

## COMPARISON OF VARIOUS APPROACHES TO MULTI-CHANNEL INFORMATION FUSION IN C-OTDR SYSTEMS FOR REMOTE MONITORING OF EXTENDED OBJECTS

A.V. Timofeev<sup>a</sup>

<sup>a</sup> LPP “EqualiZoom”, Astana, 010000, Kazakhstan  
Corresponding author: [timofeev.andrey@gmail.com](mailto:timofeev.andrey@gmail.com)

### Article info

Received 24.11.2014, accepted 17.12.2014  
doi: 10.17586/2226-1494-2015-15-1-122-129  
Article in English

**Reference for citation:** Timofeev A.V. Comparison of various approaches to multi-channel Information fusion in c-otdr systems for remote monitoring of extended objects. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2015, no. 1, vol.15, pp. 122–129 (in English)

**Abstract.** The paper presents new results concerning selection of optimal information fusion formula for ensembles of C-OTDR channels. Here C-OTDR is a coherent optical time domain reflectometer. Each of these channels provides data for appropriate automatic classifier which is designed to classify the elastic vibration sources in the multiclass case. Those classifiers form a so-called classifiers ensemble. Ensembles of Lipschitz Classifiers were considered. In this case the goal of information fusion is to create an integral classifier designed for effective classification of seismoacoustic target events. The Matching Pursuit Optimization Ensemble Classifiers (MPOEC), the Linear Programming Boosting (LP-Boost) (LP- $\beta$  and LP-B variants), the Multiple Kernel Learning (MKL), and Weighing of Inversely as Lipschitz Constants (WILC) approaches were compared. The WILC is a brand new approach to optimal fusion of Lipschitz Classifiers Ensembles. The basics of these methods have been briefly described along with intrinsic features. All of those methods are based on reducing the task of choosing convex hull parameters to a solution of an optimization problem. All of the mentioned approaches can be successfully used for using in the C-OTDR system data processing. Results of practical usage are presented.

**Keywords:** C-OTDR channels, MKL, LP-Boost, Lipschitz Classifiers Ensembles.

УДК 004.942, 001.891.57

## СРАВНЕНИЕ РАЗЛИЧНЫХ ПОДХОДОВ К МУЛЬТИКЛАССОВОМУ СЛИЯНИЮ ИНФОРМАЦИИ В С-ОТДР СИСТЕМАХ ДЛЯ УДАЛЕННОГО МОНИТОРИНГА ПРОТЯЖЕННЫХ ОБЪЕКТОВ

А.В. Тимофеев<sup>a</sup>

<sup>a</sup> ТОО «Эквализум», Астана, 010000, Казахстан,  
Адрес для переписки: [timofeev.andrey@gmail.com](mailto:timofeev.andrey@gmail.com)

### Информация о статье

Поступила в редакцию 24.11.2014, принята к печати 17.12.2014  
doi: 10.17586/2226-1494-2015-15-1-122-129  
Язык статьи – английский

**Ссылка для цитирования:** Тимофеев А.В. Сравнение различных подходов к мультиклассовому слиянию информации в С-ОТДР системах для удаленного мониторинга протяженных объектов // Научно-технический вестник информационных технологий, механики и оптики. 2015. Том 15. № 1. С. 122–129

**Аннотация.** Рассматривается проблема выбора оптимального способа слияния информации, поступающей от множества каналов системы когерентной оптической рефлектометрии во временной области (С-ОТДР). Каждый из этих каналов является источником данных для автоматических классификаторов источников упругой сейсмоакустической вибрации. Эти классификаторы образуют ансамбль классификаторов. Были рассмотрены ансамбли липшицевых классификаторов. В этом случае целью слияния информации является построение интегрального классификатора, предназначенного для эффективной классификации целевых сейсмоакустических событий. Сравнивались следующие методы: поиск совпадений на ансамбле классификаторов (МПОЕС), бустинг с линейным программированием (LP-Boost) (LP- $\beta$  и LP-B модификации), обучение с множественными ядерными функциями (МКЛ) и метод обратных констант Липшица (WILC). Выделен WILC как новый подход к оптимальному слиянию информации на ансамбле липшицевых классификаторов. Обсуждены основы работы каждого из методов с указанием специфических особенностей. Показано, что все методы основаны на сведении проблемы выбора параметров выпуклой оболочки к

оптимизационной задаче. Все рассмотренные подходы могут успешно использоваться при обработке информации в C-OTDR системах. Представлены результаты практического использования этих методов.

