

# Selection of Optimal Rule Refinements

Hans - Werner Kelbassa

E-mail: kelbassa@uni-paderborn.de

## Abstract

Refinement systems currently select the best rule refinements from a set of conflicting and alternative refinements by using heuristics. This paper presents a new operations research approach for the selection of optimal rule refinements. The problem analysis and the description of the optimization are performed using two well-defined examples. The analysis of the refinement selection problem leads to the conclusion that a binary linear maximization problem is to be solved by a operations research procedure. The mathematical solution of the refinement selection problem is a milestone in the history of rule refinement.

## Introduction

At present the available *commercial tools* for the development of rule bases do not provide a *validation system*. A study of the current state of the art in the area of rule validation and rule refinement reveals that there is no generic validation interface and no optimal selection of rule refinements. Concerning the selection of rule refinements no methodical standard so far existed, i.e. current refinement systems select the best rule refinements out of a set of conflicting and alternative ones by hill-climbing procedures.

*Rule trace validation* can ideally be subdivided into the initial rule base evaluation step (case-based validation), the identification of faulty rules, the generation of suitable rule refinements, the selection of the best rule refinements and, finally, the execution of the selected rule refinements. This article deals with the *selection of the best rule refinements*, which, in current refinement systems, is performed in a heuristic manner (Craw 1991, Ginsberg 1988, Boswell 1999, Carbonara and Sleeman 1999), and presents a *mathematical formalization* of the rule refinement selection problem. The reader who is interested in how to come up with suitable rule refinement heuristics is referred to (Kelbassa and Knauf 2003). The focus of this paper is the selection of optimal refinement heuristics out of a given set of conflicting, normal, and alternative rule refinement heuristics. The mathematical outcome is a binary maximization problem solvable by operations research procedures, i.e. a novel approach for the optimal selection of rule refinements.

Copyright © 2003, American Association for Artificial Intelligence (www.aaai.org). All rights reserved.

In the next section we present a rule refinement example which has already been analyzed in (Kelbassa 2003). The result of this conflict analysis is a set of *one-of disjunctions* which state the conflicting nature of the rule refinements considered, and enables an operations research formalization for the rule refinement selection problem. After the operations research solution for the rule refinement example with conflicting refinements has been presented, the simultaneous selection of conflicting, normal, and alternative rule refinement heuristics is discussed by using a second working example.

## Rule Refinement Example

For the elaboration of the refinement selection problem an higher order refinement example with three refinement classes is discussed now (Kelbassa 2002a, 2002b). Consider the following rules contained in a rule base  $RB_0$  of a forward-chaining inference system:

$$\begin{aligned} RB_0 &:= \{ \dots, R_{18}, \dots, R_{42}, \dots, R_{46}, R_{47}, \dots, R_{54}, \dots, R_{64}, \dots, R_{66}, \dots \} \\ R_{18} &:= \text{IF } (A \wedge B) \text{ THEN } \textit{Hypothesis\_1} \\ R_{42} &:= \text{IF } (C \vee D) \text{ THEN } \textit{Hypothesis\_3} \\ R_{46} &:= \text{IF } (\textit{Hypothesis\_1} \wedge \neg \textit{Hypothesis\_3}) \text{ THEN } I_8 \\ R_{54} &:= \text{IF } (\textit{Hypothesis\_3} \wedge E) \text{ THEN } I_2 \\ R_{64} &:= \text{IF } (A \wedge \neg K) \text{ THEN } \textit{Hypothesis\_7} \end{aligned}$$

Here  $I_2$  and  $I_8$  are two different *final* conclusions; the letter  $I$  means *interpretation*, i.e., any proposition. Assume that the production rules in this  $RB_0$  processed different problem cases and that the domain expert has entered his evaluation for every case by validation interface. Suppose the rule refinement heuristics listed below have been obtained by the validation system according to the validators rule trace evaluation. These refinement heuristics are *rule refinement expertise* – the above rules became refinement candidates:

RH1 := IF rule  $R_{46}$  is generalized by  $\phi_G^2$ ,  
THEN case set  $C_1$  gets valid reasoning  
paths (rule traces):  $|C_1| = 4$

RH1/2 := IF rule  $R_{64}$  is contextualized by  $\phi_C^1$ ,  
THEN case set  $C_2$  gets valid reasoning  
paths (rule traces):  $|C_{1/2}| = 9$

