International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:9, No:4, 2015

Lipschitz Classifiers Ensembles: Usage for Classification of Target Events in C-OTDR Monitoring Systems

Andrey V. Timofeev

***Abstract*—**This paper introduces an original method for guaranteed estimation of the accuracy for an ensemble of Lipschitz classifiers. The solution was obtained as a finite closed set of alternative hypotheses, which contains an object of classification with probability of not less than the specified value. Thus, the classification is represented by a set of hypothetical classes. In this case, the smaller the cardinality of the discrete set of hypothetical classes is, the higher is the classification accuracy. Experiments have shown that if cardinality of the classifiers ensemble is increased then the cardinality of this set of hypothetical classes is reduced. The problem of the guaranteed estimation of the accuracy for an ensemble of Lipschitz classifiers is relevant in multichannel classification of target events in C-OTDR monitoring systems. Results of suggested approach practical usage to accuracy control in C-OTDR monitoring systems are present.

***Keywords*—**Lipschitz classifiers, confidence set, C-OTDR monitoring, classifiers accuracy, classifiers ensemble.

I. Introduction

ONE of the main problems related to stochastic samples classification is to estimate the degree of similarity between a sample and those classes that are located nearby in the metric of the feature space. A question arises: what is the formal mechanism for the selection of classes, which corresponds to a sample with a priori specified lower bound of the classification accuracy value? In [1] a comparatively simple approach was suggested to solve this problem. In this approach, the classification solution was obtained as a finite closed set of alternative hypotheses, which contains an object of classification with probability no less than the specified value. The contents of the presented paper are a generalization of the mentioned approach to the case of ensemble of Lipschitz classifiers. The ensemble classifiers are commonly used to stabilize [2] and to increase the efficiency [3]. Simply speaking, the output of the classifiers ensemble represents a certain combination of outputs of the multiple classifiers, which were included in the ensemble. Using the ensemble of classifiers is a promising approach in the number of practical areas. For example, the approach of ensemble classifiers is extremely important in practical multimodal biometrics as well as in multichannel monitoring systems. This paper introduces an original method for guaranteed estimating the accuracy of the Lipschitz classifiers ensemble for a large number of classes. It should be noted that the so-called Lipschitz classifiers [2] correspond to a very wide type of classifiers. This type includes such well-known and practically effective classifiers as SVM (Support Vector Machine), Linear Programming Machines and even NN (Nearest Neighbor) method. A variety of the Lipschitz classifiers (LC) and their broad application mainly caused the major focus on this type of classifiers within the frames of this paper. The presented method to estimate the classification accuracy can be applied to any type of Lipschitz classifiers. The mechanisms of Lipschitz classifiers construction and learning, however, are out of the scope of this research. Thus, within this article, Lipschitz classifiers are considered to be given, research on their accuracy only is being carried out.

The approach suggested in this paper allows estimation of the accuracy of the Lipschitz classifiers' ensemble by determining the closed confidence subset within a priori given set of classes. This set, having a specified confidence coefficient $P_c, P_c \in (0,1)$, contains an index of classification object. The classification accuracy is determined by the number of alternative hypotheses included in this confidence set (**target set**): the smaller the cardinality of the confidence set is, the higher the classification accuracy is. The suggested approach provides the guaranteed accuracy of estimation: the constructed confidence set will contain a true object of classification with the previously specified confidence coefficient. The paper describes a practical case that demonstrates high efficiency of the proposed approach

Andrey V. Timofeev is with the LPP "EqualiZoom", Astana, 010000, Kazakhstan (phone: +7-911-191-42-67; e-mail: timofeev.andrey@gmail.com).

