Classification Learning Using All Rules

Murlikrishna Viswanathan and Geoffrey I. Webb

School of Computing and Mathematics
Deakin University, Geelong, Vic, Australia, 3217

Abstract. The covering algorithm has been ubiquitous in the induction of classification rules. This approach to machine learning uses heuristic search that seeks to find a minimum number of rules that adequately explain the data. However, recent research has provided evidence that learning redundant classifiers can increase predictive accuracy. Learning all possible classifiers seems to be a plausible ultimate form of this notion of redundant classifiers. This paper presents an algorithm that in effect learns all classifiers. Preliminary investigation by Webb (1996b) suggested that a heuristic covering algorithm in general learns classification rules with higher predictive accuracy than those learned by this new approach. In this paper we present an extensive empirical comparison between the learning-all-rules algorithm and three varied established approaches to inductive learning, namely, a covering algorithm, an instance-based learner and a decision tree learner. Empirical evaluation provides strong evidence in support of learning-all-rules as a plausible approach to inductive learning.

1 Introduction

The heuristic covering algorithm (as typified by Michalski, 1984; Clark and Niblett, 1989; Muggleton and Feng, 1990; and Quinlan, 1990) has been the predominant approach to learning classification rules. A basic characteristic of inductive learning is the use of search. In this context machine learning is often regarded as a search for generalizations and specializations of concepts. The covering algorithm seeks to develop a minimal set of rules that adequately explains the training data.

In contrast, recent research (Ali, Brunk, and Pazzani, 1994; Breiman, 1996; Dietterich and Bakiri, 1994; Domingos, 1995; Kwok and Carter, 1990; Nock and Olivier, 1995; Oliver and Hand, 1995; Schapire, 1990; Webb, 1996a; Wogulis and Langley, 1989) has provided increasing evidence in support of learning redundant classifiers. While most of this research has occurred in the context of learning decision trees rather than the classification rules with which the current research is concerned, there is no reason to believe that the results do not generalise to this latter context. Webb (1996b) presented a system that in effect infers and employs all possible classification rules, and after preliminary investigation concluded that a heuristic covering algorithm in general provided higher predictive accuracy. In this paper we present results of extensive empirical comparison of the learning-all-rules approach against a heuristic covering algorithm, a benchmark instance-based learner and a benchmark decision tree based learning algorithm. We find that the learning-all-rules approach in general gives superior performance over the traditional covering algorithm and has equivalent performance levels to the decision tree and instance based learners.

2 The Covering Approach

The covering technique for the induction of classifiers from examples has been a popular generic approach among machine learning systems that infer classification rules. The covering strategy forms a set of rules by inferring the rules one at a time. At each step it searches for a rule that covers many positive examples and few or no negative examples. The covered examples are then removed and the algorithm starts again on the remainder. This heuristic approach to learning a set of rules seeks to infer a minimal number of rules.

Although the covering algorithm has been a commonly used technique, it is subject to several well known weaknesses. The foremost among these is the application of hill climbing search. This search technique has well known limitations, including the problems of local maxima, plateaus and ridges. These can prevent it from reaching an optimal solution. Another problem with heuristic search is that it is often difficult to determine whether the search technique has introduced additional implicit biases that evade proper identification. Such implicit biases could have a profound effect on experimental results. There is also a growing body of evidence that learning a minimal set of rules is in general sub-optimal with respect to predictive accuracy. Empirical evidence from recent research (Ali, Brunk, and Pazzani, 1994; Breiman, 1996; Dietterich and Bakiri, 1994; Domingos, 1995; Kwok and Carter, 1990; Nock and Olivier, 1995; Oliver and Hand, 1995; Schapire, 1990; Webb, 1996a; Wogulis and Langley, 1989) has shown that learning classifiers that contain elements in addition to the bare minimum needed, can improve predictive accuracy.

