A RULE-BASED CLASSIFICATION OF LARYNGOPATHIES BASED ON SPECTRUM DISTURBANCE ANALYSIS

An Exemplary Study

Krzysztof Pancerz¹, Wiesław Paja¹, Jarosław Szkola¹, Jan Warchol² and Grażyna Olchowik²

¹Institute of Biomedical Informatics, University of Information Technology and Management in Rzeszów, Rzeszów, Poland
²Department of Biophysics, Medical University of Lublin, Lublin, Poland

Keywords: Laryngopathy, Classification, Spectrum analysis.

Abstract: Our research concerns data derived from the examined patient’s speech signals for a non-invasive diagnosis of selected larynx diseases. The paper is devoted to the rule-based classification of patients on the basis of a family of coefficients reflecting spectrum disturbances around basic tones and their multiples. The paper presents a proposed procedure for feature selection and classification as well as the experiments carried out on real-life data.

1 INTRODUCTION

Our research concerns designing methods for classification of patients with selected larynx diseases using a non-invasive diagnosis. Two diseases are taken into consideration: Reinke’s edema (RE) and laryngeal polyp (LP). In general, the classification is based on selected parameters of a patient’s speech signal (phonation). Gathering parameters in this way is convenient for the patient because a measurement instrument (microphone) is located outside the voice organ. This enables free articulation. Moreover, different physiological and psychological patient factors impede making a diagnosis using direct methods.

There exist various approaches to analysis of biomedical signals (cf. (Semmlow, 2009)). In general, we can distinguish three groups of methods according to a domain of the signal analysis: analysis in a time domain, analysis in a frequency domain (spectrum analysis), analysis in a time-frequency domain (e.g., wavelet analysis). Therefore, in our research, we are going to build a specialized computer tool for supporting diagnosis of laryngopathies based on a hybrid approach. A decision support system will have a hierarchical structure based on multiple classifiers working on signals in time and frequency domains.

A series of papers published earlier has presented approaches designed for the computer tool being developed. Approaches presented in (Warchol et al., 2010) and (Pancerz et al., 2011) have been based on the speech spectrum analysis. However, in the approach presented in (Szkola et al., 2010b), (Szkola et al., 2010a), (Szkola et al., 2011a), and (Szkola et al., 2011b), we have detected all non-natural disturbances in articulation of selected phonemes using the modified Elman-Jordan networks that are recurrent neural networks. They can be used for pattern recognition in time series data due to their ability to memorize some information from the past. The approach initialized in (Pancerz et al., 2011) is extended in this paper.

2 CLASSIFICATION PROCESS

In the proposed approach, selected features (parameters) reflecting patient’s speech spectrum disturbances around a basic tone and its multiples (harmonics) are calculated. Clinical experience shows that harmonics in the speech spectrum of a healthy patient are distributed approximately steadily. However, larynx diseases may disturb this distribution (cf. (Warchol, 2006)). Therefore, the analysis of a degree of disturbances can support diagnosis of larynx diseases.

Disturbances are expressed by a family of coefficients computed for neighborhoods of a basic tone \( f_0 \) and its four multiples \( (f_1, f_2, f_3, f_4) \). In a real situation frequencies \( f_1, f_2, f_3, \) etc. are not distributed steadily (cf. (Warchol, 2006)). It means, that we need to find a real distribution of harmonics. In the presented approach, it is done on the basis of the resul-
can be expressed as a sequence of the following steps:

1. Normalizing signal sample values to the interval $[-1.0, 1.0]$, i.e., for each sample $s \in S$, a normalized sample $s_{\text{norm}}$ is equal to $s_{\text{norm}} = \frac{s}{\max_{i=1}^{N}|s_i|}$.

2. Dividing the speech signal $S$ into a family $W$ of $N$-point disjoint time windows, i.e., $W = [W_1, W_2, ..., W_N]$, where $W_1 = S[1, ..., N]$, $W_2 = S[N+1, ..., 2N]$, etc.

3. Calculating harmonics (first $f_1$, second $f_2$, third $f_3$, and fourth $f_4$) of a patient’s basic tone $f_0$, i.e., $f_1 = 2f_0$, $f_2 = 3f_0$, $f_3 = 4f_0$, and $f_4 = 5f_0$.

4. Calculating a resultant discrete spectrum $Sp$ for the family $W$ of time windows excluding some initial and final ones, based on the Discrete-Time Fourier Transform (DTFT), see e.g. (Semmlow, 2009), i.e., $X[k] = \frac{1}{N} \sum_{n=1}^{N} W[n] e^{-2\pi i k n N}$, where $k = 0, 1, ..., N-1$. The spectrum $Sp(W)$ is a vector of magnitudes $|X[k]|$, i.e., $Sp(W) = [|X[0]|, |X[1]|, ..., |X[2N]|]$. The resultant discrete spectrum $Sp$ is calculated as $Sp = \sum_{W} Sp(W)$.

