

國立交通大學
統計學研究所
碩士論文

利用無母數分類方法辨識動態車輛

The Recognition of Traveling Vehicles by
Nonparametric Discrimination of Functional Data
with Different Proximity

研究生：呂永在
指導教授：周幼珍 博士

中華民國九十六年六月

利用無母數分類方法辨識動態車輛
The Recognition of Traveling Vehicles by
Nonparametric Discrimination of Functional Data
with Different Proximity

國立交通大學
統計學研究所
碩士論文

A Thesis
Submitted to Institute of Statistics
College of Science
National Chiao Tung University
in Partial Fulfillment of the Requirements
for the Degree of
Master
in
Statistics
June 2007

Hsinchu, Taiwan, Republic of China

中華民國九十六年六月

利用無母數分類方法辨識動態車輛

研究生：呂永在

指導教授：周幼珍 博士

國立交通大學統計學研究所

摘要

塞車是日常生活中常見的問題，塞車主因之一是車道不足或車道無法有效利用，所以辨識常塞車路段的車種，統計各車種流量，將有助於車道、專用車道及交通號誌的建置與控管，進而有效解決塞車問題。由於雷達微波辨識系統建置成本較影像辨識系統低廉，本文所考慮是一組雷達微波資料集，資料包含反射波強度及車種的紀錄，資料經由適當的截取及平移，此筆資料可視為函數資料，因此利用函數資料的無母數分群方法來辨識車種，並比較三種不同函數資料的近似度測度，找出辨識率最高的方法。

The Recognition of Traveling Vehicles by Nonparametric Discrimination of Functional Data with Different Proximity

Student : Yung-Tsai Lu

Advisor : Dr. Yow-Jen Jou

Institute of Statistic

National Chiao Tung University



Traffic congestion is a serious and general problem in our daily life. Real time traveling vehicle information is essential to the advanced traffic management. The recognition and statistics of traffic flow among different types of traveling vehicles would be contributive to improve traffic block. This thesis considers the dataset recorded by the Radar microwave detector with the intensity of back waves and the types of traveling vehicles. The data is treated as functional data and then classification would be proposed to be performed by nonparametric discrimination of functional data with three forms of Proximity. The proximity with lower misclassification rate would be adopted for the data. The results show that the misclassification rate is pretty satisfactory if the number of groups is two and as the number of groups increases the misclassification rate increases as expected.

誌 謝

這篇論文能順利完成，最先要感謝的是我的指導教授周幼珍老師，老師耐心且細心的指導，讓我對研究方向愈來愈清楚，論文才能逐步完成。在生活及課業上老師給予的建議都讓我獲益匪淺。再來要感謝卓訓榮老師及運輸所的學長及同學們，由於大家的協助及努力，才能實測到這些雷達車輛資料，實測過程現在回想起來還真是有趣及充滿驚奇，後續資料分析及處理也感謝大家所給予的協助。

再來要感謝洪志真、徐南蓉及胡殿中老師，老師在百忙當中抽空當我的口試委員，詳讀我的論文且給予許多的建議，讓我的論文更趨完整，謝謝各位老師。

當然也要感謝統計所的同學們，兩年來不管是課業上的討論，還是中午吃飯八卦聊是非或是運動打籃球都歷歷在目。俊睿、阿Q、柯董、建威、阿淳、益銘、侑峻、小B在我念書念得煩雜時陪我打打籃球，運動打發時間。雪芳、素梅、雅莉、苣慈、花花、怡君、小米、美惠、映伶、小賴大家常一起吃飯聊天，要是沒有你們日子還真的有點無聊。最令人印象深刻的當然是我們的沙巴之旅，我的第一次出國旅行，玩的很開心，畢業後大家雖各奔東西為自己的理想打拚，希望還能多多聯絡，日後能再有機會一起出國玩，記著大家十年後的約定。

最要感謝的是我的爸爸媽媽和女友雅雲，有你們在背後默默的支持及照顧，在我情緒低落時為我加油打氣，才能讓我無後顧之憂地的完成研究所的學業，由衷地感謝你們。



呂永在 謹誌于
國立交通大學統計研究所
中華民國九十六年六月

The Recognition of Traveling Vehicles by Nonparametric Discrimination of Functional Data with Different Proximity

Student:Yung-Tsai Lu

Advisor:Yow-Jen Jou

Institute of Statistic
National Chiao Tung University
Hsinchu, Taiwan



Traffic congestion is a serious and general problem in our daily life. Real time traveling vehicle information is essential to the advanced traffic management. The recognition and statistics of traffic flow among different types of traveling vehicles would be contributive to improve traffic block. This thesis considers the dataset recorded by the Radar microwave detector with the intensity of back waves and the types of traveling vehicles. The data is treated as functional data and then classification would be proposed to be performed by nonparametric discrimination of functional data with three forms of Proximity. The proximity with lower misclassification rate would be adopted for the data. The results show that the misclassification rate is pretty satisfactory if the number of groups is two and as the number of groups increases the misclassification rate increases as expected.

Key words: Partial least square regression; Kernel; K Nearest Neighbors; Nonparametric Discrimination; Proximity.

Contents

1	Introduction	3
2	Model Specifications and Methodology	6
2.1	Prerequisite Notions and Notation	6
2.2	Three Types of Semi-Metrics	7
2.2.1	Semi-Metrics Based on Functional PCA	7
2.2.2	Semi-Metrics Based on Functional PLS	9
2.2.3	Semi-Metrics Based on Derivatives	13
2.3	Nonparametric Classification of Functional Data	15
2.3.1	Method	15
2.3.2	K Nearest Neighbors Estimator	18
3	Practical Example	21
3.1	Data Description	21
3.2	Data Analysis	25
4	Conclusion	34

1 Introduction

The Radar identification technology is developed fast after the World War II. The technology is principally used in the recognition of aircrafts and tanks. In military scenarios, the need to reliably identify objective is even more stringent, since erroneous identification could easily result in friendly fire incidents. A common technique for identification of military objective is Identification Friend Foe (IFF). IFF identification is initiated when the interrogator transmits a challenge to the aircraft. Friendly aircraft are supposed to be equipped with a transponder, which replies to the challenge by transmitting an identification code to the interrogator. Some IFF modes of operation require more information to be included in the reply, such as the current aircraft altitude. Hostile aircraft will in general not be able to respond properly to the challenge because of the lack of a (compatible) transponder, and will therefore be identified as hostile (or at least not friendly). Various other identification techniques are used in combination with IFF. For example, friendly aircraft can be required to limit their flight path to pre-defined regions of airspace called corridors.



Traditional radars collect and transform the information to form images such as two-dimensional Inverse Synthetic Aperture Radar (ISAR) or a sequence of one-dimensional range profile from the raw data, and then classify targets based on these images. Herman (2002) bypassed the image formation and attempted target recognition directly from the received data, which is labeled as Automatic Target

Recognition (ATR). A Radio Frequency System-on-Chip (RF SoC) using the theory of Frequency Modulated Continuous Waves (FMCW) combining the military Radar identification technology can be used in traffic management. The Radar vehicle recognition system is less expensive and reliable than the recognition of image system. Our approach to recognizing (or, classifying) our targets (vehicles on the road) adopts the same idea of Hermans work, i.e., attempt automatic classification directly the received data. Therefore, we use the nonparametric discrimination of functional data to solve the recognition problem. The aim is to find a robust methodology that classifies a new object into one of a prespecified set of classes.

In many scientific disciplines, the observed response from experimentation may be viewed as a continuous curve rather than a scalar or a vector. Main sources from this kind of data are real time monitoring of processes and certain types of longitudinal studies. In the terminology of Ramsay and Silverman (1997), they are called functional data sets. Although the underlying data are made up intrinsically by curves, the observed data are discretized representations of these curves. Therefore, each data curve is represented by a (large) finite-dimensional vector. The practical Radar data set can be treated as the functional data. So the nonparametric discrimination of functional data is adopted suitably for the Radar data set.

The rest of this thesis is organized as follows: Section 2 introduces the nonparametric discrimination of functional data with three types of proximity. The three types of proximity contain the PCA-type, PLS-type and Derivatives-type semi-

metrics. The method of PLS regression is also mentioned compendiously. Section 3 describes the Radar data set and uses the method in the Section 2 to classify the vehicles and lanes. Some analysis results are also mentioned. The conclusion would be proposed in the Section 4.



2 Model Specifications and Methodology

Functional data sets appear in many areas of science. Although each data point may be seen as a large finite-dimensional vector it is preferable to think of them as functional data. First we introduce some notation of functional data, then give a brief introduction of nonparametric discrimination with three types of proximity.

2.1 Prerequisite Notions and Notation

We first introduce some common notation and the terminology that generally used in mathematics.

Definition 1. A random variable \mathcal{X} is called functional random variable (f.r.v.) if it takes values in an infinite dimensional space. An observation \mathbf{x} of \mathcal{X} is called a functional data. The \mathcal{X} denotes a random curve, $\mathcal{X} = \{\mathcal{X}(t), t \in \mathbf{T}\}$.

