Machine Learning Methods for International Conflict Databases: A Case Study in Predicting Mediation Outcome*

Johannes Fürnkranz, Johann Petrak and Robert Trappl
Austrian Research Institute for Artificial Intelligence
Schottengasse 3, A-1010 Wien
E-mail: {juffi,johann,robert}@ai.univie.ac.at

Jacob Bercovitch
Department of Political Science,
University of Canterbury, New Zealand
E-mail: POLS039@cantva.canterbury.ac.nz

Abstract

This paper tries to identify rules and factors that are predictive for the outcome of international conflict management attempts. We use C4.5, an advanced Machine Learning algorithm, for generating decision trees and prediction rules from cases in the CONFMAN database. The results show that simple patterns and rules are often not only more understandable, but also more reliable than complex rules. Simple decision trees are able to improve the chances of correctly predicting the outcome of a conflict management attempt. This suggests that mediation is more repetitive than conflicts per se, where such results have not been achieved so far.

OEFAI-TR-94-33

*This work was supported by a grant of the Austrian Federal Ministry of Science and Research

Contents

1 Introduction

Artificial Intelligence has lately been recognized as having some potential for supporting social scientists in political sciences (Hudson 1991), in particular in the construction and analysis of international conflict databases. Overviews of the potential contributions and ongoing projects using AI methods for the investigation of international relations can be found in (Mallery 1988; Schrodt 1991a; Trappl 1992b; Unseld 1994). One of the most promising approaches is to include Machine Learning methods to the set of tools used in analyzing existing conflict management databases (Schrodt 1991b; Schrodt 1991c; Mallery 1994).

The purpose of this paper is to introduce some of the possibilities that Machine Learning can offer to social scientists. We believe that Machine Learning is a valuable alternative to statistical methods that are commonly used to analyse international conflict databases. The main advantages of Machine Learning lie in its ability to automatically discover hypotheses, while the commonly used classical statistical methods are mainly designed for testing the validity of existing hypotheses. Furthermore the discovered concepts are very often quite complex and describe non-trivial regularities that would be hard to detect for a human analyst. For a very common class of problems — *classification problems* — Machine Learning offers algorithms that are able to automatically induce classification rules from a number of known and pre-classified instances. These rules can be applied to new instances, where the classification is not yet known. Most of these methods are equally applicable to numeric and symbolic data. The induced rules are easy to understand for domain experts. Finally, Machine Learning methods are already widely available. Most analyses described in this paper can be performed without significant computer background, in particular without any programming knowledge, because we have deliberately used C4.5, the best-known implemented Machine Learning tool currently available (Quinlan 1993).

This paper reports a case study where we have tried to analyze some aspects of an international conflict management database using Machine Learning methods. We will first give an overview of the facilities of state-of-the-art decision tree learning algorithms (section 2), followed by a short description of the CONFMAN data collection (section 3). Thereafter we will report on a series of experiments aimed at detecting useful knowledge for predicting the outcome of international conflict management attempts (section 4). We will analyze entire trees, single rules, and finally try to detect a subset of features in the database from which the most accurate trees can be constructed. After a discussion of the obtained results (section 5) we will summarize the most important aspects of this work in section 6.

2 Decision Tree Learning

Rule-Based Expert Systems — like the famous MYCIN program (Buchanan and Shortliffe 1984) — were the main contributing factor to the big commercial success of Artificial Intelligence at the beginning of the 80's. These systems typically incorporate a huge number of domain-specific rules with which an inference engine is able to solve a problem. However, it was soon discovered that obtaining the rule-based knowledge from human experts is the bottleneck of expert systems (Feigenbaum 1977). Research in Machine Learning therefore started to explore alternatives which aimed at automatically inducing the necessary knowledge from past cases. Decision tree learning algorithms are the most prominent result of this research. This section will give a short introduction to main concepts of this topic, enough to be able to follow the discussion in the remaining sections. More detailed descriptions of the algorithms can be found in the cited literature.

2.1 Decision Trees

A *decision tree* is a hierarchy of tests that can be performed on the data objects. Each node in the decision tree structure is associated with one test and each node has edges leading to successor nodes, one for each of the possible outcomes of this test. When a data item is tested at this node, the result will be exactly one of these outcomes. The object then follows down the edge that is labelled with its result of the test. Thus the data items that are tested at this node are divided into disjoint sets, each set consisting of objects with the same result for the performed test. Furthermore, each of these sets corresponds to one of the successor nodes and the data items in each set will be further examined with the test that is associated with the corresponding node. At the bottom of the hierarchy are designated nodes that have no test and no successors. These nodes are called the *leaves* of the tree.

Each data item is first submitted to the test at the top of the tree — the *root* of the tree. According to the outcome of the test it follows down the appropriate edge and the next test is performed. Eventually each object arrives at exactly one of the leaves. Thus a decision tree divides a set of examples into a number of disjoint subsets, one for each of the leaves. Each subset can be characterized by the conjunction of the test conditions that the objects have to fulfill in order to arrive at this leaf.

2.2 Automated Construction of Decision Trees

Decision tree learning algorithms like ID3 (Quinlan 1983) are able to construct decision trees from datasets, in which each object is described with exactly one value for each of a number of variables. One of these attributes is designated as the *class* of the object. A small hypothetical database taken from (Quinlan 1986) is shown in figure 1. It consists of 14 observations of whether a certain person likes to take a Saturday morning walk or not. Each observation is described by four variables that encode the weather outside and one class variable that indicates whether this person likes to take a walk or not.

No.	*Outlook*	*Temperature*	*Humidity*	*Windy*	*Walk?*
1	sunny	hot	high	false	no
2	sunny	hot	high	true	no
3	overcast	hot	high	false	yes
4	rain	mild	high	false	yes
5	rain	cool	normal	false	yes
6	rain	cool	normal	true	no
7	overcast	cool	normal	true	yes
8	sunny	mild	high	false	no
9	sunny	cool	normal	false	yes
10	rain	mild	normal	false	yes
11	sunny	mild	normal	true	yes
12	overcast	mild	high	true	yes
13	overcast	hot	normal	false	yes
14	rain	mild	high	true	no

Table 1: A small sample attribute-value database containing attributes describing the weather situation and a dependent variable that specifies whether the conditions are suitable for taking a walk or not.

From data like this a machine learning algorithm tries to construct a decision tree that is able to distinguish examples of one class from the examples of a different class. Tests in these decision trees usually correspond to checking the value of one attribute. Each leaf will be assigned a label that indicates the class of the instances that end up in this leaf. This class will be predicted for all examples that satisfy the conjunction of the tests from the root of the tree to this leaf.

Figure 2.2 depicts a tree that has uncovered five different clusters in the data of table 1: The person does not like to take walks when it is sunny and humid or when it rains and it is windy. On the other hand s/he does not mind walking when it is overcast, sunny and not too humid, or rainy, but not windy. The temperature has no influence on his decision. One can use this tree to predict for example what the person will do when the outlook is sunny, the temperature is cool, and it is humid but not windy, although no previous instance of this situation is known (s/he will not take a walk).

