Determination of the Optimal Threshold Value That Can Be Discriminated by Dysphonia Measurements for Unified Parkinson's Disease Rating Scale

Betul Erdogdu Sakar, C. Okan Sakar, Gorkem Serbes, Olcay Kursun

Abstract—**Recently, there is an increasing motivation to develop telemonitoring systems that enable cost-effective screening of Parkinson's Disease (PD) patients. These systems are generally based on measuring the motor system disorders seen in PD patients by the help of non-invasive data collection tools. Vocal impairments one of the most commonly seen PD symptoms in the early stages of the disease, and building such telemonitoring systems based on detecting the level of vocal impairments results in reliable motor UPDRS tracking systems. In this paper, we aim to determine the optimal UPDRS threshold value that can be discriminated by the vocal features extracted from the sustained vowel phonations of PD patients. For this purpose, we used an online available PD telemonitoring dataset consisting of speech recordings of 42 PD patients. We converted the UPDRS prediction problem into a binary classification problem for various motor UPDRS threshold values, and fed the features to k-Nearest Neighbor and Support Vector Machines classifiers to discriminate the PD patients whose UPDRS is less than or greater than the specified threshold value. The results indicate that speech disorders are more significantly seen in the patients whose UPDRS exceeds the experimentally determined threshold value (15). Besides, considering that the motor UPDRS ranges from 0 to 108, relatively low UPDRS threshold of 15 validates that vocal impairments can be used as early indicators of the disease.**

I. Introduction

PARKINSON'S Disease (PD) is the second most common neurodegenerative disorder ensuing Alzheimer [1]. It causes symptoms in the motor abilities of the patients such as tremor, dysphonia (breathiness, hoarseness, or creakiness in the voice), hypophonia (reduced voice volume), hypokinesia, rigidity, and postural instability [2]-[7]. In recent years, many clinical decision support systems based on machine learning algorithms have been proposed for the diagnosis and monitoring of PD based on the measurements of these motor system disorders. These systems generally focus on two main issues: (1) discriminating PD patients from healthy subjects [5], [6], [8], [9]; and (2) predicting the clinical evaluation metrics used to track the symptom progression of the disease [7] [10]. Considering that PD is generally seen in elderly people and so the frequent clinical visits of these patients are inconvenient, these systems are expected to be widely used in future since they are based on non-invasive data collection tools, enable the telemonitoring of the disease, and can be applied without expert intervention.

Vocal impairment is one of the most important symptoms of PD since it is seen in approximately 90% of the patients in the earlier stages of the disease [2], [7], [11]. Therefore, there is an increasing interest in building PD diagnosis and telemonitoring systems based on vocal features. The aim of this study is to determine the optimal UPDRS threshold value that can be discriminated by the vocal features extracted from the sustained vowel phonations of PD patients. For this purpose, we used an online available [12] Parkinson's Disease (PD) telemonitoring dataset consisting of speech recordings of 42 PD patients [7]. In this dataset, each input pattern consists of 16 dysphonia measurements extracted from the voice recordings of the patients.

The PD dataset used in this study was originally used to map the vocal features to UPDRS as a regression problem in order to track the progression of the disease remotely [7]. In this study, we converted the UPDRS prediction problem into a binary classification problem for various motor UPDRS threshold values by labeling an input pattern as positive if the corresponding motor UPDRS value is greater than the specified threshold value, and negative otherwise. Then, we fed the vocal features to *k*-NN and SVM classifiers for each of the binary class learning problems obtained with various UPDRS threshold values, and reported the optimal threshold value which has been agreed upon by both *k*-NN and SVM classifiers. The results indicate that speech disorders are more significantly seen in the patients whose UPDRS exceeds the experimentally determined threshold value (15). Besides, considering that the motor UPDRS ranges from 0 to 108, with 0 denoting symptom free and 108 denoting severe motor impairment [7], the lowness of the determined threshold validates that speech disorders are seen in the early stages of the disease.

The remainder of this paper is organized as follows. Section 2 presents the description of the PD dataset and the methods used to determine the optimal UPDRS threshold based on

Manuscript received August 15, 2015.

B. Erdogdu Sakar is with the Department of Software Engineering, Bahcesehir University, Istanbul, Turkey (phone: 90-212-381-05-89; fax: 90-212-381-05-50; e-mail: betul.erdogdu@eng.bahcesehir.edu.tr).

C. Okan Sakar is with the Department of Computer Engineering, Bahcesehir University, Istanbul, Turkey (e-mail: okan.sakar@eng.bahcesehir.edu.tr).

