Belief Rules vs. Decision Rules: A Preliminary Appraisal of the Problem

Jerzy W. Grzymała-Busse[1,2] and Zdzisław S. Hippe[1] Teresa Mroczek[1],

[1] University of Information Technology and Management,
ul. Sucharskiego 2, 35-225 Rzeszw, Poland,
zhippe,tmroczek@wenus.wsiz.rzeszow.pl

[2] Department of Electrical Engineering and Computer Science, University of Kansas,
Lawrence KS 66045-7523, USA
jerzy@ku.edu

Abstract. An in-house developed computer program system ***Belief*SEEKER**, capable to generate belief networks and to convert them into respective sets of belief rules, was applied in mining the melanoma database. The obtained belief rules were compared with production rules generated by **LERS** system. It was found, that belief rules can be presumably treated as a generalization of standard **IF...THEN** rules.

Keywords: belief networks, belief rules, production rules, ***Belief*SEEKER**, **LERS**

1 Introduction

In this research two distinct learning models were generated (using an in-house developed computer program system ***Belief*SEEKER** [1]), and then applied for classification of unseen cases of melanoma skin lesions. The first model was based on belief networks, whereas the second one was grounded on typical decision rules. The quality of the developed learning models was additionally compared with results gained by the program **LERS** [2], a well-known standard of rules generating systems.

2 Research tools used

– ***Belief*SEEKER** - is a computer program system, capable to generate learning models (for any type of decision table prepared in the format set about by Pawlak [3]) in a form of belief nets, applying various algorithms [4]. The development of belief networks is steadily controlled by a specific parameter, informing about the maximum dependence between variables, known as marginal likelihood:

$$ML = \prod_{i=1}^{v} \prod_{j=1}^{q_i} \frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij} + n_{ij})} \prod_{k=1}^{c_i} \frac{\Gamma(\alpha_{ijk} + n_{ijk})}{\Gamma(\alpha_{ijk})} \tag{1}$$

Here:
i=1,...,v - where v is the number of nodes in the network,
$j = 1, ..., q_i$- is the number of possible combinations of parents of the node X_i. (if a given attribute does not contain nodes of the type "parent", then q_i get the value of 1),
$k = 1, ..., c_i$ - where c_i is the number of classes within the attribute X_i,
n_{ijk} - is the number of rows in the database, for which parents of the attribute X_i have value *j*, and this attribute has the value of *k*, and
α_{ijk}, α_{ij} - are parameters of the initial Dirichlet's [5] distribution.

Characteristic features of the system, yet not described by us are: (**i**) capability to generate various exhaustive learning models (Bayesian networks) for different values of Dirichlet's parameter [5], (**ii**) capability to convert generated belief networks into relevant sets of belief rules of the type **IF...THEN**, (**iii**) built-in classification mechanism of unseen cases, and (**iv**) built-in mechanism for the assessment of obtained rules. It is worth to mention, that developed learning models (belief nets) can be converted into some sets of belief rules, characterizes by a specific parameter called by us *belief factor* **BF**, that reveals indirectly the influence of the most significant descriptive attributes on the dependent variable. Also, to facilitate the preliminary evaluation of generated rules, additional mechanism supports the calculation of their specifity, strength, generality, and accuracy [6].

– **LERS**[7] - data mining system (acronym from **L**earning from **E**xamples based on **R**ough **S**ets), developed at the University of Kansas, Lawrence, KS, USA, induces a set of rules from any kind of data (even from inconsistent data) and classifies new cases using the set of rules induced previously. These rules are more general than information contained in the original input database, since more new cases may be correctly classified by rules than may be matched with cases from the original data. More details about the system **LERS** can be find elsewhere [8].

3 Experiments. Chosen results and discussion

The well-known database [9] containing 548 cases of melanoma skin lesions was used in our research. It should be stressed that each case stored in the database was diagnosed, and confirmed histopathologically as belonging to one of the four concepts: *Benign nevus*, *Blue nevus*, *Suspicious nevus*, and *Melanoma malignant.* The database was randomly divided into two independent subsets. The first subset (**E384144.TAB**; **384** cases, **14** attributes, **4** concepts) was used for development of learning models, whereas the second (**E164144.TAB**; **164** cases, **14** attributes, **4** concepts) served for testing models. Information regarding the distribution of concepts in these sets is shown in Table 1.
In the first step of the research, computer program system ***Belief*SEEKER** was used to generate three different belief networks, having **Dirichlet's** para-

