Text-Dependent Speaker Recognition with Long-term Features Based on Functional Data Analysis

Chenhao Zhang

2012/11/18

Outline

- Introduction
- MFCC-FDA Feature Extraction Framework
- The Distance Measures
- Experiments
- Conclusions

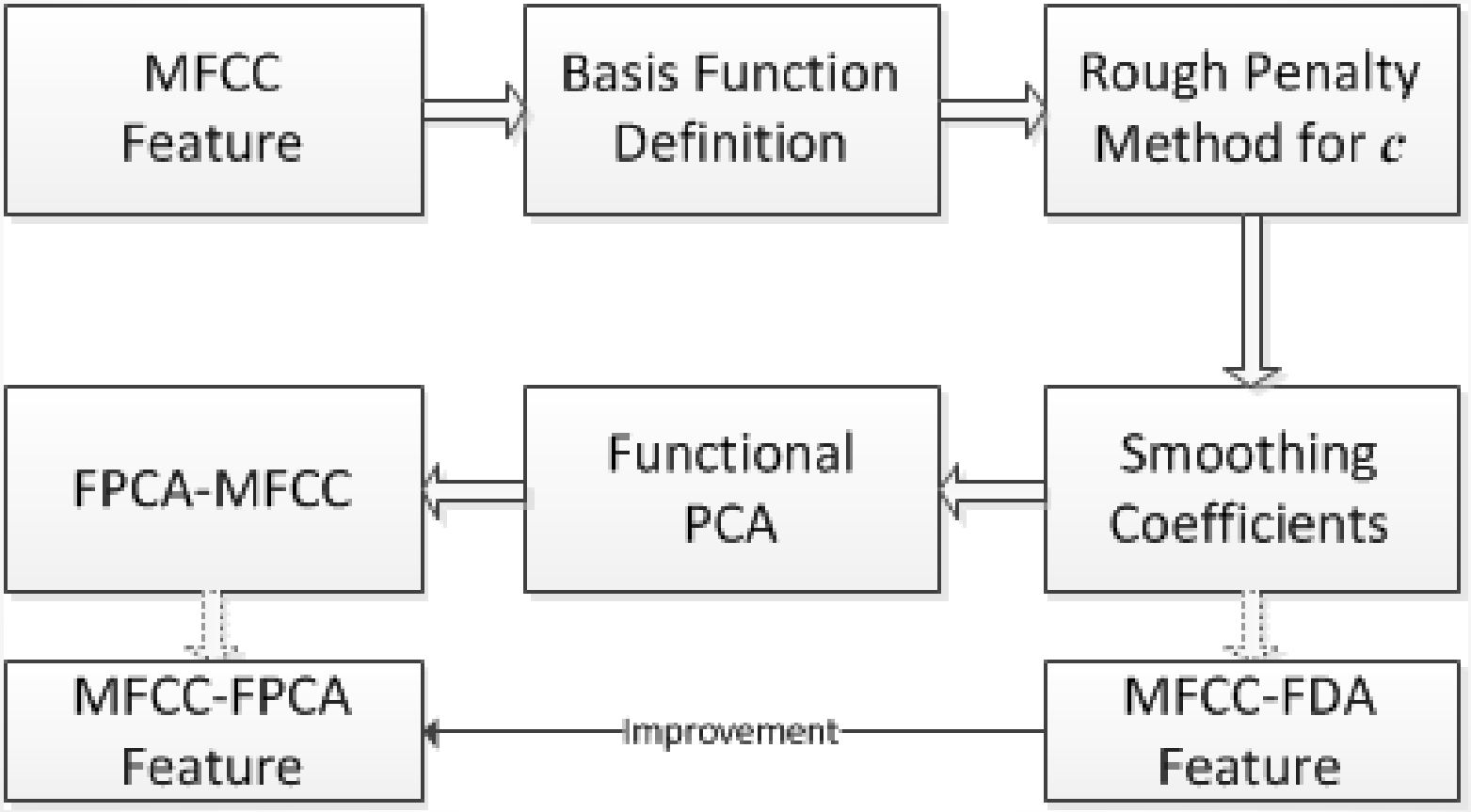
Introduction

- Text-Dependent Speaker Recognition (TDSR)
  - Speech feature dynamics are exploited to identify different speakers
  - Dynamic Time Warping (DTW)/HMM
- Speech feature vector plays a key role in TDSR
  - The short-term speech feature, like MFCC
    1. Capture the highly local portion of the significant temporal dynamics
    2. Not reflect some certain overall statistical characteristics hidden behind the sentences

