Maciej HUK
Michał SZCZEPANIK

MULTIPLE CLASSIFIER ERROR PROBABILITY FOR MULTI-CLASS PROBLEMS

PRAWDOPODOBIEŃSTWO BŁĘDU KLASYFIKATORÓW ZŁOŻONYCH DLA PROBLEMÓW WIELOKLASOWYCH*

In this paper we consider majority voting of multiple classifiers systems in the case of two-valued decision support for many-class problem. Using an explicit representation of the classification error probability for ensemble binomial voting and two class problem, we obtain general equation for classification error probability for the case under consideration. Thus we are extending theoretical analysis of the given subject initially performed for the two class problem by Hassen and Salamon and still used by Kuncheva and other researchers. This allows us to observe important dependence of maximal posterior error probability of base classifier allowable for building multiple classifiers from the number of considered classes. This indicates the possibility of improving the performance of multiple classifiers for multiclass problems, which may have important implications for their future applications in many fields of science and industry, including the problems of machines diagnostic and systems reliability testing.

Keywords: *multiple classifiers, majority voting, multi-class problems.*

W niniejszym artykule rozważamy systemy złożonych klasyfikatorów z głosowaniem większościowym dla przypadku problemów wieloklasowych, wykorzystujące wielowartościowe klasyfikatory bazowe. Stosując bezpośrednią reprezentację prawdopodobieństwa błędnej klasyfikacji dla analogicznych systemów w problemach dwuklasowych, otrzymujemy ogólny wzór na prawdopodobieństwo błędu klasyfikacji w przypadku wieloklasowym. Tym samym rozszerzamy teoretyczne analizy tego zagadnienia pierwotnie przeprowadzone dla problemów dwuklasowych przez Hansena i Salomona i ciagle wykorzystywane przez Kunchevę i innych badaczy. Pozwala nam to zaobserwować istotną zależność maksymalnego dopuszczalnego poziomu prawdopodobieństwa błędów klasyfikatorów bazowych od liczby rozważanych przez nie klas. Wskazuje to na możliwość poprawy parametrów klasyfikatorów złożonych dla problemów wieloklasowych, co może mieć niebagatelne znaczenie dla dalszych ich zastosowań w licznych dziedzinach nauki i przemysłu, z uwzględnieniem zagadnień diagnostyki maszyn oraz badania niezawodności systemów.

Słowa kluczowe: *klasyfikatory złożone, głosowanie większościowe, problemy wieloklasowe.*

1. Introduction

Multiple classifiers systems, also known as ensembles or committees, were considered in many papers [5, 10, 13, 21, 23, 29, 34] and books [6, 8, 12, 18]. Committee approaches that learn and retain multiple hypotheses and combine their decisions during classification [3, 7] are frequently regarded as one of the major advances in inductive learning in the past decade [2, 12, 19, 20, 27]. In the effect, the ensemble methodology has been used to improve the predictive performance of single models, in many fields such as: finance [22], bioinformatics [32], medicine [24], manufacturing [28], geography [4], information security [16, 25], information retrieval [10] and recommender systems [17]. On this basis many solutions were proposed to the problems of machines and electronic systems diagnostic [31, 35] as well as testing systems reliability [14, 30]. Solutions of this type can be a valuable complement to other, previously used approaches [26, 33, 36].

In the present paper we extend theoretical analysis of the ensemble classification error probability initially performed for the two class problem by Hassen and Salamon [15] and still used by Kuncheva and other researchers [18-20, 29]. We consider the general case of multi-class classification problems for ensembles using classical majority voting. We will derive general formula for multiple classifier error probability for number of classes greater than two and for any number of base classifiers with mutually equal posterior error probabilities. In the process of this we also show, what is often omitted, how the well known formula for multiple classifier error probability for two-class problems is changing when the number of base classifiers is not restricted to odd values. Analysis of the results obtained indicate the possibility of using multivalue base classifiers to improve the performance of ensembles of classifiers, even for very difficult classification problems.

