medical aggregate data mining,
2-uncertain rule, reasoning chain,
internal and global reliability

Magdalena SZYMKOWIAK[1]

SOME EXAMPLES OF REASONING WITH 2-UNCERTAIN RULES

In the paper some examples of reasoning with 2-uncertain rules are presented. First of all, they will illustrate the method for designing 2-uncertain rules from medical aggregate data. The obtained rules compose the knowledge base of a medical Rule-Based System (RBS) aiding medical diagnosis and treatment. For each obtained rule two determined factors of rules' reliability – global and internal ones – will rank it in the designed RBS. Furthermore, the presented examples will realize the influence of the reliability factors on the process of uncertain reasoning.

1. INTRODUCTION

Designing production rules with uncertainty from medical aggregate data is the subject of our previous [9, 10] and current [2, 3] research. The obtained rules compose the knowledge base of a medical Rule-Based System (RBS). The intention of the RBS is to help medical doctors to make right diagnostic and therapeutic decisions concerning diverse diseases [5, 8].

In the paper we present some examples of reasoning with 2-uncertain rules. The model for 2-uncertain rules (see [2]) is based on a classical implication, provided with two reliability factors: *internal rule reliability* (irf), stating the conditional probability of a rule's conclusion, given a certain occurrence of the rule's premises, and *global rule reliability* (grf), stating the external rule's reliability, determining the priority of the obtained rule in the designed RBS (see [3]).

The example illustrates the process of designing 2-uncertain rules on the basis of virtual data being the result of the integration of real aggregate medical data. We demonstrate that at the same time, data integration influences the rule's reliability positively (the number of patients 'caught' in the rule) and negatively (a decrease in the accuracy of the patients' attributes). In the paper we establish the influence of factor grf on the process of uncertain reasoning and calculating irf value for a hypothesis concluded by a reasoning chain. We point out that firing, in the reasoning chain, the low priority 2-uncertain rule can result in an unreasonable revision of the reliability of the concluded hypothesis.

2. EXAMPLES OF RANKING 2-UNCERTAIN RULES

The following examples will illustrate the process of ranking 2-uncertain rules. The data came from a medical repository, namely the repository of clinical trials registers.

2.1. DESIGNING 2-UNCERTAIN RULES

All the data we consider refer to patients hospitalized for the bronchial asthma exacerbation [7]. The data report the results of clinical trials carried out on three groups of patients. The first group consists of children between 1 and 20 years old, the second one – of adults between 41 and 60 years old and the last one – of patients with wide-ranging ages, between 11 and 50 years old. The data can be represented by means of the following tuples (see [10]):

T_1 = <General_Diagnosis = {asthma}⊙/108,
Current_Health_State = {acute_asthma_exacerbation, asthma_attack}⊙/108,
Drug = {short-acting_beta2_agonist, systemic_corticosteroid, inhaled_anticholin}⊙/108,

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

age_range = {1,…, 20}⊕/108,
severity_of_diagn_illness = {severe}⊕/108,
symptoms = {coughing, shortness_of_breath}⊙/108,
treatment_effects = {no_hospital_admission}⊙/82>;

T_2 = <General_Diagnosis = {asthma, diabetes}⊙/147,
Current_Health_State = {acute_asthma_exacerbation, asthma_attack}⊙/147,
Drug = {short-acting_beta2_agonist, systemic_corticosteroid, inhaled_anticholin}⊙/147,
age_range = {41,...,60}⊕/147,
severity_of_diagn_illness = {mild, moderate}⊕/147,
symptoms = {coughing, wheezing, shortness_of_breath}⊙/147,
treatment_effects = {no_hospital_admission}⊙/128> ;

T_3 = <General_Diagnosis = {asthma}⊙/189,
Current_Health_State = {acute_asthma_exacerbation}⊙/189,
Drug = {short-acting_beta2_agonist, systemic_corticosteroid}⊙/189,
age_range = {11,…, 50}⊕/189,
severity_of_diagn_illness = {mild, moderate, severe}⊕/189,
symptoms = {coughing, wheezing}⊙/189,
treatment_effects = {no_hospital_admission}⊙/160>.

