

# Diagnosis of Melanoma Using IRIM, a Data Mining System

Jerzy W. Grzymala-Busse<sup>1</sup>, Jay Hamilton<sup>2</sup>, and Zdzislaw S. Hippe<sup>3</sup>

<sup>1</sup> Department of Electrical Engineering and Computer Science, University of Kansas,  
Lawrence, KS 66045, USA

and

Institute of Computer Science Polish Academy of Sciences, 01-237 Warsaw, Poland  
Jerzy@ku.edu

<http://lightning.eecs.ku.edu/index.html>

<sup>2</sup> Department of Electrical Engineering and Computer Science, University of Kansas,  
Lawrence, KS 66045, USA

CSJay@ku.edu

<sup>3</sup> Department of Expert Systems and Artificial Intelligence, University of  
Information Technology and Management, 35-225 Rzeszow, Poland  
ZHippe@venus.wsiz.rzeszow.pl

**Abstract.** Melanoma is a very dangerous skin cancer. In this paper we present results of experiments on three melanoma data sets. Two data mining tools were used, a new system called IRIM (Interesting Rule Induction Module) and well established LEM2 (Learning from Examples Module, version 2), both are components of the same data mining system LERS (Learning from Examples based on Rough Sets). Typically IRIM induces the strongest rules that are possible for a data set. IRIM does not need any preliminary discretization or preprocessing of missing attribute values. Though performance of IRIM and LEM2 is fully comparable, IRIM provides an additional opportunity to induce unexpected and strong rules supplying an important insight helpful for diagnosis of melanoma.

## 1 Introduction

Melanoma is a very dangerous skin cancer. Moreover, the number of diagnosed cases increase each year. Any improving melanoma diagnosis will save human lives.

In our research we used three versions of a basic data set with 410 cases that was collected at the Regional Dermatology Center in Rzeszow, Poland [6]. This data set describes every case in terms of thirteen attributes: Asymmetry, Border, six binary attributes associated with Color, five binary attributes associated with Diversity, and an additional attribute called TDS (Total Dermatoscopic Score), computed on the basis of the ABCD formula [12].

We used a new component of LERS (Learning from Examples based on Rough Sets) called IRIM (Interesting Rule Induction Module). Rough set theory

was initiated in 1982 [9]. IRIM does not induce rule sets in a conventional fashion as standard rule induction algorithms such as LEM1 or LEM2 (Learning from Examples Module, version 1 and 2, respectively) [2]. IRIM resembles the ALL RULES algorithm, part of the LERS [2] and the EXPLORE algorithm [11]. First of all, rules created by IRIM may not cover all positive cases of the concept. Secondly, there is a lot of overlapping and even subsuming between rules induced by IRIM. Instead, IRIM induces the strongest possible rules, covering the most positive cases of the concept. Such rules have a potential to be not only new but also interesting and surprising to experts in the area, hence the name of the module.

An approach to data mining based on inducing very few rules, for example, a single rule, was introduced in [8], and then was continued in a number of papers, see, e.g., [7]. Rule truncation [8], i.e., removal of weak rules while keeping stronger rules in the rule set belongs to the same category. This technique proved its applicability in diagnosing melanoma [3]. A similar technique, pruning of decision trees, see, e.g., [10], is used in tree generation systems as well.

The main objective of our research was to check usefulness of IRIM in melanoma diagnosis by comparing the performance of IRIM versus a standard rule induction module LEM2. As it is known [3], [4], performance of LEM2 is fully comparable with that of melanoma diagnosticians. Results of our experiments show that performance of IRIM is comparable with performance of LEM2.

## 2 Data mining tools

The main tool for our experiments was a recent addition to LERS, called IRIM, created for inducing the strongest rules describing a concept. For every concept IRIM creates a rule set with rules satisfying some pre-defined input parameters: the minimum rule length (i.e., number of rule conditions), the maximum rule length, and minimum of conditional probability of the concept given rule domain. The rule domain is the set of all cases satisfying the left hand side of the rule. For brevity, the minimum of conditional probability of the concept given rule domain will be called a ratio parameter.

