Computational Intelligence in a Classification of Audio Recordings of Nature

Krzysztof Tyburek, Piotr Prokopowicz and Piotr Kotlarz
Institute of Mechanics and Applied Computer Science Kazimierz Wielki University, Bydgoszcz, Poland

Keywords: Fuzzy System, Linguistic Modelling, MPEG-7, Audio Descriptors, Fuzzy Classification of Audio Signals, Neural Networks.

Abstract: This paper presents different ways for a classification of sounds of birds using linguistic approach with a fuzzy system, neural network and WEKA system. Features of sounds of birds species are coded by the selected MPEG-7 descriptors. The models of classification system are based on the audio descriptors for a some chosen species of birds like: Corn Crake, Hawk, Blackbird, Cuckoo, Lesser Whitethroat, Chiffchaff, Eurasian Pygmy Owl, Meadow Pipit, House Sparrow, Firecrest. The paper proposes fuzzy models that definitely bases on the linguistic description. Moreover neural network for classification was proposed. As reference results WEKA system is used.

1 INTRODUCTION

The ability to identify individual species of animals (including birds) just by the voice has a number of practical aspects. The most important of these is the automation of identification of species present in a given area on the basis of only audio recordings.

The researches proposed in this publication have a preliminary nature. They serve a preliminary analysis and preparation for further work on more complex topics related to the application of computational intelligence in description of audio signals.

The solutions of searching of multimedia data basing on the label technique do not always give expecting results. It means that sending queries are not always in accordance with demanding of person or computer system. Correct interpretation of sound source is the main issue which occurs during recognition process. This paper show two separate conceptions of the classification of sounds of birds, based on different computational intelligence methods: fuzzy system, artificial neural network. First one strictly bases on linguistic approach and second one is typical iterative procedure of adaptation. As reference values standard algorithms classifying provided by WEKA system will be used. The data for classification comes from 10 different kind of birds: Corn Crake, Hawk, Blackbird, Cuckoo, Lesser Whitethroat, Chiffchaff, Eurasian Pygmy Owl, Meadow Pipit, House Sparrow, Firecrest.

Researching of sound of bird can be useful for high level of recognizably each other. This problem can be solved by means of the MPEG 7 standard which gives a lot of descriptors for the physical features of sound. These descriptors are defined on the base of analysis of digital signals and index of most important their factors. The MPEG 7 Audio standard contains descriptors and description schemes that can be divided (Manjunath et al., 202;Martnez, 2002; Lindsay et al. 2002) into two classes: generic low-level tools and application-specific tools. The generic tools, referred to in the standard as the audio description framework apply to any audio signal and include the scalable series, low-level descriptors (LLDs) and the un form silence segment. The application specific tools restrict their application domain as a means to afford more descriptive power and include general sound recognition and indexing tools and description tools. The low-level audio descriptors have very general applicability in describing audio. There are seventeen temporal and spectral descriptors (Lindsay et al. 2002) that can be divided into six groups. A typical LLD may be instantiated either as a single value for a segment or a sampled series. Then two names for those descriptors are used, as the application requires: AudioLLDScalarType and AudioLLDVectorType, the first type is inherited for

Tyburek K., Prokopowicz P. and Kotlarz P..
Computational Intelligence in a Classification of Audio Recordings of Nature.
DOI: 10.5220/0005153801870192
In *Proceedings of the International Conference on Fuzzy Computation Theory and Applications* (FCTA-2014), pages 187-192
ISBN: 978-989-758-053-6

scalar values and describing a segment with a single summary, such as power or fundamental frequency, the second one is inherited for vector types describing a series of sampled valued, such spectra. This paper deals with LLDs as well as application specific tools to recognize audio signal coming from a group of 10 different kind of birds. In order to find a feature vector of the group of birds the analysis has been performed in the temporal as well as in frequency domains.

2 TIME DOMAIN AND FREQUENCY DESCRIPTION

For the purpose of right describing of waveform of sound it is necessary to define descriptor. The descriptor is represented as a fraction of time of separating phases to time of all phases.

1. Log - time of the ending transient l_{tk} , which is given by:

$$ltk = \log(tpk - tmax) \quad (1)$$

where: tmax is the time at which the maximal amplitude has been reached, tpk is the time at which the level of 10 % of maximal value has been reached in the decay stage.

