Proceedings of the World Congress on Electrical Engineering and Computer Systems and Science (EECSS 2015)
Barcelona, Spain – July 13 - 14, 2015
Paper No. 327

Various Approaches to Multi-Channel Information Fusion in C-OTDR Systems

Andrey Timofeev, Dmitry Egorov
JSK "EqualiZoom"
18A Akzholov Str, Astana, 010000, Kazakhstan
timofeev.andrey@gmail.com, edv017@gmail.com

Abstract- The paper presents new results concerning selection of optimal information fusion formula for ensembles of monitoring system channels. The goal of information fusion is to create an integral classificator designed for effective classification of targeted events, which appear in the vicinity of monitored object. The LPBoost (LP-β and LP-B variants), the Multiple Kernel Learning, 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. Results of practical usage are presented.

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

1. Introduction

In common cases monitoring systems designed to control of super-extended objects (oil and gas pipelines, national borders, railways etc) operate with a stream of targeted events. For example, in the case of C-OTDR (coherent optical time domain reflectometers) monitoring systems (Taylor et al., 2003), the targeted events stream contains all seismoacoustic events, which have occurred in the vicinity of monitoring object. Monitoring systems of this class are brand new monitoring systems which are the most promising for the control of super-extended objects. Information about targeted events is transmitted via a physical field as physical waves (electromagnetic, seismic acoustic, infrared, etc). Core of C-OTDR monitoring systems (OXY) functionality is based on the modern methods of reflectometric interferential spectroscopy and high vibrosensitivity of a coherent flow of infrared energy that is injected into a dedicated fiber optic cable through a conventional infrared laser with a wavelength of 1550.116 nm. The OXY-system is designed to operate in a fully autonomous mode and allows detecting and classifying suspicious (target) seismic acoustic activities in the area along protected objects. The detection accuracy of the mechanical activity is ~ 5-10 m. Walking or running man, traffic and excavation including hand digging are typical sources of acoustic emission (structural sound waves). Monitoring system sensors register those waves, transform wave's parameters into a different format, and transmit that information via monitoring channels to data centers. Information about one and the same targeted event (TE) might be reflected in several channels simultaneously. For example, in case of OXY-system, due to the nature of the elastic oscillation, the wave from a point source of seismoacoustic emission (targeted events or signals) 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 time-frequency characteristics of the speckle pattern are largely reflected a time-frequency structure of the seismoacoustic *energy emission sources*, which occur in the vicinity of the corresponding channel. The oscillation energy is considerably attenuated and distorted during propagation in the environment. The intensity of attenuation and distortion depends on the average absorption factor of the medium and on the distance from the oscillation point to the location of channel. Noises in the channels are mutually independent if there is no signal. So, it is useful to process this information jointly. There is, however, a problem. The number of monitored channels may be huge, for example, in C-OTDR monitoring systems (Timofeev et al., 2014) numbers of channels exceed tens of thousands. Joint processing of raw data in such systems leads to huge computational costs. That is why, at the first step, monitored raw data from different channels is processed separately. And only during the consequent steps multichannel data are processed jointly. Monitoring systems perform three major tasks in the following sequence: a)Task "D" (Detection) – detection of the TE; b) Task "E" (Estimation) – estimate of the location of the TE; c) Task "C" (Classification) – classification of detected TE by means of assigning it to one of *D* priori given classes. All of those tasks are solved for each channel as a first processing step. In particular, Task "C" is solved using an automatic classifier, which is based on data from only one channel. The classification results are shifted to the next level of data processing. Multichannel data (single-channel classification results) have to be fused on this level. So, multichannel data fusion is a very important step in monitored data processing. There are number of various approaches to effective multichannel data fusion for task "C". This report describes results of comparing various multichannel data fusion approaches for TE classification including a brand new approach which is based on weighing of inversely as Lipschitz Constants (WILC) and it allows us to improve the generalization ability of the classification system.

2. Designations and Research objective

In this section we present used designations and research objectives with necessary comments.

