Speaker verification with multiple classifier fusion using Bayes based confidence measure

Fernando Huenupán, Nestor Becerra Yoma, Carlos Molina and Claudio Garretón

Speech Processing and Transmission Laboratory, Department of Electrical Engineering
Universidad de Chile, Santiago, Chile
nbecerra@ing.uchile.cl

Abstract

A novel framework based on Bayes-based confidence measure (BBCM) for Multiple Classifier System (MCS) fusion is proposed. As shown here, BBCM based MCS combination scheme corresponds to the ordinary Bayes fusion weighted by the reliability of each individual classifier. BBCM provides a formal model for heuristic weighting functions employed elsewhere. When compared with the ordinary Bayesian fusion, the proposed method leads to reductions as high as 20% and 50% in EER and the area below the ROC curve, respectively, in speaker verification.

Index Terms: speaker verification, multi-classifier system, confidence measure.

1. Introduction

In pattern recognition, the use of multiple classifier systems (MCS) has been addressed in several fields [8-10]. In speaker verification (SV), neural networks [4][13] and linear combination [11] are among the most popular approaches for MCS. Nevertheless, most of these techniques do not provide the required mathematical analysis to achieve the relevant performance evaluations [7]. In MCS, classifiers are usually combined in parallel into two levels [11]: abstract level (Fig.1) and score level. In Fig.1: O is the observed input signal; CLj is the classifier j, where 1 ≤ j ≤ J and J is the number of classifiers; Sj(O) is the score at the output of classifier j; dC is the local decision or classification due to classifier j; and, D(O) is the final decision or classification that corresponds to O. D(O) and dC indicate one of the M classes denoted by Cm, where 1 ≤ m ≤ M and M is the total number of classes.

In pattern recognition theory, the most straightforward formal strategy to fusion classifiers is the Bayes classification theory [9] (2)

D(O) = \arg \max \Pr(C_m | S(O)) = \arg \max \frac{\Pr(S(O) \mid C_m) \Pr(C_m)}{\sum \Pr(S(O) \mid C_m) \Pr(C_m)}

where \[ S(O) = \left[ S_{C_1}(O), \ldots, S_{C_J}(O), \ldots, S_{C_J}(O) \right]. \]

The a priori multivariable p.d.f.’s in (1), \Pr(S(O) \mid C_m), may require an unmanageable amount of training data to be reliably estimated [9]. The maximization of (1) can be simplified by a priori p.d.f.’s of individual classifiers. The classical techniques of simplification are [9-10]: Product Rule; Max Rule; Min rule; Mean Rule and, Majority Vote Rule (MVR). One of the most widely employed approximations is MVR [8], which combines local decision of individual classifiers as in Fig. 1. For this reason MVR is adopted in this paper as evaluation framework.

In [14], Bayes-based confidence measure (BBCM) was proposed to address the problem of assessing the accuracy in automatic speech recognition (ASR). BBCM is a probability itself and incorporates a priori information about the recognizer performance. Compared with standard confidence measures, BBCM improves the discrimination of misrecognized words. Surprisingly, the problem of assessing the accuracy of SV systems has not been exhaustively addressed in the specialized literature.

This paper proposes: a) a MCS strategy applied to SV by using simplifications of the Bayesian Fusion; b) a new classification criterion based on BBCM applicable to any pattern recognition problem; c) a new strategy for classifier fusion in MCS based on BBCM also applicable to any pattern recognition problem, and d) modeling the problem of confidence measure in SV by applying BBCM. As shown here, BBCM based MCS fusion scheme corresponds to the ordinary Bayes fusion weighted by the reliability of each individual classifier. Moreover, BBCM provides a formal model for heuristic weighting functions employed elsewhere.

Finally, the approach and analysis presented in this paper has not been found in the specialized literature.

2. Bayes classifier fusion applied to SV

In SV, the task is to decide about the user identity that claims a given identity, and two classes are possible: client, C1; and, impostor, C2. From the comparison of the input speech signal with each classifier CLj results a score \[ S_{CLj}(O) \], local decision \[ d_{CLj} \], \Pr\left[ S_{CLj}(O) \mid C_1 \right] \text{ and } \Pr\left[ S_{CLj}(O) \mid C_2 \right]. \] Consequently, the classifier array provides a set of scores \{S(O)\}, local decisions \{d(O)\}, \Pr[S(O) | C_1] \text{ and } \Pr[S(O) | C_2]. \] If the a priori probabilities of C1 and C2 are assumed uniformly distributed, then (1) can be written as:

\[
D(O) = \arg \max_n \Pr(C_n \mid S(O)) = \arg \max_n \frac{\Pr[S(O) \mid C_n]}{\sum \Pr[S(O) \mid C_n]} \tag{2}
\]