3 Induction of All Rules

Webb's (1996b) new technique learning-all-rules, in effect learns all possible rules defined by the rule-description language. The number of possible rules for a given domain or learning task may be infinite. As a result it is infeasible to develop explicit representations for all possible rules. However, explicit representations of all rules are not necessary in order to apply all possible rules. Webb (1996b) adopts an approach whereby explicit rules are only developed when an example is classified by the algorithm. Then, only the single rule that determines the class to be assigned to the unclassified item is explicitly represented. Thus, in this technique the training set is retained until actually classifying an instance. When an example is classified those rules relevant to that instance will be inferred from the training set. This can be viewed as a lazy learning (Aha, 1997) approach to learning classification rules. However, learning-all-rules differs from most lazy learners that construct temporary classifiers (such as lazy decision trees, Friedman, Kohavi, and Yun, 1996) by performing complete search, instead of heuristic search, for the best classifier with respect to the target object.

3.1 The Learning-All-Rules Algorithm

Every inductive machine learning system necessarily embodies some search technique in its quest for hypotheses. OPUS (Webb, 1993; Webb, 1995) is a search algorithm that provides efficient complete search to select individual classification rules from a search space of all possible non-disjunctive rules. OPUS takes as input a training set **t**, an evaluation function **e**, and a set of specialisation operators **o**, and outputs a set of operators from **o** which generate a classifier that maximises **e** with respect to **t**.

The evaluation function **e** specifies the inductive bias. The max consistent and Laplace accuracy estimate metrics of empirical support were used in this research. The max consistent metric favours rules that cover the most positive examples and no negative examples while the Laplace metric allows for a trade-off between the coverage of positive and negative examples. Given that N is the number of negative training examples covered, P the number of positive examples and C the number of classes from the training data, the max consistent empirical support value for a rule equals -N, if $N > 0$, else P and the Laplace empirical support for a rule equals $(P+1)/(P+N+C)$.

The OPUS algorithm is not presented here, as the search algorithm employed is not significant to the research, so long as that algorithm performs complete search.

An abstract specification of the learning-all-rules algorithm is presented as follows. Let,

T : set of training examples.

instance : an unclassified object to be classified.

OPUS(T,instance, E) : function that takes as its inputs the training set T, an example to be classified *instance*, and an inductive bias function E and returns a rule that covers instance, and has maximal support for the inductive bias function E with respect to the training set T.

X → C : a classification rule.

```
$X \rightarrow C = OPUS(T, instance, E)$
assign class $C$ to $instance$
```

The OPUS algorithm performs complete search in contrast to heuristic search. Hence, the rules it generates are always optimal with respect to the preference function.

3.2 Advantages

The learning-all-rules approach embodies a number of major advantages in comparison to the traditional covering technique.

As opposed to the covering approach, learning and employing all rules involves no search heuristics. The representation language for expressing rules used by the system determines the set of rules generated. The exclusion of heuristics from the inference process eliminates the limitations of the heuristic search.

The learning-all-rules approach employs OPUS (Webb, 1995), an admissible search algorithm which guarantees to find the nominated target as opposed to heuristic search algorithms that cannot guarantee to find the designated targets. Regarding the introduction of implicit biases discussed in the previous section, admissible search assures that the search technique is not introducing confounding unidentified implicit biases into the experimental evaluation.

For any inductive learning strategy that seeks to develop a minimal set of rules there will be a number of possible sets of rules based on the data but only one of them may be selected. According to Webb (1995) this introduces an element of uncertainty which may affect the quality of the induced classifiers. This uncertainty is eliminated in the learning all rules approach which in effect learns and employs all rules.

Exemplar-based methods have been popular in data mining due to their strong approximation properties in determining the similarity between instances (Fayyad, Piatetsky-Shapiro, Smyth, and Uthurusamy, 1996). The learning-all-rules approach offers specific advantages to data mining. Primarily the use of admissible search to explore the space of all possible rules enables a wider exploration of the instance space, which can be valuable in some data mining contexts. Learning-all-rules also offers computational advantages when the number of cases to be classified by a single classifier is low. This is because only those portions of the search space pertaining to the cases to be classified need to be explored. This may be of particular value in applications where new training examples are continually becoming available, leading to frequent updating of the inferred classifier. One of the main limitations with instance-based methods is the need to define a priori distance metric to compute similarity between instances. Often the interdependencies between attributes and the diverse measurement units of attribute values make this a formidable task. The learning-all-rules algorithm uses the inferred rules to define similarity between instances, thus eliminating the need to define an a priori distance metric.