The algorithm LEM2 needs discretization, a preprocessing, to deal with numerical attributes. Discretization is a process of converting numerical attributes into symbolic attributes, with intervals as values. On the other hand, IRIM induces rules during discretization. IRIM computes first the set of blocks for all attribute-value pairs  $(a, v)$ . If  $t = (a, v)$  is an attribute-value pair then a block of  $t$ , denoted  $[t]$ , is a set of all cases from  $U$  that for attribute  $a$  have value  $v$ . IRIM recognizes integer and real numbers as values of attributes, and labels such attributes as numerical. For numerical attributes IRIM computes blocks in a different way than for symbolic attributes. First, it sorts all values of a numerical attribute. Then it computes cutpoints as averages for any two consecutive values of the sorted list. For each cutpoint  $c$  IRIM creates two blocks, the first block contains all cases for which values of the numerical attribute are smaller than  $c$ , the second block contains remaining cases, i.e., all cases for which values

of the numerical attribute are larger than  $c$ . The search space of IRIM is the set of all blocks computed this way, together with blocks defined by symbolic attributes. Then IRIM combines attribute-value pairs relevant to a concept and creates rules describing the concept, taking into account pre-defined, by the user, input parameters. In addition, IRIM handles missing attribute values during rule induction. For any attribute with missing values, blocks are computed only from the existing attribute-value pairs [5].

The classification system of LERS [6] is a modification of the bucket brigade algorithm. The decision to which concept a case belongs is made on the basis of three factors: strength, specificity, and support. They are defined as follows: Strength is the total number of cases correctly classified by the rule during training. Specificity is the total number of attribute-value pairs on the left-hand side of the rule. The matching rules with a larger number of attribute-value pairs are considered more specific. The third factor, support, is defined as the sum of scores of all matching rules from the concept. The concept for which the support is the largest is the winner and the case is classified as being a member of that concept. Every rule induced by LERS is preceded by three numbers: specificity, strength, and rule domain size.

The output of IRIM is the set of all rules satisfying input parameters. In general, IRIM generates very big rule sets. The worst time complexity of IRIM is exponential with respect to the number of attributes. Hence, in our experiments, we selected the maximum rule length to be equal to two or three and the ratio to equal to at least 0.8.

### 3 Experiments

We conducted our experiments on three data sets on melanoma. Two data sets were discretized using divisive cluster analysis [4]. The third was the original data set, with original numerical attributes. In the first discretized data set, in [m] denoted by "0-0", the attribute TDS was optimized by using a new ABCD formula [1]. In the second discretized data set, denoted by "div", the attribute TDS before discretization was computed using the traditional ABCD formula. In the original, not discretized data set, called "original", the TDS attribute was computed using the traditional ABCD formula. In all of our experiments for testing data the LERS classification system was applied. In our first sequence of experiments the number of errors, out of 410 cases, was determined using ten-fold cross validation. Results are presented in Table 1. Note that the number of errors for "0-0" and "div" data sets and LEM2 system were equal to 10 and 15, respectively.

In the third, original data set, some attributes were numerical. This causes very time-consuming rule induction by IRIM. However, during ten-fold cross validation we observed that in different folds the same strong rules were consistently induced. Due to this observation, an additional experiment was conducted, to compare the number of errors determined by ten-fold cross validation and re-substitution. In re-substitution the same data are used for training and testing.

Number of errors				
Data set "0-0"		Data set "div"		
Number of rules	Max. rule length = 2	Max rule length = 3	Max rule length = 2	Max. rule length = 3
1	22	22	9	9
2	25	33	11	9
3	35	34	10	9
4	36	35	10	9
5	36	37	10	9
6	36	37	10	9
7	36	36	10	10
8	36	36	10	10
9	36	36	9	10
10	36	36	9	10

**Table 1.** Rule sets induced by IRIM

Obviously, resubstitution is much less time-consuming than ten-fold cross validation. Since we wanted to test up to five strongest rules induced by IRIM from the original data, we compared ten-fold cross validation with resubstitution for up to five rules, see Tables 2 and 3.

Number of errors				
Ten-fold cross validation		Resubstitution		
Number of rules	Max. rule length = 2	Max rule length = 3	Max rule length = 2	Max. rule length = 3
1	22	22	22	22
2	25	33	22	36
3	35	34	34	36
4	36	35	37	36
5	36	37	37	36

**Table 2.** Errors determined by ten-fold cross validation versus errors determined by resubstitution for data set "0-0"

Goodness-of-fit test returns  $\chi^2 = 1.027$ , this value is smaller than  $\chi_{0.05}^2 = 3.073$  for  $(4-1)(5-1) = 12$  degrees of freedom, so the null hypothesis cannot be rejected at the level 0.05 of significance and we conclude that the approximate test, using resubstitution, provided a good fit for ten-fold cross validation.