In each tuple T_1, T_2 and T_3 patients are the same with respect to General_Diagnosis, Current_Health_State, Drug, age_range, severity_of_diagn_illness, symptoms, co_intervention (common attributes), and they differ with respect to treatment_effects (a discriminatory attribute). As we can see, tuples T_1 and T_2 describe more precisely their groups of patients than tuple T_3. Among others, attributes Current_Health_State, Drug, age_range, severity_of_diagn_illness are determined more accurately in tuples T_1, T_2 than in tuple T_3.

First, we assume that T_1 is the initial tuple of the integration. In such a situation, we cannot integrate tuple T_1 with tuple T_2 or tuple T_3. In this case, we treat tuple T_1 as final tuple T_a of the integration (T_a=T_1) and we obtain the following 2-uncertain rule:

r_a : it happens with grf = 0.84:
if General_Diagnosis = {asthma}⊙ and
Current_Health_State = {acute_asthma_exacerbation, asthma_attack}⊙ and
Drug = {short-act_beta2_agonist, systemic_corticosteroid, inhaled_anticholin}⊙ and
age_range = {1,…,20}⊕ and
severity_of_diagn_illness = {severe}⊕ and
symptoms = { wheezing, shortness_of_breath}⊙
then treatment_effects = {no_hosp_admission}⊙ with irf = 0.76

Next, we assume that T_2 is the initial tuple of the integration. In this situation, similarly as before, we cannot integrate tuple T_2 with tuple T_1 or tuple T_3. In this case, we treat tuple T_2 as final tuple T_b of the integration (T_b=T_2) and we obtain the following 2-uncertain rule:

r_b : it happens with grf = 0.89:
if General_Diagnosis = {asthma, diabetes}⊙ and
Current_Health_State = {acute_asthma_exacerbation, asthma_attack }⊙ and
Drug = {short-act_beta2_agonist, systemic_corticosteroid, inhaled_anticholin}⊙ and
age_range = {41,…,60}⊕ and
severity_of_diagn_illness = {mild, moderate}⊕ and
symptoms = {coughing, wheezing, shortness_of_breath}⊙
then treatment_effects = {no_hosp_admission}⊙ with irf = 0.87

Finally, we assume that T_3 is the initial tuple of the integration. In this situation, we can integrate tuple T_3 with tuple T_1 and tuple T_2. As a result of the integration we obtain the following final integrated tuple T_c :

T_c = <General_Diagnosis = {asthma}⊙/444,
Current_Health_State = {acute_asthma_exacerbation}⊙/444,
Drug = {short-acting_beta2_agonist, systemic_corticosteroid}⊙/444,
age_range = {1,…, 60}⊕/444,
severity_of_diagn_illness = {mild, moderate, severe}⊕/444,
symptoms = {wheezing}⊙/444,
treatment_effects = {no_hospital_admission}⊙/370>.

For tuple T_c the following 2-uncertain rule will be obtained:

r_c : it happens with grf = 0.57:
if General_Diagnosis = {asthma}⊙ and
Current_Health_State = {acute_asthma_exacerbation}⊙ and
Drug = {short-act_beta2_agonist, systemic_corticosteroid}⊙ and
age_range = {1,…,60}⊕ and
severity_of_diagn_illness= {mild, moderate, severe}⊕ and
symptoms = {wheezing}⊙
then treatment_effects = {no_hosp_admission}⊙ with irf = 0.83

Further on, we will discuss the calculation of *internal reliability factor* irf and *global reliability factor* grf for the above rules. Let us notice now, that as we said before, tuples T_1 and T_2 describe precisely their groups of patients. If we treat these tuples as the initial tuples of the integration, it is not easy to integrate them with others tuples and as a result of the integration we obtain final tuples with relatively small 'attribute_count' of the common attribute – a small number of 'caught' patients. This fact will evoke a decrease in the global reliability of obtained rules r_a and r_b.

