The Optimization of Decision Rules in Multimodal Decision-Level Fusion Scheme

Andrey Timofeev, Victor Denisov, Dmitry Egorov

***Abstract*—**This paper introduces an original method of parametric optimization of the structure for multimodal decision-level fusion scheme which combines the results of the partial solution of the classification task obtained from assembly of the mono-modal classifiers. As a result, a multimodal fusion classifier which has the minimum value of the total error rate has been obtained.

***Keywords*—**Classification accuracy, fusion solution, total error rate.

I. Introduction

THE process of combining information from multiple sources is known as information fusion [1]-[8]. The fusion could be realized at three different levels: (a) fusion at the feature extraction level, (b) fusion at the matching score level and (c) fusion at the decision level [2]. From a practical point of view the case (c) is most important, since this case is very common. Many practical cases reduce to fusion at the decision level. For example: multimodal biometric fusion, multichannel data fusion in C-OTDR monitoring systems, integral solution on a classifiers ensemble etc. That is why this paper considers the case of fusion at the decision level. So, finding an optimal parametric scheme for the structure of the fusion solution is an important issue. The goal of the optimization is to maximize classification accuracy. The paper presents a solution to this problem.

II. Problem Statement

Assume the following:

- The objects to be classified can only belong to one of two classes. For the sake of mathematical convenience, they are labeled by «+1» and «−1», respectively.
- N is the number of information sources (information modes or modes);
- Each mode i, $i \in \{1,...N\}$, generates the corresponding type of multi-dimensional feature $x^{(i)}$, $x^{(i)} \in X^{(i)}$, here $X^{(i)}$ - feature space of the i-th mode;
- Thus, each object that needs to be classified is described by features of the multi-dimensional parametrical space. This parametrical space consists from N feature parametrical spaces $X^{(i)}, i = 1,...N$.

The essence of the object classification problem is: by analyzing the multidimensional features $x^{(i)}$, $x^{(i)} \in X^{(i)}$, $i \in \{1,...N\}$, classifier has to decide to which of the two classes the object belongs. In the single mode the classification result is a so-called decision function $h_i(\cdot)$ (in other words: i-classifier), here i is the index of the relevant mode. In this case, we deal with the N modes and, therefore, with N classifiers: $\mathbf{h}(\cdot) = \{h_i(\cdot) | i = 1, 2, ...N\}$. The set of the classifiers $\mathbf{h}(\cdot)$ is available for construction of the decision fusion function (DFF). Sometimes, this set is called a classifiers assembly. This study considers a widespread case when the DFF is a convex hull of the functions $h_i(\cdot)$, $i \in \{1,2,...N\}$ [2], [6]. This type of the DFF is called as "decision-level fusion scheme" [2]. So, we have the N feature spaces $X^{(i)}, i = 1,...N$ and classifiers $h_i(x^{(i)}), i = 1,...N$ each of which maps the corresponding feature vector $x^{(i)} \in X^{(i)}$ to a class label space $\mathbf{Y} = \{-1, 1\}$. In other words, each classifier $h_i(x^{(i)}), i = 1,...N$, indicates to which of the two classes the vector $x^{(i)} \in X^{(i)}$ corresponds. The following entry: $y_i = h_i(x^{(i)}), \quad i = 1,...N; y_i \in \mathbf{Y}$ - is admissible. At the same time, each of classifiers $h_i(x^{(i)}), i = 1,...N$, depends on the corresponding vector of the parameters $\delta^{(i)} \in \Delta^{(i)}$. Thus, we have: $h_i \equiv h_i\left(x^{(i)} \mid \delta^{(i)}\right)$. For each $i = 1,...N$ a training set $\lambda^{(i)}$ consists of $m^{(i)}$ samples whose associated labels are observed. Thus $\lambda^{(i)} = \left\{\left(x_j^{(i)}, y_j \mid j = 1, m^{(i)}\right)\right\}$, $i = 1,...N$. It is obvious the labels of a test samples are unknown and need to be defined during the test.