2. Multiple classifier error probability for two-class problems

Let $D=\{D_1,...D_L\}$ be a set of L classifiers such that $D_i : \Re^n \rightarrow \Omega$, where $\Omega=\{\omega_1,.., \omega_C\}$, assigning class label $\omega_j \in \Omega$ to input data vector $\mathbf{x} \in \Re^n$. It is assumed that classifiers from set D can be successfully used to form ensemble, if their mutual errors are uncorrelated or negatively correlated [1] and when for each base classifier D_i its posterior error probability P_S^i is less than 0.5. In the case of two-class problems (K=2) with use of the majority voting the situation is relatively easy and the ensemble error probability P_E of multiple classifier is then often presented to be:

(*) Tekst artykułu w polskiej wersji językowej dostępny w elektronicznym wydaniu kwartalnika na stronie ***www.ein.org.pl***

$$P_E = \sum_{j=j_0}^{L} \binom{L}{j} P_S^j [1-P_S]^{L-j} \tag{1}$$

where L is odd, all classifiers have equal posterior error probability P_S and initial value j_0 is the minimal number of classifiers giving wrong answer that leads to ensemble decision error.

But it should be remembered, that for many-class problems limiting the number of base classifiers L to odd values does not eliminate the possibility that base classifiers will draw. In such case the solution of random class label selection is often used - when no other class gains higher number of votes than the proper one but some of other classes tie with it, class label is randomly selected from this group, with equal posterior probabilities for each class. With this in mind the factor of ensemble error probability connected with ties can't be neglected. Thus looking for the guideline for further analysis of multi-class problems, we can omit the assumption that L is odd and extended the expression (1) to the form:

$$P_E = \sum_{j=j_0}^{L} \binom{L}{j} P_S^j [1-P_S]^{L-j} + \frac{1}{2}\delta(L \bmod 2, 0)\binom{L}{\frac{L}{2}} P_S^{L/2}[1-P_S]^{L/2} \tag{2}$$

where

$$j_0 = \begin{cases} (L+1)/2 & : L \bmod 2 > 0 \\ L/2+1 & : L \bmod 2 = 0 \end{cases} \tag{3}$$

and $\delta(x,y)$ is the Kronecker's delta:

$$\delta(x,y) = \begin{cases} 1 & : x = y \\ 0 & : x \neq y \end{cases} \tag{4}$$

The factor ½ before the Kronecker's delta in (2) is the probability of wrong random class selection when base classifiers draw and the Newton symbol $\binom{L}{\frac{L}{2}}$ determine the number of possible ties between base classifiers for two-class problem, when L is even.

3. Multiple classifier error probability for multi-class problems

The first step to find the general equation for multiple classifier error probability for multiclass problems can be rewriting the expression (2) to the form in which each component probability is explicite connected with votes assigned by base classifiers to individual classes.. Beacause without loosing the generality we can assume that the class with index 1 is the correct one, thus by simple algebraic transformations we can see that right side of (1) can take the form:

$$\sum_{k_1=0}^{L}\sum_{k_2=0}^{L}\left(\binom{L}{k_2} P_S^{k_2}\left(1-P_S\right)^{L-k_2} \cdot \delta(k_1+k_2, L) H(k_2-k_1)\right) \tag{5}$$

where k_1 and k_2 represent various numbers of votes that can be given by L base classifiers respectively for classes 1 and 2. The introduced Kronecker's delta ensures that only those combinations of votes are taken under consideration, for which the sum of votes for all classes equals the number of base classifiers:

$$k_1 + k_2 = L \tag{6}$$

and H is the Heaviside's step function used to select factors for which $k_2 > k_1$:

$$H(x) = \begin{cases} 1 & : x > 0 \\ 0 & : x \leq 0 \end{cases} \tag{7}$$

Finally, by further use of (6) for calculation of $L - k_2$, and by introducing that:

$$P_1 = 1 - P_S \quad \text{and} \quad P_2 = P_S \tag{8}$$

are probabilities of voting at the class 1 and 2 respectively, we can rewrite (5) in the form:

$$\sum_{k_1=0}^{L}\sum_{k_2=0}^{L}\left(\frac{L!}{k_1!k_2!} P_1^{k_1} P_2^{k_2} \delta(k_1+k_2, L) H(k_2-k_1)\right) \tag{9}$$

Similarly, the right part of the right side of expression (2) can be transformed to:

$$\sum_{k_2=0}^{L}\left(\frac{1}{2}\binom{L}{k_2} P_S^{k_2}[1-P_S]^{L-k_2}\delta(k_2, \frac{L}{2})\right) \tag{10}$$

