Multicriteria methods for identifying patterns in the analysis of the flow of "dangerous financial documents"

Andrzej Ameljańczyk[1], and *Maciej* Kiedrowicz[1,*]

[1]Military University of Technology, Faculty of Cybernetics, 2 Urbanowicza Str., 00-908 Warsaw, Poland
[*]Corresponding author: maciej.kiedrowicz@wat.edu.pl

Abstract. The article outlines a concept of applying the methodology for identifying patterns used for detecting documents suggesting the execution of some criminal financial transactions. The analogies of diagnostic processes for disease classification in medicine were used in the method. The idea of the described method consists in defining model patterns of financial documents, suggesting criminal activity in the form of the financial flow and developing mathematical models of actual financial documents which shall be used for the comparisons with the patterns at a later stage of the process. The next step is to develop similarity indicators of documents to appropriate patterns, to define and develop a multicriteria detection area for the documents and to develop a method for dividing the set of monitored documents into similar document classes. The final stage is the development of multicriteria rankings that allow to organize the set of transaction documents according to the degree of similarity to the relevant patterns and to determine the optimal cut-off threshold in the ranking of documents intended for a more detailed analysis. The described method may be used in counteracting financial crimes, and in particular in combating money laundering.
Key words: identification of patterns, multicriteria similarity model, dangerous document pattern, Pareto filtration

Introduction

The detection of criminal financial transactions (financial crimes, supporting of terrorism, money laundering) is a difficult and complex issue. The total number of financial transactions in the national financial market reaches several millions of operations per day. The scope of diversity of financial transactions is really extensive. The financial transactions are executed in different transactional environments: physical environments, local networks, postal networks and global networks such as the Internet. To "fish out" criminal transactions is, therefore, a really complex, but also important process, as it may lead to quick and efficient identification of criminal groups and prediction of financial crimes, hence, allowing their earlier detection and prevention. Therefore, particular countries (including the whole European Union) introduce a number of regulations (including regulations equivalent to Acts of Parliament) imposing certain obligations (procedures) on financial institutions to ensure the undertaking of efficient actions preventing such events [1]. In compliance with the provisions of relevant Acts, the financial institutions shall, as part of applying financial security measures, undertake the following actions: "monitor economic relations on an ongoing basis, including the inspection of transactions executed during such relations to ensure that the transactions are in accordance with the knowledge of a given institution regarding the client, business profile and the risk, including, where possible, the sources of funds, and to ensure that the documents, data or information held by the institutions are being updated regularly" [1]. Due to the scope and complexity of the financial flow system and, above all, due to the special role of detecting criminal financial transactions, it has become necessary to use appropriately designed IT systems supporting such activities. The appropriately designed and implemented IT systems may turn out to be an efficient tool supporting the detection of criminal financial transactions. The main module of such systems is the subsystem for detecting patterns of transactional documents used in criminal financial activities, which includes the multicriteria module for the similarity analysis [2-5]. The automatically selected set of "suspicious financial documents" may be then

"manually" verified by experts and used for subsequent operational activities.

The article describes the general idea of the proposed method for identifying financial documents (Chapter 1). The analogy used in medical diagnostic procedures based on the recognition of "disease patterns" (Chapter 2) was used. The next chapter presents a multicriteria model of the process for detecting patterns of "dangerous" documents. This article is aimed at presenting the model of initial detection in such a manner so as to enable the use of broad and efficient set of possibilities offered by the theory of multicriteria optimization in further studies.

1. General concept of the method for identifying financial documents

The methods for identifying financial flow of funds as well as parties and criminal groups involved in financial transactions may come down to methods for identifying financial documents related to criminal activity. The identification and detection of such documents usually allow to quickly and clearly define the offending parties and types of financial crimes. Therefore, this study shall be mainly devoted to the so-called computer methods for identifying financial documents. Such documents shall be identified based on the degree of similarily of the monitored actual documents according to specific patterns of financial documents suggesting criminal activity [1-2, 6].