2.2. CALCULATING THE FACTOR OF INTERNAL RULE'S RELIABILITY

First, we estimate *internal reliability factor* irf of the rule from the following formula (see [9]):

$$\text{irf} = \frac{L}{N}, \tag{1}$$

where $N = \sum_{i=1}^{m} N_i$ (the number of patients 'caught' in the rule) stands for the 'attribute_count' of the common attribute in final tuple T (corresponding to the premises of the rule) and number $L = \sum_{i=1}^{m} L_i$ stands for the 'attribute_count' of the chosen discriminate attribute in final tuple T (corresponding to the conclusion of the rule). Let us recall that final tuple T is obtained as a result of the integration of tuples T_i (for i = 1,..., m), where in each tuple T_i, the 'attribute_count' of common attributes is equal to N_i and the 'attribute_count' of the chosen discriminate attribute – to L_i. Factor irf is the counterpart of the confidence from association rules [1]. In statistics, irf 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 [4]. Let us recall that we assume positive monotonic dependence between a rule's premises and a rule's conclusion – the lower the level of fulfillment of the premises, the lower the level of fulfillment of the conclusion (see [2]).

Therefore, from formula (1), the internal reliabilities of exemplary rules r_a, r_b and r_c are equal:

$\text{irf}_{r_a} = \frac{82}{108} = 0.76$ $\text{irf}_{r_b} = \frac{128}{147} = 0.87$ and $\text{irf}_{r_c} = \frac{370}{444} = 0.83$, respectively.

It means that conclusion treatment_effects = {no_hosp_admission}⊙ has, for the group of patients between 41 and 60 years old, an evidently higher probability, than for the group of patients between 1 and 21 years old. It is connected with the fact that older patients, suffering from asthma, need to be hospitalized more rarely then younger ones.

Moreover, let us notice that the negative form of the rules' conclusion treatment_effects = {no_hosp_admission}⊙, guaranties a positive monotonic dependence between the rules' premises and the rules' conclusion – the lower the level of fulfillment of the premises of rule r_a, r_b and r_c, respectively, the lower the level of fulfillment of the conclusion treatment_effects = {no_hosp_admission}⊙. The property of positive monotonic dependence is not guaranteed for the opposite, positive conclusion treatment_effects = {hosp_admission}⊙.

Let us recall that in our study, a rule with extreme irf (close to 1 or close to 0) is regarded as a rule with a 'characteristic' conclusion (see [3]). The fact of being a rule with a 'characteristic' conclusion will cause an increase in the global reliability of this rule.

2.3. CALCULATING RULE'S WEIGHT

Let us now define a rule's *weight* wg – the parameter that will have an influence on the global reliability of a rule. It will point out an importance of the rule being distinguished simultaneously, by a high number of patients 'caught' in the rule and by a 'characteristic' conclusion. First from the following formula:

$$l_{1-\alpha} = 2 \cdot u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\mathrm{irf} \cdot (1-\mathrm{irf})}{N}} \tag{2}$$

we calculate lengths $l_{0.95,\,r_a}$, $l_{0.95,\,r_b}$ and $l_{0.95,\,r_c}$ of the confidence intervals [6] for factors irf_{r_a}, irf_{r_b} and irf_{r_c}, respectively. They are as follows:

$l_{0.95,\,r_a} = 0.16$, $l_{0.95,\,r_b} = 0.11$ and $l_{0.95,\,r_c} = 0.07$.

Then, by means of the following formula (see [10]):

$$\mathrm{wg} = \min\{1 - l_{0.95},\ 0.95\} \tag{3}$$

we determine their *weight* for the exemplary rules r_a, r_b and r_c:

wg_{r_a} = min {0.84, 0.95} = 0.84, wg_{r_b} = min {0.89, 0.95} = 0.89 and wg_{r_c} = min {0.93, 0.95} = 0.93.

This means that rule r_c has the highest weight compared to rules r_a and r_b. It is connected with the fact that rule r_c was obtained from the tuple with the largest number of 'caught' patients (N=444) and with a relatively 'characteristic' conclusion ($\mathrm{irf}_{r_c} = 0.83$). Rule r_a has the smallest weight compared to rules r_b and r_c, because it was obtained from the tuple with the smallest number of 'caught' patients (N=108) and with the least 'characteristic' conclusion ($\mathrm{irf}_{r_a} = 0.76$).

