An algorithm which learns multiple covers via integer linear programming

Part II: experimental results and conclusions

Krzysztof J. Cios
University of Toledo, Ohio, USA, and
Ning Liu
Dana Corporation, Ottawa Lake, Michigan, USA

Experiments and results

The CLILP2 algorithm is tested on data from three medical domains: lymphatic cancer, prognosis of breast cancer recurrence, and location of primary tumour[22]. The reason for using these particular data sets is that the results obtained by others[7, 8], are easily available for comparison. We also select the same training and test examples. Namely, for all three domains, 70 per cent of examples are selected randomly for training and the remaining 30 per cent are selected for testing. Each experiment is repeated four times, and the results are averaged.

Because CLILP2 generates multiple hypotheses, it requires the specification of the number of hypotheses to be generated. In all our experiments this number was set to 4. As a result, the average was taken from 16 results (4×4 hypotheses). The generation of a hypothesis depends on two parameters: the noise-distortion threshold and decision rule truncation threshold. A hypothesis is obtained by searching for the best result from cases in which the noise-distortion threshold was 30, 40, and 50 per cent, and in which the decision rule truncation threshold was 0.90, 0.85 and 0.80. Histograms are used to show the selection of the parameters.

Lymphatic cancer data are characterized by four decision classes, and the examples are described by 18 attributes. There are 148 examples. The data are consistent, i.e. there is no identical example in more than one class. The actual division of learning and testing examples is shown in Table I.

Editor's note: Part I appeared in *Kybernetes*, Vol. 24 No. 2, 1995, pp. 29-50. Parts I and II should be read in conjunction with the authors' previous article, "Generation of diagnostic rules via inductive machine learning", *Kybernetes*, Vol. 22 No. 5, 1993, pp. 44-56.

This work was partially supported by the NSF grant number DDM-9015333, the American Heart Association grant number NW-90-02, and the University of Toledo Research Support Program. The authors would also like to thank Dr Igor Mozetic for permission to use the lymphatic cancer, primary tumour and breast cancer data.

Kybernetes, Vol. 24 No. 3, 1995, pp. 24-36. © MCB University Press, 0368-492X

The prognosis of breast cancer data is characterized by two classes and nine attributes. The data are inconsistent: some examples, from different classes, are identical. The actual division into training and testing examples is shown in Table II.

The location of a primary tumour is characterized by 22 classes and 17 attributes. The data are also inconsistent. There are 339 examples. Some classes contain several identical examples for both training and testing. Test results from these classes are not representative since misrecognizing one or two test examples affects the accuracy greatly. The actual division into training and testing examples is shown in Table III.

To generate decision rules for a class, the learning examples of that class are treated as positive training examples, and the learning examples of other classes are treated as negative training examples. For the three medical domains there will be three sets of decision rules generated. Since CLILP2 generates unordered decision rules, the rules can be used either individually for classifying test examples or as a set for diagnosing diseases. The generated concept descriptions are tested using two methods described in the next two sections. The first one, rule verification test, in our opinion better evaluates the results. The second method is used only for comparison.

Rule verification test

The decision rules generated by CLILP2 are like classification templates, and each template can be individually used to determine the class identities of test data. The effectiveness of a decision rule is verified by how accurately it classifies test examples. In an ideal situation, a decision rule should recognize all the positive test examples but none of the negative test examples. In other words, it should draw an accurate decision boundary between the positive and negative examples.

Class	Learning	Testing	Total
Normal	1	1	2
Metastases	57	24	81
Lymphoma	43	18	61
Fibrosation	3	1	4
Total	104	44	148

Table I.
Example distribution of training and test data for lymphatic cancer

Class	Learning	Testing	Total
No recurrence	141	60	201
Recurrence	60	25	85
Total	201	85	286

Table II.
Example distribution of training and test data for breast cancer

Class	Learning	Testing	Total
Bladder	1	1	2
Breast	17	7	24
Cervical uteri	1	1	2
Colon	10	4	14
Corpus uteri	4	2	6
Duodenum/intestine	1	0	1
Oeophagus	6	3	9
Gall bladder	11	5	16
Head/neck	14	6	20
Kidney	17	7	24
Liver	5	2	7
Lung	59	25	84
Ovary	20	9	29
Pancreas	20	8	28
Prostate	7	3	10
Rectum	4	2	6
Salivary glands	1	1	2
Stomach	27	12	39
Testis	1	0	1
Thyroid	10	4	14
Vagina	1	0	1
Anus	0	0	0

Table III.
Example distribution of training and test data for primary tumour

The rule verification test is designed to investigate the actual decision boundaries set by the generated decision rules.