2. 1. Designations

- *Energy emission sources* (EES). EES's are targets of monitoring. In depending from physics field type, EES might generates various types of energy: seismicacoustic, infrared, electromagnetic etc. If EES type is a targeted type, then this EES call "targeted event" (TE) for monitoring system. For example, in C-OTDR monitoring systems which designed to monitor of railways, EES "train" is TE, but EES "pedistrain" is not TE.
- *Monitoring system channels*. $Ch(\kappa_k)$ is k-th channel, where a tuple $\kappa_k = (\mathrm{A}_k, R_k)$, here A_k is an absorption coefficient of *k*-th channel, R_k is a length of k-th channel; channel length is a distance from EES to sensor of monitoring system.
- *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 $\underline{Z}$.
- *Training Set.* $\mathbf{Z}_T = \{(Z_i, \theta_i) \mid i = 1,...N\}$, $Z_i = \{\mathbf{z}_{1i}, \mathbf{z}_{2i}, ..., \mathbf{z}_{mi}\}$, $\theta_i \in \Theta$, each of $\mathbf{z}_{ki} \in \underline{Z}$, $k \in \{1,...,m\}$, corresponds to $Ch(\kappa_k)$, and to θ_i.
- *True index of TE class*. A $\theta^* \in \Theta$ is a **true index** of the TE-class to which the samples $\mathbf{z}_k$ belong, thus θ^* is an index of a **target class.**
- *Samples to classify*. A set $Z = \{\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_m\}$ is feature sample set; each of $\mathbf{z}_k \in \underline{Z}, k \in \{1,...,m\}$ corresponds to $Ch(\kappa_k)$; in another words, we obtain the feature sample $\mathbf{z}_k$ from *k*-th channel $Ch(\kappa_k)$.
- *Lipschitz Margin Classifier (LMC)*. Let $\mathrm{f}_k(\theta \mid \mathbf{z}_k), k \in \{1,...,m\}; \theta \in \Theta,$ be a binary Lipschitz Margin classifier (Timofeev, 2012; Bousquet and Luxburg, 2004) with Lipschitz Constant (LC) L_k; $\mathrm{f}_k(\theta \mid \mathbf{z}_k) : Z \to \{\theta, \Theta \setminus \theta\}$ (concept: **one against all**); so, classifier $\mathrm{f}_k(\theta \mid \cdot)$ divides the feature space (Z, d) into two classes θ and $\Theta \setminus \theta$; $\mathrm{f}_k(\theta \mid \mathbf{z}_k) = (f_k(\theta \mid \mathbf{z}_k), R_k)$ here $f_k(\theta \mid \mathbf{z}_k) \in R^1$ 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 $f_k(\theta \mid \mathbf{z}_k)$ 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 $R_k : \tilde{\theta}_k = Arg \underset{\theta\in\Theta}{Max}\left(f_k\left(\theta \mid \mathbf{z}_k\right)\right)$; let us denote ϑ - set of LMC $f_k\left(\theta \mid \mathbf{z}_k, \vartheta\right)$ parameters, which needs to be tuned during of training process; otherwise, set ϑ will be denoted as LMCP or LMCP ϑ.

- *Ensemble of LMC*. The set $\mathbf{F}(\theta \mid Z) = \{\mathrm{f}_k\left(\theta \mid \mathbf{z}_k\right) \mid k = 1,...,m\}$ is an ensemble of the LMC.
- *Integral Classifier*. $\mathbf{F}_\theta : Z \to \theta, \Theta \setminus \theta$ is an integral classifier on the ensemble $\mathbf{F}(\theta \mid Z)$; $\mathbf{F}_\theta = \left(F\left(\theta \middle| \mathbf{F}(\theta \mid Z)\right), \mathbf{R}\right)$; here the rule $\mathbf{R} : \tilde{\theta} = Arg \underset{\theta\in\Theta}{Max}\left(F\left(\theta \middle| \mathbf{F}(\theta \mid Z)\right)\right)$ is output of integral classifier $F\left(\theta \middle| \mathbf{F}(\theta \mid Z)\right)$.
- *Discriminate function*. $F\left(\theta \middle| \mathbf{F}(\theta \mid Z)\right)$ is a discriminate function on the classifiers ensemble $\mathbf{F}(\theta \mid Z)$, $F\left(\theta \middle| \mathbf{F}(\theta \mid Z)\right) = \sum_k \beta_k f_k\left(\theta \mid \mathbf{z}_k\right)$, where $\sum_k \beta_k = 1, \forall k : \beta_k > 0$; coefficients $\{\beta_k\}$ are determined by various methods, which are object of our investigation.