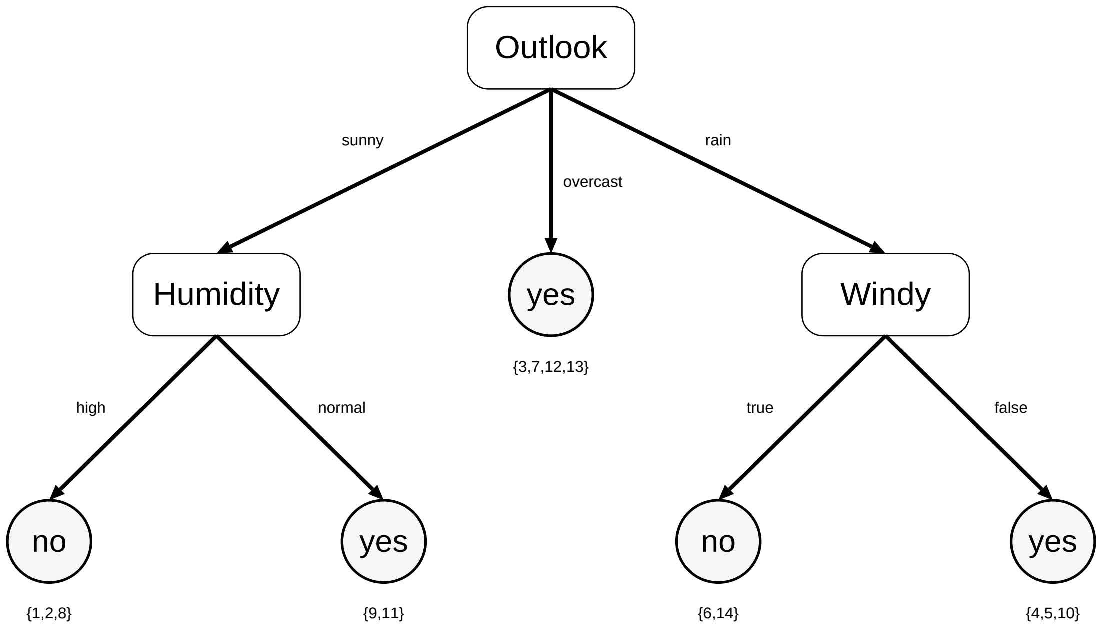

Figure 1: A decision tree that encodes a theory of the weather conditions that are suitable for taking a walk.

Of course, this is only one of several possible decision trees that are consistent with the data of table 1. A different tree might yield a different prediction for the above example. However, this tree is probably the simplest and most understandable, which makes it preferrable to its alternatives, all other things being equal. In order to keep explanations simple, ID3-like algorithms heuristically choose the attribute that yields the biggest *information gain* about the class value and test this attribute first (at the root). Subsequently the data are split into disjoint sets according to the possible values of this attribute and a subtree for each of these sets is learned. When one of the sets only contains objects of one class, no further splitting is performed and a leaf node is added to the tree. Thus the decision tree divides the original dataset into disjoint sets containing only objects of the same class. Each of these sets can be characterized by the conjunction of the conditions that are tested on the way from the root to the leaf that corresponds to this set. Each path from the root to a leaf in a decision tree can be transformed into a *rule* using the conjunction of all test outcomes that define this path as a condition and the class label of the leaf as a conclusion. The left-most leaf of figure 2.2 can thus be formulated as

> **If** *it is sunny* **and**
> *humidity is high*
> **then** *person X does not like to take walks.*

2.3 Learning Simpler Trees

Top-down induction of decision trees (TDIDT) as described in the last section will always manage to divide the datasets into disjoint sets as long as there are no two objects which have the same values for all the attributes, but not the same class values. This means that a unique class label can be assigned to each leaf, because all leaves will only contain instances that have the same class. This also means that the class label for each of the training instances can be correctly reproduced.

However, usually many of those leaves will only contain a small number of examples, most of them only one. The tests that have been chosen to discriminate these single examples from other examples are often chosen arbitrarily among a high number of candidate attributes that might discriminate this example from other examples. In particular near the leaves where only few examples are left at each node the chosen tests are often unreliable. Therefore it is preferrable to produce simpler trees which may contain leaves with examples from different classes. For these nodes no unique class label can be chosen. Programs then resort to assigning the majority class or using probabilistic classification methods at the tree leaves. Simpler trees are not only more understandable, but are very often also more accurate on unseen data.

For our analysis we have used the C4.5 program (Quinlan 1993) which is the direct successor of ID3. We have deliberately used an off the shelf (but state of the art) program, in order to illustrate what is possible without developing special-purpose algorithms. C4.5 has a variety of control parameters that can change the size and shape of the resulting decision tree. In particular the user can require

- a minimum number of examples in each leaf (`-m` option)
- that unreliable subtrees are replaced with leaves (tree *pruning*). The user can specify a significance parameter (`-c` option). Smaller values will result in smaller trees.
- that the outcomes of the tests at each node are not single values, but subsets of the set of all possible values (`-t` option).

In addition using only a small, but relevant subset of the available attributes will also produce smaller trees. C4.5 has also a variety of additional facilities that make it more suitable for handling real-world data sets, which often are imperfect. Among them are the ability to handle continuous data values and to deal with instances for which the values of some attributes are missing.

2.4 Estimating Tree Accuracy

We have already seen that — as long as there are no two examples with the same attribute values but different class values — a decision tree can be expanded until it only contains pure leaves. This means that reclassifying the examples used for learning with the resulting decision trees will always reproduce the correct class labels, i.e. the decision tree is 100% accurate in classifying its training examples. However, the same decision tree might well misclassify new examples that it has not seen during the learning phase. The percentage of new examples that a decision tree will classify correctly, its *predictive accuracy*, is often used as a criterion for evaluating the quality of a tree. A simple way for estimating this accuracy is to put aside a certain percentage of the available examples (the so-called *test set*, and test the decision tree that has been learned from the remaining examples (the *training set*) on this test set. However, this procedure has the disadvantage that not all of the available examples can be used for learning. Another approach — *bootstrapping* (Efron 1982) — learns a tree from the entire set of examples and tries to estimate its accuracy by performing a high number of experiments with randomly assigned training and test sets and averaging the results.

A computationally less expensive method is to perform a *cross-validation* (Stone 1974). A tree is learned from the entire training data, but its accuracy is estimated by splitting the available data into n subsets, learning a concept from $n-1$ of them and training on the n-th. This is repeated n times, such that each subset has been tested once. The average of these trials is a reasonable estimate for the accuracy of the original tree. The common practice is to choose $n = 10$, i.e. 10-fold cross-validation.

3 The CONFMAN Database

The *International Conflict Management (CONFMAN) Dataset* has been collected under the supervision of Jacob Bercovitch. Its primary focus is international mediation. Its aim is to both further our understanding of mediation, and facilitate the comparative investigation of different conflict management mechanisms.

Prompted by dissatisfaction with previous studies, which have rested on ideographic or normative approaches, this research project was established with the aim of furthering the much needed empirical investigation of conflict management within a sound theoretical framework. The project is founded on the contingency approach to the study of international conflict management which regards the outcome of management efforts as contingent upon a number of contextual and process variables. The contingency approach encourages systematic empirical

Attribute	*Description*
V1	Dispute Number
V2/V3	Duration (grouped/raw)
V4/V5	Fatalities (grouped/raw)
V6	Dispute Intensity
V7	System Period
V8	Geographic Region
V9–V11	Issue 1 – Issue 3
V12	Final Outcome
V13	Dispute Initiator
V14/V15	Identity Party A/B
V16/V17	Time in IS A/B
V18	Alignment
V19/V20//V21/V22	Power A/B (raw//grouped)
V23	Previous Relation
V24/V25	Political System A/B
V26/V27	Number of Parties A/B
V28/V29	Homogeneity A/B
V30/V31	Political rights A/B
V32/V33	Civil liberties A/B
V35	Conflict Management Type
V36	Third Party Identity
V37	Mediator Rank
V38	Mediation Strategies
V39	Previous Relationship
V40	Prev Attempts
V41	Prev Attempts this Mediation
V42/V43	Timing (grouped/raw)
V44	Initiated by
V45	Environment
V46	Outcome
V80	m intensity
V70	Power Disparity
V75/V76	Human Rights A/B
V77	Human Rights Disparity
V91	Political System Type
V81	Political System Difference
V92	Ally Numbers
V82	Ally Support Disparity
V93	Homogen Type
V83	Homogen Comparison
V94	Time in System
V84	Time in System Comparison
V90	Total Issues
V99	Mediation Outcome

Table 2: Attributes of the CONFMAN database

research because it recognizes variables and attributes with explicit operational criteria.

