

https://dspace.org/

Artificial intelligence for conflict management

Habtemariam, Eyasu A.

2006-10-31T08:43:47Z
http://hdl.handle.net/10539/1515

Downloaded from a DSpace repository. https://dspace.org/

ARTIFICIAL INTELLIGENCE FOR CONFLICT MANAGEMENT

EYASU A. HABTEMARIAM

A RESEARCH REPORT SUBMITTED TO THE FACULTY OF ENGINEERING AND BUILT ENVIRONMENT, UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG, IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN ENGINEERING BY COURSEWORK AND RESEARCH REPORT

June 24, 2005

Declaration

I declare that this research report is my own, unaided work. It is being submitted for the Degree of Master of Science in Engineering in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other University.

_____

Eyasu A. Habtemariam

This _____ day of _____ 2005

Abstract

One of the risks that have a great impact on society is military conflict. Militarised Interstate Dispute (MID) is defined as an outcome of interstate interactions which result in either peace or conflict. Effective prediction of the possibility of conflict between states is a good decision support tool. Neural networks (NNs) have been implemented to predict militarised interstate disputes before Marwala and Lagazio [2004]. Support Vector Machines (SVMs) have proven to be very good prediction techniques in many other real world problems Chen and Odobez [2002]; Pires and Marwala [2004]. In this research we introduce SVMs to predict MID. The results found show that SVM is better in predicting conflict cases (true positives) without effectively reducing the number of correctly classified peace (true negatives) than NN. A sensitivity analysis for the influence of the dyadic (explanatory) variables shows that NN gives more consistent and easy to interpret results than SVM. Further investigation is required with regards to the sensitivity analysis of SVM.

Acknowledgements

First of all, I would like to thank God for without his grace and help I would never have completed this research. I would like to extend my heartfelt thanks to my supervisor Prof. Tshilidzi Marwala for his constant and priceless advices and guidance in every stage of the research. I also want to extend my gratitude to my co-supervisor Dr. M. Lagazio for her invaluable comments, support and allowing me to use their previous processed MID data sets. I want to thank all my colleagues for insightful discussions we had and their helpful suggestions. I would like to thank Miss Paida Nyazvivanda and Miss Unathi Mahola for proof reading the document for me. Last but not least, my deepest gratitude goes to all of my family members and best friends for their encouragement all the way up until now.

Contents

Declaration i

Abstract ii

Acknowledgements iii

Contents iv

List of Abbreviations vii

1 Introduction 1

1.1 Introduction 1

1.2 Militarised interstate dispute 2

1.3 Challenges facing quantitative analysis of MID 3

1.4 Problem statement 5

1.5 Artificial Intelligence (AI) 6

1.6 Motivation of the research 7

1.7 Structure of the document 8

2 Background and Related Work 10

2.1 Introduction 10

2.2 Conflict Modelling 11

2.3 MID Data 13
2.3.1 Dependent variable 14
2.3.2 Independent variables 15
2.4 Standard statistical techniques for quantitative conflict analysis 19
2.5 Learning machines for pattern recognition 22
2.6 Artificial Neural Networks (NNs) 25
2.6.1 Multi-Layer Perceptron 27
2.6.2 Back-propagation training algorithm 28
2.7 Support Vector Machines (SVMs) 30
2.8 Previous implementation of NNs for MID modelling 35
2.9 Support vector machines for MID prediction 38
2.10 Summary 39

3 Research Method **41**

3.1 Introduction 41
3.2 Neural networks for MID modelling 43
3.2.1 Model selection for NN 44
3.3 Support vector machines for MID modelling 45
3.3.1 Model selection for SVM 47
3.4 MID data sets 48
3.4.1 Rare-event prediction problems 50
3.5 Data preprocessing 51
3.6 Summary 52

4 Results and Discussions **54**

4.1 Introduction 54
4.2 Prediction results 55
4.3 Receiver operating characteristic (ROC) curve 56
4.4 Goodness-of-prediction of a classifier and ROC curve . . . 58
4.5 Sensitivity analysis 60
4.5.1 Experiment one 61
4.5.2 Experiment two 62
4.6 Alternative sensitivity analysis for SVM 65
4.7 Summary 66

5 Conclusions and Recommendations 70

5.1 Conclusion 70
5.2 Future work 75
5.2.1 Feature selection / sensitivity analysis 75
5.2.2 Decision trees 76

References 85

Appendix 1 86

List of Abbreviations

MID	Militarised Interstate Disputes
AI	Artificial Intelligence
NN	Artificial Neural Networks
SVM	Support Vector Machine
COW	Correlates of War
ROC	Receiver Operating Characteristic
AUC	Area Under the Curve
VC	Vapnik-Chervenonkis
MLP	Multi Layer Perceptron
KKT	Karush-Kuhn-Tucker
RBF	Radial Basis Function
SCG	Scaled Conjugate Gradient
TC	True Conflict
FC	False Conflict
TP	True Peace
FP	False Peace

Chapter 1

Introduction

1.1 Introduction

The history of mankind can be regarded as a history of conflicts. Peoples wage wars on other peoples for different reasons, dominance of power, ideology, beliefs, and different types of political, strategic and other interests to mention but a few. Political and international studies have and are still trying to come up with a clear reasons why states go to war. What are the major determinant factors for two or more states to have conflicts? Different types of scientific studies have put their efforts with the aim of finding sound explanations for the causes of interstate conflict.

Assuming that a general consensus to the factors that determine interstate conflicts can be reached, how we measure or quantify the effect of each factor to interstate conflict is another challenge which should be dealt with. Which of the factors is the single most determinant, if there is any, or which combination of factors influence militarised interstate

conflict (MID) to a greater extent than others? Are the factors that determine MID and the MID outcome linearly related or are they interdependent and exhibit complex relationships among each other? Which scientific technique and on what constraints is more appropriate to model the complex behaviour of interstate conflicts? All these and other similar complex questions are the challenges facing political science and international relations scholars.

The following sections introduce some basic terms and concepts that come along with interstate disputes. Section 2 will discuss the definition of militarised interstate conflict. Section 3 describes some of the challenges that are being faced by political scientists and international studies scholars in quantifying MID. Section 4 presents the problem statement of the research. A brief overview of artificial intelligence is given in Section 5. The motivation for the research will be given in Section 6 and the structure of the remaining document is presented in Section 7.

1.2 Militarised interstate dispute

Militarised interstate dispute (MID) according to Gochman and Maoz [1984] is defined as a set of interactions between or among states that can result in the actual use, display or threat of using military force in an explicit way. These interactions can result in either peace or conflict. What factors or variables determine MID is an important question which needs an appropriate answer. Different scientific studies and data collection measures have been underway with the aim of listing and recording

the most determinant factors which are believed to influence the MID outcome. One such project is the Correlates of War (COW) COW [2004] which is an ongoing effort to study the conditions associated with MID. It is a data collection program mostly used for quantitative studies of political and international relations.

For the study of militarised interstate disputes, countries are paired to form dyads. A dyad in the MID context refers to a pair of two states. A dyad-year then represents all the interactions that occur between two states in a specific year. The interactions are quantified in terms of some factors or variables that are widely believed (based on international relations theories) to influence the MID outcome. That is, any militarised conflict that outbreak in a specific year is a direct outcome of the interactions between the two states, expressed in the form of dyadic variables of the preceding year. Recently, the dyad-year is becoming a dominant approach and is being widely used in the empirical analysis of interstate conflicts [Bremer 1992; Maoz and Russett 1993; Beck *et al.* 2000; Reed 2000; Russett and Oneal 2001; Lagazio and Russett 2003] etc.

1.3 Challenges facing quantitative analysis of MID

Despite the fact that a lot of effort is made towards the quantitative analysis of MID, there is no clear-cut approach in terms of the causal variables, methodologies and techniques in use. Previously, statistical methods like multivariate analysis like logit and probit were used and the results found were not satisfactory. They showed high variance which

make them difficult to be reliable and tend to be inconsistent, hence, potentially confusing [Beck *et al.* 2000; Ray 2003]. Therefore, the results have to be taken cautiously and their interpretation require prior good knowledge of the problem domain.

As Ray [2003] pointed out, the results of multivariate analysis of MID conducted by various researchers give inconsistent causal effect rankings of dyadic variables towards the MID outcome. Ray [2003] suggests that the reason for the inconsistency is the inappropriateness of multivariate analysis when there are too many explanatory variables involved. This is because the model becomes too complex to be tractable especially if they have inter-connections among them. As Beck *et al.* [2000] note interstate conflicts exhibit complex and nonlinear characteristics which are not easy to represent using linear models. Therefore, it makes sense to model interstate conflicts using techniques that are better equipped in modelling complex and nonlinear problems.

Previous studies have been trying different approaches to address the problem quantitative analysis of militarised interstate conflicts. Some of the approaches, as it is mentioned in Beck *et al.* [2000], include improving the data and measurements of international conflict, modifying the existing statistical models and looking for new techniques which take into account the different factors that affect the empirical study of international conflicts. Our study falls on the third group of approaches which strive to come up with new militarised interstate dispute modelling techniques.

1.4 Problem statement

Previous quantitative analysis of militarised interstate disputes were mainly done using standard linear regression models which assume prior knowledge about the MID data. As Beck *et al.* [2000] put it, these models do not have the ability to forecast MID with a probability of more than 0.5 using dyad-year data. A model's out-of-sample forecasting ability is usually considered as a measure of its quality [Beck *et al.* 2000]. Previous studies have pointed out that explanatory variables of MID as being non-linear, highly interdependent and context dependent [Schrodt 1991; Beck *et al.* 2000; Lagazio and Russett 2003; Beck *et al.* 2004]. In contrast to these studies, DeMarchi *et al.* [2004] argue that neural networks as not being superior to either logit or linear discriminant estimators in forecasting MIDs.

As Beck and Jackman [1998] pointed out, it is a narrow methodological practice to consider the political or social relationships to exhibit global linearity. Taking into consideration this suggestion and the argument of the first group of scholars who believe the variables that determine the MID outcome are related in a non-linear fashion, it implies that the standard linear discriminant or logit regression techniques are not good enough for modelling the complex and non-linear militarised interstate conflicts which makes it inevitable to look for other new quantitative approaches to model MIDs.

Formally and briefly the problem statement can be stated as, *Given an interstate input data set expressed in the form of dyadic variables, is it*

possible to model it using artificial intelligence techniques that can deliver better results than those of the standard linear logistic regression models. Neural networks have already been employed for this purpose. Is it possible to extend the previous work of modelling interstate conflict using support vector machines.

1.5 Artificial Intelligence (AI)

Artificial intelligence (AI) is a branch of Computer Science that incorporates the idea of human intelligence into the traditional problem solving approach of computers using algorithms. AI has many different branches used to model very complex real world problems such as speech recognition, image processing, expert systems etc. They are new problem solving techniques that try to mimic the human brain's ability to process information. They have the property of learning from previous (training) data so as to generalise data which they have never come across (test data). Neural Networks and Support Vector Machines are two such techniques which are used in this research to model militarised interstate conflict.

Although support vector machines and neural networks have their own way of formulating problems mathematically, both of them are used to solve similar problems. They learn from examples, store the knowledge and apply it in the future for similar problems as the examples. They have been used in many pattern recognition problems. Pattern recognition problem involve understanding some underlying function that govern a process or relationship among entities. In the context of the study of

international conflicts, both neural networks and support vector machines try to work out what will happen next year with regards to their peace of conflict outcome based on a set of conditions that two states are at the moment. In other words, given a set of variables of two states, the techniques recognise the pattern of the variables and classify it as conflict or peace.

1.6 Motivation of the research

Even though neural networks have been applied successfully in various real world applications, they are not without their own weakness. Their learning process is aimed at minimising the empirical risk or the training error. But a zero training error does not necessarily mean it gives minimum error for previously unseen test examples. The other problem they face is their inability to always give a global solution and the solution depends on an initial random values of a weight vector. The learning algorithm employs gradient descent to minimise the squared error between the network output values and their corresponding target values. Since the error function can have multiple local minimum values, the gradient descent does not guarantee to give the global minimum. Support vector machines, on the other hand, employ a concept of structural risk (error) minimisation which aims to find a tighter bound on the test error. This is based on the empirical error and the capacity of the function class which is discussed in later chapters. Moreover, unlike neural networks support vector machines always give the global solution.

Support Vector Machines (SVMs) have proved to be very good in modelling complex problems and have high level of generalisation ability for wide range of real world problems. Recent empirical studies show that SVMs have outperformed neural networks in many real world applications. According to Chen and Odobez [2002], SVMs have resulted in better text texture verification than neural networks with multi-layer perceptron (MLP) architecture. Pires and Marwala [2004] have also come to a similar conclusion when comparing SVMs and neural networks for American option pricing. Based on these findings, modelling interstate conflict using SVMs is expected to give better results as compared to Neural Networks (NNs). These are the major motivations of doing the research.