The method consist in:

1) defining model patterns of financial documents suggesting criminal activity in the form of the financial flow [2-4, 7-8], i.e. the "dangerous documents";
2) developing mathematical models of actual financial documents, which shall be used for the "model comparisons" with the patterns at a later stage of the process [9];
3) developing similarity indicators for comparing the documents with appropriate patterns [4-5, 10];
4) defining and developing the so-called multicriteria detection area for the documents [2-3,11];
5) developing a method for dividing the set of the monitored documents into clusters (classes) of similar documents (the use of the so-called Recurrent Pareto Filter (RPF) [3, 12]);
6) developing multicriteria rankings allowing to order the set of the transactional documents according to the degree of similarity in comparison with appropriate patterns [6, 10, 12-13];
7) determining optimum cut-off threshold in the ranking of the documents meant for a more detailed analysis [3-4, 14-15].

2. Modeling of patterns of the selected transactional document classes

In its essence, the general concept of the methodology for identifying patterns of criminal financial documents is very similar (therefore, may be successfully used) to the medical diagnostic procedures [4-5, 14], which consist in identifying "disease patterns" on the basis of:

a) external symptoms (manifestation),
b) risk factors (circumstances),
c) additional specialist studies.

In case of procedures for identifying (detecting) criminal financial documents, a similar methodology is used. The disease patterns correspond to appropriate classes of criminal financial documents. Such patterns must be also defined in terms of characteristic attributes of a given class of documents:

a) external similarity of the documents (external symptoms) – characteristics of the transaction,
b) risk factors (circumstances of the financial transaction),
c) additional specialist studies (arrangements of the expert).

Similarly, "current transactional documents" monitored to find unusual documents correspond, in this case, to patients (in medical diagnosis). The diagnosis in the process of detecting dangerous documents consists in the selection of a subset of documents most similar in their classes to specific patterns of dangerous (criminal) documents [3-4, 14].

Therefore, a typical model (pattern) of a dangerous document should contain three segments of characteristic information:

- description of characteristics (symptoms) typical for a given type of the document [1, 6, 11, 16],
- description of risk factors (circumstances) accompanying the generation and circulation procedure of such documents [1, 6, 17],
- description of types of potential, additional specialist studies (forensic examination).

Formally, the mathematical model of the $\mathrm{M}(m)$ pattern of the $m \in \mathrm{M} = \{1,...,\mathrm{M}\}$ type document may be as follows [3]:

$$\mathrm{M}(m) = \left(S^m, R^m, P^m\right) \tag{1}$$

where: S^m – a set of numbers of features (symptoms) characteristic of the $m \in \mathrm{M}$ document [1]

$$S^m = \left\{S_1^m, ..., S_k^m, ..., S_{K(m)}^m\right\} \subset S, m \in \mathrm{M} \tag{2}$$

The S - set is the set of number of all characteristics (i.e. $S \subset \mathrm{N}$).

$K(m)$ – a number of symptoms for the $m \in \mathrm{M}$ type pattern [1].

R^m - a set of number of the risk factors (circumstances), including the generation of the $m \in \mathsf{M}$ document [1].

$$R^m = \left\{ r_1^m, ..., r_l^m, ..., r_{L(m)}^m \right\} \subset R,\ m \in \mathsf{M} \quad (3)$$

The R set is a set of numbers of all types of circumstances, in which the dangerous documents included in the $(R \subset N)$ register are generated.

$L(m)$ − a number of risk factors for the $m \in \mathsf{M}$ type document.

P^m − a set of numbers of the types of specialist studies for the $m \in \mathsf{M}$ document [1].

$$P^m = \left\{ p_1^m, ..., p_n^m, ..., p_{N(m)}^m \right\} \subset P,\ m \in \mathsf{M} \quad (4)$$

The P set is a set of numbers of all types of the $(P \subset N)$ studies (whose values may be determined during the specialist studies).

$N(m)$ − a number of all types of specialist studies concerning the $m \in \mathsf{M}$ type document.

While identifying each type of a dangerous document, particular features (symptoms), risk factors and results of appropriate specialist studies have different meaning (different "characteristic gravity") [1, 5-6, 11, 16-18].