When a positive test example is true for any one of the decision rules that are logically disjoint for a class, it is defined as true positive, otherwise it is false negative. If a negative example does not satisfy any of the decision rules, it is defined as true negative, otherwise it is false positive. There are three evaluation criteria for this test[23]:

$$(1)\quad \text{sensitivity} = \frac{\text{true positive}}{\text{true positive} + \text{false positive}}$$

$$(2)\quad \text{specificity} = \frac{\text{true negative}}{\text{true negative} + \text{false positive}}$$

$$(3)\quad \text{predictive accuracy} = \frac{\text{true positive} + \text{true negative}}{\text{true positive} + \text{false positive} + \text{true negative} + \text{false positive}}$$

The sensitivity measures how many of the positive test examples are recognized, and the specificity measures how many of the negative test

examples are excluded. Predictive accuracy gives the overall evaluation. The syntactic complexity of decision rules is also evaluated in terms of how many complexes are included in a hypothesis (averaged over 16 results).

The results of the lymphatic cancer data are shown in Table IV. One may notice that two classes, fibrosation and normal, have near perfect specificities, but the sensitivities are very low. However, due to a big difference in number between positive and negative test examples (1:43), the predictive accuracies are high: 97.3 and 98.6 per cent respectively. Decision rules for lymphoma and metastases are tested using a "better" ratio of positive to negative test examples, 18:26 and 24:20 respectively, and thus yield more reliable results.

To evaluate the algorithm's performance for a given domain, the averages of the total positive/negative and true positive/negative from the four classes are used to calculate the three evaluation criteria. The number of complexes of a domain is the sum of the examples from all classes. In Tables IV-VI, four digits after the decimal point are displayed so it is easy to verify how the averages are calculated.

For breast cancer recurrence results (Table V), if a negative example was identical to a positive example, this negative example was eliminated. Test examples were used without checking for consistency.

Test results for the location of a primary tumour are shown in Table VI. Decision rules of many classes are tested to recognize the few available positive test examples and to exclude the large number of negative test examples. From Table VI we notice that the results of the classes with less than five positive test examples are very poor in terms of sensitivity, although the corresponding specificity and predictive accuracy can be quite high. From this observation we

Class	Total positive	Total negative	True positive	True negative	Sensitivity (per cent)	Specificity (per cent)	Predictive accuracy (per cent)	Number of complexes
Fibrosation	1	43	0.2500	42.5625	25.0	99.0	97.3	1.0
Lymphoma	18	26	15.5000	21.5000	86.1	82.7	84.1	7.6
Metastases	24	20	21.6875	14.5625	90.4	72.8	82.4	8.7
Normal	1	43	0.7500	42.6250	75.0	99.1	98.6	1.0
Average	11	33	9.5469	30.3125	86.8	91.9	90.6	18.3

Table IV. Test results for lymphatic cancer data

Class	Total positive	Total negative	True positive	True negative	Sensitivity (per cent)	Specificity (per cent)	Predictive accuracy (per cent)	Number of complexes
Recurrence	25	60	13.8750	47.9375	55.5	79.9	72.7	19.1
No recurrence	60	25	42.0000	13.0625	70.0	52.3	64.8	21.3
Average	42.5	42.5	27.9375	30.5000	65.7	71.8	68.8	40.4

