The South African Institute of Computer Science and Information Technology

The 1997 National Research and Development Conference

Riverside Sun
Vanderbijlpark
13 & 14 November

Hosted by

Potchefstroomse Universiteit
vir Christelike Hoër Onderwys

The Department of Computer Science and Information Systems
Potchefstroom University for Christian Higher Education
Vaal Triangle Campus

PROCEEDINGS

Edited by L.M. Venter & R.R. Lombard
The South African Institute of Computer Science and Information Technology

Proceedings of the
The 1997 National Research and Development Conference
Towards 2000

Riverside Sun
Vanderbijlpark
13 & 14 November

Edited by
L.M. Venter
R.R. Lombard
©1997 Copyrights reside with the original authors who may be contacted directly

**ISBN 1-86822-300-0**

Printed and Binded by Xerox Printers, Potchefstroom

The views expressed in this book are those of the individual authors
Foreword

This book contains a collection of papers presented at a Research and Development conference of the South African Institute of Computer Scientists and Information Technologists (SAICSIT). The conference was held on 13 & 14 November 1997 at the Riverside Sun, Vanderbijlpark. Most of the organization for the conference was done by the Department of Computer Science and Information Technology of the Vaal Triangle Campus, Potchefstroom University for Christian Higher Education.

The programming committee accepted a wide selection of papers for the conference. The papers range from detailed technical research work to reports of work in progress. The papers originate mainly from Academia, but also describe work done in and for Industry. It is hoped that the papers give a true reflection of the current research scene in Computer Science and Information Technology in South Africa. Since one of the aims of the conference is Research development, the papers were not subjected to a refereeing process.

A number of people spent numerous hours helping with the organization of this conference. In this regard, we wish to thank the members of the Organizing committee, and the Programming committee who had very little time to screen the abstracts and compile the program. A special thanks goes to the secretary of the department, Mrs Helei Jooste, whose very able work was interrupted by the birth of her first child.
Organizing Committee

Conference General Chairs
Prof. J.M. Hattingh (PU for CHE)

Organizing Chair
Prof. Lucas Venter (PU for CHE)

Organizing Committee
Mrs. S. Gilliland
Mr. J.P. Jooste
Mr. R.R. Lombard
Mrs. M. Huisman

Secretariat
Mrs. H. Jooste

Program Chair
Prof. A de Waal (PU for CHE)

Program Committee
Prof. D. Kourie (UP)
Prof. C. Bornman (UNISA)
Prof. L.M. Venter (PU for CHE)
# Table of Contents

- Foreword i
- Organizing Committee ii
- List of Contributors vii

**Software Objects Change: Problems and Solution**  
S.A. Ajila  
1

**Liming-like Curve Constructions**  
M.L. Baart and R. McLeod  
26

**A Model for Evaluating Information Security**  
L. Barnard and R. von Solms  
27

**Integrating Spatial Data Management and Object Store Technology**  
S. Berman, S. Buffler and E. Voges  
31

**Metamodelling in Automated Software Engineering**  
S. Berman and R. Figueira  
32

**Using Multimedia Technology for Social Upliftment in Deprived Communities of Southern Africa**  
L. Bester and E. de Preez  
33

**Extending the Client-Server Model for Web-based Execution of Applications**  
L. Botha, J.M. Bishop and N.B. Serbedzija  
36

**Access Control Needs in an Electronic Workflow Environment**  
R.A. Botha  
45

**The Use of the Internet in an Academic Environment to Commercially Supply and Support Software Products**  
B. Braude and A.J. Walker  
51

**Explanation Facilities in Expert Systems Using Hypertext Technology**  
T. Breetzke and T. Thomas  
63

**Theoretical Computer Science: What is it all about, and is it of any relevance to us?**  
C. Brink  
75

**Representing Quadrics on a Computer**  
M.A. Coetzee and M.L. Baart  
76
The Generation of Pre-Interpretations for Detecting Unsolvable Planning Problems
D.A. de Waal, M. Denecker, M. Bruynooghe and M. Thielscher

The Emerging Role of the Chief Information Officer in South Africa
B. Dekenah

A Java-Implemented Remote Respiratory Disease Diagnosis System on a High Bandwidth Network
A. Foster

Early Results of a Comparative Evaluation of ISO 9001 and ISO/IEC 15504 Assessment Methods Applied to a Software Project
C. Gee and A.J. Walker