The project aims to answer such fundamental questions as "How do international mediation, and other forms of conflict management work?" and "Under what conditions are respective conflict management efforts most effective?". In answering these questions it is hoped the project will make a concrete contribution to the improvement of the international conflict management process.

A mediation attempt is defined as the formal or institutionalized non-violent and non-judicial intervention of an outsider or third party willing to help both disputants seek an acceptable outcome. An offer of mediation services is included in this understanding of an intervention. Other forms of conflict management that are encompassed are negotiation, arbitration/adjudication, multilateral conference, and referral to an international organization. The referral of a dispute to an international organization is coded as a separate event from any subsequent mediation or adjudication by that organization.

The central task of this research project has been the compilation of an extensive original dataset of international conflict management events since 1945. Primary information sources included Keesings Contemporary Archives (laterly Keesings Record of World Events), The Times Index, and The New York Times Index. Whenever necessary more detailed contemporary press reports or reputable historical accounts were also utilized.

The dataset that was used in the current study encompasses 921 international disputes and management attempts from 241 disputes since 1945. The attributes we used are listed in table 2. This database — or previous versions of it — has been analyzed extensively with statistical methods, most recently in (Bercovitch and Wells 1993; Bercovitch and Houston 1993; Bercovitch and Lamare 1993; Bercovitch and Langley 1993).

4 Analyzing Conflict Management Outcome

First we tried to learn decision trees for predicting the outcome of future conflict mediation attempts. The CONFMAN database (section 3) provided the learning examples. All entries where the outcome of the conflict management attempt is unknown were removed from the database. Furthermore we grouped the 5 different types of conflict management outcome into two classes: Mediation was *successful* when it resulted in a full or partial settlement of the conflict, or in a ceasefire. It was *unsuccessful* when a mediation attempt took place, but failed, or when mediation was only offered, but not accepted by the conflict parties. The resulting dataset consisted of 718 conflict management events, 408 (56.82%) of them resulting in failure and 310 (43.18%) being successful. Each event in this dataset was encoded with 52 attributes and one class variable that indicated whether the attempt has been successful or not.

Parameters	*Tree Size*	*Purity*	*Accuracy*
`C4.5 -m 1`	547	99.7%	60.3% (± 4.8)
`C4.5 -m 2`	314	91.8%	60.1% (± 3.3)
`C4.5 -m 5`	170	82.3%	60.4% (± 5.7)
`C4.5 -m 10`	90	76.6%	60.0% (± 5.2)
`C4.5 -m 15`	62	74.1%	61.6% (± 4.7)
`C4.5 -m 20`	47	71.9%	62.7% (± 2.0)
`C4.5 -m 25`	37	71.3%	63.0% (± 2.2)
`C4.5 -m 30`	26	70.1%	65.1% (± 2.5)
`C4.5 -m 35`	22	69.9%	65.0% (± 4.2)
`C4.5 -m 40`	20	69.2%	64.8% (± 2.6)
`C4.5 -m 50`	24	69.1%	64.5% (± 3.5)
`C4.5 -c 75`	524	99.7%	61.0% (± 4.5)
`C4.5 -c 50`	357	95.3%	60.2% (± 3.6)
`C4.5 -c 25`	257	91.2%	62.3% (± 4.4)
`C4.5 -c 15`	137	81.8%	64.8% (± 4.6)
`C4.5 -c 10`	75	76.9%	65.9% (± 4.9)
`C4.5 -c 5`	53	74.7%	63.8% (± 6.0)
`C4.5 -c 1`	27	70.2%	63.4% (± 5.8)
`C4.5 -m 2 -c 25`	173	86.2%	62.5% (± 5.2)
`C4.5 -m 30 -c 10`	20	69.6%	66.7% (± 3.7)
Mode Prediction	1	56.8%	56.8%

Table 3: Decision tree learning results on the CONFMAN database.

Table 3 gives an overview of some results we have achieved with different settings of C4.5's parameters (see section 2.3). For each setting we report the number of nodes (including leaves) in the generated tree (*Size*), the percentage of the training examples that will be classified correctly using the tree (*Purity*), and the predictive accuracy (*Accuracy*) estimated by a 10-fold cross-validation (see section 2.4) and its standard deviation.

The first series of experiments in table 3 investigated the effects of varying the `-m` parameter. This option allows the user to constrain the tree generation by allowing only tests that have at least two outcomes with more than the specified number of examples. In particular this means that nodes that contain less than the specified number of examples will automatically become leaves and no further tests are considered. A setting of `-m 1` will consider all splits and grow a tree that will contain pure leaves wherever possible (hence the high value on *Purity* for this setting). Increasing the value of this parameter will significantly decrease the size of the resulting decision tree, because less and less candidate tests are likely to pass the criterion.[1] This decrease in size comes with a decrease in purity, but also with an increase in accuracy. This confirms that the tests that are chosen near the leaves of the tree to discriminate small sets of examples from each other are often very unreliable. Removing them leads to an increase in accuracy. However, a too high increase will cause the performance to decrease again, because C4.5 is forced to discard some relevant tests along with the irrelevant ones. The limiting case would be to specify `-m 718`, which would yield a decision tree with only one node that classifies all examples by guessing the majority class (i.e. *mode prediction*).

The second block of table 3 reports results obtained by pruning the generated trees to various degrees. Contrary to the minimum number of examples criterion (`-m`, see above), pruning is a post-processing method that simplifies an existing tree[2] by replacing some of its internal nodes by leaves. In real-world domains the resulting trees are often not only smaller and more understandable, but also more accurate, because the unreliable tests near the leaves of the original trees (see section 2.3) have been discarded. Thus the aim of pruning is the same as using `-m`, but pruning is more flexible, because its parameter is independent from the actual number of examples used. We can observe the same pattern as with using `-m`: Simpler trees will have a smaller purity, but a better predictive accuracy. Too simple trees, however, will yield to decrease in accuracy, because potentially relevant information is thrown away.

Of course both of the above methods can be combined: A tree can be grown with a specified minimum number of examples for at least two outcomes of each test and the resulting tree can be pruned thereafter. We report two experiments where both methods were used. The first used the default settings of C4.5, which are `-m 2` and `-c 25`. The resulting tree is not so good. Adjusting the parameters to the domain apparently pays off, which is confirmed by (Kohavi and John 1994) where an automatic approach for finding the right values for C4.5's parameters has been investigated. In another experiment we combined the best settings of the above experiments. The 26 node tree resulting from `-m 30` was simplified using the best pruning parameter tested (`-c 10`). The result is a 20 node tree that was the best tree found with an estimated accuracy of 66.7%, which is almost 10% above the accuracy that can reasonably be expected by always guessing the majority class (56.8%).

We have also tried to use the `-s` parameter, that specifies that C4.5 is allowed to group the outcomes of a test instead of generating one new node for each possible

[1] As the `-m` parameter primarily affects the tree construction phase, we report the results of unpruned trees for these 11 experiments.

[2] The original trees in this series have been learned by using `-m 1`.

outcome. The resulting trees were a little smaller, but the predictive accuracy was usually a little worse.