1.7 Structure of the document

The next chapter discusses the background and previous quantitative analysis of militarised interstate disputes. This includes an introduction to what conflict modelling involves, the description of the MID data set that is used for this study, brief introduction to learning machines followed by more elaborate description of neural networks and support vector machines. It then discusses some previously used multivariate statistical techniques to model MID and their shortcomings. Moreover, it discusses previous studies that implement neural networks for the modelling of international disputes.

The third chapter presents the methodology of the research. The discus-

sion includes the details of how the experiments are done, support vector machine and neural networks for interstate dispute modelling, how they are trained, best model selection approaches, the training and testing data sets, the problems of rare-event data modelling and the preprocessing involved.

The results found and their respective discussions are given in the fourth chapter. The results found are given in the form of confusion matrix table and receiver operating characteristic (ROC) graphs. The ROC curve and the area under the curve (AUC) is employed to compare the prediction results of the two classifiers. The sensitivity analysis results and their respective discussions for both NN and SVM are given. The last chapter is the conclusion which gives the summary of the overall research approach, the methodology used, the results found, their discussions and identified future works.

Chapter 2

Background and Related Work

2.1 Introduction

As it is mentioned in the previous chapter, the history of mankind is a history of conflicts. Wars are waged between peoples for different justifiable or unjustifiable reasons. These wars have the potential to cause an unsurmountable amount of risk to the economy and hence the lives of the people of the involved countries. The first step which should be taken in order to avoid or minimise these interstate conflicts is to better understand their root causes. As some wise people say, "knowing your sickness is half of the remedy", so also is this with the interstate conflicts.

Different studies have and still are being done to understand what are the determinant factors that make countries to have conflict among each other. By doing so, the studies try to quantify the probability of states falling into the trap of militarised disputes. The Correlates of War (COW) is a project aimed at studying and exploring the factors that

lead to the outbreak of war and militarised disputes [COW 2004].

Although international study intellectuals are putting much effort to this regard, due to the complexity of the problem, many of the studies have not come up with unified conclusion. The causes of these disagreements can be attributed to different reasons some of which are the quality of MID data, the relevance of their measurement and the available statistical analysis methods in use.

The remaining sections of this chapter are organised as follows. First it describes what interstate conflict modelling is and the preconditions that come along with it. Section 3 describes the details of the MID data, the dependent variable and the list of the independent variables that are used for this study. Previous efforts of quantitative analysis of MID are described in Section 4. Section 5 gives a brief introduction of learning machines and Sections 6 and 7 go further in elaborating the basics of neural networks and support vector machines. The last two sections look at previous research works on using neural networks and support vector machines for the quantitative analysis of international conflicts, and the conclusion of the chapter follows at the end.

2.2 Conflict Modelling

Modelling international conflicts involves quantitative and empirical analysis based on existing dyadic information of states. Dyad-year in our context refers to a pair of states in a particular year. Political scientists use

dyadic parameters as a measure of the possibility that two states might have a militarised conflict. A historic data of each dyad showing their interactions during that particular year is recorded. The interactions are expressed and quantified in the form of dyadic variables. The values of these variables are believed to be the determinant factors whether the member states of a dyad will be at peace or conflict the following year.

There are different international relations theories that are put forward which in essence are believed by their respective advocates to govern the process and interactions of states. Some among these theories include realism and liberalism. Realism theory states that the principal actors of world politics are states which always strive for power and their national interests [Morgenthau 1973]. Liberalism, on the other hand, believe that states are one among many actors in the world politics. States are inter-dependent and cooperate through international organisations to play an important role in the world politics [Baldwin 1993]. These theories use a set of their own parameters to measure the interactions among states.

Although conflict modelling is based on some predetermined parameters which are then quantitatively analysed to predict the corresponding MID result, scholars do not fully agree in listing these variables. That is to say there are competing arguments to why states have conflicts and hence various parameter lists corresponding to the various theories. Researchers tend to use their own set of variables to fit their own respective theories they want to prove. Hence, the usage of the variables vary from one research to another. Various researchers have put great effort in compiling these variables from different sources [Gochman and Maoz 1984; Tucker

1997; Russett and Oneal 2001; Jaggers and Gurr 1996]. The data collection process is an ongoing effort that strives to improve the available data which the quantitative interstate conflict analysis depends on. One such kind of work is the correlates of war (COW) project which collects and studies the conditions associated with MID [COW 2004].

Even though immense data collection effort have been made, still a lot of research is underway to come up with satisfactory and reliable conflict models. One of the major reasons why conflict modelling is complex, according to Beck *et al.* [2000], is that the causes of conflict are tiny for the vast majority of dyads. That is, international conflict is a rare event and the processes that drive it vary for each incident. This makes it to be highly nonlinear, very interactive and context dependent. Previous linear discriminant and logit regression techniques for MID forecasting give inconsistent results that vary from research to research. This implies there is a possibility of mismatch between the currently available MID data and commonly used linear and normal statistical techniques [Beck *et al.* 2000].

2.3 MID Data

Militarised interstate conflict (MID) is an outcome of states' interactions. The assumption is that these interstate interactions determine whether the corresponding states will be in peace or conflict after those interactions. These interactions are commonly represented in terms of dyadic variables. As is mentioned in Ray [2003], it was Bremer Bremer [1992]

who popularised the use of dyad-year as a unit of quantitative international conflict analysis. Later on, many more researchers followed in his footsteps to use dyadic variables for empirical studies of interstate conflicts or wars.

The dyadic variables which are of interest to our study are discussed below. They have been compiled by Russett and Oneal [2001] based on the Correlates of War and *Polity III* data sources [COW 2004; Jaggers and Gurr 1996] and have also been used by [Oneal and Russett 2001; Lagazio and Russett 2003; Marwala and Lagazio 2004] for the same purpose. The description of each variable, what it entails and the possible values it can have is described in detail.

2.3.1 Dependent variable

Our dependent variable is interstate dispute (MID). This variable can have values of either 0 or 1. If any dyad is involved in dispute in any particular year, the MID variable gets value 1, and 0 otherwise. Dispute according to COW [2004] is defined as either or both of the states of a dyad threaten to use force, make a demonstration of force or actually use military force against each other. Since the goal is to predict the onset of a conflict as opposed to its continuation, only the initial year of the militarised conflict is taken into consideration.

2.3.2 Independent variables

All the values of the independent variables lag by one year from the MID variable as it was used in [Oneal and Russett 2000]. They have done this based on previous studies about test of causality. That is, a dependent variable Y is said to be caused by dependent variable X if it is possible to predict Y on the basis of past values of X. That is why the need for lagging the independent variables by one year. It is then assumed that the interactions of these independent (explanatory) variables in a specific year determine the outcome of the MID in the following year. The other point worth mentioning is, not all studies agree on how the variables affect conflicts or wars. Almost for every variable two opposing and seemingly convincing arguments come forward on how it affects the MID outcome. The description of each variable is given below as it is described in previous studies [Oneal and Russett 2000; Lagazio and Russett 2003; Marwala and Lagazio 2004].

Democracy

This variable is a quantitative measure of the political characteristic of regimes. The value is calculated by subtracting the autocracy score of each state from its respective democracy score. The data source used to calculate this variable is *Polity III* data [Jaggers and Gurr 1996]. The values range from -10 to 10. A value -10 represents the worst autocratic state while a score of 10 means a very democratic country. The joint democracy level is then calculated as the minimum of the two scores. This

is because it is assumed that the less democratic state plays a determinant role in starting conflict as compared to the more democratic state.

Previous studies differ in their arguments about the relationship between democracy and conflict or war [Bremer 1992]. Some say democratic states are less conflict prone among each other than those who are not Maoz and Abdolali [1989] while others argue democracies are neither more nor less conflict prone than others [Small and Singer 1976]. But most of the recent studies tend to back the first argument that says democracies are more at peace with each other than others.

Economic Interdependence / Dependency

This variable is calculated based on the statistics of the bilateral trade of the countries involved. To be specific, the dependency variable is calculated as the minimum of bilateral trade-to-GDP ratio of the two involved states. It is a continuous variable that measures the level of economic interdependence of the less economically dependent state in the dyad.

As Copeland [1996] pointed out, there are two opposing arguments that come forth in regard to how economic interdependence affect war between states. The first argument comes from the liberal group of thought which says that economic interdependence lowers the likelihood of war. The realist group of thought propose an opposing argument that economic interdependence can increase the probability of war, especially in anarchic states since the most worrying factor for them is their security which can

be vulnerable as the economic ties gets stronger.

Capability Ratio

Capability ratio is the power parity between the states in a dyad. It is measured as the logarithm to the base 10 of the ratio of the total population plus number of people in urban areas plus industrial energy consumption plus iron and steel production plus number of military personnel in active duty plus military expenditure in dollars in the last 5 years measured on stronger country to weaker country. Some studies suggest that a balance of military capability between the members of a dyad deters any possible disputes while others believe the opposite.

As Bremer [1992] put it, there are two opposing and at the same time convincing arguments with regard to the relationship between conflict and capabilities of states. One side says that a weaker state would not dare to have conflict with a stronger state. Therefore, preponderance promotes peace according to them. The second group argues that if the capabilities of the two states are relatively equal, neither of them would initiate conflict since it is not certain of victory. Recent studies tend to emphasise preponderance results in peace outcome [Gochman 1990b].

Alliance

This variable measures the degree of alliance between two states. Its value becomes 1 if the two states have any mutual defence treaty or neutrality pact, and 0 otherwise. It is assumed that allied states are less

likely to have disputes as compared to those who have no alliance.

In a similar fashion Bremer [1992] continues to present how war and alliance are related. He suggested that based on the theoretical expectations, alliance has a deterring effect to conflict. But some empirical studies like Ray [1990b] found out that allies are more prone to conflict than non-allied states. Recent studies, on the other hand, show that alliance is negatively related with conflict [Maoz and Russett 1993; Oneal and Russett 1997; Reed 2000].

Contiguity

The contiguity is assigned a value of 1 if the dyad members share a common border or 0 otherwise. Countries that share borders have more probability and reason to fight (eg. territorial boundaries, natural resources, grievances of cross-border ethnic groups and so on) as opposed to those who do not. It is widely accepted that contiguous states are more conflict prone than noncontiguous. This is because it is more likely for them to have conflict of interest (eg. shared resources) to engage in dispute [Bremer 1992].

Distance

This variable is similar to the previous one. If the two states are close enough so that at least one of them can reach with effective military force, this may enhance the possibility of disputes. The variable is calculated as the natural logarithm of the distance in kilometres between the capitals

of the two states (or between the major ports for the largest countries). Distance is also believed to have a similar (positive) relationship towards conflict as contiguity.

Major Power

The correlates of war classifies countries as major powers if they have substantial destructive power globally based on a consensus of historians. The variable then assumes a value of 1 if either member of the dyad is a superpower and 0 otherwise. It is commonly believed that major powers tend to be more conflict prone than minor powers.

2.4 Standard statistical techniques for quantitative conflict analysis

Before the recent introduction of artificial intelligence techniques, specifically neural networks for the quantitative analysis of interstate conflicts or wars, researchers were commonly using different statistical techniques for the same purpose. In the following paragraphs, we look at some of the studies and their approaches in doing empirical interstate conflict or war analysis.

One of the widely referenced paper that did a bivariate and multivariate statistical analysis was [Bremer 1992]. The paper as it is pointed out by Ray [2003], has three distinct qualities from other previous studies. It adopted the dyad as a unit of analysis, the study covers quite

a wide range of temporal (from 1816 to 1965) and spacial domain (all possible dyads). Bremer carefully studied the effect of seven variables that are assumed as predictors of war. The definition of the variables with the exception of few are defined as those of the Correlates of War (COW) [COW 2004]. These variables include geographical proximity, power ratios, power status, alliance ties, democracy, development and militarisation.

Bremer [1992] used Poisson regression model instead of a standard regression model for his experiment. First he conducted a bivariate analysis of the effect of each variable in causing or deterring war. A positive or negative sign in the following rankings shows whether a variable is directly or inversely related to war. The results ranked from strongest to weakest are proximity(+), power status(+), alliance(+), militarisation(+), democracy(-), development(+) and power difference(-). A similar multivariate analysis shows some changes in the rankings. They are ranked from strongest to the weakest as proximity(+), democracy(-), development(+), power status(+), power difference(-), alliance(-) and militarisation(+). In looking at the possible effect of interactions among each other, he found out that a combination of militarisation and alliance has a very significant effect on the outcome. The ranking of the effects of the variables became (from strongest to weakest) proximity(+), alliance(-), development(-), democracy(-), power difference(-), power status(+) and militarised-alliance (+).

Another similar research which took the dyad as a unit of analysis is [Oneal and Russett 1999b]. Their goal was to study interstate conflicts

as opposed to Bremer [1992] who studied interstate wars. They employed pooled cross-sectional time-series regression analysis to see the effect of democracy, economic interdependence and joint membership in international organisations (IGO) on the onset of militarised interstate conflicts. They have also included four other variables which are capability ratio, alliance, contiguity and distance to control their effect on interstate conflicts.