**Ключевые слова:** C-OTDR каналы, MKL, LP-Boost, ансамбль липшицевых классификаторов

### Introduction

Application of the C-OTDR (Coherent Optical Time Domain Reflectometer) technology to decide various problems of extended objects remote monitoring is currently being evaluated as a very promising approach [1, 2]. In particular, this technology can be effectively used to monitor oil and gas pipelines, controlling technological processes and identifying unauthorized activities in close proximity of the monitored objects. Simplistically, a C-OTDR-system consists of an infrared laser, an optical fiber and a processing unit. The laser sends the probing signals through the optical fiber which is buried in the vicinity of the monitoring object. The processing unit is designed for comprehensive processing of the backscattered signals, which are called speckle-structures. The main item of the C-OTDR technology is a comprehensive analysis of the Rayleigh backscattered radiation characteristics, which transforms into an energetically weakened pulse and propagates constantly in the direction opposite to the direction of a pulsed laser flow. The reflected signal is created by the presence of static impurities in the optical fiber body and defects in the microstructure. Signals scattered by the centers coherently and randomly interfere with each other, forming so-called speckle patterns. Speckle patterns corresponding to different sections of the optical fibers are recorded and accumulated in the data center. The slightest change of the reflectance index value of the fiber, which occurred in a particular place, radically changes the speckle pattern corresponding exactly to this place of the fiber. These changes are reliably detected by the data center. The local changes in refractive index occur under the impact of temperature or due to mechanical action on the optical fiber surface. Let us call the optical fiber buried in the soil to a depth of 50–100 cm, a fiber optical sensor (FOS). Mechanical stress on the FOS surface is caused by seismic acoustic waves. These waves are generated of the sources of elastic vibrations (SEV) which located in vicinity of laying the FOS. Upon reaching the FOS, seismoacoustic wave causes a local longitudinal microstrain on its surface. Those microstrains in turn, cause a change in the local refractive index of light in a relatively small sector of the FOS. As a result, the speckle pattern, which corresponds to this sector, changes significantly. Thus, the FOS quite accurately reflects the state of the seismoacoustic field in its vicinity. The seismoacoustic field contains information about events that occur in the surface layers of the ground near the FOS. This field is created by structural waves, which generated due to mechanical effects on the soil or as a result of a seismic activity. Walking or running man, traffic, earthworks, including hand digging are typical sources of the seismoacoustic emission (structural acoustic wave). In this case, the frequency range of the seismoacoustic waves is in the interval of 0.1 Hz to 1000 Hz. The information, which is required for correct identify the type of SEV, is concentrated in the frequency range of 0.1 Hz to 500 Hz, while 95% of the meaningful information is in even the more narrow range of 1 Hz to 350 Hz. The spectral characteristics of the target signals which lie above and below this frequency range carry information only about the individual characteristics of the SEV. The SEV, which are subjects of interest for remote C-OTDR monitoring will be called a target SEV (TSEV). For the convenience of data processing, the entire FOS length is broken to successive portions (sites) each has length around 10-15 m. The data from those sites is processed separately. These sites will be called C-OTDR channels or just channels. Width of the channel  $\Delta_w$  depends on the probe pulse length. In practice, TSEV has its own small size and assumed point. Due to the nature of the elastic oscillation, the wave from a point source of seismoacoustic emission is usually detected simultaneously in several C-OTDR channels. At the same time, due to strongly anisotropic medium of the elastic vibrations propagation, the structure of the oscillations (speckle patterns) varies considerably between different C-OTDR channels. In each channel a time-frequency characteristics of the speckle pattern are largely reflect a time-frequency structure of the SEV, which occur in vicinity of the corresponding channel. The oscillation energy is considerably attenuated and distorted during propagation in the environment. Intensity of attenuation and distortion depend on the average absorption factor of the medium and on the distance from the oscillation point to the location of channel. Accordingly there is always one channel, wherein vibrations of the TSEV reflected substantially intense than in other channels. We will call this channel dominant. C-OTDR systems perform three major tasks in the following sequence:

1. task “D” (Detection) – detection of the TSEV;
2. task “E” (Estimation) – estimate of the location of the TSEV;
3. task “C” (Classification) – classification of detected TSEV by means of assigning it to one of  $D$  priori given classes.

In the multichannel case, the task “C” has to be solved by creation the method of effective multichannel data fusion. There are number of various approaches to effective multichannel data fusion for task “C” (classification). This paper describes results of a comparing various multichannel data fusion approaches for TSEV classification including a brand new approach which based on weighing of inversely as Lipschitz Constants (WILC) and it allows to improve the generalization ability of the classification system.

**Designations and Research objective**

Let us denote:

- **C-OTDR channels.**  $Ch(K_k)$  is  $k$ -th C-OTDR channel, where a tuple  $K_k = (A_k, R_k)$ , here  $A_k$  is an absorption coefficient of  $k$ -th channel,  $R_k$  is a length of  $k$ -th channel;
- **Feature.** A tuple  $(\underline{Z}, d)$  is a compact feature space where  $\underline{Z}$  is a set of feature values,  $d$  is a metric of  $\underline{Z}$ , data of all channels belongs to  $\underline{Z}$ ;
- **Set of SEV classes.** A set  $\Theta$  is a finite set of indexes of SEV-classes,  $|\Theta| = D$ ;
- **Training Set.**  $\mathbf{Z}_T = \{(Z_i, \theta_i) | i = 1, \dots, N\}$ ,  $|\{(Z_i, \theta_i)\}| = N$ ,  $Z_i = \{z_{1i}, z_{2i}, \dots, z_{mi}\}$ ,  $\theta_i \in \Theta$ , each of  $z_{ki} \in \underline{Z}, k \in \{1, \dots, m\}$ , corresponds to  $Ch(K_k)$ , and to  $\theta_i$ ;
- **True index of SEV class.** A  $\theta^* \in \Theta$  is a true index of the SEV-class to which the samples  $\mathbf{z}_k$  belong, thus  $\theta^*$  is an index of a target class;
- **Samples to classify.** A set  $Z = \{z_1, z_2, \dots, z_m\} [0, 1]$  is feature sample set; each of  $\mathbf{z}_k \in \underline{Z}, k \in \{1, \dots, m\}$ , corresponds to  $Ch(K_k)$ ; in another words, we obtain the feature sample  $\mathbf{z}_k$  from  $k$ -th channel  $Ch(K_k)$ ;
- **Lipschitz Margin Classifier.** Let  $f_k(\theta | \mathbf{z}_k), k \in \{1, \dots, m\}; \theta \in \Theta$ , be a binary Lipschitz Margin classifier (LMC) [3, 4] with Lipschitz Constant (LC)  $L_k$ ;  $f_k(\theta | \mathbf{z}_k) : Z \rightarrow \{\theta, \Theta \setminus \theta\}$  (concept: **one against all**); so, classifier  $f_k(\theta | \cdot)$  divides the feature space  $(Z, d)$  into two classes  $\theta$  and  $\Theta \setminus \theta$ ;  $f_k(\theta | \mathbf{z}_k) = (f_k(\theta | \mathbf{z}_k), R_k)$  here  $f_k(\theta | \mathbf{z}_k) \in R^1$  is **discriminate (stochastic) functions** (so-called score-parameters, which shows similarity degree of a sample  $\mathbf{z}_k$  regarding to class  $\theta \in \Theta$ ; discriminant function  $f_k(\theta | \mathbf{z}_k)$  explicitly dependent on the index hypothesis to be tested  $\theta$  and implicitly on the index of the target class  $\theta^*$ ;  $R_k$  is the classification decision-making rule  $R_k : \tilde{\theta}_k = \text{Arg Max}_{\theta \in \Theta} (f_k(\theta | \mathbf{z}_k))$ ; let us denote  $\psi$ -set of LMC  $f_k(\theta | \mathbf{z}_k, \psi)$  parameters, which needs to be tuned during of training process; otherwise, set  $\psi$  will be denoted as LMCP or LMCP  $\psi$ ;
- **Ensemble of LMC (Lipschitz Classifiers Ensembles).**  $\mathbf{F}(\theta | Z) = \{f_k(\theta | \mathbf{z}_k) | k = 1, \dots, m\}$  is an ensemble of the LMC;  
 $\mathbf{F}_\theta : Z \rightarrow \{\theta, \Theta \setminus \theta\}$  is an integral classifier on the ensemble  $\mathbf{F}(\theta | Z)$ ;  $\mathbf{F}_\theta = (F(\theta | \mathbf{F}(\theta | Z)), \mathbf{R})$ ;  
 $\mathbf{R} : \tilde{\theta} = \text{Arg Max}_{\theta \in \Theta} (F(\theta | \mathbf{F}(\theta | Z)))$  is output of integral classifier  $F(\theta | \mathbf{F}(\theta | Z))$ ;
- $F(\theta | \mathbf{F}(\theta | Z))$  – **discriminate function** on the classifiers ensemble  $\mathbf{F}(\theta | Z)$ ,  $F(\theta | \mathbf{F}(\theta | Z)) = \sum_k \beta_k f_k(\theta | \mathbf{z}_k)$ , where  $\sum_k \beta_k = 1, \forall_k \beta_k \geq 0$ ; coefficients  $\{\beta_k\}$  are determined by various methods, which are object of our investigation.

So there exist  $m$  statistical independent C-OTDR channels  $\mathbf{Ch} = \{Ch(K_k) | k = 1, \dots, m\}$ . Each of those channels depends of external (environmental) parameters tuple  $K_k = (A_k, R_k)$ . Simply speaking, these channels transmit signals from sources of elastic vibrations (SEV) to FOS. Thus signals  $\mathbf{z}_k \in Z, k \in \{1, \dots, m\}$  are outputs of C-OTDR channels  $\mathbf{Ch}$ . The tuple  $K_k$  defines the effectiveness of channel  $Ch(K_k)$  for signal transmission.

The signals  $Z$  are contain relevant information about SEV time-frequency parameters. Every two channels  $Ch(K_k)$  and  $Ch(K_p)$  distort the SEV time-frequency parameters by differently because of external parameters  $K_k$  and  $K_p$  are different. Accordingly we suppose every two different samples  $\mathbf{z}_k$  and  $\mathbf{z}_p$  are statistically independent if  $k \neq p$ . For each C-OTDR channel  $Ch(K_k)$  are used appropriate  $D$  binary classifiers  $f_k(\theta_i | \mathbf{z}_k)$ ,  $\theta_i \in \Theta$ . Each LMC  $f_k(\theta_i | \mathbf{z}_k)$ , is binary classifier, which divides the feature space  $(Z, d)$  into two classes  $\theta_i$  and  $\Theta \setminus \theta_i$ .