Table 1. Distribution of investigated cases in learning and training sets

Diagnosed concept	**E384144.tab** (learning set)	**E164144.tab** (testing set)
Benign nevus	160	88
Blue nevus	64	14
Suspicious nevus	80	34
Melanoma malignant	80	28

meter α=1, 30 and 60 respectively. The basic difference in the structure of these networks relied on the number of the most distinctive (important) attributes, displaying direct influence on the dependent variable (type of skin lesion, the decision). It was found, for example, that for the network with α=1 two descriptive attributes, namely $< TDS >$ and $< color\ blue >$ seemed to be the most important. Then, for networks with α=30, additional attribute ($< asymmetry >$), appeared in the network. Finally, in the case of the third network, besides attributes pointed out earlier, the attribute $< dark\ brown >$ was observed.
The significance of these attributed was of great importance. Based on this information - in the next step of the research - for variable values of **BF** parameter - various sets of rules were generated for each of the networks (for ***Belief*****SEEKER**), and one, unique set of rules was generated by **LERS** (see Table 2 and Table 3). It was found that the best results were obtained for **BF** = 0.7.

Table 2. The summary of results of classification of unseen cases by ***Belief*****SEEKER** and **LERS**

	Belief**SEEKER**			**LERS**
	Dirichlet's parameter			
	α**=1**	α**=30**	α**=60**	
Number of rules	20	31	42	49
Number of cases classified correctly [%]	95.12	95.12	87.80	87.8
Number of cases classified incorrectly [%]	4.88	4.27	4.88	6.10
Number of cases unclassified [%]	0	0.61	7.32	6.10

In the last step of the research, selected information (like specificity, strength, generality and accuracy of generated rules) allowed to fix the strongest belief rules and decision rules (Table 4).
It can be assumed that, the increase of Dirichlet's parameter causes extending the most crucial belief rules (Table 4) by merging another attributes. It was also

Table 3. Occurrences of selected attributes in belief rules (***Belief*** **SEEKER**) and decision rules (**LERS**)

Descriptive attributes		***Belief*** **SEEKER** Dirichlet's parameter α=1	***Belief*** **SEEKER** Dirichlet's parameter α=30	***Belief*** **SEEKER** Dirichlet's parameter α=60	**LERS**
	asymmetry		31	42	16
	border				25
Color	*black*				11
	blue	20	31	42	18
	dark brown			42	5
	light brown				6
	red				16
	white				11
Diversity of structure	*pigment dots*				11
	pigment globules				6
	pigment network				15
	structureless areas				9
	branched streaks				9
	TDS	20	30	42	33

Table 4. Selected examples of the most important belief rules and decision rules

Belief **SEEKER** Dirichlet's parameter α=1	***Belief*** **SEEKER** Dirichlet's parameter α=30	***Belief*** **SEEKER** Dirichlet's parameter α=60	**LERS**
RULE 5 IF TDS >= 4.080 AND TDS < 4.850 AND C_BLUE IS absent THEN DIAGNOSIS IS Benign_nev	**RULE 3** IF TDS >=4.080 AND TDS < 4.850 AND C_BLUE IS absent AND ASYMMETRY IS 1_axial_as THEN DIAGNOSIS IS Benign_nev	**RULE 6** IF TDS >= 4.080 AND TDS < 4.850 AND C_BLUE IS absent AND ASYMMETRY IS 1_axial_as AND C_d_BROWN IS present THEN DIAGNOSIS IS Benign_nev	**RULE 1** IF C_BLUE IS absent AND C_d_BROWN IS present AND C_RED IS absent AND D_PIGM_DOTS IS present AND ASYMMETRY IS sym_change THEN DIAGNOSIS IS Benign_nev
RULE 4 IF TDS >= 3.310 AND TDS < 4.080 AND C_BLUE IS absent THEN DIAGNOSIS IS Benign_nev	**RULE 7** IF TDS >= 3.310 AND TDS < 4.080 AND C_BLUE IS absent AND ASYMMETRY IS sym_change THEN DIAGNOSIS IS Benign_nev	**RULE 10** IF TDS >= 3.310 AND TDS < 4.080 AND C_BLUE IS absent AND ASYMMETRY IS sym_change AND C_d_BROWN IS present THEN DIAGNOSIS IS Benign_nev	**RULE 15** IF C_BLUE IS absent AND TDS >= 4.75 AND TDS < 5.35 AND C_WHITE IS present AND C_BLACK IS absent AND D_PIGM_NETW IS absent THEN DIAGNOSIS IS Benign_nev
RULE 19 IF TDS >= 4.850 AND TDS < 5.620 AND C_BLUE IS absent THEN DIAGNOSIS IS Suspicious	**RULE 28** IF TDS >= 4.850 AND TDS < 5.620 AND C_BLUE IS absent AND ASYMMETRY IS 1_axial_as THEN DIAGNOSIS IS Suspicious	**RULE 39** IF TDS >= 4.850 AND TDS < 5.620 AND C_BLUE IS absent AND ASYMMETRY IS 1_axial_as AND C_d_BROWN IS present THEN DIAGNOSIS IS Suspicious	**RULE 24** IF D_STREAKS IS absent AND D_PIGM_DOTS IS absent AND C_BLUE IS present THEN DIAGNOSIS IS Blue_nevus