2.4. CALCULATING RULE'S ACCURACY

Now we will precisely discuss the second parameter having, in our opinion, an influence on global rule reliability. It is *accuracy of rule* ac. To establish this parameter, we have to estimate a rule's relative and objective accuracy. *Relative accuracy of rule* rac will express a decrease in the accuracy of virtual data being designed and *objective accuracy of rule* oac will express objective precision of a rule's premises and conclusion.

To estimate *relative accuracy of rule* rac, first, for each attribute A_k, we will determine parameter $\mathrm{rat}(A_{ki})$ enabling to express *relative accuracy of attribute* in integrated tuple T_i in comparison to final tuple T. This parameter is defined as follows (see [3]):

$$\text{rat}(A_{ki}) = \begin{cases} \dfrac{|V_{ki}|}{|V_k|} & \text{for} \quad \text{disjunction} \ \oplus \\ \dfrac{|V_k|}{|V_{ki}|} & \text{for} \quad \text{conjunction} \ \odot \\ 1 & \text{for} \quad |V_{ki}| = |V_k| = 0 \end{cases} . \tag{4}$$

Next, for each potential fact F_k of designed rule (that corresponds to attribute A_k in final tuple T), we will estimate parameter $\text{raf}(F_k)$ determining *relative accuracy of fact* F_k. Apart from relative accuracies of the corresponding attribute in each integrated tuple T_i, we have to pay attention to maximal 'attribute_count' N_i of each integrated tuple T_i, which will decide about the power of this tuple's influence on $\text{raf}(F_k)$. This parameter can be calculated from the following formula (see [3]):

$$\text{raf}(F_k) = \frac{\sum_{i=1}^{m} N_i \cdot \text{rat}(A_{ki})}{N} \tag{5}$$

Lastly, for each rule r obtained from final tuple T we can define parameter rac determining *relative accuracy of rule*. If we assume that in tuple T, z attributes A_k (for $1 \le k \le z$) correspond to z potential facts F_k, being the potential premises or the conclusion in the obtained rule, then parameter rac can be calculated as the arithmetic mean of its facts' relative accuracies:

$$\text{rac} = \frac{1}{z} \sum_{k=1}^{z} \text{rf}(F_k) \tag{6}$$

Let us now determine the relative accuracy of exemplary rules r_a, r_b and r_c. Let us recall that for rule r_a and r_b, initial tuples T_1 and T_2 are treated as final tuples T_a and T_b of the integration. It means that for rule r_a and r_b no fact decreases its accuracy during the integration. Therefore, for obtained rules r_a and r_b, for each fact F_k parameter $\text{raf}(F_k)=1$ and so on, we can estimate parameters $\text{rac}_{r_a} = 1$ and $\text{rac}_{r_b} = 1$.

In order to determine rac_{r_c} – the relative accuracy of rule r_c obtained from tuple T_c (being the result of the integration of the initial tuple T_3 with tuples T_1 and T_2), first, for each attribute A_k from tuple T_c, using formula (4) we estimate its relative accuracy $\text{rat}(A_{ki})$ in integrated tuple T_i (for $i=1,2,3$) in comparison to final tuple T_c. Therefore, for each attribute A_k from tuple T_c, we determine cardinalities $|V_{ki}|$ of sets of its 'attribute_values' in each integrated tuple T_i and cardinality $|V_k|$ of a set of its 'attribute_values' in tuple T_c.

Moreover, according to formula (5), in order to establish, for each fact F_k, its relative accuracy $\text{raf}(F_k)$, we also have to pay our attention to maximal 'attribute_count' N_i of each integrated tuple T_i. The following Table 1. contains all data required to establish parameter rac_{r_c} :

Table 1. Cardinalities of 'attribute_values' sets.