Therefore, the numbers (as specified by the experts):

$$\begin{aligned} &\alpha(s_k^m) \in [0,1],\ \ s_k^m \in S^m \\ &\beta(r_l^m) \in [0,1],\ \ r_l^m \in R^m \\ &\gamma(p_n^m) \in [0,1],\ \ p_n^m \in P^m \end{aligned} \quad (5)$$

shall mean "priority" of particular parameters concerning the characteristics, risk factors and additional studies in the field of detection of documents types no. $m \in \mathsf{M}$.

Let us assume that as a result of the preliminary stage, the $S_o(x) \subset S$ set of symptoms and set of risk factors were identified in document x $R_O(x) \subset R$

$$\begin{aligned} S_O(x) &= \{ s \in S \,|\, w(x,s) > 0 \} \\ R_O(x) &= \{ r \in R \,|\, w(x,r) > 0 \} \end{aligned} \quad (6)$$

where $w(x,s)$ − a rate of "occurrence" of the $s \in S$ symptom (determined by the inspector (expert) on a [0,1] scale (often, the aforesaid rate has binary values: 0 or 1). Similarly, $w(x,r)$ − a rate of occurrence of risk factor no. r (also on a [0,1] scale).

3. Multicriteria model of the process for detecting patterns of "dangerous" documents

The $\mathsf{M}_o(S)$ set of the occurring symptoms as suggested by the set shall be defined in the following manner:

$$\mathsf{M}_o(S) = \{ m \in M \,|\, S_o(x) \cap S^m \neq \phi \} \quad (7)$$

Similarly, the $\mathsf{M}_o(R)$ set of the documents related to the occurring risk factors shall be defined in the following manner:

$$\mathsf{M}_o(R) = \{ m \in M \,|\, R_o(x) \cap R^m \neq \phi \}$$

Another step shall be to determine the total set of dangerous documents. The set may constitute preliminary estimation

$$\mathsf{M}_o = \mathsf{M}_o(S) \cup \mathsf{M}_o(R)$$

or more radically:

$$\mathsf{M}_o = \mathsf{M}_o(S) \cap \mathsf{M}_o(R)$$

However, such an approach to initial identification of the documents is quite risky due to a possibility of occurrence of risk factors or symptoms simultaneously for several types of dangerous documents and difficulties in their precise definition. With the data on the $x \in X$ document regarding occurrence of the risk symptoms and factors in the form of $w(x,s),\ s \in S_o(x)$ and $w(x,r),\ r \in R_O(x)$ numbers, it is possible to determine the "distance of document *x* " from appropriate patterns of dangerous documents included in $\mathsf{M}_o(S)$ and $\mathsf{M}_o(R)$ sets. It may be done in the following way.

The model of the $x \in X$ document, defined on the basis of the occurring risk symptoms and factors is in the form of a pair:

$$f(x) = (f_S(x), f_R(x)),\ x \in X \quad (8)$$

where:

$$\begin{aligned} f_S(x) &= (w(x,s); s \in S_o(x)), \\ f_R(x) &= (w(x,r); r \in R_o(x)) \end{aligned}$$

$\overset{*}{s}(m)$ and $\overset{*}{r}(m)$ symbols are used to designate patterns no. m , respectively, in terms of the risk symptoms and factors [4, 7].

The $d_1\left(f_S(x), \overset{*}{s}(m) \right),\ m \in \mathsf{M}$ symbol is used to designate distance (similarity) of document *x* (as resulting from the occurring symptoms) from the $m \in M$ pattern, defined on the basis of the symptoms and, analogically, the symbol $d_2\left(f_R(x),\ \overset{*}{r}(m) \right),\ m \in \mathsf{M}$ is used for marking the distance of document *x*, (as resulting from the occurring risk factors) from the pattern of a dangerous document type $m \in \mathsf{M}$, defined on the basis of risk factors.