As mentioned above, (2) can be approximated by MVR. In MCS, MVR corresponds to a straightforward scheme that
combines the outputs of classifiers. In this paper, a weighted version of MVR, WMVR-MCS, is employed. WMVR-MCS is defined as [2][9]:

\[
D(O) = \sum_{j=1}^{N} \lambda_{C_l} \left[ \Pr\left( d_{C_l} \mid S_{C_l}(O) \right) \right]
\]

where \( \lambda_{C_l} = \begin{cases} 1 & \text{if } d_{C_1}(O) = C_i \\ -1 & \text{if } d_{C_1}(O) = C_i \end{cases} \)

3. BBCM in SV

BBCM was firstly proposed in the field of speech recognition. An ASR system receives a speech signal as input and delivers a set of recognized words (i.e. \( w_1, w_2, ..., w_n \)), where \( w_i \) denotes the \( i \)th word in the string. If \( WF \) denotes a word feature, BBCM is defined in ASR as [14]:

\[
BBCM(WF) = \frac{\Pr(w_i \text{ is OK}|WF)}{\Pr(w_i \text{ is OK})}
\]

where event “OK” corresponds to the fact that word \( w_i \), which is contained at least in one of the N-best hypotheses, was properly recognized. Notice that \( BBCM(WF) \) is a probability itself. Moreover, the distribution \( \Pr(w_i \text{ is OK}) \) and the probability \( \Pr(w_i \text{ is OK}) \) provide information about the recognition engine performance.

In SV, the word feature employed in ASR can certainly be replaced with the score of a given classifier. As a consequence, applying the definition of BBCM to the SV problem leads to [6]:

\[
BBCM[S_{C_l}(O)] = \frac{\Pr\left( d_{C_l} \text{ is OK}|S_{C_l}(O) \right)}{\Pr\left( d_{C_l} \text{ is OK} \right)}
\]

where “\( d_{C_l} \) is OK” corresponds to the decision of classifier \( C_l \) is correct. By considering two classes as mentioned above, (5) can be expressed as [6]:

\[
BBCM[S_{C_l}(O)] = \Pr\left( d_{C_l} \text{ is OK}|S_{C_l}(O) \right)
\]

where \( \text{OK} \) corresponds to the A posteriori model confidence as a probability, which in turn is not an heuristic result.

4. BBCM Based Classifier Fusion in SV

BBCM as defined in (10) could also be employed as the classification criterion that needs to be maximized to optimally decide about the recognized class.

4.1 BBCM as a classification criterion

According to (10), \( BBCM\left[S_{C_l}(O) \cap C_a\right] \) can be written as:

\[
BBCM\left[S_{C_l}(O) \cap C_a\right] = \Pr\left( O \in C_a \land d_{C_l}(O) = C_a \mid S_{C_l}(O) \right)
\]

where \( \Pr\left( O \in C_a \mid S_{C_l}(O) \right) = \Pr\left( C_a \mid S_{C_l}(O) \right) \) is the a posteriori probability that appears in the maximization according to the Bayes classification rules in (2); and, \( \Pr\left( d_{C_l}(O) = C_a \mid S_{C_l}(O), O \in C_a \right) \) corresponds to an additional information incorporated by BBCM that is related to the reliability of an individual classifier. As a consequence, it sounds reasonable to classify by selecting the class that maximizes \( BBCM\left[S_{C_l}(O) \cap C_a\right] \):

\[
d_{C_l} = \arg \max \left\{ BBCM\left[S_{C_l}(O) \cap C_a\right] \right\}
\]

Observe that (13) maximizes a confidence measure instead of the conventional a posteriori probability \( \Pr\left( C_a \mid S_{C_l}(O) \right) \). Nevertheless, BBCM is a probability itself and also incorporates \( \Pr\left( C_a \mid S_{C_l}(O) \right) \), besides the information on classifier reliability. Finally, the result denoted by (13) should be applicable to any classification problem independently of the number of classes.
4.2 Fusion with BBCM

According to (13), the local decision of classifier $j$ is the class that maximizes $BBCM[S(\cdot)|C_m]$. By considering the MCS environment in Fig. 1, the selected class corresponds to:

$$D(O) = \arg \max_{C_m} \left[ BBCM[S(O) \cap C_m] \right]$$

where $BBCM[S(O) \cap C_m]$ is estimated as:

$$BBCM[S(O) \cap C_m] = \frac{Pr[D(O) = C_m | O \in C_m] \cdot Pr[S(O) | O \in C_m]}{\sum_{i=0}^{K} Pr[D(O) = C_i | O \in C_i] \cdot Pr(C_i)}$$ (15)

In (15), $Pr[S(O) | O \in C_m] \cdot Pr[D(O) = C_m | O \in C_m]$ required to estimate $BBCM[S(O) | C_m]$ also demands a high amount of training data. Consequently, the same strategy adopted for Bayesian fusion (WMVR) is also used here. So, in BBCM fusion with WMVR, BBCM-WMVR, $Pr[d_{j,1} | S_j(O)]$ is replaced with $BBCM[S_j(O) \cap d_{j,1}(O)]$ in (3):

$$D(O) = \sum_{j=1}^{K} \Delta_{d_{j,1}} \cdot BBCM[S_j(O) \cap d_{j,1}(O)]$$ (16)

According to (12), $D(O)$ in (16) corresponds to the Bayes WMVR approximation weighted by the reliability of classifier $j$ given class $C_m$.

