medical aggregate data mining,
production rule with uncertainty,
internal and global rule's reliability

Magdalena SZYMKOWIAK[1], Beata JANKOWSKA[2]

RELIABILITY OF MEDICAL PRODUCTION RULES OBTAINED BY MEANS OF AGGREGATE DATA MINING

In the paper a method for designing production rules with uncertainty from medical aggregate data is proposed. Our main goal is to define the parameters that have an influence on the reliability of obtained rules. We distinguish two factors of reliability: global and internal ones. They determine a rule's importance in comparison to other obtained rules. Those rules compose the knowledge base of a medical Rule-Based System (RBS) aiding medical diagnosis and treatment.

1. INTRODUCTION

Designing knowledge bases of specialized medical Rule-Based Systems (RBSs) is the subject of our previous [4, 10] and current [5] research. The intention of RBS is to help medical doctors to make right diagnostic and therapeutic decisions concerning diverse diseases [6, 9]. These diseases could be sometimes infrequent and not very well-known to the doctors. The knowledge base of RBS will consist of production rules with uncertainty that can be generated from medical aggregate data.

In the paper [5] we present the algorithm for designing production rules. In this paper we pay attention to determining the parameters that have an influence on the reliability of generated rules. Each production rule with uncertainty is provided with two factors of reliability. These are: factor *grf(r)* of *global rule's reliability*, determining the priority of a rule in comparison to other rules from the knowledge base of RBS, and factor *irf(r)* of *internal rule's reliability*, corresponding to the conditional probability of a rule's conclusion given the certain occurrence of its premises. Factor *irf(r)* is a counterpart of the *confidence* from association rules [1]. The problem of calculating factor *grf(r)* is more complex. This factor depends on a great number of parameters, with the following being, in our opinion, the most significant: the rule's *weight*, *exactness* and *precision*.

A detailed analysis of these parameters, the method of their estimation and the way of calculating factors *grf(r)* and *irf(r)* used in the exemplary production rules with uncertainty will be the main subject of the paper.

2. RULES AS THE WAY OF KNOWLEDGE REPRESENTATION

Let $F=\{F_1,\ldots, F_w\}$ be a set of binary facts and let $D=\{d_1,\ldots, d_n\}$ be a set of individuals. Each individual d_j in D can be represented as binary vector $\{d_{j1},\ldots,d_{jw}\}$, with $d_{jk}= 1$ if for individual d_j fact F_k occurs and $d_{jk}= 0$ otherwise. Let $\{F_a,\ldots, F_b\}$ and $\{F_u,\ldots, F_v\}$ be disjoint sets of facts from F. We will consider an implication of the form:

$$r\text{: if } F_a,\ldots, F_b \text{ then } F_u,\ldots, F_v\,, \tag{1}$$

as a rule representing the knowledge that if facts-premises $F_a,\ldots, F_b$ occur then, consequently, facts-conclusions $F_u,\ldots, F_v$ occur.

[1] Poznan University of Technology, Institute of Mathematics, ul. Piotrowo 3a, 60-965 Poznań, Poland
email: magdalena.szymkowiak@put.poznan.pl .
[2] Poznan University of Technology, Institute of Mathematics, ul. Piotrowo 3a, 60-965 Poznań, Poland
email: magdalena.szymkowiak@put.poznan.pl .

2.1. NUMERICAL MEASURE OF ASSOCIATION RULE' S IMPORTANCE

In the case of an association rule [1], set F corresponds to a set of binary items and set D corresponds to a set of transactions. In binary vector $\{d_{j1},\ldots,d_{jw}\}$ value $d_{jk}= 1$ if transaction d_j bought item F_k and value $d_{jk}= 0$ otherwise. For the given set D of transaction we can determine *support of item* F_k as a number of transactions in set D that bought item F_k:

$$sup_D(F_k) = \left|D(F_k)\right|. \tag{2}$$

For each association rule r defined in formula (1), we can determine:

- *support of the rule's premises*, as a support of conjunction of items $F_a,\ldots, F_b$, it means the number of transactions in set D, that bought items $F_a,\ldots, F_b$:

$$sup_D(F_a \wedge \cdots \wedge F_b) = \left|D(F_a \wedge \cdots \wedge F_b)\right|, \tag{3}$$

- *support of the rule*, as a support of conjunction of items $F_a,\ldots, F_b$ and $F_u,\ldots, F_v$, it means the number of transactions in set D, that bought items $F_a,\ldots, F_b, F_u,\ldots, F_v$:

$$sup_D(r) = \left|D(F_a \wedge \cdots \wedge F_b \wedge F_u \wedge \cdots \wedge F_v)\right|, \tag{4}$$

- *confidence of the rule*, as a proportion of the rule's support and the support of its premises:

$$conf_D(r) = \frac{sup_D(r)}{sup_D(F_a \wedge \cdots \wedge F_b)}. \tag{5}$$

We consider the association rule, obtained as a consequence of exploring given set D of transactions, important [2] if the rule's support is above some minimum support *min_sup*, and the rule's confidence is above some minimum confidence *min_conf*.

2.2. PRODUCTION RULES WITH UNCERTAINTY

In the case of production rules discussed in [10], set F corresponds to a set of binary attributes and set D corresponds to a set of patients. In binary vector $\{d_{j1},\ldots,d_{jw}\}$ value $d_{jk}= 1$ if patient d_j possess attribute F_k and value $d_{jk}= 0$ otherwise.

The productions rules with uncertainty used in medical RBSs for automatic reasoning take the following form:

r: it happens with *grf(r)*:
if $F_a,\ldots, F_b$ (6)
then F_c with *irf(r)*,

where attributes $F_a,\ldots, F_b$, stand for the premises of rule r, and attribute F_c – for the uncertain conclusion. Such rules are additionally provided with two factors: factor *grf(r)* of global rule's reliability, determining the priority of the rule in comparison to other rules from knowledge base of RBS, and factor *irf(r)* of internal rule's reliability, corresponding to the conditional probability of attribute F_c, given the certain occurrence of attributes $F_a,\ldots, F_b$.

Contrary to association rules, low values of these factors do not necessarily reduce production rules' importance. In the case of the absence of hypotheses with high global reliability, even a less reliable one can be useful (e.g. an initial diagnostic hypothesis made by a General Practitioner). Also, a low level of internal reliability does not decide about the low importance of the rule as a whole (e.g. a hypothesis on frequency of negative, adverse effects of treatment). The method of the factors *grf(r)* and *irf(r)* estimation by means of aggregate medical data mining will be the subject of the following section.

3. FACTORS OF PRODUCTION RULE'S RELIABILITY

Each production rule with uncertainty is provided with two factors of reliability. These are: factor *grf(r)* of global rule's reliability and factor *irf(r)* of internal rule's reliability. Now we will try to determine the parameters that should have, in our opinion, an influence on these factors.

3.1. INTERNAL RULE'S RELIABILITY

We assume that each production rule *r*, is designed on the base of tuple T (of the given in [5] reference schema) being the final result of the integration of initial tuple T_1 with attached tuples $T_2,\ldots, T_m$. We notice the maximal 'attribute_count' of each tuple T_i (for $i = 1,\ldots, m$) as N_i, and the 'attribute_count' of a "special" attribute, chosen in the subset-criterion – as L_i. In tuple T number $N = \sum_{i=1}^{m} N_i$ stands for the 'attribute_count' of the common attribute (corresponding to the premises of designing rule *r*), and number $L = \sum_{i=1}^{m} L_i$ – for the 'attribute_count' of the "special" attribute (corresponding to the conclusion of designing rule *r*). Then i*nternal reliability* of rule *r* can be determined by the formula:

$$irf\ (r) = \frac{L}{N}. \tag{7}$$

It easy to notice that numbers N and L are the counterparts of the association rule's importance measure respectively: of the support of the rule's premises, defined in (3), and of the rule's support, defined in (4). Moreover, factor *irf(r)* is the counterpart of the confidence from the association rule, defined in (5). Factor *irf(r)* takes the value from range <0;1> and in statistics it is the counterpart of the point estimate of the proportion corresponding to the conditional probability of the rule's conclusion, given the certain occurrence of the rule's premises. From the point of view of the system's expert efficiency, as the important rules we will consider these rules that are characterized by high (close to 1) or low (close to 0) internal reliability rule. High level of *irf(r)* will be characteristic for the standard hypotheses for which the conclusion is highly probable (e.g. hypothesis about the consequence of standard drug admission). However, a low level of *irf(r)* will also decide about high importance of the hypothesis (e.g. hypothesis about the appearance of adverse effects of the specific pharmacotherapy). It is connected with the fact that the low probability of the event defined in the rule's conclusion implies the high probability of the opposite one. Consequently, from the point of view of the system's expert efficiency, we will consider these rules unimportant for which factor *irf(r)* is close to 0.5 . In these rules, the occurrence of the event defined in the rule's conclusion has almost the same probability as the opposite one. We will calculate *weight* of rule using the following formula:

$$w(r) = \max\{irf(r),\ 1-irf(r)\}. \tag{8}$$

From the point of view of the system's expert efficiency we will consider important the rules with high weight (close to 1). Rule's weight *w(r)* should have an influence on, defined in following subsection, global rule's reliability.

For factor *irf(r)* we can estimate $100\%\cdot(1-\alpha)$ confidence interval [7]. Parameter $1-\alpha$, known as the confidence level of the estimation, stands for the probability of the fact that the confidence interval contains the estimated factor. Since the estimation refers to the proportion, than its confidence interval should be included in interval <0; 1>. The length of this confidence interval calculated using the formula:

$$l_{1-\alpha}(r) = \min\left\{2\cdot u_{1-\frac{\alpha}{2}}\sqrt{\frac{irf(r)\cdot(1-irf(r))}{N}},\ 1\right\}, \tag{9}$$

will decide about the accurateness of the interval estimation for this factor [10]. The accurateness of interval estimation, identified farther as the rule's *exactness* can be determined by the following formula:

$$e(r) = \min\{1-\alpha,\quad 1-l_{1-\alpha}(r)\}\,. \tag{10}$$

From the point of view of the system's expert efficiency we will consider important the rules with high exactness (close to 1). It easy to notice that rule's exactness depends on number N standing for the 'attribute_count' of the common attribute in tuple T (it is bigger if number N is bigger), and on confidence level $1-\alpha$ of the interval estimation. The high (close to 1) confidence level suggested in the interval estimation implies the increasing of the interval's length. Rule's exactness *e(r)* will be the next value that should have an influence on, defined in following subsection, global rule's reliability.

3.2. GLOBAL RULE'S RELIBILITY

The problem of calculating global rule's reliability *grf(r)* seems to be very complex. As we mentioned in subsection 2.2, this factor determines the priority of the rule in comparison to other rules from the knowledge base of RBS [10].

We suggest that in the process of designing production rules, the rules in which premises and conclusion are not fuzzified should be more important ones. In these rules the attributes corresponding to the premises and the conclusion do not lose their precision during the integration. The precision of the attribute is often determined by the medicine doctor performing the clinical trials.

Let us now define the parameter of precision of fact F_k, in rule r defined by (6), for $k = a,\ldots, b, c$. This fact corresponds to an attribute in final integrated tuple T, being the result of integration of initial tuple T_1 with tuples $T_2,\ldots, T_m$ (see [5]). Each tuple T_i (for $i = 1,\ldots, m$) has the maximal 'attribute_count' of the common attributes equal to N_i. For sake of simplicity of the consideration, let us assume that each attribute corresponding to fact F_k in tuple T_i, has 'attribute_values' from countable set of value V_{ik} of cardinality $|V_{ik}|$. In final integrated tuple T, set V_k, being the set of 'attribute_value' of the attribute corresponding to fact F_k, presents:

- an union of sets V_{ik} (for $i = 1,\ldots, m$), in the case of the 'value_qualifier' of this attribute taking the form of disjunction $\oplus$ or,
- an intersection of sets V_{ik} (for $i = 1,\ldots, m$), in the case of the 'value_qualifier' of this attribute taking the form of conjunction $\odot$.