4 Evaluation with a Covering Algorithm

The covering algorithm used is a reimplementation of CN2 (Clark and Niblett, 1989) using unordered rules and the Laplacian error estimate evaluation function (Clark and Boswell, 1991). In order to minimise possible confounds in the experimental comparision, the covering algorithm uses OPUS to provide complete search in place of the heuristic search algorithm employed in the original

CN2 algorithm. If the original heuristic search were employed, it would not be possible to determine to what extent differences between the system could be attributed to the difference between complete and heuristic search, and to what extent they were due to differences between learning all rules and the use of a covering algorithm.

An abstract specification of the covering algorithm, COV, is presented as follows. Let,

T : set of training examples.

instance : an unclassified object to be classified.

OPUS(T,class,E) : function that takes as its inputs the training set *T*, a class *class*, and an inductive bias function *E* and returns a rule for the class that has maximal support for the inductive bias function *E* with respect to the training set *T*.

```
$ruleset = \emptyset$
For $class$ = each class in turn
    $examples$ = the training examples
    while $examples$ contains objects belonging to $class$
        $rule = OPUS(examples, class, E)$
        if no rule found
            remove from $examples$ all objects of class $class$
        otherwise
            remove from $examples$ all objects of class $class$ covered by $rule$
            add $rule$ to $ruleset$
```

Note that whereas in learning-all-rules the OPUS algorithm is used to search for a rule for any class that covers a specific case, in COV it is used to search for any rule of a specific class, in both cases seeking the rule that maximises the preference function.

5 Evaluation with Instance Based Learning

Instance-based learning (IBL) is the most widely employed form of lazy learning. Instance-based learning algorithms are an offspring of nearest neighbour (NN) algorithms and k-nearest neighbour algorithms (k-NN) (Fix and Hodges, 1952). As opposed to most other supervised learning methods, instance-based learners do not construct explicit abstractions such as decision trees or rules (Aha, Kibler, and Albert, 1991). In typical instance-based learning systems the training data (either in its entirety or a selected subset thereof) is retained for use in classification. A new example is classified by finding the nearest stored example from the training data based on some similarity function/metric, and assigning its class to the new example. Basically the performance of an instance-based learner depends critically on the metric used to compute the similarity between the instances (Domingos, 1995).

The similarities between the learning-all-rules approach and instance-based learning suggested an empirical evaluation of their relative performances. The learning-all-rules approach shares certain features with instance-based learners. First, in both approaches the entire training set or a subset is retained and referenced during classification. Webb (1996b) suggests that the learning-all-rules approach could be considered to be a form of qualitative instance-based learning whereby the selected rule is used to define a similarity metric for classification in place of the use of a distance metric.

This paper includes in the comparative evaluation with the learning-all-rules approach the IB1 instance-based learning algorithm implemented by Aha (1990). IB1 is an adaptation of the k-nearest neighbour algorithm that retains all the training instances for classification. Three successors to IB1—IB2, IB3, and IB4—were also evaluated, but in general provided worse results than IB1, and hence results are not presented herein. IB2 is an edited nearest neighbour algorithm that retains only the misclassified instances. Since the instance selection method stores noisy instances and uses them for classification this algorithm is susceptible to noise in the data. The IB3 algorithm is an adaptation of IB2 and is similar in retaining the misclassified instances but in addition keeps a classification record for each instance and removes some of the stored instances that are believed to be noisy using a significance test. Finally IB4 includes the complete functionality of IB3 in addition to an attribute weight learning capability.

6 Experimental Evaluation

As mentioned above, the primary objective of this paper is to present an extensive empirical comparison between the learning-all-rules and other learning methods. Webb's (1996b) preliminary empirical evaluation led him to conclude that the covering algorithm enjoyed a statistically significant general advantage (in terms of predictive accuracy) over learning-all-rules. However, there was a flaw in the statistic analysis underlying this conclusion. Webb used a Friedman rank test comparing the number of times each approach outperformed the other. Experiments were conducted using 100 runs over 16 different domains. The Friedman rank test analysis was performed over all resulting 1600 comparisons. However, the results for each domain are not independent of one another, and hence the analysis is invalid. Such an analysis could be applied to a single domain to determine whether there was a significant difference within that domain, but should only be applied to a single result for each domain (such as the mean accuracy across all 100 runs) if it is to be used to evaluate the significance of an general advantage across domains. On extending the initial evaluation by incorporating additional data domains and analyzing mean performance on each domain, it is found that the covering algorithm outperforms learning-all-rules for

9 domains while learning-all-rules outperforms the covering algorithm for 22 domains. This does not support the claim of a general advantage for the covering algorithm.