Definition 2. $\|\cdot\|$ is a semi-norm on some space \mathbf{F} as long as:

1. $\forall \mathbf{x} \in \mathbf{F}, \|\mathbf{x}\| \geq 0$.
2. $\forall (a, \mathbf{x}) \in \mathcal{R} \times \mathbf{F}, \|a\mathbf{x}\| = |a|\|\mathbf{x}\|$.
3. $\forall (\mathbf{x}, \mathbf{y}) \in \mathbf{F} \times \mathbf{F}, \|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

Definiton 3. d is a semi-metric on some space \mathbf{F} as long as:

1. $\forall \mathbf{x} \in \mathbf{F}, d(\mathbf{x}, \mathbf{x}) = 0$.
2. $\forall (\mathbf{x}, \mathbf{y}) \in \mathbf{F} \times \mathbf{F}, d(\mathbf{x}, \mathbf{y}) \geq 0$.
3. $\forall (\mathbf{x}, \mathbf{y}, \mathbf{z}) \in \mathbf{F} \times \mathbf{F} \times \mathbf{F}, d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y})$.

Note that in fact, a semi-norm $\|\cdot\|$ does not have the property that $\|\mathbf{x}\| = 0$ imply $\mathbf{x} = \mathbf{0}$. Similarly, a semi-metric does not have the property that $d(\mathbf{x}, \mathbf{y}) = 0$ imply

$\mathbf{x} = \mathbf{y}$.

2.2 Three Types of Semi-Metrics

Because most available functional datasets are curves, we will describe three types of semi-metrics well adapted for this kind of data in the following subsection.

2.2.1 Semi-Metrics Based on Functional PCA

Functional Principal Components Analysis (FPCA) is a good tool for computing proximities between curves in a reduced dimensional space. As long as $E[\int \mathcal{X}^2(t)dt]$ is finite, the FPCA of the f.r.v. \mathcal{X} has the following expansion (Dauxois, etc. 1982):

$$\mathcal{X} = \sum_{k=1}^{\infty} \left(\int \mathcal{X}(t)v_k(t)dt \right) v_k$$

with v_1, v_2, \dots , being the orthogonal eigenfunctions of the covariance operator

$$\Gamma_{\mathcal{X}}(s, t) = E(\mathcal{X}(s)\mathcal{X}(t))$$

associated with the eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots$. Then the truncated version of the above expansion of \mathcal{X} is

$$\tilde{\mathcal{X}}^{(q)} = \sum_{k=1}^q \left(\int \mathcal{X}(t)v_k(t)dt \right) v_k.$$

The main goal is to find truncated version such that

$$E \left(\int (\mathcal{X}(t) - P_q \mathcal{X}(t))^2 dt \right)$$

is minimized over all projections P_q of \mathcal{X} into q -dimensional spaces. According to the classical L^2 -norm, we can define a parametric class of semi-norms and semi-

metrics in the following way:

$$\|\mathbf{x}\|_q^{PCA} = \sqrt{\int (\tilde{\mathbf{x}}^{(q)}(t))^2 dt} = \sqrt{\sum_{k=1}^q \left(\int \mathbf{x}(t) v_k(t) dt \right)^2}$$

$$d_q^{PCA}(\mathcal{X}_i, \mathbf{x}) = \sqrt{\sum_{k=1}^q \left(\int [\mathcal{X}_i(t) - \mathbf{x}(t)] v_k(t) dt \right)^2}$$

In general, $\Gamma_{\mathcal{X}}$ is unknown and then, the v_k 's too. However, the covariance operator can be well approximated by its empirical version

$$\Gamma_{\mathcal{X}}^n(s, t) = 1/n \sum_{i=1}^n \mathcal{X}_i(s) \mathcal{X}_i(t).$$

The eigenvectors of $\Gamma_{\mathcal{X}}^n$ are consistent estimators of eigenvectors of $\Gamma_{\mathcal{X}}$. In practice, we never observe exactly $\{\mathbf{x}_i = \{\mathbf{x}_i(t); t \in T\}\}_{i=1, \dots, n}$ but a discrete version $\{\mathbf{x}_i = (\mathbf{x}_i(t_1), \dots, \mathbf{x}_i(t_J))'\}_{i=1, \dots, n}$. So we can approximate the integral in a discrete way

$$\int [\mathcal{X}_i(t) - \mathbf{x}(t)] v_k(t) dt \approx \sum_{j=1}^J w_j (\mathcal{X}_i(t_j) - \mathbf{x}(t_j)) v_k(t_j)$$

where w_1, \dots, w_J are weights which define the approximate integration. A standard choice is $w_j = t_j - t_{j-1}$. Similarly, the semi-metric $d_q^{PCA}(\mathbf{x}_i, \mathbf{x}_{i'})$ can be approximated by its empirical version

$$d_q^{PCA}(\mathbf{x}_i, \mathbf{x}_{i'}) = \sqrt{\sum_{k=1}^q \left(\sum_{j=1}^J w_j (\mathbf{x}_i(t_j) - \mathbf{x}_{i'}(t_j)) v_k(t_j) \right)^2}$$

where v_1, v_2, \dots, v_q , are the \mathbf{W} -orthonormal eigenvectors of the covariance matrix

$$\Gamma^n \mathbf{W} = 1/n \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' \mathbf{W}$$

associated with the eigenvalues $\lambda_{1,n} \geq \lambda_{2,n} \geq \dots \geq \lambda_{q,n}$, and $\mathbf{W} = \text{diag}(w_1, \dots, w_J)$.

2.2.2 Semi-Metrics Based on Functional PLS

In this subsection we first introduce partial least squares (PLS) regression and different forms of PLS, then give a brief introduction of semi-metric based on functional PLS.

Partial Least Squares Regression:

Partial Least Squares is a widespread method for modeling relations between a set of dependent variables and a large set of predictors. PLS generalizes and combines features from principal component analysis and multiple regression. It originated in social sciences (particularly economy, Herman Wold 1966) but became popular first in chemometrics due to Herman's son Svante, (Geladi and Kowalski, 1986). It was first presented as an algorithm analog to the power method (used for computing eigenvectors) but was suitably interpreted in statistical framework. (Frank and Friedman, 1993; Hoskuldsson, 1988; Helland, 1990; Tenenhaus, 1998).

PLS usually is used to predict response variable \mathbf{Y} from predictors \mathbf{X} and describe their common structure. If \mathbf{X} is full rank, ordinary multiple regression can be applied. When the number of predictors is larger than the number of observations, \mathbf{X} would be singular and the ordinary multiple regression is no longer practicable. Several methods have been developed to solve this problem, e.g. principal component regression. But the method is chosen to explain the variation of \mathbf{X} , nothing guarantees that the choice to explain \mathbf{X} are suitable for \mathbf{Y} . PLS regression finds components from \mathbf{X} that are also connected with \mathbf{Y} . It searches for a

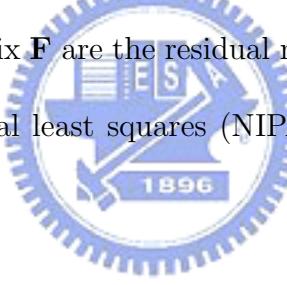
set of components (latent vectors) that performs a simultaneous decomposition of \mathbf{X} and \mathbf{Y} with the constraint that these components explain as much as possible of the covariance between \mathbf{X} and \mathbf{Y} .

Let \mathbf{X} be the zero-mean ($n \times N$) matrix and \mathbf{Y} be the zero-mean ($n \times M$) matrix, where n denotes the number of data sample. PLS decomposes \mathbf{X} and \mathbf{Y} into the form

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E}$$

$$\mathbf{Y} = \mathbf{UQ}' + \mathbf{F} \quad (1)$$

where the \mathbf{T} , \mathbf{U} are ($n \times p$) matrix of the p extracted components, the ($N \times p$) matrix \mathbf{P} and the ($M \times p$) matrix \mathbf{Q} represent loading matrices and the ($n \times N$) matrix \mathbf{E} and the ($n \times M$) matrix \mathbf{F} are the residual matrices. PLS which is based on the nonlinear iterative partial least squares (NIPALS) algorithm finds weight vectors \mathbf{w} , \mathbf{c} such that



$$[cov(\mathbf{t}, \mathbf{u})]^2 = [cov(\mathbf{Xw}, \mathbf{Yc})]^2 = \max_{(|\mathbf{r}|=|\mathbf{s}|=1)} [cov(\mathbf{Xr}, \mathbf{Ys})]^2$$

where $cov(\mathbf{t}, \mathbf{u}) = \mathbf{t}'\mathbf{u}/n$ denotes the sample covariance between the components \mathbf{t} and \mathbf{u} . The NIPALS algorithm starts with random initial value of the component \mathbf{u} and repeats a sequence of following steps until convergence.