Next, because in the case of a tie $k_1 = k_2 = L/2$, formula (10) can be rewritten as:

$$\sum_{k_1=0}^{L}\sum_{k_2=0}^{L}\left(\frac{1}{2}\delta(k_1+k_2, L)\delta(k_1,k_2)\frac{L!}{k_1!k_2!} P_1^{k_1} P_2^{k_2}\right) \tag{11}$$

In the result, after combining (5) and (11) and reorganizing, the formula for ensemble error probability for two-class problem (2) can be given by:

$$P_E = \sum_{k_1=0}^{L}\sum_{k_2=0}^{L}\left(\delta(k_1+k_2, L)\left[H(k_2-k_1) + \frac{1}{2}\delta(k_1,k_2)\right]\frac{L!}{k_1!k_2!} P_1^{k_1} P_2^{k_2}\right) \tag{12}$$

The expression (12) shows the natural method of determining the ensemble error probability for multi-class problems (K>2) – by adding further summations connected with other classes. It is easy to notice, that in such case only the part of (12) taken in square brackets require special analysis. The Heaviside's function gives information if the proper class received fewer votes than the wrong class. Thus for many classes it should be replaced by the form:

$$H_E = H\left(\sum_{i=2}^{K} H(k_i - k_1)\right) \tag{13}$$

which has value 1 if one or more classes received more votes form base classifiers than the correct class and zero in other cases. The second, right part in square brackets in (12) - the Kronecker's delta - can be identified as an element holding the number of classes that tie with the correct one, additionally multiplied by the probability of wrong random class selection. In the general case (K>2) the number of ties can be represented by the formula:

$$H_D = \sum_{i=2}^{K} \delta(k_1, k_i) \tag{14}$$

and due to that the probability of wrong random class selection during tie is given by:

$$\frac{H_D}{H_D + 1} \tag{15}$$

Now it is easy to calculate that the ensemble error probability for multi-class problems is given by:

$$P_E = \sum_{k_1=0}^{L}\sum_{k_2=0}^{L}\cdots\sum_{k_K=0}^{L}\left(\delta\left(\sum_{i=1}^{K} k_i, L\right)\left[H_E + (1-H_E)\left(1 - \frac{1}{1+H_D}\right)\right] L!\prod_{i=1}^{K}\frac{P_i^{k_i}}{k_i!}\right) \tag{16}$$

where the sum of the probabilities of assigning votes for each class:

$$\sum_{i=1}^{K} P_i = 1 \tag{17}$$

But it is noteworthy that factor:

$$L!\prod_{i=1}^{K}\frac{1}{k_i!} \tag{18}$$

is a multinomial coefficient P_{MF} of the multinomial probability distribution, thus the expression (16) can be written finally as:

$$P_E = \sum_{k_1=0}^{L}\sum_{k_2=0}^{L}\cdots\sum_{k_K=0}^{L}\left(P_{MF}\left[H_E + (1-H_E)\left(1-\frac{1}{1+H_D}\right)\right]\right) \tag{19}$$

where:

$$P_{MF} = f(k_1, k_2, ..., k_K, L, P_1, P_2, ..., P_K) \tag{20}$$

is the probability mass function of the multinomial distribution for non-negative integers k_1, k_2,...,k_K.

4. Simulations and i discussion of results

Formula (19) derived in previous section was at first verified experimentally with the use of statistical simulations of the system with multiple base classifiers. Due to the high computational cost of such simulations, we considered only cases of classes numbers K from 2 to 10, numbers of base classifiers from 1 to 100 and selected values of base classifiers classification error probabilities P_S (0; 0,1; 0,3; 0,5; 0,7; 0,9 i 1). During simulations for each set of parameters 10^6 votings were performed where answers of individual base classifiers were generated randomly with use of standard random generator included in Borland Object Pascal System library.

Obtained results have shown high consistency between outcomes of conducted simulations and values of formula (19). For all considered values of parameters the difference between results of simulations and calculated error probabilities was not greater than 2,7% of computed values (average 0,043%). Additionally, for the case of two class problems both methods have given results consistent also with values of expression (2).

