COMPUTER SCIENCE AND MATHEMATICAL MODELLING **1** 11–16 (2015)

Property analysis of multi-label classifiers in the example of determining the initial medical diagnosis

A. AMELJAŃCZYK
andrzej.ameljanczyk@wat.edu.pl

Military University of Technology, Faculty of Cybernetics,
Institute of Computer and Information Systems
Kaliskiego Str. 2, 00-908 Warsaw, Poland

In the paper analysis of properties of medical diagnoses acquired using simple and complex classifiers was performed. The introduced terms are illustrated with a comprehensive example in the field of medical diagnostics.

Keywords: medical diagnostics, committee classifiers, synthesis of classifiers, not dominated diagnosis classifier reliability.

1. Introduction

The work below presents the analysis of properties of a classifier being the synthesis (fusion) of two simple (ranking) domain classifiers applied to medical diagnostics. The presented work is a continuation of the theoretical results included in [2].

The example concerns determination of initial diagnosis for the patient on the basis of diagnosed symptoms of disease and risk factors with regard to their intensification. The applied diagnostic model, described in the papers [6, 7] also considers the significance levels for the applicable symptoms and risk factors in diagnosing individual diseases. Data on the patient's health condition (observation $x \in X$) were divided by domains into two areas: data on symptoms presence and data on risk factors and their intensification. These data formed the basis for developing two domain classifiers. The third classifier, typical for medical diagnostics processes and 'operating' on data concerning the specialist results of laboratory tests is included further in the iteration diagnostic process upon obtaining the initial diagnosis.

2. Pareto filter as a complex classifier

The classifiers used in the example were developed on the basis of ranking functions [2], and their values may be interpreted as the distance (fitting rate, similarity) of observation $x \in X$ (in the area of symptoms or risk factors, respectively) to the applicable disease units $l \in \mathcal{L}$, [6, 9].

$f_x^1(l)$ – similarity rate for symptoms

$f_x^2(l)$ – similarity rate for risk factors (1)

Set $\mathcal{L} = \{l_1,...,l_m,..,l_{20}\}$ in the analyzed example is a set of twenty disease units (labels) indexed with $m \in \mathcal{M}$, presented in Table 1. This table also contains (for the adopted observation of medical results $x \in X$) the values of both ranking functions (1). As we see, these values are not injective, therefore the rankings $r(f_x^1)$ and $r(f_x^2)$ developed on the basis there of will be not linear [2, 20].

Table 1. Values of similarity rates

m	1	2	3	4	5	6	7	8	9	10
f_x^1	0.3	0.4	0.5	0.6	0.6	0.5	0.5	0.3	0.2	0.5
f_x^2	0.6	0.7	0.6	0.4	0.3	0.2	0.1	0.1	0.1	0.5
m	11	12	13	14	15	16	17	18	19	20
f_x^1	0	0.1	0.2	0.4	0.5	0.4	0.3	0.2	0.3	0.4
f_x^2	0.4	0.6	0.6	0.5	0.4	0.3	0.2	0.4	0.5	0.6

The applied medical data (observation *x*) refer to the so-called "difficult diagnostic case" since the calculated values of the similarity function (1) according to the model adopted in works [6, 7] achieved relatively low and repeatable for many labels values.

On the basis of the a/m functions, the following two simple classifiers were developed:

$$C_1(x) = \arg\max_{l \in \mathcal{L}} f_x^1(l)$$
$$C_2(x) = \arg\max_{l \in \mathcal{L}} f_x^2(l) \qquad (2)$$

Classifications acquired with these classifiers (initial diagnoses) are as follows:

$$D_1(x) = C_1(x) = \{l_4, l_5\},\ D_2(x) = C_2(x) = \{l_2\} \qquad (3)$$

Diagnostic concluding on the basis of these results is most probably hindered and doubtful, for example due to that

$$C_1(x) \cap C_2(x) = \varnothing \qquad (4)$$

This undoubtedly results from the fact that both classifiers are too simple (only the symptoms or only risk factors and of low accuracy and the ranking functions are not injective) [6, 7]).

