Odyssey 2014:
The Speaker and Language Recognition Workshop
16-19 June 2014, Joensuu, Finland

i-Vector Selection for Effective PLDA Modeling in Speaker Recognition

Sangeeta Biswas, Johan Rohdin, and Koichi Shinoda

Tokyo Institute of Technology, Japan
{sangeeta,johan}@ks.cs.titech.ac.jp, shinoda@cs.titech.ac.jp

Abstract

Data selection is an important issue in speaker recognition. In previous studies, the data selection for universal background model (UBM) training and for the background dataset of support vector machines (SVM) have been addressed. In this paper, we address the data selection for a probabilistic linear discriminant analysis (PLDA) model which is one of the state-of-the-art methods for i-vector scoring. We first show that the data selection using the conventional k-NN method indeed improves the speaker verification performance. We then propose a robust way of selecting k by using a *local distance-based outlier factor* (LDOF). We name our method as *flexible* k-NN or *fk*-NN. Our *fk*-NN obtained significant performance improvements on both male and female trials of the NIST speaker recognition evaluation (SRE) 2006 core task, NIST SRE 2008 core task (condition-6) and NIST SRE 2010 coreext-coreext task (condition-5).

1. Introduction

A background model plays an important role in speaker verification either by providing a prior distribution for the parameters of the speaker-specific model or by acting as an alternate model for scoring of a trial. In previous studies, it has been shown that for UBM training [1, 2] and SVM background models [3, 4], relevant training data is more important for better performance than the amount of training data. In this paper, we address the relevant data selection issue for PLDA [5] modeling.

PLDA is one of the state-of-the-art methods for separating speaker factors of i-vectors[6] from irrelevant factors such as the transmission channels or the speaker's emotion. In order to train parameters of PLDA models, multi-session recordings from several hundred speakers, resulting in several thousands of recordings, are typically used. For example, research groups involved in the NIST speaker recognition evaluation (SRE) typically use utterances from all NIST 2004-2005 data along with the Switchboard II, Phases 1, 2 and 3; Switchboard Cellular, Parts 1 and 2 data and Fisher data. However, there is no evidence that using all the data available would guarantee the best PLDA model. Based on the experiences from the other models such as UBM, SVM or joint factor analysis (JFA), researchers typically use gender-dependent PLDA models. In [7], it was empirically shown that gender-dependent PLDA models outperformed gender-independent PLDA models. Kanagasundaram *et al.* [8] showed that the PLDA model trained by utterances whose lengths matched with those utterances in the evaluation set performed better than that trained by full-length utterances. These studies indicate that if we have information about the target evaluation set, it is better to take into account that information for selecting training data of PLDA models.

In many NIST SREs, we cannot access information about the whole evaluation set consisting of speakers enrolled to the system, i.e., *enrollment set*, and speakers participated in the authentication phase, i.e., *test set*, during the development phase of the system. However, in many applications such as on-line bank services for registered customers, we can access the enrollment set during the development phase of the system. Targeting that kind of applications, we use the enrollment set for selecting suitable training data for the PLDA model in this study.

We show that by selecting a training set whose i-vectors are close to the i-vectors of the enrollment set, we can improve the PLDA modeling. We first use the conventional k-NN method in order to choose the k-nearest neighbors of each enrollment speaker in the training set of the PLDA model. We show that this method performs remarkably well when the optimal k is known. However, it is difficult to estimate the optimal k. We, therefore, propose a robust way of selecting k which uses *local distance-based outlier factor* (LDOF).

Although our data selection method does not offer an impressive reduction in computational expense as some data selection methods for UBM training, it helps improving the verification accuracy. Since the training time of the PLDA model is very small (a couple of seconds), we can re-train the PLDA model quickly after adding relevant data for newly added enrollment speakers, which is not practical for UBM training or total variability matrix training.

Using the enrollment set for system development was allowed in the NIST SRE 2012 but not in the earlier NIST SREs. However, we evaluated our proposed data selection method for the NIST SRE 2006 core task, NIST SRE 2008 core task (condition-6) and NIST SRE 2010 coreext-coreext task (condition-5), because we wanted to focus on the data selection process rather than paying extra attention on processing the huge amounts of SRE 2012 data. Our experiments showed that our method obtained significant performance improvements on both male and female trials.