Table V. Test results for breast cancer data

Class	Total positive	Total negative	True positive	True negative	Sensitivity (per cent)	Specificity (per cent)	Predictive accuracy (per cent)	Number of complexes
Bladder	1	101	0.6525	98.7500	65.3	97.8	97.4	1.0
Breast	7	95	6.0625	81.3750	86.6	85.7	85.7	4.4
Cervical uteri	1	101	0.0000	99.4375	0.0	98.5	97.5	1.0
Colon	4	93	1.6525	76.3125	39.1	77.9	76.3	6.8
Corpus uteri	2	100	0.3750	85.2500	37.5	85.3	84.0	3.8
Duodenum/ intestine	0	102	0.0000	99.0625	0.0	97.1	97.1	1.0
Oesophagus	3	99	0.9375	88.6875	31.3	89.6	87.9	4.1
Gall-bladder	5	97	3.3750	75.3125	76.5	77.6	77.1	6.1
Head/neck	6	96	5.2500	93.9375	87.6	97.9	97.2	1.8
Kidney	7	95	4.6250	67.8125	66.1	71.4	71.0	8.6
Liver	2	100	0.0625	86.0625	3.3	86.1	84.4	4.3
Lung	25	77	18.9375	61.5625	75.8	80.0	78.9	17.3
Ovary	9	93	7.9375	79.9375	88.2	86.0	86.2	4.5
Pancreas	8	94	4.3125	58.1250	53.9	61.2	61.2	12.3
Prostate	3	99	1.6250	85.2500	54.2	86.1	85.2	3.5
Rectum	2	100	1.2500	88.6250	62.5	88.6	88.1	2.8
Salivary glands	1	101	0.5625	98.8750	56.3	97.9	97.5	1.0
Stomach	12	90	6.3750	53.2500	53.1	59.2	58.5	15.9
Testis	0	102	0.0000	101.8750	0.0	99.9	99.9	1.0
Thyroid	4	98	1.3750	85.0000	34.4	86.7	84.7	5.5
Vagina	0	102	0.0000	101.4375	0.0	99.5	99.5	1.0
Anus	0	0	0	0	0	0	0	0
Average	4.6364	92.7273	2.9700	80.2653	64.1	86.6	85.5	107.8

Table VI.
Test results for primary tumour location data

draw a conclusion that a higher level of confidence can be placed only for results which yield high values for all three measures (sensitivity, specificity, and diagnostic accuracy).

Histograms depicted in Figures 5 and 6 show how the hypotheses are selected for different parameter combinations. Most of the hypotheses are generated using a 0.90 decision rule truncation threshold but since the decision rules are short, it truncates no more than one feature test from a complex. From Figure 6 we notice that the high noise distortion threshold does not directly relate to low accuracy.

It is obvious that too few learning examples do not provide enough information about a concept to be learned. On the other hand, a test on only a few examples is even more difficult. Recognition of one or two positive test examples frequently results in dozens of negative test examples not being excluded. This is the major reason that affects the accuracy of the primary