5. Finding real multiples ($f_2^1, f_2^2, f_3^1$, and $f_4^1$) of the basic tone $f_0$. A real multiple $f_k^1$ is assumed to correspond to the maximum value of $Sp$ in a neighborhood of $f_k$, i.e., in the interval $[f_k - d_m, f_k + d_m]$, where $d_m$ is the input parameter and $k = 1, ..., 4$.

6. Calculating the regularity coefficient $R_k$, which can also be treated as some kind of shape parameter, for each frequency $f_k$ (the basic tone $f_0$ or its multiple $f_0^1, f_0^2, f_0^3$, or $f_4^1$). Two discrete integrals $I_1$ and $I_2$ of the spectrum are calculated, the first one for the frequency interval $[f_1 - d_2, f_1 + d_2]$, the second one for the frequency interval $[f_3 - d_3, f_3 + d_3]$, where $d_2$ and $d_3$ are input parameters and $d_2 < d_3$. The discrete integral $I$ of a fragment (between points $k_1$ and $k_2$) of the spectrum $Sp$ is equal to $I = \sum_{j=k_1}^{k_2} |X[j]|$. The ratio of $I_1$ to $I_2$ constitutes the regularity coefficient $R_k$, i.e., $R_k = \frac{I_1}{I_2}$. If both integrals are equal (ideal case), then the regularity coefficient $R_k = 1$. In a real situation, some fuzziness of a spectrum around a given harmonic frequency can be observed. It causes that the regularity coefficient $R_k < 1$, because $I_1 < I_2$. It is easy to see, that the greater the fuzziness of the spectrum, the smaller the regularity coefficient (the smaller the slenderness of the spectrum distribution).

7. Calculating the deviation coefficient $D_k$ for each multiple $f_k$ of the basic tone $f_0$. The deviation coefficient $D_k$ is the ratio $D_k = \frac{|f_k^1 - f_0|}{f_0}$, where $|x|$ denotes the absolute value of $x$.

After execution of the procedure consisting of steps presented above, we obtain a decision table which is an input for classification algorithms. A sample of the decision table is presented in Table 1.

In the decision table being a training set of cases for classifiers, we assign to each patient one of the two classes: norm - a norm - for the patient from the control group, i.e., without disturbances of phonation confirmed by a phoniatrist opinion, path - pathology - for the patient either with laryngeal polyp, or with Reinke’s edema (both clinically confirmed).

The approach presented in this paper has been tested using classification algorithms available in the popular data mining and machine learning software tools: WEKA (Witten and Frank, 2005), Rough Set Exploration System (RSES) (Bazan and Szczuka, 2005), NGTS (Hippe, 1997). Four of them are rule-based algorithms: exhaustive (RSES) (Bazan et al., 2000), LEM2 (RSES) (Grzymala-Busse, 1997), Genetic (RSES) (Wróblewski, 1998), and NGTS (Hippe, 1997). Two of them are decision-tree based algorithms: J48 (WEKA) - an implementation of C4.5 (Quinlan, 1993), CART (WEKA) (Breiman et al., 1993).

Values of features can be treated as continuous quantitative data. Building classification rules for such data can be difficult and/or highly inefficient. Therefore, for RSES generation algorithms, the so-called discretization was a necessary preprocessing.
Table 1: A sample of the decision table.

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>$R_0$</th>
<th>$R_1$</th>
<th>$R_2$</th>
<th>$R_3$</th>
<th>$R_4$</th>
<th>$D_1$</th>
<th>$D_2$</th>
<th>$D_3$</th>
<th>$D_4$</th>
<th>CLASS</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>0.90</td>
<td>0.89</td>
<td>0.89</td>
<td>0.82</td>
<td>0.75</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>norm</td>
</tr>
<tr>
<td>#2</td>
<td>0.85</td>
<td>0.82</td>
<td>0.75</td>
<td>0.59</td>
<td>0.58</td>
<td>0.04</td>
<td>0.03</td>
<td>0.08</td>
<td>0.08</td>
<td>path</td>
</tr>
</tbody>
</table>
| ...        | ...   | ...   | ...   | ...   | ...   | ...   | ...   | ...   | ...   | ...

step (Cios et al., 2007). Its overall goal was to reduce the number of values by grouping them into a number of intervals. Some discretization techniques based on rough sets and Boolean reasoning have been presented in (Bazan et al., 2000).