On the above basis, we observed how the multiple classifier error probability changes with increasing number of classes under consideration (see fig.1). For typical example of $L = 21$ and $P_S = 0,3$ for two classes the error probability is $P_E \approx 0,0264$, but for three and five classes it amounts just to 0,00202 and 0,000126. This is the result of growing number of classes other than the correct one - missed votes are dispersed over all K - 1 wrong classes. In the effect the average cumulative number of votes for individual wrong class decreases with increase of K, which do not apply to the correct class.

It is also very interesting that for number of classes K greater than 2, the upper limit of base classifier posterior error probability, that allows successfull building of multiple classifier is greater than 0.5 (compare fig. 2a and fig. 2b). Due to practical difficulties in creating large sets of base classifiers with a low errors probabilities and also with a high degree of lack of correlation between errors committed by them, observed result suggests the possibility of easier ensembles of classifiers building for complex multiclass problems by admission to the considerations also base classifiers that commit errors more frequently than in the half the cases.

For example - when the number of classes $K = 5$ and the number of base classifiers $L = 21$, error probability of base

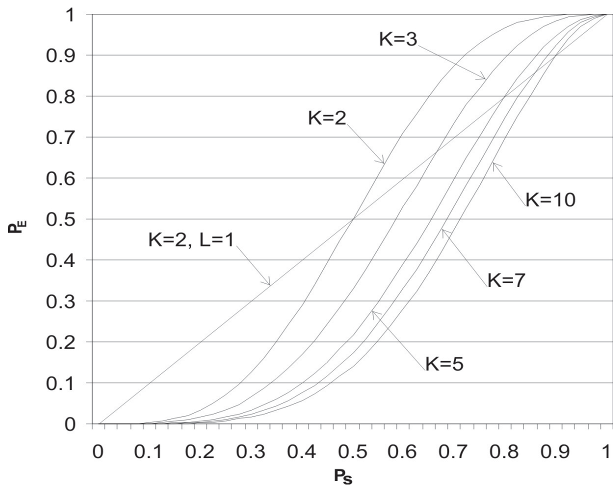
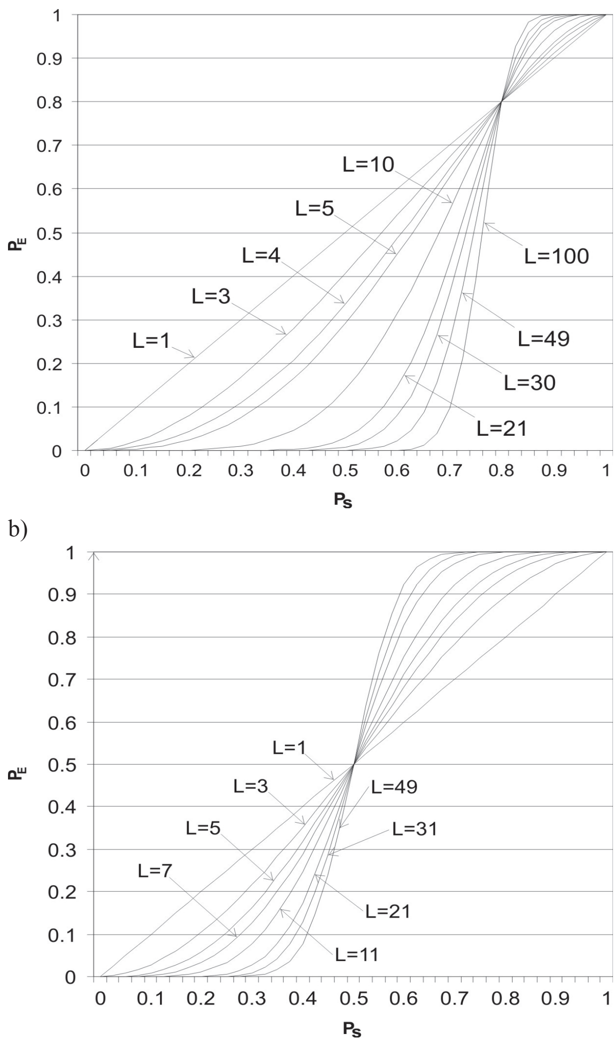


Fig. 1. Multiple classifier error probability P_E as a function of the error probability P_S of seven base classifiers (L=7), with negatively correlated mutual errors,for different numbers of classes K

a)



b)