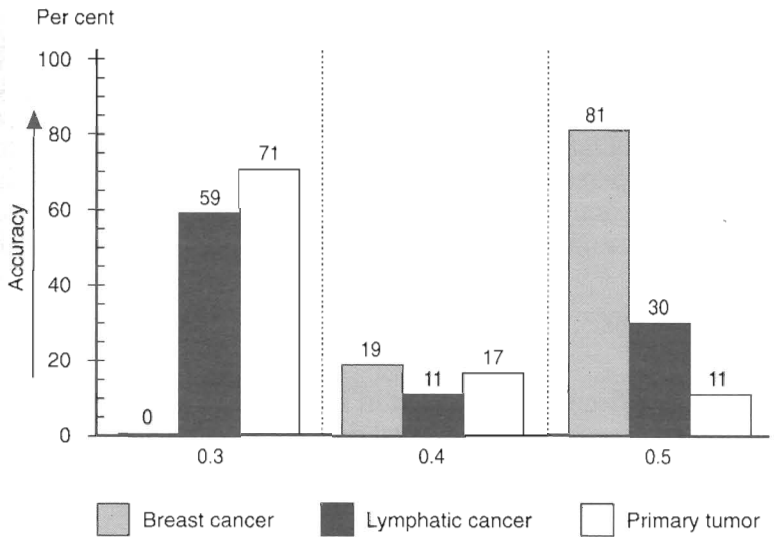

Figure 5.
Histograms for a noise distortion threshold

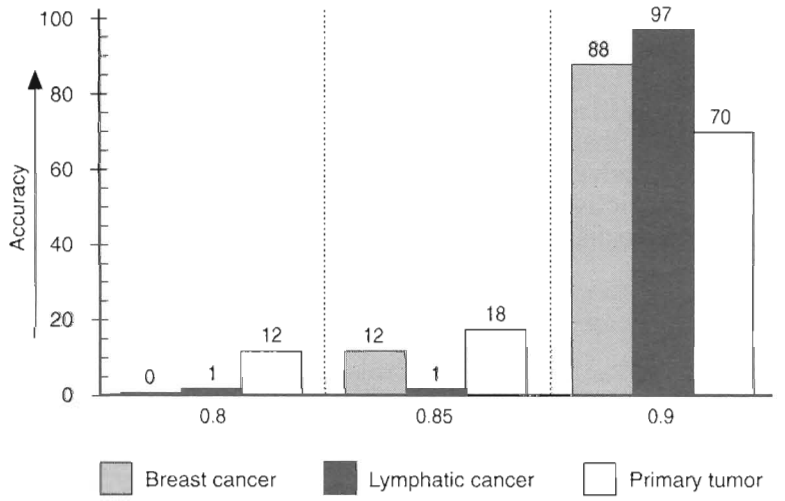

Figure 6.
Histograms for a noise truncation threshold

tumour data. We wish to point out that all this information is revealed if we use the measures of specificity and sensitivity. In our other work[24] we applied CLILP2 algorithm to scintigraphic data and generated a small number of rules used for diagnosing patients with coronary artery diseases, which on test data yielded high recognition rates in terms of both specificity and sensitivity. We shall see that the more popular diagnostic test, used in machine learning, introduced in the next section "hides" a lot of information concerning the correctness of the learned concepts.

Diagnostic test

This test uses a similarity match of a test example with the generated decision rules. Because it is widely used for presenting machine learning research results we shall compare the results obtained in this section with those reported by Michalski[7] and Clark and Niblett[8]. Rather than testing how accurate a rule is, this test compares how close a test example is to a concept description. To diagnose a test example, the decision rule that matches it best determines the example's class membership. Diagnostic accuracy of a class is defined as:

$$\text{diagnostic accuracy} = \frac{\text{true positive}}{\text{all positive}}$$

Applying the same decision rules used in Part I to the test examples for the three domains, each example is classified differently than when using the rule verification test where an example was recognized only when it satisfied the conditions of a rule with a truth value (the sum of weights) higher than a given threshold. No thresholds are used in this section and an example is applied to the rules of all classes. In an extreme case, if the example, for instance, is 30 per cent true of class 1, 10 per cent true of class 2, and 20 per cent true of class 3, then it is diagnosed as belonging to class 1 although the highest percentage is much below the thresholds used in the rule verification test.

The test results on lymphatic cancer, breast cancer and primary turnout are given in Tables VII, VIII and IX respectively.

In order to explain the reason for the low accuracy for the primary tumour data and to show the advantage of using the rule verification tests of sensitivity, specificity and predictive accuracy, we shall first discuss the relationship between them and the diagnostic test on a hypothetical set of test examples.

Table VII.
Test results for lymphatic cancer data

Class	Test examples	True positive	Diagnostic accuracy (per cent)
Fibrosation	1	0.3125	31.3
Lymphoma	18	14.8125	82.3
Metastases	24	21.5625	89.8
Normal	1	0.5625	56.35
Average	11	9.3125	84.7

Table VIII.
Test results for breast cancer data

Class	Test examples	True positive	Diagnostic accuracy (per cent)
Recurrence	25	12.6825	50.8
No recurrence	60	51.5625	85.9
Average	42.5	32.1250	75.6

Let us assume that diagnostic results are given in Table X. Each test example is tested by the rules of all three classes and the numbers in Table X are truth values.

The rule verification test evaluates the concept description of each class, i.e. the test is conducted on each individual column. If the truth value of an example

Class	Test examples	Recognized	Diagnostic accuracy (per cent)
Bladder	1	0.1250	12.5
Breast	7	4.6250	66.1
Cervical uteri	1	0.0000	0.0
Colon	1	0.0000	0.0
Corpus uteri	2	0.0000	0.0
Duodenum intestine	0	0.0000	–
Oesophagus	3	0.1875	6.3
Gall-bladder	5	0.6875	1.4
Head/neck	6	4.9375	82.3
Kidney	7	1.5000	21.4
Liver	2	0.0000	0.0
Lung	25	10.1875	40.8
Ovary	9	6.3125	70.1
Pancreas	8	1.4375	18.0
Prostate	3	0.2500	8.3
Rectum	2	0.5000	25.0
Salivary glands	1	0.0625	6.3
Stomach	12	1.0625	8.9
Testis	0	0.0000	–
Thyroid	4	0.0625	1.6
Vagina	0	0.0000	–
Anus	0	0.0000	–
Average	4.6364	1.4517	31.3