So, we need to classify of the SEV type using observation  $Z$  of C-OTDR channels  $\mathbf{Ch}$ . An obvious approach to solving this problem is to use the Lipschitz Classifiers Ensembles ( $\mathbf{F}(\theta | Z)$ ). But the problem of effective multichannel data fusion arises. There are number of various approaches to multichannel data fusion.

The goal of this paper is to compare some data fusion methods effectiveness. A number of known

approaches and one a brand new method were studied. The brand new method is based on use of Lipschitz constants of LMC's.

### Some Approaches to C-OTDR Multichannel Data Fusion for Multiclass Classification of TSEV

So the classification problem TSEV is reduced to the task of creating an effective multiclass classifier (MC) which is based on a LMC classifiers ensemble (Lipschitz Classifiers Ensembles). We remark that an ensemble of classifiers is a set of classifiers whose individual decisions are combined in some way (typically by weighted or unweighted voting) to classify new examples [5]. At any rate a MC learning method choice is a dominant problem. Usually the problem of learning a MC from training data is often addressed by means of kernel method (KM) [6–8]. In this case each kernel corresponds to an appropriate channel of the set  $\mathbf{Ch}$ . For brevity we will not describe the baseline of this well-known method but we are going to pay attention to some KM modifications which are designed to work with LMC classifiers ensembles.

For the sake of simplicity, we will consider as a LMC a classic SVM [9]. By definition a SVM discriminant function  $f_k(\theta | \mathbf{z}_k, \psi)$  depends on the parameters  $\alpha \in R^N$  ( $N$  is a power of the training set  $\mathbf{Z}_T$ ) and  $b \in R^1$ . Here  $\alpha$  is a normal vector to the hyperplane,  $|b|/\|\alpha\|$  is the perpendicular distance from the hyperplane to the origin. Thus we have  $\psi = \{\alpha, b\}$ , and each tuple  $\psi$  defines the hyperplane in the feature space.

#### Multiple Kernel Learning (MKL)

In contrast to baseline kernels selection (“averaging kernels” and “product kernels” [8]), MKL kernel selection is to learn a kernel combination during the training phase of the algorithm. So, the MKL objective is to optimize jointly over a linear combination of kernels  $q(Z^{(i)}, Z^{(j)}) = \sum_{k=1}^m \beta_k k_k(\mathbf{z}_{ki}, \mathbf{z}_{kj})$  with LMCP  $\psi = \{\alpha, b\}$ .

Here  $Z^{(i)} = \{\mathbf{z}_{1i}, \mathbf{z}_{2i}, \dots, \mathbf{z}_{mi}\}$ ,  $Z^{(j)} = \{\mathbf{z}_{1j}, \mathbf{z}_{2j}, \dots, \mathbf{z}_{mj}\}$ ,  $\sum_{k=1}^m \beta_k = 1, \beta_k \geq 0$ ,  $k_k(\cdot, \cdot)$  is kernel function. MKL was originally introduced in [6]. Let us denote  $Q_i(Z) = (q_i(Z, Z_1), q_i(Z, Z_2), \dots, q_i(Z, Z_N)) \in R^N, i = 1, \dots, m$ . The final decision has form

$F_{MKL}(Z) = \text{Arg Max}_{\theta \in \Theta} \left( \sum_{k=1}^m \beta_{k,\theta} \left( Q_k(Z)^T \alpha_\theta + b_\theta \right) \right)$ . The choice of parameters MKL is made by using for each  $\theta$  the following scheme:

$$\min_{\alpha, b, \beta} \left[ 0.5 \left( \sum_{k=1}^m \beta_{k,\theta} \alpha_\theta^T Q_k \alpha_\theta \right) + C \sum_{i=1}^N \left( D(\theta_i, b_\theta + \sum_{k=1}^m \beta_{k,\theta} Q_k^T(Z_i) \alpha \beta_\theta) \right) \right] \text{ sb.t. } \sum_{k=1}^m \beta_{k,\theta} = 1, \beta_{k,\theta} \geq 0.$$

$D(\theta, t) = \max(0, 1 - \theta t)$ ,  $(Z_i, \theta_i) \in \mathbf{Z}_T$ . In other words, in MKL case we optimize jointly the convex hull of kernels. Here for each  $\theta$  we have the same LMCP  $\psi_\theta = \{\alpha_\theta, b_\theta\}$  for different  $k$ .