Introduction

- Functional Data Analysis (FDA)
  - Formulate the problem to statistical thinking and analysis
  - The way to combine the short-term features and the extra information obtained by looking at all the data as a whole
  - FDA theory shows good performance on the speech feature analysis and speech re-synthesis
    - Jim Ramsay and Bernard Silverman introduced FDA in the late 1990's
    - M. Gubian, "Functional Data Analysis for Speech Research", InterSpeech 2011

MFCC-FDA Feature Extraction Framework

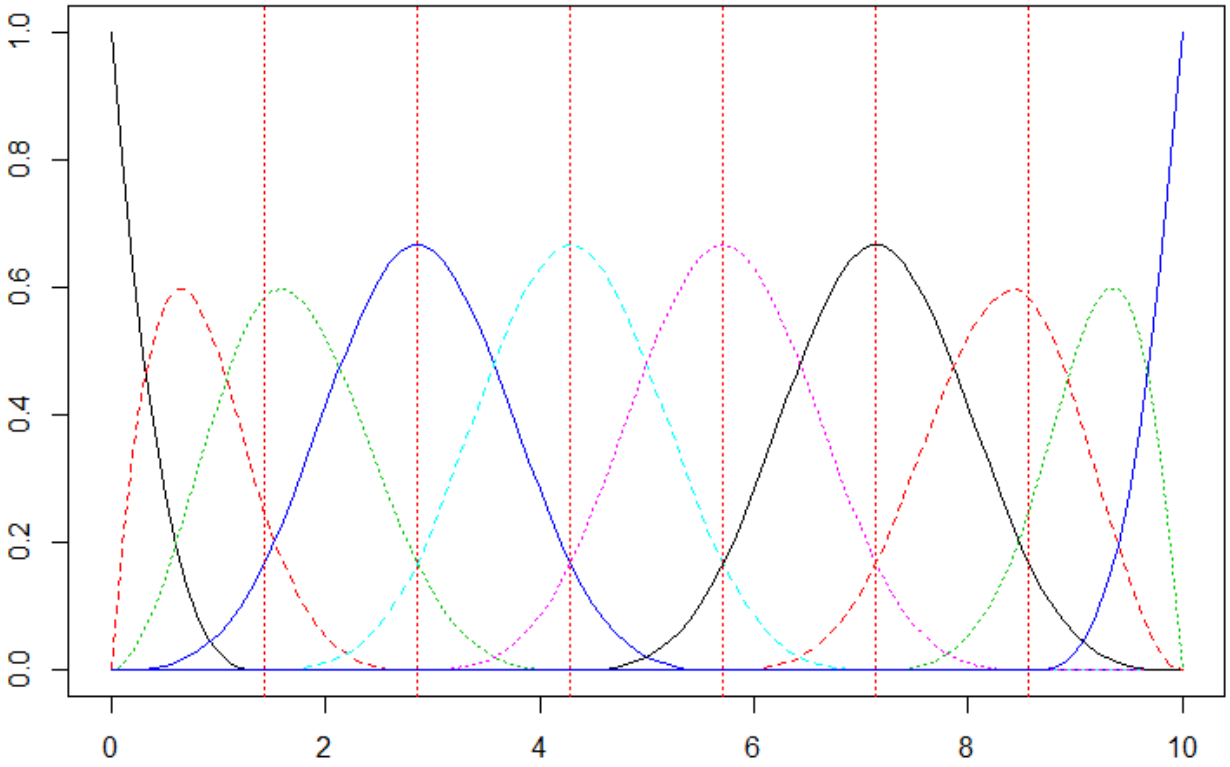


The Basis Functions

- Th… d
li… nt
th… s

ar…
…, c…

- Th… ate
th… e:
ϕ…



The Calculation of Coefficients c

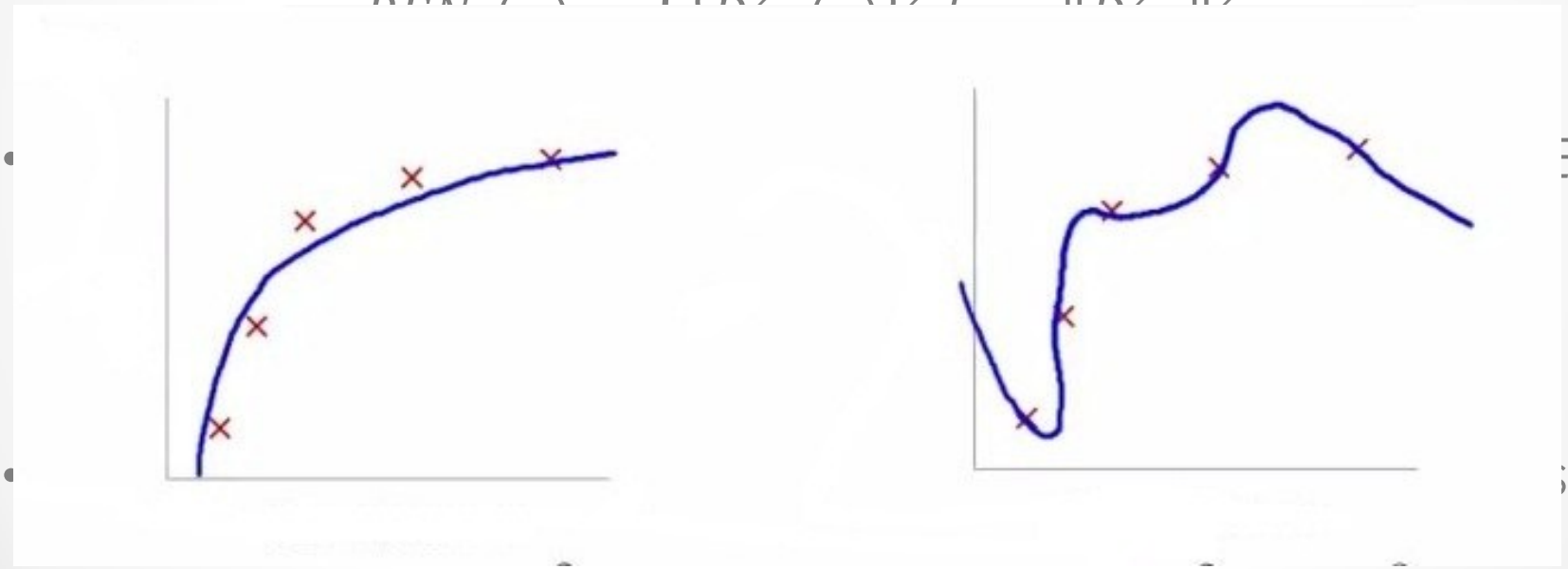
- Defining data fitting as the minimization of the sum of squared errors or residuals:

$$\min SSE(y_i|c) = \sum_{j=1}^{n} \left[y_{ij} - \sum_{k=1}^{K} c_{ik}\phi_k(t_j) \right]^2$$

- The overfitting problem or the underfitting problem:
  - Make sure the closeness of the fit is good enough.
  - Make sure the overfitting will not exist, that is, no dramatic changes in a local range.

Roughness Penalty Method

- The integration of square of the second derivate



  - G. H. Golub, M. Heath and G. Wahba, "Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter", Technometrics ,Vol. 21, No. 2, May, 1979

Functional Principle Component Analysis

- The original data $x_{i1}, x_{i2}, \cdots, x_{ip}$ is converted into function $x_i(t)$, FPCA provides a way to compress the functional data information as :

$$f_i = \int_o^T \beta(s)\, x_i(s) ds = \int \boldsymbol{\beta x_i}$$

- f_i is the i-th functional principal component for the functional data $x_i(t)$, so this problem can be abstracted as the formulate as:

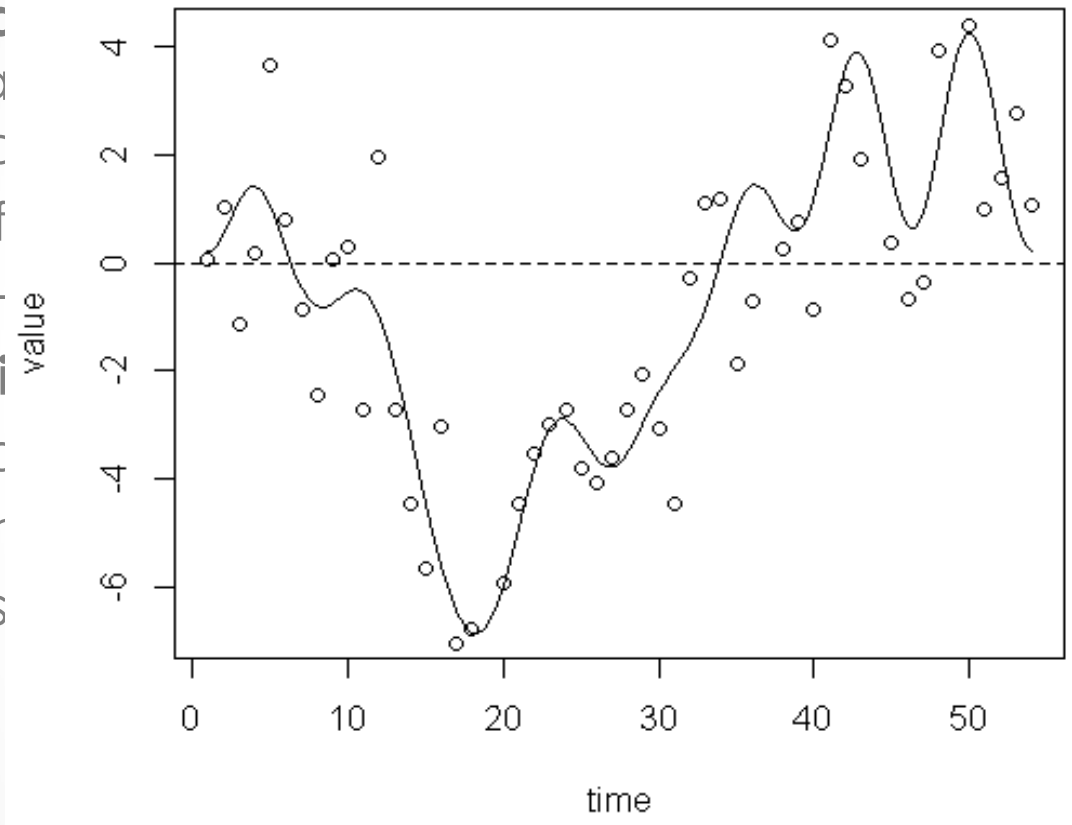
$$\begin{cases} \max\ Var(f') = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\left(\int \boldsymbol{\beta x_i}\right)^2 \\ s.t \int [\beta(s)]^2 ds = \|\boldsymbol{\beta}\|^2 = 1 \end{cases}$$

The Distance Meatures

- The MFCC-FDA feature
  - The short-term time dynamics
  - The overall envelop information
- Different distance measures have different space description properties (Amplitude distance/ Angle)

$$\begin{cases} d_{Mink} = \sqrt[P]{\sum_{k=1}^{n} |x_{1k} - x_{2k}|^P} \\ S_{cos} = \dfrac{\sum_{k=1}^{n} x_{1k} x_{2k}}{\sqrt{\sum_{k=1}^{n} {x_{1k}}^2} \sqrt{\sum_{k=1}^{n} {x_{2k}}^2}} \end{cases}$$

- **Databa** omputer
  Enterta -meter
  distanc
  - 5 diff times
  - 16 kH
- **Conditi**
  - 32 –c
  - The r
  - The s



Results and Analysis

- Baseline System: Dynamic Time Warping method with MFCC feature

Method	EER
MFCC-DTW (Baseline)	6.13%
MFCC-FDA	9.54%

- It can be found that compared with the classic MFCC-DTW system, the performance of directly using the FDA coefficients as the new feature is not good enough