Table IX.
Test results for primary tumour location data

	Diagnostic results		
	Class 1	Class 2	Class 3
Test examples of class 1	0.91	0.37	0.97
	0.92	0.42	0.97
	0.94	0.14	0.99
	0.93	0.31	0.98
Test examples of class 2	0.21	0.95	0.98
	0.13	0.93	0.97
	0.34	0.90	0.98
	0.53	0.96	0.99
Test examples of class 3	0.36	0.17	0.93
	0.28	0.32	0.98

Table X.
Results on ten hypothetical test examples

is higher than a threshold, say 0.90, then the test example is recognized as belonging to a class corresponding to the column. For the first two classes, the sensitivity, specificity and predictive accuracy are all 100 per cent.

For class 3, sensitivity is 100 per cent but specificity is 0 per cent and predictive accuracy is only 50 per cent. It can be concluded from the rule verification test that the rules generated from the first two classes are valid in recognizing test examples, but the rules for class 3 are not. In this test, if the rules for one class fail to describe a concept, the failure has no effect on the other classes.

Now, let us look at how a diagnostic test evaluates the performance on this hypothetical set of results. It evaluates the test results between different classes: i.e. the test is conducted in individual rows comparing the similarity of an example to a set of concept descriptions. In this case, if the rules of one class fail to describe a concept, the overall test results will be affected. From Table X, we notice that all examples of the first two classes are incorrectly diagnosed as belonging to class 3. The overall diagnostic accuracy is only 20 per cent (only two out of ten examples are correctly diagnosed). The incorrect concept description of class 3 prevents obtaining a better test result although the rules for recognizing the first two classes are available. If we knew that the concept descriptions of class 3 were "defective", and thus could have been discarded, the test examples of the first two classes would be recognized at the cost of not recognizing any of the examples of class 3. The diagnostic accuracy then increases to 80 per cent. This phenomenon was also observed and reported by Michalski[6,7]. If we use the rule verification test, however, we do not have to assume any such knowledge and will still be able to evaluate the results correctly.

In our data, especially for the lymphatic cancer and the primary tumour, there are many classes with a very limited number of positive training examples. The concept descriptions generated from those small classes may be unreliable and can be discarded. After discarding the concept descriptions generated from the classes containing less than ten positive training examples, the diagnostic test results for the lymphatic cancer data and primary tumour data are shown in Tables XI and XII respectively. Because the breast cancer data has only two classes and each class contains more than ten positive training examples, none of the generated concept descriptions are discarded.

Classes	Test examples	True positive	Diagnostic accuracy (per cent)
Fibrosation	1	0.0000	0.0
Lymphoma	18	14.9375	83.0
Metatases	24	21.8750	91.1
Normal	1	0.0000	0.0
Average	11	9.2031	83.7

Table XI.
Test results for lymphatic cancer data

Classes	Test experiments	Recognized	Diagnostic accuracy (per cent)
Bladder	1	0.0000	0.0
Breast	7	5.2500	75.0
Cervical uteri	1	0.0000	0.0
Colon	4	0.0625	1.6
Corpus uteri	2	0.0000	0.0
Duodenum/intestine	0	0.0000	–
Oesophagus	3	0.0000	0.0
Gall-bladder	5	1.0625	21.3
Head/neck	6	5.2500	87.5
Kidney	7	2.4375	34.8
Liver	2	0.0000	0.0
Lung	25	12.8125	51.3
Ovary	9	6.8750	76.4
Pancreas	8	2.1875	27.3
Prostate	3	0.0000	0.0
Rectum	2	0.0000	0.0
Salivary glands	1	0.0000	0.0
Stomach	12	1.5000	12.5
Testis	0	0-0000	0.0
Thyroid	4	0.2500	6.3
Vagina	0	0.0000	–
Anus	0	0.0000	-
Average test experiment of 22 classes is: 4.6364			
Average test experiments recognized: 1.7131			
Diagnostic accuracy for the domain: 36.9 per cent			

Table XII.
Test results for primary tumour location data

Next, we shall compare the results obtained by CLILP2 with those obtained by using other algorithms[7,8]. Unfortunately, the results shown in Table XIII are calculated in terms of diagnostic accuracy and complexity, widely used in machine learning, thus not revealing the true power of CLILP2 algorithm. As noticed by Clark and Niblett[8]: "cross-algorithm comparisons of the complexity of concept description are difficult due to the differences in representation and the degree of subjectivity involved in judging complexity". The comparison of complexity is conducted by counting the number of nodes in the decision-tree-based algorithms and the number of complexes generated by the rule-based algorithms.