Let us denote:

- $\mathbf{X} = X^{(1)} \otimes X^{(2)} ... \otimes X^{(N)}$ - common feature space;
- $\mathbf{X} \setminus X^{(k)} = X^{(1)} \otimes X^{(2)} ... \otimes X^{(k-1)} \otimes X^{(k+1)} ... \otimes X^{(N)}$;
- $\boldsymbol{\delta} = \left(\delta^{(1)}, \delta^{(2)}, ..., \delta^{(N)}\right) \in \Delta^{(1)} \otimes \Delta^{2} \otimes ... \otimes \Delta^{(N)} = \boldsymbol{\Delta}$;
- $\bar{x} = \left(x^{(1)}, x^{(2)}, ...x^{(N)}\right) \in \mathbf{X}$;
- $\bar{x} \setminus x^{(k)} = \left(x^{(1)}, x^{(2)}, .., x^{(k-1)}, x^{(k+1)} ... x^{(N)}\right) \in \mathbf{X} \setminus X^{(k)}$;
- $\mathbf{h}\left(\bar{x} | \boldsymbol{\delta}\right) = \left(h_1\left(x^{(1)} \mid \delta^{(1)}\right), ..., h_N\left(x^{(N)} \mid \delta^{(N)}\right)\right) \in R^N$;

- $\boldsymbol{\alpha}^*(\boldsymbol{\delta}) = \left(\alpha_1^*\left(\delta^{(1)}\right), \alpha_2^*\left(\delta^{(2)}\right), ..., \alpha_N^*\left(\delta^{(N)}\right)\right)$;
- $h_k\left(x^{(k)} \middle| \bullet\right) = h_k\left(x^{(k)} \mid \delta^{(k)}\right)$;
- event: $\varpi_i(\delta^{(i)}) : \left\{y \neq h_i\left(x^{(i)} \mid \delta^{(i)}\right)\right\}$;
- event: $\omega_i(\delta^{(i)}) : \left\{y = h_i\left(x^{(i)} \mid \delta^{(i)}\right)\right\}$;
- $\varepsilon(k \mid \delta^{(k)}) = \mathbf{E}_{x^{(k)} \sim X^{(k)}}\left(\varpi_k\left(\delta^{(k)}\right)\right)$ is the average **total error** for the k-th classifier;
- $A(i \neq k) = \mathbf{E}_{\bar{x} \setminus x^{(k)} \sim \mathbf{X} \setminus X^{(k)}}\left[\mathbf{E}_y\left\{\left\langle \prod_{i \neq k}^{N} e^{-y\alpha_i h_i\left(x^{(i)} \middle| \bullet\right)} \right\rangle \middle| \bar{x} \setminus x^{(k)}\right\}\right]$;
- $\mathbf{1}_E(\omega)$ is the indicator function of the event ω.

So, we consider:

- both training and test samples are drawn *i.i.d.* from underlying distribution $\boldsymbol{\Lambda}$;
- the DFF $H(\bar{x} \mid \boldsymbol{\delta}) = H\left(h_1\left(x^{(1)} \mid \delta^{(1)}\right), h_2\left(x^{(2)} \mid \delta^{(2)}\right), ...\right)$ is a mapping from $\mathbf{X}$ to $\mathbf{Y}$ after training on a data set $\lambda^{(i)}$ $i = 1, ...N$;
- the DFF $H(\bar{x} \mid \boldsymbol{\delta})$ has the following form: $H(\bar{x} \mid \boldsymbol{\delta}) = \boldsymbol{\alpha}\mathbf{h}^T\left(\bar{x} \middle| \boldsymbol{\delta}\right)$;
- Using DFF, the classification procedure is realized by the following simple rule: $y(\bar{x} \mid \boldsymbol{\delta}) = SIGN\left(H(\bar{x} \mid \boldsymbol{\delta})\right)$.

Our goal is to build the DFF $H(\bar{x} \mid \boldsymbol{\delta})$ that would minimize the posterior expected loss

$$\mathbf{E}_{\bar{x} \sim \mathbf{X}, y}\left(L\left(y, \boldsymbol{\alpha}\mathbf{h}^T\left(\bar{x} \middle| \boldsymbol{\delta}\right)\right)\right).$$

Here $L(\cdot)$ is a convex loss function. Thus with fixed $\boldsymbol{\delta} \in \boldsymbol{\Delta}$ we have to solve the following optimization task:

$$\boldsymbol{\alpha}^*\left(\boldsymbol{\delta}\right) = Arg \inf_{\boldsymbol{\alpha}}\left(\mathbf{E}_{\bar{x} \sim \mathbf{X}, y}\left(L\left(y, \boldsymbol{\alpha}\mathbf{h}^T\left(\bar{x} \middle| \boldsymbol{\delta}\right)\right)\right)\right) \quad (1)$$

It is obvious in this case that DFF $H^*(\bar{x} \mid \boldsymbol{\delta}) = \boldsymbol{\alpha}^*(\boldsymbol{\delta})\mathbf{h}^T\left(\bar{x} \middle| \boldsymbol{\delta}\right)$ will fully meet the requirements of the problem statement.