2. i-vector and PLDA based speaker verification

In an i-vector based speaker verification system [6], it is assumed that a GMM-*supervector*, $\boldsymbol{\mu}$, corresponding to an utterance can be modeled as

$$\boldsymbol{\mu} = \bar{\boldsymbol{\mu}} + \boldsymbol{T}\boldsymbol{\omega}, \quad (1)$$

where $\boldsymbol{\omega}$ is a random vector known as the *i-vector*, $\boldsymbol{T}$ is a basis for the *total variability space* for speaker and channel variability of $\boldsymbol{\mu}$, and $\bar{\boldsymbol{\mu}}$ is the mean of $\boldsymbol{\mu}$. It is assumed that $\boldsymbol{\omega}$ follows a standard normal distribution and its dimension, d, i.e., the rank of $\boldsymbol{T}$, is lower than that of $\bar{\boldsymbol{\mu}}$.

In [9], it was proposed to use PLDA in speaker verification with i-vectors as features. In that study, a modification of the original PLDA model [5], suitable for low-dimensional features was suggested. According to that modification, i-vectors, $\boldsymbol{\omega}$, can be modeled as

$$\boldsymbol{\omega} = \boldsymbol{m} + \boldsymbol{V}\boldsymbol{y} + \boldsymbol{\epsilon}, \tag{2}$$

where $\boldsymbol{m}$ is the mean of $\boldsymbol{\omega}$ and $\boldsymbol{y}$ is a random variable depending on speaker factors. The elements of $\boldsymbol{y}$ follow standard normal distribution and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$. $\boldsymbol{V}$ is a basis for the *between-speaker subspace*.

3. i-Vector selection

Given the i-vectors of the enrollment set, we would like to find i-vectors in the training set that are useful for learning a good PLDA model, *i.e.*, a model that clearly separates speaker and channel effects of the i-vectors in the enrollment set as well as in the unseen test set. Our assumption is that the i-vectors in the training set of the PLDA model which have similar characteristics as in the enrollment set will be a better choice for training a good PLDA model over the whole training set. Further, we assume that i-vectors that have high cosine similarity have similar characteristics. In [6], it was shown that cosine similarity can be applied to decide whether two i-vectors are from the same speaker or not. Session compensation techniques like linear discriminant analysis (LDA), within-class covariance normalization (WCCN), and nuisance attribute projection (NAP) were applied in order to keep only the information about the speaker factors. However, we are more interested in finding i-vectors that are similar in the total variability space. Therefore, we do not apply any session compensation techniques before our i-vector selection process. Empirically, we show that k-nearest neighbors (k-NN) based on cosine similarity can find the best training set of a PLDA model (Table 4).

In the conventional k-NN, the optimum value of k may vary between different enrollment sets. If we use a k optimized for a different enrollment set, we sometimes end up in the situation that the selected i-vectors in the training set are much closer to each other than to the enrollment i-vector. In other words, the enrollment i-vector is an *outlier* compared to its k-nearest neighbors. In such situations, those neighbors cannot be expected to be similar to the enrollment i-vector. In order to solve this problem, we propose a modification of the k-NN method which we denote *flexible* k-NN (fk-NN). In this method we use the local distance outlier factor (LDOF) defined in the next section to measure to what extent the enrollment i-vector lies inside the cluster made by its k-nearest neighbors. We then increase k until all enrollment i-vectors lie inside the clusters of nearest neighbors according to the LDOF criteria.

3.1. LDOF

In [10], Zhang *et al.*, proposed the *local distance-based outlier factor* or *LDOF*. LDOF captures the degree to which an object deviates from its neighborhood system. When LDOF of any object is smaller than a threshold, t, we can say that object is surrounded by a data cloud. In data mining applications, LDOF is used for capturing the outlierness of an object among a scattered neighborhood. In this paper, we use it to control the value of k in the k-NN based data selection process by checking how far any i-vectors of the enrollment set lie from their k-nearest training i-vectors of the PLDA model.