RH2 := IF rule  $R_{18}$  is contextualized by  $\phi_C^2$ ,  
and  
rule  $R_{42}$  is specialized by  $\phi_S^1$ ,  
THEN case set  $C_2$  gets valid reasoning  
paths (rule traces):  $|C_2| = 1$

RH3 := IF rule  $R_{46}$  is generalized by  $\phi_G^1$ ,  
and  
rule  $R_{54}$  is generalized by  $\phi_G^1$ ,  
and  
rule  $R_{64}$  is contextualized by  $\phi_C^2$ ,  
THEN case set  $C_3$  gets valid reasoning  
paths (rule traces):  $|C_3| = 8$

RH3/2 := IF rule  $R_{46}$  is specialized by  $\phi_S^1$ ,  
and  
rule  $R_{54}$  is specialized by  $\phi_S^2$ ,  
and  
rule  $R_{64}$  is generalized by  $\phi_G^3$ ,  
THEN case set  $C_{3/2}$  gets valid reasoning  
paths (rule traces):  $|C_{3/2}| = 15$

In these refinement heuristics the symbols  $\phi_C^1$ ,  $\phi_C^2$ ,  $\phi_G^1$ ,  $\phi_G^2$ ,  $\phi_G^3$ ,  $\phi_S^1$ ,  $\phi_S^2$  characterize *elementary* rule refinement operations (Kelbassa 2003). The index C means contextualization ( $\phi_C$ ), the index G means generalization ( $\phi_G$ ), and the index S means specialization ( $\phi_S$ ); the superscript is the class index. The set of all rule refinements is  $\Phi := \{\phi_C, \phi_G, \phi_S\}$ .

It is to be emphasized that the above refinement heuristics are *not alternative ones*, because every case appears once only:  $C_1 \cap C_{1/2} \cap C_2 \cap C_3 \cap C_{3/2} = \emptyset$ . If refinement heuristics are alternative ones, so, for example, that we should apply either heuristic RH4 or RH5 in order to validate a certain case set, then this case set intersection is not empty:  $C_4 \cap C_5 \neq \emptyset$ . The optimal selection of alternative rule refinement heuristics is formalized in a section below.

### Rule Refinement Conflict Analysis

The elementary refinement operations above are stated with regard to the reference rule  $R_x$ , i.e. the refinements all are referring to an unrefined faulty rule  $R_x \in RB_0$  of the same rule base. Thus there are difficulties if we try to get a good sequence of elementary refinements for the refinement of any rule  $R_x$  which failed in several cases. This *sequence problem* or *refinement reference problem* is not sufficiently solved yet (Boswell 1999).

Let  $CS(R_x)$  be the refinement conflict set for rule  $R_x \in RB_0$ , which contains all demanded refinements the rule  $R_x$  is subject to, i.e. every required refinement operation for this rule. Accordingly the conflict sets CS for our specific problem are the following:

$CS(R_{18}) = \{\phi_C^2\}$	Comment: no conflict
$CS(R_{42}) = \{\phi_S^1\}$	Comment: no conflict
$CS(R_{46}) = \{\phi_G^1, \phi_G^2, \phi_S^1\}$	
$CS(R_{54}) = \{\phi_G^1, \phi_S^2\}$	
$CS(R_{64}) = \{\phi_C^1, \phi_C^2, \phi_G^3\}$	

The conflict sets  $CS(R_{18})$  and  $CS(R_{42})$  do not reveal any refinement conflict, because these sets both have one element only. The example conflict sets to be investigated are the  $CS(R_{46})$ , the  $CS(R_{54})$ , and the  $CS(R_{64})$ .

Altogether the conflict analysis regarding the refinement candidates  $R_{18}, R_{42}, R_{46}, R_{54}, R_{64}$  leads to the recognition that the conflict sets of the latter three rules have to be described by ONE-OF DISJUNCTIONS. A detailed analysis of this sample is presented in (Kelbassa 2003); due to space limit this analysis cannot be repeated here. The resulting ONE-OF restrictions for  $CS(R_{46})$ ,  $CS(R_{54})$ , and  $CS(R_{64})$  have the following form:

ONE-OF( $RH1(R_{46}), RH3/2(R_{46}), RH3(R_{46})$ ),  
[ ONE-OF( $RH3(R_{54}), RH3/2(R_{54})$ ), ]  
ONE-OF( $RH1/2(R_{64}), RH3(R_{64}), RH3/2(R_{64})$ ).