#### LP-Boost (LP- $\beta$ )

So, we will consider a case when classifiers  $f_k(\theta | \mathbf{z}_k)$  of ensemble  $F(\theta | \mathbf{F}(\theta | Z))$  are not trained jointly, but coefficients  $\{\beta_k\}$  are determined jointly. Here we have a situation where LMCP tuples  $\psi_\theta$  are different for different  $k$ . This method is called the  $\beta$ -LP-Boost [10, 11], and here the final decision has the form

$$F_{LP\beta}(Z) = \text{Arg Max}_{\theta \in \Theta} \left( \sum_{k=1}^m \beta_k \left( Q_k(Z)^T \alpha_{\theta,k} + b_{\theta,k} \right) \right).$$

The training phase comes down to an optimal choice of parameters  $\{\beta_k\}$ . This choice is performed by using standard optimization method (linear programming – LP) according to the following scheme:

$$\min_{\beta, \xi, \rho} \left( -\rho + \frac{1}{vN} \left( \sum_{i=1}^N \xi_i \right) \right),$$

$$\text{under the condition } \sum_{k=1}^m \beta_k \left( k_k(Z, Z_i) \alpha_{\theta,k} + b_{\theta,k} \right) - \text{Arg Max}_{\theta \neq \theta^*} \sum_{k=1}^m \beta_k \left( k_k(Z, Z_i) \alpha_{\theta^*,k} + b_{\theta^*,k} \right) + \xi_i \geq \rho,$$

$i = 1, \dots, N$ ,  $\sum_{k=1}^m \beta_k = 1, \beta_k \geq 0, k = 1, \dots, m$ . Here  $\xi$  – slack variables,  $v$  – regularization constant, which is chosen using Cross Validation (CV). In frame of this approach not need provide the normalization of kernels  $k_k(\cdot)$ . Moreover, features for which  $\beta_k = 0$  need not to be computed for the final decision function.

#### LP-Boost (LP-B)

Another version of LP approach to choice  $\{\beta_k\}$  was called B-LP-Boost [11, 12]. In this case, each class has its own weight vector. So, we have  $(m \times D)$  weighting matrix  $\mathbf{B}$ . The final decision has the form

$$F_{LPB}(Z) = \text{Arg Max}_{\theta \in \Theta} \left( \sum_{k=1}^m \mathbf{B}_k^\theta \left( Q_k(Z)^T \alpha_{\theta,k} + b_{\theta,k} \right) \right).$$

Choice of parameters  $\{\beta_k\}$  we make in such way:

$$\min_{\beta, \xi, \rho} \left( -\rho + \frac{1}{vN} \left( \sum_{i=1}^N \xi_i \right) \right),$$

under the condition  $\sum_{k=1}^m \mathbf{B}_k^{\theta'} (k_k(Z, Z_i) \alpha_{\theta', k} + b_{\theta', k}) - \sum_{k=1}^m \mathbf{B}_k^{\theta''} (k_k(Z, Z_i) \alpha_{\theta'', k} + b_{\theta'', k}) + \xi_i \geq \rho$ ,

$i = 1, \dots, N, \theta' \neq \theta''$ ,  $\forall \theta, m: \left( \sum_{k=1}^m \mathbf{B}_k^{\theta} = 1, \mathbf{B}_k^{\theta} \geq 0, k = 1, \dots, m \right)$ . As above, here  $\xi$  are slack variables,  $v$  – regularization constant,  $v$  is chosen using CV. Here we have a linear programming problem too, but this problem is more expensive because of dimension increasing.

**Matching Pursuit Optimization Ensemble Classifiers**

The matching pursuit optimization approach was described in [13]. In [14] a new method that Matching Pursuit Optimization Ensemble Classifiers (MPOEC) was proposed in order to eliminate some useless or similar classifiers from the whole ensemble to improve the performance of ensembles without decreasing the diversity among classifiers. So, in frame of MPOEC, we adopt a greedy iterative approach [13] to search for an optimal combination of classifiers. We describe this approach very briefly. So, let  $\mathbf{Z}_T = \{(Z_i, \theta_i)\}$  is the training set and we have a finite dictionary  $D = \{d_1, d_2, \dots, d_M\}$  including  $M$  functions. The  $D$  is given in given in the Hilbert Space  $H$ . We can try to find some target function  $q \in H$ ,  $Z \xrightarrow{q} \Theta$ . The sparse approximations of  $q$  can be found out in the dictionary  $D$ , which is expansion of the form  $q_p = \sum_{n=1}^P \alpha_n g_n, \forall \alpha_n \in R^1$ ,  $P$  is the number of the basis functions (BF) in the expansion,  $\{g_n\} \subset D$  is basis of the expansion,  $\{\alpha_n\}$  is the set of coefficient of this expansion. The  $q_N$  is sparse approximation of function  $q$ . The approximation  $q_N$  is built with usage of  $P$  basis functions from  $D$ . The approximation error has the following simple for