II. The Research Objective

Let us denote:

- $\mathbf{Z} = \ Z_k, d_k \ _k \,\big|\, k \in \ 1,\dots,m$ - set of the compact feature spaces;
- $Z(k) = \ Z_k, d_k \ _k$ - k-th compact feature space, d_k – metric of this space, Z_k – set of feature values;
- Θ is a set of indexes of classes;
- $z_k \ \theta$ - feature of the θ class, $z_k \ \theta \ \in Z_k, k \in \ 1,\dots,m \ , \theta \in \Theta$;
- $d_k \ \theta_2, \theta_1 \ \equiv d_k \ z_k \ \theta_2 \ , z_k \ \theta_1$, $\theta_2 \neq \theta_1 \in \Theta : \forall_k d_k \ z_k \ \theta_2 \ , z_k \ \theta_1 \ > 0$;
- $\mathbf{z}_k^{(\theta^*)} \in Z_k$ - sample, which need to classify; $Z^{(\theta^*)} = \ \mathbf{z}_1^{(\theta^*)}, \mathbf{z}_2^{(\theta^*)}, \dots, \mathbf{z}_m^{(\theta^*)}$;
- $\theta^* \in \Theta$ is a true index of the class to which the sample $\mathbf{z}_k^{(\theta^*)}, k \in \ 1,\dots,m$, belongs; in another words, θ^* is an index

of a target class; $f_k : Z_k \to \Theta$, $k \in 1,...,m$, are Lipschitz classifiers, which divides the spaces $Z(k)$ into $m = |\Theta|$ classes; $f_k(\theta \mid \theta^*) = f_k\ \theta \mid \mathbf{z}_k^{(\theta^*)}\ , R_k$ here $f_k\ \theta \mid \mathbf{z}_k^{(\theta^*)} \in R^1$, $k \in 1,...,m$, are discriminate functions (so-called score-parameters), which shows degree of similarity of a sample $\mathbf{z}_k^{(\theta^*)}$ regarding to class $\theta \in \Theta$; for every $k \in 1,...,m$ function $f_k\ \theta \mid \mathbf{z}_k^{(\theta^*)}$ represent a stochastic function, which explicitly dependent on the index hypothesis to be tested θ and implicitly on the index of the target class θ^*;

- R_k is the classification decision-making rule, $R_k : \tilde{\theta}_k = Arg \underset{\theta \in \Theta}{Max}\ f_k\ \theta \mid \mathbf{z}_k^{(\theta^*)}$;
- $\mathbf{F}(\mathbf{Z} \mid \theta^*) = f_k(\theta \mid \theta^*) \mid k = 1,...,m$ - ensemble of the Lipschitz classifiers formed at the value of the target class $\theta^* \in \Theta$; $f_k^{(N)}\ \theta \mid \mathbf{z}_k^{(\theta^*)} = Norm\ f_k\ \theta \mid \mathbf{z}_k^{(\theta^*)}$ is normalization of the function $f_k\ \theta \mid \mathbf{z}_k^{(\theta^*)}$, which provides $f_k^{(N)}\ \theta \mid \mathbf{z}_k^{(\theta^*)} \in (0,1)$;
- $F\ \theta \big| \mathbf{F}(\mathbf{Z} \mid \theta^*)$ is discriminate function on an ensemble of classifiers $\mathbf{F}(\mathbf{Z} \mid \theta^*)$, $F\ \theta \big| \mathbf{F}(\mathbf{Z} \mid \theta^*) = \sum_k \alpha_k f_k^{(N)}\ \theta \mid \mathbf{z}_k^{(\theta^*)}$, where $\sum_k \alpha_k = 1, \forall_k \alpha_k \geq 0$; $\mathsf{F} : \bigcup_k Z_k \circledR\ \mathrm{Q}$ is integral classifier on an ensemble $\mathbf{F}(\mathbf{Z} \mid \theta^*)$;
- $\mathsf{F} = F\ \theta \big| \mathbf{F}(\mathbf{Z} \mid \theta^*)\ , \mathsf{R}$;
- $\chi\ \mathbf{z}_k^{(\theta^*)} \in \Theta$, $k \in 1,...,m$, is the class to which actually the sample $\mathbf{z}_k^{(\theta^*)}$ belongs; it is obvious that $\theta^* = \chi\ \mathbf{z}_k^{(\theta^*)}$;
- $|X|$ is a cardinality of set X;
- $Diam\ Z_k = \sup_{z_1, z_2 \in Z} \|z_1 - z_2\|$, $k \in 1,...,m$.