4.1 The Unpruned Tree

The tree that has been generated with a setting of `-m 1` — the *unpruned* tree — is the most specific tree. It tries to group the management attempts into clusters of successful and unsuccessful events. The resulting tree contains 547 nodes and splits the data into more than 100 disjoint sets, most of them only containing one example. Of course, these one-element sets are not of interest, because the conditions that separate these events from other events might be arbitrary as discussed in section 2.3. This is why the predictive accuracy of this tree is relatively bad compared to simpler trees (table 3). On the other hand the accuracy is still better than the default accuracy that can be achieved by always guessing the majority class. Thus it is quite likely that the tree contains some useful information about the domain.

One way of extracting this information is to discard unreliable leaves near the leaves of the tree. This improves predictive accuracy as we have seen from table 3). Alternatively one could only use some branches of the tree. In particular, leaves that contain a relatively high number of examples for conflict management attempts that resulted in the same outome might yield some useful information about the nature of international conflict management. We have looked for clusters that contain 10 or more conflict management events, all of them having the same outcome (success or failure). The unpruned tree contains 12 rules that fulfill this criterion, 5 of them describing successful conflict management attempts (figure 2), the other 7 covering failures (figure 3).

Table 4 contains a summary of how many successful or unsuccessful conflict mediation events from how many different conflicts each of these rules describes. The rules starting with S denote clusters of successful conflict managements attempts, while the rules starting with F represent unsuccessful attempts. Together these 12 rules explain 185 events (25.77%) of the data set. Some of the rules are rather complicated, and it is unlikely that these regularities could have been detected by a human analyst. However, there are some simple rules testing only a few relevant conditions. For example **rule S1** states that

> **If** *there have been less than 400 fatalities* **and**
> *party B's raw power index is* $\leq$ *33* **and**
> *the conflict management type was mediation* **and**
> *the conflict lasted between 1 and 3 months*
> **then** *the conflict management was always successful*
> *in 15 mediation attempts in 8 different conflicts.*

```
Rule S1:
if (1 < V3_DUR <= 3) &&
   (V5_FAT <= 400) &&
   (V20_POWERB <= 33) &&
   (V35_MGMTACT == "MEDIATION")
then SUCCESS: 15  FAILURE: 0 (8 Conflicts)

Rule S2:
if (V2_DUR_G > 2) &&
   (V5_FAT <= 400) &&
   (V20_POWERB <= 33) &&
   (V35_MGMTACT == "MEDIATION") &&
   (V43_TIM <= 35) &&
   (V44_REQINI == "BOTH_PARTIES")
then SUCCESS: 15  FAILURE: 0 (11 Conflicts)

Rule S3:
if (400 < V5_FAT <= 700000) &&
   (V19_POWERA <= 31) &&
   (V20_POWERB <= 22) &&
   (V32_LIBA > 3) &&
   (V35_MGMTACT == "MEDIATION") &&
   (V38_MEDSTR == "DIRECTIVE") &&
   (V41_NRMEDM <= 4) &&
   (V43_TIM > 145) &&
   (V84_TISC == "DIFF_TIME_SYS")
then SUCCESS: 14  FAILURE: 0 (5 Conflicts)

Rule S4:
if (V3_DUR > 13) &&
   (400 < V5_FAT <= 700000) &&
   (V19_POWERA <= 31) &&
   (V20_POWERB <= 33) &&
   (V32_LIBA > 3) &&
   (V35_MGMTACT == "MEDIATION") &&
   (V38_MEDSTR == "PROCEDURAL") &&
   (V43_TIM <= 18) &&
   (V94_TIS <= 4)
then SUCCESS: 10  FAILURE: 0 (5 Conflicts)

Rule S5:
if (815 < V5_FAT <= 700000) &&
   (V19_POWERA <= 31) &&
   (V20_POWERB <= 10) &&
   (V30_RIGHTSA > 4) &&
   (V32_LIBA > 3) &&
   (V35_MGMTACT == "MEDIATION") &&
   (V38_MEDSTR == "DIRECTIVE") &&
   (V39_RELMED == "SAME_BLOC_BOTH") &&
   (V40_NRMED <= 5) &&
   (V41_NRMEDM <= 2) &&
   (V43_TIM <= 61)
then SUCCESS: 12  FAILURE: 0 (7 Conflicts)
```

Figure 2: Rules that cover 10 or more successful conflict management attempts

Rule	*Success*	*Failure*	*Conflicts*
S1	15	0	8
S2	15	0	11
S3	14	0	5
S4	10	0	5
S5	12	0	7
F1	0	12	2
F2	0	19	4
F3	0	14	3
F4	0	16	5
F5	0	14	8
F6	0	13	3
F7	0	31	3
Total	66	119	–

Table 4: Rules that cover only successful (S1–S5) or only unsuccessful (F1–F7) conflict management attempts

```
Rule F1:
if (400 < V5_FAT <= 700000) &&
   (V20_POWERB <= 33) &&
   (V32_LIBA <= 3) &&
   (V33_LIBB == 1) &&
   (V35_MGMTACT == "MEDIATION")
then SUCCESS: 0  FAILURE: 12 (2 Conflicts)

Rule F2:
if (400 < V5_FAT <= 700000) &&
   (V20_POWERB <= 33) &&
   (V24_POLSYSA == "MULTI-PARTY") &&
   (V32_LIBA <= 3) &&
   (V33_LIBB > 1) &&
   (V35_MGMTACT == "MEDIATION") &&
   (39 < V43_TIM <= 67)
then SUCCESS: 0  FAILURE: 19 (4 Conflicts)

Rule F3:
if (V3_DUR > 6) &&
   (400 < V5_FAT <= 700000) &&
   (V20_POWERB <= 33) &&
   (V24_POLSYSA == "MULTI-PARTY") &&
   (V32_LIBA <= 3) &&
   (V33_LIBB > 1) &&
   (V35_MGMTACT == "MEDIATION") &&
   (V40_NRMED > 3) &&
   (67 < V43_TIM <= 256)
then SUCCESS: 0  FAILURE: 14 (3 Conflicts)

Rule F4:
if (V3_DUR > 76) &&
   (400 < V5_FAT <= 700000) &&
   (V19_POWERA <= 31) &&
   (V20_POWERB <= 22) &&
   (V30_RIGHTSA > 4) &&
   (V32_LIBA > 3) &&
   (V35_MGMTACT == "MEDIATION") &&
   (V38_MEDSTR == "DIRECTIVE") &&
   (V41_NRMEDM <= 4) &&
   (61 < V43_TIM <= 136)
then SUCCESS: 0  FAILURE: 16 (5 Conflicts)

Rule F5:
if (V3_DUR <= 190) &&
   (400 < V5_FAT <= 700000) &&
   (V19_POWERA <= 31) &&
   (V20_POWERB <= 33) &&
   (V32_LIBA > 3) &&
   (V33_LIBB > 3) &&
   (V35_MGMTACT == "MEDIATION") &&
   (V38_MEDSTR == "COMM-FACIL") &&
   (V39_RELMED == "NO_PREV_REL") &&
   (V41_NRMEDM <= 2) &&
   (V93_HOMT == "MAJORITY")
then SUCCESS: 0  FAILURE: 14 (8 Conflicts)

Rule F6:
if (400 < V5_FAT <= 700000) &&
   (V19_POWERA > 31) &&
   (V20_POWERB <= 33) &&
   (V32_LIBA > 3) &&
   (V35_MGMTACT == "MEDIATION") &&
   (V41_NRMEDM <= 4) &&
   (V43_TIM <= 158)
then SUCCESS: 0  FAILURE: 13 (3 Conflicts)

Rule F7:
if (V5_FAT > 700000) &&
   (V43_TIM <= 73) &&
   (V84_TISC == "DIFF_TIME_SYS")
then SUCCESS: 0  FAILURE: 31 (3 Conflicts)
```

Figure 3: Rules that cover 10 or more successful conflict management attempts

On the other hand, **rule F1** shows us that

> **If** *there have been between 400 and 700,000 fatalities* **and**
> *party B's raw power index is* $\leq$ *33* **and**
> *both conflict parties have comparably high civil liberties* **and**
> *the conflict management type was mediation*
> **then** *the conflict management was never successful*
> *in 12 mediation attempts in 2 different conflicts.*

In general, the rules for explaining failed conflict management attempts are based on episodes from a smaller number of different conflicts. The reason for this is probably based on the fact that in conflicts with many mediation attempts usually the majority of them has failed.