In their analysis, they have looked at the influence of the variables for a specific time ranges (eg. before World War II) and specific dyads (eg. relevant dyads). Relevant dyads are dyads that are either contiguous or include major power. Their main goal was to evaluate the Kantian peace which focuses on the effect of democracy, interdependence and membership in IGO on interstate conflict. They used two data sets, one that includes all dyads while the other only relevant dyads. Their results show both democracy and interdependence have a significant effect in deterring conflicts in both data sets. However, the effect of the IGO variable was significant only in the case where the data set includes relevant dyads.

Ray [2003] has done an extensive comparison of different multivariate models for MID. The comparison mainly focuses on the studies that employ dyadic analysis of interstate conflicts or wars. Each study has its own set of variables to deal with on the causal impact of war or conflict. Ray [2003] gives the results of each research that was considered in a tabular form. His main point of argument lies on the question of whether multivariate models for MID analysis are being implemented in the way

they should. The point he tries to make is that the multivariate models of MID results in varying rankings of the significance of the variables on dispute or war. A small change of a variable usually has a chain effect on other variables which results in changing the rankings. Therefore, he emphasises that multivariate MID analysis should be simplified by reducing the number of explanatory variables.

2.5 Learning machines for pattern recognition

A pattern according to Jain *et al.* [2000], is something that behaves in some kind of order or rule as opposed to a chaotic way. Pattern recognition then simply means searching for an underlying rule that governs a process, event or a data source. There are some common aspects that are taken into consideration when designing a pattern recognition system. These include data acquisition and preprocessing, data representation and decision making [Jain *et al.* 2000]. It should be noted that each aspect can influence the subsequent steps positively or negatively. A properly defined and sufficiently constrained recognition problem results in compact pattern representation and simple decision making strategy. The data acquisition and preprocessing mainly deals with how the data is acquired and transformed so as to make the system more robust. An example of a simple preprocessing may be normalising the input/output data or removing any noise.

Any statistical pattern recognition system can be modelled as in figure 2.1. The modelling process has two major steps which are the training

and classification. In the training mode, the feature extraction/selection module finds any features or patterns for representing the input (training) data which the classifier uses to partition feature data. In the classification mode, the preprocessing and feature extraction is similar to the training mode. The trained classifier then classifies the input (test) data once they are extracted.

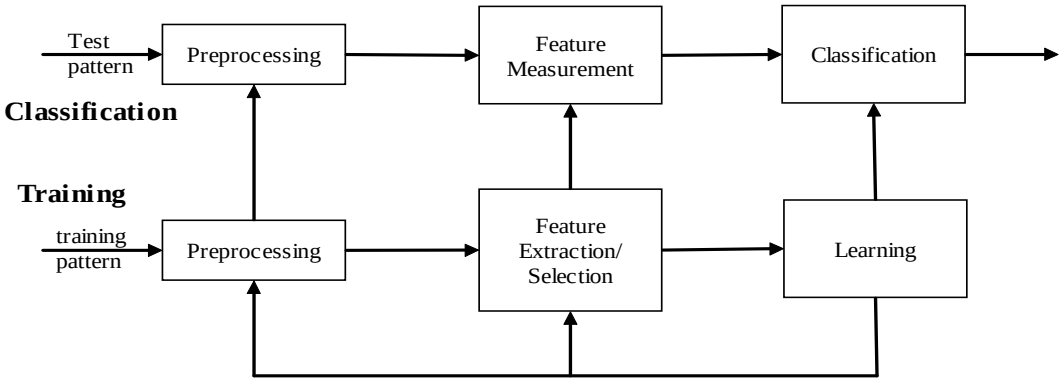


Figure 2.1: Pattern recognition model [Jain *et al.* 2000]

.

There are various types of learning machines used for different kinds of purposes. The two major applications of learning machines for pattern recognition are classification and regression. Learning problem for classification can be defined as finding a rule that assigns an object or entity into different classes [M*ü*ller *et al.* 2001]. The rule that governs the classification is devised based on an acquired knowledge about the objects from some examples. The process of knowledge acquisition from given examples is called training. In this paper, we look at two different types of learning machines for the purpose of classifying input patterns of MID data. Both techniques learn a model or pattern based on a training MID data to classify previously unseen MID test data.

As Jain *et al.* [2000] put it, learning machines are statistical prediction and modelling algorithms. Besides the classification and regression models described above, there is also a third modelling problem which is called density estimation. Classification and regression models are called supervised learning problems while density estimation is called unsupervised learning problem. Classification problem involves assigning a set of input vectors $\mathbf{x}$ into n number of classes $C_1, C_2, ...C_n$. Regression problem, on the other hand, entails estimating the values of continuous variables.

In the perspective of probability, classification and regression problems involve the estimation of conditional densities. It can be assumed that the main goal of a learning machine as being the estimation of these densities [Jordan and Bishop 1996]. Given a set of input patterns $\mathbf{x}$ and target outputs $\mathbf{t}$, the joint distribution is given by:

$$p(\mathbf{x},\mathbf{t}) = p(\mathbf{t}|\mathbf{x})p(\mathbf{x}) \tag{2.1}$$

The joint distribution depends on the probability distribution of the input and a conditional density. Thus a pattern vector $\mathbf{x}$ belonging to class C_i is assumed as an observation drawn randomly from a class conditional probability function $p(\mathbf{x}|C_i)$. Considering a simple two-class classification problem, the probability of the target equals one of the possible targets is given using Bayes rule as:

$$p(t_i|x) = \frac{p(x|t_i)p(t_i)}{p(x)} \tag{2.2}$$

where $p(t_i|\mathbf{x})$ is the *posterior probability* of class i given the input $\mathbf{x}$

[Jordan and Bishop 1996]. This posterior probability can be written in the form of a *logistic function*

$$y = \frac{1}{1 + e^{-z}} \tag{2.3}$$

The z is called a *discriminant function* which is usually used to decide on a class membership [Jordan and Bishop 1996].

2.6 Artificial Neural Networks (NNs)

A neural network is a processor that resembles the brain in its ability to acquire knowledge from its environment and store the information in some synaptic weights [Haykin 1999]. As its name tells, it was first inspired by the functionality of the brain's neurons. The objective was then to develop a simplified mathematical models of brain-like systems [Rumelhart *et al.* 1994]. It is composed of simple extremely large number of neurons with many interconnections that are capable of processing information in a massively parallel fashion. Neural networks can do computations in a highly parallelised fashion and also can make generalisations once they are trained using some input example data [Haykin 1999; Jain *et al.* 2000].

In the perspective of statistics, neural networks can be regarded as a generalisation of conventional pattern recognition statistical techniques [Bishop 1995; Jain *et al.* 2000]. An example of a pattern could be fingerprint image, human face, speech signal etc. Pattern recognition, as

described by Bishop [1995], is a wide variety of information processing problems like handwritten character classification, fault detection and speech recognition. It is the study of machines in relation to how they study an environment, learn for any pattern of interest and make important decisions about the patterns [Jain *et al.* 2000]. Human beings make sound and better decisions based on the degree of their knowledge about a pattern [Jain *et al.* 2000]. Pattern recognition is not an easy task for computers to do while humans employ them in their daily activities without putting any effort. Efficient pattern recognition solutions with robust theoretical basis are then required to address pattern recognition problem .

Neural networks can also be thought of as labelled acyclic directed graphs. They have at least one node with no inputs and one node without outputs. Numerical values attached to each node are used to represent the patterns. The transformation between patterns is achieved by message-passing algorithms [Jordan and Bishop 1996]. In other words, each node except the input nodes has a label which is calculated as some type of transformation of all the inputs coming into that node [Vidyasagar 1997].

Neural networks are ways of mapping input vectors $\mathbf{x}$ using a set of M nonlinear functions $\phi(\mathbf{x})$, $j = 1, ..., M$, and combine them linearly.

$$y_k(\mathbf{x}) = \sum_j w_{kj}\phi_j(\mathbf{x}) \tag{2.4}$$

where $\phi_j(\mathbf{x})$ are *basis functions* which are adaptive and have weight parameters that can be adjusted based on the observed data input. It is

the choice of the basis function that determines the type of network.

There are different types of neural networks. The most commonly used include the feed-forward network and radial basis function (RBF) networks. Single and multi-layer perceptron (MLP) are examples of a feed-forward network. Both network types are organised into layers with unidirectional connections between nodes of subsequent layers. The learning process basically is conducted by adjusting the connection weights to ensure the ability to classify or regress with as much accuracy as possible.

2.6.1 Multi-Layer Perceptron

The most widely used feed-forward neural network is multi-layer perceptron (MLP) with two adjustable layers of weights. It has input, hidden and output layers as shown in figure 2.2. The input layer represents independent variables, the hidden layer latent variables and the output layer the dependent variables [Zeng 1999]. Feed-forward neural networks provide a framework to represent a non-linear functional mapping of a set of d input variables x_i, $i = 1, ..., d$ into a set of c output variables y_j, $j = 1, ..., c$ [Bishop 1995].

The relationship between the input and output units of the neural network is represented by the following function [Bishop 1995]:

$$y_k = f_{outer}\left(\sum_{j=1}^{M} W_{kj}^{(2)} f_{inner}\left(\sum_{i=1}^{d} W_{ji}^{(1)} x_i + W_{j0}^{(1)}\right) + W_{k0}^{(2)}\right) \tag{2.5}$$

where $W_{ji}^{(1)}$ and $W_{kj}^{(2)}$ are the first and second layer weights going from

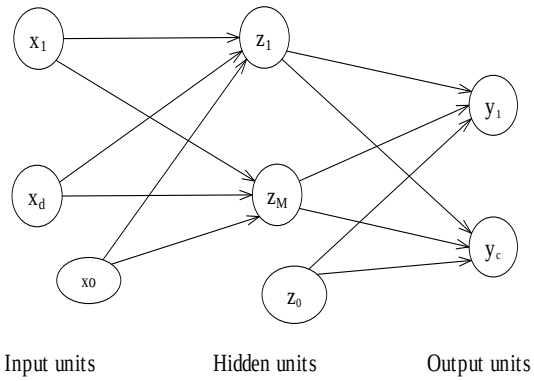


Figure 2.2: A feed-forward network with two layers of adaptive weights [Marwala and Lagazio 2004]

.

input i to hidden unit j and hidden unit j to output unit k respectively, M is the number of the hidden units, d is the number of input units, while $W_{j0}^{(1)}$ and $W_{k0}^{(2)}$ represent the biases of the hidden and output units respectively. f_{outer} represents the output activation function and f_{inner} corresponds to the activation function for the hidden unit.

2.6.2 Back-propagation training algorithm

Training of a neural network means adjusting the weights of the network based on the input-output data sets. There are two types of training methods which are called supervised and unsupervised. In the case of supervised training, the network is fed with *target* values along with each input vector which the network is trained to predict. MLP is trained by supervised learning using the iterative back-propagation (BP) algorithm Bishop [1995]. Although there are different varieties of BP algorithms, the basic idea of how the algorithm works is described below.

The algorithm begins by first assigning random values to all the weights.

It then passes the input patterns iteratively through the processing units of the network to get an output result. The aim is to adjust the weights so that the difference between the output and the corresponding target is minimised. This difference is usually called error function, E. The training involves two distinct steps first of which is the evaluation of the derivatives of the error function with respect to the weights and second computing adjustments for the weights based on the evaluated derivatives [Bishop 1995]. The required derivative is given by:

$$\frac{\partial E^n}{\partial w_{ji}} = \delta_j z_i \tag{2.6}$$

where z_i is the activation of unit i and δ_j is an error term for each node j and applying chain rule for partial derivatives is calculated as:

$$\begin{aligned} \delta_j &\equiv \frac{\partial E^n}{\partial a_j} \\ &\equiv g'(a_j) \sum_k w_{kj} \delta_k \end{aligned} \tag{2.7}$$

where $g(a_j)$ is the activation function at unit j and g' is its derivative. The input at each unit is calculated as $a_j = \sum_i w_{ji} z_i$ Bishop [1995].

Since evaluation of δ_j for the output units is straightforward, δ for a particular hidden unit can be obtained by propagating the δ's backward from upper units in the network down to the lower units. That is why the name backward propagation is used.

Once the derivatives are evaluated, we need a way of updating the weights based on the derivatives. There are several types of training strategies (parameter optimisation techniques) in use. One simple strategy, a fixed-

step gradient descent, updates the weight as:

$$\Delta W_{ji} = -\eta \delta_j x_i \tag{2.8}$$

2.7 Support Vector Machines (SVMs)

Given an empirical data $(x_1, y_1), ..., (x_m, y_m) \in \mathcal{X}$ x $\{\pm 1\}$, where x_i are drawn from a nonempty set $\mathcal{X}$, a learning machine is one that can generalise unseen data points based on the empirical data [Schölkopf and Smola 2003]. This means, given a new pattern $x' \in \mathcal{X}$ find a value of $y \in \{\pm 1\}$ that corresponds to x based on some similarity measure between x and x'.