Thus, we have an ensemble of Lipschitz classifiers $\mathbf{F}(\mathbf{Z} \mid \theta^*)$. Every classifier $f_k(\theta \mid \theta^*) \in \mathbf{F}(\mathbf{Z} \mid \theta^*)$ is defined on the appropriate feature space $Z(k)$, $k \in 1,...,m$. On the ensemble $\mathbf{F}(\mathbf{Z} \mid \theta^*)$ was formed the integral classifier F, which has the output in the form of $\tilde{\theta} = Arg \underset{\theta \in \Theta}{Max}\ F\ \theta \big| \mathbf{F}(\mathbf{Z} \mid \theta^*)$.

The goal of this paper is, while observing samples $Z^{(\theta^*)}$, to determine for classifier $\mathsf{F} = F\ \theta \big| \mathbf{F}(\mathbf{Z} \mid \theta^*)\ , \mathsf{R}$ such a confidence set of indices $\Xi(\tilde{\theta})$ (target set), $\Xi(\tilde{\theta}) \subseteq \Theta$, for which the following statement is true: $\mathbf{P}\ \theta^* \in \Xi(\tilde{\theta}) \geq P_c$. The confidence coefficient P_c is a priori specified. Thus the desired class θ^* will belong to the target set $\Xi(\tilde{\theta})$ with a probability at least equal to the previously prescribed value $P_c, P_c \in (0,1)$. In this case, the classification results are yielded not as one class, but as a set of classes that together constitute the target set.

III. Solution Method

Since classifiers $f_k(\theta^*) \mid k = 1,...,m$ have a stochastic nature, their outputs are random variables. The output of any classifier $f_k(\theta^*)$ can be described by the following statement: $f_k\ \theta \mid \mathbf{z}_k^{(\theta^*)} = \mathbf{E}_{\theta^*} f_k\ \theta \mid \mathbf{z}_k^{(\theta^*)} + \eta_k\ \theta \mid \mathbf{z}_k^{(\theta^*)}$. Here, $\theta^* \in \Theta$ is index of the target object class; $\theta \in \Theta$ is index of the testing hypothesis class; $\mathbf{E}_{\theta^*} f_k\ \theta \mid \mathbf{z}_k^{(\theta^*)}$ is the expected value of a random function $f_k\ \theta \mid \mathbf{z}_k^{(\theta^*)}$ with specified parameters $\theta^*, \theta \in \Theta$; $\eta_k\ \theta \mid \mathbf{z}_k^{(\theta^*)}$ is a random function of the parameters $\theta^*, \theta \in \Theta$. Let us, for the specified parameter $c \in (0, \infty)$, consider the following auxiliary set:

$$\Xi\ \tilde{\theta} \mid F, \mathbf{Z}, c\ = \ \theta \in \Theta \mid \Big| F\ \tilde{\theta} \big| \mathbf{F}(\mathbf{Z} \mid \theta^*)\ - F\ \theta \big| \mathbf{F}(\mathbf{Z} \mid \theta^*)\ \Big| \leq c\ .$$

The main result of this paper is the following theorem:

Theorem 1. Let the following conditions be true:

1. The random values $\eta_k\ \theta \mid \mathbf{z}_k^{(\theta^*)}\ \big| k \in 1,...,m$ are independent. $\forall k; \theta, \theta^* \in \Theta : \mathbf{E}_{\theta^*}\ \eta_k\ \theta \mid \mathbf{z}_k^{(\theta^*)} = 0$;
2. $\underset{k}{\forall} \exists a_k, b_k \in R^1 :\ f_k\ \theta \mid \mathbf{z}_k^{(\theta^*)} \in [a_k, b_k] \subseteq R^1, |a_k - b_k| < \infty$, $f_k^{(N)}\ \theta \mid \mathbf{z}_k^{(\theta^*)} = f_k\ \theta \mid \mathbf{z}_k^{(\theta^*)} - b_k\ |a_k - b_k|^{-1}$;
3. $\forall k; \theta^* \in \Theta :\ \underset{\theta \in \Theta}{Max}\ \mathbf{E}_{\theta^*} f_k\ \theta \mid \mathbf{z}_k^{(\theta^*)} = \mathbf{E}_{\theta^*} f_k\ \theta^* \mid \mathbf{z}_k^{(\theta^*)}$, $\Big|\ \theta^\# \in \Theta \big| \mathbf{E}_{\theta^*} f_k\ \theta^\# \mid \mathbf{z}_k^{(\theta^*)} = \mathbf{E}_{\theta^*} f_k\ \theta^* \mid \mathbf{z}_k^{(\theta^*)}\ \Big| = 1$;
4. $\forall k; \theta_1, \theta_2, \theta^* \in \Theta \exists L_k \in]0, \infty[$: $\Big\| f_k\ \theta_1 \mid \mathbf{z}_k^{(\theta^*)} - f_k\ \theta_2 \mid \mathbf{z}_k^{(\theta^*)}\ \Big\| \leq L_k d_k\ \theta_1, \theta_2$ a.s.;
5. $c(P_c) = \left(\sum_k \alpha_k^2\ L_k^{(N)} Diam\ Z_k\ ^2 \right)^{0.5} 1 - P_c\ ^{-0.5}$, where $P_c \in]0,1[$ and $L_k^{(N)} = L_k\ |a_k - b_k|^{-1}$.

Then we have

$$\forall \theta^* \in \Theta : \mathbf{P}_{\theta^*}\ \theta^* \in \Xi\ \tilde{\theta} \mid F, \mathbf{Z}, c(P_c)\ \geq P_c\ .$$

The proof is similar to the proof of Theorem 1 in [1].

Remark: it can be shown that the decrease in the value of $\big| \Xi\ \tilde{\theta} \mid F, \mathbf{Z}, c(P_c)\ \big|$ will contribute to decreasing of the value of the $c(P_c)$, with constant value P_c. It causes decrease of the value of $\mathbf{E}_{q^*} |\Xi(\cdot)|$ hence accuracy of the classification procedure is increased (because, we have a decrease in the quantity of the alternative hypotheses which were contained in the target set.

IV. Practical Example: Multichannel C-OTDR Monitoring Systems

Multichannel C-OTDR monitoring systems are a good example of the usefulness of the proposed approach, in particular, when applied to the task of target event type classification. Let us describe the common principle of C-OTDR monitoring systems operate. We will call the systems of this class as optical fiber classifier of seismoacoustic pulses (OXY). An OXY-approach based on the use of the high vibrosensitivity of the infrared energy stream injected into ordinary optical fiber (buried in the ground near the monitoring object) by means of semiconductor laser of low power. This optical fiber will be called a fiber optic sensor (FOS). Typically FOS length is 40-50 km. In the systems of this class, all relevant information is transferred to Processing Center (PC) by the optical fiber which is not only a sensor (FOS) but at the same time an effective and reliable channel for ordinary data transmission. The basis of the described method underlying OXY is the use of the vibrosensitive infrared stream injected into a standard monomode fiber (FOS) by means of a coherent semiconductor laser at the wavelength of 1550 nm. Thus, the laser probes the FOS with usage of infrared stream. This probing is carried out in the pulsed mode. Pulses have a length of ~ 50-200 ns, with an interval of ~ 50-300 μs. The optical fiber is put into the ground, at the depth of 30-50 cm, at the distance of 5-10 m from the monitoring object and, as a matter of fact, it is an optical fiber sensor. When a pulse is moving along the optical fiber, the Rayleigh elastic backscattering is realized on its natural irregularities (impurities), which due to high coherence of the used laser of 3B class leads to formation of the so-called stable interference structures of chaotic type, otherwise called speckles or speckle images. A sequence of speckles is received in the point of emanation using an ordinary welded coupler or a circulator.

