IE Interfaces
Vol. 23, No. 4, pp. 365-371, December 2010.

Remote Health Monitoring of Parkinson's Disease Severity Using Signomial Regression Model

Young-Seon Jeong[1] · **Chungmok Lee**[2] · **Norman Kim**[3†] · **Kyungsik Lee**[4]

[1]Department of Industrial and Systems Engineering, Rutgers University
[2]Department of Industrial & Systems Engineering, KAIST
[3]Rutgers Center for Operations Research, Rutgers University
[4]Department of Industrial and Management Engineering, Hankuk University of Foreign Studies

파킨슨병 원격 진단을 위한 Signomial 회귀 모형

정영선[1] · **이충목**[2] · Norman Kim[3] · **이경식**[4]

[1]Rutgers University, Department of Industrial and Systems Engineering / [2]KAIST, 산업 및 시스템공학과 /
[3]Rutgers University, Rutgers Center for Operations Research / [4]한국외국어대학교 산업경영공학과

In this study, we propose a novel remote health monitoring system to accurately predict Parkinson's disease severity using a signomial regression method. In order to characterize the Parkinson's disease severity, sixteen biomedical voice measurements associated with symptoms of the Parkinson's disease, are used to develop the telemonitoring model for early detection of the Parkinson's disease. The proposed approach could be utilized for not only prediction purposes, but also interpretation purposes in practice, providing an explicit description of the resulting function in the original input space. Compared to the accuracy performance with the existing methods, the proposed algorithm produces less error rate for predicting Parkinson's disease severity.

Keyword: remote health monitoring, telemedicine, parkinson's disease, signomial regression, column generation

1. Introduction

Recently, there has been growing interest in remote health monitoring, so called telemedicine, by using advanced telecommunication systems such as video-conferencing and long distance home care tools (Handschu *et al*., 2003; Kong and Feng, 2001; Little *et al*., 2009; Tsanas *et al*., 2009; Westin *et al*., 2010). For telemedicine, diverse diseases including acute stoke, cancer, and brain injury detection have been taken into consideration (Handschu *et al*., 2003; Kong and Feng, 2001; Ricke and Bartelink, 2000). Parkinson's disease (PD) is the second largest neuro-logical disorders and it affects more than one million people in North America (Little *et al*., 2009; Tsanas

This research for the fourth author (Kyungsik Lee) was supported by Hankuk University of Foreign Studies Research Fund.
The authors acknowledge the support of Dr. Max Little in opening us the Parkinson's Telemonitoring data set.
†Corresponding author : Norman Kim, 640 Bartholomew Rd., Piscataway, NJ 08854, U.S.A., Fax : +1-732-445-5472,
E-mail : norman.kim@gmail.com
Received 12 October 2010, Revised 21 October 2010, Accepted 14 November 2010.

et al., 2009). Even though people with Parkinson disease (PWP) are rapidly increasing over an aging population, there is little knowledge about the root-cause and available cure (Tsanas *et al*., 2009). Therefore, it is critical to have regular checkups and early detection and diagnosis for offering effective alleviation of symptoms. Although most of PWP require regular clinic visits for monitoring and treatment, clinic examinations are costly and time-consuming for patients, and those are sometimes not possible because of distance limitations.

In order to overcome these problems, remote health monitoring methods of Parkinson's disease have been paid attention. Westin *et al*. (2010) proposed the telemonitoring assessment procedures of PWP state by using drawing impairment. Tsanas *et al*. (2009) presented the telemonitoring model for prediction of the Unified Parkinson's Disease Rating Scale (UPDRS) by using conventional regression models such as least-squares regression and classification and regression tree method (CART). In their paper, in order to predict the UPDRS, they used the dysphonia measures, which are assessed by acoustic signals. In addition, Little *et al*. (2009) proposed a detection model for discriminating healthy people from PWP by using dysphonia measures. They used the radial basis function (RBF) kernel-based support vector machines (SVMs) classifier.