According to Müller *et al.* [2001], the classification problem can be formally stated as estimating a function $f : \mathbb{R}^N \rightarrow \{-1, 1\}$ based on an input-output training data generated from an independently, identically distributed unknown probability distribution $P(\mathbf{x}, y)$ such that f will be able to classify previously unseen $(\mathbf{x}, y)$ pairs. It is known, however, that there are many such functions that can learn well the training data. We have also to keep in mind that the best fit for the training data does not necessarily mean it is the best to generalise previously unseen data (test data). We have to restrict our choice of functions based on a predefined criteria. One such criteria takes into account the expected test error which is commonly called *risk*. The best such function is the one

that minimises the *risk* which is given by

$$R[f] = \int L(f(x), y) dP(x, y) \tag{2.9}$$

where L represents a loss function. Since the underlying probability distribution P is unknown, equation 2.9 cannot be solved directly. The best we can do is to find an upper bound for the risk function which depends on both the empirical risk and the capacity of the function class. The capacity of the function class, the Vapnik-Chervonenkis (VC) dimension, is defined as the largest number of h points that can be separated in all possible ways using functions of the given class [Vapnik 1995]. The empirical risk is calculated as:

$$R[f]_{emp} = \frac{1}{n} \sum_{i+1}^{n} l(f(x_i), y_i) \tag{2.10}$$

The upper bound for the risk is then given by:

$$R[f] = R[f]_{emp} + \sqrt{\frac{h\left(\ln \frac{2n}{h} + 1\right) - \ln\left(\frac{\delta}{4}\right)}{n}} \tag{2.11}$$

where $h \in \mathbb{N}^+$ is the VC dimension of the function class F, $f \in F$ and $\delta > 0$ holds true for all δ.

The second term of the right hand side of the equation is a $confidence\ term$ which is introduced in the structural risk minimisation concept by Vapnik [1982].

Even though it is not always the case in real data, let us assume the data is linearly separable. This means that there is a set of hyperplane that

can separate the data into two classes which are represented as:

$$f(x) = (\mathbf{w}.\mathbf{x}) + b \tag{2.12}$$

where $\mathbf{w} \in \mathcal{R}^N$ is an adjustable weight vector and $\mathbf{b} \in R$ is an offset. A simple separating hyperplane is shown in figure 2.3 [Müller *et al.* 2001].

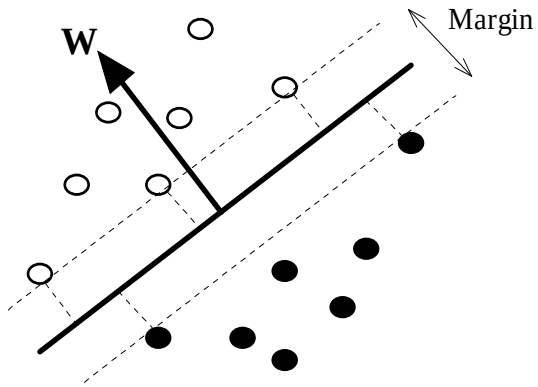


Figure 2.3: A linear SVM classifier and margins: A linear classifier is defined by a hyperplane's normal vector $\mathbf{w}$ and an offset b, i.e. the decision boundary is $\{\mathbf{x}|\mathbf{w}.\mathbf{x} + \text{b} = 0\}$ (thick line). Each of the two half spaces defined by this hyperplane corresponds to one class, i.e. $f(x) = sign((\mathbf{w}.\mathbf{x}) + b)$. [Müller *et al.* 2001]

The *margin* is defined as the shortest distance between any two points on either side of the hyperplane measured perpendicular to the hyperplane [Müller *et al.* 2001]. Among all the possible separating hyperplanes, the one with the maximum *margin* of separation is selected in order to get a tighter bound.

Since most of real world problems are very complex in their nature, using linear classifiers may not give good results. Therefore, there is a need for complex classifiers. But on the other hand, complex classifiers are not easy to deal with. The concept of feature space then comes into place to play a major role. This means, the input data is mapped into a higher dimensional feature space $\mathcal{F}$ which can make the learning process much

easier. This is achieved mainly by reducing the complexity of the classifier function. In other words, instead of using a very complex classifier in the input space, first map the input data into a higher dimensional feature space f and then use a simple class of decision rule (eg. linear classifiers) to classify the data [M*ü*ller *et al.* 2001]. A mapping function called *kernel function* that maps the input vector into the feature space is introduced. This implies that each training example $\mathbf{x}_i$ is substituted with the mapping function $\phi(\mathbf{x}_i)$ so that equation 2.12 becomes:

$$y_i((\mathbf{w}.\Phi(\mathbf{x}_i) + b), i = 1, 2, ..., n \tag{2.13}$$

The VC dimension h in the feature space $\mathcal{F}$ is bounded according to $h \leq \|W\|^2 R^2 + 1$ where R is the radius of the smallest sphere around the training data M*ü*ller *et al.* [2001]. Hence minimising the expected risk can be stated as an optimisation problem as:

$$\begin{array}{ll} \min\limits_{w,b} & \dfrac{1}{2}\|W\|^2 \end{array} \tag{2.14}$$

$$\text{subject to} \quad y_i((\mathbf{w}.\mathbf{x}_i) + b) \geq 1,\ i = 1, ..., m.$$

Assuming that we can only access the feature space by only using dot products, (2.14) is transformed into a dual optimisation problem by introducing Lagrangian multipliers $\alpha_i \geq 0$ and a Lagrangian of the form Sch*ö*lkopf and Smola [2003]:

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \alpha_i(y_i((\mathbf{x}_i.\mathbf{w}) + b) - 1) \tag{2.15}$$

The Lagrangian is then minimised with respect to α_i and maximised

with respect to $\mathbf{w}$ and b. Besides, the fact that the derivatives of L with respect to the $\mathbf{w}$ and b becomes zero at the saddle point gives us

$$\sum_{i=1}^{n} \alpha_i y_i = 0 \tag{2.16}$$

and

$$\mathbf{w} = \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i \tag{2.17}$$

Substituting (2.16) and (2.17) in (2.15) gives us the optimisation problem [Burges 1998; M*ü*ller *et al.* 2001; Sch*ö*lkopf and Smola 2003]:

$$\max_{\alpha} \quad \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j) \tag{2.18}$$

$$\text{subject to} \quad \alpha_i \geq 0, i = 1, ..., n \quad \text{and} \quad \textstyle\sum_{i=1}^{n} \alpha_i y_i = 0.$$

The Lagrangian coefficients α_i are obtained by solving equation (2.18) which in turn is used to solve $\mathbf{w}$ to give the non-linear decision function [M*ü*ller *et al.* 2001; Sch*ö*lkopf and Smola 2003]:

$$\begin{aligned} f(x) &= sgn\left(\sum_{i=1}^{n} y_i \alpha_i (\Phi(\mathbf{x}).\Phi(\mathbf{x}_i)) + b\right) \\ &= sgn\left(\sum_{i=1}^{n} y_i \alpha_i k(\mathbf{x}, \mathbf{x}_i) + b\right) \end{aligned}$$

In the case when the data is not linearly separable, slack variables ξ_i, $i = 1, ..., n$ are introduced to relax the constraints of the margin as

$$y_i((\mathbf{w}.\xi(\mathbf{x}_i)) + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, ..., n \tag{2.19}$$

A trade off is made between the VC dimension and the complexity term

of (2.11) which gives the optimisation problem

$$\min_{\mathbf{w}, b, \xi} \qquad \frac{1}{2}\|W\|^2 + C\sum_{i=1}^{n}\xi_i$$

where $C > 0$ is a regularisation constant that determines the above mentioned trade-off. The dual optimisation problem is then given by [Müller *et al.* 2001]:

$$\max_{\alpha} \qquad \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j) \tag{2.20}$$

$$\text{subject to} \qquad 0 \leq \alpha_i \leq C, i = 1, ..., n \quad \text{and} \quad \textstyle\sum_{i=1}^{n}\alpha_i y_i = 0.$$

A Karush-Kuhn-Tucker (KKT) condition which says only the α_i's associated with the training values $\mathbf{x}_i$'s on or inside the *margin* area have non-zero values, is applied to the above optimisation problem to find the α_i's and the threshold variable b reasonably and then the decision function f [Müller *et al.* 2001].

2.8 Previous implementation of NNs for MID modelling

Prediction and classification problems, as mentioned before, involve estimating an underlying function that generates some type of real world data observations. A good model is the one which can predict or estimate the data with as little error as possible. Forecast accuracy is regarded as a major criteria to evaluate the goodness of a model Beck *et al.* [2000]

and a good model is one that can generalise previously unseen data [Zeng 1999]. If a model can forecast new data with good accuracy, it means that it has been able to find the true causal underlying structure of the data.

The most commonly used statistical models for classification like logit and probit require *a priori* assumptions of the data. They assume that there is a linear form of distribution for the underlying functions [Zeng 1999]. Neural networks on the other hand do not put any *prior* assumptions which enable them to approximate any arbitrary functional mappings. The above assumption has its own limitations which makes neural networks more attractive as classification and forecasting models.

Beck *et al.* [2000]; Lagazio and Russett [2003], describe interstate conflict as a complex phenomenon with highly non-linear and interactive attributes which makes it hard to model with the available statistical modelling techniques. Neural networks, on the other hand, are massively interactive and highly nonlinear models [Haykin 1999]. This means, neural networks have a better-fit characteristic to model interstate conflicts than the linear or logit statistical models [Zeng 1999]. Neural networks were applied to model international conflicts by [Schrodt 1991; Beck *et al.* 2000]. In line with the previous studies, Lagazio and Russett [2003]; Marwala and Lagazio [2004] have also built neural network models which gave better results in predicting MID.

Beck *et al.* [2000] discuss what they believe the problems quantitative interstate conflict studies are faced with. They emphasise the point which

is overlooked by researchers that the effects of the causes of conflict differ by dyads. They describe the effects as being small for the most majority of the dyads and very big for a very small dyads which are more prone to conflict. They then suggest a need for an appropriate model to be introduced to deal with the problem. They then introduced neural networks as a generalisation of the commonly used logit models. Their results show that neural networks were far better than logit in learning the underlying structure of conflict data.

Another empirical study of interstate conflict using neural networks is given by [Lagazio and Russett 2003]. In their paper they point out that the various factors that determine MID are highly interactive and influence each other. They give an example of the reciprocal relationships that exist between democracy and interdependence and also among the other variables that makes it more unrealistic to consider individual causal relationships in isolation. They add more reasons why they favour using neural networks because neural networks do not require independent observations and they are flexible enough to discover the undefined causal interactions by themselves.

Marwala and Lagazio [2004] used neural networks that employ Bayesian framework Neal [1992]; Bishop [1995] to model interstate conflict. The training was done using evidence framework based on Gaussian approximation and Monte Carlo methods. The liberal variables (x) were mapped into the MID (y) using multi-layer perceptron (MLP) and supervised learning. A Bayesian framework was used to identify the weights and biases using the function:

$$P(wD) = \frac{1}{Z_s} exp(\beta \sum_{n}^{N} \sum_{k}^{K} \{t_{nk} ln(y_{nk}) + (1 - t_{nk}) ln(1 - y_{nk})\} - \sum_{j} w \frac{\alpha_j}{2} w_j^2) \tag{2.21}$$

where

$$Z_s(\alpha, \beta) = (\frac{2\pi}{\beta})^{\frac{N}{2}} + (\frac{2\pi}{\alpha})^{\frac{w}{2}} \tag{2.22}$$

n is the index for the training pattern, hyperparameter β is the data contribution to the error, k is the index for the output units, t_{nk} is the target output corresponding to the n^{th} training pattern and k^{th} output unit and y_{nk} is the corresponding predicted output. The parameter α_j is hyperparameter and it determines the relative contribution of the regularisation term on the training error.

2.9 Support vector machines for MID prediction

Support vector machines, to our best knowledge, have never been employed before for the purpose of militarised interstate conflict modelling. They have proved themselves to be very good techniques for modelling and classification in so many real world applications like text texture verification Chen and Odobez [2002] and option pricing Pires and Marwala [2004] to mention but a few. We are introducing them for the purpose of interstate conflict modelling.

2.10 Summary

International conflicts have the capacity of putting the involved parties into serious risk. Modelling and predicting interstate conflicts are then quite crucial as a decision support tool. Knowing what the major causes of disputes and if possible, ranking them in terms of their importance with regards to avoiding conflicts is an area of research which is demanding a lot of effort. A successful outcome of the study will enable governments and international relations policy makers to make better decisions.

With regards to the above mentioned goals, to better understand international conflicts and wars, political science and international relations scholars have and are still doing quantitative analysis of interstate conflicts. The quantitative research focuses on three main areas which include improving the data collection, improving the available statistical methods and coming up with new quantitative analysis techniques [Beck *et al.* 2000].

In recent years, interstate conflict analysis have been mainly done based on a dyad-year unit of analysis Ray [2003]. Using dyad-year as a unit of analysis means that all the variables that are believed to have an effect on interstate conflict or war are recorded for all pairs of states. These list of causal variables used in one study usually vary from other studies because the list is still debatable or there is no complete list of the variables available as yet. One such effort with the aim of defining and recording the possible dyadic variable is the Correlates of War project [COW 2004]. Likewise, the techniques and approaches in use for the quantitative anal-

ysis of interstate conflict vary from researcher to researcher.