Step 1. $\mathbf{w} = \mathbf{X}'\mathbf{u}/(\mathbf{u}'\mathbf{u})$ (estimate \mathbf{X} weights)

Step 2. $\|\mathbf{w}\| \rightarrow 1$

Step 3. $\mathbf{t} = \mathbf{Xw}$ (estimate \mathbf{X} component)

Step 4. $\mathbf{c} = \mathbf{Y}'\mathbf{t}/(\mathbf{t}'\mathbf{t})$ (estimate \mathbf{Y} weights)

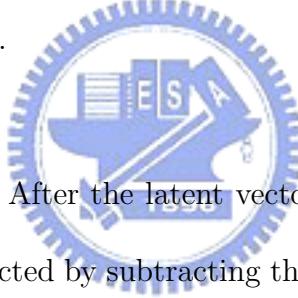
Step 5. $\|\mathbf{c}\| \rightarrow 1$

Step 6. $\mathbf{u} = \mathbf{Yc}$ (estimate \mathbf{Y} component)

Specially, $\mathbf{u} = \mathbf{y}$ if $M = 1$, \mathbf{Y} is a one-dimensional vector that is denoted by \mathbf{y} . In this case the NIPALS procedure converges in a single iteration.

It can be shown that the weight vector \mathbf{w} also corresponds to the first eigenvector of the following series of equations: $\mathbf{w} \propto \mathbf{X}'\mathbf{u} \propto \mathbf{X}'\mathbf{Yc} \propto \mathbf{X}'\mathbf{YY}'\mathbf{t} \propto \mathbf{X}'\mathbf{YY}'\mathbf{Xw}$. This shows that the weight vector \mathbf{w} is the right singular vector of the matrix $\mathbf{X}'\mathbf{Y}$. Similarly, the weight vectors \mathbf{c} is the left singular vector of $\mathbf{X}'\mathbf{Y}$. Then the latent vectors \mathbf{t} and \mathbf{u} are given as $\mathbf{t} = \mathbf{Xw}$ and $\mathbf{u} = \mathbf{Yc}$, where the weight vector \mathbf{c} is defined in steps 4 and 5 as stated above. Similarly, the extraction of \mathbf{t} , \mathbf{u} or \mathbf{c} estimates can be derived.

Forms of PLS:



PLS is an iterative process. After the latent vectors \mathbf{t} and \mathbf{u} being extracted, the matrices \mathbf{X} and \mathbf{Y} are extracted by subtracting their rank-one approximations based on \mathbf{t} and \mathbf{u} . Different extractions form several variants of PLS. By equation (1) the loading vectors \mathbf{p} and \mathbf{q} are computed as coefficients of regressing \mathbf{X} on \mathbf{t} and \mathbf{Y} on \mathbf{u} , respectively. Then, the loading vectors can be solved by $\mathbf{p} = \mathbf{X}'\mathbf{t}/(\mathbf{t}'\mathbf{t})$ and $\mathbf{q} = \mathbf{Y}'\mathbf{u}/(\mathbf{u}'\mathbf{u})$.

1.) PLS Mode A:

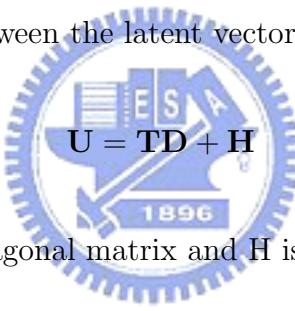
The PLS Mode A is based on rank-one deflation of individual matrices using the corresponding latent and loading vectors. In this case, the \mathbf{X} and \mathbf{Y} matrices are extracted $\mathbf{X} = \mathbf{X} - \mathbf{tp}'$ and $\mathbf{Y} = \mathbf{Y} - \mathbf{uq}'$. This method was originally proposed

by Herman Wold (1966) to model the relations between the different sets of data.

2.) PLS1 and PLS2:

PLS1 (either dependent variable or independent variable consists of a single variable) and PLS2 (both variables are multidimensional) are used as PLS regression method. The form of PLS is the most popular PLS approach. The main feature of the approach is that the relation between of \mathbf{X} and \mathbf{Y} is asymmetric. The main assumptions of the form of PLS are:

- (i) The latent vectors $\{\mathbf{t}_i\}_{i=1}^p$ are good predictors of \mathbf{Y} ; p denotes the number of extracted latent vectors.
- (ii) a linear inner relation between the latent vectors \mathbf{t} and \mathbf{u} exists; that is,



where \mathbf{D} is the $(p \times p)$ diagonal matrix and \mathbf{H} is the residual matrix.

The asymmetric assumption of the independent and dependent variable relation is transformed into a deflation scheme. The latent vectors $\{\mathbf{t}_i\}_{i=1}^p$ are good predictors of \mathbf{Y} . Then the latent vectors are used to deflate \mathbf{Y} , that is, a component of the regression of \mathbf{Y} on \mathbf{t} is removed from \mathbf{Y} at each iteration of PLS.

$$\mathbf{X} = \mathbf{X} - \mathbf{t}\mathbf{p}' \text{ and } \mathbf{Y} = \mathbf{Y} - \mathbf{t}\mathbf{t}'\mathbf{Y}/(\mathbf{t}'\mathbf{t}) = \mathbf{Y} - \mathbf{t}\mathbf{c}'$$

where the weight vector \mathbf{c} is defined in step 4 of NIPALS. This way of deflation will ensure that the extracted latent vectors $\{\mathbf{t}_i\}_{i=1}^p$ are mutually orthogonal.

3.) SIMPLS:

In order to avoid deflation steps at each iteration of PLS1 and PLS2, Jong (1993) has introduced another form of PLS denoted SIMPLS. The SIMPLS approach directly finds the weight vectors $\{\tilde{\mathbf{w}}_i\}_{i=1}^p$ which are applied to the original matrix \mathbf{X} . The criterion of the mutually orthogonal latent vectors $\{\tilde{\mathbf{t}}_i\}_{i=1}^p$ is kept.

Semi-Metrics Based on Functional PLS:

Let v_1^q, \dots, v_p^q be the vectors of \mathcal{R}^J performed by multivariate partial least squares regression (MPLSR) where q denotes the number of the factors and p the number of scalar responses. Then the semi-metric based on the MPLSR is defined as

$$d_q^{PLS}(\mathbf{x}_i, \mathbf{x}_{i'}) = \sqrt{\sum_{k=1}^p \left(\sum_{j=1}^J w_j (\mathbf{x}_i(t_j) - \mathbf{x}_{i'}(t_j)) v_k^q(t_j) \right)^2}$$

where w_1, \dots, w_J are weights which define the approximate integration. A standard choice is $w_j = t_j - t_{j-1}$. When we consider only one scale response ($p = 1$), the proximity between two discrete curves is due to only one direction, which seems inadequate with regard to the complexity of functional data. However, as soon as we consider multivariate response, such a family of semi-metrics allows to obtain very good results, which is the case in the curves discrimination context.

2.2.3 Semi-Metrics Based on Derivatives

One semi-metrics is to consider a distance between curves among their derivatives. The semi-metric based on derivatives of two observed curves \mathbf{x}_i and $\mathbf{x}_{i'}$ can

be defined as

$$d_q^{deriv}(\mathbf{x}_i, \mathbf{x}_{i'}) = \sqrt{\int (\mathbf{x}_i^{(q)}(t) - \mathbf{x}_{i'}^{(q)}(t))^2 dt}$$

where $\mathbf{x}^{(q)}$ denotes the q th derivative of \mathbf{x} . Note that $d_0^{deriv}(\mathbf{x}_i, \mathbf{0})$ is the classical L^2 -norm of \mathbf{x} . The computation of successive derivatives is very sensitive numerically.

In order to overcome the numerical stability problem, we can use a B-spline (Boor, 1978) approximation for the curves. Once we have obtained an analytical B-spline expansion for each curve, the successive derivatives are directly computed by differentiating several times their analytic form. Let $\{B_1, \dots, B_B\}$ be a B-spline basis, then the discrete approximative form of the curve $\mathbf{x}_i = (\mathbf{x}_i(t_1), \dots, \mathbf{x}_i(t_J))'$ is as follows:

$$\hat{\beta}_i = (\hat{\beta}_{i1}, \dots, \hat{\beta}_{iB}) = \arg \inf_{(\alpha_1, \dots, \alpha_B) \in \mathcal{R}^B} \sum_{j=1}^J \left(\mathbf{x}_i(t_j) - \sum_{b=1}^B \alpha_b B_b(t_j) \right)^2.$$

This produces a good approximation of the solution of the minimization problem

$$\arg \inf_{(\alpha_1, \dots, \alpha_B) \in \mathcal{R}^B} \int \left(\mathbf{x}_i(t) - \sum_{b=1}^B \alpha_b B_b(t) \right)^2 dt.$$

Therefore, the approximative form of the curve $\mathbf{x}_i = (\mathbf{x}_i(t_1), \dots, \mathbf{x}_i(t_J))'$ is

$$\hat{\mathbf{x}}_i(\cdot) = \sum_{b=1}^B \hat{\beta}_{ib} B_b(\cdot)$$

Because the analytic expression of the B'_b s is well-known, the successive derivatives can be exactly computed and we can differentiate easily the approximated curves:

$$\hat{\mathbf{x}}_i^{(q)}(\cdot) = \sum_{b=1}^B \hat{\beta}_{ib} B_b^{(q)}(\cdot).$$

Then semi-metric based on derivatives of two observed curves \mathbf{x}_i and $\mathbf{x}_{i'}$, can be computed by

$$d_q^{deriv}(\mathbf{x}_i, \mathbf{x}_{i'}) = \sqrt{\int (\hat{\mathbf{x}}_i^{(q)}(t) - \hat{\mathbf{x}}_{i'}^{(q)}(t))^2 dt}.$$

2.3 Nonparametric Classification of Functional Data

Classification or discrimination of functional data is in the situation when we observe a f.r.v. \mathcal{X} and a categorical response \mathbf{Y} which gives the group of each functional observation. The main purpose is to predict the group when we observe a new functional data. In next section we hope to find a robust method for assigning each functional observation to some homogeneous group. We review the nonparametric discrimination method that is proposed by Ferraty and Vieu (2006).