However, previous existing procedures should extract several single or combination features characterizing UPDRS well prior to predicting UPDRS by using regression models. An exhaustive search of all possible combinations of these features is time-consuming and tedious works. In addition, previous kernel-based models represent implicit function description in high dimensional feature space, which is not suitable for interpretation purposes. Therefore, this paper proposes a novel model for remote health monitoring of Parkinson's disease severity. The proposed method is a modified version of signomial regression (SR) model firstly developed in Lee *et al*. (2010). One important advantage of SR is that an explicit description of the resulting function in the original input space can be obtained, which may give significant information to clinic experts to reveal the relationship between Parkinson's disease severity and its related measurements. In other words, SR can be utilized for not only prediction purposes, but also interpretation purposes in clinic. Here, we extend SR to permit more general loss function and regularization. <Figure 1> illustrates the framework of the proposed telemonitoring model for prediction of a Parkinson's disease severity.

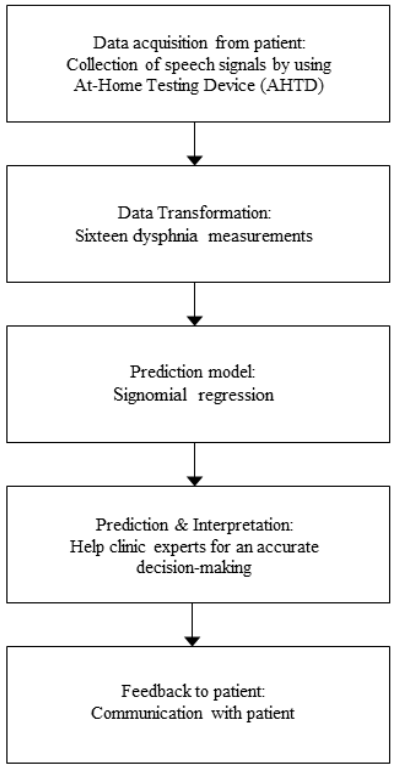


Figure 1. Framework of the proposed telemonitoring procedure for prediction of PD severity

Section 2 gives a detail description of the proposed model. Section 3 provides experimental results that compare the accuracy of the proposed method with that of previous methods. Concluding remarks and future studies are offered in Section 4.

2. Methods

Consider a real vector $x\in R^n_{++}$, and let x_i denote the ith component of x, where $i\in N$ and $N:=\{1,\cdots,n\}$. Define a function of d, $g_d := \prod_{i\in N} x_i^{d_i}$, where d_i is the ith component of a real vector $d\in R^n$. Then, a signomial function of x is defined as

$$f(x)=\sum_{d\in D} w_d g_d(x)+b \tag{1}$$

where $b\in R$, D is a finite subset of R^n such that $0\not\in D$, and $w_d\in R$ for all $d\in D$. A well-known special case of the signomial function is the polynomial function with degree k, where D is given as $\{d\in$

$Z_+^n | 1 \le \sum_{i \in N} d_i \le k \}$. Any signomial function f is defined by the weight vector (w, b) and the set of exponents D.

In SVMs, the original feature space is mapped to some high dimensional-space by the *kernel trick*. When a proper kernel is used, the hidden structure of given data may be revealed as some higher dimensional coupling function of the original variables. The SVMs are widely used recently, because of its simplicity and flexibility. For comprehensive tutorials on SVMs, we refer the readers to Burges (1998), Gunn (1998) and Smola and Schölkopf (2004). Nevertheless, the SVMs have some shortcomings : (i) Even the high dimensional interpretation is possible by the kernel trick, the interpretation is restricted to some *surface* in the high dimensional space because the weights between higher terms are fixed by the kernel parameters; (ii) It is not easy to obtain an original-space view from the kernel-tricked resulting regression function; (iii) The quality of the resulting solution may heavily depend on the choice of kernel function and its parameters. This work aims to handle these issues by proposing a systematical procedure constructing the signomial regression function. The signomial regression was firstly developed in Lee *et al*. (2010). We extend their results to permit more general loss function and regularization, and assert validity of the sigmonial regression by a computational study.