4.3 BBCM versus classic Bayesian score

BBCM incorporates information about the classifier performance, confidence, and the a posteriori probability itself. In contrast, Bayes classification employs only the a posteriori probability. Moreover, BBCM score is between 0 and 1 (see Fig.2), but the sum over the classes is not necessarily equal to one. On the other hand, in Bayes rule a posteriori probability. Moreover, BBCM score is between 0 and 1 (see Fig.2), but the sum over the classes is not necessarily equal to one. Consequently, the same strategy adopted for Bayesian fusion (WMVR) is also used here. So, in BBCM fusion with WMVR, BBCM-WMVR, $Pr[d_{j,1} | S_j(O)]$ is replaced with $BBCM[S_j(O) \cap d_{j,1}(O)]$ in (3):

$$D(O) = \sum_{j=1}^{K} \Delta_{d_{j,1}} \cdot BBCM[S_j(O) \cap d_{j,1}(O)]$$ (16)

According to (12), $D(O)$ in (16) corresponds to the Bayes WMVR approximation weighted by the reliability of classifier $j$ given class $C_m$.

5. Tuned weighting function (TWF)

As can be seen in (11), $BBCM[S_j(O) \cap C_m]$ is the sum of $BBCM[S_j(O) \cap C_m]$ over all classes. Usually, when two classes are employed:

$$BBCM[S_j(O) \cap C_m = d_{j,1}] >> BBCM[S_j(O) \cap C_m \neq d_{j,1}]$$

Consequently, the following approximation is considered:

$$BBCM[S_j(O) \cap C_m = d_{j,1}] >> BBCM[S_j(O) \cap C_m \neq d_{j,1}]$$ (17)

By replacing (17) in (16) BBCM-WMVR can be written as:

$$D(O) = \sum_{j=1}^{K} \Delta_{d_{j,1}} \cdot BBCM[S_j(O) \cap d_{j,1}(O)]$$ (18)

$BBCM[S_j(O) \cap d_{j,1}(O)]$ requires p.d.f.’s estimated with likelihood maximization. However, these a priori p.d.f.’s do not necessarily minimize the error rate. As shown in Fig.2, $BBCM[S_j(O) \cap d_{j,1}(O)]$ curve could be approximated by an inverted Gaussian curve. For this reason, an empirical tuned weighting function (TWF) is proposed and evaluated:

$$TWF[S_j(O)] = 1 - \frac{K_{j_{TWF}}}{\sigma_{j_{TWF}}} \cdot N[TH_{j_{TWF}}, \sigma_{j_{TWF}}]$$ (19)

where $K_{j_{TWF}}$ and $\sigma_{j_{TWF}}$ are estimated with a training database (Fig. 3), and $TH_{j_{TWF}}$ is the threshold of equal error rate (TEER) for classifier $j$. In (18) $BBCM[S_j(O) \cap C_m = d_{j,1}(O)]$ is replaced with $TWF[S_j(O)]$:

$$D(O) = \sum_{j=1}^{K} \Delta_{d_{j,1}} \cdot TWF[S_j(O)]$$ (20)

Figure 3: EER versus $K_{j_{TWF}}$ and $\sigma_{j_{TWF}}$ in TWF as defined in (19) with a forced Viterbi based SV system.

6. Experiments

The database is composed of 40 speakers (20 males and 20 females). The vocabulary corresponds to Spanish digits. Each speaker pronounced the 10-digit sequence “0-1-2-3-4-5-6-7-8-9” three times for enrolling. For verification, every speaker uttered the four-digit sequences “1-8-6-4”, “4-5-2-0” and “9-8-9” three times for enrolling. For verification, every speaker uttered the four-digit sequences “1-8-6-4”, “4-5-2-0” and “9-8-9” three times for enrolling. The enrollment and verification speech signals were recorded on the same telephone. The database was divided in two groups, A and B. Database A, composed of 30 speakers (15 males and 15 females), is used for testing. Database B, composed of 10 speakers (5 males and 5 females), is employed to estimate the required a priori p.d.f.’s, for BBCM and Bayes curves, and the parameters $K_{j_{TWF}}$ and $\sigma_{j_{TWF}}$ in (19). The database employed in this paper is similar to those mentioned elsewhere [13]. Enrolling and verification utterances are decomposed as a sequence of triphones. Thirty-three cepstral coefficients are computed per frame: the frame energy plus ten static coefficients and their first and second time derivatives. The HMM’s were trained with the Viterbi algorithm.