T_1, N_1=108		T_2, N_2=147		T_3, N_3=189		T_c, N=444	
V_{13}={asthma}⊙	$\|V_{13}\|$=1 $\text{rat}(A_{11})$=1	V_{12}={asthma, diabetes}⊙	$\|V_{12}\|$=2 $\text{rat}(A_{12})$=0.5	V_{11}={asthma}⊙	$\|V_{13}\|$=1 $\text{rat}(A_{13})$=1	V_1={asthma}⊙	$\|V_1\|$=1 $\text{raf}(F_1)$=0.84
V_{21}={acute_asthma_exacerb asthma_attack}⊙	$\|V_{21}\|$=2 $\text{rat}(A_{21})$=0.5	V_{22}={acute_asthma_ exacerb, asthma_attack}⊙	$\|V_{22}\|$=2 $\text{rat}(A_{22})$=0.5	V_{23}={acute_asthma_ exacerb}⊙	$\|V_{23}\|$=1 $\text{rat}(A_{23})$=1	V_2={acute_asthma_exacerb}⊙	$\|V_2\|$=1 $\text{raf}(F_2)$=0.7
V_{31}={short-act_beta2_agonist, systemic_corticosteroid inhaled_anticholin}⊙	$\|V_{31}\|$=3 $\text{rat}(A_{31})$=0.67	V_{32}={short-act_beta2_agonist, systemic_corticosteroid inhaled_anticholin}⊙	$\|V_{32}\|$=3 $\text{rat}(A_{32})$=0.67	V_{33}={short-act_beta2_agonist, systemic_corticosteroid}⊙	$\|V_{33}\|$=2 $\text{rat}(A_{33})$=1	V_3={short-act_beta2_agonist, systemic_corticosteroid}⊙	$\|V_3\|$=2 $\text{raf}(F_3)$=0.8

V_{41}={1 ,..., 21}⊕	$\|V_{41}\|$=20 rat(A_{41})=0.33	V_{42}={41 ,..., 60}⊕	$\|V_{42}\|$=20 rat(A_{42})=0.33	V_{43}={11 ,..., 50}⊕	$\|V_{43}\|$=40 rat(A_{41})=0.67	V_4={1 ,..., 60}⊕	$\|V_4\|$=60 raf(F_4)=0.48
V_{51}={severe}⊕	$\|V_{51}\|$=1 rat(A_{51})=0.33	V_{52}={mild, moderate}⊕	$\|V_{52}\|$=2 rat(A_{52})=0.67	V_{53}={mild, moderate, severe}⊕	$\|V_{53}\|$=3 rat(A_{53})=1	V_5={ mild, moderate, severe}⊕	$\|V_5\|$=3 raf(F_5)=0.73
V_{61}={ wheezing, shortness_of_breath}⊙	$\|V_{61}\|$=2 rat(A_{61})=0.5	V_{62}={coughing, wheezing, shortness_of_breath}⊙	$\|V_{62}\|$=3 rat(A_{62})=0.33	V_{63}={coughing, wheezing}⊙	$\|V_{63}\|$=2 rat(A_{63})=0.5	V_6={ wheezing}⊙	$\|V_6\|$=1 raf(F_6)=0.45
V_{71}={no_hosp_admission}⊙	$\|V_{71}\|$=1 rat(A_{71})=1	V_{72}={no_hosp_admission}⊙	$\|V_{72}\|$=1 rat(A_{72})=1	V_{73}={no_hosp_admission}⊙	$\|V_{73}\|$=1 rat(A_{73})=1	V_7={no_hosp_admission}⊙	$\|V_7\|$=1 raf(F_8)=1

Then, from formula (6), relative accuracies of exemplary rules r_a , r_b and r_c are equal:
(as we mentioned above, $rac_{r_a}=1$ and $rac_{r_b}=1$) and $rac_{r_c}=0.71$.