Let $\mathcal{N}_p$ be the set of the k-nearest neighbors of the enrollment i-vector ω_p. The LDOF of an i-vector, ω_p, can be defined as

$$\mathrm{LDOF}_k(\omega_p) = \frac{\bar{d}_{\omega_p}}{\bar{D}_{\omega_p}}, \tag{3}$$

where $\bar{d}_{\omega_p}$ is the k-NN distance of ω_p and $\bar{D}_{\omega_p}$ is the k-NN inner distance of ω_p defined as

$$\bar{d}_{\omega_p} = \frac{1}{k} \sum_{\omega_i \in \mathcal{N}_p} \mathrm{dist}(\omega_i, \omega_p), \tag{4}$$

$$\bar{D}_{\omega_p} = \frac{1}{k(k-1)} \sum_{\omega_i, \omega_j \in \mathcal{N}_p, i \neq j} \mathrm{dist}(\omega_i, \omega_j), \tag{5}$$

3.2. Algorithm of fk-NN

Let the i-vector sets for training the PLDA model and enrollment speakers be $\mathcal{R}$ and $\mathcal{S}$, respectively. The steps of our data selection process using *fk*-NN are given as :

1. Estimate $\mathrm{dist}(\omega_i, \omega_p)$ where $\omega_i \in \mathcal{R}$ and $\omega_p \in \mathcal{S}$.
2. Sort $\mathrm{dist}(\omega_i, \omega_p)$ in the ascending order.
3. Estimate $\mathrm{dist}(\omega_i, \omega_j)$ where $\omega_i, \omega_j \in \mathcal{R}$ and $i \neq j$.
4. Initialize k.
5. Put the k-nearest neighbors of ω_p from $\omega_i \in \mathcal{R}$ into $\mathcal{N}_p$.
6. For each $\omega_p \in \mathcal{S}$, do
   (a) Estimate $\bar{d}_{\omega_p}$ by using Eq. (4).
   (b) Estimate $\bar{D}_{\omega_p}$ by using Eq. (5).
   (c) Estimate $\mathrm{LDOF}_k(\omega_p)$ by using Eq. (3).
   (d) i. If $\mathrm{LDOF}_k(\omega_p) \geq t$, then
      - $k = k + \delta$
        where δ is any integer value for increasing the value of k.
      - Go to Step-5.

      ii. Else, go to the Step-6a for the next i-vector.
7. $\mathcal{N}_p$ will be the cloud of k-nearest neighbors of ω_p for each p. Take unique set of i-vectors from $\{\mathcal{N}_p\}_{p \in \mathcal{S}}$ to make relevant training data of PLDA model.

The initial value of k can be set to 2. Our proposed fk-NN will increase k until all i-vectors of the enrollment set are surrounded by their neighborhood. With fk-NN, it is possible to set different k for different i-vectors of the enrollment set. However, if initial k is too small for all enrollment i-vectors, then the selected training set will end up having insufficient number of i-vectors. By controlling t we can solve this problem. However, for simplicity, in this paper, we set t equal to 1 and keep k the same for all i-vectors of an enrollment set in our experiments.

4. Experiments

4.1. Experimental setup

We conducted experiments on the NIST SRE 2006 core task (SRE06), NIST SRE 2008 core task condition-6 (SRE08) and NIST SRE 2010 core task condition-5 (SRE10). The number of enrollment i-vectors, E, and the number of trials, R, of three evaluation sets that we used after discarding corrupted files, are shown in Table 1.

For UBM and $\boldsymbol{T}$ training, NIST SRE 2004 (sre04) and NIST SRE 2005 (sre05), Switchboard II Phase 1(sb2p1), Switchboard II Phase 2 (sb2p2) and Switchboard II Phase 3 (sb2p3), Switchboard Cellular Part 1 (sbCp1) and Switchboard Cellular Part 2 (sbCp2) were used. For PLDA training we used the same sets with the number of i-vectors as shown in Table 2.