2.3.1 Method



Let $(\mathcal{X}_i, \mathbf{Y}_i)_{i=1, \dots, n}$ be a sample of n independent pairs, identically distributed as $(\mathcal{X}, \mathbf{Y})$ and valued in $\mathbf{F} \times \mathbf{G}$, where $\mathbf{G} = \{1, \dots, G\}$ and (\mathbf{F}, d) is a semi-metric vector space (\mathcal{X} is a f.r.v. and d is a semi-metric). The notation $(\mathbf{x}_i, \mathbf{y}_i)$ denotes the observation of the pair $(\mathcal{X}_i, \mathbf{Y}_i)$.

General classification rule (Bayes rule). Given a functional observation \mathbf{x} , the purpose is to estimate the posterior probability

$$p_g(\mathbf{x}) = P(\mathbf{Y} = g | \mathcal{X} = \mathbf{x}), g \in \mathbf{G}.$$

Once the G probabilities are estimated $(\hat{p}_1(\mathbf{x}), \dots, \hat{p}_G(\mathbf{x}))$, the classification rule consists of assigning an incoming functional observation \mathbf{x} to the group with highest estimated posterior probability

$$\hat{\mathbf{y}}(\mathbf{x}) = \arg \max_{g \in \mathbf{G}} \hat{p}_g(\mathbf{x}).$$

This classification rule is called Bayes rule. To use a suitable kernel estimator make precise discrimination of functional data.

Kernel estimator of posterior probabilities. Before defining the kernel-type estimator of the posterior probabilities, we remark that

$$p_g(\mathbf{x}) = P(\mathbf{I}_{[\mathbf{Y}=g]} | \mathcal{X} = \mathbf{x}),$$

with $\mathbf{I}_{[\mathbf{Y}=g]}$ equal to 1 if $\mathbf{Y} = g$ and 0 otherwise. Therefore we can use kernel-type estimator introduced for the prediction via conditional expectation.

$$\hat{p}_g(\mathbf{x}) = \hat{p}_{g,h}(\mathbf{x}) = \frac{\sum_{i=1}^n I_{[\mathbf{Y}_i=g]} K(h^{-1}d(\mathbf{x}, \mathcal{X}_i))}{\sum_{i=1}^n K(h^{-1}d(\mathbf{x}, \mathcal{X}_i))},$$

where K is the kernel and h is the bandwidth (a strictly positive smoothing parameter). The kernel posterior probability estimate can be rewritten as

$$\hat{p}_{g,h}(\mathbf{x}) = \sum_{\{i: \mathbf{Y}_i=g\}} w_{i,h}(\mathbf{x}) \quad \text{with} \quad w_{i,h}(\mathbf{x}) = \frac{K(h^{-1}d(\mathbf{x}, \mathcal{X}_i))}{\sum_{i=1}^n K(h^{-1}d(\mathbf{x}, \mathcal{X}_i))}.$$

In order to compute the quantity $\hat{p}_{g,h}(\mathbf{x})$ we use only the \mathcal{X}_i 's belonging to both the group g and the ball centered at \mathbf{x} and of radius h .

$$\hat{p}_{g,h}(\mathbf{x}) = \sum_{i \in \mathbf{I}} w_{i,h}(\mathbf{x}) \quad \text{where} \quad \mathbf{I} = \{i : \mathbf{Y}_i = g\} \cap \{i : d(\mathbf{x}, \mathcal{X}_i) < h\}.$$

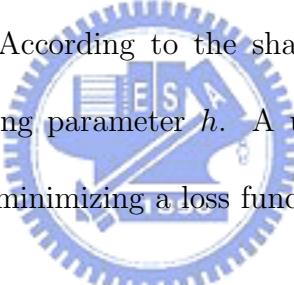
The closer \mathcal{X}_i is to \mathbf{x} the larger the quantity $K(h^{-1}d(\mathbf{x}, \mathcal{X}_i))$. Hence, the closer \mathcal{X}_i is to \mathbf{x} the larger the weight $w_{i,h}(\mathbf{x})$. So, among the \mathcal{X}_i 's lying to the g th group, the closer \mathcal{X}_i is to \mathbf{x} and the larger is its effect on the g th estimated posterior probability. As long as K is nonnegative, the kernel estimator has the following interesting properties :

$$(i) \ 0 \leq \hat{p}_{g,h}(\mathbf{x}) \leq 1,$$

$$(ii) \ \sum_{g \in \mathbf{G}} \hat{p}_{g,h}(\mathbf{x}) = 1,$$

which ensure that the estimated probabilities are forming a discrete distribution.

Choosing the bandwidth. According to the shape of the kernel estimator, we have to choose the smoothing parameter h . A usual way for an automatic choice of h is constructed from minimizing a loss function $Loss$ as:



$$h_{Loss} = \arg \min_h Loss(h),$$

where the function $Loss$ can be built from $\hat{p}_{g,h}(\mathbf{x})$'s and \mathbf{y}_i 's. The misclassification rate is a nature choice among different types of $Loss$ functions. Therefore, the functional classification can be performed by the following procedure:

Learning step:

for $h \in \mathcal{H}$

for $i = 1, 2, \dots, n$

$g = 1, 2, \dots, G$

$$\hat{p}_{g,h}(\mathbf{x}_i) = \frac{\sum_{i': \mathbf{y}_{i'}=g}^n K(h^{-1}d(\mathbf{x}_i, \mathbf{x}_{i'}))}{\sum_{i'=1}^n K(h^{-1}d(\mathbf{x}_i, \mathbf{x}_{i'}))}$$

enddo

enddo

enddo

$$h_{Loss} = \arg \min_{h \in \mathcal{H}} Loss(h)$$

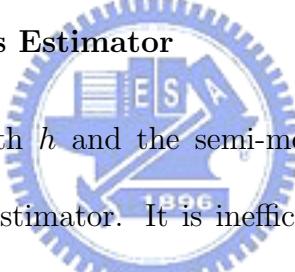
Predicting step:

Let \mathbf{x} be a new functional observation and $\hat{\mathbf{y}}(\mathbf{x})$ be its estimated group:

$$\hat{\mathbf{y}}(\mathbf{x}) \leftarrow \arg \max_{g \in \mathcal{G}} \{\hat{p}_{g,h_{Loss}}(\mathbf{x})\}$$

where $\mathcal{H} \subset \mathcal{R}$ is a set of suitable values for h and K is a known kernel.

2.3.2 K Nearest Neighbors Estimator



The choices of the bandwidth h and the semi-metric d have great influence on the behavior of the kernel estimator. It is inefficient to choose bandwidth h among the positive real number subset from a computational point of view. So, we consider a general and simple way which is the k nearest neighbors (kNN) version of kernel estimator. Then we can replace a choice of real parameter among an infinite number of values with an integer parameter k (among a finite subset). The main idea of the kNN estimator is to replace the parameter h with h_k which is the bandwidth allowing us to take into account k terms in the weighted average. The p_g at \mathbf{x} is estimate by

$$\hat{p}_{g,k}(\mathbf{x}) = \frac{\sum_{i: \mathbf{y}_i=g}^n K(h_k^{-1}d(\mathbf{x}, \mathbf{x}_i))}{\sum_{i=1}^n K(h_k^{-1}d(\mathbf{x}, \mathbf{x}_i))}$$

where h_k is a bandwidth such that the number of $\{i : d(\mathbf{x}, \mathbf{x}_i) \leq h_k\}$ is k . That is, the minimization problem on h over a subset of \mathcal{R} is replaced with a minimization on k over a finite subset $\{1, 2, \dots, K\}$:

$$k_{Loss} = \arg \min_{k \in \{1, 2, \dots, K\}} Loss(k) \quad \text{and} \quad h_{Loss} \leftarrow h_{k_{Loss}},$$

where the loss function $Loss$ is now built from $\hat{p}_{g,k}(\mathbf{x}_i)$'s and \mathbf{y}_i 's.

