are shifted slightly and horizontally to avoid crossovers.

and unregistered speaker is expected for the value of $p_{\rm G}$ larger than 0.86 and 0.04, respectively, Furthermore, from the right hand side of Fig. 3(a) for text verification, the ratio of correct digits is expected to be more than 5/5, 4/5, 3/5, 2/5, 1/5 for the value of $p_{\rm G}$ larger than 0.97, 0.93, 0.62, 0.19, 0.03, respectively, On the other hand, it is hard to obtain the property of the probability for BI as shown in Fig. 3(b). This is owing to the fluctuation of the mean value and the large variance of $p_{\rm B}$ as shown in Fig. 2 and a mathematical analysis is shown in [2].

Experimental Result of Losses and Remarks We show experimental results of losses in Table 1, where $L_{{\rm AVE}_\theta}$ indicates AVE (average verification error) obtained for the method using the thresholds given above. From the comparison of the losses between GEBI and BI, we can see that GEBI has achieved smaller losses (bold face figures) for almost all classes than BI, especially, it has achieved smaller mean values for all losses. From the columns of $L_{{\rm AVE}_\theta}$ for GEBI, we can see that the mean verification error $L_{{\rm AVE}_\theta}$ is 0.004 and 0.032 for speaker and text verification, respectively, and they seem small enough.



(a) GEBI



(b) BI

Fig. 3. Posterior probability $p\left(c_i \;\middle|\; p_{\mathrm{G}}\right)$ for (a) p_{G} obtained by GEBI and (b) $p_{\mathrm{B}}\ (= p_{\mathrm{G}}|_{\beta_t=1})$ by BI for speaker (left) and text (right) classification. The horizontal axis indicates p_{G} or p_{B} and the vertical length of a colored bar indicates the probability of a class c_i corresponding to the color.

Next, for the class index $i = 4$ in text verification, we can see that $L_{\mathrm{ACE}} = 0.844$ is very larger than others. This indicates that the probabilistic prediction for the class has very low reliability. As shown in [2], these errors are owing that the discrimination of the data in $c_4^{[D]}$ and $c_5^{[D]}$ are difficult which we can see in Fig. 2(a)(right) for the curves of $i/N = 5/5$ and $4/5$.

To solve this problem for more reliable classification, we combine the class $c_4^{[D]}$ and $c_5^{[D]}$ into a class $c_{4\uplus 5}^{[D]}$. Then, for a test data in $c_{4\uplus 5}^{[D]}$, we have $L_{\mathrm{ACE}} = 0.021$ for the prediction using GEBI. As a result, by means of using the classes $c_i^{[D]}$ for $i = 0, 1, 2, 3, 4 \uplus 5$, we have achieved L_{ACE} less than 0.076 with the mean 0.044. These analysis and modification indicate that we have to understand and reduce the risk of probability loss in using probabilistic prediction. From this point of view, we hardly use the probabilistic prediction obtained by BI. We would like to analyse other losses in our future research.

Table 1. Experimental result of losses for multiclass classification derived for speaker and text verification. The losses are obtained for the test dataset consisting of input and reference sequences in the classes of speakers, $c_i^{[S]}$ for $i = -1, 0, 1$, and texts (digit sequences), $c_i^{[D]}$ for $i = 0, 1, 2, \cdots, 5$.

	class	L_{AVE_θ}		L_{ACE}		L_{NLP}		L_{NLPD}	
	index i	GEBI	BI	GEBI	BI	GEBI	BI	GEBI	BI
speaker verification	1	**0.002**	0.046	**0.002**	0.029	**223.2**	5549.5	**224.6**	6538.9
	0	**0.011**	0.050	**0.061**	0.748	**7730.9**	42052.3	**7730.8**	42308.0
	-1	**0.000**	0.000	0.033	**0.019**	**4420.7**	18796.5	**4427.7**	19332.4
	mean	**0.004**	0.032	**0.032**	0.265	**4125.0**	22132.8	**4127.7**	22726.5
text verification	5	**0.011**	0.012	**0.009**	0.015	**2708.1**	5331.8	**2736.4**	5669.5
	4	**0.180**	0.216	**0.844**	1.000	7378.3	**4161.6**	**7455.7**	8656.6
	3	**0.000**	0.581	**0.061**	0.774	**1293.7**	4844.1	**1476.9**	11600.1
	2	**0.000**	0.339	**0.043**	0.862	**911.8**	4694.2	**996.1**	12072.9
	1	**0.000**	0.187	**0.076**	0.705	**1291.0**	5768.7	**1323.5**	8475.6
	0	**0.000**	0.001	0.042	**0.004**	**1297.7**	3984.0	**1345.5**	4098.9
	mean	**0.032**	0.223	**0.179**	0.560	**2480.1**	4797.4	**2555.7**	8428.9