A Neural Network Model of a Fluidised Bed
M. Hajek

The Effects of Virtual Banking on the South African Banking Industry
M.L. Hart and M. Dunley-Owen

Linear Response Surface Analysis and Some Applications
J.M. Hattingh

Model Checking Software with Symbolic Trajectory Evaluation
A. Hazelhurst

A Risk Model to Allocate Resources to Different Computerized Systems
H.A. Kruger and J.M. Hattingh

Returns on the Stock Exchange
J.W. Kruger

Cardinality Constrained 0-1 Knapsack Problems
M.F. Kruger, J.M. Hattingh and T. Steyn

An Investigation in Software Process Improvement in the Software Development of a large Electricity Utility
M. Lang and A.J. Walker

Design and Implementation of a C++ Package for Two-Dimensional Numerical Integration
D.P. Laurie, L Pluym and Ronald Cools

Algebraic Factorization of Integers Using BDE's
H. Messerschmidt and J. Robertson
Global Optimization of Routes after the Process of Recovery
M. Mphahlele and J. Roos

Using a Lattice to Enhance Adaptation Guided Retrieval in Example Based Machine Translation
G.D. Oosthuizen and S.L. Serutla

Information Systems Development and Multi Criteria Decision Making / Systems Thinking
D. Petkov, O. Petkova

The Development of a Tutoring System to Assist Students to Develop Answering Techniques
N Pillay

Combining Rule-Based Artificial Intelligence with Geographic Information Systems to Plan the Physical Layer of Wireless Networks in Greenfield Areas
K. Prag, P. Premjeeth and K. Sandrasegaran

A Distributed Approach to the Scheduling Problem
V. Ram and P. Warren

More readings than I thought : Quantifier Interaction in Analysing the Temporal Structure of Repeated Eventualities
S. Rock

Ray Guarding Configuration of Adjacent Rectangles
I. Sanders, D. Lubinsky and M. Sears

Developing Soft Skills in Computer Students
C Schröder, T. Thomas

Information Security Awareness, a Must for Every Organization
M. Thomson and R. von Solms

Pla Va: A Lightweight Persistent Java Virtual Machine
S Tjasink and S. Berman

Beliefs on Resource-Bounded Agent
E. Viljoen

Object-Orientated Business Modelling and Re-engineering
M. Watzenboeck
On Indexing in Case Based Reasoning Applied to Pre-Transportation Decision Making for Hazardous Waste Handling
K.L. Wortmann, D. Petkov and E. Senior

Author Index
List of Contributors

S.A. Ajila
Department of Mathematics and Computer Science
National University of Lesotho
Roma, 180
Lesotho

L. Baart
Department of Mathematics
Vaal Triangle Campus of the PU for CHE
PO Box 1174
Vanderbijlpark, 1900

L. Barnard
Faculty of Computer Studies
Port Elizabeth Technikon
Private Bag X6011
Port Elizabeth, 6000

S. Berman
University of Cape Town
Rondebosch, 7701

L. Bester
Faculty of Computer Studies
Port Elizabeth Technikon
Private Bag X6011
Port Elizabeth, 6000

J.M. Bishop
Computer Science Department
University of Pretoria
Pretoria, 0002

L. Botha
Computer Science Department
University of Pretoria
Pretoria, 0002

R.A. Botha
Faculty of Computer Studies
Port Elizabeth Technikon
Private Bag X6011
Port Elizabeth, 6000

B. Braude
Software Engineering Applications Laboratory,
Electrical Engineering
University of the Witwatersrand
Private Bag 3
Wits, 2050

T. Breetzke
Faculty of Computer Studies
Port Elizabeth Technikon
Private Bag X6011
Port Elizabeth, 6000

C. Brink
University of Cape Town
Rondebosch, 7700

M. Bruynooghe
Department Computerwetenschappen
Katholieke Universiteit Leuven
Celestijnenlaan 200A
B-3001 Heverlee
Belgium

S. Buffler
University of Cape Town
Rondebosch, 7701

M.A. Coetzee
Department of Mathematics
PU for CHE
Private Bag X6001
Potchefstroom, 2520