Low relative accuracy of rule r_c is connected, first of all, with the evident decrease in accuracy of attribute A_4=age_range and attribute A_6=symptoms, which takes place while integrating.
To estimate parameter rac determining objective accuracy of a rule, first, we have to establish domains for each attribute corresponding to the potential rule's fact. We assume that for rules r_a , r_b and r_c domains of the corresponding attributes, are defined as follows:

A_1=General_Diagnosis	D_1={asthma, diabetes}⊙	$\|D_1\|$=2
A_2= Current_Health_State	D_2= {acute_asthma_exacerbation, asthma_attack}⊙	$\|D_2\|$=2
A_3=Drug	D_3={short-act_beta2_agonist, systemic_corticosteroid inhaled_anticholin}⊙	$\|D_3\|$=3
A_4=age_range	D_4={1 ,..., 100}⊕	$\|D_4\|$=100
A_5=severity_of_diagn_illness	D_5={intermittent, mild, moderate, severe}⊕	$\|D_5\|$=4
A_6=symptoms	D_6={coughing, wheezing, shortness_of_breath }⊙	$\|D_6\|$=3
A_7=treatment_effects	D_7={no_hosp_admission, stab_of_FEV1, good_sleep_in_night }⊙	$\|D_7\|$=3

Furthermore, for each designed rule r_a , r_b and r_c, for each rule's fact, by comparison of cardinality $|V_k|$ of the set values of corresponding attribute A_k in final tuple T and cardinality $|D_k|$ of this attribute's domain, we will estimate parameter oaf(F_k), determining *objective accuracy of fact* F_k. This calculation is carried out by the following formula:

$$oaf(F_k)=\begin{cases}1-\dfrac{|V_k|}{|D_k|} & \text{for disjunction } \oplus\\ \dfrac{|V_k|}{|D_k|} & \text{for conjunction } \odot\end{cases}. \qquad (7)$$

Let us recall that for rules r_a and r_b , initial tuples T_1 and T_2 are treated as final tuples T_a and T_b of the integration. This means that, for each fact F_k:
for rules r_a, $|V_k|=|V_{k1}|$ (determined in the first column of Table 1.) and by formula (7):
oaf(F_1)=0.5 oaf(F_2)=1 oaf(F_3)=1 Oaf(F_4)=0.8 Oaf(F_5)=0.75 oaf(F_6)=0.67 oaf(F_7)=0.33;
for rules r_b, $|V_k|=|V_{k2}|$ (determined in the second column of Table 1.) and by formula (7):
oaf(F_1)=1 oaf(F_2)=1 oaf(F_3)=1 Oaf(F_4)=0.8 Oaf(F_5)=0.5 oaf(F_6)=1 oaf(F_7)=0.33;
for rules r_c, cardinality $|V_k|$ is determined in the last column of Table 1. and by formula (7):
oaf(F_1)=0.5 oaf(F_2)=0.5 oaf(F_3)=0.67 oaf(F_4)=0.4 Oaf(F_5)=0.25 oaf(F_6)=0.33 oaf(F_7)=0.33.

Next for each rule r obtained from final tuple T we can estimate *objective accuracy of rule* oac. If we assume that in tuple T, z attributes A_k (for $1\le k\le z$) correspond to z potential facts F_k, which are the potential premises or the conclusion in the obtained rule, then parameter oac can be calculated from the following formula as the arithmetic mean of its facts' objective accuracies:

$$oac = \frac{1}{z}\sum_{k=1}^{z} oaf(F_k) \tag{8}$$

Formula (8) above allows us to determine objective accuracy of exemplary rules r_a, r_b and r_c:
$oac_{r_a} = 0.72$, $oac_{r_b} = 0.8$ and $oac_{r_c} = 0.43$, respectively.

It means that rules r_a and r_b have a higher objective accuracy than rule r_c which is connected with the previously mentioned fact that corresponding tuples $T_a=T_1$ and $T_b=T_2$ describe precisely their groups of patients and that accuracy of premises and conclusion of rules r_a and r_b are objectively high. The low objective accuracy of rule r_c is connected, among others, with the low objective precision of attributes A_4=age_range, A_5=severity_of_diagn_illness and A_6=symptoms in final tuple T_c.

Lastly, for each rule r we can define parameter ac determining *accuracy of rule*. It can be calculated by the following formula as the arithmetic mean of a rule's relative and objective accuracies:

$$ac = \frac{1}{2}(rac + oac) \tag{9}$$

Formula (9) above allows us to determine accuracy of exemplary rules r_a, r_b and r_c:
$ac_{r_a} = 0.86$, $ac_{r_b} = 0.9$ and $ac_{r_c} = 0.57$, respectively.