III. Solution Method

The following statement is obvious:

$$\frac{\partial \mathbf{E}_{\bar{x} \sim \mathbf{X}, y}\left(L\left(y, \boldsymbol{\alpha}\mathbf{h}^T\left(\bar{x} \middle| \boldsymbol{\delta}\right)\right)\right)}{\partial \boldsymbol{\alpha}} = \left(\frac{\partial \mathbf{E}_{\bar{x} \sim \mathbf{X}, y}\left(L\left(y, \boldsymbol{\alpha}\mathbf{h}^T\left(\bar{x} \middle| \boldsymbol{\delta}\right)\right)\right)}{\partial \alpha_1}, \frac{\partial \mathbf{E}_{\bar{x} \sim \mathbf{X}, y}\left(L\left(y, \boldsymbol{\alpha}\mathbf{h}^T\left(\bar{x} \middle| \boldsymbol{\delta}\right)\right)\right)}{\partial \alpha_2}, ...\right) = 0.$$

Thus, we have the system of the non-linear equations:

$$\frac{\partial \mathbf{E}_{\bar{x} \sim \mathbf{X}, y}\left(L\left(y, \boldsymbol{\alpha}\mathbf{h}^T\left(\bar{x} \middle| \boldsymbol{\delta}\right)\right)\right)}{\partial \alpha_i} = 0, i = 1, ...N \quad (2)$$

This system must be solved relative to the variables $\alpha_1, \alpha_2, ..., \alpha_N$ with the chosen type of the loss function $L(\cdot)$. Thus we have the solution of the (2) with some $\boldsymbol{\alpha}^*(L) = (\alpha_1^*(L), \alpha_2^*(L), ..., \alpha_N^*(L))$. It is obvious that the obtained vector $\boldsymbol{\alpha}^*$ to (2) yields the solution of (1).

Now consider the choice of a loss function $L(\cdot)$. There are many types of the convex loss functions:

- $L(x) = \max(1 - x, 0)$
- $L(x) = \exp(-x)$
- $L(x) = \log(1 + \exp(-x))$,

and other. Let us use the $L(x) = \exp(-x)$ loss function which has the excellent analytical characteristics. In additional, as a loss function argument we will use the following product $y\boldsymbol{\alpha}\mathbf{h}^T\left(\bar{x} \middle| \boldsymbol{\delta}\right)$. There the $y\boldsymbol{\alpha}\mathbf{h}^T\left(\bar{x} \middle| \boldsymbol{\delta}\right)$ is called as the classification margin of the hypothesis $\mathbf{h}^T\left(\bar{x} \middle| \boldsymbol{\delta}\right)$. In this case we can write $L\left(y\boldsymbol{\alpha}\mathbf{h}^T\left(\bar{x} \middle| \boldsymbol{\delta}\right)\right) = \exp\left(-y\boldsymbol{\alpha}\mathbf{h}^T\left(\bar{x} \middle| \boldsymbol{\delta}\right)\right)$. The point wise loss, which can be decomposed to each instance $\bar{x}$, is $\mathbf{E}_y\left(\exp\left(-y\boldsymbol{\alpha}\mathbf{h}^T\left(\bar{x} \middle| \boldsymbol{\delta}\right)\right) \middle| \bar{x}\right)$.