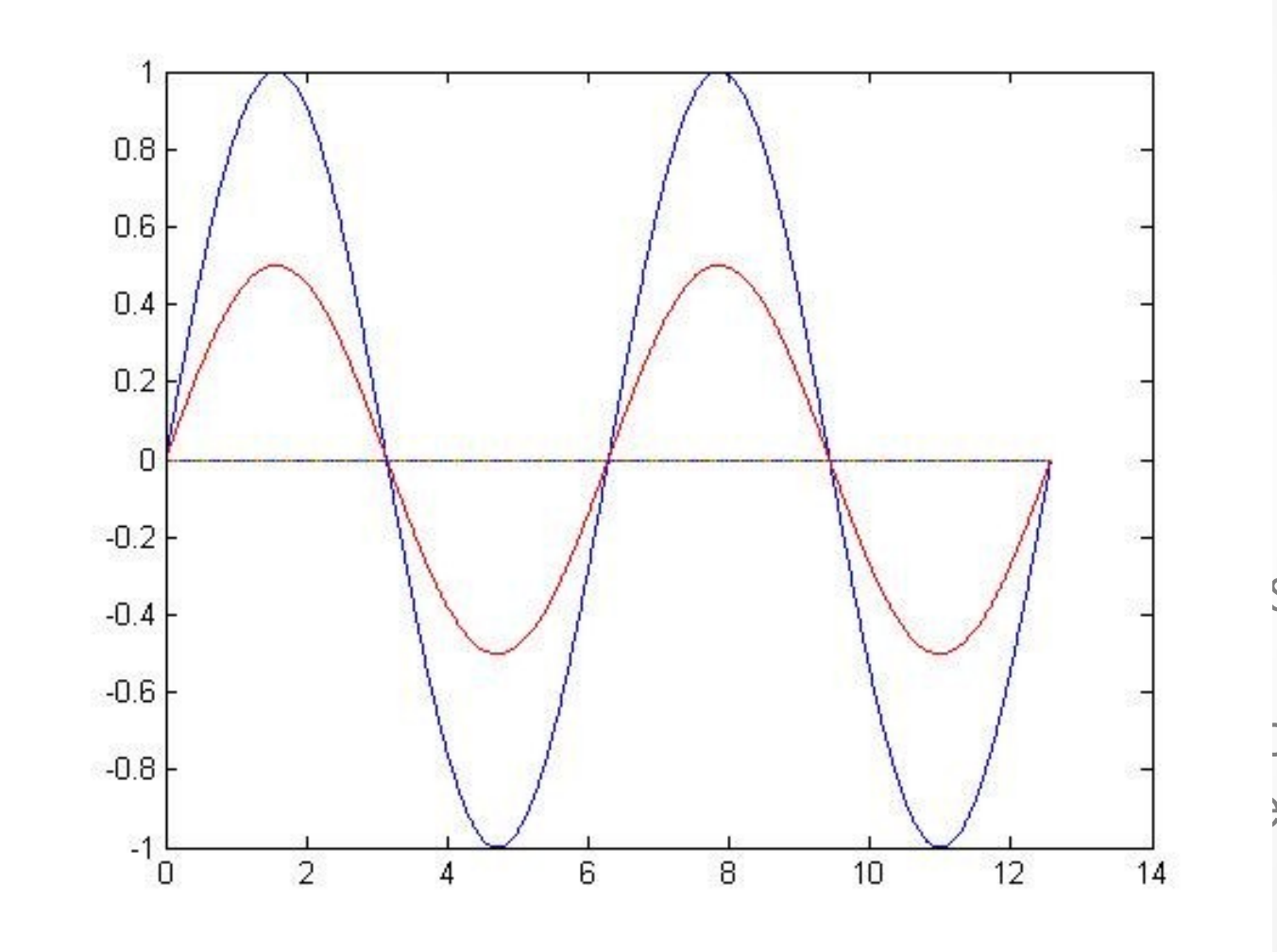
FPCA Results

- The Performance after applying the functional PCA

MFCC-FPCA (nharm)	1	3	5	7	9
EER (%)	11.80	7.95	6.31	6.15	6.01

- FPCA shows great improvement over just FDA coefficients, it effectively reduces the redundant information
- The MFCC-FPCA plus Euclidean Distance can achieve an equivalent performance as the classic MFCC-DTW TDSR system.

The influence of distance



- MF st
  pe
- Th ows
  mo en the
  an

Conclusions

- The new text-dependent speaker recognition method based on the FDA theory and the MFCC feature gave better performance over DTW method
- This method combines the advantage of short-term speech dynamics and the global statistical information
- The MFCC-FDA feature will work better with the cosine similarity measure

FDA are much more than that

- Pre-Processing
  - Local weighting smoothing methods
  - Roughness penalty: Regularization Basis Function
- Registration
  - Continuous Registration
- Functional linear models for Scale/Functional Responses
- Functional canonical correlation and discriminant analysis
- Functional cluster analysis

References

- [1]. Ramsay, J. O. and B. W. Silverman (2005). Functional Data Analysis, Second Edition.New York: Springer.
- [2]. Functional Data Analysis with R and Matlab, Ramsay, JO, Hooker, Giles, and Graves, Spencer, Springer 2009
- [3]. A Tutorial on Functional Data Analysis for Speech Research, Michele Gubian, interspeech 2011, http://lands.let.ru.nl/FDA/FDA_Tutorials.htm
- [4].The study on the methods of functional data analysis and their application, Liurui Qin, Xiamen University, PHD thesis

Thank you!