4.2 Rules from Pruned Trees

One of the drawbacks of using unpruned trees to look for pure rules only as has been done in section 4.1 is that the program tries to discriminate between successful and unsuccessful mediation outcomes at every price. This approach can cause several problems:

- the resulting tree will have many clusters containing only one instance
- the rules describing the clusters will be very complex
- some conditions in the tree might be chosen arbitrarily among different candidates

We have already seen that simpler trees that avoid the above problems by replacing small, but pure clusters with large, but impure clusters at the leaves. As an illustration consider **rule F7** from the unpruned decision tree:

> **If** *there have been more than 700,000 fatalities* **and**
> *not more than 73 months of the conflict have elapsed* **and**
> *the parties have spent a different amount of time*
> *in the international system*
> **then** *the conflict management was not successful*
> *in 31 mediation attempts in 3 different conflicts.*

A closer examination of the underlying decision tree reveals that the second and third test in this rule have only been added to discriminate the 31 unsuccessful conflict management attempts from 4 other events, 2 of which has also been unsuccessful. By dropping these conditions we can get **rule P1** (P stands for Pruned):

If *there have been more than 700,000 fatalities*
then *the conflict management was successful in only 2*
and failed in 33 mediation attempts in 4 different conflicts.

This new rule P1 is much simpler than its predecessor F7. The two conditions that have been chosen in rule F7 to filter a cluster of 31 failed mediation attempts out of the total 35 mediation attempts with more than 700,000 fatalities were probably chosen arbitrarily among several others that would have equally well separated the two successful from the 33 failed attempts. Thus rule P1 is not necessarily less accurate than rule F7, because only presumably irrelevant conditions have been removed. In fact, almost all simplified decision trees of table 3 replace rule F7 and the 4 additional rules, each one of them covering only one example, with the compact rule P1.

Rule	*Success*	*Failure*	*Conflicts*
P1	2	33	4
P2	4	41	27
P3	25	18	29
P4	12	19	9
P5	44	101	37
P6	19	44	11
P7	123	53	58
P8	23	56	15
P9	56	40	9

Table 5: Rules from a pruned decision tree

Figure 4 shows the best decision tree from table 3. It consists of 9 rules, each of them covering both, successful and unsuccessful conflict management attempts. Table 5 contains a summary of the number of successes and failures that meet the conditions of each rule, and the number of different conflicts in which these mediation attempts occurred. Three rules (P3, P7, and P9) cover a majority of successful conflict management attempts. In particular rule P7 which describes 176 attempts, among them 123 (69.9%) successes, looks interesting:

If *there have been less than 700,000 fatalities* **and**
mediation is the chosen conflict management activity **and**
the civil liberties of party A are comparably restricted **and**
party A's power is comparably low **and**
the conflict lasts no longer than 8 years
then *the conflict management was successful*
in 69.9% of 176 mediation attempts in 58 different conflicts.

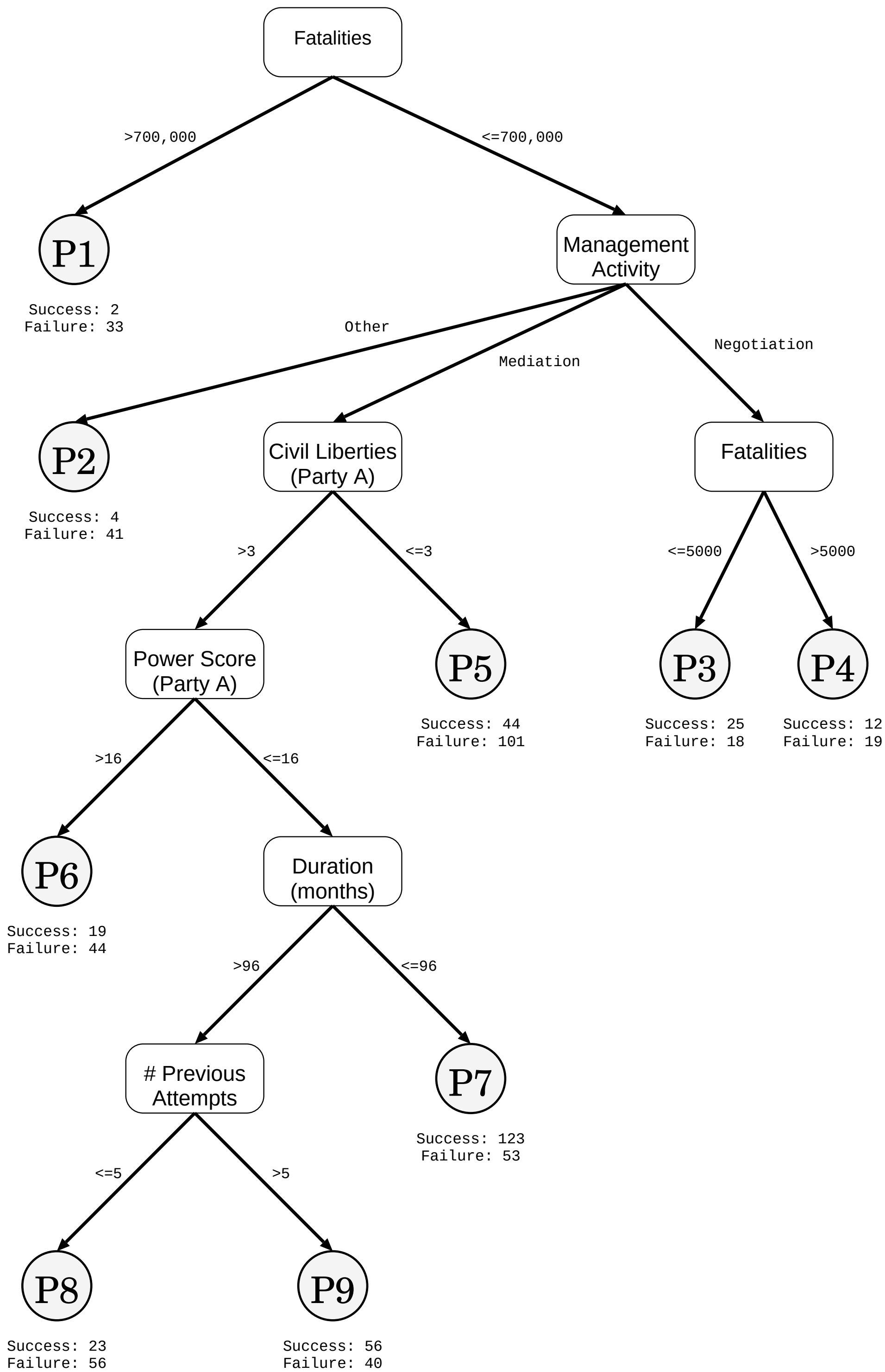

Figure 4: The best simplified decision tree

Rules P8 and P9 complement this rule with specifying that chances for success in conflicts that meet the above conditions, but last longer than 8 years, are higher when more conflict management attempts are tried. An interesting finding is that both rule F1 (see section 4.1) and rule P5 seem to indicate that high civil liberties are not compatible with mediation.