The $M(S_0(x))$ set of "the most probable" patterns "matching" the symptoms shall be established in the following manner:

$$M(S_o(x)) = \left\{ \overset{*}{m} \in \mathsf{M}_o \,\middle|\, d_1\left(f_s(x), \overset{*}{s}\left(\overset{*}{m}\right) \right) = \min_{m \in \boldsymbol{M}_o} d_1\left(f_S(x),\ \overset{*}{s}(m) \right) \right\} \quad (9)$$

On the other hand, the $M(R_o(x))$ set of "the most probable" patterns in terms of the occurring risk factors shall be established in the following manner:

$$M\left(R_o(x)\right)=\left\{m^{*}\in \mathsf{M}_o \middle| d_2\left(f_R(x), r^{*}\left(m^{*}\right)\right)=\min_{m\in \boldsymbol{M}_o} d_2\left(f_R(x),\ r^{*}(m)\right)\right\} \quad (10)$$

An empty set is often a common part of such $\overset{*}{\mathsf{M}}=\mathsf{M}\left(S_o(x)\right)\cap\mathsf{M}\left(R_o(x)\right)$ sets [7]. Below is an algorithm for setting preliminary similarity drawing based on the idea of multidimensional similarity described in the previous point of this study. The similarity indices shall be defined as properly understood distances of the document from the patterns of dangerous documents. They constitute a certain modification of the Jaccard distance (similarity). The computer assistance process is executed on the basis of the software algorithms for diagnostic conclusion. The basis for the construction of such algorithms are document models and models (patterns) of dangerous documents. The suggestion (proposal) of subsequent detection activities (if necessary) constitutes the result of the implemented algorithm. The general idea of the supporting mechanism, depending on the adopted modeling concept (e.g. Bayesian network [12, 19], fuzzy sets [3, 20-21], proximate sets, cobweb models or the concept of patterns [4]), consists in selecting the list of the most probable identified documents and then choosing the optimum set of additional specialist studies. The theory of multicriteria optimization and relational structures [7, 22] is an interesting proposal in terms of identifying sets of patterns, most probable from the point of view of the set of the occurring risk symptoms and factors. When establishing an appropriate model of "detection preferences" $\overline{R}$, such task may be defined in the form of

$$\left(\mathsf{M}_o, d(m), \overline{R}\right) \quad (11)$$

where $d(m)$ function is a vector function measuring the distance (similarity) of the document from the pattern of dangerous document no. m

$$d(m)=\left(d_1(m), d_2(m)\right),\ m\in M \quad (12)$$

The distances for document x are defined in the following manner [3]:

$$d_1(x,m)=1-\sum_{s_k^m\in S_o^m(x)} w\left(x, s_k^m\right)\alpha\left(s_k^m\right),\ m\in\mathsf{M}$$
$$d_2(x,m)=1-\sum_{r_l^m\in R_o^m(x)} w\left(x, r_l^m\right)\beta\left(r_l^m\right),\ m\in\mathsf{M} \quad (13)$$

$\overline{R}$ − model of detection preferences (similarity relation, e.g. Pareto [7]).

In practice, the following three options of the detection preferences are taken into account:

1) risk symptoms and factors are equally important (Pareto relation),
2) risk symptoms are more important (hierarchical relationship),
3) risk factors are more important (hierarchical relationship).

In case of two criteria and the M_o set with relatively "small numbers", the above sentence may be easily illustrated in graphic form. The illustration is in fig. 1.

The image of the set of M_o patterns in terms of distance from document x shall be set Y (fig. 1):

$$Y=d(M_o)=\left\{d(m)\in R^2 \middle| m\in M_o\right\}$$

Therefore, the solution to the task shall be the so-called Pareto's set [7], i.e. the set of patterns from the set of initial estimation M_o, with respect to which no "more similar" objects can be found. The set shall be marked with the following symbol:

$$M_N^{\overline{R}}=\left\{m^{o}\in \mathsf{M}_o \mid m\in \mathsf{M}_o-\left\{m^{o}\right\} \text{ does not exist that } d(m)\le d\left(m^{o}\right)\right\}$$

The $M_N^{\overline{R}}$ set is the inverse image [3] of the $Y_N^{\overline{R}}$ Pareto's set.

$$M_N^{\overline{R}}=d^{-1}\left(Y_N^{\overline{R}}\right)=\left\{m^{o}\in M_o \middle| d\left(m^{o}\right)\in Y_N^{\overline{R}}\right\} \quad (14)$$

In this case, the so-called "compromise solution" [22], which usually leads to an unambiguous solution, may be the final determinative factor.

