Middle-East Journal of Scientific Research 21 (9): 1659-1662, 2014
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2014.21.09.21704

Production-Tabular Knowledge Bases. Tools for Assessing and Checking of Correctness

R.A. Karayev and N.Y. Sadikhova

Institute of Cybernetics of the National Academy of Sciences of Azerbaijan, Az1141, 9, B. Vahabzade str., Baku, Azerbaijan

Abstract: The production-tabular knowledge bases widely used in commercial expert systems. One of the main problems arising from the operation of this kind of knowledge bases is a problem their correctness. The reliability of the inference mechanism and the robustness of the expert system at “shift the paradigm” depend largely on successful solution of this problem. The paper gives a formal definition of “correctness” of extended entry production-tabular knowledge bases and proposes an algorithm to control their correctness. The obtained results create the theoretical preconditions to ensure the reliability and robustness of the production-tabular technologies, widely used in expert systems of diagnostics, monitoring, management, forecasting, decision-making.

Key words: Production-tabular knowledge bases · Correctness · Algorithms check

INTRODUCTION

Considered production-tabular knowledge bases is the class of hybrid structures of knowledge representation, in which production systems described in terms of extended entry decision tables [1, 2]. The formalism of decision tables can significantly extend the capabilities of applications of popular expert systems based on production rules.

In particular, to solve critical for these systems issue associated with check of correctness (completeness and consistency) of knowledge bases.

The paper attempts to solve this issue in the framework of mathematical model of decision tables, regarded as the isomorphism of production-tabular knowledge base.

Used a modified technique of analyzing the correctness, developed in [3, 4] for decision tables with limited input (tables with double-digit terms "Yes / No").

Basic Concepts and Definitions: Formally, the decision table is given [4] by a set $T = \langle C, D, \tilde{C}, \tilde{D} \rangle$, where

$C = \{C_i\}$, $i=\overline{1,m}$, - is a set of conditions or identifiers of conditions considered as the coordinates of a set of data vectors represented the elementary states of problem area;

$D = \{D_r\}$, $r=\overline{1,k}$ - is a set of solutions or identifiers of solutions, considered as a coordinates of any totality of solution vectors;

$\tilde{C} = \| c_{ij} \|$, $i=\overline{1,m}$, $j=\overline{1,n}$, $\tilde{D} = \| d_{rj} \|$, $r=\overline{1,k}$, $j=\overline{1,n+1}$ - are the matrices that interrelates of the data vectors (or states) and solutions.

The general structure of a decision table shown on the Table 1.

The pair $R_j = \langle \tilde{C}_j, \tilde{D}_j \rangle$, $j=\overline{1,n}$ where $\tilde{C}_j$, $\tilde{D}_j$ - are the vector-columns of matrices $\tilde{C}$ and $\tilde{D}$ called the solutions rule (rule R).

Pair $E = \langle *, \tilde{D}_{n+1} \rangle$, where the symbol " * " means that the first element of the pair is not defined, called the rule "or else" (rule E).

Rule E used for fixing the situations anomalous terms of semantics of problem area and input into decision table for elimination of possible incomplete of knowledge base.

Corresponding Author: R.A. Karayev, Institute of Cybernetics of the National Academy of Sciences of Azerbaijan, Az1141, 9, B. Vahabzade str., Baku, Azerbaijan.

Table 1: The general structure of a decision table

	Rule				
Table Name	1	2	…	n	
Condition 1	A	B	…	C	← Condition Values
Condition 2	Y	Y	…	N	
…	…	…	…	…	
Condition m	Y	N	…	…	
Solution 1	X	X	…		← X for each Solution
Solution 2		X	…		
...			…	X	
Solution k	X		…		

Set of states of a problem area mean the set consisting of the data vectors

$$s_q = (C_i^q),\ I = \overline{1,m},\ where\ C_i^q \in \hat{C}_i; q = \left\{1, 2, ..., \prod_{i=1}^{m} \left|\hat{C}_i\right|\right\}$$

Matrices $\tilde{C} = \left\| c_{ij} \right\|,\ i = \overline{1,m},\ j = \overline{1,n},\ \tilde{D} = \left\| d_{rj} \right\|,\ r = \overline{1,k},\ j = \overline{1, n+1},$ where $c_{ij} \in \{\lambda \bigcup \hat{c}_i\},\ d_{rj} \in \{0, 1, ..., k\}$ establish the relationship between the data vectors (or states) and solutions.