A logical and safe approach in this case would be adopting the initial diagnosis $D(o)$ in the form

$$D(o) = C_1(x) \cup C_2(x) = \{l_2, l_4, l_5\} \qquad (5)$$

without any guarantee, however, that the actual diagnosis will be included in set $D(o)$. An alternative manner will be synthesis of the applied simple classifiers (2). Figure 1 presents the area of synthesis Y as well as ranking image Y_x of set for observation $x \in X$ and the so-called ideal point $\overset{*}{y}(x) = (0,6;0,7)$.

The coordinates of such point shall be: $\overset{*}{y}_n(x) = \max_{y \in Y_x} y_n = \max_{l \in \mathcal{L}} f_x^n(l), n \in \mathcal{N}$

This point is a ranking image of virtual label (utopian label) of such a disease unit, which would have the highest similarity rate in terms of symptoms and risk factors under observation $x \in X$ [2, 3, 26, 27].

As the synthesis relation, the previously discussed [2] Pareto relation was adopted – the most common in such cases [3, 13, 15]. According to [2], an integrated classifier (generated as a result of synthesis) shall be the classifier:

$$C_R(x) = f_x^{-1}\left(Y_x^{RN}\right) \subset \mathcal{L}, \qquad (6)$$

Set $Y_x^{RN} = \{2,3,4\}$ was marked on Figure 1.

Thus

$$C_R(x) = f_x^{-1}\left(\left\{(0,4;0,7),(0,5;0,6),(0,6;0,4)\right\}\right)$$
$$= \left\{l_m \in \mathcal{L} \,\middle|\, f_x(l_m) \in Y_x^{RN}\right\} = \left\{l_2, l_3, l_4\right\}$$

that is

$$C_R(x) = \mathcal{L}_x^{RN} = \{l_2, l_3, l_4\} \subset \mathcal{L}$$

This is a set of disease units (labels), from which there are no 'more fitted' units in set $\mathcal{L}$ with regard to observation $x \in X$ in the area of diagnosed symptoms and risk factors. This is the effect of operation of the integrated classifier.

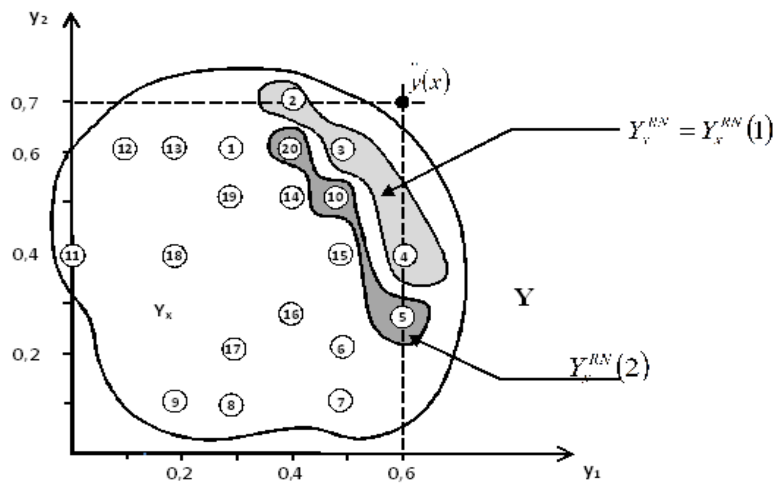

Fig. 1. Space of synthesis **Y** and set Y_x

Table 2 presents the list of indications of the individual classifiers on the basis of observation $x \in X$ with reference to set $\mathcal{L}$.

The last two columns of the table present information on conformity of the conformity rate of the indication of the given classifier with "baseline indication" $\mathcal{L}_x^{RN}$, concerning the set of labels, from which there are no other better fitted [2]. The last column of the table contains the values of Jaccard's conformity (similarity) index [24, 28] of indication of a given classifier with a set of diagnosis, from which there are no better fitted.