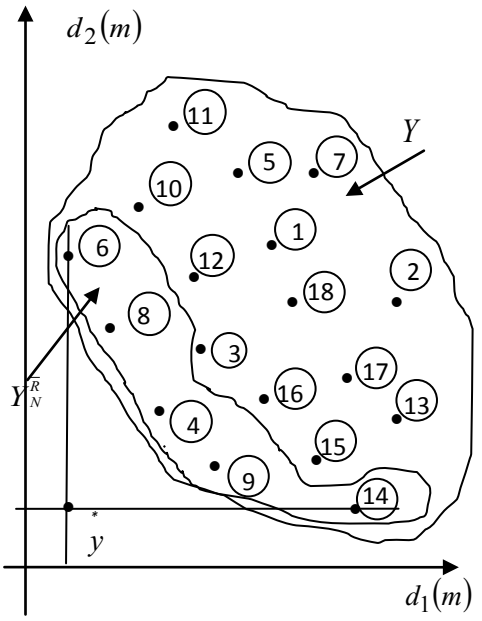

Fig. 1. The concept of determining the $M_N^{\overline{R}}$ set of patterns of dangerous documents, with respect to which there are no more similar documents [6]

The $\mathsf{M}_o=\{1,...,18\}$ set in the above-mentioned example constitutes preliminary estimation of the set of possible patterns. The patterns with numbers {4,6,8,9,14} (inverse image of the $Y_N^{\overline{R}}$ set) form the $M_N^{\overline{R}}$ set. Therefore, in case of document x, there are "suspicious occurrences" of dangerous documents numbers $m\in M_N^{\overline{R}}$. When calculating the distance of the images of such document from the y^{*} "utopian" (virtual) image,

the "most probable" (in terms of ascertained risk symptoms and factors) pattern, it is possible to create a ranking of potential patterns for further detection activities.

The closest "most probable pattern", as resulting from the ascertained risk symptoms and factors, is the pattern of dangerous document type no. 4. However, in practice, for the expert to be able to make the final decision, the "whole Pareto's set" and ranking of its elements are necessary.

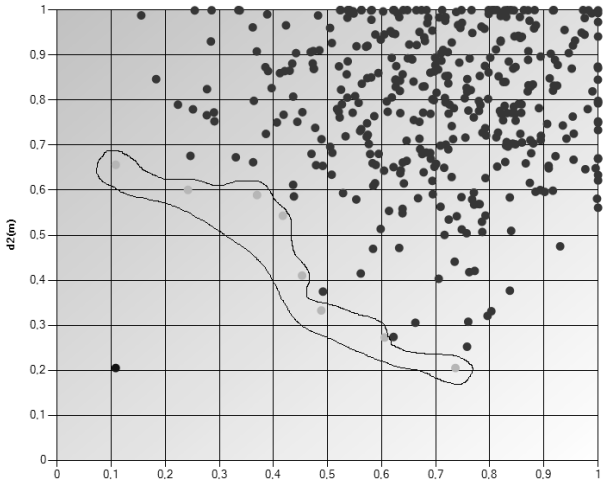

Fig. 2 Typical image of the set of patterns of the documents, with respect to which there are no "more similar" documents using a computer example.