Since y and $h_i\left(x^{(i)} \mid \delta^{(i)}\right)$ have to be +1 or -1, assuming the independence of the $h_i\left(x^{(i)} \mid \delta^{(i)}\right), x^{(i)}$ $(i = 1, ...N)$, the following expansion is valid for any $k \in \{1, 2, ...N\}$:

$$\frac{\partial \mathbf{E}_{\bar{x} \sim \mathbf{X}}\left[\mathbf{E}_y\left(\exp\left(-\sum_{i=1}^{N} y\alpha_i h_i\left(x^{(i)} \middle| \bullet\right)\right) \middle| \bar{x}\right)\right]}{\partial \alpha_k} =$$

$$\mathbf{E}_{\bar{x} \sim \mathbf{X}}\left[\mathbf{E}_y\left(\frac{\partial \exp\left(-\sum_{i=1}^{N} y\alpha_i h_i\left(x^{(i)} \middle| \bullet\right)\right)}{\partial \alpha_k} \middle| \bar{x}\right)\right] =$$

$$\mathbf{E}_{\bar{x} \sim \mathbf{X}}\left[\mathbf{E}_y\left(\left(-yh_k\left(x^{(k)} \middle| \bullet\right)\right)\exp\left(-\sum_{i=1}^{N} y\alpha_i h_i\left(x^{(i)} \middle| \bullet\right)\right) \middle| \bar{x}\right)\right] =$$

$$\mathbf{E}_{\bar{x} \sim \mathbf{X}}\left[\mathbf{E}_y\left(\left(-yh_k\left(x^{(k)} \middle| \bullet\right)\right)\prod_{i=1}^{N}\exp\left(-y\alpha_i h_i\left(x^{(i)} \middle| \bullet\right)\right) \middle| \bar{x}\right)\right] =$$

$$\mathbf{E}_{\bar{x} \sim \mathbf{X}}\left[\mathbf{E}_y\left\{\left(-yh_k\left(x^{(k)} \middle| \bullet\right)\right)\prod_{i=1}^{N}\left\langle e^{-y\alpha_i h_i\left(x^{(i)} \middle| \bullet\right)}\mathbf{1}_E\left(\omega_i\left(\delta^{(i)}\right)\right) + e^{-y\alpha_i h_i\left(x^{(i)} \middle| \bullet\right)}\mathbf{1}_E\left(\varpi_i\left(\delta^{(i)}\right)\right)\right\rangle \middle| \bar{x}\right\}\right] =$$

$$\mathbf{E}_{\bar{x} \sim \mathbf{X}}\left[\mathbf{E}_y\left\{\left(-yh_k\left(x^{(k)} \middle| \bullet\right)\right)\prod_{i=1}^{N}\left\langle e^{-\alpha_i}\mathbf{1}_E\left(\omega_i\left(\delta^{(i)}\right)\right) + e^{\alpha_i}\mathbf{1}_E\left(\varpi_i\left(\delta^{(i)}\right)\right)\right\rangle \middle| \bar{x}\right\}\right] =$$

$$\mathbf{E}_{\bar{x}\sim\mathbf{X}}\left[\mathbf{E}_y\left\{\left\langle\prod_{i\neq k}^{N}\left(e^{-\alpha_i}\mathbf{1}_E\left(\omega_i\left(\delta^{(i)}\right)\right)+e^{\alpha_i}\mathbf{1}_E\left(\varpi_i\left(\delta^{(i)}\right)\right)\right)\right\rangle\cdot\right.\right.$$
$$\left.\left.\left\langle\left(-yh_k\left(x^{(k)}|\bullet\right)\right)\cdot\left(e^{-\alpha_k}\mathbf{1}_E\left(\omega_k\left(\delta^{(k)}\right)\right)+e^{\alpha_k}\mathbf{1}_E\left(\varpi_k\left(\delta^{(k)}\right)\right)\right)\right\rangle\middle|\bar{x}\right\}\right]=$$
$$\mathbf{E}_{\bar{x}\sim\mathbf{X}}\left[\mathbf{E}_y\left\{\mathbf{E}_y\left\{\left\langle\prod_{i\neq k}^{N}e^{-y\alpha_i h_i\left(x^{(i)}|\bullet\right)}\right\rangle\cdot\right.\right.\right.$$
$$\left.\left.\left.\left\langle\left(-yh_k\left(x^{(k)}|\bullet\right)\right)e^{-\alpha_k}\mathbf{1}_E\left(\omega_k\left(\delta^{(k)}\right)\right)+\left(-yh_k\left(x^{(k)}|\bullet\right)\right)e^{\alpha_k}\mathbf{1}_E\left(\varpi_k\left(\delta^{(k)}\right)\right)\right\rangle\middle|\bar{x}\right\}\right]=$$