Three standard SV techniques are addressed: forced Viterbi based score (LL) [5]; Maximum Vote Rate for sequence of feature vectors, MVR-FV [12]; and, Support Vector Machines, SVM [2]. In LL system, the HMM’s were trained with the Viterbi algorithm. Each triphone was modeled with a three-state left-to-right HMM topology without skip-state transition, with one multivariate Gaussian density per state in speaker dependent models, and eight multivariate Gaussian densities per state in the speaker dependent but provide a sum over all the classes equal to one.

According to (12), $D(O)$ in (16) corresponds to the Bayes WMVR approximation weighted by the reliability of classifier $j$ given class $C_m$.
independent model. Both models employed diagonal covariance matrices. In MVR-FV, the size of window is equal to two frames and the threshold for local decisions is estimated with database B. In SVM, Gaussian kernel is employed [1] and K-means algorithm is used to group the training data with 256 and 512 codewords for client and impostor class, respectively. The optimal $k_{\alpha}$ and $\sigma_{\alpha}$ in (19) were estimated with the training database by minimizing EER: $k=0.08$, 0.003, 0.002 and $\sigma=0.0035$, 0.02 and 0.003 for LL, MVR-FV and SVM systems, respectively. FA/FR error rates are computed with database A. FR curves are estimated with 30 speakers x 9 verification signals per client = 270 signals; and, FA curves are computed by avoiding cross-gender impostor trials with (15 impostors) x 9 verification signals per impostor x 30 users = 3780 experiments. The baseline system in this paper is given by LL that is more accurate than MVR-FV and SVM. The baseline system gives an EER equal to 5.9%. The EER’s provided by MVR-FV and SVM are 7.42% and 14.82%, respectively. Results are presented in Table 1 and Fig. 4.

Table 1: EER and integral below the ROC curve with individual classifiers and MCS fusion approximation.

<table>
<thead>
<tr>
<th>Individual Classifier</th>
<th>MCS Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>LL</td>
<td>MVR</td>
</tr>
<tr>
<td>SVM</td>
<td>WMVR Eq. (3)</td>
</tr>
<tr>
<td>Bayes</td>
<td>WMVR Eq. (16)</td>
</tr>
<tr>
<td>SVM</td>
<td>WMVR Eq. (18)</td>
</tr>
<tr>
<td>approx LL</td>
<td>BBCM</td>
</tr>
<tr>
<td>WMVR Eq. (19)</td>
<td>WMVR</td>
</tr>
<tr>
<td>WFMVR Eq. (20)</td>
<td>WMVR</td>
</tr>
</tbody>
</table>

EER: 5.93, 7.42, 14.82, 4.07, 3.43, 3.58, 3.00, 3.00

ROC: 258, 229, 782, 81, 41, 48, 38, 38

Table 1: EER and integral below the ROC curve with individual classifiers and MCS fusion approximation.

Figure 4: DET curves with LL, Bayes-WMVR, BBCM-WMVR and TWF-WMVR.

7. Discussion and conclusions

A new framework based on BBCM is presented for MCS fusion. The BBCM fusion scheme corresponds to ordinary Bayes combination weighted by the reliability of each classifier. Also, BBCM provides a formal model instead of the heuristic weighting functions employed elsewhere. As can be seen in Table 1 and Fig.4, the BBCM-WMVR fusion method defined in (16) outperformed the baseline system and the ordinary Bayes-WMVR fusion approximation. Compared with the baseline system, BBCM-WMVR leads to reductions in EER and in the integral below the ROC equal to 42% and 65%, respectively. When compared with the ordinary Bayes-WMVR fusion, BBCM-WMVR gives reductions of 16% and 41% in EER and in the integral below the ROC, respectively. The approximation of BBCM-WMVR according to (18) provide slightly lower improvements than BBCM-WMVR in (16). However, the BBCM based empirical weighting function, TWF-WMVR, leads to reductions as high as 26% and 53% in EER and the area below ROC, respectively, when compared with Bayes-WMVR.

Although tested with a SV task on the telephone line, the approaches presented here should easily be generalized to any classification problem independently on number of classes. Finally the applicability of confidence based MCS fusion to SV in mismatch conditions and to other classification problems is proposed as future work.

8. Acknowledgments

This work was funded by Conicyt-Chile under grants Fondef N° D05I-10243 and Fondecyt N° 1070382.

9. References