Tab. 2. Proposals of initial diagnoses determined by the selected classifiers

Lp.	Classifier	Classifier indication	Common part with the set $\mathcal{L}_N^R$	The compatibility factor Jacckard set and $\mathcal{L}_N^R$
1.	$C_1(x)$	$\{l_4, l_5\}$	$\{l_4\}, 1/3$	$1/4$
2.	$C_2(x)$	$\{l_2\}$	$\{l_2\}, 1/3$	$1/3$
3.	$C_3(x)$	$\{l_2, l_3, l_4, l_5, l_{10}, l_{20}\}$	$\{l_2, l_3, l_4\}, 1$	$1/2$
4.	$C_R(x)$	$\{l_2, l_3, l_4\}$	$\{l_2, l_3, l_4\}, 1$	1
5.	$C_{MG}(x)$	$\{l_2, l_4\}$	$\{l_2, l_4\}, 2/3$	$2/3$
6.	$C_{WG}(x)$	$\{l_2, l_4\}$	$\{l_2, l_4\}, 2/3$	$2/3$

The next to last column contains information on “intersection” of the indication of a given classifier with the set of nondominated labels and the coverage index with regard to intersection (proportion of a number of nondominated labels contained in the classifier indication to the total number of nondominated labels).

Referring to the need of possible further extension of classification we may [2], additionally determine the following sets on recurrent basis:

$\mathcal{L}_x^{RN}(1)$, $\mathcal{L}_x^{RN}(2)$ (the “most important” two, subsequent clusters of label set $\mathcal{L}$ (see Figure 1) [5] and perform the ranking of the total of these sets $\mathcal{L}_x^{RN}(1 \cup 2)$

$$\mathcal{L}_x^{RN}(1) = \{l_2, l_3, l_4\}$$

$$\mathcal{L}_x^{RN}(2) = \{l_5, l_{10}, l_{20}\}$$

$$\mathcal{L}_x^{RN}(1) \cup \mathcal{L}_x^{RN}(2) = \{l_2, l_3, l_4, l_5, l_{10}, l_{20}\} = \mathcal{L}_x^{RN}(1 \cup 2)$$

developing another classifier $C_3(x) = \mathcal{L}_x^{RN}(1 \cup 2)$.

This set may be further subject to ranking using the distance of its elements from the ideal point (hypothetical ideal diagnosis) $\overset{*}{y}(x)$ [6, 7, 26].

We shall obtain (see Figure 1) the following ranking:

$$r(\ \mathcal{L}_x^{RN}(1 \cup 2)\ ) = \langle 3,2,10,20,4,5 \rangle$$

Let’s notice a significant difference between the ranking indications presented above with the ranking indications of only the first Pareto filtration:

$$r(\mathcal{L}_x^{RN}) = \langle 3, 2, 4 \rangle$$

Using the recurrent Pareto filtration formula [2], extended the set of initial indications with the new labels numbered 10, 20 and 5, of which two l_{10} and l_{20} overtook label l_4 in the ranking determined on the basis of distance from the ideal point (ideal label) $\overset{*}{y}(x)$. This is a significant diagnostic conclusion implying great caution in excessively restrictive narrowing of the initial diagnosis. The specific nature of this situation is clearly illustrated in Figure 1.

3. Classifier committee voting

For completeness of analysis of the discussed example the indications resulting from ‘classifier voting’ will be determined. In the case of such utilization of the classifier the fusion of classification is frequently applied, consisting in different ‘classifier voting’ technologies [15]. The new complex classifiers are obtained using this method, for example, ‘maximum number of votes’ or ‘majority of votes’.

Let

$C = \{C_n,\ n \in \mathcal{N}\}$ – committee (set) of classifiers such that

$$C_n : X \to 2^{\mathcal{L}}, \quad n \in \mathcal{N}$$

that is such that

$$C_n(x) \subset \mathcal{L}, \quad n \in \mathcal{N}$$

Set of ‘voting classifiers’ for label $l \in \mathcal{L}$ shall be defined as

$$\mathcal{N}_x(l) = \{C_n \in C \mid l \in C_n(x)\}$$

Let $N_x(l)$ be the cardinality of set $\mathcal{N}_x(l)$, $l \in \mathcal{L}$

The classifier of “maximum number of votes” selected (generated) by the classifier committee shall be the classifier $C_{MG}: X \to 2^{\mathcal{L}}$, such that

$$C_{MG}(x) = \arg\max_{l\in\mathcal{L}} N_x(l) \subset \mathcal{L}, \qquad (7)$$

that is

$$C_{MG}(x) = \left\{ \overset{*}{l} \in \mathcal{L} \middle| N_x\left(\overset{*}{l}\right) = \max_{l\in\mathcal{L}} N_x(l) \right\} \subset \mathcal{L}$$