4.3 Finding Relevant Variables

Of all methods for obtaining simpler trees that we have discussed in section 2.3 one of them cannot be performed automatically within C4.5, namely the selection of relevant variables. If C4.5 is given only a relevant subset of the possible attributes it will not be able to include irrelevant distinctions of instances near the leaves of the tree. We performed an additional experiment aimed at determining a set of relevant variables from the many variables in the database.

(John, Kohavi, and Pfleger 1994) have developed a program that uses C4.5 and determines the set of attributes from which the best decision tree can be learned. It starts with an empty set of attributes and greedily adds the attribute that gives the highest increase in estimated predictive accuracy for the tree grown from the new set of attributes. Alternatively, the algorithm can also choose to delete an existing attribute from the current set of attributes. Predictive accuracy is estimated with consecutive 10-fold cross-validation experiments (with different random splits) until the standard deviation of the resulting estimate is below 1%. If no feature can be added or deleted without decreasing the estimated accuracy of the tree for two consecutive tries, the program stops with the current set of features. In order to avoid to be too short-sighted a one-time decrease is not sufficient for stopping the algorithm. In this case two features may be added at a time if this increases accuracy.

Table 6 reports the results from two experiments, one that was performed with the default settings for the parameters, and one with the `-s` option turned on (see section 2.3). We have tried a few different settings, in particular the best parameter choice from table 3, but in this case the default choices seemed to be very good, which indicates that only relevant attributes are used and that therefore too high settings of the `-m` parameter and too low settings of the `-c` parameter may force C4.5 to throw away relevant information. For each of the two experiments we report the purity of the final tree, the number of cross-validations needed to get the standard deviation below 1% and most importantly the estimated accuracy of the tree. The tables have to be read from the top to the bottom.

The final decision tree in both cases consisted of 8 variables and had an accuracy of above 67%. It is interesting that in the experiment where the `-s` parameter was activated the program at one point had 9 variables in the tree, but the

C4.5 -m 2 -c 25				
Choice	*Variable*	*Purity*	*# X-vals*	*Accuracy*
1	V39 Previous Relationship	64.2%	3	63.2%
2	V20 Raw Power Score B	68.1%	4	65.5%
3	V37 Mediator Rank	71.7%	4	65.9%
4	V27 Number of Parties B	74.1%	4	67.0%
5	V22 Grouped Power Score B	73.4%	4	67.3%
6	V06 Dispute Intensity	—	—	—
6	V45 Environment	70.3%	4	67.5%
8	V11 Issue 3	70.5%	4	67.7%

C4.5 -m 2 -c 25 -s				
Choice	*Variable*	*Purity*	*# X-vals*	*Accuracy*
1	V39 Previous Relationship	64.3%	2	63.3%
2	V19 Raw Power Score A	67.5%	2	65.4%
3	V27 Number of Parties B	70.3%	4	65.8%
4	V10 Issue 2	70.6%	4	66.7%
5	V70 Power Disparity	—	—	—
5	(V82 Ally Support Disparity)	71.3%	3	66.9%
7	V90 Total Issues	71.3%	3	67.1%
8	V21 Grouped Power Score A	—	—	—
8	V45 Environment	70.9%	4	67.1%
10	(V82 is deleted again)	70.9%	3	67.7%

Table 6: Relevant attributes detected by feature subset selection

feature concerning the disparity of the support of each party's allies could be deleted again at the end with a further increase of accuracy. This shows that the algorithm does not necessarily converge towards an optimal subset of features. It may for example be the case that adding a combination of certain attributes yields a better tree, while adding only one of them results in a worse tree.

Table 7 lists the relevant aspects of a conflict management attempt that are encoded in the variables judged important in both experiments. In both cases (and in all other experiments that we have performed) the variable that describes the previous relationship of the mediator to the two conflict parties proved to be most important. Using a decision tree with only this one variable can raise the predictive accuracy from about 57% for always predicting the majority class to about 63%. Adding the next variable that in both cases reflects the power score of one of the conflict parties (although different parties are used in both cases) further increases the predictive accuracy to above 65%. These decision trees that test only two variables are already competitive with the best trees of table 3.

Previous Relation of Mediator
Power Score and Disparity
Number of involved Parties
Mediation Environment
Issues

Table 7: Relevant aspects for predicting mediation outcome

4.4 Mediation Outcome

The purpose of this section is to show on an example how one can use Machine Learning methods to investigate more complex questions. So far we have only investigated which factors influence the outcome of conflict management attempts. In the experiment described in this section we tried to answer a more complex question:

> "If the chosen conflict management activity is mediation, which attributes are likely to have an influence on successfulness of a chosen mediation strategy?"

We distinguish three different strategies a mediator might use (with increasing level of activity on the mediator's side):

Communicative: The mediator is in a fairly passive role, acting largely as a channel of communication or go-between for the conflict parties.

Procedural: The mediator can control factors like the environment, number, type and agenda of meetings with the adversaries.

Directive: The mediator has influence on the contents and the process of the negotiations, for example by making substantive suggestions or by pressuring one of the parties to accept them.

The database contains 548 episodes where mediation has been chosen as the conflict management activity, all other instances were removed. Table 8 shows the frequency distributions of the mediation strategies in the remaining dataset.

Our goal was to get very simple rules. We applied all simplification methods described in section 4.2 including feature subset selection. From the original set of variables we have manually selected ten attributes that describe the most relevant aspects (see table 9) of mediation attempts (Bercovitch and Lamare 1993).