For features we used 15 PLP coefficients and log-energy plus their first-order and second-order derivatives, resulting in 48 features per frame. For removing non-speech, we used a

Table 1: *The number of enrollment i-vectors, E, and the number of trials, R, of three evaluation sets, SRE06, SRE08 and SRE10.*

Eval set	Male		Female	
	$\#E$	$\#R$	$\#E$	$\#R$
SRE06	349	22123	459	28945
SRE08	648	12356	1140	22957
SRE10	1906	179338	2361	236781

Table 2: *The number of i-vectors, V, and the number of speakers, S, of datasets used for training gender-dependent PLDA models.*

Dataset	Male		Female	
	$\#V$	$\#S$	$\#V$	$\#S$
sb2p1	391	125	455	191
sb2p2	1868	283	2134	307
sb2p3	1399	277	1921	337
sbCp1	236	78	290	94
sbCp2	1038	157	1595	232
sre04	1906	126	2651	188
sre05	2705	245	3792	336

spectral subtraction-based voice activity detector (VAD) [11]. We applied feature warping [12] before applying VAD. We used gender-dependent systems. The dimension of the i-vector, d, was set to 400. The i-vectors were centered, whitened, and length-normalized as proposed in [13] prior to PLDA training. We trained the parameters of the PLDA model by the ML criteria. The rank of $\boldsymbol{V}$ was set to 250. We used *equal error rate* (EER) and the minimum value of the normalized *detection cost function* (DCF) as evaluation metrics. We used DCF as defined in the evaluation plans of NIST SRE 2006, 2008, and 2010 [14, 15, 16] for SRE06, SRE08 and SRE10, respectively.

4.2. Results

First we tried to find a relevant database for the SRE06, SRE08 and SRE10 evaluation sets. We trained two gender-dependent PLDA models for each database. For sb2p1 and sbCp1, PLDA training failed due to an insufficient amount of training data. This problem can be avoided by applying regularization to the channel covariance during PLDA training, but we did not attempt this in this study. From the results shown in Table 3, we can assume that SRE04 and SRE05 have more relevant data for the male enrollment speakers of SRE06 and SRE08 than the Switchboard databases. Using only NIST SRE data (allsre), we got the lowest EER and the minimum DCF. It reveals that using all the data available did not guarantee the best PLDA model for the target evaluation set. The presence of unrelevant data in the training set of the PLDA model may deteriorate the system's performance. However, for SRE10, we got the lowest EER and the minimum DCF when we combined Switchboard data with NIST SRE data (sre+sb). The same phenomena was also seen for female trials. It indicates that relevant data differs in different target evaluation sets.

Instead of checking each combination of different training sets one by one, we tried to select relevant data from the whole training set of the PLDA model by choosing k neighbors in three ways. At first we optimized k for the k-nearest neighbors

Table 3: *EER and minimum DCF of male trials of SRE06, SRE08 and SRE10. Empty entries mean that PLDA training failed due to insufficient amount of training data. For SRE06 and SRE08, DCF is in 10^{-2} whereas for SRE10, DCF is in 10^{-4}. For all tasks, EER is in %.*

Data	SRE06		SRE08		SRE10	
	EER	DCF	EER	DCF	EER	DCF
sb2p1	-	-	-	-	-	-
sb2p2	13.26	5.54	13.26	5.54	17.96	8.56
sb2p3	14.30	5.65	14.31	5.65	18.19	8.78
sbCp1	-	-	-	-	-	-
sbCp2	15.16	5.79	15.16	5.79	12.47	9.81
allsb	8.17	3.86	9.65	4.83	5.38	6.54
sre04	6.96	3.29	6.96	3.29	4.88	7.51
sre05	5.73	2.99	5.74	2.99	2.81	5.65
allsre	**2.07**	**0.97**	**4.58**	**2.36**	2.28	4.20
sre+sb	2.30	1.16	4.92	2.55	**2.01**	**3.73**

Table 4: *The performance of the gender-dependent PLDA models trained by selecting k neighbors of the enrollment set in the training set for SRE06. k-NN: k-nearest neighbors were selected, k-FN: k-farthest neighbors were selected, and k-RN: k neighbors were selected randomly. For male, $k = 37$, and for female, $k = 25$. EER is in % and DCF is in 10^{-2}.*