The optimization will be on the heuristic level. It is not relevant for the following which rules occur in ONE-OF DISJUNCTIONS so that the second ONE-OF is discarded (stated by [...]), because the third ONE-OF is more restrictive. These constraints can be converted into linear inequalities which are the basis of an operations research solution.

### Operations Research Approach

In the scientific world the discipline which is researching mathematical optimization is called operations research (acronym: OR). As far as it is known to the author the optimal selection of *conflicting or alternative* rule refinements has not been solved and published by other scientists yet.

The various heuristics have different success in the validation of cases, hence the expected total case gain of the heuristics is maximized. The question whether a certain heuristic is an element of the optimal heuristic set can be answered by a binary decision variable  $x \in \{0, 1\}$ . Accordingly the optimization result  $x_j = 1$  ( $j \in \mathbb{N}$ ) means that the  $j$ -th heuristic is optimal; whereby the result  $x_j = 0$  ( $j \in \mathbb{N}$ ) means that the  $j$ -th refinement heuristic is suboptimal and therefore not to be executed. With respect to our special rule refinement problem we define the following variables:

$x_1 \in \{0, 1\} :=$  Decision variable for heuristic RH1  
 $x_2 \in \{0, 1\} :=$  Decision variable for heuristic RH1/2  
 $x_3 \in \{0, 1\} :=$  Decision variable for heuristic RH2  
 $x_4 \in \{0, 1\} :=$  Decision variable for heuristic RH3  
 $x_5 \in \{0, 1\} :=$  Decision variable for heuristic RH3/2

The gain  $g_j \in \mathbb{R}$  of the refinement heuristics here are the number of debugged cases (THEN-part); hence we assign the following gain values:  $g_1 = |C_1| = 4$ ,  $g_2 = |C_{1/2}| = 9$ ,  $g_3 = |C_2| = 1$ ,  $g_4 = |C_3| = 8$ ,  $g_5 = |C_{3/2}| = 15$ .

Using the gain values  $g_j$  ( $j = 1, \dots, n$ ) we can also take into account that different cases have different weights. So, for example, if any gain variable  $g_{42}$  is referring to a refinement heuristic which validates three cases  $C_{42} := \{c_4, c_{21}, c_{36}\}$  having the weights  $w_4 = 17$ ,  $w_{21} = 25$ , and  $w_{36} = 44$ , then the gain is  $g_{42} = 17 + 25 + 44 = 86$ . For the sake of clarity different case weights are not used in our specific refinement

example. Different case weights are mentioned here, since uniform case weights are a drawback of several refinement systems (Carbonara and Sleeman 1999, Ginsberg 1988).

The *objective function* for the rule refinement selection problem RSP is:

$$\sum_{j=1}^n g_j x_j \rightarrow \text{maximum!}$$

$$x_j \in \{0, 1\}; g_j \in \mathbb{R}; j = 1, \dots, n (n \in \mathbb{N}).$$

Although in our sample all gain values are positive, it is stated that  $g_j \in \mathbb{R}$  (– relevant in the case of side effects).

Next we come up with the *linear inequalities* for the ONE-OF DISJUNCTIONS. Let  $x_1, \dots, x_n$  be binary decision variables involved in any one-of conflict, then for a ONE-OF DISJUNCTION the following inequality holds:

$$x_1 + \dots + x_n \leq 1 (n \in \mathbb{N}).$$

Solving our refinement problem we process the above two ONE-OF DISJUNCTIONS:

$$\text{ONE-OF}(RH1, RH3, RH3/2)$$

$$\text{ONE-OF}(RH1/2, RH3, RH3/2)$$

Based on the problem specific definition of the binary decision variables  $x_j (j = 1, \dots, 5)$  the one-of inequalities are:

$$x_1 + x_4 + x_5 \leq 1 \quad (I)$$

$$x_2 + x_4 + x_5 \leq 1 \quad (II)$$

Thus the optimization approach for our special rule refinement selection problem is:

$$4x_1 + 9x_2 + 1x_3 + 8x_4 + 15x_5 \rightarrow \text{maximum!}$$

$$x_1 + \quad \quad \quad x_4 + \quad \quad \quad x_5 \leq 1 \quad (I)$$

$$x_2 + \quad \quad \quad x_4 + \quad \quad \quad x_5 \leq 1 \quad (II)$$

$$x_1, x_2, x_3, x_4, x_5 \in \{0, 1\}$$

The mathematical approach for the conflicting rule refinement selection problem RSP is:

$$\text{RSP} := \begin{cases} \sum_{j=1}^n g_j x_j \rightarrow \text{maximum!} \\ \text{subject to} \\ \sum_{j=1}^n a_{ij} x_j \leq b_i \quad (i = 1, \dots, m) \\ g_j \in \mathbb{R}, x_j \in \{0, 1\} \quad (j = 1, \dots, n) \\ a_{ij} \in \{1, 0, -1\}, b_i \in \{1, 0\} \quad (i = 1, \dots, m) \end{cases}$$

Due to the restriction  $x_j \in \{0, 1\}$  this optimization problem is called a *binary integer problem*. In our numeric example  $m = 2$  holds since finally there are only two one-of constraints. As there are no predefined dependencies between the decision variables here  $a_{ij} \in \{0, 1\}$  and  $b_i \in \{1\}$  holds, else  $a_{ij} \in \{1, 0, -1\}$  and  $b_i \in \{0, 1\}$  can be present.

This refinement selection problem RSP is solvable by several well known OR procedures; in particular a RSP can be solved by using the

- ADDITIVE BALAS' ALGORITHM (Balas 1967, Burkhard 1972, Neumann 1975);
- BRANCH AND BOUND PROCEDURE (Neumann and Morlock 2002, Hillier and Lieberman 2001);
- GOMORY PROCEDURE (Neumann 1975, Neumann and Morlock 2002, Schrijver 2000);
- BRANCH AND CUT PROCEDURE (Hoffman and Padberg 1991, Hillier and Lieberman 2001, Jünger and Naddef 2001).

Due to scientific progress, predictions about the future roles of the above OR procedures are risky <sup>1</sup> (Joseph 2002).

The matrix representation of the RSP is:

$$\max \{gx | Ax \leq b, x \in \{0, 1\}^n, n \in \mathbb{N}\}.$$

For GOMORYS PROCEDURE A must be an integral  $m \times n$  matrix ( $A \in \mathbb{Z}^{m \times n}$ ), b an integral m-vector ( $m \in \mathbb{N}$ ), and g an integral n-vector ( $g \in \mathbb{Z}^n$ ).

For the BRANCH AND CUT PROCEDURE A must be a rational  $m \times n$  matrix ( $A \in \mathbb{Q}^{m \times n}$ ) and b an m-vector of rationals ( $m \in \mathbb{N}$ ).

The optimal solution for our specific RSP is  $G^{opt}(0, 0, 1, 0, 1) = 16$  cases. As the outcome is  $x_3 = x_5 = 1$ , the heuristics RH2 and RH3/2 are optimal.

## Alternative Refinements

Concerning RSP no alternative refinement heuristics have been considered here yet. However, this topic is crucial for the optimization of rule base refinement. The simplest form of alternative refinements is present if there are two rule refinement heuristics validating the same case set. In this situation we have to ensure that maximal one of them is realized. Let  $x_8$  and  $x_9$  be the binary decision variables for the alternative heuristics RH4 and RH5, then the ONE-OF constraint for these alternative refinements is:  $x_8 + x_9 \leq 1$ .

A powerful refinement optimization should combine the selection of conflicting refinements with the selection of alternative refinements. As not all five refinement heuristics of our first example are optimal ones, we have to find alternatives for the heuristics RH1, RH1/2, and RH3, after the optimal refinements RH2 and RH3/2 have been executed. Assume that this is done side effect free and that these refinements are listed in a revision protocol which ensures that the rules just refined not become refinement candidates in the following refinement stage(s) again (pendulum problem).

When we start to find out alternative refinement heuristics for the remaining case sets, we are going to solve the alternative rule trace problem for a part of the entire refinement problem. In general it should be possible to find all the alternatives for every refinement heuristic (by using a refinement generator (Carbonara and Sleeman 1999)) so that it is possible to compute the optimal refinement set for known competing and alternative refinement heuristics at the beginning. However, this requires more work to be done by the validator and larger financial and time expenditures for rule base development.