$$\|R_p\|^2 = \|q - q_p\|^2 = \sum_{i=1}^P (\theta_i - q_p(Z_i))^2.$$

The basis function  $\{g_n\}$  and the coefficients  $\{\alpha_n\}$  are selected by applying the iterative greedy method, in the interest to minimizing the  $R_p$ . So, in frame of MPOEC approach, the ensemble of classifiers  $\mathbf{F}(\theta | Z)$  is regarded as a BF of the dictionary  $D$ , and each LMC  $f_k(\theta | \mathbf{z}_k) \in \mathbf{F}(\theta | Z)$  obtains the coefficient  $\alpha_k$  by minimizing  $R_p$ . As result of MPOEC procedure usage, the content of ensemble  $\mathbf{F}(\theta | Z)$  is changed. If  $\alpha_k \neq 0$ , the corresponding classifier  $f_k(\theta | \mathbf{z}_k)$  remains in ensemble, otherwise the  $f_k(\theta | \mathbf{z}_k)$  is eliminated of ensemble  $\mathbf{F}(\theta | Z)$ . On each of MPOEC procedure iteration  $t$  we have the various combination  $\mathbf{F}_t(\theta | Z)$  of LMC from  $\mathbf{F}(\theta | Z)$ . For convenience let us denote an index of LMC inside the temporary combination  $\mathbf{F}_t(\theta | Z)$  (temporary ensemble) as  $j$ . On each iteration  $t$  the prediction error is calculated as following

$$e_t = N^{-1} \sum_{i=1}^N (\theta_i - q_t(Z_i)) = N^{-1} \sum_{i=1}^N \left( \theta_i - \text{sign} \left( \sum_{j=1}^t \alpha_j f_j(\theta | \mathbf{z}_{ji}) \right) \right).$$

Here  $Z_i = \{\mathbf{z}_{i1}, \mathbf{z}_{i2}, \dots, \mathbf{z}_{mi}\}$ ,  $t$  – number of iteration,  $\alpha_j$  is the coefficient gained in  $j$ -th iteration. On each iteration  $t$  the residues  $R_t$  refined by the formula  $R_t = R_{t-1} - \alpha_t f_t(\cdot)$ , and coefficients  $\{\alpha_j\}$  are calculated as following formula  $\alpha_j = \langle f_j, R_{t-1} \rangle / \|f_j\|^2$ ,  $j = 1, \dots, |F_t(\cdot)|$ . The MPOEC procedure iterations are repeated until prediction error  $e_t$  is more than prior given threshold  $\lambda > 0$ . So, the MPOEC method selects LMC according to diversity between a pair of LMC to increase the diversity of the ensemble  $F(\cdot)$ . On the other hand, the increase in the diversity will not reduce the accuracy of individual classifiers in the optimization process of MPOEC approach.

**MKL Weighing of Inversely as Lipschitz Constants (WILC-MKL)**

Let us consider the brand new modification of the MKL that differ from classical MKL by method choice of linear combination parameters. The motivation of this approach is using some intrinsic properties of LMC. The fact is that value of Lipschitz Constant significantly determines of the LMC properties. Simply speaking, the Lipschitz classifier decision function has to a small Lipschitz constant. This feature comes from well-known regularization principle, which recommends avoid using discriminative functions with a high variation. So, LMC's with small LC are more preferable for providing of stable classification process. In other words, classifiers with small LC provide the greater generalization ability of classification system: such difficulties have lower complexity to avoid overfitting. Hence, in formula of  $F(\theta | \mathbf{F}(\theta | Z))$  LMC's with small LC must get weight coefficients with bigger value. Let us call this approach to modification of MKL as Weighing of Inversely as to value of the Lipschitz Constant (WILC) or WILC-MKL. In frame of WILC-MKL approach to LMC-ensemble  $\mathbf{F}(\theta | Z)$  we have the following discriminative function:

$$F_{WILC}(Z) = \text{Arg Max}_{\theta \in \Theta} \left( \sum_{k=1}^m \sigma_{\theta,k} \beta_{k,\theta} \left( Q_k(Z)^T \alpha_{\theta} + b_{\theta} \right) \right).$$

Here  $\sigma_{\theta,k} = L_{0,k}^{-1} \left( \sum_{j=1}^m L_{0,j}^{-1} \beta_{j,\theta} \right)^{-1}$ ,  $k=1, \dots, m$ ,  $\sigma_{\theta,k} \beta_{k,\theta} > 0$ ,  $\sum_{k=1}^m \sigma_{\theta,k} \beta_{k,\theta} = 1$ ,  $L_{0,k}$  is Lipschitz Constant of discrimination function  $K_k(Z)^T \alpha_{\theta} + b_{\theta}$ . Thus, using WILC-MKL, we make attempt to improve the generalization ability of MKL by considering information about variation characteristics of classifiers discrimination functions. As it was shown in series of practical experiments, usage of WILC-MKL allows considerably improve the performance of LMC-ensemble in some practical cases.

### Results of practical usage

All of above described methods were used for multichannel TSEV classification in C-OTDR system of railways monitoring. This system was successfully installed on the railways test area (RTA) of Kazakhstan Railways Company (JSC NC "KTZ") in august of 2014, and this system continues to operate. The RTA is located at a distance of 10 km from Astana City. The FOS which has been installed on RTA has length around 2000 m, depth of the FOS laying is 50 cm approximately, and FOS offset from rails is 5 m. Parameters of the C-OTDR system:

- duration of the probe pulse is 50–200 ns;
- period of probe pulse belongs to the range 50–300  $\mu$ s;
- laser wavelength is 1550 nm.