Data	Male		Female	
	EER	DCF	EER	DCF
Baseline	2.30	1.16	3.42	1.85
k-NN	**1.84**	**1.05**	**2.71**	**1.43**
k-FN	2.47	1.30	4.02	2.27
k-RN	2.37	1.20	3.56	1.96

(k-NN) case and it was 37 for male and 25 for female trials of SRE06, respectively. Then we chose the k-farthest neighbors (k-FN) and the k-random neighbors (k-RN). We used the system trained by the whole training set as the *baseline* in order to compare the performance of k-neighbor based systems. As shown in Table 4, the PLDA model trained either by k-FN or by k-RN selected data was worse than the k-NN based PLDA model. This result indicates that i-vectors of the training set of the PLDA model that were close to the enrollment set were more relevant than any other i-vectors; especially more relevant than those far from the enrollment set.

As shown in Fig. 1 and Fig. 2, the optimum k varies from data set to data set. For SRE06 male, the optimum value of k was 37, whereas for SRE10, the optimum k was 18. As shown in Fig. 3, if we choose 37 as the value of k for SRE10, we would end up covering almost the whole training set. On the other hand, if we choose k around 18, for SRE06 we would not get a sufficient amount of i-vectors for training a good PLDA model. One of the reasons behind this phenomena may be that there is a large difference between the number of enrollment i-vectors of SRE06 and SRE10. SRE06 has 349 male enrollment i-vectors, whereas SRE10 has almost three times more enrollment i-vectors.

From Table 5, we can see that data selection by either k-NN or fk-NN improved the PLDA model. k-NN performed remarkably well on the development set, SRE06, where k was

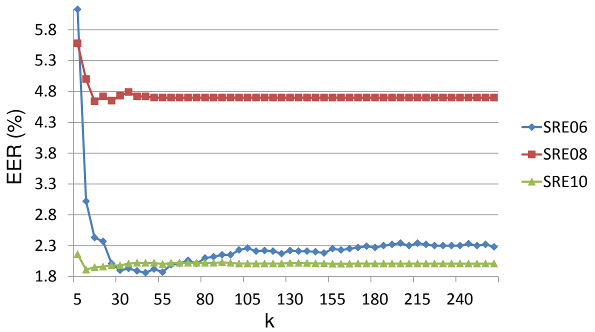

Figure 1: *The effect of the value of k on the EER(%) of male trials of SRE06, SRE08 and SRE10.*

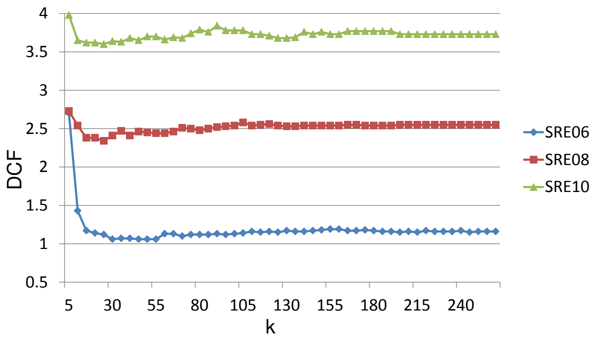

Figure 2: *The effect of the value of k on the DCF of male trials of SRE06, SRE08 and SRE10.*

Table 5: *EER and DCF for baseline, k-NN and fk-NN. In k-NN, k was optimized considering SRE06 as the development set. For male, $k = 37$, and for female, $k = 25$. For SRE06 and SRE08, DCF is in 10^{-2} whereas for SRE10 DCF is in 10^{-4}. For all tasks EER is in %.*

Male	SRE06		SRE08		SRE10	
	EER	DCF	EER	DCF	EER	DCF
Baseline	2.30	1.16	4.92	2.55	2.01	3.73
k-NN	1.84	1.05	4.76	2.44	2.05	3.68
fk-NN	2.08	1.12	4.73	2.43	**1.92**	**3.53**
Female	SRE06		SRE08		SRE10	
	EER	DCF	EER	DCF	EER	DCF
Baseline	3.42	1.85	5.97	2.85	3.02	4.94
k-NN	2.71	1.43	5.81	2.82	2.93	4.74
fk-NN	2.71	1.43	5.78	2.84	2.91	4.74

optimized. fk-NN, which did not need any tuning, was better or equal to k-NN on all evaluation sets and equally good on the development set for female.