G. Serbes is with the Department of Biomedical Engineering, Yildiz Technical University, Istanbul, Turkey (e-mail: gserbes@yildiz.edu.tr).

Olcay Kursun is with the Department of Computer Engineering, Istanbul University, Istanbul, Turkey (e-mail: okursun@istanbul.edu.tr)

vocal features. Section 3 presents the experimental setup and results. We conclude in Section 4.

II. MATERIALS AND METHODS

A. Dataset Description

In this study, we used Parkinson's Disease (PD) telemonitoring dataset consisting of speech recordings of 42 PD patients, which is available online at UCI machine-learning archive [12]. This dataset was collected remotely at patient's home and transmitted over the internet [7]. The PD patients were diagnosed with PD within the previous five years at trial onset (mean ±std. 72 ± 69, min. 1, max. 260, median 48 weeks since diagnosis) if he had at least two of the following symptoms: rest tremor, bradykinesia or rigidity, without evidence of other forms of parkinsonism (mean age, 64.4 ± 9.24; age range, 36-85; median, 65) [7]. We refer the reader to [7] for more detailed description of the dataset.

The PD patients were monitored for a six-month period, and remained un-medicated during the duration of the study. The voice recordings of the subjects were obtained at weekly intervals for the six-month duration of the study whereas motor and total UPDRS were assessed only three times by the medical staff: at baseline (onset of trial), and after three and six months. Therefore, Tsanas et al. [7] used linear interpolation to obtain weekly UPDRS estimates corresponding to the weekly voice recordings. During the six months data collection period, in each trial, six sustained phonations of the vowel “ahhh…” were recorded summing up to 5875 voice recordings.

The feature set extracted from the voice recordings consists of 16 dysphonia measurements. The feature set includes several measures of fundamental frequency, several measures of variation in amplitude, noise-to-harmonics and harmonics-to-noise ratios, nonlinear dynamical complexity measure, signal fractal scaling exponent, and pitch period entropy.

B. Determination of UPDRS Threshold based on Dysphonia Measurements

In this study, the aim is to determine the optimal UPDRS threshold value that can be discriminated with the lowest possible error rate using dysphonia measurements. For this purpose, we converted the original UPDRS prediction problem into a binary classification problem for various motor UPDRS threshold values. In each setting, if the motor UPDRS value of an input pattern is greater than the threshold value, it is labeled as positive, otherwise negative. The interval of the UPDRS threshold value that has been evaluated was determined so that each of the classes contains at least 10% of the total number of samples.

The features are fed to SVM and k-NN classifiers for each of the binary class learning problems obtained with various UPDRS threshold values. Although we present the results in terms of accuracy and Matthew’s Correlation Coefficient (MCC) evaluation metrics, since the binary classification problems obtained according to various UPDRS threshold values may result in imbalanced datasets in which at least one of the classes contains is represented by only a small number of training examples when compared to the number of training examples has sample from one class is in higher number than other, we take the MCC metric into account to determine the maximally predictable UPDRS threshold value. The MCC metric is a balanced measure which can be used even if the classes are of very different sizes. It gets a value between –1 and +1. The formulation of MCC metric is given below:

$$MCC = \frac{TPxTN - FPxFN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (1)$$

where TP and TN represents the number of correctly classified positive and negative examples, respectively, and FP and FN represents the number of incorrectly classified positive and negative examples, respectively. MCC gets the value of +1 when the classifier makes perfect predictions, –1 when the predictions and actual values totally disagree, and 0 when the classification is no better than a random prediction.

III. EXPERIMENTS

A. Experimental Setup

We first normalized the features of the PD dataset so that each has a zero mean and unit variance. Then, the features are fed into SVM and k-NN classifiers for various motor UPDRS threshold values. For cross-validation, we used 70% percent of the samples for training, and rest for validation. For k-NN classifier, we used Euclidean, city-block, and correlation as distance metrics with various number of nearest neighbor values (*k*). For SVM classifier, we used LIBSVM implementation [13] with linear and Radial Basis Function (RBF) kernels with different values of cost (C) and kernel width (g) parameters. Depending on the value of the UPDRS threshold value, the PD dataset becomes highly imbalanced, and SVM classifier tends to label the samples as majority class to minimize the error on the training set. Therefore, in order to tackle with the class imbalance problem, we used the "class-weight" parameter of LIBSVM, *w*, which is used to increase the cost of errors made on the training samples of minority class.