For the given m data points $(x_j, y_j) \in R_{++}^n \times R$ for all $j \in M := \{1, \cdots, m\}$, the goal is to find a signomial function which minimizes the regularized loss function $F := R_f + L_f$, where R_f and L_f are the terms for regularization and loss for signomial function f, respectively. Then, for the given exponents set D, the ϵ-insensitive regression problem is stated as follows :

$$\begin{aligned} & min\, R_f + L_f \\ & s.t.\ y_j - f_{w,b}^D(x_j) \le \epsilon + u_j,\ j \in M \\ & \quad f_{w,b}^D(x_j) - y_j \le \epsilon + v_j,\ j \in M \\ & \quad w \in R^{|D|},\ b \in R,\ u \in R_+^m,\ v \in R_+^m \end{aligned}$$

where the variables u_j and v_j are used to measure prediction errors.

For the ease of later exposition, we rewrite $w = w^+ - w^-$, where $w^+, w^- \in R_+^{|D|}$. Then, signomial regression problem (SRP) is given as follows :

$$\begin{aligned} & min\, R_f + L_f \\ & s.t.\ y_j - \sum_{d \in D} (w_d^+ - w_d^-) g_d(x_j) - b \le \epsilon + u_j,\ j \in M \\ & \quad \sum_{d \in D} (w_d^+ - w_d^-) g_d(x_j) + b - y_j \le \epsilon + v_j,\ j \in M \\ & \quad w^+,\ w^- \in R_+^{|D|},\ b \in R,\ u \in R_+^m,\ v \in R_+^m \end{aligned}$$

Note that the feasible set of SRP is a polyhedron, since the exponents set D is finite. We consider following four cases of the regularized loss functions :

$$\begin{aligned} [F_{L=1}^{R=1}] &:= \| w \|_1 + C \sum_{j \in M} (\| u \|_1 + \| v \|_1) \\ [F_{L=2}^{R=1}] &:= \| w \|_1 + C \sum_{j \in M} (\| u \|_2^2 + \| v \|_2^2) \\ [F_{L=1}^{R=2}] &:= \| w \|_2^2 + C \sum_{j \in M} (\| u \|_1 + \| v \|_1) \\ [F_{L=2}^{R=2}] &:= \| w \|_2^2 + C \sum_{j \in M} (\| u \|_2^2 + \| v \|_2^2) \end{aligned}$$

where C is a given constant, $\| w \|_1 := \sum_{d \in D} w_d$, and $\| w \|_2^2 := \sum_{d \in D} w_d^2$. $\| u \|_1$ and $\| u \|_2^2$ ($\| v \|_1$ and $\| v \|_2^2$) are defined in a similar manner. Note that we can replace $\| w \|_2^2$ with $\| w^+ \|_2^2 + \| w^- \|_2^2$, since either w_d^+ or w_d^- is zero at the optimal solution for any $d \in D$. Let $(SRP)_b^a$ denote regression problem SRP with objective function $F_{L=b}^{R=a}$. Note that $(SRP)_1^1$ is a linear programming (LP) problem, while others are quadratic programming (QP) problems. The LP and QP problems can be solved efficiently (when $|D|$ is polynomially bounded) by some well-known algorithms such as Simplex and Barrier algorithm.

2.1 An Implementation of SRP

Unfortunately, the exponents set D can be exponentially large, which makes solving the problem directly intractable. In this section, we propose a column generation method for SRP. The column generation method iteratively finds the *beneficial* exponents not included in the current problem, so that we do not need to solve the problem with all of possible exponents explicitly. Assume we have $\widetilde{D} \subset D$. Then, the column generation *master problem*, $(MSR)_b^a$, is given as follows :

$$\begin{aligned} & min\, F_{L=b}^{R=a} \\ & s.t.\ y_j - \sum_{d \in \widetilde{D}} (w_d^+ - w_d^-) g_d(x_j) - b \le \epsilon + u_j,\ j \in M && (2) \\ & \quad \sum_{d \in \widetilde{D}} (w_d^+ - w_d^-) g_d(x_j) + b - y_j \le \epsilon + v_j,\ j \in M && (3) \\ & \quad w^+,\ w^- \in R_+^{|\widetilde{D}|},\ b \in R,\ u \in R_+^m,\ v \in R_+^m \end{aligned}$$

Let $\alpha_j \geq 0$ and $\beta_j \geq 0$ denote the dual variables corresponding with (2) and (3). The following theorem says that the optimal solution of $(MSR)_b^a$ becomes also the optimal solution for $(SRP)_b^a$ also.