Fig. 2. Multiple classifier error probability P_E as a function of the error probability P_S for different numbers L of base classifiers, with negatively correlated mutual errors, for five a) and two b) classes

classifiers $P_S = 0,6$ results in an error probability of a multiple classifier $P_E \approx 0,146$, what is the better value than randomly guessing. In addition, by increasing the number of base classifiers to 100, the above probability of multiple classicier error can be reduced to just 0,000815. However, it should be remembered that presented results were obtained under the assumption that the underlying mutual errors of base classifiers are fully uncorrelated or negatively correlated, which is difficult to achieve in practice. Partial correlation of errors can cause changes in individual values of the above probabilities, however, should not affect the basic properties of the results.

5. Summary and future work

In this work the formula for multiple classifier error probability for multi-class problems was formally presented. Its detailed derivation was based on the widely known analogous formula for two-class problems, which was additionally extended for even numbers of base classifiers.

Simulations during analysis of obtained formula indicate that increasing the number of considered classes lowers ensemble error probability. But what is more interesting, under assumption that mutual errors of base classifiers are uncorrelated or negatively correlated, the upper limit of base classifier posterior error probability P_S that allows successfull building of multiple classifier is increasing with considered number of classes.

As a consequence, the transition from the schema of bivalued to multivalued hypotheses, facilitates the creation of large collections of diverse base classifiers, and thus - even finer ensembles of classifiers. This could be of great importance for further applications of such methods in many fields of science and industry - including the issues of machines maintenance and diagnostics and systems reliability testing.

In future works we will investigate how the partial correlation between errors of multivalued base classifiers modifies error probabilities of multiple classifiers for numbers of classes greater than 2. We will also try to find computationally efficient expressions for estimation of derived formula for number of classes above 100.

6. References

1. Ali K, Pazzani M. Error reduction through learning multiple descriptions. Machine Learning 1996; 24(3): 173-206.
2. Bian S, Wang W. On diversity and accuracy of homogeneous and heterogeneous ensembles. IOS Press Amsterdam: 2007, 4(2): 103-128.
3. Brown G, Wyatt J, Harris R, Yao X. Diversity creation methods: A survey and categorization. Journal of Information Fusion 2005; 6(1).
4. Bruzzone L, Cossu R, Vernazza G. Detection of land-cover transitions by combining multidate classifiers. IOS Press Amsterdam: 2007, 25(13): 1491-1500.
5. Buhlmann P, Hothorn T. Boosting algorithms: Regularization, Prediction and Model Fitting. Statistical Science 2007; 22(4): 477-505.
6. Claeskens G, Hjort N. Model Selection and Model Averaging. Volume 27 of Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press 2008.
7. Dietterich T. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization., Machine Learning 2000; 40(2): 139-157.
8. Dietterich T. (Ensemble learning, in M. ARBIB (Ed.)) The Handbook of Brain Theory and Neural Networks. Second ed., Cambridge: 2002, 405-408.
9. Elovici, Y, Shapira B, Kantor P. A decision theoretic approach to combining information filters: Analytical and empirical evaluation., Journal of the American Society for Information Science and Technology 2006; 57(3): 306-320.
10. Evgeniou T, Pontil M, Elisseef A. Leave one out error, stability, and generalization of voting combinations of classifiers, Machine Learning 2004; 55(1): 71-97.
11. Freund Y, Lu J, Schapire R. Boosting: Models, Applications and Extensions., Chapman and Hall/CRC 2010.
12. Freund Y, Schapire R. Experiments with a new boosting algorithm., Machine Learning: Proceedings of the Thirteenth International Conference (ICML 96). SanFrancisco: 1996, 148-156.
13. Fumera G, Roli F. A theoretical and experimental analysis of linear combiners for multiple classifier systems., IEEE Transactions on Pattern Analysis and Machine Intelligence 2005; 27(6): 942-956.
14. Halbiniak Z, Jóźwiak I.: Deterministic chaos in the processor load. Chaos, Solitons and Fractals 2007; 31(2): 409-416.
15. Hansen L, Salamon P. Neural network ensembles, IEEE Transactions on Pattern Analysis and Machine Intelligence 1990; 12(10): 993-1001.
16. Jacak J, Jóźwiak I, Jacak L.: Composite fermions in braid group terms. Open Systems and Information Dynamics 2010; 17(1): 53-71.
17. Jahrer M, Tscher A, Legenstein R. Combining predictions for accurate recommender systems, Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, Washington, DC, USA: 2010, 693-702.
18. Kuncheva L. Combining Pattern Classifiers. Methods and Algorithms., Wiley 2004
19. Kuncheva L, Bezdek J, Sutton M. On combining multiple classifiers by fuzzy templates, Proceedings of the on combining multiple classifiers by fuzzy templates. Conference, Pensacola, Florida USA: 1990, 193-197.
20. Kuncheva L, Whitaker C, Shipp C, Duin R. Limits on the majority vote accuracy in classifier fusion, Pattern Analysis and Applications 2003; 6: 22-31.