R. Cools
Katholieke Universiteit Leuven
Celestijnenlaan 200A
B-3001 Heverlee
Belgium

E. de Preez
Faculty of Computer Studies
Port Elizabeth Technikon
Private Bag X6011
Port Elizabeth, 6000

D.A. De Waal
Department of Computer Science and
Information Systems
PU for CHE
Private Bag X6001
Potchefstroom, 2531

B. Dekennah
The Board of Executors

M. Denecker
Department Computerwetenschappen
Katholieke Universiteit Leuven
Celestijnenlaan 200A
B-3001 Heverlee
Belgium

M. Dunley-Owen
Department of Information Systems
University of Cape Town
Rondebosch, 7700

R. Fiqueira
University of Cape Town
Rondebosch, 7701

A. Foster
Department of Computer Science
University of Cape Town
Rondebosch, 7701

C. Gee
Software Engineering Applications Laboratory,
Electrical Engineering
University of the Witwatersrand
Private Bag 3
Wits 2050
K. Sandrasegaran  
Department of electrical Engineering  
University of Durban-Westville  
Private Bag X54001  
Durban, 4000

C. Schoder  
Faculty of Computer Studies  
Port Elizabeth Technikon  
Private Bag X6011  
Port Elizabeth, 6600

M. Sears  
Department of Mathematics  
University of the Witwatersrand  
Private Bag 3  
Wits, 2050

E. Senior  
International Center for Waste Technology  
University of Natal, Pietermaritzburg  
Private Bag X01  
Scotsville, 3209

N.B. Serbedzija  
GMD FIRST  
Rudower Chaussee 5  
D-12489 Berlin  
Germany

S.L. Serutla  
Department of Computer Science  
The University of Pretoria  
Pretoria, 0002

T. Steyn  
PU for CHE  
Private Bag X6001  
Potchefstroom, 2520

M. Thielscher  
Fachgebiet Intellektik, Fachgebiet Informatik  
Technische Hochschule Darmstadt  
Alexanderstrasse 10  
D-64283 Darmstadt  
Germany

T. Thomas  
Faculty of Computer Studies  
Port Elizabeth Technikon  
Private Bag X6011  
Port Elizabeth, 6000

M. Thomasy  
Faculty of Computer Studies  
Port Elizabeth Technikon  
Private Bag X6011  
Port Elizabeth, 6000

S. Tjasink  
University of Cape Town  
Rondebosch, 7700

E. Viljoen  
Department of Computer Science and  
Information Systems  
University of South Africa  
PO Box 392  
Pretoria, 0001

E. Voges  
University of Cape Town  
Rondebosch, 7701
Returns on the Stock Exchange

J. W. Kruger
jkruger@concave.cs.wits.ac.za

Abstract

The McGregor database supplies the accounting information of companies listed on the Johannesburg stock exchange. These attributes are not significantly correlated to the Return.

The Pearl's algorithm in Bayesian belief networks induces a belief network from data. With a solid grounding in probability theory, the Pearl algorithm allows belief updating by propagating likelihoods of leaf nodes (variables) and the prior probabilities.

The Pearl algorithm was originally developed for binary variables and my research was a generalization to more states.

The data used to test this new method, in a Portfolio Management context, are the Return and various attributes of companies listed on the Johannesburg Stock Exchange (JSE).

The results of this model was then compared to a Linear Regression model. The Bayesian method performed much better than stepwise Linear Regression.

1 Datamining compared to Statistics

Statistics (noun). The mathematics of the collection, organization, and interpretation of numerical data, especially the analysis of population characteristics by the inference from sampling (American Heritage Dictionary).

In Statistics a population is defined, then a sample is collected to make inferences about the population. This means that data cannot be re-used. They define a model before looking at the data. (If you look at the data beforehand, the significance is affected [ Glymour 1997 ]).

Datamining does not attempt generalizations to a population. The database is considered as the population. With the computing power of modern computers can use the whole database, making sampling redundant. Data can be re-used (the data in the database are used over and over). In datamining it is a common practice to try hundreds of models and find the one that fits best. This makes the interpretation of the significance difficult. [Glymour 1997].

Machine Learning is the datamining equivalent to Regression. (Pattern Recognition is the Engineering equivalent). In Machine Learning we use a training set to train the system to find the dependent variable.