The most common quantitative analysis techniques for classification of MID in use are multivariate statistical techniques. Some of these include logit and probit regression analysis methods. But these techniques usually put some *a priori* assumption about the density distribution of the data. They also consider the underlying functions to be expressed in a linear relationship with the variables. As Beck *et al.* [2000] point out, these techniques also assume the effect of the dyadic variables is the same for every dyad. Beck *et al.* [2000] argue that the effects of each variable are different for each dyad and are small for the majority of the dyads while very big for a small set of dyads. Besides, the relationship between the variables are highly nonlinear, massively interactive and vary for each specific conflict. That is why they say, that the results of previous studies are inconsistent and differ for every study. Therefore, there is a need for other techniques which overcome the above shortcomings.

Artificial intelligence is a collection of recent techniques that have proved to be effective modelling tools for complex and nonlinear real world problems. Neural networks and support vector machines are two such examples which have been effectively used for modelling and forecasting purposes. Neural networks have been used in previous studies to model interstate conflicts and have given better results than other previous statistical techniques. This research employs support vector machines for the first time to model interstate conflicts and compare the results with those of neural networks. The research methodology is discussed in the next chapter.

Chapter 3

Research Method

3.1 Introduction

Militarised interstate conflict, as mentioned in the previous chapters, is a state that occurs between countries based on their previous interactions. These interactions are represented in the explanatory variables that determine the outcome of the MID. It has also been pointed out that the explanatory variables are interdependent which usually have a loop like relationships. Modelling these interrelationships using linear modelling techniques gives non-consistent and unreliable results which are hard to interpret without having prior domain knowledge.

The previous chapter looked at various efforts that have and still are underway to make quantitative analysis of interstate conflict possible. These efforts range from collecting relevant data about the explanatory variables, improving on the existing empirical and quantitative techniques and venturing into new quantitative methodologies. The most

common of these empirical analysis methodologies are statistical multivariate and regression analysis techniques. Unfortunately these statistical methods have shortcomings which make them give unsatisfactory results for classification or forecasting of militarised interstate conflicts.

In order to remedy the weakness of the statistical methodologies, artificial neural techniques have been introduced recently into this challenging field of study. Neural networks have been used in many other scientific and engineering studies like various types of pattern recognition applications, fault detection in an engineering process, speech recognition, signature verification, financial forecasting and so forth and have delivered satisfactory results. NNs have given much better results as compared to previously used statistical techniques for MID modelling [Beck *et al.* 2000; Lagazio and Russett 2003; Marwala and Lagazio 2004]. This research is an extension of previous studies on neural networks with the introduction of another artificial intelligence technique, support vector machine, for MID modelling. Neural networks are compared to support vector machines.

This chapter discusses the research method in much detail. Section 2 describes the approach that was followed in using neural networks for MID classification. In Section 3 the approach of the support vector machines including the model selection process is presented. The description and the source of the MID data, how the training and testing data sets are generated and the preprocessing for the data are described in Section 5 and Section 6 concludes the chapter by giving the summary of the methodology.

3.2 Neural networks for MID modelling

As pointed out in the previous chapter, there are different varieties of neural network models. The most common are the feed forward and radial basis function (RBF) neural networks. The feed forward neural network is also classified into single and multi-layer perceptron. Multi-layer perceptron (MLP) has proven itself to approximate any function to arbitrary accuracy provided that it has sufficient enough number of hidden units [Jordan and Bishop 1996]. MLP is used in this study for modelling militarised interstate conflicts.

In order to have a good understanding of the MLP, it is worth describing and elaborating on its main components. MLP is a network that is comprised of simple processing units grouped into layers. A two layer MLP with an input, one hidden and output layer is used for this study. The input layer has 7 nodes to represent the seven explanatory variables. The output layer has one node to represent the MID outcome which can either be peace or conflict. The number of hidden nodes should be selected in such a way that the network model can approximate the underlying input-output relationship with good accuracy. Too few hidden nodes imply inadequate room for flexibility of the model. Similarly, too many hidden nodes result in an over sensitive model that tends to pick unnecessary details as opposed to the basic input-output relationships. Choosing the best number of the hidden units node is done using a model selection technique discussed in the following subsection.

All the seven processing units of the input layer are connected to all nodes

of the hidden layer, and those of the hidden layer to the node of the output layer. An adjustable weight is associated with each connection which determines the behaviour of the network after training. The activation function commonly used at the hidden layer of MLP is the *sigmoid* which is given by:

$$f(x) = \frac{1}{1 + exp(-x)} \tag{3.1}$$

A variety of activation functions are available for use for the output layer. Their choice varies according to which one performs better for a specific problem. Like the number of hidden units, the best choice of the activation function is done based on the outcome of the model selection stage.

3.2.1 Model selection for NN

As is the case for any modelling technique, neural networks require selecting the best model to give good classification results. In our context, model selection means searching the best combination of parameters from the set of possible parameters in order to construct an optimal neural network architecture for the classification of militarised interstate dispute data. The variables that need to be tuned include the number of hidden units, the activation functions for the output units, the training algorithm and the number of cycles for the network to be trained. Small numbers of training cycles can result in an under-trained network, while training the network for too many cycles might result in an over-trained network. Similarly, very few hidden units can result in a network not

flexible enough to pick the input-output relationship of test data while a network with too many hidden units can give a network that classifies the training data perfectly and performs badly on the test data. Both these cases should be avoided in order for the network to be able to generalise the unseen test cases as best as possible.

A multi-layer perceptron (MLP) trained with scaled conjugate gradient method Møller [1993] was used in this study. Logistic and hyperbolic activation functions for the output and hidden layer respectively and an $M = 10$ of hidden units resulted in an optimal architecture.

3.3 Support vector machines for MID modelling

Given a set of input-output MID patterns of the form $(\mathbf{x}, y)$, $\mathbf{x} \in \mathbb{R}^n$ and $y \in \{1, -1\}$, where the $\mathbf{x}$ are the explanatory variables and y the MID outcome, support vector machine for binary classification is a classifier that classifies the $\mathbf{x}$ explanatory input variables into two classes. That is, it seeks to estimate a function $f : \mathcal{X} \to \{\pm 1\}$. In so doing, a separating hyperplane with optimum margin of separation is searched for from all the classes of separating hyperplanes. SVM employs a method of mapping the input space into a feature space $\mathcal{H}$ of higher dimensionality and then finds a linear separating hyperplane with maximum margin of separation.

For the militarised interstate conflict data, it is assumed that there is an underlying function that governs the relationship between the explanatory variables and the MID result. In other words, given a set of

explanatory variables of a dyad in a particular year, the support vector machine would classify the MID outcome of the following year as either peace or conflict. The support vector classifier learns the underlying interrelationship among the various variables from the training data so that it can generalise previously unseen test data.

Support vector machines usually employ a similarity measure $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$, a function that quantifies the similarity of two input vectors as a real number [Schölkopf and Smola 2003]. This similarity measure is usually calculated as a dot product of the vectors. Since there is no clear assumption that the input patterns live in a dot product space, the need for mapping into a feature space arises. This mapping into a feature space is done using kernel functions.

There are different kernel functions available for use, the most common of which are linear, polynomial, radial basis function (RBF) and sigmoid. As some studies Hsu *et al.* [2004] suggest, RBF: $K(x_i, x_j) = exp(-\gamma\|x_i - x_j\|^2), \gamma > 0$ can handle non-linear data better than the linear kernel function. The polynomial kernel has a number of hyper-parameters which influence the complexity of the model and has more mathematical difficulty than RBF, sometimes its values may become infinity or zero as the degree becomes large. Due to these reasons, Hsu *et al.* [2004] suggest, RBF kernel function is the best choice for common use. RBF was employed for this study.

3.3.1 Model selection for SVM

Similar to the discussion of neural networks above, SVM also requires selection of a model that gives an optimal result. Our experiment shows that RBF gives best results for the classification of MID data with a better efficiency. Efficiency is the time it takes to train and test the data. Model selection in this context means selecting parameters that give the best results for the test data. When RBF is used as the kernel function, there are two parameters which influence the results far greater than the other parameters. These are the penalty parameter of the error term C and the γ parameter of the RBF kernel function. This means that they have to be adjusted to give the best results. In so doing, two straight-forward methods, cross-validation and grid-search, are used and their description follows.

Cross-validation

The end goal in training a learning machine is to be able to create a machine that can generalise an unknown input data based on the knowledge it acquired from the training. That is, the aim is not minimising the training error rather it is minimising the generalisation error. Therefore, a set of parameters that give good results for the training data does not necessarily mean it gives good results for the test data. Cross-validation is a way of going around this problem by first classifying the training data into two sets. One set is used as a training while the other as a test set so that the machine can learn previously unseen data better. Taking

further this notion of dividing the training set into k equal subsets and using each one of them as a test while training the SVM using the rest is known as k-fold cross-validation [Hsu *et al.* 2004]. This procedure helps to avoid a common problem of over-fitting.

Grid-search

As it was mentioned earlier, support vector machines use various parameters which need to be adjusted depending on the choice of the user. Some of the parameters include the type of kernel function, epsilon (tolerance for termination), C cost, k-fold cross-validation and so forth. There should be a way of selecting which combination of these parameters gives the best results. Grid-search is a simple search for a set of parameters and pick the set with the best cross-validation accuracy [Hsu *et al.* 2004].

For the experiment, a 10 fold cross-validation with a simple grid-search for the penalty coefficient C and the variable γ of the RBF kernel function technique is used. The pair of parameters with the best accuracy for the training data is then picked up. The training is again done using those parameters for the whole training data set. The trained SVM is then used for the classification of the test data.

3.4 MID data sets

The data sets which are used for this study came from the Correlates of War (COW) and Polity data set compiled by Russett and Oneal Russett

and Oneal [2001] and was used by [Oneal and Russett 2001; Lagazio and Russett 2003; COW 2004]. It includes politically relevant dyads for the cold war and immediate post-cold war period (CW), from 1946 to 1992. Politically relevant population refers to all dyads which are contiguous and which contain major power Marwala and Lagazio [2004]; Lagazio and Russett [2003]; Oneal and Russett [2001 1999b]. That is, dyads that share any kind of border (land or water) as is defined in the Correlates of War [COW 2004]. The reason for omitting distant and weak dyads is because it is less probable they would exert to policy intervention in each other and so get into conflict. This choice of relevant variables reduces the prediction power of the omitted variables but on the other hand it makes the prediction to be more interesting [Lagazio and Russett 2003].

The unit of analysis for the study is dyad-year. Since the aim of this research as its previous studies Lagazio and Russett [2003]; Marwala and Lagazio [2004] is to predict the onset of conflict rather than its continuation, the dyads include only those with no disputes or only the initial year of the militarised conflict. After the omission, a total dyad cases of 27,737 with 26,845 peace dyad-years and 892 conflict dyad-years were filtered out.

The dyadic data is classified into two sets which are the training and testing sets. In their study Lagazio and Russett [2003] have given a robust discussion on how the training set should be chosen. They have found out that a balanced set, equal number of conflict and peace dyads, gives best results as a training set for the neural network. The training set contains 1000 randomly chosen dyads, 500 from each group. The test

set contains 26737 dyads of which 392 are conflict and 26345 non-conflict dyads.

3.4.1 Rare-event prediction problems

Some of the problems that face statistical analysis (eg. linear regression models) for rare-event data include underestimating their probability and poor data collection strategies associated with them [King and Zeng 2000]. Militarised interstate conflicts are such kind of data with conflict cases being very small in number as compared to number of peace cases. Even after only relevant dyads are selected, the proportion of peace to that of conflict is very large. Hence inferences made from such data sets can be biased unless careful strategy is taken to address the problem.

Training neural networks using the above mentioned data sets with very huge difference in the occurrences of the cases has its own problems. Neural networks training process involves adjusting weights of the network based on the training samples. This means that the weight estimates tend to represent the commonly encountered (modal) values [Garson 1998]. Therefore, there is need for some reasonable strategies to address this problem and improve neural networks prediction ability for rare events [Lagazio and Russett 2003].

Based on the suggestion of King and Zeng [2000] for logistic regression which employs the idea of selecting data based on the dependent variable and using statistical correction to avoid the selection bias, Lagazio and Russett [2003] have adapted it for neural networks. In their study, they

train the neural network using a balanced data from both classes and then use an unbalanced cross-validation set to correct any possible selection biases. The same principle of balanced training was adopted for this study.

3.5 Data preprocessing

Preprocessing is transforming the data into a format that is workable with the classifier and improves the overall efficiency and quality of results. Examples of preprocessing include, representing the features as vector of real numbers and scaling the data [Hsu *et al.* 2004]. Since support vector machines treat all the inputs as real numbers vectors, any categorical features should be transformed into real number vector format. Scaling is done with the aim of avoiding dominance of features with large numeric range over those with small range and also to avoid numerical difficulties during calculation [Hsu *et al.* 2004].