21. Leapa N, Clemansa P, Bauer K, Oxley M. An investigation of the effects of correlation and autocorrelation on classifier fusion and optimal classifier ensembles, International Journal of General Systems 2008; 37(4): 475-498.
22. Leigh W, Purvis R, Ragusa J. Forecasting the nyse composite index with technical analysis, pattern recognizer, neural networks, and genetic algorithm: a case study in romantic decision support. Decision Support Systems 2002; 32(4): 361-377.
23. Liu Y, Yao X, Higuchi T. Evolutionary ensembles with negative correlation learning. IEEE Transactions on Evolutionary Computation 2000; 4(4): 380-387.
24. Mangiameli P, West D, Rampal R. Model selection for medical diagnosis decision support systems, Decision Support Systems 2004; 36(3): 247-259.
25. Menahem E, Shabtai A, Rokach L, Elovici Y. Improving malware detection by applying multi-inducer ensemble. Computational Statistics and Data Analysis 2009; 53(4): 1483-1494.
26. Niewczas A, Pieniak D, Bachanek T, Surowska B, Bieniaś J, Pałka K. Prognosing of functional degradation of bio-mechanical systems exemplified by the tooth-composite filling system. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2010; 45(1): 23-34.
27. Opitz D, Shavlik J. Generating accurate and diverse members of a neural-network ensemble, Advances in Neural Information Processing Systems. MIT Press, Denver: 1996, 535-543.
28. Rokach L. Mining manufacturing data using genetic algorithm-based feature set decomposition. International Journal of Intelligent Systems Technologies and Applications 2008; 4(1): 57-78.
29. Rokach L. Taxonomy for characterizing ensemble methods in classification tasks: a review and annotated bibliography. Computational Statistics and Data Analysis 2009; 53(12): 4046-4072.
30. Santhanam P, Bassin K. Managing the maintenance of ported, outsourced, and legacy software via orthogonal defect classification. In proc. IEEE International Conference on Software Maintenance 2001; 726-734
31. Shahrtash S, Jamehbozorg A. A Decision-Tree-Based Method for Fault Classification in Single-Circuit Transmission Lines. IEEE Transactions on Power Delivery 2010; 25(4): 2190-2196.
32. Tan A, Gilbert D, Deville Y. Multi-class protein fold classification using a new ensemble machine learning approach. Genome Informatics 2003; 14: 206-217.
33. Tao J, Zhang Y, Chen X, Ming Z. Bayesian reliability growth model based on dynamic distribution parameters. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2010; 46(2): 13-16.
34. Valentini G, Masulli F. Ensembles of learning machines. in M. M. and T. R. (Eds), Neural Nets: 13th Italian Workshop on Neural Nets, Vol. 2486 of Lecture Notes in Computer Science, Springer, Berlin: 2002, 3-19.
35. Xu D, Wu M, An J. Design of an expert system based on neural network ensembles for missile fault diagnosis. In Proc. IEEE International Conference on Robotics, Intelligent Systems and Signal Processing 2003, 2: 903-908.
36. Yu T, Cui W, Song B, Wang S. Reliability growth estimation for unmanned aerial vechicle during flight-testing phases. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2010; 46(2): 43-47.

Dr inż. Maciej HUK
Mgr inż. Michał SZCZEPANIK
Instytut Informatyki
Politechnika Wrocławska
Ul. Wyb. Wyspiańskiego nr 27
50-370 Wrocław, Polska
e-mail: Maciej.Huk@pwr.wroc.pl
e-mail: Michal.Szczepanik@pwr.wroc.pl