From Table 6, we can see that for the male, fk-NN reduced the training data more than k-NN. For female, the reduction is similar. We can conclude that fk-NN finds relevant data without the need for parameter tuning.

Table 7 shows that we can get k close to the optimum k in k-NN by using fk-NN. Finding optimum k by using k-NN for each evaluation set was more time consuming than selecting k by fk-NN. For k-NN, we need to train the PLDA model

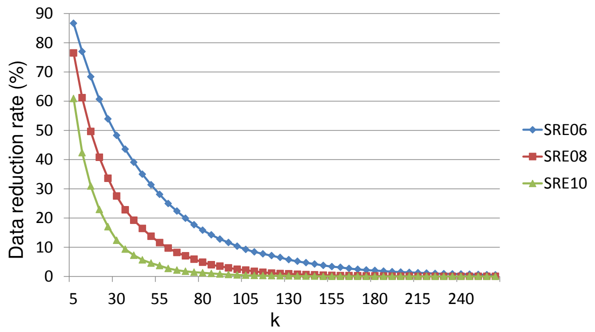

Figure 3: *The effect of the value of k on the data reduction rate of training data of PLDA model by k-NN method for male trials of SRE06, SRE08 and SRE10.*

Table 6: *Data reduction rate of k-NN and fk-NN. In k-NN, k was optimized considering SRE06 as the development set. For male, $k = 37$, and for female, $k = 25$.*

Male	SRE06	SRE08	SRE10
Baseline	0	0	0
k-NN	41.69	21.27	8.34
fk-NN	55.16	28.64	25.77
Female	SRE06	SRE08	SRE10
Baseline	0	0	0
k-NN	57.90	26.31	14.66
fk-NN	57.90	22.96	15.70

Table 7: *Optimal k and processing time. k finding by k-NN and fk-NN for male and female systems for SRE06, SRE08, and SRE10. Time is in sec.*

Male	SRE06		SRE08		SRE10	
	k	Time	k	Time	k	Time
k-NN	37	593	28	1298	18	1092
fk-NN	24	35	29	64	18	64
Female	SRE06		SRE08		SRE10	
	k	Time	k	Time	k	Time
k-NN	25	906	26	2656	22	1671
fk-NN	25	61	28	136	24	137

every time after setting k in the range 10-45, whereas for fk-NN, we trained the PLDA model only once after deciding k by LDOF. The experiments were performed on a 3.4GHz Intel i7-3770 CPU.

As shown in Fig. 4, k-NN and fk-NN selects data from all data sets. The most noticable trend is that SRE08 and SRE10 use much more of the Switchboard corpora than SRE06. In particular, they use almost all of swbCp1. Table 8 shows the number of speakers and the number of i-vectors in the original training data and the data selected by fk-NN. Overall, fk-NN reduces the number of i-vectors more than the number of speakers.