Theorem 2.1 : A solution $\left(\widehat{w^+}, \widehat{w^-}, \hat{b}, \hat{u}, \hat{v}, \hat{\alpha}, \hat{\beta}\right)$ is an optimal solution for $(SRP)_b^a$, if the optimal solution of $(MSR)_b^a$, $\left(\widetilde{w}^+, \widetilde{w}^-, \hat{b}, \hat{u}, \hat{v}, \hat{\alpha}, \hat{\beta}\right)$ satisfies the following conditions :

$$\sum_{j\in M}\left(\hat{\alpha}_j g_d(\hat{x}_j) - \hat{\beta}_j g_d(\hat{x}_j)\right) \leq \xi,\ \forall d \in D \setminus D$$
$$\sum_{j\in M}\left(\hat{\alpha}_j g_d(\hat{x}_j) - \hat{\beta}_j g_d(\hat{x}_j)\right) \geq -\xi,\ \forall d \in D \setminus D$$
$$\widehat{w_d^+} = \begin{cases} \widetilde{w_d^+}, & if\ d \in \widetilde{D} \\ 0, & otherwise \end{cases},\quad \widehat{w_d^-} = \begin{cases} \widetilde{w_d^-}, & \text{if}\ d \in \widetilde{D} \\ 0, & otherwise \end{cases}$$

where $\xi = 1$ if $a = 1$, 0 if $a = 2$ (See <Appendix 1> for the proof).

We consider set D that is defined by four parameters, d_{max}, d_{min}, T, and S, as follows :

$$D = \left\{d \in R^n \mid d_{min} \leq d_i \leq d_{max}, i \in N, \sum_{i\in N}|d_i| \leq S, Td \in Z^n\right\}$$

where $T > 0$ and $S > 0$. We have a polynomial model of degree less than or equal to k, when $d_{max} = k, d_{min} = 0, T = 1$ and $S = k$ for some positive integer k.

Let $z(d) := \sum_{z\in M}(\hat{\alpha}_j g_d(\hat{x}_j) - \hat{\beta}_j g_d(\hat{x}_j))$. The column generation problem is given as follows :

$$(CG+)\ \ z^+ := max\{z(d) \mid d \in D \setminus \widetilde{D}\}$$
$$(CG-)\ \ z^- := min\{z(d) \mid d \in D \setminus \widetilde{D}\}$$

The column generation is terminated if $z^+ \leq \xi$ and $z^- \geq -\xi$, where $\xi = 1$ if $a = 1$, 0 if $a = 2$. Otherwise, an exponent d^* such that $z(d^*) > \xi$ (or $z(d^*) < -\xi$) is added to the $(MSR)_b^a$, and the procedure is repeated. Note that, when $a = 1$, we can replace $D \setminus \widetilde{D}$ with D in the problems above. This is counterintuitive when we solve the column generation problems by enumerating all exponents in D - the brute force method. It is, however, helpful when we consider an optimization algorithm for the column generation problems. In the following section, we propose a heuristic algorithm to solve $(CG+)$ and $(CG-)$, when $a = 1$.

2.2 Column Generation Algorithm

From now, we only concern $(CG+)$, since the following result can be readily generalized to $(CG-)$. When $a = 1$, $(CG+)$ can be formulated as follows :

$$max \sum_{j\in M}\hat{\alpha}_j \prod_{i\in N}\hat{x}_j^{d_i} - \sum_{j\in M}\hat{\beta}_j \prod_{i\in N}\hat{x}_j^{d_i} \tag{4}$$
$$s.t.\ d_{min} \leq d_i \leq d_{max},\ \forall i \in N \tag{5}$$
$$\sum_{i\in N}|d_i| \leq S \tag{6}$$
$$Td_i \in Z,\ \forall i \in N \tag{7}$$

Theorem 2.2 : $(CG+)$ *is NP-hard* (Lee *et al.*, 2010).