Conclusions

The major advantage of CLILP2 is that it generates several hypotheses to describe a given concept. By testing unknown data with several different hypotheses of a concept we have a better chance of recognizing the test data

Algorithm	Lymphatic cancer Accuracy (per cent)	Lymphatic cancer Complexity	Breast cancer Accuracy (per cent)	Breast cancer Complexity	Primary tumour Accuracy (per cent)	Primary tumour Complexity
Assistant						
No pruning	76	38	67	120	41	188
Pruning	77	25	72	16	46	35
Bayes	83	–	65	–	39	
AQR	76	76	72	208	35	562
CN2						
90 per cent threshold	78	24	70	23	37	33
95 per cent threshold	81	22	70	20	36	42
99 per cent threshold	82	12	71	4	36	19
AQTT-15						
Complete	81	12	66	41	39	104
Unique > I	80	10	68	32	41	42
Top rule	82	4	68	2	29	22
CLILP2						
No discard	85	18	76	40	31	108
Discard	84	16	–	–	37	83

Table XIII. Diagnostic accuracy and complexity of knowledge structures acquired by diverse algorithms for the data of three medical domains

correctly. The power of CLILP2 comes from its two-step learning process: it partitions the training examples into almost every possible way and then selects the best subsets to cover all the positive training examples. The CLILP2 algorithm is very efficient because it does not require a large amount of memory to store the descriptions of subsets and it generates If-Then rules only for the complexes included in a cover.

CLILP2 generates multiple hypotheses using integer linear programming models. The first ILP model is used to induce multiple hypotheses, and the second is used to generate concept descriptions with the minimum number of feature tests. Different hypothesis and role generation preference criteria can be implemented easily by changing the weights of the objective functions. The multiple solutions of the second ILP model allow for convenient determination of the *key/auxiliary features* of a concept description.

While learning from noisy data, CLILP2 uses a noise-distortion threshold to eliminate redundant subsets so that an exponential partitioning of training data can be avoided. By varying the noise-distortion threshold, one may obtain different partitions of training data and observe the impact of noise. The CLILP2 algorithm also demonstrates that the apparent noise caused by poor selection of hypotheses can be eliminated by considering multiple hypotheses, thereby preserving consistency when possible. For all three experimental data

sets rules were generated to cover the training examples of all classes consistently .

The algorithm was evaluated using two different approaches: the rule verification test and the diagnostic test. We would like to advocate strongly the use of sensitivity, specificity and predictive accuracy, in testing machine learning results, over using a single test of diagnostic accuracy. The former three reveal more information about the correctness of obtained results.

CLILP2 uses two parameters which allow the user to tailor the algorithm for his or her specific needs. However, in many applications the use of CLILP2 default parameters is sufficient. Future research should provide means for determining these parameters by the algorithm itself.