In this case, the main problem is to fusion of multichannel data to classify the TSEV with maximum accuracy. As was said above, for each C-OTDR channel  $Ch(K_k)$  are used appropriate  $D$  binary classifiers  $f_k(\theta_i | \mathbf{z}_k)$ ,  $\theta_i \in \Theta$ . Each LMC  $f_k(\theta_i | \mathbf{z}_k)$  is binary classifier, which divides the feature space  $(\underline{Z}, d)$  into two classes  $\theta_i$  and  $\Theta \setminus \theta_i$ . Each LMC  $f_k(\cdot)$  was trained independently, and each LMC uses the same set of features in the space  $(\underline{Z}, d)$ . The  $(\underline{Z}, d)$  is the ordinary GMM-vector space [15]. We describe the procedure for calculation of the GMM-vectors very briefly. On feature extraction phase for each speckle pattern obtained in the probing period  $T$  for each of the channel are built Linear-Frequency Spaced Filterbank Cepstrum Coefficients (LFCC). In our case these features are based on 10 linear filter-banks (from 0.1 to 500 Hz) derived cepstra. Thus, 10 static and 10 first-order delta coefficients were used, giving the feature order  $m = 20$ . Further, approximation of the probability distribution function of the feature vectors (LFCC) by semi-parametric multivariate probability distribution model, so-called Gaussian Mixture Models (GMM), was carried. Presently, the GMM is one of the principal methods of modeling broadband acoustic emission sources (including TSEV) for their robust identification. The GMM of TSEV feature vectors distribution is a weighted sum of  $J$  components densities [15] and given by the equation  $P(x|\lambda_s) = \mathbf{w}_s \mathbf{B}_s^T(x)$ , where  $x$  is a random  $m$ -vector,  $\mathbf{w}_s = (w_{s1}, \dots, w_{sJ}) \in R^J$ ,

$$\mathbf{B}_s(x) = (B_{s1}(x), \dots, B_{sJ}(x)) \in R^J, \forall_{s,i} B_{s,i}(x) = \left( (2\pi)^{m/2} |\Sigma_{si}|^{1/2} \right)^{-1} \exp \left( -\frac{1}{2} (x - \mu_{si})^T \Sigma_{si}^{-1} (x - \mu_{si}) \right),$$

$\lambda_s = \{(w_{si}, \mu_{si}, \Sigma_{si}) | i=1, J\}$ . In general, diagonal covariance matrices  $\Sigma_{si}$  are used to limit the model size. The model parameters  $\lambda_s$  characterize a SEV in the form of a probabilistic density function. During training, those parameters are determined by the well-known expectation maximization (EM) algorithm [15]. In the described experiments value  $J$  was equal to 1024. Thus, for identification of TSEV class, each TSEV is modeled by a GMM-vector and is referred to as his model parameters  $\lambda \in \underline{Z}$ . The classic SVM with Bhattacharyya-kernel [3] was used as the LMC.

Priori defined target classes of TSEV, which collectively makes up a finite set  $\Theta$ . For example, in case of railways monitoring the array  $\Theta$  consists of the following TSEV classes: "train", "hand digging the soil", "chiseling ground scrap", "pedestrian", "group of pedestrian", "passenger car", "heavy equipment excavator", "easy excavation equipment". Five alternative approaches for multichannel data fusion were compared on stage of TSEV classification. In particular, MKL, LP- $\beta$ , LP-B, MPOEC, and WILC-MKL approaches were used. The results of using these methods as parts of the C-OTDR system are presented in Table.

In the process of using the method WILC-MKL values of Lipschitz Constants were evaluated numerically for each LMC from ensemble  $\mathbf{F}(\theta | Z)$ . The volumes of training sets were equal for each of various data fusion approaches, but those volumes were different for various TSEV types. Presented results prove that the WILC-MKL and LP ( $\beta$  and B) are more effective with respect to MKL and MPOEC approaches. At the same time, WILC-MKL is more effective compared to MKL, but the LP-B is the best approach for a fusion of multichannel data in C-OTDR monitoring systems. It is important; the LP-B approach requires more computing resources than the WILC-MKL approach, wherein the accuracies of those methods are close. That is why the WILC-MLK approach is preferable from the practical point of view.