To get an approximate solution to $(CG+)$, we first solve the continuous relaxation of $(CG+)$, which is obtained by dropping the integrality constraints in the problem. Since the objective function of $(CG+)$ is not convex nor concave, we use a heuristic algorithm based on the Frank-Wolfe algorithm (Frank and Wolfe 1956) to obtain a local optimal solution. After obtaining a local optimal solution, we round the solution to make the exponent vector multiplied by T, Td, integral. In $(CG+)$, the parameters d_{min}, d_{max}, S, T need to be appropriately selected. The selected values of four parameters may have an impact on the performance of the resulting functions. However, our computational study shows that regression functions obtained by SR, with $d_{min} = -1$, $d_{max} = 1$, $S = 1$, and $T = 10$ show competitive prediction accuracy for the tested data set.

To obtain the steepest accent direction, we have to solve the following problem :

$$max\, c_0 + \sum_{i\in N} c_i d_i,\ s.t.\ d \in \{d \in R^n \mid (5)\,\text{and}\,(6)\}$$

where c_0 and c_is are some constants given from the current iteration's solution point. It is noteworthy that this problem can be easily solved by sorting c_is in the decreasing order of absolute values and then assigning the value of d_is in a greedy manner.

3. Experimental Results

Computational experiments are conducted using Parkinsons Telemonitoring data set from the UCI Machine Learning Repository (Murphy and Aha, 1992) to compare the proposed SR with the existing methods in terms of the mean absolute error (MAE) of prediction.

Although the etiology of PD is usually unknown, there are several symptoms of PD such as tremor, rigidity, and voice impairment (Ho *et al*., 1998). Among them, vocal impairment is a particular important symptom for physical examinations in clinic office (Gamboa *et al*., 1997). For example, the symptoms in voice such as hypophonia (reduced voice volume) and dysphonia (breathiness, hoarseness or creakiness in the voice) may happen as PD progresses. Thus, speech signals have been widely used for telemonitoring for PD detection (Little *et al*., 2009; Tsanas *et al*., 2009). After collecting those signals from people with early stage PD, the speech signals is transformed to 16 biomedical voice measurements for Parkinson's telemonitoring. The transformed data are aimed to predict the motor and total UPDRS scores ('motorUPDRS' and 'totalUPDRS'), which reflects the presence and severity of PD symptoms, with a variety of voice measurements. For details on data sets, see the reference (Tsanas *et al*., 2009). The Parkinsons Telemonitoring data set originally consists of 5875 observations. Among them, we selected 400 data points randomly for the training data. Validation and testing samples of 300 data points, respectively, were sampled in the same manner.

The performance of SR is compared to that of CART (Brieman *et al*., 1984) and support vector regression (SVR) (Vapnik, 1995) in terms of prediction accuracy (i.e. MAE). In SR, $(SRP)_1^1$ and $(SRP)_2^1$ with the terms for regularization $\| w \|_1$ are considered. For comparisons, SR was implemented by using Xpress Mosel language (Xpress-MP, 2008). All computations for CART and SVR were conducted in the Matlab environment (Matlab statistics toolbox, 2008).

For SR, we defined the set D by $d_{min} = -1$, $d_{max} = 1$, $S = 1$, and $T = 10$, which means polynomial terms up to degree 10 are considered in the column generation by scaling implicitly the original data x to $x^{1/10}$. Also, we tested various model parameters C such that $C = 10^{-3}$, 10^{-2}, 10^{-1}, 1, 10^1, 10^2 and 10^3. The value of ϵ was varied such that $\epsilon = 10^{-3}$, 10^{-2}, 10^{-1}, 1, 10^1, 10^2 and 10^3.

For CART, the pruning step is necessary to avoid overfitting the training data. Therefore, we obtained the optimal regression tree by varying the pruning level.