2. The Log-Attack-Time Descriptor characterizes the “attack” of a sound, the time it takes for the signal to rise from silence to the maximum amplitude. This feature signifies the difference between a sudden and a smooth sound. The Log-Attack-Time is given by:

$$ltp = \log(tmax - tpp) \quad (2)$$

where: tmax is the time at which the maximal amplitude has been reached, tpp is the begin of rise of signal.

3. ZC – (zero crossing) which describes the number of crossing the X axis in the analyzed window in the time domain. In these experiments the length of window was n = 1000 samples. The onset of the window was in the begin of rise of signal. The length of the window was sufficiently long for all samples in experiments.

Since the frequency domain may contain important information concerning features of the sound it is worthwhile to introduce its parametrization. The base of parametrization of sound spectrum are Fourier transform, wavelet analysis, cepstrum or Wigner-Ville'a transform. The following parameters describing frequency domain of signal were applied:

4. Brightness

$$Br = \frac{\sum_{i=0}^{n} A(i) \cdot i}{\sum_{i=0}^{n} A(i)} \quad (3)$$

where: A(i) is amplitude of the i-th partial (harmonic) i - the frequency of the i-th partial

5. Irregularity of spectrum

$$Ir = \log\left(20 \sum_{i=2}^{N-1} \left| \log \frac{A(i)}{\sqrt[3]{A(i-1) \cdot A(i) \cdot A(i+1)}} \right| \right) \quad (4)$$

where: A(i) is amplitude of the i-th partial (harmonic), N - number of available harmonic

3 RESEARCH METHODOLOGY

The purpose of experiment is searching of vector of features which allow to automatic classification of sound of bird. For parametrization of frequency domain state window length was proposed (Tyburek, 2006; Tyburek et al.2008). It was applied for all samples in experiment. State window length is the fragment of signal (in time domain) which was taken in the same point of time. State window length contains constant amount of samples. The beginning of this window was taken when the level of 10 % of maximal value has been reached. The length of window is determined by resolution of spectrum, according to the formula:

$$f_r = \frac{f_s}{n} \quad (5)$$

where: fr is the spectrum resolution, fs - sampling frequency (44100 Hz), n - number of samples.

In this paper fr equal to 10Hz was assumed. It means that number of samples which are assigned to experiment is equal 4410. If testing sound is shorter then length of window (n=4410) then absent values should be supplemented with zeros to n=4410 (Tyburek, 2006; Tyburek et al.2008). Selecting fragment of signals in time domain were treated DFT and this spectrum was analyzed. Moreover for this study, Blackman window has been used.

4 RESULTS FROM WEKA

To properly conduct an experiment we need reference values. Well known and popular system WEKA can be used at this point. For the further study and compare the cross-validation method and algorithms: k-Nearest Neighbors, Random forest, Jrip were chosen. The tables 1 to 3 shows a Weka results for mentioned methods.

Table 1: Error matrix for classification of sound of 10 kind of bird. Used k-NN, cross-validation method (k=10). General recognition 95,45%.

a	b	c	d	e	f	g	h	i	j	Classified
93	0	0	0	0	0	0	0	7,3	0	a = Corn Crake
0	98	1,8	0	0	0	0	0	0	0	b = Hawk
0	1,8	85	3,6	0	0	7,3	0	1,8	0	c = Blackbird
0	0	0	100	0	0	0	0	0	0	d = Cuckoo
0	0	0	0	100	0	0	0	0	0	e = Lesser Whitethroat
0	0	0	0	0	100	0	0	0	0	f = Chiffchaff
0	0	1,8	0	0	0	98	0	0	0	g = Eurasian Pygmy Owl
0	0	0	0	0	1,8	0	98	0	0	h = Meadow Pipit
18	0	0	0	0	0	0	0	82	0	i = House Sparrow
0	0	0	0	0	0	0	0	0	100	j =Firecrest

Table 2: Error matrix for classification of sound of 10 kind of bird. Used Random forest, cross-validation method (k=10). General recognition 97,1%.