$$\mathbf{E}_{\bar{x}\sim\mathbf{X}}\left[\mathbf{E}_y\left\{\left\langle\prod_{i\neq k}^{N}e^{-y\alpha_i h_i\left(x^{(i)}|\bullet\right)}\right\rangle\cdot\left\langle-e^{-\alpha_k}\mathbf{1}_E\left(\omega_k\left(\delta^{(k)}\right)\right)+e^{\alpha_k}\mathbf{1}_E\left(\varpi_k\left(\delta^{(k)}\right)\right)\right\rangle\middle|\bar{x}\right\}\right]=$$

$$\mathbf{E}_{\bar{x}\setminus x^{(k)}\sim\mathbf{X}\setminus X^{(k)}}\left[\mathbf{E}_y\left\{\left\langle\prod_{i\neq k}^{N}e^{-y\alpha_i h_i\left(x^{(i)}|\bullet\right)}\right\rangle\middle|\bar{x}\setminus x^{(k)}\right\}\right]\cdot$$
$$\mathbf{E}_{x^{(k)}\sim X^{(k)}}\left[\mathbf{E}_y\left\{\left(-e^{-\alpha_k}\mathbf{1}_E\left(\omega_k\left(\delta^{(k)}\right)\right)+e^{\alpha_k}\mathbf{1}_E\left(\varpi_k\left(\delta^{(k)}\right)\right)\right)\middle|x^{(k)}\right\}\right].$$

Following (2) we have

$$\mathbf{E}_{\bar{x}\setminus x^{(k)}\sim\mathbf{X}\setminus X^{(k)}}\left[\mathbf{E}_y\left\{\left\langle\prod_{i\neq k}^{N}e^{-y\alpha_i h_i\left(x^{(i)}|\bullet\right)}\right\rangle\middle|\bar{x}\setminus x^{(k)}\right\}\right]\cdot$$
$$\mathbf{E}_{x^{(k)}\sim X^{(k)}}\left[\mathbf{E}_y\left\{\left(-e^{-\alpha_k}\mathbf{1}_E\left(\omega_k\left(\delta^{(k)}\right)\right)+e^{\alpha_k}\mathbf{1}_E\left(\varpi_k\left(\delta^{(k)}\right)\right)\right)\middle|x^{(k)}\right\}\right]=$$
$$A(i\neq k)\mathbf{E}_{x^{(k)}\sim X^{(k)}}\left[\mathbf{E}_y\left\{\left(-e^{-\alpha_k}\mathbf{1}_E\left(\omega_k\left(\delta^{(k)}\right)\right)+e^{\alpha_k}\mathbf{1}_E\left(\varpi_k\left(\delta^{(k)}\right)\right)\right)\middle|x^{(k)}\right\}\right]=0.$$

Further

$$A(i\neq k)\mathbf{E}_{x^{(k)}\sim X^{(k)}}\left[-e^{-\alpha_k}\mathbf{P}(\omega_k\mid x^{(k)},\delta^{(k)})+e^{\alpha_k}\mathbf{P}(\varpi_k\mid x^{(k)},\delta^{(k)})\right]=$$
$$A(i\neq k)\left(\mathbf{E}_{x^{(k)}\sim X^{(k)}}\left[-e^{-\alpha_k}\mathbf{P}(\omega_k\mid x^{(k)},\delta^{(k)})\right]+\right.$$
$$\left.\mathbf{E}_{x^{(k)}\sim X^{(k)}}\left[e^{\alpha_k}\mathbf{P}(\varpi_k\mid x^{(k)},\delta^{(k)})\right]\right)=A(i\neq k)\cdot$$
$$\left(\mathbf{E}_{x^{(k)}\sim X^{(k)}}\left[-e^{-\alpha_k}\mathbf{P}(\omega_k\mid x^{(k)},\delta^{(k)})\right]+\mathbf{E}_{x^{(k)}\sim X^{(k)}}\left[e^{\alpha_k}\mathbf{P}(\varpi_k\mid x^{(k)},\delta^{(k)})\right]\right)=$$
$$A(i\neq k)\left(-e^{-\alpha_k}\cdot\left(1-\varepsilon(k\mid\delta^{(k)})\right)+e^{\alpha_k}\varepsilon(k\mid\delta^{(k)})\right)=0.$$

and it easy to see that

$$\alpha_k^*\left(\delta^{(k)}\right)=0.5\ln\left(\left(1-\varepsilon(k\mid\delta^{(k)})\right)/\varepsilon(k\mid\delta^{(k)})\right).$$