observed that strongest rules generated from the first network (α=1), became a base for rules generated from other networks (α=30 and 60).
The strongest decision rules generated by **LERS** system present more detailed study of cases, mainly because they contained greater number of descriptive attributes than respective belief rules, moreover, the attributes used were often completely different. Only the attribute <*color blue*> was common for both sets of rules. Considering the overall results of classification (Table 2) the best results were obtained for belief networks with α=1 or α=30. Expanding the set of rules (α=60) with another attribute - <*color dark brown*> - did not improve the classification process. It may be pointed out, that generalizing of the rule sets yielded positive effects in term of classification process. Results of the carried out research has also showed that belief rules (produced by ***Belief*** **SEEKER**) seemed to be some kind of generalization of rules developed by **LERS** system. In the future research, the detailed foundation of this interesting finding will be dealt with.

References

1. P. Błajdo, J.W. Grzymała-Busse, Z.S. Hippe, M. Knap, T. Marek, T. Mroczek, M. Wrzesień: *A suite of machine learning tools for knowledge extraction from data*, **In:** R. Tadeusiewicz, A. Ligęza, M. Szymkat (Eds.), Computer Methods and Systems in Scientific Research, Edition of "Oprogr. Naukowe", Cracow 2003, pp. 479-484 (in Polish).
2. M.R. Chmielewski, J.W. Grzymała-Busse, Neil W. Peterson, S. Than: *The Rule Induction System LERS - A Version for Personal Computers*, **Foundations of Computing and Dec. Sciences 1993** (18, No 3-4) pp. 181-211.
3. Z. Pawlak: *Knowledge and rough set*, **In**: W. Traczyk (Eds.), Problem of artificial intelligence, Wiedza i życie - Warsaw 1995, pp. 9-21 (in Polish).
4. T. Mroczek, J.W. Grzymała-Busse, Z.S. Hippe: *Rules from belief networks: A Rough Set Approach*, **In**: S. Tsumoto, R. Słowiński, J. Komorowski, J.W. Grzymała-Busse (Eds.), Rough Sets and Current Trends in Computing, Springer-Verlag, Uppsala, Sweden 2004, pp.483-487
5. D. Heckerman: *A Tutorial on Learning Bayesian Networks*, Technical report MSR-TR-95-06. heckerman@microsoft.com.
6. Z.S. Hippe: *Design and Application of New Knowledge Engineering Tool for Solving Real World Problems*, **Knowledge-Based Systems 9**(1996)509-515.
7. J.W. Grzymała-Busse: *A New Version of the Rule Induction System* ***LERS***, **Fundamenta Informaticae 31**(1997)27-39.
8. J.W. Grzymała-Busse, W. Ziarko: *Data Using Based on Rough Sets*, **In**: J.Wang (Eds.), Data Mining: Opportunities and Challenges, Idea Group Publishing, Hershey 2003, pp. 142-173.
9. A. Alvarez, S. Bajcar, F.M. Brown, J.W. Grzymała-Busse, Z.S. Hippe: *Optimization of the ABCD Formula Used for Melanoma Diagnosis*, **In**: M.A. Kłopotek, S.T. Wierzchoń, K. Trojanowski (Eds.), Advances in Soft Computing, Springer-Verlag, Heidelberg 2003, pp.233-240.