In order to choose the tuning parameter k , we must introduce a loss function ($Loss$). The loss function allows us to build a local version of the kNN estimator. The main goal is to compute the quantity:

$$p_g^{LCV}(\mathbf{x}) = \frac{\sum_{\{i: \mathbf{y}_i=g\}}^n K\left(\frac{d(\mathbf{x}, \mathbf{x}_i)}{h_{LCV}(\mathbf{x}_{i_0})}\right)}{\sum_{i=1}^n K\left(\frac{d(\mathbf{x}, \mathbf{x}_i)}{h_{LCV}(\mathbf{x}_{i_0})}\right)}$$

where $i_0 = \arg \min_{i=1, 2, \dots, n} d(\mathbf{x}, \mathbf{x}_i)$ and $h_{LCV}(\mathbf{x}_{i_0})$ is the bandwidth corresponding to the optimal number of neighbors at \mathbf{x}_{i_0} obtained by the following cross-validation procedure:

$$k_{LCV}(\mathbf{x}_{i_0}) = \arg \min_k Loss_{LCV}(k, i_0),$$

where

$$Loss_{LCV}(k, i_0) = \sum_{g=1}^G \left(\mathbf{I}_{[\mathbf{y}_{i_0}=g]} - p_{g,k}^{(-i_0)}(\mathbf{x}_{i_0}) \right)^2$$

and

$$p_{g,k}^{(-i_0)}(\mathbf{x}_{i_0}) = \frac{\sum_{\{i: \mathbf{y}_i=g, i \neq i_0\}}^n K\left(\frac{d(\mathbf{x}_{i_0}, \mathbf{x}_i)}{h_k(\mathbf{x}_{i_0})}\right)}{\sum_{i=1, i \neq i_0}^n K\left(\frac{d(\mathbf{x}_{i_0}, \mathbf{x}_i)}{h_k(\mathbf{x}_{i_0})}\right)}.$$

As long as we set up the appropriate semi-metric and kernel function $K(\cdot)$, the prediction procedure is finished.

The misclassification rate for the learning sample $(\mathbf{x}_i, \mathbf{y}_i)_{i=1,2,\dots,n}$ (the sample of curves for which the groups are observed) will be used to assess the performance of the predicted results, the procedure is described as follows:

for $i = 1, 2, \dots, n$

$$\mathbf{y}_i^{LCV} \leftarrow \arg \max_{g \in \{1,2,\dots,G\}} p_g^{LCV}(\mathbf{x}_i)$$

enddo

$$\text{Misclas} \leftarrow \frac{1}{n} \sum_{i=1}^n I[\mathbf{y}_i \neq \mathbf{y}_i^{LCV}]$$



3 Practical Example

The procedures described in previous sections will be applied to a real data set. We consider a real data set which is recorded by Radar microwave detector. The collection of the data would be introduced in section 4.1. Then some analysis results would be shown in section 4.2.

3.1 Data Description

The practice data set are collected near the section 1 of Singlong Road, Jhubei City, Hsinchu County by vehicle Radar microwave detector, which is a side-looking configuration. There are four lanes on the Singlong Road. The collection time is from 10:00 AM to 5:00 PM on October 11, 2006. These vehicles with speed 40 km/hour are recorded on the four lanes of Singlong Road. There are 248 observed curves of vehicles recorded. We do registration in the original dataset, and take the peak of intensity of back wave as the marker. So the observed curves are shifted and truncated, then we take the adjusted curves as the analytic curves. Each of curves represents the intensity of back wave at 30 discrete observed points and the types of vehicles. So we observe $n = 248$ pairs $(\mathbf{x}_i, \mathbf{y}_i)_{i=1,\dots,n}$ where $\mathbf{x}_i = (\mathbf{x}_i(t_1), \mathbf{x}_i(t_2), \dots, \mathbf{x}_i(t_{30}))'$ correspond to the i th discrete functional data and \mathbf{y}_i give the class of the i th observation. Given a new functional observation \mathbf{x} , our main goal is to predict the corresponding class of vehicle and lane \mathbf{y}^{LCV} .

The classes of vehicles contain small and large vehicles. The small vehicles are the sedans. However, the large vehicles contain cranes, buses, trucks, goods

wagons, etc. The Figure 1 displays intensity of back wave for the two vehicle classes on the lane 3. The 10 solid curves and 10 dashed curves represent the sample of small and large vehicles, individually. The functional data of different lanes and vehicles are shown in Figure 2 and Figure 3.

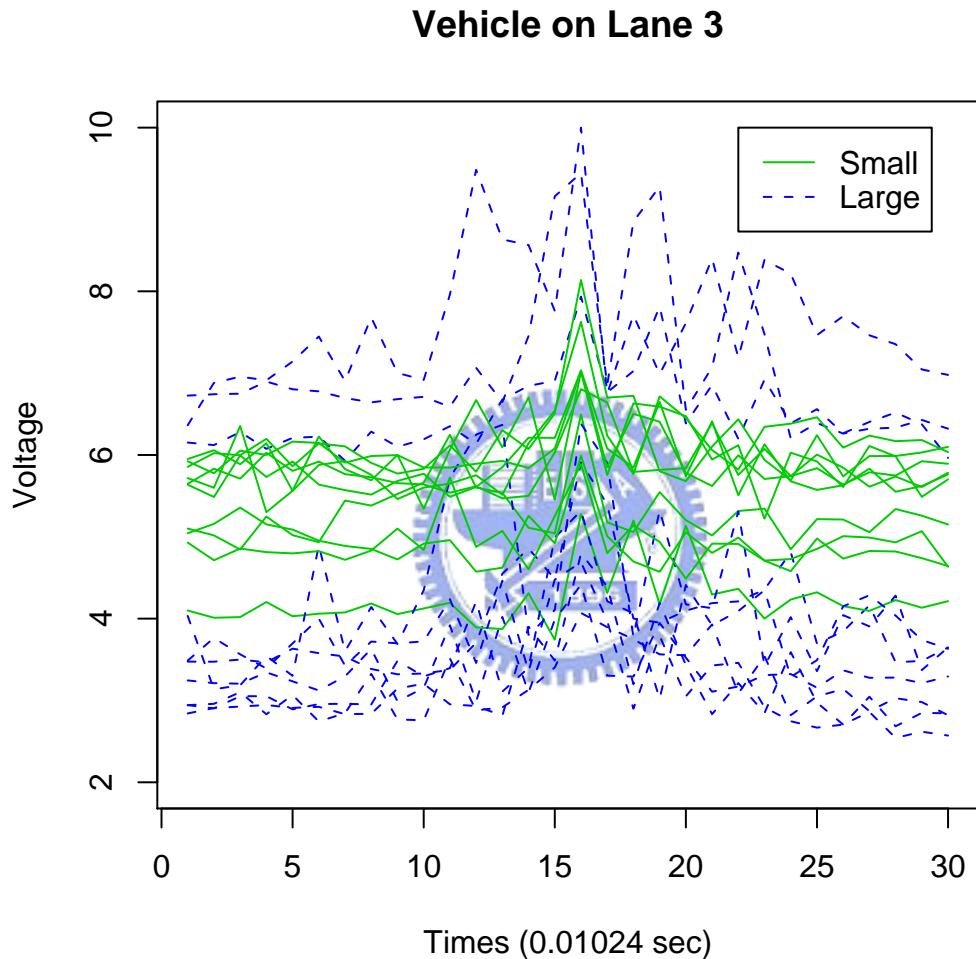
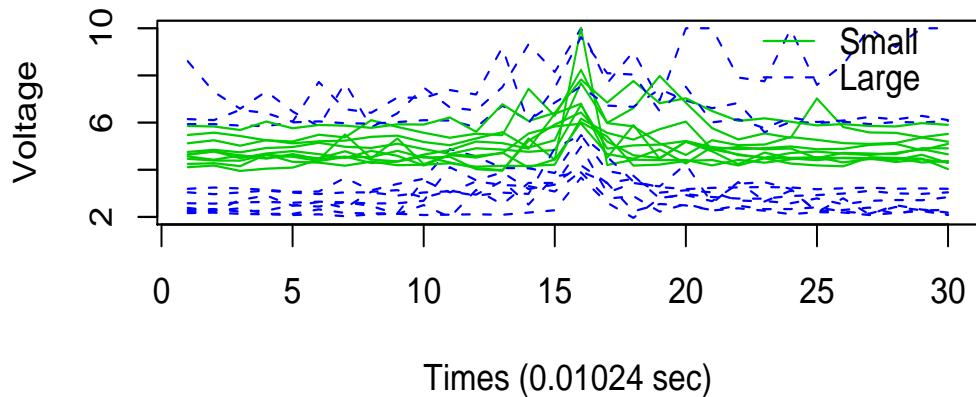


Figure 1: The intensity of back wave for the two vehicle classes on the lane 3. The 10 solid curves and 10 dashed curves represent the sample of small and large vehicles, individually.