Coordinates of the $\overset{*}{y}=\left(\overset{*}{y}_1,\overset{*}{y}_2\right)$ utopian pattern are set in the following manner:

$$\overset{*}{y}_1=\min_{m\in \boldsymbol{M}_o} d_1(m),\ \overset{*}{y}_2=\min_{m\in \boldsymbol{M}_o} d_2(m) \qquad (15)$$

An important aspect of the modeling process is the choice of the forms of similarity functions (d_1 and d_2 distances) and decision on whether to accept the appropriate $\overline{R}$ preference model. Specific mathematical formulas defining the so-called "distance functions" are based on the adopted modeling concepts [2-3, 19]. For example, in case of models based on the Bayesian networks, such concepts are in the form of appropriate distributions of conditional probabilities. In case of models based on the theory of fuzzy sets [20-21, 23-24], the concepts refer to the functions of belonging to the set of initially identified documents, and in case of models based on patterns - the appropriately defined metrics in the so-called detection area [2-3]. In special cases, the models of detection preferences do not have to be based on Pareto's relations or "lexicography". The relations may be based on the pessimist (optimist) model or the so-called "collective preference relations". This article was aimed at presenting the model of initial detection in such a manner so as to enable the use of broad and efficient set of possibilities offered by the theory of multicriteria optimization in further studies. The procedure outlined in the article may be considered the preliminary detection process initiating each identification process of dangerous documents [25-27]. The procedure leads to the generation of the set (relatively with small numbers) of the so-called patterns, with respect to which there are no more probable documents. Another step of the detection process is to potentially (if necessary) choose the "optimum set" of additional specialist studies allowing to make a final decision in terms of identifying the dangerous document, and then to choose the optimum strategy for further operational (preventive) activities.

Summary

The above-described filtering methodology (the application of the so-called Pareto's filter [3, 6]) of the set of actual documents allows obtaining the detection area of to compare documents with particular patterns. The result of such procedure is the so-called "Pareto's front" [2-3, 7], which constitutes the subset of identified patterns of criminal documents, with respect to which there are no more similar documents (in terms of the Pareto's relations) to the analyzed document. In practice, it is usually a quite extensive set. In such case, it seems reasonable to use the methodology for filtering transactional documents, based on the appropriately designed multicriteria-ranking algorithm [3, 6, 14]. The final subset of the "most suspicious" documents may be selected using the so-called optimum cut-off threshold [3, 5-6]. The multicriteria ranking methods outlined in the article [2-3, 6-7] may be successfully used, upon modification, for the purpose of the monitored filtering of the set of transactional documents in terms of their similarity to the applied patterns of dangerous documents.

Bibliography

[1] Poradnik – *Przeciwdziałanie praniu pieniędzy i finansowaniu terroryzmu*, Wydanie III zmienione i poprawione, praca zbiorowa, Ministerstwo Finansów, Warszawa, 2009

[2] A. Ameljańczyk, Wielokryterialne mechanizmy wspomagania podejmowania decyzji klinicznych w modelu repozytorium w oparciu o wzorce, *Biuletyn Instytutu Systemów Informatycznych* Nr 5, 2-8,(2010).

[3] A. Ameljańczyk, „Metoda podziału zbioru obiektów na wielokryterialne klastry jakościowe”, *Biuletyn Instytutu Systemów Informatycznych*, Nr 12, 1–7 (2013).

[4] A. Ameljańczyk, „Multicriteria similarity models in medical diagnostics support algorithms”, *Bio-Algorithms and Med.-Systems*, Vol. 21, No.1, 33–39 (2013).

[5] A. Ameljańczyk, „Wiarygodność komputerowych systemów wspomagania diagnostyki medycznej”,

w: *Problemy modelowania i projektowania opartych na wiedzy systemów informatycznych na potrzeby bezpieczeństwa narodowego*, 23–39, WAT, Warszawa, 2014.

[6] A. Ameljańczyk, „Metryki Minkowskiego w tworzeniu uniwersalnych algorytmów rankingowych", *Biuletyn WAT*, Vol. LXIII, Nr 2, 324–336 (2014).

[7] A. Ameljańczyk, Analiza wpływu przyjętej koncepcji modelowania systemu wspomagania decyzji medycznych na sposób generowania ścieżek klinicznych, *Biuletyn Instytutu Systemów Informatycznych,* Nr 4, 1-6 (2009).

[8] M. Kiedrowicz, *Interoperability and Globalization of Information Models*, Conference: Geographic Information Systems Conference and Exhibition "GIS ODYSSEY 2017", 4-8 of September 2017, Trento–Vattaro, Italy (2017).