Method	Type of SEV	Accuracy	Volume of training set
MKL	"hand digging the soil"	76%	60
	"chiselling ground scrap"	79%	60
	"pedestrian"	78%	80
	"group of pedestrians "	79%	30
	"passenger car"	79%	50
	"train"	100%	150
	"heavy equipment excavator"	81%	20
	"easy excavation equipment"	83%	20
	"shrew digging the ground"	81%	30
LP-β	"hand digging the soil"	81%	60
	"chiselling ground scrap"	83%	60
	"pedestrian"	81%	80
	"group of pedestrians "	83%	30
	"passenger car"	80%	50
	"train"	100%	150
	"heavy equipment excavator"	85%	20
	"easy excavation equipment"	86%	20
	"shrew digging the ground"	84%	30
LP-B	"hand digging the soil"	82%	60
	"chiselling ground scrap"	85%	60
	"pedestrian"	79%	80
	"group of pedestrians "	84%	30
	"passenger car"	81%	50
	"train"	100%	150
	"heavy equipment excavator"	86%	20
	"easy excavation equipment"	88%	20
	"shrew digging the ground"	88%	30
MPOEC	"hand digging the soil"	72%	60
	"chiselling ground scrap"	79%	60
	"pedestrian"	75%	80
	"group of pedestrians "	80%	30
	"passenger car"	79%	50
	"train"	100%	150
	"heavy equipment excavator"	81%	20
	"easy excavation equipment"	82%	20
	"shrew digging the ground"	83%	30
WILC-MKL	"hand digging the soil"	81%	60
	"chiselling ground scrap"	82%	60
	"pedestrian"	78%	80
	"group of pedestrians "	83%	30
	"passenger car"	84%	50
	"train"	100%	150
	"heavy equipment excavator"	84%	20
	"easy excavation equipment"	87%	20
	"shrew digging the ground"	86%	30

Table. The practical detection results

### Conclusions

This paper describes results of comparison of various multichannel data fusion approaches for TSEV classification including MKL, LP-β, LP-B, MPOEC and WILC-MKL. The practical usage of these approaches proves better effectiveness of LP-B approach to fusion of multichannel data for classification of TSEV type. A brand new approach, WILC-MKL, was suggested for multichannel data fusion. This approach is simple to use and performs well in a C-OTDR classification subsystem.

## References

1. Choi K.N., Juarez J.C., Taylor H.F. Distributed fiber-optic pressure/seismic sensor for low-cost monitoring. *Proceedings of SPIE - The International Society for Optical Engineering*, 2003, vol. 5090, pp. 134–141. doi: 10.1117/12.484911
2. 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. *Proc. MVML-2014*. Prague, 2014, vol. 1, pp. 22–32.
3. Timofeev A.V. The guaranteed estimation of the Lipschitz classifier accuracy: confidence set approach. *Journal of the Korean Statistical Society*, 2012, vol. 41, no. 1, pp. 105–114. doi: 10.1016/j.jkss.2011.07.005
4. Von Luxburg U., Bousquet O. Distance-based classification with Lipschitz functions. *Journal of Machine Learning Research*, 2004, vol. 5, pp. 669–695.
5. Murty M.N., Devi V.S. *Pattern Recognition: an Algorithmic Approach (Undergraduate Topics in Computer Science)*. London, Springer, 2011, 263 p.
6. Hofmann T., Sholkopf B., Smola A.J. Kernel methods in machine learning. *Annals of Statistics*, 2008, vol. 36, no. 3, pp. 1171–1220. doi: 10.1214/009053607000000677
7. Lanckriet G.R.G., Cristianini N., Bartlett P., El Ghaoui L., Jordan M.I. Learning the kernel matrix with semidefinite programming. *Journal of Machine Learning Research*, 2004, vol. 5, pp. 27–72.
8. Bach F.R., Lanckriet G.R.G., Jordan M.I. Multiple kernel learning, conic duality, and the SMO algorithm. *Proc. 21<sup>st</sup> Int. Conf. on Machine Learning, ICML 2004*. Banff, Canada, 2004, pp. 41–48.
9. Hears M.A., Dumais S.T., Osman E., Platt J., Scholkopf B. Support vector machines. *IEEE Intelligent Systems*, 1998, vol. 13, no. 4, pp. 18–28. doi: 10.1109/5254.708428
10. Demiriz A., Bennett K.P., Shawe-Taylor J. Linear programming boosting via column generation. *Machine Learning*, 2002, vol. 46, no. 1–3, pp. 225–254. doi: 10.1023/A:1012470815092
11. Gehler P., Nowozin S. On feature combination for multiclass object classification. *Proc. ICCV*, 2009, pp. 47–50.
12. Weston J., Watkins C. Support vector machines for multi-class pattern recognition. *Proc. 7<sup>th</sup> European Symposium on Artificial Neural Networks*, 1999, pp. 219–224.
13. Mallat S.G., Zhang Z. Matching pursuit with time-frequency dictionaries. *IEEE Transaction on Signal Processing*, 1993, vol. 41, no. 12, pp. 3397–3415. doi: 10.1109/78.258082
14. Mao S., Jiao L.C., Xiong L., Gou S. Greedy optimization classifiers ensemble based on diversity. *Pattern Recognition*, 2011, vol. 44, no. 6, pp. 1245–1261.
15. Blimes J.A. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. *Technical Report TR-97-021*. Berkeley, California, International Computer Science Institute, 1998, pp. 97–021.

**Andrey V. Timofeev** – Dr. Habil. Ing., CSO, LPP “EqualiZoom”, Astana, 010000, Kazakhstan, timofeev.andrey@gmail.com

**Тимофеев Андрей Владимирович** – доктор технических наук, научный директор, ТОО «Эквализум», Астана, 010000, Казахстан, timofeev.andrey@gmail.com