In the analyzed example, considering the first four classifiers (with no selection), we will obtain (see Table 2)

$$C_{MG}(x) = \arg\max_{l\in\mathcal{L}} N_x(l) =$$

$$\left\{ \overset{*}{l} \in \mathcal{L} \middle| N_x\left(\overset{*}{l}\right) = \max_{l\in\mathcal{L}} N_x(l) = 3 \right\} = \{l_2, l_4\} \subset \mathcal{L}$$

The other voting principle is the majority of votes principle, leading to synthesis of the "majority of votes classifier":

$G_{WG}: X \to 2^{\mathcal{L}}$

of the following exemplary form:

$$C_{WG}(x) = \left\{ l \in \mathcal{L} \middle| N_x(l) \geq \tfrac{1}{2}|C| + 1 \right\}, \qquad (8)$$

In the analyzed case:

$$C_{WG}(x) = \left\{ l \in \mathcal{L} \middle| N_x(l) \geq 3 \right\} = \{l_2, l_4\}.$$

Table 2 also lists the results of indications of these classifiers and their conformity indices with the indication obtained as a result of Pareto filtration.

The diagnosis proposals obtained on the basis of indications of classifiers may be additionally assessed using many other quality indices, including: classification error, indication ambiguity or distinctness [6, 7, 24, 28].

4. The reliability of the Pareto classifier

In the case of meta-classifiers obtained as a result of synthesis one may also define a "global quality index", so-called "indication reliability". Using the definition of the diagnostic process reliability index [7], one may propose the following formula specifying the value of the reliability index for meta-classifier $C_R(x)$:

$$\overline{w}(x, C_R(x)) = \left\| \overset{*}{y} \right\| - \left\| \overset{*}{y} - \overset{*}{y}(x) \right\|, \qquad (9)$$

where

$\overset{*}{y} = \left(\overset{*}{y}_1, ..., \overset{*}{y}_n, ..., \overset{*}{y}_N \right)$, such that

$$\overset{*}{y}_n = \max_{x\in X} \max_{l\in\mathcal{L}} f_x^n(l), \; n \in \mathcal{N}, \qquad (10)$$

the global ranking reference point (global ideal point, greatest lower bound of set *Y*). Upon normalization of the index value, we will obtain:

$$w(x, C_R(x)) = 1 - \frac{\left\| \overset{*}{y} - \overset{*}{y}(x) \right\|}{\left\| \overset{*}{y} \right\|} = 1 - \alpha\left(\left\| \overset{*}{y} - \overset{*}{y}(x) \right\| \right) \qquad (11)$$

where

$\alpha = \frac{1}{\left\| \overset{*}{y} \right\|} = \frac{1}{\sqrt{N}}$ – normalization index (for normalized ranking areas it is assumed that $\overset{*}{y} = (1, ..., 1)$).

In the analyzed example, we will obtain:

$$w(x, C_R(x)) = 1 - \tfrac{1}{\sqrt{2}} \left\| (1,1) - (0{,}6\,;0{,}7) \right\| =$$

$$= 1 - \tfrac{1}{\sqrt{2}} \cdot \sqrt{(0{,}4)^2 + (0{,}3)^2} = 1 - \tfrac{1}{\sqrt{2}} \sqrt{0{,}25} =$$

$$= 1 - \tfrac{1}{\sqrt{2}} \cdot \tfrac{1}{2} = 1 - \tfrac{1}{2\sqrt{2}} = 1 - 0{,}35 = 0{,}65$$

Thus the reliability of indication $(x, C_R(x))$ of classifier $C_R(x)$ for observation $x \in X$ in the analyzed example is only 65%. This confirms the fact that the diagnostic data used in the example and resulting from observation *x* were insufficiently expressive and unambiguous (see Table 1).