Then we can determine *precision* of fact F_k using the following formula:

$$v(F_k) = \begin{cases} \sum_{i=1}^{m} \frac{N_i}{N} \cdot \frac{|V_{ik}|}{|V_k|} & \text{for the disjunction} \\ \sum_{i=1}^{m} \frac{N_i}{N} \cdot \frac{|V_k|}{|V_{ik}|} & \text{for the conjunction} \end{cases}, \tag{11}$$

where $N = \sum_{i=1}^{m} N_i$ stands for the 'attribute_count' of the common attributes in final integrated tuple T. The precision of the fact takes the value from range (0; 1> and it gets the maximal value 1 if the corresponding attribute is not fuzzified in none of the integrated tuples. During the estimation of this parameter we have to pay attention to the maximal 'attribute_count' of each integrated tuple which will decide about the power of this tuple's influence on the precision of the fact.

Average of the rule's precision v(r) for the rule defined by (5) can be calculated by the formula:

$$v(r) = \frac{\sum_{k=a}^{b} v(F_k) + v(F_c)}{b-a+2}, \tag{12}$$

This parameter takes values from range (0; 1> and gets the maximal value 1 if none of the attributes – corresponding to the rule's conclusion and premises – is fuzzified.

And finally, we propose to estimate *global rule's reliability grf(r)* as the minimal value of the parameters: rule's weight *w(r)* defined by (8); rule's exactness *e(r)* defined by (10); and average of the rule's precision *v(r)* defined by (12):

$$grf(r) = \min\{w(r), e(r), v(r)\}. \quad (13)$$

This means that we will consider rule *r* as the rule with high global reliability if this rule has, at the same time, the high weight, exactness and precision.

4. EXAMPLES OF DESIGNING PRODUCTION RULES

The following example will illustrate the designing of production rules with uncertainty, especially the method of estimation of the rules' internal and global reliabilities. The data came from a medical repository, namely the repository of clinical trials registers.

4.1. DESIGNING PRODUCTION RULES WITH UNCERTAINTY

All the data we consider, refer to young patients hospitalized for the bronchial asthma exacerbation [8]. The data report the results of clinical trials carried out on three groups of patients. They can be represented by means of the following tuples:

T_1 = <General_Diagnosis:{pediatric_asthma}⊙/17,
Current_Health_state:{acute_asthma_exacerbation}⊙/17,
Standard_Drug:{short-acting_beta2_agonist}⊙/17,
Additional_Drug:{inhaled_anticholin_multi_doses}⊙/17,
co_intervention:{systemic_corticosteroid}⊙/17,
age_range:{1,…, 7} ⊕/17,
severity_of_diagn_illness:{mild, moderate}⊕/17,
symptoms: {coughing}⊙/17,
treatment_effects:{no_hospital_admission}⊙/13,
adverse_effects: {vomiting}⊙/3,
relapse: {next_asthma_exacerbation_in_72_hours}⊙/1> ;

T_2 = <General_Diagnosis:{pediatric_asthma, diabetes}⊙/18,
Current_Health_State:{acute_asthma_exacerbation}⊙/18,
Standard_Drug:{short-acting_beta2_agonist}⊙/18,
Additional_Drug:{inhaled_anticholin_multi_doses}⊙/18,
Co_Intervention:{systemic_corticosteroid}⊙/18,
age_range:{4,..., 9}⊕/18,
severity_of_diagn_illness:{moderate}⊕/18,
symptoms: {coughing, wheezing}⊙/18,
treatment_effects:{no_hospital_admission, stability_of_FEV1}⊙/18,
adverse_effects: {vomiting}⊙/2> ;