[9] M. Kiedrowicz, *Objects identification in the information models used by information systems*, Conference: Geographic Information Systems Conference and Exhibition "GIS ODYSSEY 2016", Perugia, Italy (2016).

[10] H. Courtney, J. Kirkland, and P. Viguerie, *Strategia w warunkach niepewności*, in: *Zarządzanie w warunkach niepewności*, Harvard Busines Review, 316–324, Helion, Gliwice, (2006).

[11] J.P. Brans, Ph. Vincke, *A preference ranking organization method: The PROMETHEE method for Multiple Criteria Decision-Making*, *Management Science*, Vol. 31, No. 6, 647–656 (1985).

[12] M.F. Balcan, N. Bansal, A. Beygelzimer, D. Coppersmith, J. Langford, and G.B. Sorkin, *Robust reductions from ranking to classification*, *Machine Learning*, 72(1–2):139–153 (2008).

[13] M. Kiedrowicz, *Multi-faceted methodology of the risk analysis and management referring to the IT system supporting the processing of documents at different levels of sensitivity*, *MATEC Web of Conferences*, vol. 125 (2017)

[14] A. Walczak, B. Bieniek, K. Różyk-Jahnz, E. Paluchowska , *Fuzja klasyfikatorów w diagnostyce chorób skóry*, in: *Problemy modelowania i projektowania opartych na wiedzy systemów informatycznych na potrzeby bezpieczeństwa narodowego* (eds. T. Nowicki and Z. Tarapata), 143–152, WAT, Warszawa, 2014.

[15] M. Kiedrowicz, J. Stanik, *Models and method for the risk assessment of an intellectual resource*, *WSEAS Transactions on Information Science and Applications*, vol. 14(2017), pp.: 174-183 (2017).

[16] D. Larose, *Metody i modele eksploracji danych*, Wydawnictwo Naukowe PWN, Warszawa, 2008.

[17] D. Bouyssou, T. Marchant, *An axiomatic approach to noncompensatory sorting methods in MCDM*, *I: The case of two categories*, *EJOR*, 178(1): 217–245 (2007).

[18] Z. Pawlak, *Systemy informacyjne – podstawy teoretyczne*, WNT, Warszawa, 1983.

[19] J. Furnkranz , E. Hullermeier , E. Mencia, and K. Brinker, *Multilabel classification via calibrated label ranking*, *Machine Learning*, 73:133–153 (2008).

[20] S. Acid, L.M. Campos, "A comparison of learning algorithms for Bayesian Networks: a case study based on data from an emergency medical service", *Artificial Intelligence in Medicine*, 30, 215–232 (2004).

[21] Z. Pawlak, *Rough Sets*, *International Journal of Computer and Information Sciences*, Vol. 11, 341–356 (1965).

[22] H. Rasiowa, *Wstęp do matematyki współczesnej*, PWN, Warszawa, 2005.

[23] C. Ruiz, *Illustration of the K2 Algorithm for Learning Bayes Net Structures*, Department of Computer Science WPI, Bayesian Network Power Constructor, Worcester, MA, 2009.

[24] T.L. Saaty, *Rank from comparisons and from ratings in the analytic hierarchy/network processes*, *EJOR*, 168(2):557–570 (2006).

[25] Choi Seung-Seok, Cha Sung-Hyuk, Charles C. Tappert, *A Survey of Binary Similarity and Distance Measures*, Pace University, New York, 2006.

[26] C.A. Shipp , L.I. Kuncheva , *Relationships between combination methods and measures of diversity in combining classifiers*, *Information Fusion*, Vol. 3, No. 2, 135–148 (2002).

[27] D. Pierzchała, R. Antkiewicz, M. Dyk, R. Kasprzyk, A. Najgebauer, Z. Tarapata: „*Modelling, simulation and computer support of the Polish criminal procedure*", in: *Information Systems Architecture and Technology. The Use of IT Technologies to Support Organizational Management in Risky Environment*, ed. Z. Wilimowska, L. Borzemski, A. Grzech, J. Świątek, ISBN 978-83-7493-858-7, pp. 51-60, Wrocław, 2014