Thus

$$\boldsymbol{\alpha}^*\left(\boldsymbol{\delta}\right)=\left(\frac{1}{2}\ln\left(\frac{1-\varepsilon(1\mid\delta^{(1)})}{\varepsilon(1\mid\delta^{(1)})}\right),...,\frac{1}{2}\ln\left(\frac{1-\varepsilon(N\mid\delta^{(N)})}{\varepsilon(N\mid\delta^{(N)})}\right)\right).$$

and with fixed $\boldsymbol{\delta}$ we have

$$H^*(\bar{x}\mid\boldsymbol{\delta})=\boldsymbol{\alpha}^*(\boldsymbol{\delta})\mathbf{h}^T\left(\bar{x}\middle|\boldsymbol{\delta}\right). \qquad (3)$$

Thus DFF $H^*(\bar{x}\mid\boldsymbol{\delta})$ is optimal in the sense of minimum of the posterior expected loss $\mathbf{E}_{\bar{x}\sim\mathbf{X},y}\left(L(y,H(\bar{x}\mid\boldsymbol{\delta}))\right)$ with $L(x)=\exp(-x)$ and fixed $\boldsymbol{\delta}$.

IV. Practical Example

A good example of a practical use of proposed strategy is a task of multichannel data fusion in C-OTDR monitoring systems (OXY).This system has been designed to monitor of railways (detection of technological activity on railroad tracks).System OXY was installed at the railroad testing area of Kazakhstan Railways Company (JSK "Kazakh Temir Zholy"). Distributed fiberoptic sensor (DFOS) of OXY was buried at the distance of 5 m from railways, at the depth of 30-50 cm. For classification *N*of C-OTDR channels were used only. Value of *N* have been determined empirically, *N*=11. Channel with number 6 has been closest to the source of seismoacoustic emission; channels with numbers 1 and 11 have been the most distant to the source of seismoacoustic emission. Parameters of the C-OTDR monitoring system:

- the probe pulse duration – 10..100 ns;
- frequency sensing – 3..5 kHz;
- the probe signal power - 15 mW;
- DFOS length – 1 200 m;
- laser wavelength - 1550 nm.

In this case, we have problem of optimal fusion for ensemble of single-mode classifiers $\left\{h_i\left(x^{(i)}\right)\right\}$, each of them is correspond to relevant C-OTDR channel (Fig. 1). So, in other form, we can write:

$$\mathbf{h}\left(\bar{x}\middle|\boldsymbol{\delta}\right)=\left(h_1\left(x^{(1)}\right),..,h_i\left(x^{(i)}\right),.,h_N\left(x^{(N)}\right)\right)\in R^N,$$

where i is C-OTDR channel number, the $x^{(i)}$ is i-th channel output. There are objects of two classes:

- Technological activityon railroad tracks: label is "1"
- Common activity on railroad tracks: label is "-1"

By analyzing the multidimensional features $x^{(i)}$, $x^{(i)}\in X^{(i)}$ $i\in\{1,...N\}$, classifier $h_i\left(x^{(i)}\mid\delta^{(i)}\right)$ (i-thC-OTDR channel) has to decide to which of the two classes the object belongs. So, we have the N feature spaces $X^{(i)},i=1,...N$ and classifiers $h_i(x^{(i)}),i=1,...N$ each of which maps the corresponding feature vector $x^{(i)}\in X^{(i)}$ to a class label space $\mathbf{Y}=\{-1,1\}$.

We denote DDF, which has been calculated according to (3), as DDF*. For $\boldsymbol{\alpha}=\left(N^{-1},...,N^{-1}\right)$ the DDF is $H^+(\bar{x}\mid\boldsymbol{\delta})=\boldsymbol{\alpha}\mathbf{h}^T\left(\bar{x}\right)$, and we have denoted this DDF as DDF$^+$. In additional, classification results of "marginal" single-mode classifiers were investigated: classifier of channel 1 $h_1\left(x^{(1)}\right)$ (the most distant) and classifier of channel 6 $h_6\left(x^{(6)}\right)$ (the closest). Table I contains results of practical experiments.