T_3 = <General_Diagnosis:{pediatric_asthma}⊙/89,
Current_Health_State:{acute_asthma_exacerbation,
asthma_attack }⊙/89,
Standard_Drug:{short-acting_beta2_agonist}⊙/89,
Additional_Drug:{inhaled_anticholin_multi_doses}⊙/89,

T = <General_Diagnosis:{pediatric_asthma}⊙/124,
Curent_Health_State:{acute_asthma_exacerbation}⊙/124,
Standard_Drug:{short-acting_beta2_agonist}⊙/124,
Additional_Drug:{inhaled_anticholin_multi_doses}⊙/124,
age_range:{1,…, 18}⊕/124,
severity_of_diagn_illness:{mild, moderate, severe}⊕/124,
symptoms: {coughing}⊙124,
treatment_effects:{no_hospital_admission}⊙/101
adverse_effects: {vomiting}⊙/10> .

For final integrated tuple T and the "special" attribute: treatment_effects, chosen in the subset-criterion (a), the following production rule with uncertainty will be obtained:

r_a: it happens with $grf(r) = 0.81$:

if (pediatric_asthma) and
(acute_asthma_exacerbation) and
(short-acting_beta2_agonist) and
(inhaled_anticholinergic_multi_doses) and
(age_range = {1,…, 18}) and
(severity_of_diagn_illness =
= (mild or moderate or severe)) and
(coughing)
then (no_hospital_admission) with $irf(r) = 0.81$.

For final integrated tuple T and the "special" attribute: adverse_effects, chosen in the subset-criterion (b), the following production rule with

co_intervention:{no_corticosteroid}⊙/89,
age_range:{6,…, 18} ⊕/89,
severity_of_diagn_illness:{mild, moderate, severe}⊕/89,
symptoms: {coughing}⊙/89,
treatment_effects:{no_hospital_admission}⊙/70>
adverse_effects: {vomiting}⊙/5>.

We assume that T_1 is the initial tuple of the integration and for this tuple we determine two subset-criteria:

(a) $K_1 \cup \{\} \cup \{\text{treatment\_effects}\}$,

(b) $K_1 \cup \{\text{symptoms}\} \cup \{\text{adverse\_effects}\}$.

We can integrate all three tuples T_1, T_2 and T_3, to both criteria and we obtain the following final integrated tuple T:

uncertainty will be obtained:

r_b: it happens with $grf(r) = 0.87$:

if (pediatric_asthma) and
(acute_asthma_exacerbation) and
(short-acting_beta2_agonist) and
(inhaled_anticholinergic_multi_doses) and
(age_range = {1,…, 18}) and
(severity_of_diagn_illness =
= (mild or moderate or severe)) and
(coughing)
then (vomiting) with $irf(r) = 0.08$.

Reliability factors *irf(r)* and *grf(r)*, given to these rules were calculated by means of formulas (7) and (13), under the assumption of confidence interval level $1-\alpha = 0.95$. A method of calculation of these factors and their detailed analysis will be performed in the following subsections.

4.2. CALCULATION OF THE RELIABILTY FACTORS USED IN EXEMPLARY PRODUCTION RULES

Rules r_a and r_b presented in subsection 4.1 were generated from final integrated tuple T. In each tuple: T_1, T_2 and T_3 the maximal 'attribute_count' of common attributes is equal respectively: $N_1 = 17$, $N_2 = 18$, $N_3 = 89$ and the 'attribute_count' of attributes: treatment_effects, and adverse_effects chosen in the subset-criteria (a) and (b) as the "special" ones, is equal respectively: $L_{1a} = 13$, $L_{2a} = 18$, $L_{3a} = 70$ and $L_{1b} = 3$, $L_{2b} = 2$, $L_{3b} = 5$. In final integrated tuple T the maximal 'attribute_count' of common attributes (corresponding to the rule's premises) is equal $N = 124$, and the 'attribute_count' of the "special" attribute (corresponding to the rule's conclusion) is, in the subset-criterion (a), equal $L_a = 101$ and in the subset-criterion (b), $L_b = 10$. For each rules r_a and r_b, its internal reliability defined by formula (7), is equal respectively: $irf(r_a) = 0.81$ and $irf(r_b) = 0.08$, and its weight, defined by formula (8), is equal respectively: $w(r_a) = 0.81$ and $w(r_b) = 0.92$.