We can notice that the accuracy of rule r_c is much smaller in comparison with the accuracies of rule r_a and r_b. It is connected with the reason that some facts of rule r_c, while integrating, radically decrease their accuracies as well as with the reason that precision of some facts of rule r_c is objectively small.

2.5. CALCULATING THE FACTOR OF GLOBAL RULE'S RELIABILITY

To conclude, *global reliability factor* grf, defined by the following formula, will depend simultaneously, on a rule's weight and accuracy:

$$grf = \min\{wg, ac\} \tag{10}$$

Formula (10) above allows us to determine global reliabilities of exemplary rules r_a, r_b and r_c:
$grf_{r_a} = \min\{0.84, 0.86\} = 0.84$, $grf_{r_b} = \min\{0.89, 0.9\} = 0.89$ and $grf_{r_c} = \min\{0.93, 0.57\} = 0.57$.

As we can see rule r_b has the highest factor grf. It means that while integrating, the number of patients 'caught' in rule r_b is relatively high and the decrease in the accuracy of the patients' attributes in rule r_b is relatively low. Simultaneously, rule r_b has a relatively 'characteristic' conclusion and the objective precision of facts in rule r_b is relatively high.

2.6. RULE'S PRIORITY

To summarize, we have concluded that for each 2-uncertain rule, factor $\lfloor grf \cdot 100 \rfloor$ (the integral part of $grf \cdot 100$) will be the one deciding about a rule's priority. It means that rule r_c has a much lower priority in designed RBS then rules r_b and r_a. This statement has to be taken into consideration during the process of uncertain reasoning illustrated in Chapter 3.

3. EXAMPLES OF UNCERTAIN REASONING

To realize the influence of factor grf on the process of uncertain reasoning and calculating irf value for a hypothesis concluded by reasoning chains, let us continue our example from Chapter 2. Let us assume that in a sick room, a boy, aged 7, suffering from severe asthma, with current acute asthma exacerbation and asthma attack, with wheezing and shortness of breath, treated with short-act beta2 agonist, systemic corticosteroid, and inhaled anticholinergic, is being seen by a surgeon on duty. The doctor is considering if the patient does not need to be hospitalized. Let us notice that our patient satisfies premises of 2-uncertain rules: r_a and r_c, in the designed RBS (see Chapter 2). Suppose that reasoning is in

progress at the moment and none of the rules concluding no_hosp_admission has been fired up to this moment.

If rule r_a with $grf_{r_a} = 0.84$ is fired for this boy then hypothesis treatment_effects would be:

treatment_effects = no_hosp_admission with irf = $p_{c\,r_a} = 0.76$

If rules r_a and r_c are now successively fired (based on the rules' activities and the advantage of factor $grf_{r_a} = 0.84$ over factor $grf_{r_c} = 0.57$) then after firing, these two rules in the reasoning chain, hypothesis treatment_effects would be as follows:

treatment_effects = no_hosp_admission with irf = $p_{c\,ch(r_a r_c)}$, where an internal reliability factor of reasoning chain $irf = p_{c\,ch(r_1 \ldots r_w)}$ is calculated by the following formula:

$$p_{c\,ch(r_1 \ldots r_w)} = f(H, \mathbf{KB}) = f(H, \mathbf{KB}, w) \qquad \text{where:}$$

$$f(H, \mathbf{KB}, i) = \begin{cases} p_{c\,r_1} & \text{for} \quad i = 1 \\ (1 - v_i) \cdot f(H, \mathbf{KB}, i-1) + v_i \cdot p_{c\,r_i} & \text{for} \quad 2 \le i \le w \end{cases} . \tag{11}$$

From the following formula:

$$v_i = \begin{cases} 1 & \text{for} \quad i = 1 \\ \dfrac{v_{i-1}}{t + v_{i-1}} & \text{for} \quad 2 \le i \le w \end{cases} \tag{12}$$

and assumption that $t = 1.1$, we obtain: $v_2 = \dfrac{1}{1.1 + 1} = 0.48$.