Bayesian Statistics use prior knowledge to improve estimates. Bayesian Belief nets claim to give information about the causes [Pearl 1987] and in the Statistical literature latent structure is sometimes equated to causal structure [Lazarfeld 1966].
In the behavioral sciences, epidemiology, economics, market research, engineering, and even applied physics statistical methods are routinely used to justify causal inferences from data not obtained from randomized experiments [Spirtes 1993]. The ideas of causal structure is controversial and interested readers are referred to [Spirtes 1993], [Pearl 1995], and [Humphries 1995].

In summary datamining changes data into information, while Statistics makes inferences to a population.

Note: Datamining and datawarehousing are two different concepts.

2 Graphical language

Bayesian probabilities can be represented graphically [Buntine 1996]. The Markov principle allows a probability to be factorised.

3.1 Pearl’s Causal Structure Algorithm

The Causal Structure is the belief network that is induced by the data. It identifies the variables that cause other variables. Before evidence can be propagated through the network, as discussed in the previous Section, the directed acylical graph (DAG) has to be found.

From correlation analysis we know that if \( p(a, c) = p(a)p(c) \) then \( ac = (ab)(bc) \). This means that node \( b \) lies between nodes \( a \) and \( c \).

With three nodes (the hidden node) \( ad = (ab)(bd) \), \( ac = (ab)(bc) \) and \( cd = (cb)(bd) \).

If these equations do not hold, the variables are not star decomposable.

Solving by substitution gives:

\[ bd_2 = (ad)(cd)/(ac), \]
\[ ab_2 = (ac)(ad)/(cd) \] and \( bc_2 = (ac)(de)/(ad) \).

To link up a fourth node, \( e \), in a star formation, and if the node links on to:

\[ arc-ba \rightarrow (ac)(de) = (ad)(ce), \]
\[ arc-bc \rightarrow (ac)(de) = (cd)(ae) \] and \( arc-bd \rightarrow (ad)(ce) = (cd)(ae) \).

By following this reasoning any new node, that can be presented in a star formation, can be linked onto the belief network.

3.3 Number of variables needed to solve the parameters

The [Pearl 1987] algorithm was developed for binary variables. [Kruger 1997] developed a heuristic to handle four states.

4.1 Pearl’s Algorithm for evidence propagation

In a tree structure:

The belief of \( x \), \( \text{BEL}(x) \), is defined as: \( P(x|e) \), where \( e \) is the combined effect of all instantiated variables.

\( \text{BEL}(x) \) is a vector, where each component represents the belief that \( x \) is in that state. For instance, the belief that the weather tomorrow will be \( x_1 = \text{sunny}, x_2 = \text{rainy} \).
cloudy, x3 = raining, can be represented by the vector, \( \text{BEL}(x) = (\text{BEL}(x_1), \text{BEL}(x_2), \text{BEL}(x_3)) \), where \( x \) is the vector \((x_1, x_2, x_3)\).

The likelihood vector, or diagnostic support, \( ((x)) \), is \( P(e \mid x) \). Where \( e \) is the evidence. In other words, the likelihood vector is the probability that the evidence will occur given the state of the vector \( x \).

4.2 Consider a node \( x \), with \( m \) children \( y_1, y_2, \ldots, y_m \), and parent \( u \):

In a tree structure, a typical node \( x \), has \( m \) children \( y_1, y_2, \ldots, y_m \), and parent \( u \).

Call the causal support \( ((x)) \), and diagnostic support \( ((x)) \), the belief distribution of \( x \) is:

\[
\text{BEL}(x) = (P(e^+ \mid x), P(x \; e^-)) = ((x)) \; ((x)).
\]

\[
((x)) = (P(e^- \mid x) = P(\text{evidence from the children \( x \)}) = \text{E} \text{MBED \ Equation.2}
\]

\[
\text{EMBED \ Equation.2} \quad ((y_i \mid x)) = \text{E} \text{MBED \ Equation.2} \quad \text{EMBED \ Equation.2}
\]

because \( x \) separates its children and the siblings are conditionally independent.

\[
((x)) = \text{E} \text{MBED \ Equation.2} \quad ((x)) \quad P(x \; u) (x(u)) = ((u)) \; P(x \; u)
\]

and \((x)\) is a normalizing constant to make \( \text{EMBED \ Equation.2} \quad \text{EMBED \ Equation.2} \quad \text{BEL}(x) = 1 \).