<sup>1</sup>Hillier and Lieberman 2001, p. 167

Hence we prefer to solve a *remainder problem* instead of determining alternatives for all refinement heuristics. The complexity of the refinement problems discussed here is not high. Though industrial rule bases can have more than 13 000 rules with many conditions and therefore it can be difficult and time consuming to develop all refinement alternatives for large scale knowledge bases (Puppe et al. 2001). However, for small real-world refinement problems it is possible to optimize a set of alternative and competing rule refinements at the beginning. This will be elucidated now.

In order to come up with a refinement solution for the remaining falsified case sets associated with heuristics RH1, RH1/2, and RH3, recall that the refinement of the rule base  $RB_0 \rightarrow RB_1$  yields improved inferences, so that we can suppose that the number of suboptimal elementary rule refinement operations (here: 5) is larger than the number of alternative elementary refinement operations needed still (Ginsberg 1988). Facing this situation the following two refinement heuristics are presented, which correct rules having the same conclusion as two of the rules just refined:

RH1/3 := IF rule  $R_{47}$  is generalized ( $\phi_G^4$ ),  
 THEN case set  $C_1$  gets a valid reasoning  
 path (rule trace):  $|C_1| = 4$   
 $R_{47} := \text{IF } (W \wedge \text{Hypothesis}_1) \text{ THEN } I_8$   
 $\phi_G^4(R_{47}) := \text{Insertion of a disjunct into a given conjunc-}$   
 tion; here:  $W \wedge (\text{Hypothesis}_1 \vee K)$   
 RH1/4 := IF rule  $R_{66}$  is generalized ( $\phi_G^3$ ),  
 THEN case sets  $C_{1/2}$  and  $C_3$  gets valid  
 reasoning paths:  $|C_{1/2}| + |C_3| = 17$   
 $R_{66} := \text{IF } (K \wedge L) \text{ THEN } \text{Hypothesis}_7$   
 $\phi_G^3(R_{66}) := \text{Enlargement of an interval in a}$   
 numerical threshold condition;  
 here:  $L := (l > 25) \rightarrow L^* := (l > 10)$

As there is no conflict between these two refinement heuristics (different rules), these heuristics can be considered as optimal ones which validate all remaining cases. Although our special refinement problem is completely solved now, we show here how to come up with an OR solution for the selection between *alternative, normal, and competing* rule refinement heuristics. Assuming we have to solve a selection problem in which the five heuristics of our conflict example are involved, and furthermore the alternative heuristics RH1/3 and RH1/4 together with the alternative heuristics RH4 and RH5, which both realize the gain  $g_8 = |C_4| = 18$ . All heuristics involved in this second example are presented in table 1.

Supposed that these alternative heuristics will not cause a new refinement conflict, hence concerning CS analysis it is sufficient to state the above constraints (I, II). The columns in table 1 reveal that for the case sets  $C_2$  and  $C_{3/2}$  there are no alternative refinements. So we can make a *preselection*, for example, represented by  $x_3 = 1$ . As heuristic RH2 is neither a conflicting one nor an alternative one, we classify it as a *normal* heuristic. If one refinement heuristic appears in more than one case set column, as RH1/4 does, every one of these columns can yield a single ONE-OF DISJUNCTION; here we have two:

Table 1: Selection table for conflicting, normal, and alternative rule refinement heuristics.

Cases $C_1$	Cases $C_{1/2}$	Cases $C_2$	Cases $C_3$	Cases $C_{3/2}$	Cases $C_4$
RH1 ( $x_1$ )	RH1/2 ( $x_2$ )	RH2 ( $x_3$ )	RH3 ( $x_4$ )	RH3/2 ( $x_5$ )	RH4 ( $x_8$ )
RH1/3 ( $x_6$ )	RH1/4 ( $x_7$ )		RH1/4 ( $x_7$ )		RH5 ( $x_9$ )

ONE-OF(RH1/2, RH1/4) and ONE-OF(RH1/4, RH3). But these two restrictions enable redundant results as  $x_2 = x_7 = 1$  and  $x_4 = x_7 = 1$ , although it is sufficient to generate  $x_7 = 1$  and  $x_2 = x_4 = 0$ . This minimal refinement outcome can be ensured by the following *either-or constraints* which are mutually exclusive:

(either)  $x_2 + x_4 < x_7$ , i.e.,  $x_2 = x_4 = 0, x_7 = 1$ ;  
 (or)  $x_7 = 0$ .