For SVR, Gaussian RBF kernels with $\sigma = 10^{-3}$, 10^{-2}, 10^{-1}, 1, 10^1, 10^2 and 10^3 are all tested. In addition, we varied ϵ for the ϵ-insensitive loss function such that $\epsilon = 10^{-3}$, 10^{-2}, 10^{-1}, 1, 10^1, 10^2 and 10^3. We also tested various values of C such that $C = 10^{-3}$, 10^{-2}, 10^{-1}, 1, 10^1, 10^2 and 10^3.

For each combination of parameters in each method, we performed three independent experimental runs. Then, the average MAE for the best parameters with respect to the average MAE for validation data are summarized in <Table 1>.

Table 1. Prediction accuracy-average test error in terms of MAE for the best parameter combination(standard deviations in parentheses)

Methods	motorUPDRS	totalUPDRS
CART	7.211(0.379)	8.964(0.511)
SVR	7.085(0.366)	8.684(0.586)
$(SRP)_1^1$	6.690(0.475)	8.397(0.587)
$(SRP)_2^1$	6.632(0.576)	8.650(0.568)

From <Table 1>, we observe that the proposed SR approaches yield better results (i.e. smaller MAE values) than existing methods regardless of the loss functions, and more importantly, SR performs slightly better than SVR, which has have shown promising empirical results in many practical applications.

Note that SR also gives an explicit description of the resulting function in the original input space while, SVR is not capable of defining the function explicitly. That is, the proposed SR, unlike SVR, can provide an explicit regression function with the selected input variables. For the Parkinsons Telemonitoring data set, the best regression functions for motorUPDRS and totalUPDRS were obtained from $(SRP)_2^1$ and $(SRP)_1^1$, respectively, as follows.

$$
\begin{aligned}
& f_{motor\,UPDRS} = \\
& \quad 0.100 \times APQ3^{-1} \\
& - 0.016 \times Jitter(\%)^{-1} \\
& - 0.002 \times Jitter(Abs)^{-1} \\
& - 0.013 \times Jitter(Abs)^{-0.9} \\
& - 0.002 \times Jitter(Abs)^{-0.9} \times APQ3^{-0.1} \\
& + 0.018 \times Jitter(Abs)^{-0.9} \times DFA^{-0.1} \\
& + 0.004 \times Jitter(Abs)^{-0.9} \times Shimmer(dB)^{-0.1} \\
& + 0.003 \times Jitter(Abs)^{-0.9} \times NHR^{-0.1} \\
& - 0.002 \times Jitter(Abs)^{-0.8} \times NHR^{-0.1} \times PPE^{0.1} \\
& - 0.081 \times Jitter(Abs)^{-0.8} \times PPE^{0.2} \\
& + 0.010 \times Jitter(Abs)^{-0.8} \times RAP^{-0.1} \\
& - 0.005 \times Jitter(Abs)^{-0.8} \times APQ5^{-0.1} \times APQ11^{-0.1} \\
& + 0.212 \times Jitter(Abs)^{-0.7} \times PPE^{0.3} \\
& - 0.039 \times NHR^{-1} \\
& - 0.014 \times PPQ5^{-1} \\
& - 0.012 \times RAP^{-1} + 2.22
\end{aligned}
$$

$$\begin{aligned}
f_{totalUPDRS} = \\
& 0.001 \times Jitter(Abs)^{-1} \\
& - 0.023 \times Jitter(Abs)^{-0.9} \\
& + 0.008 \times Jitter(Abs)^{-0.9} \times DFA^{-0.1} \\
& + 0.006 \times Jitter(Abs)^{-0.9} \times Jitter(\%)^{0.1} \\
& + 0.002 \times Jitter(Abs)^{-0.9} \times PPE^{0.1} \\
& - 0.008 \times Jitter(Abs)^{-0.9} \times APQ11^{0.1} \\
& + 0.002 \times Jitter(Abs)^{-0.8} \times Shimmer^{-0.2} + 26.46
\end{aligned}$$

We observe that both resulting functions are comprised of several measures such as Jitter (Abs), shimmer, DFA and PPE. This is an encouraging result since the the measures were selected as significant variables based on information criteria (AIC and BIC) in Tsanas *et al*. (2009). As expected, the regression functions have small numbers of terms due to the L_1 regularization-16 and 7, respectively. It is noteworthy that the each term is composed of small number of original features (3 features at most). Recall that, in the column generation procedure, we iteratively update the exponent vector d^k. When updating each component of d^k, the steepest direction is obtained by a greedy manner. So, bigger components in d^k are getting bigger while many other components approach to zero, which yields some sparse vector d.