The sets of rules obtained using NGTS were improved using the RuleSEEKER system (Blajdo et al., 2004). The main optimizing process was based on an exhaustive application of a collection of generic operations (Paja and Hippe, 2005): finding and removing redundancy, finding and removing incorporative rules, merging rules, finding and removing unnecessary rules, finding and removing unnecessary conditions, creating missing rules, discovering hidden rules, rule specification, selecting final set of rules.

3 EXPERIMENTS

In the experiments, sound samples were analyzed. The experiments were carried out on two groups (Warchoł, 2006). The first group included persons without disturbances of phonation - the control group (CG). They were confirmed by a phoniatrist opinion. All persons were non-smoking, so they did not have contact with toxic substances which can have an influence on the physiological state of vocal folds. The second group included patients of Otolaryngology Clinic of the Medical University of Lublin in Poland. They had clinically confirmed dysphonia as a result of Reinke’s edema (RE) or laryngeal polyp (LP). Experiments were carried out by a course of breathing exercises with instruction about the way of articulation. The task of all examined patients was to utter separately different Polish vowels with extended articulation as long as possible, without intonation, and each on separate expiration. Each sound sample was recorded on MiniDisc MZ-R55 (Sony). Effectiveness of such an analysis was confirmed by Winholtz and Titze in 1998 (Winholtz and Titze, 1998).

For experiments, we have tested various combinations of input parameters. The best classification results have been obtained for the following parameters: $N = 8192$ - the number of points (samples) taken for DTFT, $d_m = 12$ - deviation for searching maximum, $d_2 = 4$, $d_3 = 8$ - deviations for calculating spectrum regularity coefficients.

To determine accuracy of generated rules by classification algorithms a cross-validation method was used. Cross-validation is frequently used as a method for evaluating classification models. It comprises of several training and testing runs. First, the data set is split into several, possibly equal in size, disjoint parts. Then, one of the parts is taken as a training set for rule generation and the remainder (sum of all other parts) becomes the test set for rule validation. In our experiments, a standard 10 cross-validation test was used (CV-10). Each classification algorithm was evaluated via testing the classification accuracy of unseen cases. Results of evaluation are collected in Table 2.

Table 2: Results of experiments: classification accuracy.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Classification accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exhaustive (RSES)</td>
<td>0.8430</td>
</tr>
<tr>
<td>LEM2 (RSES)</td>
<td>0.8700</td>
</tr>
<tr>
<td>Genetic (RSES)</td>
<td>0.8430</td>
</tr>
<tr>
<td>J48 (WEKA)</td>
<td>0.8441</td>
</tr>
<tr>
<td>CART (WEKA)</td>
<td>0.7922</td>
</tr>
<tr>
<td>NGTS</td>
<td>0.6786</td>
</tr>
<tr>
<td>NGTS (partial matching)</td>
<td>0.7786</td>
</tr>
<tr>
<td>NGTS (after optimization)</td>
<td>0.7214</td>
</tr>
<tr>
<td>NGTS (after optimization, partial matching)</td>
<td>0.8071</td>
</tr>
</tbody>
</table>

Almost all popular rule-based algorithms of machine learning and data mining manifest very similar classification accuracy. The obtained result can be treated as promising. In general, the considered problem is not simple. An important role is played by the quality of speech recording. The quality of results is also dependent on chosen preprocessing methods (e.g. filtration) and signal processing methods, e.g. Discrete-Time Fourier Transform (DTFT). An especially difficult task is the extraction of the correct signal.

4 CONCLUSIONS

On the basis of experiments described in the paper, we can notice that a family of coefficients, calculated on the basis of the analysis of spectrum shapes and deviations around a basic tone and its multiples for the
A RULE-BASED CLASSIFICATION OF LARYNGOPATHIES BASED ON SPECTRUM DISTURBANCE ANALYSIS
- An Exemplary Study

examined patient’s speech signal, may be helpful for classification of a patient. In general, the speech analysis of patients with larynx diseases is a difficult task. Therefore, the obtained results seem to be promising.

In the future, we plan to tune parameters reflecting regularity and deviations in the speech spectrum, especially by selecting proper regions of interest of the spectrum and we are going to add some new coefficients characterizing spectrum disturbances as well. The proposed approach based on analysis of speech signals in a frequency domain can be a part of a computer tool based on multicriteria decision making process. It should be treated as a supplementary element for other techniques. An important challenge is to design methods enabling distinction between different larynx diseases (for example, laryngeal polyp and Reinke’s edema). The approach presented in this paper does not enable us to make this distinction.

ACKNOWLEDGEMENTS

This research has been supported by the grant No N N516 423938 from the Polish Ministry of Science and Higher Education.

REFERENCES