References

1. Quinlan, J.R., "Learning effective classification procedures and their application to chess ending games", in Michalski, R.S., Carbonell, J.G. and Mitchell, T.M. (Eds), *Machine Learning: An Artificial Intelligence Approach*, Morgan Kaufmann, San Mateo, CA, 1983.
2. Cestnik, B., Konoenko, I. and Bratko, I., "ASSISTANT 86: a knowledge elicitation tool for sophisticated users", in Bratko, I. and Lavrac, N. (Eds), *Proceedings of the Second European Working Session on Learning*, Bled, May 1987.
3. Quinlan, J.R., "Probabilistic decision trees", in Kodratoff, Y. and Michalski, R.S. (Eds), *Machine Learning: An Artificial Intelligence Approach*, Vol. III, Morgan Kaufmann, San Mateo, CA, 1990, pp. 140-52.
4. Quinlan, J.R., "Generating production rules from decision trees", *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, Milan, 1987, Morgan Kaufmann, San Mateo, CA, pp. 304-7.
5. Michalski, R.S., "On the quasi-minimal solution of the general covering problem", *Proceedings of the 5th International Symposium on Information Processing (FCIP 69)*, Vol. A3 (switching circuits), Bled, 1969, pp. 125-8.
6. Michalski, R.S., Mozetic, I., Hong, J. and Lavrac, N., "The multipurpose incremental learning system AQ15 and its testing application to three medical domains", *Proceedings of the Fifth National Conference on Artificial Intelligence*, Philadelphia, 1986, Morgan Kaufmann, San Mateo, CA, pp. 1041-5.
7. Michalski, R.S., "Learning flexible concepts: fundamental ideas and a method based on two-tiered representation", in Kodratoff, Y. and Michalski, R.S. (Eds), *Machine Learning: An Artificial Intelligence Approach*, Vol. III, Morgan Kaufmann, San Mateo, CA, 1990, pp. 63-102.
8. Clark, P. and Niblett, T., "The CN2 algorithm", *Machine Learning*, Vol. 3, 1989, pp. 261-83.
9. Michalski, R.S. and Larson, J., "Selection of most representative training examples and incremental generation of VL1 hypotheses: the underlying methodology and the description of programs ESEL and AQ11", *Reports of the Department of Computer Science*, No. 867, University of Illinois, Urbana, IL, 1978.
10. Quinlan, J.R., "The effect of noise on concept learning", in Michalski, R.S., Carbonell, J.G. and Mitchell, T.M. (Eds), *Machine Learning: An Artificial Intelligence Approach*, Vol. II, Morgan Kaufmann, San Mateo, CA, 1986.
11. Quinlan, J.R., "Simplifying decision trees", *International Journal of Man-Machine Studies*, Vol. 27, 1987, pp. 221-334.
12. Niblett, T., "Constructing decision trees in noisy domains", *Proceedings of the Second European Working Session on Learning*, Sigma Press, Bled, 1987, pp. 67-78.

13. Cios, K.J. and Liu, N., "Machine learning in generation of a neural network architecture: a continuous ID3 approach", *IEEE Transactions on Neural Networks*, Vol. 3 No. 2, 1992, pp. 280-91.
14. Michalski, R.S., "Variable-valued logic: system VL1", *Proceedings of the 1974 International Symposium on Multiple-valued logic on Pattern Recognition*, West Virginia University, Morgantown, VA, 1974, pp. 323-46.
15. Bar Yehuda, R. and Even, S., "A linear-time approximation algorithm for the weighted vertex cover problem", *Journal of Algorithms*, Vol. 2, 1981, pp. 198-203.
16. Cios, K.J. and Torkzadeh, G., "Fuzzy clustering and neural networks for instrument validation", *Heuristics, the Journal of Knowledge Engineering*, Vol. 5 No. 1, 1992, pp. 90-6.
17. Grafinked, R.S. and Nembauser, G.L., *Integer Programming*, John Wiley & Sons, New York, NY, 1972.
18. Bellmore, M. and Ratliff, H.D., "Set covering and involutory bases", *Management Science*, Vol. 18 No. 3, November 1971.
19. Chvatal, V., "A greedy-heuristic for the set-covering problem", *Mathematical Operations Research*, Vol. 4 No. 3, 1979.
20. Hochbaum, D.S., "Approximation algorithm for the weighted set-covering and vertex cover problems", *Siam Journal of Computing*, Vol. 11 No. 3, August 1982.
21. Aspvall, B. and Stone, R.E., "Khachiyan's linear programming algorithm", *Journal of Algorithms*, Vol. 1 No. 1, 1980, pp. 1-13.
22. Kononenko, I., Bratko, L. and Roskar, E., "Experiments in automatic learning of medical' diagnostic rules", Technical Report, E. Kardelj University, Faculty of Electrical Engineering, Ljubljana, Yugoslavia, 1984.
23. Cios, K.J., Shin, I. and Goodenday, L.S., "Using fuzzy sets to diagnose coronary artery stenosis", *IEEE Computer Magazine*, Special Issue on Biomedical Engineering, Vol. 24 No. 3, 1991, pp. 57-63.
24. Cios, K.J., Liu, N. and Goodenday, L.S., "Generation of diagnostic rules using inductive machine learning", *Kybernetes*, Vol. 22 No. 5, 1993, pp. 44-56.