For the MID data, the only preprocessing required is scaling because all the variables are already represented as real numbers and can be used straight for neural network or support vector machine classifiers. Some of the explanatory variables fall into a very small range (eg. 0 to 0.1719 in the case of *dependency*) while others have much bigger range (eg. -10 to 10 for *democracy level*). The training and test data has been scaled using different scaling ranges like [0,1], [-1,1], [-2,2]...[-10,10] and picking the one that delivers the best results. In this case, the range [-2,2] was the best and it was used for this study.

3.6 Summary

This chapter discusses the research methodology used for the research. Neural networks have been employed in previous studies to model interstate conflicts. For this study, a neural network of type multi layer perceptron (MLP) with seven input dyadic variables, ten hidden units and one output node that represents the MID outcome which is either peace or conflict was employed. The activation functions for the hidden and output layers are sigmoid and logistic, respectively. The network is trained with one variant of back-propagation training algorithm which the scaled conjugate gradient.

The second artificial intelligence technique that was employed to model MID is support vector machine for classification. This technique maps the input data into a higher dimensional feature space and finds a separating hyperplane with optimal margin of separation. Radial basis function (RBF) is used as a kernel function to calculate a similarity measure which is the criteria for classification. When RBF is used as a kernel function, there are two parameters that need to be adjusted which are γ and the regularisation factor C.

As any other models, the performance of both neural networks and support vector machines depend on how the model was set up and trained. Good performance of the training set does not necessarily imply good prediction results for the test data set. Hence, there is a need for a method that enables to choose the best combination of parameters for the learning machine to generalise previously unseen data. This process

of choosing the best parameters is called model selection. The two model selection techniques that were used in our study are cross-validation and grid-search. Cross-validation is a way of breaking down the training set into a number of equal subsets and iteratively using one subset as a test set while the rest as training. Grid-search is a simple search for a combination of parameters that give good results for the training with cross-validation.

Preprocessing is a way of adjusting the input/output data so that it can be readily used with the NN or SVM tools and also enhances the tool's performance. This include coding non-numeric features as real numbers and normalising or scaling the data sets with the aim of avoiding dominance by those variables that have bigger ranges of values. For the purpose of this study all the variables have been scaled so they fall within the same range. A range between -2 and 2 gave the best results for the support vector machine. The results and their respective discussions of the experiments are given in the next chapter.

Chapter 4

Results and Discussions

4.1 Introduction

The previous chapter looks at how the experiments were done for both NN and SVM. This includes how both methodologies address the classification problem, the various parameters involved, the optimisations algorithms involved, how to select best models, the data inputs and the preprocessing required to make them give the best results. This chapters gives the results of the experiments that were conducted based on the previous chapter's methodologies.

The main results include how the two techniques perform in classifying the militarised interstate conflict data. That is, the true and false predictions for both peace and conflict is given in a table. Besides, the receiver operating characteristic (ROC) curves together with the areas under the curve (AUC) and their respective standard deviation values for both NN and SVM are given.

Sensitivity analysis of the explanatory variables are also other results presented in this chapter. The sensitivity analysis in this context entails understanding how each variable influences the MID outcome. For the NN this is done by looking at the MID outcome when one explanatory variable is changed. The chapter is organised as follows. Section 2 gives the results and the discussion of the classification of both NN and SVM. The receiver operating characteristic curve analysis results and their discussion are given in Section 3. Section 4 provides the results of the sensitivity analysis conducted for both NN and SVM.

4.2 Prediction results

Neural networks and support vector machine were employed to classify the MID data. Once each technique was trained with the balanced training data set, it was tested with the test data. The test set represents the overall reality of the data in which the occurrence of conflict is very rare as compared to that of peace. The main focus of the result is to look at the percentage of correct MID prediction of the test data set by each technique.

When talking about the correct prediction of peace and conflict, there is a crucial point to be addressed. The MID data set is a data that is biased towards peace. This means that even a classifier without any classification ability can give 98% correct results even though all the data is classified as peace. Therefore, the goodness of the classifier should be determined based on how well it can classify both cases correctly. It has

to look at both the correct prediction of peace and conflict. Table 4.1 depicts the confusion matrix of the results.

Although NN performed as good as SVM in predicting true conflicts (true positives), this is achieved at the expense of reducing the number of correct peace prediction (true negatives). SVM picked up the true conflicts (true positives) better than NN without effectively minimising the number of true peace (true negatives). That is, SVM is able to predict peace and conflict with accuracies of 79% and 75%, respectively. The corresponding results for NN are 74% for peace and 76% for conflict. The combined results are 79% and 74% for SVM and NN, respectively.

Table 4.1: NN and SVM classification results

Method	**TC**	**FP**	**TP**	**FC**
Neural Network	297	95	19464	6881
Support Vector Machine	295	97	20914	5431

TC = true conflict (true positive), FC=false conflict (false positive), TP=true peace (true negative), and FP = false peace (false negative)

4.3 Receiver operating characteristic (ROC) curve

Receiver operating characteristic graphs have long been used in signal detection theory to show the hit rates against false alarm rates Egan [1975] and as clinical diagnosis tools [Zweig and Campbell 1993]. Recently, they have been extended to the use of comparing the prediction ability of binary classifiers [Provost and Fawcett 1997]. The ROC curve is calculated based on the *sensitivity* and *specificity* of the classifier.

In the context of our MID classifiers, *sensitivity* is defined as the probability of a classifier predicting conflict correctly. It is also referred as *true positive rate*. *Specificity*, on the other hand, is the probability of a classifier predicting peace correctly [Westin 2001]. That is,

$$\begin{aligned} sensitivity &= TC/(TC+FP) \quad and \\ specificity &= TP/(FP+TP) \end{aligned} \tag{4.1}$$

where TC, FP and TP represent true conflict, false peace and true peace, respectively. *False positive rate*, the percentage of peace incorrectly classified as conflict, is then calculated as 1-*specificity*. The ROC curve analysis takes into consideration the number of true conflicts (*true positive rate*) and false peace (*false positive rate*) to determine the goodness of a classifier.

The ROC curve is a graph that plots the *sensitivity* on the vertical-axis and 1-*specificity* on the horizontal-axis. The area under the curve (AUC) is used as a measure to compare the performance of each classifier. The AUC for NN and SVM found are 0.81 and 0.84 with standard errors of 0.00998 and 0.01022, respectively. According to Hanley and McNeil [1983], the normal distribution z value which is used to compare if there is a significant difference between AUCs of two classifiers that are derived from the same cases is given by:

$$z = \frac{A_1 - A_2}{\sqrt{SE_1^2 + SE_2^2 - 2rSE_1SE2}} \tag{4.2}$$

where A_1, A_2, SE_2 and SE_2 are the areas and standard errors of the

respective curves. The value r represents the estimated correlation between A_1 and A_2 [Hanley and McNeil 1983]. The value of z is 2.697 which gives significant difference in a 95% confidence interval. The results of the SVM are much better in predicting the conflicts without affecting the prediction of peace as it is clearly shown in figure 4.1. The ROC graphs of the NN and SVM results are given in figure 4.1.

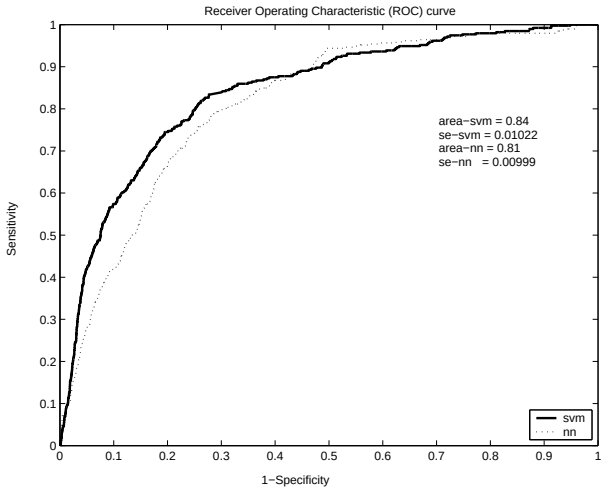


Figure 4.1: ROC curve for both NN and SVM. Area-svm and area-nn signify the areas under the curves while se-svm and se-nn are their respective standard errors.

.

4.4 Goodness-of-prediction of a classifier and ROC curve

Comparing the goodness-of-prediction of classifiers is not a straight forward task. Prediction results in the context of MID can be summarised in four values which are the number of true peace, false peace, true conflict and false conflict. The goal of the two classifiers NN and SVM is to increase the number of true conflicts and peace while reducing the

number of false peace and conflicts. Both NN and SVM are similar in picking up the true conflicts. But SVM was far better in picking the true conflicts without effectively reducing the number of correctly predicted peace cases. SVM is able to pick 1450 more cases of true peace than NN. This implies that any policy measures that might be taken to address the conflicts based on the results of NN would mean the allocation of unnecessary extra resources for those 1450 cases.

Receiver operating characteristic (ROC) curve is a technique that takes true peace, false peace, true conflict and false conflict values and represents it in the form of a graph. The area under the curve is used as a measure for the goodness-of-prediction of the classifier. *Area* of 1 means perfect classification results and $area = 0.5$ implies the classifier has no classification ability at all. The comparison of the classifiers is made possible using the AUCs. Hanley and McNeil [1983] have developed a method for comparing ROC curves that are derived from the same data cases. They employed a correspondence between the AUC and the Wilcoxon statistic and the underlying Gaussian distributions (binomial) based on the observed ratings to find a correlation between the AUCs. Lack of significant difference between the two areas does not necessarily imply absence of significant difference between goodness of the classifiers [Westin 2001]. It only means that it is not possible to state that they are significantly different. Significant difference, on the other hand, imply that there is a difference on the goodness-of-prediction of the classifiers. The two AUCs for SVM and NN are found to be significantly different. This is because SVM gives better results in classifying the MID data than

NN.

4.5 Sensitivity analysis

Sensitivity analysis in the context of the discussion means finding out which of the explanatory variables affect the MID outcome the most. As it was mentioned in the background chapter, there are quite a number of different studies that looked at this issue of sensitivity analysis. It is not easy though to say which variables play a significant role in determining the MID outcome based on their results. Even the results of the same study can differ significantly if minor assumptions or changes are made to the way the analysis is done. One good example that confirms this fact is the study done by [Bremer 1992]. Addition of one explanatory variable to their study caused a fundamental change in the ranking of the causal impact.

In the discussion of sensitivity analysis of the MID, keeping in mind of the above mentioned facts helps on how to make use of the rankings of a particular study. Two separate experiments were done to see the causal effects of the explanatory variables on the MID outcome for both NN and SVM. The two techniques agree in picking up the influences of some of the variables while they differ on others. The discussion of the results follows.

4.5.1 Experiment one

This experiment looked at how assigning each variable to its possible maximum value while keeping the rest at their possible minimum values and vise versa affect the MID outcome. The approach makes it possible to see if the effect of each variable is strong enough to reverse the outcome when all others are on the opposite extreme. The results for NN show that only *democracy level* and *capability ratio* are able to deliver a peaceful outcome while all the other variables are kept minimal. This means, dyadic preponderance has a deterring effect on conflict as is the joint democracy level of the states involved. On the other hand, keeping all the variables to their maximum values while assigning one variable to its minimum value resulted in a peaceful outcome. In other words, no single variable is able to change the outcome if all the other variables are set to their possible maximum values for NN.

Although the experiments were not done in exactly the same way, Maoz and Russett [1993]; Oneal and Russett [1997] ranked *democracy* to be the most influential variable in affecting the outcome of conflict. In other studies, Russett and Oneal [2001]; Reed [2000], *Capability* was ranked as first. Therefore, it can be said that for the experiment to pick these two variables as the only ones that are able to change the outcome agrees with previous studies that ranked the two variables on top of the list as the most influential variables in the respective studies.

A similar experiment conducted for SVM shows that it is not able to pick the influence of a variable as it is possible with NN. The results were the

same in both cases. This is because it was found that setting variables to their minimums or maximums always gives a peace outcome.

4.5.2 Experiment two

This sensitivity analysis is similar to the first one except that when one variable is assigned to its possible maximum / minimum values the remaining variables are kept fixed. The experiment was done to measure the sensitivity of the variables in the spirit of partial derivatives as Zeng [1999] puts it. The idea is basically to see the change in the output for a small change in one of the input variables. The experiment looks at how the MID varies when one variable is assigned to its possible maximum and minimum values while keeping all the other variables constant. The results found for both NN and SVM are shown in table 4.2. The test data set has 26737 cases of peace and 392 cases of war. The first line of the table shows the correct number of peace and war prediction when all variables are used. Different testing data sets were then generated by assigning each variable to its possible maximum and minimum values while keeping the other variables fixed. Each subsequent line of the table shows the number of correct prediction for peace and war.

The results for NN are consistent and much more easier to interpret than those of SVM. The trend of the change in the outcome was in either direction depending on the change in the variable. If the variable has a positive relationship with peace, then maximising its value increases the number of peace outcomes. Conversely, if the variable has a negative

relationship with peace, maximising the variable results in minimising the number of peaceful outcomes. Hence, it can be said that this sensitivity analysis better fits for NN.

