Prediction of Chemical Carcinogenicity in Rodents By Machine Learning of Decision Trees and Rule Sets

Dennis Bahler
Department of Computer Science
North Carolina State University
Raleigh NC 27695-8206

Douglas W. Bristol
National Institute of Environmental Health Sciences
Research Triangle Park, NC 27709

Abstract

We constructed toxicity models in the form of decision trees and rule sets by machine learning, using as a training set recent results from rodent carcinogenicity bioassays conducted by the National Toxicology Program (NTP) on 226 test agents. We performed 10-way cross-validation on each of these models to approximate their expected error rates on unseen data. We are using the models to offer prospective predictions of the carcinogenicity of the thirty test agents in the second phase of the NIEHS Predictive Toxicology Evaluation experiment (PTE-2) now underway.

Introduction

Determining which chemicals in the environment are carcinogenic to humans is of clear public benefit. The rodent bioassay program currently being performed by the U. S. National Toxicology Program (NTP) [Huff & Haseman 91] to determine which chemicals cause cancer in rodents, while clearly important to human risk assessment and public policy [Rodericks 94], is time-consuming and expensive.

It is our hypothesis that inductive analysis of biological information by a variety of machine learning techniques will discover patterns, co-occurrences, and correlations that have not come to light using other techniques. Further, we believe that such inductive analysis could lead to development of information management systems that are useful, first, for assisting researchers in the task of predicting the presence or absence of carcinogenic effects of chemical compounds, and, second, in developing mechanistic hypotheses that explain such effects. We are engaged in a long-term project to test our hypothesis. Aside from testing machine predictions against the results of ongoing empirical studies, a major goal of the project is to provide predictions that can help guide the selection of chemicals for testing. In the longer term, this approach may also help reduce the use of laboratory animals in such testing.

Methods

The Training Data

We gathered a set of 904 rodent bioassay experiments conducted on 226 test agents by the NTP and reported in NTP Technical Reports 201-458. Of these experiments, 160 were classified as Equivocal Evidence and were eliminated. Of the remaining 744 experiments, 276 were classified as No Evidence (which we refer to as Negative) and 468 were classified as either Clear Evidence or Some Evidence. These latter two classifications were combined into a Positive class for purposes of training our models.

The information available on the experiments consisted of a set of 261 attributes. (Six attributes in the dataset were used for bookkeeping and ignored during learning.) The attributes fell into several categories:

Salmonella Mutagenesis Test Result (1 attribute) The Ames test for mutagenesis has been performed on *Salmonella typhimurium* bacteria on most of the test agents in the training set.

Structural Alerts (18 attributes) Human expertise has identified certain functional-group substructures of organic molecules that may predispose the parent molecule towards causing chemical mutagenesis and carcinogenesis, because they represent the potential for either entering into electrophilic reaction with DNA or being converted by metabolism into an electrophilic functionality that can react with DNA [Ashby and Paton 93]

Physical Chemical Parameters (20 attributes) These are properties of a test agent, and can be determined either computationally or experimentally. The physical chemical parameters used were: electronegativity (Ke) rate (computed by two methods); octanol-water partition coefficient (logp, computed by two methods); highest occupied and lowest unoccupied molecular orbits; molecular weight; Pka; rectangular area (RA); planar area (PA); Z-depth (D); square of the RA-to-D ratio; square of the PA-to-D ratio; molecular hardness (computed by two

methods, plus an indicator attribute if the two methods produced different results); molecular volume; molecular area; molecular ovality; and dipole moment.

Subchronic Histopathology (209 attributes) These were represented as pairs of organ site and morphology. A total of 38 organs exhibited pathologic change in at least one experiment, and 72 different morphologies were observed at least once.

Sex and Species Exposed (2 attributes)

Route of Administration (1 attribute) Routes are feed, water, skin painting, inhalation, and gavage.

Maximally Tolerated Dose (4 attributes) Using molecular weight and standard conversion formulas, doses were normalized to micromoles per kilogram per day.

Building the Models Via Machine Learning

Tree Models Decision tree models are constructed by computer using a greedy, divide-and-conquer algorithm, which at each step has the goal of selecting from a set of attributes the one whose values best discriminate a set of examples according to the classification.

Rule Models Decision tree structures can be large, difficult for humans to understand, and can contain redundant subtrees which hide the underlying structure of information. Production rules can avoid these difficulties. Therefore, after the tree induction phase, a set of production rules was generated from the trees.

Cross-Validation of the Models The tree and rule models were each cross-validated using ten-way cross-validation. After generating tree models using a variety of attribute selection criteria and pruning methods, the highest cross-validated accuracy obtainable was found to be 90.7%. The best rule model generated had a cross-validated accuracy of 90.2%. After cross-validation, the tree and rule models described below were then built by using the entire training set of 744 nonequivocal experiments.

Results

The results of classification of the training examples by the overall cross-validated tree model are summarized in Table 1. The results of classification of the training examples by the overall cross-validated rule model are summarized in Table 2.