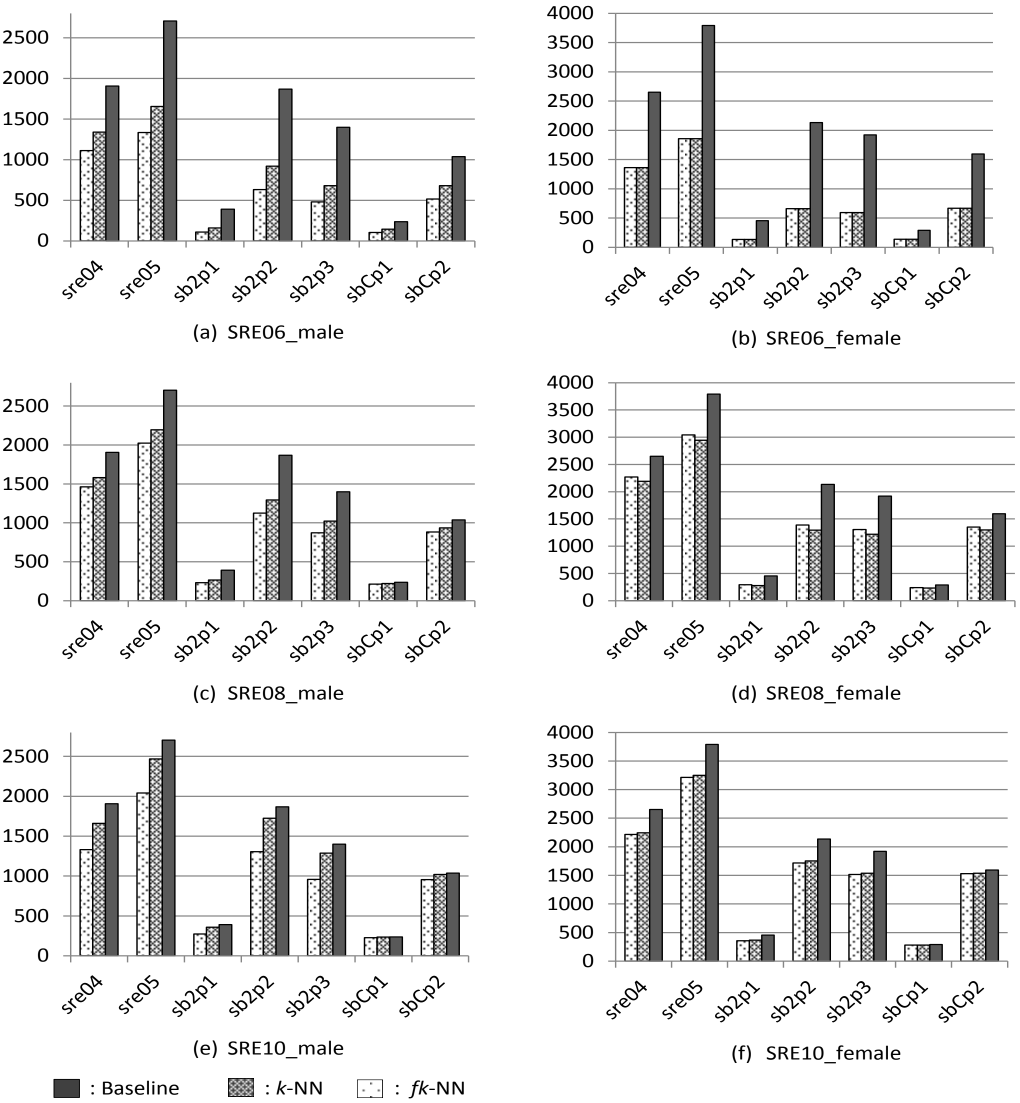

Figure 4: *y-axis shows the number of i-vectors used by the baseline, k-NN and fk-NN based systems for male and female trials of SRE06, SRE08 and SRE10.*

5. Conclusions

In this paper, we proposed a data selection method for a PLDA model, which is one of the state-of-the-art methods for i-vector scoring. We showed that by using k-NN we can reduce the amount of training data substantially and improve the system performance for both male and female trials of the NIST SRE 2006 core task, NIST SRE 2008 core task (condition-6) and NIST SRE 2010 coreext-coreext task (condition-5). In order to avoid the difficulty of optimizing k on a development set, we developed a robust way of selecting k, named fk-NN, which uses a *local distance-based outlier factor* (LDOF). Our proposed fk-NN reduced the amount of training data as much as the conventional k-NN without depending on any parameter tuning. In NIST SRE 2010 male trials, fk-NN performed significantly better than the conventional k-NN.