a	b	c	d	e	f	g	h	i	j	Classified
93	0	0	0	0	0	0	0	7,3	0	a = Corn Crake
0	98	1,8	0	0	0	0	0	0	0	b = Hawk
0	1,8	95	0	0	0	1,8	0	1,8	0	c = Blackbird
0	0	1,8	96	0	0	1,8	0	0	0	d = Cuckoo
0	0	0	0	100	0	0	0	0	0	e = Lesser Whitethroat
0	0	0	0	0	98	0	0	1,8	0	f = Chiffchaff
0	0	0	0	0	0	100	0	0	0	g = Eurasian Pygmy Owl
0	0	0	0	0	0	0	98	0	1,8	h = Meadow Pipit
5,5	0	0	0	0	0	0	0	95	0	i = House Sparrow
0	0	0	0	0	0	0	1,8	0	98	j =Firecrest

5 LINGUISTIC MODELLING OF FUZZY CLASSIFIER

Although the research described in this publication relate to other descriptors and other birds species, the idea of a classifier is based on the concept presented in (Tyburek et al.,2014) as proposition 2.

Table 3: Error matrix for classification of sound of 10 kind of bird. Used JRip, cross-validation method (k=10). General recognition 90,2%.

a	b	c	d	e	f	g	h	i	j	Classified
82	0	0	3,6	0	1,8	0	0	13	0	a = Corn Crake
0	89	3,6	0	1,8	3,6	0	0	0	1,8	b = Hawk
0	7,3	85	0	1,8	1,8	0	1,8	1,8	0	c = Blackbird
0	0	1,8	96	0	0	1,8	0	0	0	d = Cuckoo
0	0	0	0	93	3,6	0	0	3,6	0	e = Lesser Whitethroat
5,5	0	0	0	0	95	0	0	0	0	f = Chiffchaff
0	1,8	1,8	1,8	0	0	93	0	0	1,8	g = Eurasian Pygmy Owl
0	0	0	0	1,8	1,8	0	96	0	0	h = Meadow Pipit
7,3	0	3,6	0	0	3,6	0	0	84	1,8	i = House Sparrow
0	0	0	0	1,8	0	0	9,1	0	89	j =Firecrest

5.1 Defining Linguistic Variables

We define fuzzy sets for input variable as characteristic values for each species. Similar solution to another problem - the classification of flowers (irises) - was presented in (Siler and Buckley, 2005) . In this way, we have a linguistic values which clearly indicates the considered species i.e.: “Hawk”, “Blackbird”, “Cuckoo”, etc. Each of them is a triangular fuzzy set (see LR fuzzy sets notation in (Dubois and Prade, 1980) and is determined with the use of the available data.

The method for construction an input variable fuzzy set indicating a given specie is presented by formula:

$$Bird_i = \Lambda\left(x; x_{mean} - 2\Delta_L; x_{mean} + 2\Delta\right)$$
$$\Delta_L = x_{mean} - x_{min} \quad \Delta_L = x_{max} - x_{mean} \tag{6}$$

where i – number of bird species, xmin,xmax,xmean- the minimum/maximum/mean value of the descriptor for the given species.

In this way we construct all the fuzzy sets for the every input linguistic variables, which represents the audio descriptors. It may be surprising, but we ignore the properties usually expected from the fuzzy model such as the completeness or the continuity (see Driankov et al., 1996). We also do not expect that the values of the fuzzy membership functions sum to unity within a linguistic variable. The outputs are simplified to the two valued linguistic variables which recognize or reject given species..

5.2 Rule Base

For such linguistic variables the rule base is very intuitive, and is composed of the rules, each of which is intended to recognize one particular species. Model for the defining rules can be expressed linguistically as follows:

$$\begin{gathered} IFi_1 isB_1 \wedge i_2 isB_1 \wedge ... i_i isB_1 \\ THENB_1 isrec \wedge B_2 isrej \wedge ... B_i isrej \end{gathered} \tag{7}$$

ik - input number k, Bl - bird (species) number i rec – fuzzy set for the recognition and rej – for the rejection.

Such modeling of fuzzy systems has one basic advantage, related above all, to the simplicity of expansion defined classifier. Adding the next species of birds does not violate existing structure. It needs just adding the next fuzzy sets characteristic for the new species, and add a rule that recognizes a new class of data. Similarly, if we want to introduce another descriptor to the model, the changes also will be simply and intuitive.

5.3 Testing the Linguistic Classification

Testing of efficiency presented proposition in general will be realized as follows:
- divide all data for two sets: Training and Testing.
- creating fuzzy classifier basing on the Training data set,
- applying created fuzzy classifier for the Testing set, and calculating wrong recognitions.