The central moment of the concept is the phenomenon that any seismic vibration arising on the surface of the optical fiber due to propagation of seismoacoustic waves from the sources of elastic oscillations, changes its local refractive index. Changes of the local refractive index are reflected in the time-and-frequency structure (TFS) of the respective speckle. Knowing the pulse duration and the velocity of wave propagation in the optical fiber, it is easy to determine the section where the TFS speckle deviation took place. Analysis of the sequence of speckle structures using wavelet conversion apparatuses (the phase of singling out of primary signs of target signals) and Lipschitz classifiers (the phase of classification of target signals) makes it possible not only to reliably detect the target source of seismoacoustic radiation, but also to determine its type and area of occurrence. In particular, location of the target source of seismoacoustic radiation is determined with the accuracy of up to 5 m at the distance of up 40 km from the laser location. Actually, as a result of logical processing, several thousands of the so-called C-OTDR channels are formed on the monitoring distance, each of which transfers information on seismoacoustic activity at the well-defined point of the space. It is obvious that the width of the typical C-OTDR channel is 5 m.

The following problems are solved in the process of analysis of seismic activity:

- Target Seismic-Acoustic Event (TSAE) detection;
- SAE location assessment;
- TSAE type classification.

All these problems are solved on the basis of the so-called “front-end speckle patterns processing” (FESPP). As a result of FESPP, multidimensional information invariants, otherwise called primary signs or features, are singled out from the sequence of the speckle structures corresponding to various C-ORDR channels. Naturally, this analysis is carried out not on a stationary speckle corresponding to the channel state in absence of external disturbances, but on the difference between the speckles adjacent by probing time intervals, which is substantially different from zero. It is this information that is significant for the system and will be called “C-OTDR signal”. And it is C-OTDR signals that are subjected to profound processing in order to solve a complex of problems of remote monitoring. The so-called multidimensional GMM vectors [4] built either by spectral or wavelets coefficients computed above the speckles are used as primary signs in the OXY system. Then, in the space of primary signs, the problems of TSAE detection, location assessment and classification are solved. Use of such multilevel approach allows significantly increase in the anti-interference ability of the system, making it robust against the impact of noises of internal and external nature.

TSAE detection is carried out within the widespread concept of guaranteed detection of statistical disorder of observed processes in C-OTDR channels [5]. TSAE location assessment is based on solving an ordinary triangulation problem using measurements of the adjacent group of channels. The TSAE classification problem seems to be the most difficult. For solving this problem, the approach is used which is based on the ensemble of Lipschitz classifiers, namely the Support Vector Machine ensemble [6]. As a rule, a multitude of target classes has the cardinal number of $m > 10$. Thus, the classification problem has to be solved in the multiclass formulation. In practice, the “one-against-all” approach has turned out to be efficient, within the framework of which the m-class problem is replaced by a series of m binary classification problems, each of which is solved efficiently with the help of the SVM ideology. The Bhattacharyya kernel [7] having good smoothing characteristics was used as the SVM kernel function. To assess reliability of the classification solution, the confidence set approach [8] was used. Use of the phenomenon of multichanneling, consisting of registration of data from channels of the adjacent group, made it possible to significantly increase reliability of solution of the classification problem, at the same time minimizing the impact of the medium of propagation of seismoacoustic waves. The multi-class classifier of TSAE was built based on the conventional SVM (Support Vector Machine). The classification result was formed on the basis of the widespread "one-against-all" approach. The coefficients α_k were determined for each classifier f_k independently during the training phase, $|\Theta| = 20$. The results of the usage of suggested approach in real C-OTDR system are represented in the Table I. In experiment, the targeted sets $\Xi\ \tilde{\theta}_k \mid f_k^{(N)}\ \cdot\ ,\cdot, c_k(P_c)$ were formed for each classifier $f_k(\theta \mid \theta^*)$ with usage of the method which has been described in [1]. After that, two classifiers have been selected. First of them is $f_{\min}$ and it corresponds to the target set with minimal power. Second of them is $f_{\max}$ and it corresponds to the target set with maximal power. Here, in order to avoid bulkiness, we used the following notation:

- $\Xi\ F, P_c$ is the target set which corresponds to integral classifier;
- $\Xi\ f_{\min}^{(N)}, P_c$ is the target set which corresponds to $f_{\min}$;
- $\Xi\ f_{\max}^{(N)}, P_c$ is the target set which corresponds to $f_{\max}$.