B. Experimental Results

We present the classification performance of SVM with linear and RBF kernels, and k-NN classifiers for various UPDRS threshold values. We tried various settings for *k*-NN and SVM classifiers. In Figures 1 and 2, the accuracies and MCC values that were obtained with the optimal parameter values of SVM-linear, SVM-RBF, and k-NN classifiers are shown.

As it is seen in Figure 1, the highest test set accuracy is achieved with k-NN (city-block distance, k=7) when the UPDRS threshold is specified around 10. Similarly SVM (both with linear and RBF kernel) achieved highest accuracy for the comparably lower values of UPDRS threshold (around 12). However, we should note that the binary classification problem obtained by setting UPDRS threshold to 10 is highly imbalanced (number of positive examples is 1868 whereas number of negative examples is 207), and using accuracy on such imbalanced datasets may lead to false inferences regarding the success of classifiers [14] [15]. For example, with these settings a simple strategy of labeling all the test set examples as positive class gives an accuracy of 90.02%. However, as it is seen in Figure 2, the MCC performed by *k*-NN (0.1854) with the same UPDRS threshold is comparably lower than that of SVM-linear and SVM-RBF classifiers. When Figures 1 and 2 are evaluated together, we see that the accuracy and MCC performances of SVM-linear and SVM-RBF classifiers change similarly with respect to the UPDRS threshold value. On the other hand, while accuracy of *k*-NN decreases as UPDRS threshold increases up to around 23 and then tends to increase again, MMC of *k*-NN increases as UPDRS threshold increases up to around 15 and thereafter shows a decreasing trend. These results show that SVM performs more consistently than *k*-NN on imbalanced datasets when the costs of errors made on the training samples of majority and minority class are tuned well.

It is seen in Figure 2 that SVM-RBF achieves higher MCC for all UPDRS threshold values than *k*-NN and SVM-linear classifiers. The highest MCC (0.4197) on test set is obtained with SVM-RBF when UPDRS threshold is set to 15. The corresponding test set accuracy of SVM-RBF is 75.86%. Figure 2 also shows that all the classifiers gave the highest MCC when the UPDRS threshold is set to 15.

IV. CONCLUSION

In this study, we aimed to determine the optimal UPDRS threshold value that can be discriminated by dysphonia measurements. For this purpose, we used Parkinson's Disease (PD) telemonitoring dataset consisting of speech recordings of 42 PD patients. We converted the UPDRS prediction problem into a binary classification problem for various motor UPDRS threshold values, and fed the features to *k*-NN and SVM classifiers to discriminate the PD patients whose UPDRS is less than or greater than the specified threshold value.

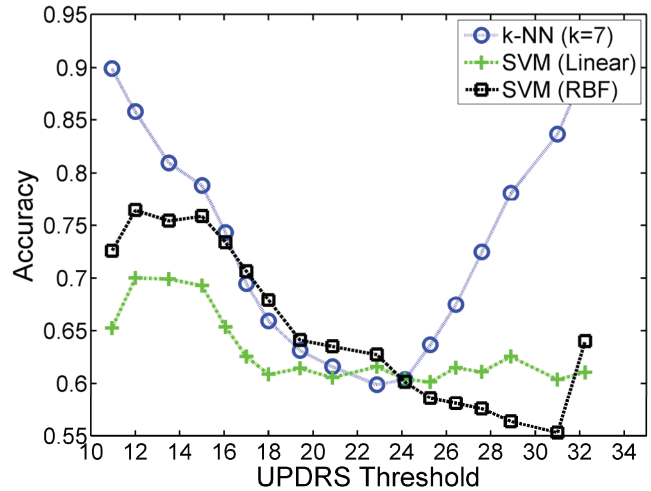

Fig. 1. Test set classification accuracies obtained with k-NN and SVM classifiers under various UPDRS threshold values

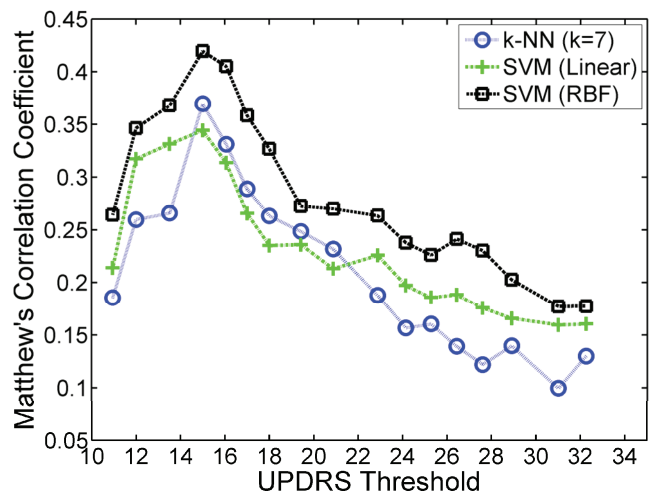

Fig. 2. Matthew's correlation coefficients obtained on test set with k-NN and SVM classifiers under various UPDRS threshold values