Bottom-up propagation:

Node \( x \) uses the \( ((x)) \) message from the children to compute a new message \( (x(u)) \) to send to its parent \( u \).

\[
(x(u)) = (x) \quad P(x \; u)
\]

Top-down propagation:

The new \( (\text{message sent to by node } x \text{ to its } j\text{-th child } y_j \) is:

\[
((y_j)) = ((x)) \quad \text{EMBED \ Equation.2} \quad (y_k \mid x)
\]

but,

\[
\text{BEL}(x) = ((x)) \quad ((x)), \text{ therefore } ((y_j)) = (\text{BEL}(x) / ((y_j))
\]

Readers interested in Belief propagation in more general networks are referred to [Pearl 1987].

From the above it is evident that we only need the prior probabilities of the root and the likelihood functions of the leaf nodes to calculate the belief of all the nodes.

5 Evaluating the belief networks.

66 accounting attributes were discretesized into four states [Kruger 1997]. The Return was added as an additional variable.

To evaluate the belief nets, the likelihood of each state of the variables, except Return, is set (instantiated) for each of the companies and the belief of the Return is determined by propagating the likelihoods through the net. If the true state of Return is the same as the state with the maximum belief of Return then it is counted as a success.

6 Results from Stock exchange data
Use 240 companies to create, and select, the net. This leaves 117 companies to test the net. This section shows that this method holds, when the Belief net is created, and selected, with one set of data and tested with another set of data.

Seventeen Belief nets have an accuracy of more than 50%, when measured on the 240 companies. Testing the accuracy by looking at the remaining 117 companies, fourteen of the Belief nets still have an accuracy of more than 50%. The difference in accuracy between the 240 companies used to create the nets and the 117 companies used to test the nets are only marginal.

Six Belief nets have measured accuracy of more than 70% on the 240 companies, of these six only one dropped to below 70% when tested by the remaining 117 companies. This one Belief net dropped from 73% to 69.57%.

7 Linear Regression

To compare the modified Pearl algorithm to Linear Regression, the parameters are also estimated on the same 240 companies and then tested on the remaining 117 companies. Stepwise Linear Regression was done after the data had been converted to their respective states.

The accuracy for the Linear Regression model was measured as 0.346 on the 240 companies that was used to estimate the parameters. When the Linear Regression model was tested on the 117 remaining companies, the accuracy was only 0.188. The modified Pearl algorithm clearly outperforms Linear Regression.

8 Conclusions

If the states of variables can be generalized from binary to four states, then it can be further generalized to more states. The number of data items tended to limit further generalization in this case; Empty cells kept on cropping up, which caused singular matrices. In the binary case, with three variables, we have to solve nine parameters: Two for each of the link matrices and one for the prior probability of one of the states of each variable. For three states this is 24 parameters, and with four states the number of parameters goes up to 45. In general we have to solve \(3(n - 1)(n + 1)\) parameters, where \(n\) is the number of states for each variable. This is to get three variables onto a belief network. For every additional variable that must link onto a network a triplet of variables, two already in the network and the new variable, has to be calculated. Any error in a link matrix to a leaf node is further propagated when the link matrices between internal nodes are calculated.

The different states of the variables were used to get the prior probabilities and link matrices, while the actual values of the variables were used to create the structure and to get the correlations to the hidden variables. The correlation and causal structure seem in order, but the link matrices, calculated from the different states of the variables, tend to give problems. For further research better heuristics to find the link matrices will have to be developed. What is actually needed is a generalization to a continuous state scenario.

The Pearl algorithm clearly works for binary variables [Pearl 1987]. [Musick 1994], and [Heckerman 1997] approximate variables by estimating the probability density
function. This makes the prior probability a binary probability density function of a given variable. It is possible to extend this thinking to the Pearl algorithm. With this approach, link matrices, that have elements that are functions with a stochastic variable as argument, can be found.

Before the generalized Pearl algorithm can be promoted as a Forecasting technique, the method will have to be evaluated on different time periods. How well do the causal structures hold over different time periods? [Makridakis 1983], [Elton 1995].
References:


BUNTE, W. L., "Graphical Methods for discovering knowledge", in Fayyad: Advances in Knowledge Discovery and Data Mining 1996, AAI-Press, USA.


KRUGER, J. W., "Generalizing the number of states in Bayesian Belief propagation, as applied to Portfolio Management", Masters Research Report, University of the Witwatersrand, 1996.