2. 2. Research Objective

So there exist *m* statistical independent monitoring channels $\mathbf{Ch} = \{Ch(\kappa_k) \mid k = 1,...,m\}$. Each of those channels depends of external (environmental) parameters tuple $\kappa_k = (\mathrm{A}_k, R_k)$. Simply speaking, these channels transmit signals from EES to sensors of monitoring system. Thus signals $\mathbf{z}_k \in Z_k, k \in \{1,...,m\}$ are outputs of monitoring channels $\mathbf{Ch}$. The tuple κ_k defines the effectiveness of channel $Ch(\kappa_k)$ for signal transmission. The signals Z are contain relevant information about TE time-frequency parameters. Every two channels $Ch(\kappa_k)$ and $Ch(\kappa_p)$ distort the TE time-frequency parameters by differently because of external parameters κ_k and κ_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 channel $Ch(\kappa_k)$ are used appropriate *D* binary classifiers $\mathrm{f}_k\left(\theta_i \mid \mathbf{z}_k\right)$, $\theta_i \in \Theta$. Each LMC $\mathrm{f}_k\left(\theta_i \mid \mathbf{z}_k\right)$ is binary classifier, which divides the feature space (Z, d) into two classes θ_i and $\Theta \setminus \theta_i$

So, we need to classify of the TE type using observation Z of monitoring system channels $\mathbf{Ch}$. An obvious approach to solving this problem is to use the ensemble of LMC ($\mathbf{F}(\theta \mid Z)$). 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.

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

So the classification problem TE is reduced to the task of creating an effective multiclass classificator (MC) which is based on a LMC classifiers ensemble. 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 (Narasimha and Devi, 2011; Barlett et al., 2004). 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) (Smola et al., 2008; Barlett et al., 2004). 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 (Platt et al., 1998). By definition a SVM discriminant function $f_k(\theta \mid \mathbf{z}_k, \vartheta)$ 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 α is a normal vector to the hyperplane, $|b| / \|\alpha\|$ is the perpendicular distance from the hyperplane to the origin, Thus we have $\vartheta = \{\alpha, b\}$, and each tuple ϑ defines the hyperplane in the feature space.

3. 1. Multiple Kernel Learning (MKL)

In contrast to baseline kernels selection ("averaging kernels" and "product kernels" [Jebara, 2004]), 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 $\mathbf{k}\left(Z^{(i)}, Z^{(j)}\right) = \sum_{k=1}^{m} \beta_k \mathrm{k}_k\left(\mathbf{z}_{ki}, \mathbf{z}_{kj}\right)$ with LMCP $\vartheta = \{\alpha, b\}$. Here $Z^{(i)} = \{\mathbf{z}_{1i}, \mathbf{z}_{2i}, ..., \mathbf{z}_{mi}\}$, $Z^{(j)} = \{\mathbf{z}_{1j}, \mathbf{z}_{2j}, ..., \mathbf{z}_{mj}\}$, $\sum_{k=1}^{m} \beta_k = 1, \beta_k \geq 0$. MKL was originally introduced in (Bousquet and Luxburg, 2004). Let us denote $K_i(Z) = \left(\mathrm{k}_i(Z, Z_1), \mathrm{k}_i(Z, Z_2), ..., \mathrm{k}_i(Z, Z_N)\right) \in R^N, i = 1, ...N$. The final decision has form $F_{MKL}(Z) = Arg \underset{\theta \in \Theta}{Max}\left(\sum_{k=1}^{m} \beta_{k,\theta}\left(K_k(Z)^T \alpha_\theta + b_\theta\right)\right)$. The choice of parameters MKL is made by using for each θ the following scheme:

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

$D(\theta, t) = \max(0, 1 - \theta t)$, $(Z_i, \theta_i) \in \mathbf{Z}_T$.

In other words, in MKL case we optimize jointly the convex hull of kernels. Here for each θ we have the same LMCP $\vartheta_\theta = \{\alpha_\theta, b_\theta\}$ for different *k*.

3. 2. LP-Boost (LP-β)

So, we will consider a case when classifiers $f_k(\theta \mid \mathbf{z}_k)$ of ensemble $F\left(\theta \middle| \mathbf{F}(\theta \mid Z)\right)$ are not trained jointly, but coefficients $\{\beta_k\}$ are determined jointly. Here we have a situation where LMCP tuples ϑ_θ are different for different *k*. This method is called the β-LP-Boost (Demiriz and Bennet, 2002), and here the final decision has the form

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

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:

$$\min_{\beta,\xi,\rho}\left(-\rho + \frac{1}{\nu N}\left(\sum_{i=1}^{N} \xi_i\right)\right), \tag{2}$$

Under the condition

$$\sum_{k=1}^{m} \beta_k\left(\mathrm{k}_k(Z, Z_i)\alpha_{\theta',k} + b_{\theta',k}\right) - \arg\max_{\theta' \neq \theta''} \sum_{k=1}^{m} \beta_k\left(\mathrm{k}_k(Z, Z_i)\alpha_{\theta'',k} + b_{\theta'',k}\right) + \xi_i \geq \rho,$$