The ranking of the variables according to their causal effect on the outcome is given in table 4.3. In conformance to the previous experiment, both *democracy* and *capability* are ranked at the top. As mentioned above, the acquired results agree with previous studies Maoz and Russett [1993]; Oneal and Russett [1999b]; Mousseau [2000] to rank *democracy* as number one. *Capability* was ranked as number one in [Reed 2000; Russett and Oneal 2001]. The NN sensitivity analysis is able to pick the two variables ranked higher in other studies and their effects are 100% and 98%, respectively. *Contiguity, distance and alliance* are the next three variables in the ranking with their effect as 45%, 31% and 20%, respectively. The remaining two variables with the least effect on the MID onset are *dependency* and *major power* with their respective influence being at 6% and 3%.

The results of the same experiment for SVM is not as consistent as that of NN. The effect of changing a variable is not consistently reflected on the MID outcome. Both maximising and minimising a variable can reduce the number of conflicts and peace at the same time and vise versa. The observation is that the experiment does not give meaningful results or the results are hard to interpret. Our suggestion is that further investigation is required to come up with sensitivity analysis that better fits SVM and can give results that can be interpreted easily. In light of this idea, an alternative sensitivity analysis was done for SVM and this is discussed in

the next section.

Table 4.2: The effect of changing one variable while keeping the other variables fixed

	NN		SVM	
Variable	**Peace**	**War**	**Peace**	**War**
Test set results	19464	297	20914	295
Dem-min	16263	325	22327	205
Dem-max	26345	-	23761	35
Allies-min	18555	313	20469	274
Allies-max	21034	237	21999	153
Contig-min	23682	164	24745	60
Contig-max	12463	342	18939	281
Dist-min	5351	370	25067	34
Dist-max	22525	206	26284	3
Capab-min	6929	373	19840	180
Capab-max	26322	3	26345	-
Depnd-min	19455	297	20498	305
Depnd-max	20411	277	26345	-
Majpow-min	19686	289	22345	-
Majpow-max	19428	299	23583	136

NN result: It shows *democracy level* has the maximum effect in reducing conflict while *capability ratio* is second in conformance to the first experiment. Allowing *democracy* to have its possible maximum value for the whole data set was able to avoid conflict totally. *Capability ratio* reduced the occurrence of conflict by 98%. Maximising *alliance* between the dyads reduced the number of conflicts by 20%. Maximising *dependency* has a 6% effect in reducing possible conflicts. Reducing *major power* was able to cut the number of conflicts by 3%. Minimising the *contiguity* of the dyads to their possible lower values and maximising the *distance* reduced the number of conflicts by 45% and 31% respectively.

SVM result: The results of the experiment show inconsistency on how the MID outcome is affected when the variables are maximised and minimised. Further investigation is required to understand more clearly the influence of each variable (eg. exploring some other sensitivity analysis technique). Therefore, a sensitivity analysis that involves using only one explanatory variable to predict the MID and see the goodness of the accuracy is used. This means, training and test data sets with only one variable at a time were generated and the experiment conducted. The ROC curves were drawn and the area under the curve (AUC) calculated for the purpose of ranking [Guyon and Elisseeff 2003]. The variables were then ranked in descending order of their respective AUC. The ranking of the effects of variables on the MID by NN and SVM is given in table 4.3.

Table 4.3: Rankings of the influence of variables

Rank	**NN**	**SVM**
1	Democracy	Contiguity
2	Capability	Alliance
3	Contiguity	Dependency
4	Distance	Democracy
5	Alliance	Distance
6	Dependency	Capability
7	Major power	Major power

4.6 Alternative sensitivity analysis for SVM

This sensitivity analysis as is described in Guyon and Elisseeff [2003] involves ranking the influence of the variables based on the performance of

the classifiers built with a single variable. The SVM classifier was built with only one explanatory variable at a time and then their goodness-of-classification compared. The comparison is done based on the area under the curve (AUC) of the receiver operating characteristic (ROC) curves of the variables. The rankings are depicted in table 4.3. The respective values of the AUC for *contiguity, distance, major power, capability, democracy, dependency and alliance* are 0.728, 0.681, 0.679, 0.651, 0.567, 0.530 and 0.513.

The rankings of the SVM show significant difference compared to those of the NN. Both *contiguity* and *distance* are ranked on the top. Previous study that has a similar result is Bremer [1992] which ranked *proximity* as the most influential factor. One of the reasons for such a difference may be because of the single variable classifiers used in SVM. The explanatory variables are believed to be highly interdependent [Beck *et al.* 2000]. Using a single variable might not make it possible to measure the effect of the variables have toward each other. In summary, what can be said from the sensitivity analysis results of the SVM is that it is not easy to take them as they are or they are not easy to interpret. The results can be taken as triggers for further sensitivity analysis.

4.7 Summary

The main focus of this chapter is to present the results of the experiments done and discuss them. First of all, the prediction results of both SVM and NN is given in a table which depicts the number of correctly predicted

peace (true negatives), correct conflict (true positives), false peace (false negatives) and false conflict (false positives). Fist it looks at how both techniques performed in classifying the MID data with the help of receiver operating characteristic curve and the area under the curve measurement. The AUC of the ROC curve is used to measure the predictive power of classifiers. This means that a standard normal distribution z value is calculated as in Hanley and McNeil [1983] to see if the two AUCs differ significantly. The results show the probability that the difference is accounted to randomness is less than 5%. That is to say the two classifiers differ significantly in a 95% confidence interval. The AUC of the SVM is significantly better than that of the NN which means SVM performed better than NN in classifying the MID data.

Sensitivity analysis of the explanatory variables was done for both NN and SVM. The sensitivity analysis looks at how the MID outcome is affected for an amount of change in the explanatory variables. Some of the questions the experiments try to answer include, how does assigning the variables to their possible maximum or minimum values affect the outcome. What happens if one variable has its maximum or minimum value while all the rest are kept at their respective minimum or maximum values? How about if one variable becomes maximum or minimum while all the rest are kept constant? Two separate experiments were done to answer these questions.

Experiment One looks at how a variable affects the MID outcome when its value becomes maximum or minimum while keeping the other variables at their respective minimum or maximum values. The results of the

NN show that only *democracy level* and *capability* are able to affect the outcome to be peace when they are assigned to their maximum values while the other variables are kept at their possible minimum values. On the other hand, neither variable was able to change the outcome to be conflict when all the other variables are kept at their maximum values. Unlike the NN, similar experiments done on SVM could not pick the difference. The results become peace whether the variables are maximum or minimum.

The second experiment keeps the other variables fixed when one variables is assigned to its maximum or minimum values. This experiment makes use of the idea of partial derivatives which involves looking at how the outcome is affected for a change in the explanatory variables. Similar to the previous experiment, the NN gives consistent results on how each variable affects the MID outcome. The results of the two experiments show that NN was able to give a more consistent and easily interpretable results than the results of SVM.

Similar experiments conducted on SVM show a significant difference with the results of NN. The results of the SVM did not show consistency and another sensitivity analysis was conducted as a remedy. Although further investigation is required with regard to the sensitivity analysis of SVM, the observation made is that the experiments which focus on monitoring the change in the MID outcome for a small change in an explanatory variable do not give consistent and easily interpretable results. An alternative sensitivity analysis looks at how each variable affects the MID on a one-to-one basis. That is, how does each variable affect the goodness

of accuracy when it is considered separately and the AUC of the ROC curves is used to rank the effect of the variables. The alternative sensitivity analysis that was done for SVM still shows significant difference in the rankings obtained by the NN.

Chapter 5

Conclusions and Recommendations

5.1 Conclusion

The history of mankind has witnessed various types of conflicts and wars between or among states. Although states may go to war for justifiable or unjustifiable reasons, it can be said that there is some kind of interest involved when states have disputes. Political science and international study intellectuals have and are still trying to understand the main reasons why states have disputes or go to war. One of the major studies include identifying the main explanatory or determinant factors of conflicts.

Despite the fact that there are some commonly used variables like *distance, contiguity, democracy, alliance, capability ratio, interdependence etc*, different studies add their own set of explanatory variables for their studies. They then quantify the variables in terms of real numbers so that they can be analysed quantitatively. Once the variables are quantified,

various statistical methods that range from standard linear discriminant to logistic regression are used for the analysis. The studies try to rank the variables in order of their causal effect on the MID outcome.

According to Gochman and Maoz [1984], militarised interstate conflict (MID) is defined as a set of interactions between or among states that can result in the actual use, display or threat of using military force in an explicit way. These interactions between states within a specific year are quantified in the form of dyadic variables. The assumptions is that any militarised conflict between two states is the outcome of the interactions which they had in the previous dyad-year. The Correlates of war project COW [2004] is an ongoing effort that focuses on identifying and defining the major explanatory variables for war or conflict and administering a repository of interstate data.

Despite the vast efforts of data collection and standard statistical analysis of interstate conflicts, the results found are far from satisfactory. Every quantitative study comes up with its own results that differ significantly from other studies and this makes the results difficult to interpret. As Beck *et al.* [2000] pointed out, the quality of a model is mainly determined by its out-of-sample forecasting ability, which is the model's ability to generalise. Previously used statistical techniques for quantitative analysis of interstate conflicts have a prior assumption that the underlying function exhibits a linear relationship. However, it is believed that relationships that govern interstate conflicts show very complex nonlinear interactions with great interdependence among the variables. That is the main reason for the variation in the quantitative analysis results besides

the quality of the data. The need for new techniques that can address the complexity of the problem domain becomes inevitable.

Artificial intelligence techniques have proven themselves to be excellent for problems involving pattern recognition. They have given promising results for modelling different real world problems like face recognition Heisele *et al.* [2001], speech recognition Fritsch [1996], fault detection in scientific processes Dietz *et al.* [1988], financial forecasting Pires and Marwala [2004] and others. Two artificial intelligence techniques, neural networks and support vector machines, are used to model interstate disputes in this study. Neural networks have already been applied for modelling interstate conflicts Schrodt [1991]; Beck *et al.* [2000]; Lagazio and Russett [2003]; Marwala and Lagazio [2004]. Support vector machines have never been applied for the use of international conflict studies.

There are different types of neural networks. The most commonly used include the feed-forward network and radial basis function (RBF) networks. The multi-layer perceptron (MLP) with two adjustable layers of weights with input, hidden and output layers is the dominant one. The input layer represents independent variables, the hidden layer latent variables and the output layer the dependent variables. The supervised learning process basically works by adjusting the connection weights based on an error minimisation technique so as the network may be able to classify new inputs with as much accuracy as possible. Feed-forward neural networks provide a framework to represent a non-linear functional mapping of a set of d input variables x_i, $i = 1, ..., d$ into a set of c output variables y_j, $j = 1, ..., c$ [Bishop 1995].

In this study, multi-layer perceptron (MLP) with scaled conjugate gradient method Møller [1993] training strategy is used to train the network. Logistic and hyperbolic activation functions for the output and hidden layer respectively and an $M = 10$ of hidden units resulted in an optimal architecture.

Given a set of input-output patterns, a support vector machine for binary classification is a classifier that classifies the input patterns into two classes. That is, it seeks an estimate function $f : \mathcal{X} \rightarrow \{\pm 1\}$. In so doing, a separating hyperplane with optimum margin of separation is searched for from all the classes of separating hyperplanes. Most of time, real world input data exhibit linearly inseparable behaviour, complex decision functions are required to classify them. Finding this non-linear and complex decision function makes the problem very hard to solve. SVM employs a method of mapping the input space into a feature space $\mathcal{F}$ of higher dimensionality and then finds a simple linear separating hyperplane with maximum margin of separation.

Our experiments confirm that SVM with RBF kernel function gives best result for the classification of MID data with a better efficiency. Model selection involves selecting parameters that give best results for the test data based on the training process. When RBF is used there are two major parameters which need be adjusted. These are the penalty parameter of the error term C and the γ parameter of the RBF kernel function. A simple grid-search with cross-validation techniques was used to select the best models.

The classification results show a better performance by SVM than NN. Although NN performed as good as SVM in predicting true conflicts (true positives), this is achieved at the expense of reducing the number of correct peace prediction (true negatives). SVM was able to predict peace and conflict with 79% and 75%, respectively. The corresponding results for NN are 74% for peace and 76% for conflict. The combined results are 79% and 74% for SVM and NN, respectively. Comparing SVM and NN using area under the curve of the receiver operating characteristic confirms that SVM outperforms NN with a 95% confidence of interval.

Sensitivity analysis was conducted for both NN and SVM to see how the explanatory variables influence the MID outcome. Ranking the variables according to the strength of their effect on the MID outcome helps policy makers to take the right decision in avoiding the consequence that may arise due to interstate conflicts. Sensitivity analysis that was conducted with the spirit of observing how a change in the input variables affects the MID outcome is more appropriate for NN than SVM. The results of NN show better consistency than those of SVM on how the MID outcome changes for every amount of change in an explanatory variable. The same sensitivity analysis done for SVM gave results that are not easy to interpret. An alternative sensitivity analysis conducted for SVM gives a significant difference in the ranking of the influence of the variables. Further investigation is required to look at other sensitivity analysis techniques for SVM.