Next, for each determined factors $irf(r_a)$ and $irf(r_b)$, we can estimate 95% confidence interval and calculate its length by formula (9). Then for each rule r_a and r_b we obtain its exactness, defined by formula (10), which is equal respectively: $e(r_a) = 0.86$ and $e(r_b) = 0.9$.

Let us notice that both rules r_a and r_b have the same set of premises. The precisions of these premises defined by formula (11) are equal respectively: $v(\text{pediatric\_asthma}) = 0.93$, $v(\text{acute\_asthma\_exacerbation}) = 0.64$, $v(\text{short-acting\_beta2\_agonist}) = 1$, $v(\text{inhaled\_anticholinergic\_multi\_doses}) = 1$, $v(\text{age\_range = \{1,…,18\}}) = 0.62$, $v(\text{severity\_of\_diagn\_illness = (mild or moderate or severe)}) = 0.86$, $v(\text{coughing}) = 0.93$.

Let us also demonstrate the method of calculating these parameters for two different premises, the first one (age_range = {1,…,18}) corresponding to the attribute age_range with the 'value_qualifier' taking the form of disjunction ⊕, and the second one (coughing) corresponding to the attribute symptoms with the 'value_qualifier' taking the form of conjunction ⊙.

For premise F_k = (age_range = {1,…,18}) the set values of the corresponding attribute age_range in tuples T_1, T_2 and T_3 are respectively: $V_{1k} = \{1,\ldots,7\}$, $V_{2k} = \{4,\ldots,9\}$ and $V_{3k} = \{6,\ldots,18\}$ of the cardinalities equal: $|V_{1k}| = 7$, $|V_{2k}| = 6$, $|V_{1k}| = 13$. In final integrated tuple T, the "updated" set value of the attribute age_range is $V_k = \{1,\ldots,18\}$ of the cardinality equal $|V_k| = 18$. Then, using formula (11), we can calculate: $v(\text{age\_range = \{1,…,18\}}) = \frac{17}{124}\cdot\frac{7}{18}+\frac{18}{124}\cdot\frac{6}{18}+\frac{89}{124}\cdot\frac{13}{18} = 0.62$.

For premise F_k = (coughing) the set values of the corresponding attribute symptoms in tuples T_1, T_2 and T_3 are respectively: $V_{1k} = \{\text{coughing}\}$; $V_{2k} = \{\text{coughing, wheezing}\}$ and $V_{3k} = \{\text{coughing}\}$ of the cardinalities equal: $|V_{1k}| = 1$,

$|V_{2k}| = 2$, $|V_{3k}| = 1$. In final integrated tuple T the set value of the attribute symptoms is V_k = {coughing} of the cardinality $|V_k| = 1$. Then using formula (11), we can calculate: $v(\text{coughing}) = \frac{17}{124}\cdot\frac{1}{1}+\frac{18}{124}\cdot\frac{1}{2}+\frac{89}{124}\cdot\frac{1}{1} = 0.93$.

Both rules r_a and r_b differ in their conclusions, and the precision of these conclusions are equal: $v(\text{no\_hospital\_admission}) = 0.64$, $v(\text{vomiting}) = 1$. The average of the precision for rules r_a and r_b, defined by formula (12), is equal respectively: $v(r_a) = 0.83$ and $v(r_b) = 0.87$.