In addition, <Table 2> shows the average number of selected terms for the resulting signomial functions in three experimental runs. From <Table 2>, we observe that $(SRP)^1_1$ has fewer terms than $(SRP)^1_2$. This is consistent with the common understanding that the quadratic cost function leads to more support vectors in SVMs (Chung *et al*., 2003). This result tells us that it is more preferable to use L_1 loss function than L_2 loss function when we are interested in the number of terms in the regression function.

Table 2. The number of terms in the resulting signomial function

	motorUPDRS	totalUPDRS
$(SRP)^1_1$	22.7(16.6)	31.7(25.3)
$(SRP)^1_2$	51.0(30.4)	59.7(47.2)

4. Conclusion

A novel model to predict Parkinson's disease severity for remote health monitoring is developed by extending signomial regression model, and is compared to the existing methods in terms of MAE using Parkinsons Telemonitoring data set. Experimental results show that the proposed approach performs better than the existing methods. This is an encouraging result since the proposed approach, unlike SVR, gives an explicit function description in original input space, which can be used for prediction as well as interpretation. Fruitful areas of future research may include conducting more extensive computational experiments to further verify the findings and extending the proposed approach to the case of the variable selection.

References

Brieman, L., Friedman, J., Olshen, R. and Stone, C. (1984), *Classication and Regression Trees*, Chapman and Hall.

Burges, C. J. C. (1998), A Tutorial on Support Vector Machines for Pattern Recognition, *Data Mining and Knowledge Discovery*, 2, 121-167.

Chung, K. M., Kao, W. C., Sun, C. L., Wang, L. L. and Lin, C. J. (2003), Radius Margin Bounds for Support Vector Machines with the RBF Kernel, *Neural Computation*, 15(11), 2643-2681.

Frank, M. and Wolfe, P. (1956), An Algorithm for Quadratic Programming, *Naval Research Logistics Quarterly*, 3(1-2), 95-110.

Gamboa, J., Jiménez-Jiménez, F. J., Nieto, A., Montojo, J., Ortí-Pareja, M., Molina, J. A., García-Albea, E. and Cobeta, I. (1997), Acoustic Voice Analysis in Patients with Parkinson's Disease Treated with Dopaminergic Drugs, *Journal of Voice*, 11(3), 314-320.

Gunn, S. R. (1998), Support Vector Machines for Classification and Regression, *Technical Report of School of Electronics and Computer Science, University of Southampton*.

Handschu, R., Littmann, R., Reulbach, U., Gaul, C., Heckmann, J., Neundörfer, B., and Scibor, M. (2003), Telemedicine in Emergency Evaluation of Acute Stroke, *Journal of the American Heart Association*, 34, 2842-2846.

Ho, A., Iansek, R., Marigliani, C., Bradshow, J. and Gates, S. (1998), Speech Impairment in a Large Sample of Patients with Parkinson's Disease, *Behavioral Neurology*, 11, 131-137.

Kong, X. and Feng, R. (2001), Watermarking Medical Signals for Telemedicine, *IEEE Transactions on Information Technology in Biomedicine*, 5(3), 195-201.

Lee, K., Kim, N., and Jeong, M. (2010), Sparse Signomial Classification and Regression, *RUTCOR Research Reports*.

Little, M., McSharry, P., Hunter, E., Spielman, J. and Ramig, L. (2009), Suitability of Dysphonia Measurements for Telemonitoring of Parkinsons Disease, *IEEE Transactions on Biomedical Engineering*, 56(4), 1015-1022.

Matlab statistics toolbox (2008), URL : http://www.mathworks.com.