The results of using resubstitution for the strongest rules induced by IRIM from the original data are presented in Table 4. Due to enormous computational complexity we restricted our attention to rules of the maximal length equal to two and the ratio equal to 0.9. Also, selection of rules to the rule sets was not automatic, as in the previous experiments, with discretized data sets. This

Number of errors				
Ten-fold cross validation		Resubstitution		
Number of rules	Max. rule length = 2	Max rule length = 3	Max rule length = 2	Max. rule length = 3
1	9	9	9	9
2	11	9	11	9
3	10	9	10	9
4	10	9	10	9
5	10	9	11	9

**Table 3.** Errors determined by ten-fold cross validation versus errors determined by resubstitution for data set "div"

Number of rules	Number of errors
1	9
2	13
3	11
4	11
5	10

**Table 4.** Errors determined by IRIM and resubstitution from the original data set

time, selection was conducted manually since consecutive rules in the sorted list differed only slightly. For example, for rules describing benign cases of melanoma, the two strongest rules, at the beginning of the sorted list, were

2, 144, 149

(C\_BLUE, 0) & (TDS, 1..4.85) -> (MELANOMA, Benign\_nev)

2, 144, 149

(C\_BLUE, 0) & (TDS, 1.25..4.85) -> (MELANOMA, Benign\_nev)

Obviously, the second rule does not introduce a lot of new knowledge, so that rule was ignored in the process of identifying the strongest rules.

## 4 Conclusions

From the results of our experiments it is clear that IRIM may induce high quality rules. For the data set "0-0" IRIM was worse than LEM2, but for the data set "div" it was better. The two data sets differ by the way TDS was computed. Since in the data set "0-0" the optimal ABCD formula was selected particularly for LEM2, it explains why the performance of LEM2 was better.

Moreover, inducing rule sets directly from raw data, not discretized, IRIM again performed very well. It is surprising that the number of errors may grow with the number of rules in the data set. Nevertheless, IRIM has a potential to induce the best possible rule set, because the rule set induced by any other rule induction system is a subset of the set of all possible rules induced by IRIM, the only problem is selecting this rule set from a very large rule set induced by

IRIM. Thus, the problem needed to be solved in the future is how to automate selection of the best rule set from the superset of all rules induced by IRIM. Finally, rules induced by IRIM may be used not only for classification but also to provide a new insight into the problem for domain experts. Such experts may learn strong regularities, hidden in the data, that were not known before use of data mining.

## References

1. Alvarez, A., Brown, F. M., Grzymala-Busse, J. W., and Hippe, Z. S.: Optimization of the ABCD formula used for melanoma diagnosis. Proc. of the IIPWM'2003, Int. Conf. On Intelligent Information Processing and WEB Mining Systems, Zakopane, Poland, June 2-5, 2003, 233-240.
2. Grzymala-Busse, J. W.: LERS—A system for learning from examples based on rough sets. In Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory. Slowinski, R. (ed.), Kluwer Academic Publishers, Dordrecht, Boston, London, 1992, 3-18.
3. Grzymala-Busse J. W. and Hippe Z. S.: Postprocessing of rule sets induced from a melanoma data set. Proc. of the COMPSAC 2002, 26th Annual International Conference on Computer Software and Applications, Oxford, England, August 26-29, 2002, 1146-1151.
4. Grzymala-Busse J. W. and Hippe Z. S.: A search for the best data mining method to predict melanoma. Proceedings of the RSCTC 2002, Third International Conference on Rough Sets and Current Trends in Computing, Malvern, PA, October 14-16, 2002, Springer-Verlag, 538-545.
5. Grzymala-Busse, J. W. and Wang A. Y.: Modified algorithms LEM1 and LEM2 for rule induction from data with missing attribute values. Proc. of the Fifth International Workshop on Rough Sets and Soft Computing (RSSC'97) at the Third Joint Conference on Information Sciences (JCIS'97), Research Triangle Park, NC, March 2-5, 1997, 69-72.
6. Hippe, Z. S.: Computer database NEVI on endargement by melanoma. Task Quarterly 4, 1999, 483-488.
7. R. C. Holte. Very simple classification rules perform well on most commonly used datasets. Machine Learning 11 (1993) 63-90.
8. R. S. Michalski, I. Mozetic, J. Hong, N. Lavrac.: The AQ15 Inductive Learning System: An Overview and Experiments. Intelligent System Group, University of Illinois at Urbana-Champaign, ISG 86-20, 1986.
9. Pawlak, Z.: Rough Sets. International Journal of Computer and Information Sciences, 11 (1982) 341-356.
10. Quinlan, J. R.: C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, San Mateo, CA, 1988.
11. J. Stefanowski.: Algorithms of Decision Rule Induction in Data Mining. Poznan University of Technology Press, Poznan, Poland, 2001.
12. Stolz, W., Braun-Falco, O., Bilek, P., Landthaler, A. B., Cogneta, A. B.: Color Atlas of Dermatology, Blackwell Science Inc., Cambridge, MA, 1993.