Flexible Verification Using Probabilistic Prediction For text-prompted speaker verification, we can use the class index thresholds $i_\theta^{[S]} = 1$ and $i_\theta^{[D]} = 4 \uplus 5$ for speaker and text verification, respectively. Here, however, when the probability $p\left(c_1^{[S]} \mid p_{\mathrm{G}}\right)$ or $p\left(c_{4\uplus 5}^{[D]} \mid p_{\mathrm{G}}\right)$ for an input sequence is not so bigger than 0.5, a decision maker has a possibility to ask additional question to obtain much larger or much smaller probability than 0.5.

For text verification, we can tune the threshold $i_\theta^{[D]}$ for accepting the input sequence satisfying $i \geq i_\theta^{[D]}$ indicating that more than or equal to $i_\theta^{[D]}$ correct digits out of N-length sequence are expected. Here, the tuning of $i_\theta^{[D]}$ in the present method is easier and understandable than in the previous method requiring the tuning of thresholds $p_\theta^{[D]}$ in (3). Therefore, as an example of application, the tuning of $i_\theta^{[D]}$ has a possibility to be flexibly used in verifying spoken digits of a specific speaker in a recorded tape, where we do not need high security level.

4 Conclusion

We have presented a method of probabilistic prediction for flexible verification without using thresholds for acceptance and rejection. After introducing multiclass classification for classifying the verification results, the method utilizes BI to obtain the probability. The method also uses loss functions for evaluating the performance of probabilistic prediction. By means of numerical experiments using recorded real speech data, we have examined the properties of the present method using GEBI and BI, and show the effectiveness and the risk of probability loss in the present method.

References

1. Kurogi, S.,Ueki, T.,Mizobe, Y. and Nishida, T.: Text-prompted multistep speaker verification using Gibbs-distribution-based extended Bayesian inference for reducing verification errors *Proc. ICONIP2013, Part III LNCS 8228*, pp.184–192 (2013)
2. Kurogi, S., Ueki, T.,Takeguchi, S., Mizobe, Y.: Properties of text-prompted multistep speaker verification using Gibss-distribution-based extended Bayesian inference for rejecting unregistered Speakers, *Proc. ICONIP2014, Part II, LNCS8835*, pp.35–43 (2014)
3. Beigi, H.: Fundamentals of speaker recognition. *Springer-Verlag New York Inc (C)* (2011)
4. Slingo, J., Palmer, T.: Uncertainty in weather and climate prediction, *Phil. Trans. R. Soc. A* Vol. 369, pp.4751–4767 (2011)
5. Kurogi, S., Ueno, T. and Sawa, M.: A batch learning method for competitive associative net and its application to function approximation. *Proc. SCI2004*, Vol. V, pp.24–28 (2004)
6. Kurogi, S., Mineishi, S. and Sato, S.: An analysis of speaker recognition using bagging CAN2 and pole distribution of speech signals. *Proc. ICONIP2010, Part I, LNCS 6443*, pp.363–370 (2010)
7. Campbell, J.P.: Speaker Recognition: A Tutorial. *Proc. the IEEE*, Vol. 85, No.9, pp.1437–1462 (1997)
8. Quiñonero-Candela, Joaquin, Dagan, Ido, Magnini, Bernard, d'Alche-Buc, Florence: Evaluating Predictive Uncertainty Challenge, Machine Learning Challenges.Evaluating Predictive Uncertainty,Visual Object Classification, and Recognising Tectual Entailment, vol.3944,pp.1-27(2006)