The values of the matrix elements $\tilde{C}$ and $\tilde{D}$ has the following meaning:

$$c_{ij} = \begin{cases} c \in \hat{C}_i, \textit{if the conditions } C_i \textit{ for the rule } R_j \textit{ is } C \\ \lambda, \quad \textit{if the conditions } C_i \textit{ for the rule } R_j \textit{ is} \\ \quad\quad \textit{immaterial}; \end{cases}$$

$$d_{rj} = \begin{cases} d \in \{1,2,...,k\}, \textit{if the decision } D_r \textit{ safisfied for} \\ \quad\quad \textit{the rule } R_j\ (E\ \textit{rules,if } j = n+1) \\ \quad\quad \textit{and has priority and the order } d; \\ 0, \quad \textit{if the action } D_r \textit{ is not performed} \\ \quad\quad \textit{for rules } R_j\ (\textit{rule } E, \textit{if } j = n+1) \end{cases}$$

Usually, the elements $d_{ri} = 0$ are assumed to "default" and not recorded in the decision table and instead elements $c_{ij} = \lambda$ is put the symbol "–".

Definition 1: A decision table that does not contain rules E is called *complete* concerning S, if $(\forall s_q \exists R_j)(s_q \to R_j)$.

Otherwise, the decision table is called *incomplete* concerning S.

Definition 2: A decision table is called *consistent* concerning S, if $(\exists s_q, R_j, R_p)\left[(s_q \to R_j)\,\&\,(s_q \to R_p) \Rightarrow (\tilde{D}_j = \tilde{D}_p)\right]$.

Accordingly, the decision table is called contradictory concerning S, if $(\exists s_q, R_j, R_p)\left[(s_q \to R_j)\,\&\,(s_q \to R_p)\,\&\,(\tilde{D}_j \neq \tilde{D}_p)\right]$.

In this case, we say that the data vector lead to inconsistency of decision tables for rules R_j and R_p.

Definition 3: A decision table is called *correctness* concerning S, if it is complete and consistent concerning S. Otherwise, the decision table is called *incorrectness* concerning S.

The correctness of a decision table concerning S are also called semantic correctness or correctness relative to a given problem interpretation.

Definition 4: The set of syntactically possible (assuming independence conditions C_i) situations N we mean the set consisting of the data vectors $s_q = (C_i^q),\ i = \overline{1,m},\ q = 1, 2, ..., \prod_{i=1}^{m} \left|\hat{C}_i\right|$.

The correctness of a decision table concerning to the set N is called the syntactic correctness or correctness concerning to any problem interpretation.

Before turning to the description of the algorithms check for correctness, we will make some remarks.

R e m an r k 1. Since S is determined by the specifics of a solved problem and given, usually implicit (through a system of constraints), in order to universality as S we take the set N.

Accordingly, check the correctness of decision tables will perform relative to the set N.

R e m an r k 2. In the event of inconsistency or incomplete decision table against N assume that there is a processor (e.g., the compiler of decision tables), capable of the output of the algorithm to establish the correctness or incorrectness of decision tables concerning S. Thus, the question of the semantic correctness in this case rests on the processor.

Check Consistency: Let $R^k \subseteq R$. Vectors conditions S^k of rules R^k form a matrix $\tilde{C}^k = \left\| c_{ij} \right\|,\ i = \overline{1,m},\ j \in J_k$, where J_k – is a set of indexes of rules, that are included in R^k.

Definition 5: Vectors conditions S^k will be called *equivalen*t ("~"), if in each row of the matrix $\tilde{C}^k$, all the essential elements ($c_{ij} \neq \lambda$) are equal each other or all of the elements, except one, is not essential ($c_{ij} \neq \lambda$).

Accordingly, the combination of equivalent vectors S^k will be called an equivalent combination and labeling as $\hat{K}$.

Lemma: In order to no contradictory a decision table has been concerning to S, necessary and sufficient is performance of ratio $(\forall j\, p)\left[(S_j \sim S_p)\,\&\,(\tilde{D}_j = \tilde{D}_p)\right],\ j \neq p.$

Accordingly, a necessary and sufficient condition for the contradictory of decisions table concerning S for the rules R_j and R_p is the performance ratio $(S_j \sim S_p)\,\&\,(\tilde{D}_j \neq \tilde{D}_p),\ j \neq p.$

Scheme of the proof of the lemma is borrowed from [5].