Vehicle on Lane 1



Vehicle on Lane 2

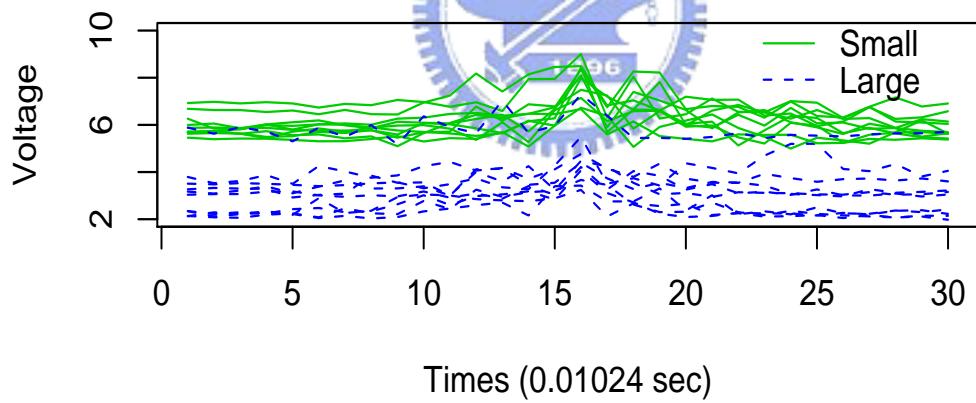
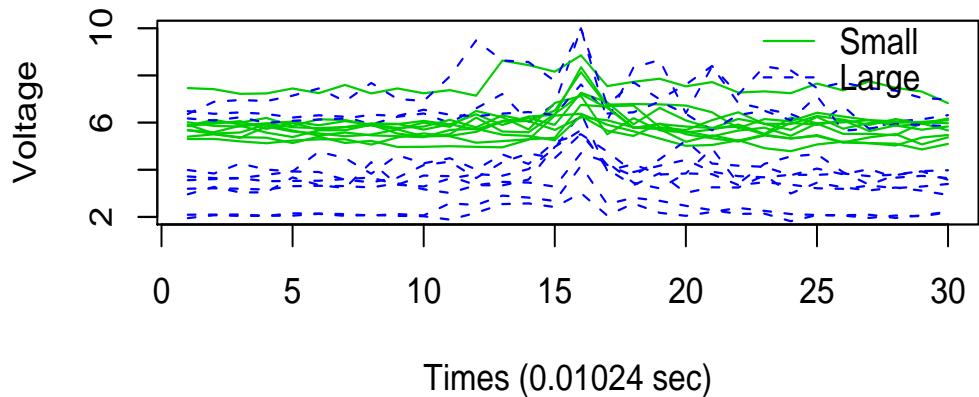


Figure 2: The intensity of back wave for the two vehicle classes on the lane 1 and lane 2. The 10 solid curves and 10 dashed curves represent the sample of small and large vehicles, individually.

Vehicle on Lane 3



Vehicle on Lane 4

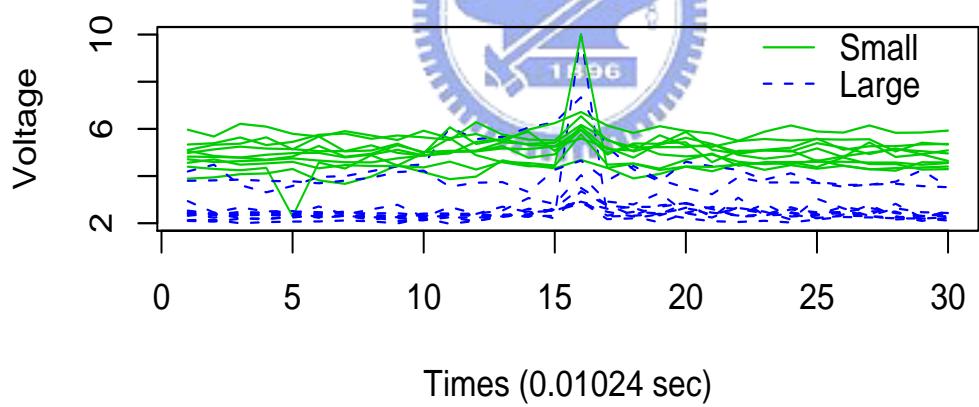
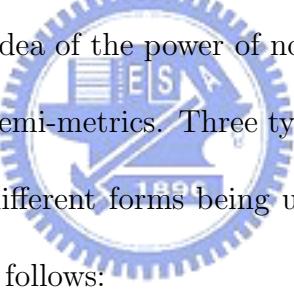


Figure 3: The intensity of back wave for the two vehicle classes on the lane 3 and lane 4. The 10 solid curves and 10 dashed curves represent the sample of small and large vehicles, individually.

3.2 Data Analysis

Our principal aim is to predict the class of vehicle and lane relative to the new observed functional data. In order to measure the performance of our functional nonparametric discrimination method, we set up two samples from the collected data sets. One is the learning sample, and the other is the testing sample. The learning sample allows us to estimate the posterior probabilities. The testing sample is useful for assessing the discriminant power. We can get the predicted groups of testing sample. The misclassification rate can be evaluated by the proportion of predicted groups not equal to the observed groups. Repeating this procedure 5000 times, we will get 5000 misclassification rates. The distribution of these misclassification rates can give a good idea of the power of nonparametric discrimination of functional data with various semi-metrics. Three types of semi-metrics are considered, and each of them has different forms being used. The different forms of three proximity measures are as follows:



- (i) PLS-type semi-metrics with a number of factors taking its values in 2, 3, 4, 5, 6, 7, 8 and 9 successively.
- (ii) PCA-type semi-metrics with a number of components taking its values in 2, 3, 4, 5, 6, 7 and 8 successively.
- (iii) Derivative-type semi-metrics with a number of derivatives equal to zero (classical L_2 -norm).

We first consider the recognition of traveling vehicles with fixed lane and speed.

In order to measure the performance of our functional nonparametric discrimination method, we build randomly two samples from the original dataset. The learning sample and testing sample have 18 functional observations for each of classes of vehicles. Repeating the procedure 5000 times, we will get 5000 misclassification rates. The Figure 4 displays the box-plot of misclassification rates for three types of semi-metrics proximity. The Figure 5 displays the scatter plot of mean and variance of misclassification rates for three types of semi-metrics proximity. Our goal is to predict the classes of traveling vehicles correctly. So we hope to find the robust methodology with lower misclassification rate. From the Figure 4 and Figure 5 we prefer PLS-type semi-metrics with 2 factors and PCA-type semi-metrics with 3 components to the others. The proportion that we can correctly recognize the class of traveling vehicle is about 84%. The Figure 6 displays the classification rate over 5000 runs with 2 classes of vehicles. The columns denote the real classes and the rows denote the predicted classes. The correct classification rates and misclassification rates are both exhibited in the Figure 6. The diagonal rates denote the correct classification rate over the 5000 runs. Others denote the misclassification rate over the 5000 runs.

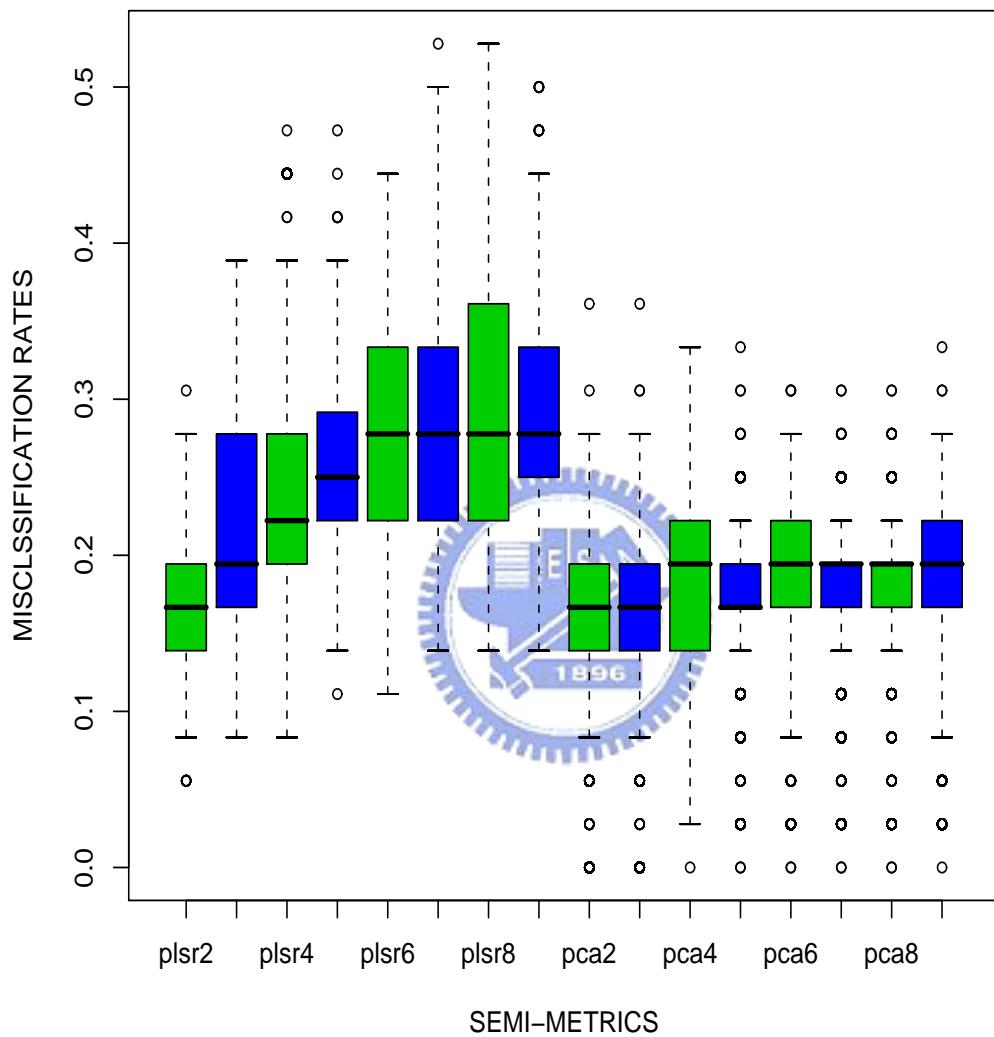


Figure 4: Dynamic vehicle recognition data discrimination over 5000 runs with 2 classes of vehicles.