In the future, it would be interesting to see whether it is possible to replace a gender-dependent PLDA model by a gender-independent PLDA model which is trained using data selected by our proposed fk-NN. In that case, we would be able to get more relevant data for the PLDA model for the enrollment speakers who are closer to the opposite gender. Our proposed data selection method does not depend on any channel compensation techniques as used in i-vector based systems. Therefore, it would be a good idea to explore whether an i-vector based system using methods such as WCCN, NAP, LDA etc., can be benefited by our data selection procedure. We should also explore how much data is required for training an efficient PLDA model. Taking the number of times an i-vector from the training set was selected by our proposed method into account can also be an interesting consideration for the future.

Table 8: *Comparison of the data in the whole training set and the training set selected by fk-NN. #S: number of speakers, #SS: number of selected speakers by fk-NN, #V: number of i-vectors and #SV: number of selected i-vectors by fk-NN.*

Male	SRE06	SRE08	SRE10
#S	1278		
#SS	1029	1187	1224
#V	9543		
#SV	4279	6810	7084
Female	**SRE06**	**SRE08**	**SRE10**
#S	1665		
#SS	1328	1569	1615
#V	12838		
#SV	5404	9891	10822

6. References

[1] C-L. Huang and B. Ma, "Maximum entropy based data selection for speaker recognition," in *INTERSPEECH*, pp. 2713–2716, 2011.

[2] T. Hasan and J. H. L. Hansen, "A study on universal background model training in speaker verification," *IEEE Transactions on Audio, Speech & Language Processing*, vol. 19, no. 7, pp. 1890–1899, 2011.

[3] M. L. McLaren, R. J. Vogt, B. J. Baker, and S. Sridharan, "Data-driven background dataset selection for SVM-based speaker verification," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 18, no. 6, pp. 1496–1506, 2010.

[4] J-W. Suh, Y. Lei, W. Kim, and J. H. L. Hansen, "Effective background data selection in SVM speaker recognition for unseen test environment: More is not always better," in *ICASSP*, pp. 5304–5307, 2011.

[5] S. J. D. Prince and J. H. Elder, "Probabilistic linear discriminant analysis for inferences about identity," *IEEE International Conference on Computer Vision*, pp. 1–8, 2007.

[6] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *IEEE Transactions on Audio, Speech & Language Processing*, vol. 19, no. 4, pp. 788–798, 2011.

[7] M. Senoussaoui, P. Kenny, N. Brmmer, E. de Villiers, and P. Dumouchel, "Mixture of plda models in i-vector space for gender-independent speaker recognition," in *INTERSPEECH*, pp. 2528, 2011.

[8] A. Kanagasundaram, R. Vogt, D. Dean, and S. Sridharan, "Plda based speaker recognition on short utterances," in *Odyssey Speaker and Language Recognition Workshop*, Singapore, 2012.

[9] P. Kenny, "Bayesian speaker verification with heavy-tailed priors," in *Odyssey Speaker and Language Recognition Workshop*, 2010.

[10] K. Zhang, M. Hutter, and H. Jin, "A new local distance-based outlier detection approach for scattered real-world data," pp. 813–822, 2009.

[11] M. W. Mak and H. B. Yu, "Robust voice activity detection for interview speech in NIST speaker recognition evaluation," in *APSIPA ASC*, 2010.

[12] J. Pelecanos and S. Sridharan, "Feature warping for robust speaker verification," in *Odyssey Speaker Recognition Workshop*, pp. 213–218, 2001.

[13] D. Garcia-Romero and C. Y. Espy-Wilson, "Analysis of i-vector length normalization in speaker recognition systems," in *INTERSPEECH*, pp. 249–252, 2011.

[14] NIST, "The NIST year 2006 speaker recognition evaluation plan," Available: `http://www.itl.nist.gov/iad/mig//tests/sre/2006/sre-06_evalplan-v9.pdf`.

[15] NIST, "The NIST year 2008 speaker recognition evaluation plan," Available: `http://www.itl.nist.gov/iad/mig//tests/sre/2008/sre08_evalplan_release4.pdf`.

[16] NIST, "The NIST year 2010 speaker recognition evaluation plan," Available: `http://www.itl.nist.gov/iad/mig//tests/sre/2010/NIST_SRE10_evalplan.r6.pdf`.