Discussion

The broad overall goal of predictive toxicology research is to develop methods that provide accurate predictions for as many chemicals as possible in the universe of structurally diverse, noncongeneric chemicals. Accordingly, the design of methods that can address this

		Predicted Bioassay		
		Positive	Negative	Total
Actual Bioassay	Positive	462	6	468
	Negative	66	210	276
	Total	528	216	744
Sensitivity:		0.99		
Specificity:		0.76		
+ Predictivity:		0.88		
− Predictivity:		0.97		
Accuracy:		0.90		
False +/False −:		11.00		
Mathews Corr. Coeff.:		0.80		
Yates-Corrected χ^2:		467.90 ($p < 0.001$)		

Table 1: Training Set Classification Accuracy (Tree Model)

		Predicted Bioassay		
		Positive	Negative	Total
Actual Bioassay	Positive	468	0	468
	Negative	104	172	276
	Total	572	172	744
Sensitivity:		1.00		
Specificity:		0.62		
+ Predictivity:		0.82		
− Predictivity:		1.00		
Accuracy:		0.86		
False +/False −:		infinite		
Mathews Corr. Coeff.:		0.71		
Yates-Corrected χ^2:		375.85 ($p < 0.001$)		

Table 2: Training Set Classification Accuracy (Rule Model)

noncongeneric prediction problem must ideally avoid *a priori* restrictions of the concept space used to describe the biological activity being predicted so as to maximize the extent to which the method covers the universe of chemicals.

Our general approach to developing predictive toxicology methods is designed to use inductive approaches for recognizing patterns and relationships as an effective way to address the noncongeneric chemical prediction problem. We believe induction is the most appropriate approach that can be applied to the development of noncongeneric prediction methods, because it enables the discovery of relationships in knowledge domains that lack formal models; i. e., induction requires no specific knowledge of the multiple biological processes or mechanism(s) that determine relationships between the noncongeneric universe of chemicals and a complex biological endpoint such as carcinogenesis.

As evidence of the power of inductive models, our models use a combination of microbial assay results, route and dose information, physical chemical parameters, alerting chemical substructures, and subchronic histopathology. In other words, our models can utilize any parameter, enabling it to exploit a learning set containing biological as well as chemical data. In terms of their ability to use whatever information may be appropriate to their task, our models resemble the human heuristic approach more than they do many other published computer systems in this domain. In addition, tree and rule models are readily understood by human experts in toxicology, since the rules use largely familiar terms and concepts. Moreover, our models can handle inorganic molecules as well as organic, noncongeneric organic molecules as well as congeneric, and mixtures.

Many of the rules that our models have generated have never been explicitly enunciated by human experts. We believe it is plausible that such novel patterns may serve to stimulate the formation of new mechanistic hypotheses and further research. Mechanisms of carcinogenesis almost surely involve a variety of factors, ranging from molecular structure, to metabolic factors, to the genotoxic effects of electrophilic chemicals on DNA, to the modulation of hormones or various receptors that regulate gene expression.

Taking the long view, purely inductive techniques are unlikely to be the entire solution to the larger problem of predictive toxicology. At the same time, however, it is unlikely that any single domain theory can be constructed that will constitute an adequate explanation for all the data on the universe of chemicals, let alone warrant confidence in its predictive accuracy. Little is known about causal mechanisms in toxicology, so the codification of domain knowledge is a hard problem, and such theories will necessarily be partial and uncertain. Detailed mechanisms that explain the causality of cancer remain largely a mystery even to expert researchers. We are convinced that a combination of rule induction and knowledge-based methods promise ultimately to make a significant contribution in this field.

Comparison with Our Previous Studies

To demonstrate the feasibility of using inductive machine learning methods for predictive toxicology, in 1993 we conducted a preliminary series of experiments in supervised tree and rule induction from a training set of 301 example chemicals [Ashby and Tennant 91, Ashby and Paton 93] whose carcinogenicity had been determined by long-term rodent studies [Huff and Haseman 91, Tennant 93] producing a system called TRIPT (Tree and Rule Induction for Predictive Toxicology). We then compared the resulting predictive accuracy with a set of published human and computer predictions for a common set of 44 test chemicals. A table giving the results of this prediction experiment, which has come to be known as PTE-1, is given in [Bahler and Bristol 93a]. The accuracy (concordance) of this system was comparable to the most accurate human expert prediction [Tennant *et al.* 90], and exceeded that of any of the computer-based predictions. Moreover, these trees and rules in part "rediscovered" existing human expert knowledge in this domain, and thus provided confirmation of the promise of this approach. Readers interested in more detail may consult [Bahler and Bristol 93a].

Beginning in 1993 [Bahler and Bristol 93b, Bristol and Bahler 93, Bristol, Tennant, and Bahler 93, Bristol 95, Bristol and Bahler 95a, Bristol and Bahler 95b] we analyzed the information content of the attributes employed by the original TRIPT learning system. Among other results, it was found that, of 189 attributes, the most informative 28 represented more than 99% of the information needed to discriminate the example chemicals into three classes: positive, negative, or equivocal. Moreover, in more than 400 sets of rules generated in that series of experiments under a wide variety of conditions, no more than 16 attributes survived the rule-pruning process to be included in any final set of rules. This leads us to believe that the information, and by extension the power to discriminate biologically active from inactive chemicals, is distributed quite unevenly among attributes about which information is available.