Murphy, P. M. and Aha, D. W. (1992), UCI Machine Learning Repository, URL : http://archive.ics.uci.edu/ml/.

Ricke, J. and Bartelink, H. (2000), Telemedicine and its Impact on Cancer Management, *European Journal of Cancer*, 36 (7), 826-833.

Smola, A. J. and Schölkopf, B. (2004), A Tutorial on Support Vector Regression, *Statistics and Computing*, 14, 199-222.

Tsanas, A., Little, M., McSharry, P. and Ramig, L. (2009), Accurate Telemonitoring of Parkinsons Disease Progression by Non-Invasive Speech Tests, *IEEE Transactions on Biomedical Engineering*, 57(4), 884-893.

Vapnik, V. N. (1995), The Nature of Statistical Learning Theory, Springer.

Westin, J., Ghiamati, S., Memedi, M., Nyholm, D., Johansson, A., Dougherty, M. and Groth, T. (2010), A New Computer Method for Assessing Drawing Impairment in Parkinson's Disease, *Journal of Neuroscience Methods*, 190(1), 143-148.

Xpress-MP (2008), URL : http://www.dashoptimization.com.

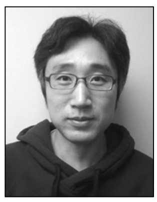
<Appendix 1> Proof of Theorem 2.1

Proof : When $a=b=1$, the result follows immediately by the LP duality theory. We only consider when $a=b=2$ here, since the other cases can be easily shown in a similar manner. Let $\pounds$ be the Lagrangian dual function of $(MSR)_2^2$, then by the optimality condition, the optimal solution $\left(\widehat{w^+}, \widehat{w^-}, \hat{b}, \hat{u}, \hat{v}, \hat{\alpha}, \hat{\beta}\right)$ should satisfy the following conditions :

$$\frac{\partial \pounds}{\partial w_d^+} \geq 0 \Rightarrow \sum_{j \in M} (\hat{\alpha}_j g_d(\hat{x}_j) - \hat{\beta}_j g_d(\hat{x}_j)) \leq 2\hat{w}_d^+, \ \forall d \in D$$
$$\frac{\partial \pounds}{\partial w_d^-} \geq 0 \Rightarrow \sum_{j \in M} (\hat{\alpha}_j g_d(\hat{x}_j) - \hat{\beta}_j g_d(\hat{x}_j)) \geq -2\hat{w}_d^-, \ \forall d \in D$$
$$\frac{\partial \pounds}{\partial u^j} \geq 0 \Rightarrow \hat{\alpha}_j \leq 2C\hat{u}_j, \ \forall j \in M$$
$$\frac{\partial \pounds}{\partial v^j} \geq 0 \Rightarrow \hat{\beta}_j \leq 2C\hat{v}_j, \ \forall j \in M$$

By assumption, $\left(\hat{w}^+, \hat{w}^-, \hat{b}, \hat{u}, \hat{v}, \hat{\alpha}, \hat{\beta}\right)$ satisfies the conditions above, so it is dual feasible. Moreover, clearly the solution is primal feasible to $(SRP)_2^2$, and the objective values are coincide ($F_{L=2}^{R=2} = \pounds$) at the solution. The result follows immediately.

Young-Seon Jeong
Ph.D. Candidate, Department of Industrial and Systems Engineering, Rutgers, The State University of New Jersey
MS : Korea University
BS : Chonnam National University
Research Topics : Statistical Data Mining and its Applications



Chungmok Lee
Postdoctoral Fellow, Department of Industrial and Systems Engineering, KAIST
Ph.D and MS : KAIST
BS : Korea University
Research Topics : Optimization Theory and Applications

Norman Kim
Postdoctoral Fellow, Rutgers Center for Operations Research (RUTCOR), Rutgers, The State University of New Jersey
Ph.D and MS and BS : KAIST
Research Topics : Datamining and its Applications

Kyungsik Lee
Professor, Department of Industrial & Management Engineering, Hankuk University of Foreign Studies
Ph.D and MS : Korea Advanced Institute of Science and Technology
BS : Seoul National University
Research Topics : Optimization Theory and Applications