Webb's (1996b) results also suggested that the learning-all-rules algorithm performed better with the max-consistent metric while the covering algorithm performed optimally with Laplace metric. Therefore the learning-all-rules algorithm employing the max consistent metric along with the covering algorithm employing the Laplace metric; C4.5 (Quinlan, 1993), the decision tree learner and the instance-based learner were applied to a representative collection of 33 datasets from the UCI Machine Learning Repository (Merz and Murphy, 1997) that were considerably diverse in size, number and type of attributes and number of classes.

It would also have been interesting to compare learning-all-rules with some further lazy learning algorithms, but did not have access to implementations of these systems for this purpose. This remains an interesting subject for future research.

6.1 Discretization of Continuous-Valued Attributes

Due to the limitation of the current implementation of the OPUS search algorithm to searching for categorical attribute-value rules, all the data domains that contained continuous values for the attributes had to be discretized. The discretization system provided by Ting (1995) employs Fayyad and Irani's (1993) discretisation method that considers all possible cut-points (i.e., all values in the training set) and selects the cut-point that gives the highest information gain. The method is recursively applied to the subsets of the previous split until the stopping criterion is met. The stopping criterion is based on the minimum description length principle, MDLP (Rissanen, 1989).

Note that C4.5 was applied to the discretized version of the data even though it has the capacity to perform its own discretization. This was done in order to minimize the number of possible confounding factors in the comparison. After all, in theory, the learning-all-rules approach could be applied to continuous valued data. The only reason that it was not is our lack of access to a search algorithm capable of performing complete search through such data.

6.2 Description of Experiment

The final experiment included the following steps.

- Each data set containing continuous attributes was discretized.
- Each data set was randomly divided into training (80%) and evaluation (20%) sets 100 times and for each pair of training and evaluation sets so formed all learning methods which included the covering algorithm with the Laplace metric (COV), learning-all-rules (LAR) with the max consistent metric, IB1 and C4.5 were applied to the training set and the predictive accuracy of the resulting classifiers was evaluated on the evaluation set.

All systems were run with their default settings in all experiments.

6.3 Analysis of Results

The mean predictive accuracy achieved by each treatment on each domain is presented in Table 1. In order to analyze these raw figures, Table 2 outlines the win/loss ratio. As can be seen, learning-all-rules with the max consistent metric achieves a higher mean predictive accuracy than the covering algorithm in 22 out of 33 domains. Learning-all-rules also outperforms the instance-based learner. Learning-all-rules achieves higher predictive accuracy than IB1 in 19 of the 33 domains. In comparison to the pruned version of C4.5, learning-all-rules achieves higher predictive accuracy in 11 of the 33 domains.

A multiple comparisons test was performed in order to compare each combination of pairs of treatments. The test was used in evaluating the statistical significance of the observed differences between the different learning methods. This test indicates the learning methods whose rankings significantly differ and also the direction of the difference. In the Table 3, '>' indicates that the treatment for the row has obtained a higher rank (at the 0.05 level) more often than the treatment for the column. '<' indicates that the treatment for the row has obtained a lower rank significantly (at the 0.05 level) more often than the treatment for the column. Finally, '=' indicates no overall significant difference in ranking. In the light of this experiment the table suggested that:

- Learning all rules with the max consistent metric was ranked significantly higher than the covering algorithm. In contrast to Webb's (1996b) earlier erroneous conclusions, this supports the existence of a general advantage to learning-all-rules over the covering algorithm.
- There was no significant difference between the general rankings of learning-all-rules with the max consistent metric and C4.5 or IB1. This suggests that these algorithms perform at similar levels.

7 Conclusions

The empirical analysis conducted so far has clearly and significantly yielded evidence against Webb's (1996b) hypotheses on the superior performance of the covering algorithm. The comparison with a benchmark decision tree learner and an instance based learning system suggests that the learning-all-rules algorithm is a plausible alternative to state-of-the-art heuristic inductive learning systems. In domains where constant acquisition of novel training examples results in frequent updating of the inferred classifier, a classifier will only be used to classify a small number of cases each time. In such domains, the learning-all-rules approach, owing to its use of lazy learning, enjoys a computational advantage over conventional machine learning. Therefore in such a context if a suitable metric is not available to support reliable instance-based learning, learning-all-rules provides an attractive alternative.