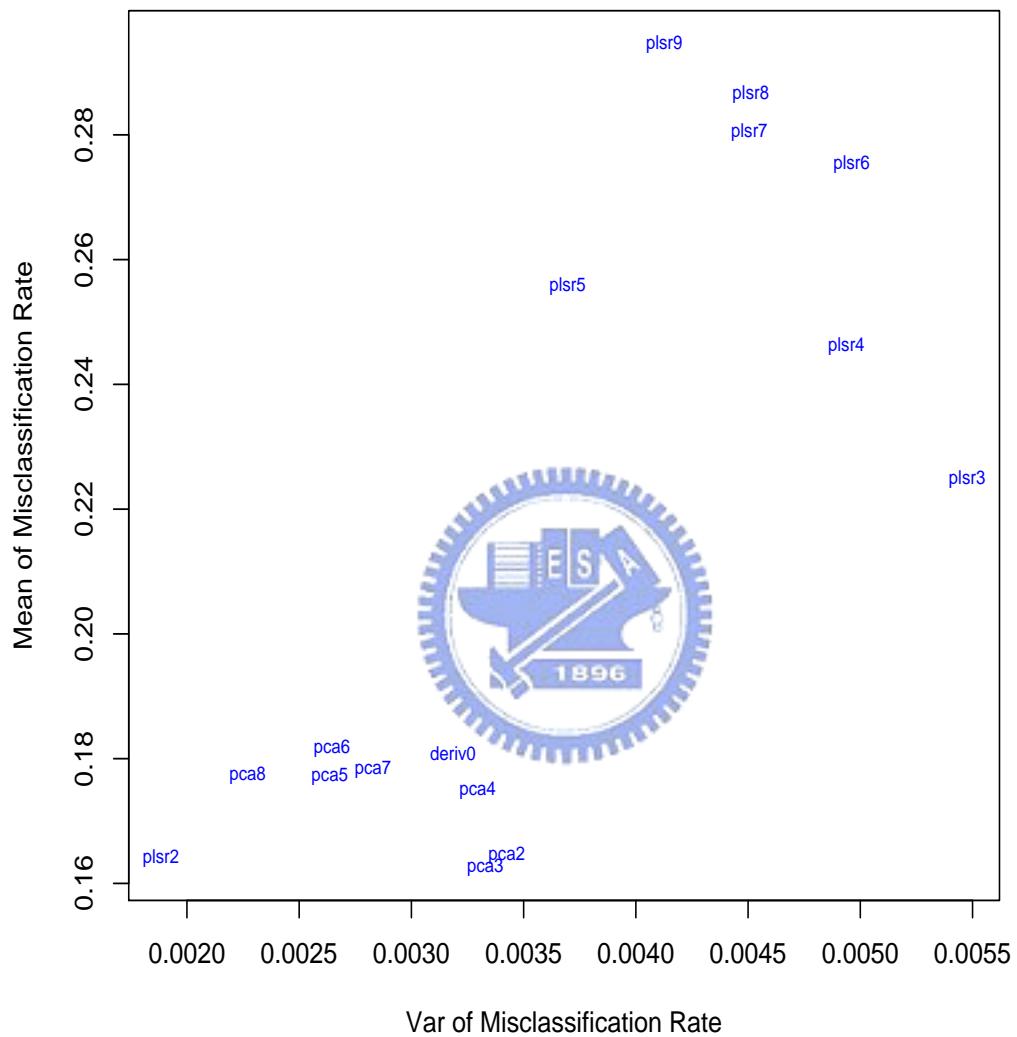


Figure 5: Scatter plot of mean and variance of misclassification rate over 5000 runs with 2 classes of vehicles.

PCA2		PCA3		PCA4	
S3	L3	S3	L3	S3	L3
45.61%	11.92%	45.95%	12.12%	45.96%	13.30%
4.39%	38.08%	4.05%	37.88%	4.04%	36.70%
PCA5		PCA6		PCA7	
S3	L3	S3	L3	S3	L3
46.18%	13.74%	46.19%	14.17%	46.20%	13.87%
3.82%	36.26%	3.81%	35.83%	3.80%	36.13%
PCA8		PLS2		PLS3	
S3	L3	S3	L3	S3	L3
46.51%	14.08%	47.82%	14.22%	43.87%	16.22%
3.49%	35.92%	2.18%	35.78%	6.13%	33.78%
PLS4		PLS5		PLS6	
S3	L3	S3	L3	S3	L3
42.25%	16.76%	42.12%	17.72%	41.34%	18.84%
7.75%	33.24%	7.88%	32.28%	8.66%	31.16%
PLS7		PLS8		PLS9	
S3	L3	S3	L3	S3	L3
40.66%	18.73%	39.87%	18.58%	39.08%	18.64%
9.34%	31.27%	10.13%	31.42%	10.92%	31.36%
Deriv0					
S3	L3				
46.28%	14.23%				
3.72%	35.77%				

Figure 6: The classification rate over 5000 runs with 2 classes of vehicles. The columns denote the real classes and the rows denote the predicted classes.

Our second aim is to classify the lanes and dynamic vehicles with fixed speed. In order to measure the performance of our functional nonparametric discrimination method, we also build randomly two samples from the original dataset. The numbers of functional observations of small vehicles on lane 1, 2, 3 and 4 take its value in 10, 10, 10 and 10. Then the numbers of functional observations of large vehicles on lane 1, 2, 3 and 4 take its value in 10, 10, 10 and 5. The size of learning sample and testing sample are the same. We also repeat 5000 times above procedure to get 5000 misclassification rates. The Figure 7 and 8 are analogous to the Figure 4 and Figure 5. We tend to regard PLS-type semi-metrics with 2 or 3 factors as the proximity for these data. The correct classification rate is about 42%, a dramatic drop due to increasing in number of classes. The Figure 9 displays the classification rate over 5000 runs with 8 classes of vehicles and lanes. The columns denote the real classes and the rows denote the predicted classes. The correct classification rates and misclassification rates are both exhibited in the Figure 9. The diagonal rates denote the correct classification rate over the 5000 runs. Others denote the misclassification rate over the 5000 runs.

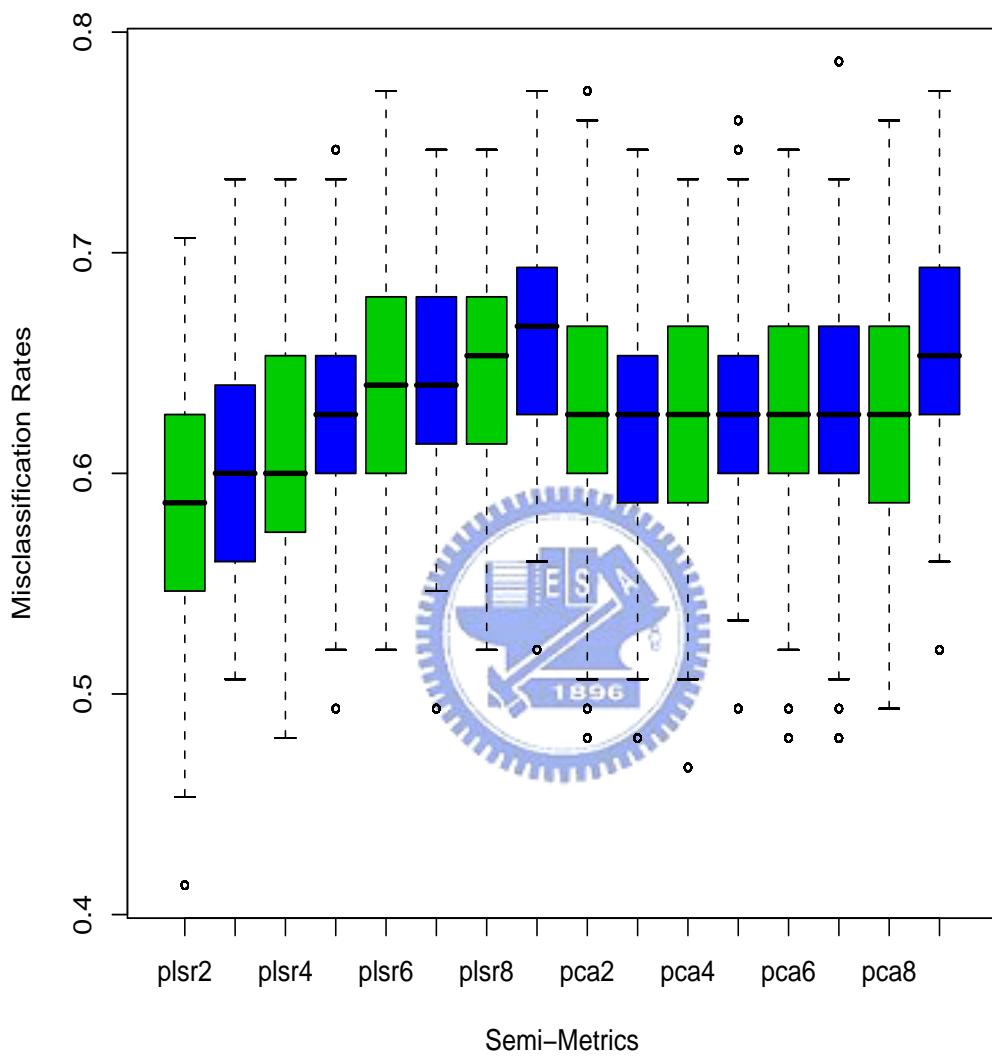


Figure 7: Dynamic vehicle recognition data discrimination over 5000 runs with 8 classes of vehicles and lanes.