Finally, using formula (11), we have:

$p_{c\,ch(r_a r_c)} = (1 - v_2) \cdot p_{c\,r_a} + v_2 \cdot p_{c\,r_c} = (1 - 0.48) \cdot 0.76 + 0.48 \cdot 0.84 = 0.8.$

As we can notice, after firing both rules, probability of the considered hypothesis no_hosp_admission increases. We can ask if such an increase in the probability of hypothesis no_hosp_admission for this concrete boy is reasonable and required. Let us recall that rule r_c has really small priority in the RBS, among others due to the reason that the objective precision of attribute A_4=age_range={1,…,60}⊕ is low. To avoid this situation we can establish a threshold value τ (see [2]) for factor grf and fire only the rules with $grf \ge \tau$.

Furthermore, for hypothesis concluded by reasoning chains of rules with determined irf, we can establish a global reliability factor of reasoning chain $grf = p_{r\,ch(r_1 \ldots r_w)}$ by the following formula:

$$p_{r\,ch(r_1 \ldots r_w)} = \min \{grf_{r_1}, \ldots, grf_{r_w}\} \tag{13}$$

It means that in the considered example, by formula (13), we obtain:

with grf =0.84 (treatment_effects = no_hosp_admission with irf =0.76) or

with grf = min {0.84, 0.57} = 0.57 (treatment_effects = no_hosp_admission with irf =0.8)

As we can notice, in the second case grf of reasoning chain is relatively small and our assumption of firing only the rules with sufficiently high $grf \ge \tau$ is fully justified.

4. CONCLUSIONS

The examples presented in the paper have illustrated the method of designing 2-uncertain rules on the basis of medical aggregate real data and determining their reliability factors irf and grf. The obtained rules compose the knowledge base of the medical Rule Based System (RBS). Consequently, in the paper the process of uncertain reasoning and calculating reliabilities of a hypothesis concluded by a reasoning chain was demonstrated. We pointed out that firing, in the reasoning chain, the low priority 2-uncertain rule can result in an unreasonable revision of the internal reliability of the concluded hypothesis. Therefore we proposed to establish a threshold value for factor grf and fire only the rules with sufficiently high global

reliability. It means that global reliability of rules in the RBS affecting the course of the reasoning chain will indirectly influence the internal reliability of the concluded hypothesis.

The 2-uncertain rules presented in the paper were obtained on the basis of aggregate data. Obviously, the presented method can be adapted to individual data too. Unfortunately, the possibility of obtaining access to such data is still a big problem.

BIBLIOGRAPHY

[1] AGRAVAL R., IMIELIŃSKI T., SWAMI A., Database mining: A performance perspective, IEEE Transactions on Knowledge Engineering, 1993, Vol. 5, No. 6, pp. 914-925.

[2] JANKOWSKA B., Evidence-Based Model for 2-Uncertain Rules and Inexact Reasoning, submitted to the conf. MIT 2012.

[3] JANKOWSKA B., SZYMKOWIAK M., On Ranking Production Rules for Rule-Based Systems with Uncertainty, Proc. 3rd Int. Conf. on Collective Computational Intelligence (Part I), Springer, Gdynia, 2011, LNCS 6922, pp. 546-556.

[4] KRYSICKI W. et al., Mathematical Statistics, PWN, Warszawa, 2006, (in polish).

[5] 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.

[6] PETRIE A., SABIN C., Medical Statistics at a Glance, Blackwell Science Ltd, London, 2000.

[7] PLOTNICK L.H., DUCHARME F.M., Combined inhaled anticholinergics and beta2-agonists for initial treatment of acute asthma in children, The Cochrane Library, 2005.

[8] SHORTLIFFE E., Computer-Based Medical Consultations: MYCIN, American Elsevier, 1976.

[9] SZYMKOWIAK M., JANKOWSKA B., Discovering Medical Knowledge from Data in Patients' Files, ICCCI Wrocław 2009, Springer-Verlag, Berlin/Heidelberg, LNAI 5796, pp. 128-139.

[10] SZYMKOWIAK M., JANKOWSKA B., Reliability of Medical Production Rules Obtained by means of Aggregate Data Mining, Journal of Medical Informatics & Technologies, 2010, Vol.14, pp.103-110.