Corollary 1: A decision table is consistent concerning to S for rules R_j and R_p if at least $\tilde{D}_j = \tilde{D}_p$ one of the essential elements $(c_{i\,j}, c_{i\,p} \neq \lambda)$ of the rows is not equal $c_{i\,j} \neq c_{i\,p},\ j \neq p.$

Corollary 2: The collection of data vectors that cause inconsistency decision table concerning to S for rules R_j and R_p are defined of pair $\langle S_j \sim S_p, D_i \neq D_p \rangle;$ for each pair of the pair number of vectors causing inconsistency, is $\prod_{i \in I_\beta} |\hat{c}_{i\,j}|,$ where I_β- is the index set of rows in which both elements are not essential $(c_{i\,j}, c_{i\,p} = \lambda)$.

An l g o r I t h m check consistency determined.

Check Completeness: According to Definition 1 (with the substitution of N instead of S), decision table is complete concerning to N, if $N \subseteq S^1$. Strict inclusion means that there are some non-empty intersections of elements from S_i^1 . A decision table is then called *redundant* concerning to N.

Check completeness of a decision table (containing no rule E) will be carried out, comparing the number of solutions rules presented in the table, with the number of H syntactically possible rules of solving.

Proposition 1: $H = \prod_{i=1}^{m} |\hat{C}_i|,$ Where $\hat{C}_i$ - is the set of values of conditions C_i?

Proposition 2: $G = \sum_{z=2}^{u} (-1)^z B(z),\ u \leq n,$ Where $B\ (z)$ – is number of data vectors, contained in the various intersection the elements from S^1 to Z, or el0……………………………..se the number of data vectors satisfying simultaneously to Z rules of solving.

Corollary 3: A decision table is complete concerning to N, if $F – G = H$; incomplete concerning to N, if

F - $G<H$; excess concerning to N, if F- $G>H$.

We now give a method of calculating $B\ (z)$ by a matrix $\tilde{C}^k$. using the principle of mathematical induction and the lemma; we can prove the following theorem.

Theorem: In order that there be a data vector S_q such that $(\forall i)(s_q \to R_{j\,i}),\ R_{j\,i} \in R,\ i \in \{1, 2, ..., z\},\ z \leq n,$ it is necessary and sufficient to satisfy the relation $(S_1^k \sim ... \sim S_j^k \sim ... \sim S_z^k),$ where $K_z = \{K\}$ – is set of combinations from n to t vectors $|K_z| = \binom{z}{n},\ z \leq n, S_j^k - j$ -the vector-column of k-the combination of the vectors S_j, $k \in K_z$.

An l g o r I t h m check completeness determined.

CONCLUSION

The isomorphism between the decision tables and production structures making allows us to consider the proposed correctness control scheme as a base for production-tabular systems in general. Moreover, for systems with limited input (limited-entry) and also for systems with extended inlet (extended-entry).

It should also be noted that the scheme can be used both at the stage of development of production-tabular systems and their possible modifications during operation. This is important when working in an the open and dynamic problem areas characterized by high demands on reliability and timeliness of decisions

The proposed correctness check algorithms are used in the “System reactive diagnostics of LAN Ethernet” [6], in the “System on-line diagnostics of power plants" [7] and in the "System of predicting the preservation of sinus rhythm after the elimination of a ciliary arrhythmia”.

REFERENCES

1. Russell, S. and P. Norvig, 1997. Artificial Intelligence: a modern approach. N.J.: Prentice Hall.
2. Hamby, E., 1976. Programming decision tables. N.Y. J. Wiley & Sons Inc.
3. The current status of Expert System development and related technologies, 1995. IEEE Expert, pp: 4.
4. Pollack, S., 1971. Decision Tables: Theory and Practice. N.Y. J. Willey & Sons Inc.
5. King, P., 1986. Ambiguity in limited entry decision tables. Comm. ACM, 2(10): 21-29.

6. Mohamed Gamal, I.O., 2009. Reference Maps for the reactive diagnosis of LAN Ethernet. In the Proceedings of the Intern. Conf. on Telecommunications Technology. Minsk: Belarus State University.
7. El-Sayed, A., 2010. The Application of Diagnosis Knowledge-Based System to Power Plant Operations. In the Proceedings of the 15th Intern. Power Systems Conf. IPSC'2010. Karlsruhe, Germany.