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

3. 3. LP-Boost (LP-B)

Another version of LP approach to choice $\{\beta_k\}$ was called B-LP-Boost (Nowozin and Gehler, 2009). In this case, each class has its own weight vector. So, we have (m x *D*) weighting matrix B. The final decision has the form

$$F_{LPB}(Z)=Arg\ \underset{\theta\in\Theta}{Max}\left(\sum\nolimits_{k=1}^{m}B_k^{\theta}\left(K_k\left(Z\right)^T\alpha_{\theta,k}+b_{\theta,k}\right)\right) \tag{3}$$

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

$$\min_{\beta,\xi,\rho}\left(-\rho+\frac{1}{\nu N}\left(\sum_{i=1}^{N}\xi_i\right)\right), \tag{4}$$

Under the condition $\begin{array}{l}\sum\nolimits_{k=1}^{m}B_k^{\theta'}\left(\mathrm{k}_k\left(Z,Z_i\right)\alpha_{\theta',k}+b_{\theta',k}\right)-\\ \sum\nolimits_{k=1}^{m}B_k^{\theta''}\left(\mathrm{k}_k\left(Z,Z_i\right)\alpha_{\theta'',k}+b_{\theta'',k}\right)+\xi_i\ge\rho,\end{array}$ $i=1,...,N;\theta'\ne\theta''$,

$\forall\theta,\mathrm{m}:\left(\sum_{k=1}^{m}B_k^{\theta}=1, B_k^{\theta}\ge 0, k=1,..m\right)$. As above, here ξ are slack variables, ν - regularization constant, ν is chosen using CV. Here we have a linear programming problem too, but this problem is more expensive because of dimension increasing.

3. 4. 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. In other words, classifiers with small LC provide the greater generalization ability of classification system: such classifiers have lower complexity to avoid overfitting. Hence, in formula of $F\left(\theta\middle|\mathbf{F}(\theta\mid Z)\right)$ 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 $\mathbf{F}(\theta\mid Z)$ we have the following discriminative function.

$$F_{WILC}(Z)=Arg\ \underset{\theta\in\Theta}{Max}\left(\sum\nolimits_{k=1}^{m}\sigma_{\theta,k}\beta_{k,\theta}\left(K_k\left(Z\right)^T\alpha_{\theta}+b_{\theta}\right)\right). \tag{5}$$

Here $\sigma_{\theta,k} = L_{\theta,k}^{-1}\left(\sum_{j=1}^{m} L_{\theta,j}^{-1}\beta_{j,\theta}\right)^{-1}$, $k=1,...m$, $\sigma_{\theta,k}\beta_{k,\theta} > 0$, $\sum_{k=1}^{m}\sigma_{\theta,k}\beta_{k,\theta} = 1$, $L_{\theta,k}$ 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.

4. Results of Practical Usage

All of above described methods were used for multichannel TE 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 September of 2014, and this system continues to operate. Parameters of the C-OTDR system: a) duration of the probe pulse is 50-200 ns; b) period of probe pulse ~ 50-300 μs; c) laser wavelength - 1550 nm. In this case, the main problem is to fusion of multichannel data to classify the seismoacoustic TE with maximum accuracy. As was said above, for each C-OTDR channel $Ch(\kappa_k)$ are used appropriate D binary classifiers $\mathrm{f}_k(\theta_i \mid \mathbf{z}_k)$, $\theta_i \in \Theta$. Each LMC $\mathrm{f}_k(\theta_i \mid \mathbf{z}_k)$ is binary classifier, which divides the feature space $(\underline{Z}, d)$ into two classes θ_i and $\Theta \setminus \theta_i$. Each LMC $\mathrm{f}_k(\cdot)$ 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 (Gentle and Blimes, 1998). 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 TE) for their robust identification. The GMM of TE feature vectors distribution is a weighted sum of J components densities [Gentle and Blimes, 1998] and given by the equation $P(x|\lambda_s) = \mathbf{w}_s\mathbf{B}_s^T(x)$, where x is a random m-vector, $\mathbf{w}_s = (w_{s1},...w_{sJ}) \in R^J$, $\mathbf{B}_s(x) = (B_{s1}(x),...B_{sJ}(x)) \in R^J$, $\underset{s,i}{\forall} B_{si}(x) = \left((2\pi)^{m/2}|\Sigma_{si}|^{1/2}\right)^{-1}\exp\left(-\frac{1}{2}(x-\mu_{si})^T\Sigma_{si}^{-1}(x-\mu_{si})\right)$, $\lambda_s = \{(w_{si}, \mu_{si}, \Sigma_{si}) \mid i = 1, J\}$. In general, diagonal covariance matrices Σ_{si} are used to limit the model size. The model parameters λ_s characterize a TE in the form of a probabilistic density function. During training, those parameters are determined by the well-known expectation maximization (EM) algorithm (Gentle and Blimes, 1998). In the described experiments value J was equal to 1024. Thus, for identification of TE class, each TE is modelled by a GMM-vector and is referred to as his model parameters $\lambda \in \underline{Z}$. The classic SVM with Bhattacharyya-kernel (Timofeev and Egorov, 2014) was used as the LMC.