Finally, global reliability of rules r_a and r_b, defined by formula (13), is equal respectively:
$grf(r_a) = \min\{0.81,0.86,0.83\} = 0.81$ and $grf(r_b) = \min\{0.92,0.9,0.87\} = 0.87$.

4.3. ANALYSIS OF THE RELIABILTY FACTORS USED IN EXEMPLARY PRODUCTION RULES

Let us notice that, however, in rules r_a and r_b internal reliability $irf(r_b)$ is much smaller than internal reliability $irf(r_a)$, global reliability $grf(r_b)$ is higher than $grf(r_a)$. This means that in the knowledge base of RBS with uncertainty, rule r_b will have the higher priority in comparison to rule r_a. In the case of rule r_a, the significant parameter for determination of $grf(r_a)$ is its weight $w(r_a)$, whereas in the case of rule r_b the significant parameter for determination of $grf(r_b)$ is its average precision $v(r_b)$. In both cases rules r_a and r_b, their exactness is high (close to 1) because of large maximal 'attribute_count' N, in final integrated tuple T.

5. CONCLUSIONS

Designing knowledge bases of specialized medical Rule-Based Systems (RBSs), aiding medical doctors in their everyday practice, while taking care of patients and making treatment decisions, is the subject of our research. In the paper we presented the method of designing production rules that compose the knowledge base of RBS. We paid attention to determining the parameters that have an influence on reliability of generated rules. So far, we have considered the confidence of the rule, its weight, exactness and precision as the significant for the rule's reliability. Attempts to find other parameters that could have an influence on global reliability of the rule will be the subject of our future research.

We consider the possibility of using the proposed method for individual patients' data mining. Unfortunately, the possibility of gaining access to such data, despite numerous talks with the medicine doctors, is still a big problem. Nevertheless, the possibility of the data coding, by means of standard electronic notations, such as HL7 [3] or EHR, that can guarantee, among others, the full anonymity of patients, seems to be very hopeful.

ACKNOWLEDGEMENTS

The research has been partially supported by the Polish Ministry of Science and Higher Education under grant N516 369536 and by Poznan University of Technology under grants PB 43-070/10 and DS 45-083/10.

BIBLIOGRAPHY

[1] AGRAVAL R., IMIELIŃSKI T., SWAMI A., Database mining: A performance perspective, IEEE Transactions on Knowledge Engineering, Vol. 5(6), 1993, pp. 914-925.
[2] CICHOSZ P., Learning systems, WNT, Warszawa, 2007.
[3] DOLIN R.H., ALSCHULER L., BOYER S., BEEBE C., Kona Editorial Group., An Update on HL7's XML-based Document Representation Standards. Proceedings of AMIA Symp., 2000, pp. 190-194.
[4] JANKOWSKA B., SZYMKOWIAK M., How to Acquire and Structuralize Knowledge for Medical Rule-Based Systems? Studies in Computational Intelligence, J. Kacprzyk (ed.), Springer Series, Berlin /Heidelberg , 2008, pp. 115-130.

[5] JANKOWSKA B., SZYMKOWIAK M., Designing Medical Production Rules from Semantically Integrated Data, submitted to the conf., MIT 2010.
[6] LUCAS P.J.F., SEGAR R.W., JANSSENS A.R., HEPAR: an expert system for diagnosis and disorders of the liver and biliary tract, Liver 9, 1989, pp. 266-275.
[7] PETRIE A., SABIN C., Medical Statistics at a Glance., Blackwell Science Ltd, London, 2000.
[8] PLOTNICK L.H., DUCHARME F.M., Combined inhaled anticholinergics and beta2-agonists for initial treatment of acute asthma in children., The Cochrane Library, 2005.
[9] SHORTLIFFE E., Computer-Based Medical Consultations: MYCIN, American Elsevier, 1976.
[10] SZYMKOWIAK M., JANKOWSKA B., Discovering Medical Knowledge from Data in Patients' Files. ICCCI Wrocław 2009, LNAI 5796, Springer-Verlag, Berlin/Heidelberg, 2009, pp. 128-139.