Table 1. Mean Predictive Accuracy for each Treatment and Domain

DOMAIN	LAR	COV	C4.5-U	C4.5-P	IB1
Australian	84.79	79.88	82.63	85.63	81
AutoS	71.75	68.87	75.95	73.85	73.33
Cleveland	80.91	74.01	74.39	76.57	76.38
Credit screening	84.97	80.81	82.89	86.01	81.01
Diabetes	73.02	70.98	72.33	74.85	69.64
Echocardiogram	67.60	65.80	73.8	74.53	69.33
Glass	65.88	62.44	67.11	67.25	69.32
Heart	80.46	76.01	75.81	77.14	78.22
Hepatitis	80.61	82.77	77.90	81.80	81.47
Horse-colic	85.41	80.05	83.67	86.58	82.47
Hungarian	81.90	78.20	79.87	79.63	70.83
Hypothyroid	97.82	98.17	98.79	98.82	97.91
Ionosphere	91.15	89.14	88.84	89.25	86.53
Iris	93.93	90.53	93.26	93.20	93.23
Pima-diabetes	72.08	69.71	71.83	73.93	69.42
Satimage	79.01	74.83	79.29	80.98	85.54
Segment	90.91	91.17	94.42	94.08	95.11
Shuttle	99.64	99.75	99.77	99.74	99.76
Vehicle	63.38	60.25	68.23	68.86	68.48
Slovenian breast cancer	69.08	69.17	66.65	71.03	66.77
Winconscin breast cancer	95.49	91.67	93.58	94.58	95.62
House-votes-84	94.25	92.43	94.42	94.94	92.70
KR-vs-KP	96.90	97.79	99.29	99.35	95.67
Lymphography	80.93	76.56	74.99	77.13	79.26
Monk1	100	100	93.70	96.71	84.02
Monk2	81.14	80.74	61.33	64.34	87.87
Monk3	97.37	97.73	98.61	98.89	82.75
MP11	98.62	98.61	87.05	86.45	85.17
Mushroom	100	100	100	100	100
Promoters	60.59	68.68	76.77	76.45	77.27
Primary tumor	39.52	35.01	39.52	39.63	36.85
Soybean-large	59.53	76.41	79.58	79.45	76.48
tic-tac-toe	96.50	96.49	83.77	83.05	96.15

Acknowledgments

Thanks to Zijian Zheng for helpful comments and suggestions on previous drafts of this paper. We are grateful to Kai Ming Ting for providing the discretization system and to David Aha for providing the IB family of instance-based learning systems.

Table 2. General Performance Summary

Learning-all-rules Verses	Won	Lost	Tied
C4.5-Unpruned	18	13	2
C4.5-Pruned	11	21	1
Cover	22	9	2
IB1	19	13	1

Table 3. Multiple Comparisons Test

Method	LAR	COV	C4.5-U	C4.5-P	IB1
LAR	na	>	=	=	=
COV	<	na	<	<	<
C4.5-U	=	>	na	<	=
C4.5-P	=	>	>	na	>
IB1	=	>	=	<	na

References

Aha, D.W. (1990). *A Study of Instance-Based Algorithms for Supervised Learning Tasks.* PhD Thesis, Department of Information and Computer Science, University of California, Irvine, Technical Report 90-42.

Aha, D. W. (1997). Editorial on Lazy Learning. *Artificial Intelligence Review*, **11**: 7-10.

Aha, D. W., Kibler, D., and Albert, M. (1991). Instance-based learning algorithms. *Machine Learning*, **6**: 37-66.

Ali, K., Brunk, C., and Pazzani, M. (1994). On learning multiple descriptions of a concept. In *Proceedings of Tools with Artificial Intelligence.* New Orleans, LA.

Breiman, L. (1996) Bagging predictors. *Machine Learning*, **24**: 123-140.

Clark, Peter and Niblett, T. (1989). The CN2 induction algorithm. *Machine Learning*, **3**: 261-284.