5.2 Future work

This research to our best knowledge is the first one to employ support vector machine for modelling of international conflicts. There is still more work to be done especially in the sensitivity analysis which was not fully explored in this research. Besides, there is also the possibility of exploring some other classification techniques for MID studies. One such technique is decision tree.

5.2.1 Feature selection / sensitivity analysis

The sensitivity analysis that was conducted for SVM gave inconsistent results which are difficult to interpret. This implies that there is room for more work in this regard. Some of the areas of exploration include looking for other variable ranking and feature selection techniques. As it is described in Guyon and Elisseeff [2003], feature selection involves selecting the most determinant subset of variables from a set of explanatory variables. Variable ranking, on the other hand, focuses only on ranking the variables based on their significance of influence on the output.

Feature selection for the international conflict problem involves determining which subset of the dyadic variables are more significant in determining the MID outcome. Since the dyadic variables are interdependent, a variable which is considered as redundant on its own may give good results when it used with other variables. Basically the feature selection to be explored is finding out which combination of the dyadic variables

result in a better prediction of the MID outcome.

As it is discussed in Guyon and Elisseeff [2003], there are various variable subset selection methods which can be divided into wrappers, filters and embedded. The wrapper methodologies use the prediction performance of the learning machine to compare the variable subsets. For the MID variables, exhaustive search of the subsets can be performed since we have relatively small number of variables. In embedded methods the subset selection is done within the training process. This technique has been employed in decision trees like CART.

5.2.2 Decision trees

Decision trees are one type of learning machines which are used to approximate discrete-valued target functions [Mitchell 1997]. The internal nodes represent the input patterns while leaf nodes the output or classification category. Input patterns are filtered down the tree based on the values of their attributes. Once the tree is created during the training phase, it can be used as a classifier for previously unseen test patterns. Decision trees have been used for classification applications [Lin *et al.* 2003]. Using decision trees for the classification of dyadic data would be an interesting option for future work.

References

[Baldwin 1993] D. Baldwin. *Neorealism and Neoliberalism: The Contemporary Debate*. Columbia University Press, New York, 1993.

[Beck and Jackman 1998] N. Beck and S. Jackman. Beyond Linearity by Default: Generalised Additive Models. *American Journal of Political Science*, 42(2):596–627, 1998.

[Beck *et al.* 2000] N. Beck, G. King, and L. Zeng. Improving Quantitative Studies of International Conflict: A Conjecture. *American Political Science Review*, 94(1):21–33, March 2000.

[Beck *et al.* 2004] N. Beck, G. King, and L. Zeng. Theory Evidence in International Conflict: A Response to de Marchi, Gelpi, and Gryanaviski. *American Political Science Review*, 98(2):379–389, May 2004.

[Bishop 1995] C.M Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, Oxford, UK, 1995.

[Bremer 1992] S. Bremer. Dangerous Dyads: Conditions Affecting the LIkelihood of Interstate War, 1816-1965. *Journal of Conflict Resolution*, 36(2):309–341, 1992.

[Burges 1998] C. Burges. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.

[Chen and Odobez 2002] D. Chen and J. Odobez. Comparison of Support Vector Machine and Neural Network for Text Texture Verification. IDIAP-RR-02 19, IDIAP, Martigny, April 2002. ftp://ftp.idiap.ch/pub/reports/2002/rr-02-19.pdf.

[Copeland 1996] D. Copeland. Economic Interdependence and War: A Theory of Trade Expectations. *International security*, 20(4):5–41, 1996.

[COW 2004] COW. Correlates of War Poject, 2004. http://www.umich.edu/ cowproj/index.html
Last date accessed: Sept, 2004.

[DeMarchi *et al.* 2004] S. DeMarchi, C. Gelpi, and J. Grynaviski. Untangling Neural Nets. *American Political Science Review*, 98(2):371–378, May 2004.

[Dietz *et al.* 1988] W. Dietz, E. Kiech, and M. Ali. Pattern-based fault diagnosis using neural networks. In *IEA/AIE '88: Proceedings of the first international conference on Industrial and engineering applications of artificial intelligence and expert systems*, pages 13–23. ACM Press, 1988.

[Egan 1975] J. Egan. *Signal Detection Theory and ROC Analysis*. Academic Press, New York, 1975.

[Fritsch 1996] J. Fritsch. Modular Neural Networks for Speech Recognition. Master's thesis, Carnegie Mellon University & University of Karlsruhe, August 1996.

[Garson 1998] D. Garson. *Neural Networks: An Introductory Guide for Social Scientists*. Sage, London, 1998.

[Gochman and Maoz 1984] C. Gochman and Z. Maoz. Militarized Interstate Disputes 1816-1976. *Journal of Conflict Resolution*, 28(4):585–615, 1984.

[Gochman 1990b] C. Gochman. *Prisoners of war: Nation-states in the moder era*, chapter Capability-driven disputes. Lexington Books, Lexington, MA, 1990b.

[Guyon and Elisseeff 2003] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182, 2003.

[Hanley and McNeil 1983] J. Hanley and B. McNeil. A Method of Comparing the Areas under the Receiver Operating Characteristic Curves Derived from the Same Cases. *Radiology*, 148(3):839–843, September 1983.

[Haykin 1999] S. Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall International, Inc, Upper Saddle River, New Jersy 07458, second edition, 1999.

[Heisele *et al.* 2001] B. Heisele, P. Ho, and T. Poggio. Face recognition with support vector machines: Global versus component-based ap-

proach. In *8th International Conference on Computer Vision*, volume 2, pages 688–694, Vancouver, Canada, 2001.

[Hsu *et al.* 2004] C. Hsu, C. Chang, and C. Lin. *A Practical Guide to Support Vector Classification*. Department of Computer Science and Information Engineering, National Taiwan University, Taipei 106, Taiwan, 2004. http://www.kernel-machines.org
last accessed date 02-2005.

[Jaggers and Gurr 1996] K. Jaggers and T. Gurr. *POLITY III: Regime Change and Political Authority, 1800-1994 [Computer file]*. Boulder, CO: Keith Jaggers/College Park, MD: Ted Rober Gurr [producers], 1995. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1996, 1996.

[Jain *et al.* 2000] A. Jain, R. Duin, and J. Mao. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–37, 2000.

[Jordan and Bishop 1996] M. Jordan and C. Bishop. Neural networks. *ACM Comput. Surv.*, 28(1):73–75, 1996.

[King and Zeng 2000] G. King and L. Zeng. Logistic Regression in Rare Events Data. *Political Analysis*, 9(1):1–27, 2000.

[Lagazio and Russett 2003] M. Lagazio and B. Russett. *Temporal Stability and Causal Complexity*, chapter A Neural Network Analysis of Militarised Disputes, 1885-1992: Temporal Stability and Causal Complexty, pages 28–62. University of Michigan Press, 2003.

[Lin *et al.* 2003] S. Lin, S. Patel, A. Duncan, and L. Goodwin. Using Decision Trees and Support Vector Machines to Classify Genes by Names. In *European Workshop on Data Mining and Text Mining for BioInformatics*, Dubrovnik, Croatia, 2003. held in conjunction with ECML/PKDD.

[Maoz and Abdolali 1989] Z. Maoz and A. Abdolali. Regime types and international conflict, 1816-1976. *Journal of Conflict Resolution*, 33:3–35, March 1989.

[Maoz and Russett 1993] Z. Maoz and B. Russett. Normative and Structural Causes of Democratice Peace, 1946-1986. *American Political Science Review*, 87(3):624–638, 1993.

[Marwala and Lagazio 2004] T. Marwala and M. Lagazio. Modelling and Controlling Interstate Conflict. In *IEEE International Joint Conference on Neural networks 2004*, Budapest, Hungary, July 2004.

[Müller *et al.* 2001] K. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf. An Introduction to Kernel-Based Learning Algorithms. *IEEE Transactions on Neural Networks*, 12(2), March 2001.

[Mitchell 1997] T. Mitchell. *Machine Learning*. McGraw-Hill, New York, 1997.

[Møller 1993] M.F. Møller. A scaled Conjugate Gradient Algorithm for Fast Supervised Learning. *Neural Networks*, 6(4):525–533, 1993.

[Morgenthau 1973] H. Morgenthau. *Politics Among Nations: The struggle for power and peace*. A.A. Knopf, New York, 5th edition, 1973.

[Mousseau 2000] M. Mousseau. Market Prosperity, Democratic Consolidation, and Democratic Peace. *Journal of Conflict Resolution*, 44(4):472–507, 2000.

[Neal 1992] R. Neal. Bayesian training of backpropagation networks by the hybrid Monte Carlo method. Technical Report CRG-TR-92-1, Department of Computer Science, University of Toronto, M5S 1A1, Ontario, Canada, April 1992. http://www.cs.toronto.edu/ radford/bbp.abstract.html.

[Oneal and Russett 1997] J. Oneal and B. Russett. The Classical Liberals were right: Democracy, Interdependence, and Conflict, 1950-1985. *International Studies Quarterly*, 41:267–294, 1997.

[Oneal and Russett 1999b] J. Oneal and B. Russett. The Kantian Peace: The Pacific Benefits of Democracy, Interdependence, and International Organization. *World Politics*, 52(1):1–37, October 1999b.

[Oneal and Russett 2000] J. Oneal and B. Russett. *Causes of Peace: Democracy, Interdependence, and International Organisations, 1885-1992*. New Haven, CT, October 2000. Paper presented at the annual meeting of the Peace Science Society (International).

[Oneal and Russett 2001] J. Oneal and B. Russett. Clear and Cean: The Fixed Effects of Liberal Peace. *International Organization*, 52(2):469–85, June 2001.

[Pires and Marwala 2004] M. Pires and T. Marwala. Option pricing using neural networks and support vector machines. In *IEEE Inter-*

national Conference on Systems, Man and Cybernetics, The Hague, Holland, 2004. IEEE Computer Society TCC.

[Provost and Fawcett 1997] F. Provost and T. Fawcett. Analysis and Visualization of Classifier Performance: Comparison under Imprecise Class and Cost Distributions. In *Knowledge Discovery and Data Mining*, pages 43–48, 1997.

[Ray 1990b] J. Ray. *Prisoners of war: Nation-states in the moder era*, chapter Friends as foes: International conflict and wars between formal allies. Lexington Books, Lexington, MA, 1990b.

[Ray 2003] J. Ray. Constructing Mulitvariate Analysis (of Dangerous Dyads). Technical report, Department of Political Science, Vanderbilt University, TN 37235, Nashville, Sweden, 2003. http://sitemason.vanderbilt.edu/files/g/gDf5Ty/CONSTRUCTING%20MULTIVARIATE%20ANALYSES3.pdf.

[Reed 2000] W. Reed. A Unified Statistical Model of Conflict Onset and Escalation. *American Journal of Political Science*, 44:84–93, 2000.

[Rumelhart *et al.* 1994] D. Rumelhart, B. Widrow, and M. Lehr. The basic ideas in neural networks. *Communications of the ACM*, 37(3):87–92, 1994.

[Russett and Oneal 2001] B. Russett and J. Oneal. *Triangulating peace: Democracy, Interdependence, and International Organizations*. W.W. Norton, New York, 2001.

[Schölkopf and Smola 2003] B. Schölkopf and A. Smola. A short introduction to learning with kernels. pages 41–64, 2003.

[Schrodt 1991] P. Schrodt. Prediction of Interstate Conflict Outcomes Using a Neural Network. *Social Science Computer Review*, 9(3):359–80, 1991.

[Small and Singer 1976] M. Small and J. Singer. The war-proneness of democratic regimes, 1816-1965. *Jerusalem Journal of International Relations*, 1(4):50–68, 1976.

[Tucker 1997] R. Tucker. Dyad-Hard: The Interstate Dyad-Year Dataset Creator. *Political Methodologist*, 8(1):28–9, 1997.

[Vapnik 1982] V. Vapnik. *Estimation of Dependencies Based on Empirical Data*. Springer, New York, 1982.

[Vapnik 1995] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer Verlag, New York, 1995.

[Vidyasagar 1997] M. Vidyasagar. *A Theory of Learning and Generalization*. Springer, London, 1997.

[Westin 2001] L. Westin. Receiver operating characteristic (ROC) analysis: Evaluating discriminance effects among decision support systems. Technical Report UMINF 01.18, ISSN-0348-0542, Department of Computer Science, Umea University, SE-90187, Umea, Sweden, 2001. http://www.cs.umu.se/research/reports/2001/018/part1.pdf.

[Zeng 1999] L. Zeng. Prediction and Classification With Neural Network Models. *Sociological Methods & Research*, 27(4):499–524, May 1999.

[Zweig and Campbell 1993] M. Zweig and G. Campbell. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 39(4):561–577, April 1993.

Appendix 1

Conference Paper

A summary of the paper has been submitted to the IEEE International Joint Conference on Neural Networks, July 31-August 4, 2005, Montreal, Canada. See below for the full reference.

Eyasu A. Habtemariam, Tshilidzi Marwala and M. Lagazio. **Artificial Intelligence for Conflict Management** *International Joint Conference on Neural Networks*, Montreal, Canada, 2005.