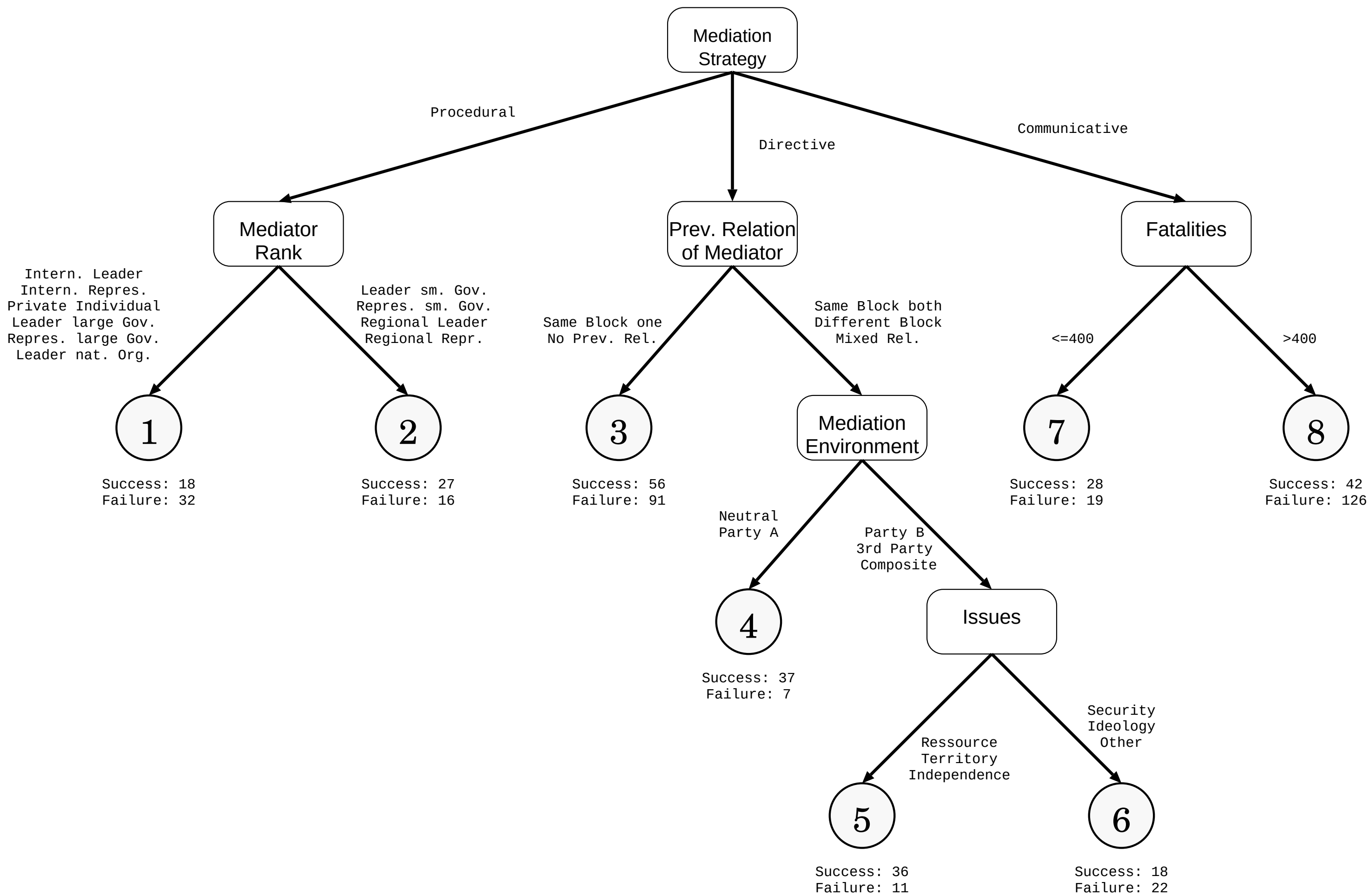

Figure 5: Predicting mediation outcome based on mediation strategy

Strategy	N	S	F	%
procedural	93	45	48	48.4
directive	278	146	132	52.5
communicative	215	70	145	32.6
total	586	261	325	44.5

Table 8: Mediation strategies

Fatalities
Mediation Environment
Mediation Strategy
Previous Relations of Mediator
Issues
Mediator Rank

Table 9: Relevant features for mediation outcome detected by statistical analysis

In addition we have forced C4.5 to choose mediation strategy as the root attribute and have set the parameters in a way to obtain a very simple tree (`-s -m 30 -c 25`). The resulting tree is shown in figure 5. It has an estimated predictive accuracy of 63.11% (± 4.0), which is a significant improvement over the mode prediction accuracy for this task (55.5%, see table 8).

The simple tree with only five decision nodes separates the data into 8 different clusters. Each cluster contains successful as well as unsuccessful mediation attempts. Interestingly, the characterizations of the clusters contain different variables for different mediation strategies, which indicates that different factors influence different mediation strategies. Communicative strategies only have a good chance for success when the number of fatalities in the conflict is low. Procedural strategies highly depend on the rank of the mediator: Leaders or representatives of small governments or regional organizations have much higher chance of success than international leaders, representatives or leaders of large governments or individuals. The success of directive strategies finally depends on the choice of the mediation environment and on the previous relation of the mediator to the conflict parties. Rule 6 for example specifies that

> **If** *the mediator has mixed relationships with the conflict parties,*
> *or is from the same block as both of them,*
> *or from a different block as both of them,*
> **and**

the mediation environment is party B's territory,
a third party's territory or a composite
and
the issues at stake are security or ideology
then *a choice of a directive mediation strategy*
is not likely to change the outcome of the mediation attempt.

In order to obtain an evaluation of the significance of these rules, we have performed a χ^2 test to see if the distribution of successful and unsuccessful mediation events in the 8 clusters is significantly different from their prior distribution.

Leaf	N	$E(S)$	$E(F)$	S	F	Δ	$\frac{\Delta^2}{E(S)}$	$\frac{\Delta^2}{E(F)}$
1	50	22.3	27.7	18	32	+4.3	0.83	0.67
2	43	19.2	23.8	27	16	−7.8	3.17	2.56
3	147	65.5	81.5	56	91	+9.5	1.38	1.11
4	44	19.6	24.4	37	7	−17.4	15.47	12.41
5	47	20.9	26.1	36	11	−15.1	10.91	8.74
6	40	17.8	22.2	18	22	−0.2	0.00	0.00
7	47	20.9	26.1	28	19	−7.1	2.41	1.93
8	168	74.8	93.2	42	126	+32.8	14.38	11.54

$\chi^2 = 87.51, df = 7, p << 1\%$

Table 10: Cluster analysis

Table 10 reveals that the hypothesis that the distribution of successful and unsuccessful mediation attempts over all 8 clusters is the same as the a priori distribution can be rejected with a very low probability of error. A closer analysis shows that almost all of the high χ^2 score can be attributed to rules 4, 5 and 8. Apparently the quality of the rules in the decision tree is not equal. However, bad rules can also be informative. It is easy to see that if the conditions of rule 6 are met the chances of being successful do not change significantly. However, other things being equal, different issues may significantly improve the chance of being successful (rule 5).

5 Discussion of Results

Rule induction via decision tree learning has been previously tried in international relations (see (Mallery 1988; Schrodt 1991a) for overviews). However, this case study employs new methods and is performed on a new dataset. From a methodological point of view we have used a commonly available[3] state of the art decision tree induction program. We have deliberately made this choice in order to illustrate the possibilites of Machine Learning without having to resort to programming skills.

Schrodt (1991b) has performed similar experiments in predicting interstate conflict outcomes using the Butterworth "Interstate Security Conflicts, 1945–1974" (Butterworth 1976). He used his own implementation of ID3, the predecessor of C4.5, to learn decision trees for predicting the effects of management efforts with respect to five different outcomes. In all his experiments the estimated predictive accuracy of the learned trees was below mode prediction accuracy, i.e. below the accuracy that one would achieve by always predicting the majority class. However, his implementation of ID3 was not capable of dealing with numeric data and, more importantly, did not have C4.5's extensive facilities for simplifying decision trees. The only method used for getting simpler trees was to restrict the number of used variables, which did not result in gains in accuracy. In our study, on the other hand, simple decision trees usually were able to achieve a higher predictive accuracy than an unpruned decision tree. However, even the unpruned tree obtained a significant gain in predictive accuracy compared to mode prediction. Predicting the outcome of conflict management attempts seems to be an easier task than to predict aspects of the outcome of the conflict itself. A reason for this might be that mediation events are more repetitive than international conflicts.

In section 4.1 we have seen that the rules contained in decision trees may be of very different quality. Table 4 has shown that only 12 rules of the more than 100 rules contained in the unpruned tree can correctly explain more than a fourth of the conflict management attempts in the database. Similarly, in section 4.4 we have seen that the simple tree of figure 5 also contains rules of very different quality. In general, dealing with single rules instead of whole trees seems to be a more promising approach. Mallery and Sherman (1993) report a variety of rules that have been learned with I^2D (Unseld and Mallery 1993), an improved version of ID3 that was specifically developed to deal with the structured nature of the SHERFACS dataset (Sherman 1988). However, a big problem is assessing the quality of the rules. Mallery and Sherman report the percentage of examples that are from the majority class and the total percentage of examples that are covered by a learned rule. Using the percentage of majority class examples as an evaluation is problematic, because it does not take into account the prior distribution of the examples and (if there are more than two classes) only gives information about the percentage of examples of one class. For this reason (Schrodt 1991b) has used an entropy ratio, similar to the information score proposed in (Kononenko and Bratko 1991) which gives a higher weight to correct predictions

[3]The book (Quinlan 1993) comes with the C source code of the programs. The programs are also available on disk or tape.

of rare classes. Alternatively, we have used a χ^2 cluster analysis to filter out rules that cover examples with a distribution of class values that is significantly different from the prior distribution. This method is only applicable for analyzing rules from a complete decision tree, because it needs a disjoint clustering of the examples. Some of the rules we have found contain aspects of previous findings with statistical analyses. Rule P1 from section 4.2 for example might in fact be considered as a special case of one of the findings of (Bercovitch, Anagnoson, and Wille 1991), where it has been shown that as the number of fatalities increases, the likelihood that mediation initiatives will prove successful suffers a corresponding decline.