Let  $M \in \mathbb{N}$  be any large integer and  $y \in \{0, 1\}$  a binary auxiliary variable, then the above either-or-constraints are represented by the inequalities (VI) to (XI) as shown in the extended RSP\* approach stated below.

In our first example it is not assumed that any decision variable  $x_l$  depends on another decision variable  $x_j$  ( $x_l \neq x_j$ ;  $l, j \in \{1, \dots, 5\}$ ). This situation may arise when the validation of the case(s) associated with decision variable  $x_j$  gets a higher priority than the validation of the case(s) associated with decision variable  $x_l$ ; a priority can be stated by the binary constraint  $x_l \leq x_j$ . Assume that for our second specific RSP\* the preference  $x_8 \leq x_1$  holds.

Then according to table 1 the OR approach for our extended (second) specific RSP\* is:

$$\begin{aligned}
 &4x_1 + 9x_2 + 1x_3 + 8x_4 + 15x_5 + 4x_6 + \\
 &\quad + 17x_7 + 18x_8 + 18x_9 \rightarrow \text{maximum!} \\
 &\quad x_1 + x_4 + x_5 \leq 1 \quad (I) \\
 &\quad x_2 + x_4 + x_5 \leq 1 \quad (II) \\
 &\quad x_1 + x_6 \leq 1 \quad (III) \\
 &\quad x_8 + x_9 \leq 1 \quad (IV) \\
 &\quad x_3 = 1 \quad (V) \\
 &\quad x_7 - My \leq 1 \quad (VI) \\
 &\quad -x_7 - My \leq -1 \quad (VII) \\
 &\quad x_2 + x_4 - My \leq 0 \quad (VIII) \\
 &\quad -x_2 - x_4 - My \leq 0 \quad (IX) \\
 &\quad x_7 + My \leq M \quad (X) \\
 &\quad -x_7 + My \leq M \quad (XI) \\
 &\quad -x_1 + x_8 \leq 0 \quad (XII)
 \end{aligned}$$

$M \in \mathbb{N}$  must be a large integer

$$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, y \in \{0, 1\}$$

The optimal solution for this specific RSP\* is  $G^*(0, 0, 1, 0, 1, 1, 1, 0, 1) = 55$  cases. As this optimum is  $x_3 = x_5 = x_6 = x_7 = x_9 = 1$ , the heuristics RH2, RH3/2, RH1/3, RH1/4, and RH5 are the optimal ones. This means, for example, that again as in the first sample the heuristics RH1 and RH1/2 are suboptimal. Regarding the preference neither  $x_1$  nor  $x_8$  is optimal:  $0 \leq 0$ .

REVISION. Rule *refinement* means that the cardinality of the refined rule base is kept constant:  $|RB_0| = |RB_1|$ . If there is no possibility to fix one or more present bugs by contextualizations, by generalizations or by specializations, then we should apply *revision* operations:  $|RB_0| \neq |RB_1|$ . This means that new knowledge is integrated by *new* rules and that *obsolete* rules are deleted.

## Conclusion

At present the selection of alternative and conflicting rule refinements by refinement systems is performed without global optimization by greedy heuristics. The novel optimization approach described in this article enables a new generation of powerful refinement systems.

Evaluating the current state of the art we cannot ascertain whether reduction approaches or reasoning path validation approaches will prove most relevant for future expert system developments, since progress is being made in both validation research domains (Knauf 2000, Carbonara and Sleeman 1999, Boswell and Craw 2000, Kelbassa and Knauf 2003).

A. GINSBERG asserted that SEEK2's rule representation would not be amenable to the application of mathematical optimization; he did not believe that global optimization is possible in the domain of rule refinement.<sup>2</sup> It has been shown in this article that the opposite is true. The result of every rule refinement conflict analysis is amenable to mathematical optimization by OR procedures. The foundation of the mathematical optimization as presented in this paper is an analysis of conflicting and alternative rule refinement restrictions, and the possibility of assigning validation gain values to rule refinement heuristics. It has been shown by discussing two defined examples that the outcome of the formal analysis of the rule refinement selection problem is a *binary linear maximization problem* solvable by a binary OR procedure. The application of exact procedures for coping with special rule refinement selection problems is innovative, since refinement systems usually employ heuristics and hill-climbing procedures. This article reveals that the selection of rule refinements can be optimized, so that *large* high-performance rule bases can be developed more rapidly, and, moreover, suboptimal techniques for rule refinement and rule validation are becoming obsolete.