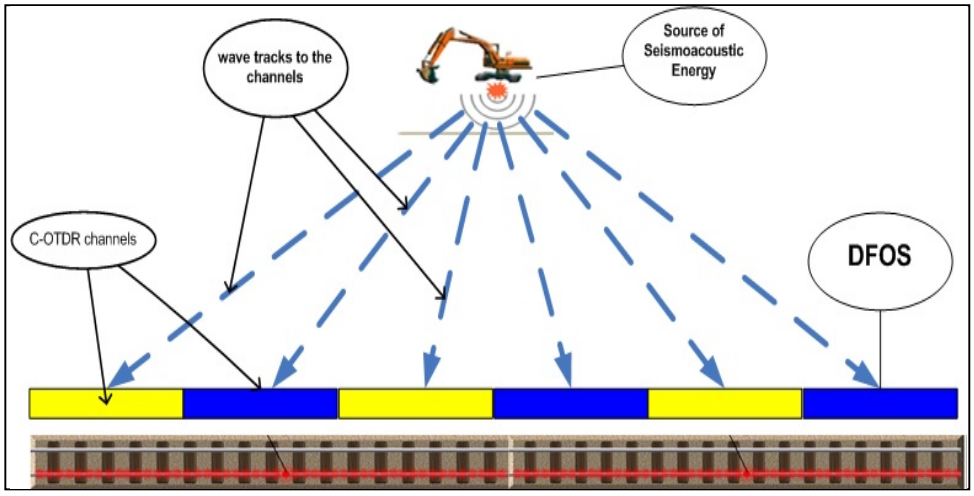

Fig. 1 C-OTDR Channels

TABLE I
THE PRACTICAL RESULTS

Classifier	Total Error (%)
DDF*	0,03
DDF^{+}	0,06
$h_1\left(x^{(1)}\right)$	0,12
$h_6\left(x^{(6)}\right)$	0,07

Those results are enough sufficient for practical usage, and they are well interpreted. Indeed, the best classification quality has the DDF*. On the other hand, the worst classification quality has the classifier $h_1\left(x^{(1)}\right)$, which is the most distant to the source of seismoacoustic emission, hence, in relevant channel we had the lowest signal to noise ratio.

V. CONCLUSION

In this paper we consider a new method of determination an optimal structure of the fusion solution, which combines the results of the partial solution of the classification task, obtained from corresponding single-mode classifiers. By the term "optimization" we mean the minimum of the total error rate. The suggested method allows finding the values of the parameters of the fusion solution structure which provide the best classification accuracy of the fusion classifier. Results of the practical usage have shown a good performance of suggested method. This method was developed for use in multichannel C-OTDR monitoring system when for obtain the classification solution we need to combine data from various C-OTDR channels.

Republic of Kazakhstan.

REFERENCES

[1] Dave L. Hall and James Linas, “Introduction to Multisensor Data Fusion”, Proc. of IEEE, Vol. 85, No. 1, pp. 6 – 23, Jan 1997.
[2] ISO/IEC JTC 1/SC 37 N 1506, Biometrics, 2006-02-28.
[3] Erik Blasch, Ivan Kadar, John Salerno, Mieczyslaw Kokar, Subrata Das, Gerald Powell, Daniel Corkill, and E. Euspini, “Issues and Challenges in Situation Assessment (Level 2 Fusion)”, Journal of Advances in Information Fusion, Vol 1, No 2, Dec. (2006).
[4] Liggins, Martin E., David L. Hall, and James Llinas. “Multisensor Data Fusion, Second Edition Theory and Practice (Multisensor Data Fusion)”. CRC, (2008).
[5] David L. Hall, Sonya A. H. McMullen, “Mathematical Techniques in Multisensor Data Fusion”, Artech House (2004)
[6] H. B. Mitchell, “Multi-sensor Data Fusion – An Introduction” Springer-Verlag, Berlin 2007)
[7] A. C. Kak, Su-Shing, Spatial reasoning and multi-sensor fusion: proceedings of the 1987 workshop, American Association for Artificial Intelligence, 1987: Saint Charles III.
[8] L., Xu, A., Kryzak, C.Y., Suen, "Methods of Combining Multiple Classifiers and Their Application to Handwriting Recognition", IEEE Trans. on Systems, Man and Cyber.,1992, vol. 22, no. 3, pp. 418-435.