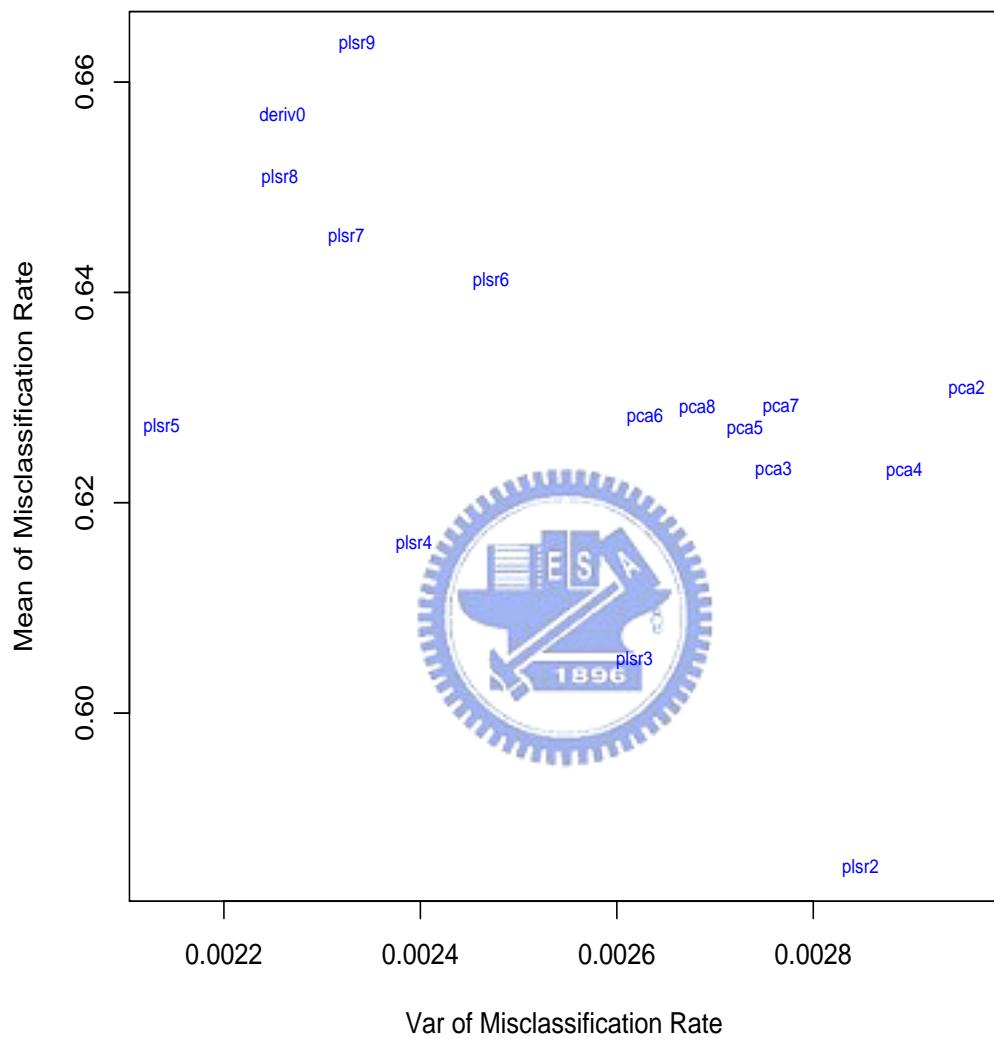


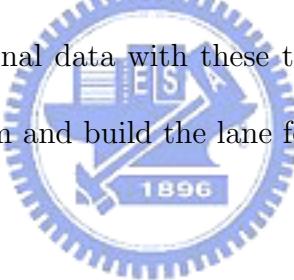
Figure 8: Scatter plot of mean and variance of misclassification rate over 5000 runs with 8 classes of vehicles and lanes.

PLS2							
S1	S2	S3	S4	L1	L2	L3	L4
7.38%	1.61%	0.54%	2.34%	0.41%	0.59%	1.09%	0.52%
3.22%	6.80%	3.36%	1.29%	1.00%	0.96%	1.55%	0.00%
0.69%	4.33%	7.29%	3.85%	0.96%	0.61%	1.35%	0.00%
1.51%	0.55%	2.09%	5.68%	0.07%	0.23%	0.29%	0.04%
0.18%	0.00%	0.01%	0.08%	5.34%	2.90%	2.68%	1.80%
0.19%	0.01%	0.00%	0.07%	2.73%	4.85%	3.99%	2.07%
0.15%	0.03%	0.04%	0.03%	2.39%	2.85%	2.22%	0.88%
0.00%	0.00%	0.00%	0.00%	0.43%	0.34%	0.16%	1.36%
PLS3							
S1	S2	S3	S4	L1	L2	L3	L4
7.15%	1.67%	0.77%	2.03%	1.09%	0.90%	1.27%	0.30%
2.46%	6.28%	2.61%	1.00%	0.92%	0.54%	1.43%	0.00%
1.06%	4.48%	7.15%	3.99%	0.84%	0.74%	1.38%	0.00%
1.85%	0.77%	2.80%	6.05%	0.07%	0.43%	0.48%	0.14%
0.14%	0.01%	0.00%	0.04%	3.91%	2.39%	2.02%	1.46%
0.29%	0.07%	0.00%	0.12%	3.15%	5.42%	3.63%	1.35%
0.39%	0.05%	0.00%	0.10%	2.71%	2.59%	2.73%	1.41%
0.00%	0.00%	0.00%	0.00%	0.65%	0.32%	0.41%	2.01%

Figure 9: The classification rate over 5000 runs with 8 classes of vehicles and lanes. The columns denote the real classes and the rows denote the predicted classes.

4 Conclusion

This research uses three semi-metrics proximity to deal with nonparametric discrimination of functional data. The result shows PLS-type semi-metrics proximity is more appropriate for the Radar microwave data set of the recognition of traveling vehicles. The vehicles functional data can be treated as multivariate data with 30 predictors, then PCA and PLS methodology implement data reduction with 2 or 3 dimension. But their recognition rate is higher than that with more factors. The large vehicles contain cranes, buses, trucks, goods wagons, etc. These vehicles have distinct forms, but we classify them analog. This may cause the recognition rate to be not as high as we expect. No matter how we expect that the nonparametric discrimination of functional data with these three semi-metrics proximity contribute to improve traffic jam and build the lane for a special purpose.



References

- [1] Boor, C. A. (1978). *A practical guide to splines*. Springer, New York.
- [2] Dauxois, J., Pousse, A., and Romain, Y. (1982) Asymptotic theory for the principal component analysis of a random vector function: some application to statistical inference. *Journal of Multivariate analysis*, 12, 136-154.
- [3] Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis*. New York: Springer.
- [4] Frank, I.E., and Friedman, J.H. (1993). A statistical view of chemometrics regression tools. *Technometrics*, 35, 109-148.
- [5] Geladi,P., and Kowlaski, B. (1986). Partial least square regression: A tutorial. *Analytica Chemica Acta*, 35, 1-17.
- [6] Helland, I.S. (1990). PLS regression and statistical models. *Scandivian Journal of Statistics*, 17, 97-114.
- [7] Herman, S. (2002). *A Particle Filtering Approach to Joint Passive Radar Trucking and Target Classification*, Doctoral Dissertation, Department of Electrical and Computer Engineering, Univ. of Illinois at Urbana-Champaign, Urbana, IL.

[8] Herman, S. and Moulin P. (2002). A particle filtering approach to joint radar tracking and automatic target recognition. in *Proc. IEEE Aerospace Conference*, (Big Sky, Montana), March 10-15.

[9] Hoskuldsson, P. (1988). PLS regression methods. *Journal of Chemometrics*, 2, 211-288.

[10] Jong, S. (1993) SIMPLS: an alternative approach to partial least squares regression. *chemometrics and Intelligent Laboratory Systems*, 18, 251-263.

[11] Ramsay, J., and Silverman, B. (1997). *Functional data analysis*. New York: Springer.

[12] Tenenhaus, M. (1998). *La regression PLS*. Paris: Technip.

[13] Wold, H. (1966). Estimation of principal components and related models by iterative least squares. In P.R. Krishnaiah (Ed.). *Multivariate Analysis*, 391-420. New York: Academic Press.

[14] Wold, H. (1975). Path models with latent variables: The NIPALS approach. In H.M. Blalock et al., editor, *Quantitative sociology: international perspectives on mathematical and statistical model building*, 307-357. New York: Academic Press.