## References

- Balas, E. 1967. Discrete Programming by the Filter Method. *Operations Research* 15 (5): 915 - 957
- Boswell, R. 1999. *Knowledge Refinement for a Formulation System*. Ph.D. Thesis. The Robert Gordon University, Department for Computing and Mathematical Science, Aberdeen / Scotland, U. K.
- Boswell, R., Craw, S. 2000. Experiences with a Generic Refinement Toolkit. Dieng, R., Corby, O. (Eds.): *Knowledge Engineering and Knowledge Management: Methods, Models, and Tools (EKAW 2000)*. Lecture Notes in Artificial Intelligence 1937, 249-256. Berlin, Germany: Springer-Verlag 2000
- Burkhard, R. E. 1972. *Methoden der Ganzzahligen Optimierung*. Vienna, Austria: Springer-Verlag
- Carbonara, L., Sleeman, D. 1999. Effective and Efficient Knowledge Base Refinement. *Machine Learning* 37: 143 - 181
- Craw, S. 1991. Judging Knowledge-base Quality. Ayel, M., Laurent, J.-P. (Eds.): *Validation, Verification and Test of Knowledge-Based Systems*, 207 - 219. New York: John Wiley & Sons
- Ginsberg, A. 1988. *Automatic Refinement of Expert System Knowledge Bases*. London, U.K.: Pitman Publishing
- Hillier, F.S., Lieberman, G.J. 2001. *Introduction to Operations Research*. 7th Edition. New York, USA: McGraw-Hill, Inc.
- Hoffman, K. L., Padberg, M. 1991. Improving LP-Representations of Zero-One Linear Programs for Branch and Cut. *ORSA Journal on Computing* 3 (2): 121 - 134
- Joseph, A. 2002. *A concurrent processing framework for the set partitioning problem*. Computers & Operations Research 29: 1375 - 1391
- Jünger, M., Naddef, D. (Eds.). *Computational Combinatorial Optimization*. Lecture Notes in Computer Science 2241. Berlin, Germany: Springer-Verlag 2001
- Kelbassa, H.-W. 2002a. Higher Order Refinement Heuristics for Rule Validation. *Proceedings of the 15th International Florida Artificial Intelligence Research Society Conference 2002 (FLAIRS 2002)*, 211 - 215. Pensacola, Florida: May 14 - 16, 2002
- Kelbassa, H.-W. 2002b. Context Refinement - Investigating the Rule Refinement Completeness of SEEK/SEEK2. *Proceedings of the 15th European Conference on Artificial Intelligence 2002 (ECAI 2002)*, 205 - 209. Lyon, France: July 21 - 26, 2002
- Kelbassa, H.-W. 2003. Optimal Refinement of Rule Bases. *Forthcoming*. Paderborn, Germany
- Kelbassa, H.-W., Knauf, R. 2003. The Rule Retranslation Problem and the Validation Interface. *Proceedings of the 16th International Florida Artificial Intelligence Research Society Conference 2003 (FLAIRS 2003)*. St. Augustine, Florida: May 12-14, 2003
- Knauf, R. 2000. *Validating Rule-Based Systems - A Complete Methodology*. Habilitation Thesis. Aachen, Germany: Shaker Verlag
- Neumann, K. 1975. *Operations Research Verfahren*. Volume I. Munich, Germany: Carl Hanser Verlag
- Neumann, K., Morlock 2002. *Operations Research*. 2nd Edition. Munich, Germany: Carl Hanser Verlag
- Puppe, F., Ziegler, S., Martin, U., Hupp, J. 2001. *Wissensbasierte Diagnose im Service-Support*. Berlin, Germany: Springer-Verlag
- Schrijver, A. 2000. *Theory of Linear and Integer Programming*. New York: John Wiley & Sons

<sup>2</sup>Ginsberg 1988, p. 5 and p. 108f.