The first column of Table I contains the values of the confidence coefficients, the rest - the average cardinality of the confidence sets: $\Xi\ F\ \cdot\ , P_c$, $\Xi\ f_{\min}^{(N)}, P_c$ and $\Xi\ f_{\max}^{(N)}, P_c$. Cardinalities of those sets predictably decrease with decreasing value of the confidence coefficient. The highest accuracy of identification comes from integral classifier **F** with the target set $\Xi\ F\ \cdot\ , P_c$.

TABLE I
THE DEPENDENCE OF THE TARGET SET CARDINALITY ON THE CONFIDENCE COEFFICIENT P_c

P_c	$\lvert\Xi(F, P_c)\rvert$	$\lvert\Xi(f_{\min}^{(N)}, P_c)\rvert$	$\lvert\Xi(f_{\max}^{(N)}, P_c)\rvert$
0.95	1	2	3
0.90	1	2	3
0.85	1	2	2
0.75	1	1	2

V. CONCLUSION

The guaranteed estimates accuracy of the ensemble of Lipschitz classifiers suggested in this paper are primarily designed to be used in case of large number of classes. A practical example specified in this paper clearly illustrates the perspectives of practical use of the proposed estimation.

ACKNOWLEDGMENT

This investigation has been produced under the project "Development of a remote monitoring system to protect backbone communications infrastructure, oil and gas pipelines and other extended ", financed under the project "Technology Commercialization", supported by the World Bank and the Government of the Republic of Kazakhstan.

REFERENCES

[1] A.V. Timofeev, "The guaranteed estimation of the Lipschitz classifier accuracy: Confidence set approach." Journal Korean Stat. Soc. 41(1), 2012, pp.105-114.

[2] L. Breiman, "Bagging predictors", Mach. Learn. 24(2), 1996, pp. 123–140.

[3] P. Rangel, F. Lozano, E. García, "Boosting of support vector machines with application to editing", Proceedings of the 4nd International Conference of Machine Learning and Applications ICMLA'05, Springer 2005.

[4] C.H. You, K.A. Lee, H. Li, "A GMM supervector Kernel with the Bhattacharyya distance for SVM based speaker recognition", ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing, 2009, pp. 4221-4224.

[5] A.V. Timofeev, "The guaranteed detection of the seismoacoustic emission source in the C- OTDR systems", ICDSP 2014: International Conference on Digital Signal Processing Conference Proceedings, Barcelona, Vol.8 N. 10, 2014, pp. 979 – 982.

[6] U. Luxburg, O. Bousquet, "Distance-based classification with Lipschitz functions", Journal of Machine Learning Research, 5, 2004, pp. 669-695.

[7] Kailath T. The Divergence and Bhattacharyya Distance Measures in Signal Selection, IEEE Transactions on Communication Technology 15 (1), 1967, pp. 52-60.

[8] Timofeev A.V., Egorov D.V., Multichannel classification of target signals by means of an SVM ensemble in C-OTDR systems for remote monitoring of extended objects, MVML-2014 Conference Proceedings V.1, Prague, 2014.

Timofeev Andrey V. was born in Chita (Russia). He received Dr. habil. sc. ing.in Computer and Information Sciences from Tomsk State University of Control Systems and Radioelectronics, Russia, in 1994. A number of research publications in the International journals (JKSS, Stat. Methodology, Automation and Remote Control etc.) and International/National conferences are at his credit. He is on the editorial board of several journals and conferences and a referee of several others. His research interests include non-asymptotic nonlinear methods of confidence estimation of multidimensional parameters of stochastic systems; machine learning, large margin classification in Banach Spaces; confidence Lipschitz classifiers; technical diagnostics, C-OTDR systems; data mining; change-point problem; alpha-stable laws; statistical classification in application to biometrics and seismic.