5. Summary

Analysis of the results obtained on the basis of the diagnostic example described in Clause 4 fully confirms the benefits achievable thanks to the application of classifier synthesis. Even a brief analysis of the results obtained in the example (including analysis of Figure 1) demonstrates obvious benefits resulting from synthesis of classifiers leading to an increased value of many indexes used for assessment of quality of classifiers, such as: ranking function injectivity index, ambiguity index, expressiveness index and reliability of indications [1, 6, 7, 28]. The analysis of the example also demonstrates the possible diagnostic benefits from the repeated application of the Pareto filtration leading to the extension of the initial diagnosis. Surprisingly, high conformity of the indications of "classifier voting" in a form of "maximum number of votes" (7) and "majority of votes" (8) classifiers with Pareto meta-classifier (66 % conformity with baseline indication $\mathcal{L}_x^{RN}$ (see Table 2) appeared to be an unexpected effect of the analyzed example.

6. Bibliography

[1] Acid S., Campos L.M., "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).

[2] Ameljańczyk A., "Pareto filter in the process of multi-label classifier synthesis in medical diagnostics suport algorithms", *Computer Science and Mathematical Modelling*, No. 1, 3–10 (2015).

[3] Ameljańczyk A., *Multiple optimization*, WAT, Warszawa, 1986.

[4] Ameljańczyk A., *Optymalizacja wielo-kryterialna w problemach sterowania i zarządzania*, Ossolineum, Wrocław, 1984.

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

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

[7] Ameljańczyk A., „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.

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

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

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

[11] Courtney H., Kirkland J., Viguerie P., „Strategia w warunkach niepewności", w: *Zarządzanie w warunkach niepewności*, Harvard Busines Review, 316–324, Helion, Gliwice, 2006.

[12] Dietterich G.T., "Machine learning research: four current directions", *AI Magazine*, 18, 97–136 (1997).

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

[14] Kahneman D., Tversky A., "Prospect theory: An Analysis of Decision Under Risk", *Econometrica*, Vol. 47, No. 2, 67–72 (1978).

[15] Kuncheva L.I., Whitaker C.J., "Measures of diversity in classifier ensembles", *Machine Learning*, Vol. 51, No. 2, 181–207 (2003).

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

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

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

[19] Perl J., *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, Los Angeles, California, 1988.

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

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

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

[23] Seo F., Sakawa M., *Multiple Criteria Decision Analysis in Regional Planning*, D. Reidel-Kluwer, Dordrecht-Boston-Lancaster–Tokyo, 1988.

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

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

[26] Yu P.L., Leitmann G., "Compromise solutions, domination structures and

Salukwadze's solution", *JOTA*, Vol. 13, 14–21 (1974).
[27] Yu P.L., Leitmann G., "Nondominated decision and cone convexity in dynamic multicriteria decision problems", *JOTA*, Vol. 14, 195–203 (1974).
[28] Walczak A., Bieniek B., Różyk-Jahnz K., Paluchowska E., „Fuzja klasyfikatorów w diagnostyce chorób skóry", w: *Problemy modelowania i projektowania opartych na wiedzy systemów informatycznych na potrzeby bezpieczeństwa narodowego* (red. T. Nowicki i Z. Tarapata), 143–152, WAT, Warszawa, 2014.

Analiza własności klasyfikatorów wieloetykietowych na przykładzie określania wstępnej diagnozy

A. AMELJAŃCZYK

W pracy dokonano analizy własności wstępnych diagnoz medycznych uzyskiwanych z zastosowaniem klasyfikatorów prostych i złożonych. Wprowadzone pojęcia zilustrowano obszernym przykładem z zakresu diagnostyki medycznej.

Słowa kluczowe: diagnostyka medyczna**,** klasyfikatory jednoetykietowe i wieloetykietowe, klasyfikatory proste i złożone, komitet klasyfikatorów, synteza klasyfikatorów, diagnoza niezdominowana, wiarygodność klasyfikatora.