Such procedure should be done repeatedly with random dividing data.

5.4 Fuzzy Numbers in Signal Description

Choosing descriptors of sound signal play a key role in a classification proposed in this work. The authors plan in future work with descriptors based on computational intelligence, especially fuzzy numbers.

Usually MPEG-7 descriptors are mathematical formulas operating on the exact parameters of the signals. One of the directions of future work is the use of fuzzy numbers represented by model Ordered Fuzzy Numbers (OFN) to define a signal descriptors.

OFNs (Kosiński et al., 2003, Kosiński and Prokopowicz, 2007, Kosiński et al., 2013, Prokopowicz and Malek, 2014) are a quite recent proposition for modelling the calculations on imprecise values in similar way as the fuzzy numbers. Most important property of OFNs is that they allow for calculations as flexibly as real numbers (Prokopowicz, 2013), so they can be applied in different mathematical formulas, as for example MPEG-7 descriptors.

6 NEURAL NETWORK FOR CLASSIFICATION

The second method of classification we are intending to employ in the planned research is utilization of a neural network classifier. It is meant as a support for the results obtained via the fuzzy logic with the purpose of acquiring possibly good classification results. The first preliminary step of the research, which has already been carried out, is determining whether the existing set of bird sound samples actually comprises separate classes (species). At the outset, a single-layer network was examined. The network consists of the same neurons as the model shown below:

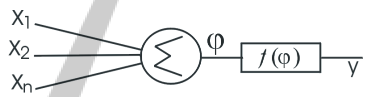

Figure 1: Used in this paper a model neuron.

Model of this neural network is presented by formula:

$$\varphi = \sum_{i=1}^{n} x_i \cdot w_i = x \cdot w^T \quad , \quad f = \{ \begin{matrix} 0{,}5 \varphi > p \\ -0{,}5 \varphi < p \end{matrix} \tag{8}$$

where: n - index of input neuron (n=5), ϕ - membrane potential, w- matrix of weights

$$w = [w_1, w_2, \ldots w_i] \quad , \quad x = [x_1, x_2, \ldots x_i] \quad ,$$

where: x - matrix inputs
f- active function, p - threshold (p=0).

The choice of the p-value as well as the neural response via the function f was selected by numerical experiments, which are associated with the selection of proper parameters of the training process. The process of neural network training employed a continuous activation function; a threshold function is applied in the process of testing. The delta rule in the version for the linear activation function was used in the training process. The general form delta rule for permanent active function ($f(\varphi)$) (see Hannagan T., 2013) :

$$r = \sigma \stackrel{def}{=} [d_i - f(w_i^t x)] f'(w_i^t x) \tag{9}$$

Adjustment of neuron weights:

$$\sigma w_t = c(d_t - y_t) f'(net_t) x_t \quad (10)$$

where: r - signal learner, σw - correction weights c - constant learning process, d - error learning, y - matrix outputs, active function, x - matrix inputs

The target of mentioned research is to confirm that learning set is composed of signal, which are separate class and to confirm that mentioned classes are connected with different bird species. The learning session were in three main phases: learning set was testing set in the same time, the learning set and testing set were composed from twenty five different signals. The third case – learning set fifteen and testing set thirty five signals. Below presented result of tests:

Table 4: The learning and testing set used the same fifty signals.

The amount of class	The best result of learning (figures of proper recognized)	Testing
5(Eurasian Pygmy Owl, Meadow Pipit, Hawk, Cuckoo, Lesser Whitethroat)	a) 88% b) 78% c) 73%	a) 88% b) 71% c) 71%
6(Eurasian Pygmy Owl, Meadow Pipit, Hawk, Cuckoo, Lesser Whitethroat, Chiffchaff)	a) 82% b) 74% c) 69%	a) 82% b) 73% c) 65%
7(Eurasian Pygmy Owl, Meadow Pipit, Hawk, Cuckoo, Lesser Whitethroat, Chiffchaff, House Sparrow)	a) 55% b) 55% c) 50%	a) 55% b) 53% c) 48%
8(Eurasian Pygmy Owl, Meadow Pipit, Hawk, Cuckoo, Lesser Whitethroat, Corn Crake, Blackbird, Chiffchaff)	a) 43% b) 41% c) 35%	a) 43% b) 34% c) 29%

a - The learning and testing set used the same fifty signals, b - The learning and testing set using 25 signals, c - The 15 learning set tested 35 different signals

As a result of the research, it is concluded that the examined set of signals contains separate classes. Although the classification results obtained are not very good, they provide a basis for further research work. In the essential part of the planned research, the possibilities of classification with the usage of different neural network models will be tested. The results obtained at the preliminary stage confirm the legitimacy of the adopted concept of research. The results warrant further research work, whose purpose is to develop a hybrid tool based on the methods of computational intelligence for the classification of audio signals using low-level descriptors.