Priori defined target classes of TE, which collectively makes up a finite set Θ. For example, in case of railways monitoring the array Θ consists of the following TE classes: "hand digging the soil", "chiseling ground scrap", "pedestrian". Four alternative approaches for multichannel data fusion were compared on stage of TE classification. In particular, MKL, LP-β, LP-B, and WILC-MKL approaches were used. The results of using these methods as parts of the C-OTDR system are presented in Table 1. In the process of using the method WILC-MKL values of Lipschitz Constants were evaluated numerically for each LMC from ensemble $\mathbf{F}(\theta \mid Z)$. The volumes of training sets were equal for each of various data fusion approaches, but those volumes were different for various TE types.

Table 1. The Practical Detection Results.

Method	Type of TE	Accuracy	Volume of training set
MKL	"hand digging the soil"	76%	60
	"chiselling ground scrap"	79%	60
	“pedestrian"	78%	80
LP-β	"hand digging the soil"	81%	60
	"chiselling ground scrap"	83%	60
	“pedestrian"	81%	80
LP-B	"hand digging the soil"	82%	60
	"chiselling ground scrap"	85%	60
	“pedestrian"	79%	80
WILC-MKL	"hand digging the soil"	81%	60
	"chiselling ground scrap"	82%	60
	“pedestrian"	78%	80

Presented results prove that LP (β and B) are more effective with respect to MKL, and WILC-MKL 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.

5. Conclusion

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 as part of a classification subsystem in a C-OTDR monitoring system.

Acknowledgements

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

References

Barlett, P., Lanckriet, G. R. G., Cristianini, N., Ghaoui, L. E., & Jordan, M. I. (2004). Learning the Kernel Matrix with Semidefinite Programming. *JMLR*, *5*, 27–72.

Bach, F. R., Lanckriet, G. R. G., & Jordan, M. I. (2004). Multiple Kernel Learning, Conic Duality, and the SMO Algorithm. In *ICML*.

Luxburg, U., Bousquet, O. (2004). Distance-Based Classification with Lipschitz Functions. *J. of Mach. Learning Research*, *5*, 669-695.

Demiriz, A., Bennett, K. P., & Taylor, S. (2002). Linear Programming Boosting via Column Generation. *Mach. Learning*, *46*(1-3), 225-254.

Blimes, J. A., Gentle, A. (1998). A Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models. *Tech. Rep., Int'l Computer Science Institute*, 97-02.

Jebara, T., Kondor, I., & Howard, A. (2004). Probability Product Kernels. *J. of Mach. Learning Research*, *5*, 819–844.

Narasimha, M., Devi, V. S. (2011). Combination of Classifiers in Pattern Recognition. *Undergraduate Topics in Computer Science* (pp. 188-206). London: Springer.

Gehler, P., Nowozin, S. (2009). On Feature Combination For Multiclass Object Classification. *Proceedings of ICCV*.

Platt, J., Hears, S. T., Dumais, O., & Scholkopf, B. (1998). Support Vector Machines. *IEEE Intelligent Syst.*, *13*(4), 18-28.

Smola, J., Hofmann, T., & Sholkopf, D. (2008). Kernel Methods in Machine Learning. *Anal. of Statistics*, *36*(3), 1171-1220.

Taylor, H. F., Choi, K. N., & Juarez, J. C. (2003). Distributed Fiber-Optic Pressure/Seismic Sensor For Low-Cost Monitoring. *SPIE 5090*, 134-141.

Timofeev, A.V., Egorov, D.V. (2004). Multichannel Classification of Target Signals by Means of an SVM Ensemble in C-OTDR Systems for Remote Monitoring of Extended Objects, *Proc. of Int. Conf. on MVML*.

Timofeev, A.V. (2012). The Guaranteed Estimation of the Lipschitz Classifier Accuracy: Confidence Set Approach. *J. Korean Stat. Soc.*, *41*(1), 105-114.