We have also presented a method for automatic discovery of a relevant attribute subset using a Machine Learning method (John, Kohavi, and Pfleger 1994). Schrodt (1991b) has tried to achieve this by observing which attributes his algorithm typically selects near the root of the tree. Performing this type of analysis reveals that for example the attributes chosen in the tree of figure 4 are very different from the attributes that appear in table 6 which have produced a more accurate tree. It is interesting to compare the results of the Machine Learning method for detecting relevant features (table 7) with the results produced with classical statistical methods (table 9). There is obviously a considerable overlap. Almost all of the variables of table 9 appear in one of the two experiments of table 6, most of them in both. The most notable exception is the absence of mediation strategy. The number of fatalities, which is also not considered by Machine Learning, is partially reflected in the intensity of the conflict, which has been recognized as important, although only in one experiment. Using only fatalities for generating a decision tree would only yield 62.4% accuracy using the same grouping as in (Bercovitch and Lamare 1993). On the other hand, the Machine Learning method has attributed a higher significance to the previous relation of the mediator (63.3%). In addition aspects concerning the power of the conflict parties and about the number of parties involved on each side have been considered.

6 Summary

This paper presented a case study in using Machine Learning methods for analyzing the outcome of conflict management attempts based on the CONFMAN database. We believe that the main contributions of this research are:

- Commonly available inductive learning algorithms, in particular the C4.5 program (Quinlan 1993), are powerful tools for data analysis.
- Unsimplified rules and trees that have mostly been used in previous studies are likely to contain irrelevant tests that decrease their quality.

- Simple trees and rules are not only more understandable, but their use may also increase the chances of a successful prediction of international conflict management attempts.

- Automatic feature subset selection is an interesting alternative to using statistical analyses for determining relevant variables.

From a Machine Learning point of view, the most important finding that this analysis reveals is that not all rules that can be generated from the leaves of a decision tree are equally good. A separate analysis of the quality of different rules should be performed. We have used a method that filters out the relevant rules from a decision tree by examining the disjoint clustering a decision tree induces on the example space.

Compared to previous studies that applied Machine Learning methods to international conflict datasets, the CONFMAN database encoding conflict management attempts seems to be better suited for prediction than other datasets. The behavior of international mediators might be more repetitive than the conflicts per se. Using Machine Learning methods for a further analysis of other variables in this database is a promising topic for further research. A natural choice would be to examine the factors that influence a mediator to choose a certain mediation strategy (Bercovitch and Wells 1993).

Acknowledgements

We are very grateful to Peter Billing, John C. Mallery, Dwain Mefford, Frank Pfetsch, Philip A. Schrodt, Frank L. Sherman, and Sigrid Unseld for their many helpful comments and their continuous encouragement throughout our work.

The support of the project "The Potential Contribution of Artificial Intelligence to the Avoidance of Crises and Wars" (principal investigator: Robert Trappl) by the Federal Ministry for Science and Research is greatly appreciated.

References

BERCOVITCH, J., J. T. ANAGNOSON, & D. L. WILLE (1991). Some conceptual issues and empirical trends in the study of successful mediation in international relations. *Journal of Peace Research 28*(1), 7–17.

BERCOVITCH, J. & A. HOUSTON (1993, October). Influence of mediation characteristics and behavior on the success of mediation in international relations. *Journal of Conflict Resolution 4*(4), 297–321.

BERCOVITCH, J. & J. W. LAMARE (1993). The process of international mediation: An analysis of the determinants of successful and unsuccessful outcomes. *Australian Journal of Political Science 28*, 290–305.

BERCOVITCH, J. & J. LANGLEY (1993, December). The nature of dispute and the effectiveness of international mediation. *Journal of Conflict Resolution 37*(4), 670–691.

BERCOVITCH, J. & R. WELLS (1993, January). Evaluating mediation strategies - a theoretical and empirical analysis. *Peace & Change 18*(1), 3–25.

BUCHANAN, B. & E. SHORTLIFFE (Eds.) (1984). *Rule-Based Expert Systems — The MYCIN Experiments of the Stanford Programming Project*. Reading, MA: Addison-Wesley.

BUTTERWORTH, R. L. (1976). *Managing Interstate Conflict, 1945-74: Data with Synopses*. Pittsburgh: University of Pittsburgh University Center for International Studies.

EFRON, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans*. Society for Industrial and Applied Mathematics.

FEIGENBAUM, E. (1977). The art of artificial intelligence 1: Themes and case studies of knowledge engineering. Technical Report STAN-CS-77-621, Department of Computer Science, Stanford University.

HUDSON, V. M. (1991). *Artificial Intelligence and International Politics*. Boulder, CO: Westview Press.

JOHN, G. H., R. KOHAVI, & K. PFLEGER (1994). Irrelevant features and the subset selection problem. In *Machine Learning: Proceedings of the Eleventh International Conference*, Rutgers University, New Brunswick, NJ, pp. 121–129. Morgan Kaufmann Publishers, Inc.

KOHAVI, R. & G. H. JOHN (1994). Automatic parameter selection by minimizing estimated error. Submitted for publication. Available from `ftp://starry.stanford.edu/pub/ronnyk/c45ap.ps`.

KONONENKO, I. & I. BRATKO (1991). Information-based evaluation criterion for classifier's performance. *Machine Learning 6*, 67–80.

MALLERY, J. C. (1988). Thinking about foreign policy: Finding an appropriate role for artificial intelligence computers. Paper presented at the 1988 Annual Meeting of the International Studies Association.

MALLERY, J. C. (1994). Beyond correlation: Bringing artificial intelligence to events data. *International Interactions 20*(1-2), 101–145.

MALLERY, J. C. & F. L. SHERMAN (1993). Learning historical rules of major power intervention in the post-war international system. Paper prepared for presentation at the 1993 Annual Meeting of the International Studies Association.

QUINLAN, J. R. (1983). Learning efficient classification procedures and their application to chess end games. In R. S. Michalski, J. G. Carbonell, and T. M. Mitchell (Eds.), *Machine Learning. An Artificial Intelligence Approach*, pp. 463–482. Tioga Publishing Co.

QUINLAN, J. R. (1986). Induction of decision trees. *Machine Learning 1*, 81–106.

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

SCHRODT, P. A. (1991a). Artificial intelligence and international relations: An overview. See Hudson (1991).

SCHRODT, P. A. (1991b). Classification of interstate conflict outcomes using a bootstrapped ID3 algorithm. *Political Analysis* (2).

SCHRODT, P. A. (1991c). Pattern recognition of international event sequences: A machine learning approach. See Hudson (1991).

SHERMAN, F. L. (1988). Sherfacs: A new cross-paradigm, international conflict dataset. Paper written for presentation at the 1988 annual meeting of the International Studies Association.

STONE, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society B 36*, 111–147.

TRAPPL, R. (Ed.) (1992a). *Cybernetics and Systems '92*, Singapore. World Scientific.

TRAPPL, R. (1992b). The role of artificial intelligence in the avoidance of war. See Trappl (1992a), pp. 1667–1672.

UNSELD, S. (1994, November). A selective overview of recent projects in Artificial Intelligence / International Relations. Paper prepared for the Second Workshop on the Potential Contribution of AI to the Avoidance of Crises and Wars, Vienna.

UNSELD, S. D. & J. C. MALLERY (1993). Interaction detection in complex datamodels. MIT A.I. Memo No. 1298.