Clark, P. and Boswell, R. (1991). Rule induction with CN2: Some recent improvements. In *Proceedings of the Fifth European Working Session on Learning*, pp. 151-163.

Dietterich, T. G. and Bakiri, G. (1994). Solving multiclass learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research*, **2**: 263-286.

Domingos, P. (1995). Rule induction and instance-based learning: A unified approach. In *Proceedings of the 13th International Joint COnference on Artificial Intelligence*, Montreal, Morgan Kaufmann, pp. 226-1232.

Fix, E. and J.L. Hodges (1952). *Discriminatory analysis - Nonparametric discrimination: Consistency properties.* From Project 21-49-004, Report Number 4, USAF School of Aviation Medicine, Randolph Field, Texas, pp. 261-279.

Fayyad, U.M. and Irani, K.B. (1993). Multi-interval discretization of continuous-valued attributes for classification learning. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, pp. 1022-1027, Morgan Kaufmann publishers.

Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R. (1996). *Advances in knowledge discovery and data mining.* MIT Press, Menlo Park, Ca.

Friedman, J. H., Kohavi, R., and Yun, Y. (1996). Lazy decision trees. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence.* AAAI Press, Portland, OR, pp. 717-724.

Kwok, S. W. and Carter, C. (1990). Multiple decision trees. In Shachter, R. D. and Levitt, T. S. and Kanal, L. N. and Lemmer, J. F. (Eds.) *Uncertainty in Artificial Intelligence 4.* North Holland, Amsterdam, pp. 327-335.

Michalski, R. S. (1984) A theory and methodology of inductive learning. In Michalski, R. S. and Carbonell, J. G. and Mitchell, T. M. (Eds.) *Machine Learning: An Artificial Intelligence Approach.* Springer-Verlag, Berlin, pp. 83-129.

Merz, C.J., and Murphy, P.M. (1997). UCI Repository of machine learning databases [http://www.ics.uci.edu/ mlearn/MLRepository.html]. Irvine, CA: University of California, Department of Information and Computer Science.

Muggleton, Stephen and Feng, C. (1990). Efficient induction of logic programs. In *Proceedings of the First Conference on Algorithmic Learning Theory*, Tokyo.

Nock, R. and Olivier G. (1995). On learning decision committees. In *Proceedings of the Twelfth International Conference on Machine Learning*, pp. 413-420, Taho City, Ca. Morgan Kaufmann publishers.

Oliver, J. J. and Hand, D. J. (1995). On pruning and averaging decision trees. In *Proceedings of the Twelfth International Conference on Machine Learning*, Morgan Kaufmann, Taho City, Ca., pp. 430-437.

Quinlan, J.R. (1990) Learning logical definitions from relations. *Machine Learning*, **5**: 239-266.

Quinlan, J.R. (1993). *C4.5: Programs for Machine Learning.* Morgan Kaufmann, San Mateo, CA.

Rissanen, J. (1989). *Stochastic Complexity in Statistical Inquiry.* World Scientific, Singapore.

Schapire, R. E. (1990). The strength of weak learnability. *Machine Learning*, **5**: 197-227.

Ting K. M., (1995). *Common Issues in Instance-based and Naive Bayesian Classifiers.* PhD thesis, Basser Dept of Computer Science, University of Sydney.

Webb, G. I. (1993). Systematic search for categorical attribute-value data-driven machine learning. In *AI'93 - Proceedings of the Sixth Australian Joint Conference on Artificial Intelligence*, World Scientific, Melbourne, pp. 342-347.

Webb, G. I.(1995). OPUS: An efficient admissible algorithm for unordered search. *Journal of Artificial Intelligence Research*, **3**: 431 -465.

Webb, G. I. (1996a). Further experimental evidence against the utility of Occam's razor. *Journal of Artificial Intelligence Research*, **4**: 397-417.

Webb, G. I. (1996b). A heuristic covering algorithm has higher predictive accuracy than learning all rules. In *Proceedings of Information, Statistics and Induction in Science*, Melbourne, pp. 20-30.

Wogulis, J. and Langley, P. (1989). Improving efficiency by learning intermediate concepts. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, Morgan Kaufmann, San Mateo, CA, pp. 657-662.