Much more recently, there has been an explosion of interest in the field of predictive toxicology among researchers in such diverse fields as artificial intelligence, biology, chemistry, toxicology, environmental science, pharmacology, and medicine [Lewis 94, Bristol *et al.* 96 (and the reminder of that volume), Srinivasan *et al.* 97, Lee *et al.* 98].

Acknowledgements

We, and all members of Dr. Bahler's group at NCSU, wish to thank Dr. Ann Richard of USEPA and her staff for compiling the physical chemical parameters in our

data sets. Thanks also to Dr. Ray Tennant and the staff at NIEHS for their support and encouragement. This work was partially supported by NIEHS/NIH and the Procter & Gamble Co. International Program on Animal Alternatives.

References

Ashby and Tennant 91 Ashby, J. and R. W. Tennant 1991. Definitive Relationships among chemical structure, carcinogenicity, and mutagenicity for 301 chemicals tested by the U.S. NTP. *Mutation Research* 257: 229-306.

Ashby and Paton 93 Ashby, J. and D. Paton 1993. The influence of chemical structure on the extent and sites of carcinogenesis for 522 rodent carcinogens and 55 different human carcinogen exposures. *Mutation Research* 286: 3-74.

Bahler and Bristol 93a Bahler, D. and D. W. Bristol 1993. The Induction of Rules for Predicting Chemical Carcinogenesis in Rodents. In L. Hunter, J. Shavlik, and D. Searls (eds.), *Intelligent Systems for Molecular Biology*, Cambridge, MA: AAAI/MIT Press.

Bahler and Bristol 93b Bahler, D. and D. W. Bristol 1993. A Quantitative Comparison of the Utility of Characteristics for Predicting Chemical Carcinogenesis. *4th Annual Keck Symposium on Computational Biology*, Pittsburgh.

Bristol 95 Bristol, D.W. 1995. Summary and recommendations for Session B: activity classification and structure-activity relationship modeling for human health risk assessment of toxic substances. *Toxicology Letters 79*, 265-280.

Bristol and Bahler 93 Bristol, D. W. and D. Bahler 1993. An Inductive Approach to Predicting Biological Activity of Noncongeneric Chemicals. Poster session, *Gordon Research Conference on Quantitative Structure-Activity Relationships*, Tilton, NH.

Bristol and Bahler 95a Bristol, D. W. and D. Bahler 1995. Inductive approaches to predicting toxicity. *Proc. 20th an.. Summer Toxicology Forum*, Givin Institute of Pathobiology, Aspen, CO.

Bristol and Bahler 95b Bristol, D. W. and D. Bahler 1995. Database Analysis to Identify Features, Provide Heuristic Information about Biological Factors, Manage Information, and Classify Chemicals. *Proc. 2nd European Conf. on High-Throughput Screening and Molecular Diversity*, Budapest.

Bristol, Tennant, and Bahler 93 Bristol, D. W., R.W. Tennant, and D. Bahler 1993. Predicting Chemical Carcinogenicity: Progress, Pitfalls, and Promise. *26th Ann. Symposium, Society of Toxicology of Canada.*

Bristol *et al.* 96 Bristol, D. W., J.T. Wachsman, and A. Greenwell 1996. "The NIEHS Predictive-Toxicology Evaluation Project," *Environmental Health Perspectives* **104** (Supplement 5), 1001-1010.

Huff and Haseman 91 Huff, J. and J. Haseman 1991. Long-term chemical carcinogenesis experiments for identifying potential human cancer hazards: Collective data base of the National Cancer Institute and National Toxicology Program (1976-1991). *Environmental Health Perspectives* 96: 23-31.

Lee *et al.* 98 Lee, Y., B. G. Buchanan, and J. M. Aronis 1998. "Knowledge-Based Learning in Exploratory Science: Learning Rules to Predict Rodent Carcinogenicity," *Machine Learning* **30**, 217-240.

Lewis 94 Lewis, David F.V. 1994. "Comparison between Rodent Carcinogenicity Test Results of 44 Chemicals and a Number of Predictive Systems," *Regulatory Toxicology and Pharmacology*, 215-222.

Rodericks 94 Rodericks, J.V. 1994. Risk Assessment, the Environment, and Public Health. *Environmental Health Perspectives 102*, 258-264.

Srinivasan *et al.* 97 A. Srinivasan, S.H. Muggleton, R.D. King, and M.J.E. Sternberg 1997. "The Predictive Toxicology Evaluation Challenge," *Proc. IJCAI-97*, 4-9.

Tennant 93 Tennant, R.W. 1993. Stratification of rodent carcinogenicity bioassay results to reflect relative human hazard. *Mutation Research* 286: 111-118.

Tennant *et al.* 90 Tennant, R.W., J. Spalding, S. Stasiewicz, and J. Ashby 1990. Prediction of the outcome of rodent carcinogenicity bioassays currently being conducted on 44 chemicals by the National Toxicology Program. *Mutagenesis* 5(1): 3-14.