7 CONCLUSIONS

The proposals presented in this work are a preliminary step for further research related to the application of computational intelligence in the analysis of the audio signal described by MPEG-7. Choosing descriptors of sound signal play a key role in a classification, so descriptors based on computational intelligence are also important part of future research.

REFERENCES

Driankov D, Hellendoorn H., Reinfrank M., (1996),An Introduction to fuzzy control, Springer-Verlag, Berlin, Heidelberg, 1996.

Dubois, D., Prade, H. M. (1980), Fuzzy sets and systems: Theory and applications, New York: Academic Press, 1980.

Hannagan T., 2013 The delta rule does Bubbles, Journal of Vision 13(8):17, 1–11.

Kosiński W., Prokopowicz P., Ślęzak D., 2003. Ordered fuzzy numbers. Bulletin of the Polish Academy of Sciences, Ser. Sci. Math., 51(3): 327–338.

Kosiński W., Prokopowicz P., 2007. Fuzziness - Representation of Dynamic Changes, Using Ordered Fuzzy Numbers Arithmetic, New Dimensions in Fuzzy Logic and Related Technologies. In: Martin Stepnicka, Vilem Nova, Ulrich Bodenhofer (eds.) Proc. of the 5th EUSFLAT Conference, vol I, Ostrava, Czech Republic, September 11-14, 2007, pp. 449-456.

Kosiński W., Prokopowicz P., Rosa A., 2013,. Defuzzification Functionals of Ordered Fuzzy Numbers. IEEE Transactions on Fuzzy Systems, 21(6): 1163-1169.

Lindsay, A. T., Burnett, I., Quackenbush, S., Jackson, M. (April 2002): Fundamentals of audio descriptions. In: Manjunath, B.S., Salembier, P., Sikora, T. (eds.) Introduction to MPEG-7: Multimedia Content Description Interface, pp. 283 298. John Wiley and Sons, Ltd.

Martnez, J. M. (July 2002): MPEG-7 Overview, Klangenfurt Descriptors, In: Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) Proc. of ICAISC 2014, Part II. LNCS (LNAI), vol. 8468, pp. 700-709.

Pathak K. K., Panthi S., and Ramakrishnan N., April 2005 Application of Neural Network in Sheet Metal Bending Process , Defence Science Journal, Vol. 55, No. 2,, pp. 125-131.

Prokopowicz P., 2013. Flexible and Simple Methods of Calculations on Fuzzy Numbers with the Ordered

Fuzzy Numbers Model. In: Rutkowski L., Korytkowski, M., Scherer R., Tadeusiewicz R., Zadeh L. A., Zurada J. M. (eds.) Proc. of ICAISC 2013, Part I. LNCS (LNAI), vol. 7894, pp. 365–375. Heidelberg: Springer.

Prokopowicz P., Malek S., 2014. Aggregation Operator for Ordered Fuzzy Numbers Concerning the Direction, In: Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) Proc. of ICAISC 2014, Part I. LNCS (LNAI), vol. 8467, pp. 267-278.

Siler W., Buckley J.J., (2005), Fuzzy Expert Systems and Fuzzy Reasoning, Wiley, 2005.

Tyburek, K., Cudny, W., Kosiński, W. 2006: Pizzicato sound analysis of selected instruments In the freguency domain. Image Processing & Communications 11(1), 53–57.

Tyburek, K. (November 2008): Classification of string instruments in multimedia database especially for pizzicato articulation, Ph. D. thesis. Institute of Fundamental Technological Research Polish Academy of Sciences, Warsaw (in Polish).

Tyburek K., Prokopowicz P., Kotlarz P.,The Fuzzy Classification of Sounds Based on the Audio. Artificial Intelligence and Soft Computing Lecture Notes in Computer Science Volume 8468, 2014, pp 700-709.