The experimental results show that both *k*-NN and SVM performed the highest MCC when UPDRS threshold is set to 15. These results indicate that speech disorders are more significantly seen in the patients whose UPDRS exceeds this threshold value. Besides, considering that the motor UPDRS ranges from 0 to 108, relatively low UPDRS threshold of 15 validates that vocal impairments can be used as early indicators of the disease. The highest MCC was obtained with SVM-RBF classifier, and SVM-RBF gave higher MCC values than *k*-NN and SVM-linear for all UPDRS threshold values. Besides, it has also been observed that the performances of *k*-NN and SVM-RBF classifiers converge to each other as the imbalance in the dataset decreases, and the difference between k-NN and SVM-RBF tends to increases in favor of SVM-RBF as the imbalance in the dataset increases. This shows that SVM-RBF is more successful than *k*-NN on imbalanced datasets especially when the cost of errors made on the majority and minority classes are tuned well.

REFERENCES

[1] M. C. de Rijk, L. J. Launer, K. Berger, M. M. Breteler, J. F. Dartigues, M. Baldereschi, L. Fratiglioni,A. Lobo, J. Martinez-Lage,C.Trenkwalder, and A. Hofman, “Prevalence of Parkinson’s disease in Europe: A collaborative study of population-based cohorts,” Neurology, vol. 54, pp. 21–23, 2000.W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.

[2] B. Harel, M. Cannizzaro, and P. J. Snyder, “Variability in fundamental frequency during speech in prodromal and incipient Parkinson’s disease: A longitudinal case study,” *Brain Cogn.*, vol. 56, pp. 24–29, 2004.

[3] S. Skodda, H. Rinsche, and U. Schlegel, “Progression of dysprosody in Parkinson’s disease over time—A longitudinal study,” *Movement Disord.*, vol. 24, no. 5, pp. 716–722, 2009.

[4] B. R. Bloem, J. P. Van Vugt, and D. J. Beckley. "Postural instability and falls in Parkinson's disease." *Advances in neurology,* vol. 87 pp. 209, 2001.

[5] C. O. Sakar and O. Kursun, "Telediagnosis of Parkinson’s disease using measurements of dysphonia," *Journal of medical systems,* vol. 34(4), pp. 591-599, 2010.

[6] M. A. Little., P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig. "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *Biomedical Engineering, IEEE Transactions on,* vol. 56(4), pp. 1015-1022, 2009.

[7] A. Tsanas, M. Little, P. E. McSharry, and L. O. Ramig. "Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests," *Biomedical Engineering, IEEE Transactions on,* 57(4), pp. 884-893, 2010.

[8] K. Polat, "Classification of Parkinson's disease using feature weighting method on the basis of fuzzy C-means clustering." *International Journal of Systems Science,* vol. 43(4), pp. 597-609, 2012.

[9] A. Unlu, R. Brause, and K. Krakow, "Handwriting analysis for diagnosis and prognosis of parkinson’s disease," In *Biological and Medical Data Analysis*, pp. 441-450. Springer Berlin Heidelberg, 2006.

[10] B. E. Sakar, and O. Kursun, "Telemonitoring of changes of unified Parkinson’s disease rating scale using severity of voice symptoms." in *Proc. 2nd International Conference on E-Health and TeleMedicine*, Istanbul, 2014, pp. 114-119.

[11] A. Ho, R. Iansek, C. Marigliani, J. Bradshaw, and S. Gates, ‘‘Speech impairment in a large sample of patients with Parkinson’s disease,’’ *Behav. Neurol.*, vol. 11, pp. 131–37, 1998.

[12] Lichman, M. (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

[13] C.W. Hsu and C.-J. Lin, "A Comparison of Methods for Multiclass Support Vector Machines," IEEE Transactions On Neural Networks, vol. 13, no. 2, pp. 415-425, 2002.

[14] R. Akbani, S. Kwek, and N. Japkowicz, "Applying support vector machines to imbalanced datasets," In *Machine Learning: ECML 2004*, pp. 39-50. Springer Berlin Heidelberg, 2004.

[15] N. V. Chawla, "Data mining for imbalanced datasets: An overview," In *Data mining and knowledge discovery handbook,* Springer US., *2005,* pp. 853-867.